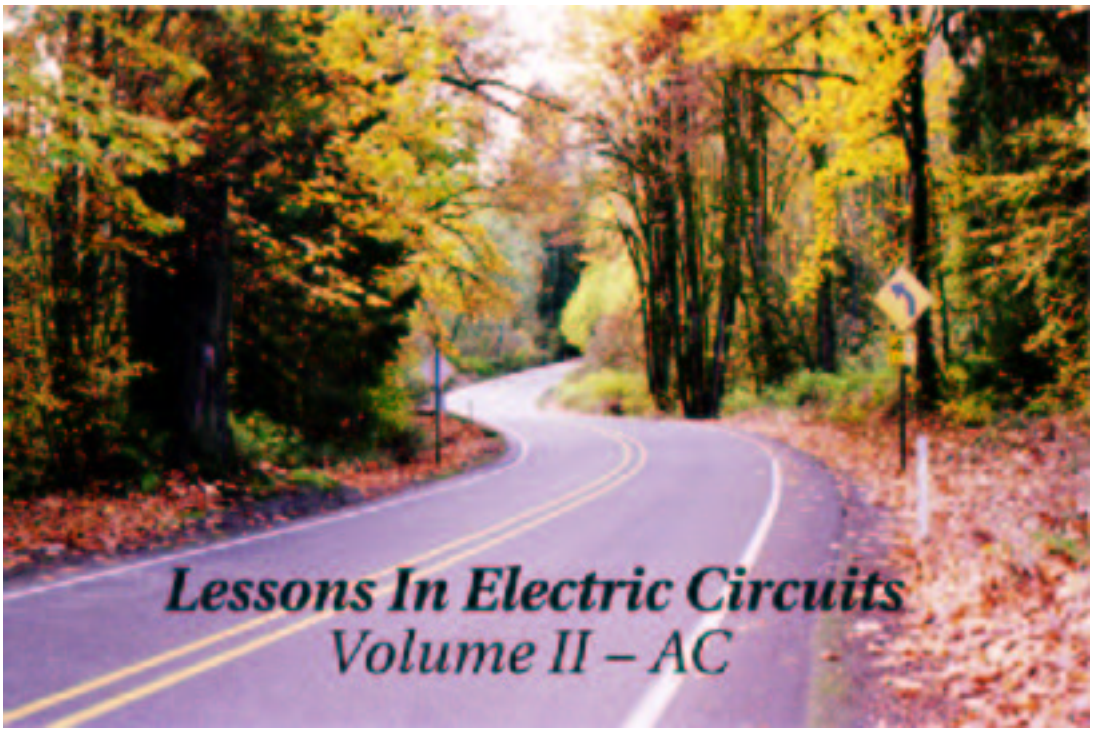


Fifth Edition, last update January 10, 2004

Lessons In Electric Circuits, Volume II – AC

By Tony R. Kuphaldt

Fifth Edition, last update January 10, 2004

© 2000-2004, Tony R. Kuphaldt

This book is published under the terms and conditions of the Design Science License. These terms and conditions allow for free copying, distribution, and/or modification of this document by the general public. The full Design Science License text is included in the last chapter.

As an open and collaboratively developed text, this book is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the Design Science License for more details.

Available in its entirety as part of the Open Book Project collection at `http://www.ibiblio.org/obp`

PRINTING HISTORY

- First Edition: Printed in June of 2000. Plain-ASCII illustrations for universal computer readability.
- Second Edition: Printed in September of 2000. Illustrations reworked in standard graphic (eps and jpeg) format. Source files translated to *Texinfo* format for easy online and printed publication.
- Third Edition: Equations and tables reworked as graphic images rather than plain-ASCII text.
- Fourth Edition: Printed in November 2001. Source files translated to *SubML* format. SubML is a simple markup language designed to easily convert to other markups like LaTeX, HTML, or DocBook using nothing but search-and-replace substitutions.
- Fifth Edition: Printed in November 2002. New sections added, and error corrections made, since the fourth edition.

Contents

1 BASIC AC THEORY **1**

1.1 What is alternating current (AC)? . . . 1
1.2 AC waveforms . . . 6
1.3 Measurements of AC magnitude . . . 12
1.4 Simple AC circuit calculations . . . 18
1.5 AC phase . . . 20
1.6 Principles of radio . . . 22
1.7 Contributors . . . 24

2 COMPLEX NUMBERS **27**

2.1 Introduction . . . 27
2.2 Vectors and AC waveforms . . . 30
2.3 Simple vector addition . . . 32
2.4 Complex vector addition . . . 34
2.5 Polar and rectangular notation . . . 35
2.6 Complex number arithmetic . . . 40
2.7 More on AC "polarity" . . . 42
2.8 Some examples with AC circuits . . . 47
2.9 Contributors . . . 53

3 REACTANCE AND IMPEDANCE – INDUCTIVE **55**

3.1 AC resistor circuits . . . 55
3.2 AC inductor circuits . . . 56
3.3 Series resistor-inductor circuits . . . 61
3.4 Parallel resistor-inductor circuits . . . 67
3.5 Inductor quirks . . . 70
3.6 More on the "skin effect" . . . 72
3.7 Contributors . . . 75

4 REACTANCE AND IMPEDANCE – CAPACITIVE **77**

4.1 AC resistor circuits . . . 77
4.2 AC capacitor circuits . . . 78
4.3 Series resistor-capacitor circuits . . . 82
4.4 Parallel resistor-capacitor circuits . . . 87

4.5 Capacitor quirks . . . 89
4.6 Contributors . . . 91

5 REACTANCE AND IMPEDANCE – R, L, AND C **93**
5.1 Review of R, X, and Z . . . 93
5.2 Series R, L, and C . . . 95
5.3 Parallel R, L, and C . . . 100
5.4 Series-parallel R, L, and C . . . 103
5.5 Susceptance and Admittance . . . 111
5.6 Summary . . . 112
5.7 Contributors . . . 113

6 RESONANCE **115**
6.1 An electric pendulum . . . 115
6.2 Simple parallel (tank circuit) resonance . . . 119
6.3 Simple series resonance . . . 123
6.4 Applications of resonance . . . 126
6.5 Resonance in series-parallel circuits . . . 128
6.6 Contributors . . . 136

7 MIXED-FREQUENCY AC SIGNALS **139**
7.1 Introduction . . . 139
7.2 Square wave signals . . . 144
7.3 Other waveshapes . . . 153
7.4 More on spectrum analysis . . . 159
7.5 Circuit effects . . . 170
7.6 Contributors . . . 174

8 FILTERS **175**
8.1 What is a filter? . . . 175
8.2 Low-pass filters . . . 176
8.3 High-pass filters . . . 181
8.4 Band-pass filters . . . 185
8.5 Band-stop filters . . . 188
8.6 Resonant filters . . . 190
8.7 Summary . . . 200
8.8 Contributors . . . 200

9 TRANSFORMERS **201**
9.1 Mutual inductance and basic operation . . . 201
9.2 Step-up and step-down transformers . . . 215
9.3 Electrical isolation . . . 220
9.4 Phasing . . . 222
9.5 Winding configurations . . . 226
9.6 Voltage regulation . . . 231
9.7 Special transformers and applications . . . 234

9.8 Practical considerations . . . 246
9.8.1 Power capacity . . . 246
9.8.2 Energy losses . . . 247
9.8.3 Stray capacitance and inductance . . . 249
9.8.4 Core saturation . . . 250
9.8.5 Inrush current . . . 252
9.8.6 Heat and Noise . . . 255
9.9 Contributors . . . 258

10 POLYPHASE AC CIRCUITS 259
10.1 Single-phase power systems . . . 259
10.2 Three-phase power systems . . . 265
10.3 Phase rotation . . . 271
10.4 Polyphase motor design . . . 275
10.5 Three-phase Y and Δ configurations . . . 281
10.6 Three-phase transformer circuits . . . 287
10.7 Harmonics in polyphase power systems . . . 293
10.8 Harmonic phase sequences . . . 315
10.9 Contributors . . . 317

11 POWER FACTOR 319
11.1 Power in resistive and reactive AC circuits . . . 319
11.2 True, Reactive, and Apparent power . . . 323
11.3 Calculating power factor . . . 326
11.4 Practical power factor correction . . . 330
11.5 Contributors . . . 334

12 AC METERING CIRCUITS 335
12.1 AC voltmeters and ammeters . . . 335
12.2 Frequency and phase measurement . . . 342
12.3 Power measurement . . . 350
12.4 Power quality measurement . . . 352
12.5 AC bridge circuits . . . 354
12.6 AC instrumentation transducers . . . 363
12.7 Contributors . . . 372

13 AC MOTORS 373

14 TRANSMISSION LINES 375
14.1 A 50-ohm cable? . . . 375
14.2 Circuits and the speed of light . . . 376
14.3 Characteristic impedance . . . 378
14.4 Finite-length transmission lines . . . 384
14.5 "Long" and "short" transmission lines . . . 390
14.6 Standing waves and resonance . . . 393
14.7 Impedance transformation . . . 412

14.8 Waveguides 419

15 ABOUT THIS BOOK **425**
15.1 Purpose 425
15.2 The use of SPICE 426
15.3 Acknowledgements 427

16 CONTRIBUTOR LIST **429**
16.1 How to contribute to this book 429
16.2 Credits 430
16.2.1 Tony R. Kuphaldt 430
16.2.2 Jason Starck 430
16.2.3 Your name here 431
16.2.4 Typo corrections and other “minor” contributions 431

17 DESIGN SCIENCE LICENSE **433**
17.1 0. Preamble 433
17.2 1. Definitions 433
17.3 2. Rights and copyright 434
17.4 3. Copying and distribution 434
17.5 4. Modification 435
17.6 5. No restrictions 435
17.7 6. Acceptance 435
17.8 7. No warranty 435
17.9 8. Disclaimer of liability 436

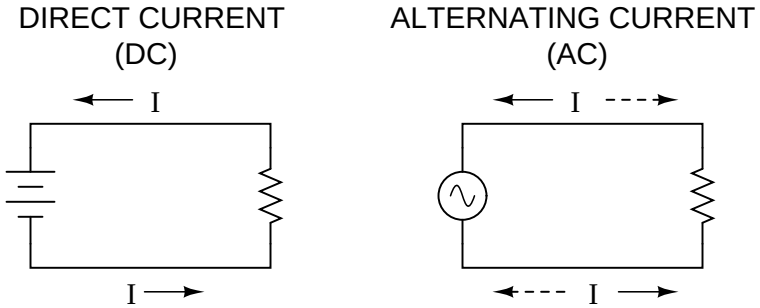
Chapter 1

BASIC AC THEORY

1.1 What is alternating current (AC)?

Most students of electricity begin their study with what is known as *direct current* (DC), which is electricity flowing in a constant direction, and/or possessing a voltage with constant polarity. DC is the kind of electricity made by a battery (with definite positive and negative terminals), or the kind of charge generated by rubbing certain types of materials against each other.

As useful and as easy to understand as DC is, it is not the only "kind" of electricity in use. Certain sources of electricity (most notably, rotary electro-mechanical generators) naturally produce voltages alternating in polarity, reversing positive and negative over time. Either as a voltage switching polarity or as a current switching direction back and forth, this "kind" of electricity is known as Alternating Current (AC):

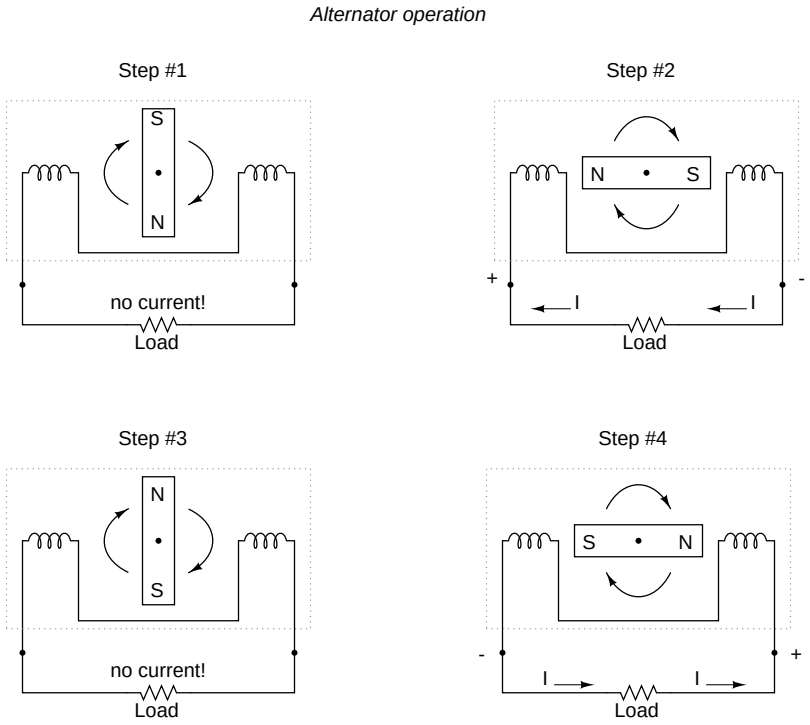


Whereas the familiar battery symbol is used as a generic symbol for any DC voltage source, the circle with the wavy line inside is the generic symbol for any AC voltage source.

One might wonder why anyone would bother with such a thing as AC. It is true that in some cases AC holds no practical advantage over DC. In applications where electricity is used to dissipate energy in the form of heat, the polarity or direction of current is irrelevant, so long as there is enough voltage and current to the load to produce the desired heat (power dissipation). However, with AC it is possible to build electric generators, motors and power distribution systems that are far more efficient than DC, and so we find AC used predominately across the world in high power applications. To explain the details of why this is so, a bit of background knowledge about AC is

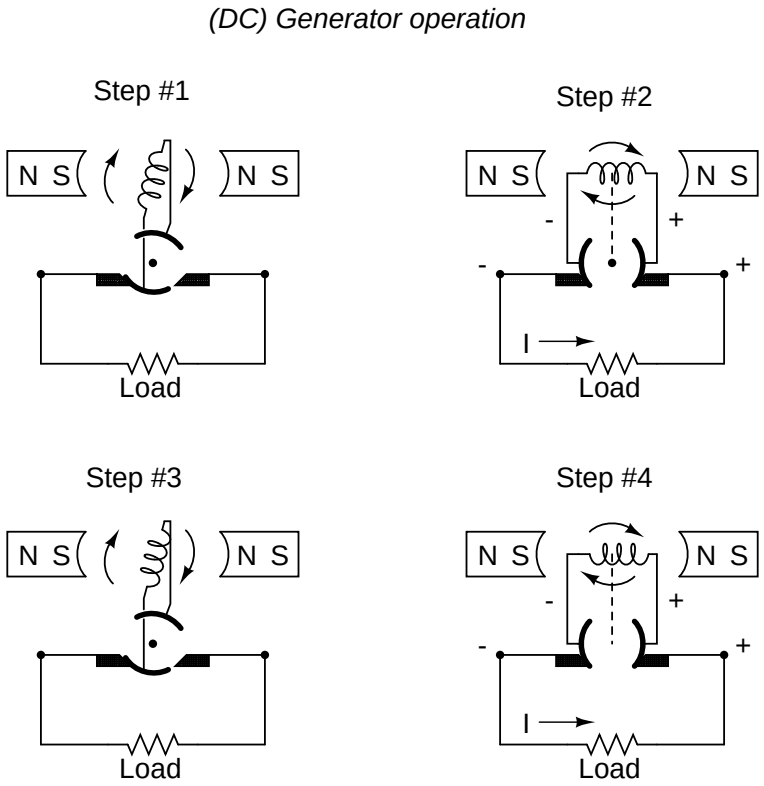
necessary.

If a machine is constructed to rotate a magnetic field around a set of stationary wire coils with the turning of a shaft, AC voltage will be produced across the wire coils as that shaft is rotated, in accordance with Faraday's Law of electromagnetic induction. This is the basic operating principle of an AC generator, also known as an *alternator*:



Notice how the polarity of the voltage across the wire coils reverses as the opposite poles of the rotating magnet pass by. Connected to a load, this reversing voltage polarity will create a reversing current direction in the circuit. The faster the alternator's shaft is turned, the faster the magnet will spin, resulting in an alternating voltage and current that switches directions more often in a given amount of time.

While DC generators work on the same general principle of electromagnetic induction, their construction is not as simple as their AC counterparts. With a DC generator, the coil of wire is mounted in the shaft where the magnet is on the AC alternator, and electrical connections are made to this spinning coil via stationary carbon "brushes" contacting copper strips on the rotating shaft. All this is necessary to switch the coil's changing output polarity to the external circuit so the external circuit sees a constant polarity:



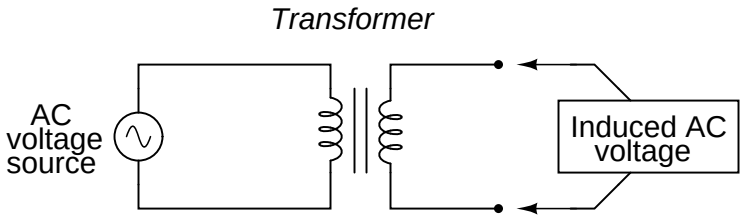
The generator shown above will produce two pulses of voltage per revolution of the shaft, both pulses in the same direction (polarity). In order for a DC generator to produce *constant* voltage, rather than brief pulses of voltage once every 1/2 revolution, there are multiple sets of coils making intermittent contact with the brushes. The diagram shown above is a bit more simplified than what you would see in real life.

The problems involved with making and breaking electrical contact with a moving coil should be obvious (sparking and heat), especially if the shaft of the generator is revolving at high speed. If the atmosphere surrounding the machine contains flammable or explosive vapors, the practical problems of spark-producing brush contacts are even greater. An AC generator (alternator) does not require brushes and commutators to work, and so is immune to these problems experienced by DC generators.

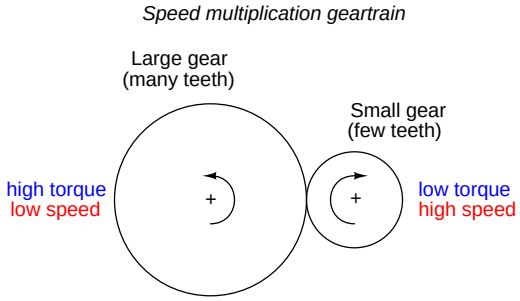
The benefits of AC over DC with regard to generator design is also reflected in electric motors. While DC motors require the use of brushes to make electrical contact with moving coils of wire, AC motors do not. In fact, AC and DC motor designs are very similar to their generator counterparts (identical for the sake of this tutorial), the AC motor being dependent upon the reversing magnetic field produced by alternating current through its stationary coils of wire to rotate the rotating magnet around on its shaft, and the DC motor being dependent on the brush contacts making and breaking connections to reverse current through the rotating coil every 1/2 rotation (180 degrees).

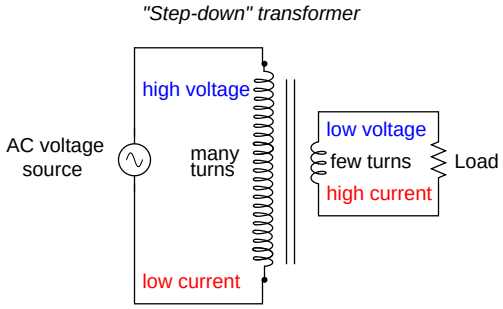
So we know that AC generators and AC motors tend to be simpler than DC generators and DC motors. This relative simplicity translates into greater reliability and lower cost of manufacture. But what else is AC good for? Surely there must be more to it than design details of generators and

motors! Indeed there is. There is an effect of electromagnetism known as *mutual induction*, whereby two or more coils of wire placed so that the changing magnetic field created by one induces a voltage in the other. If we have two mutually inductive coils and we energize one coil with AC, we will create an AC voltage in the other coil. When used as such, this device is known as a *transformer*:

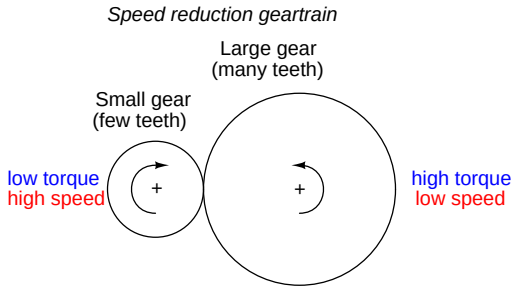


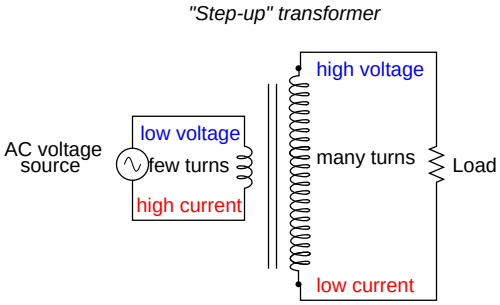
The fundamental significance of a transformer is its ability to step voltage up or down from the powered coil to the unpowered coil. The AC voltage induced in the unpowered ("secondary") coil is equal to the AC voltage across the powered ("primary") coil multiplied by the ratio of secondary coil turns to primary coil turns. If the secondary coil is powering a load, the current through the secondary coil is just the opposite: primary coil current multiplied by the ratio of primary to secondary turns. This relationship has a very close mechanical analogy, using torque and speed to represent voltage and current, respectively:



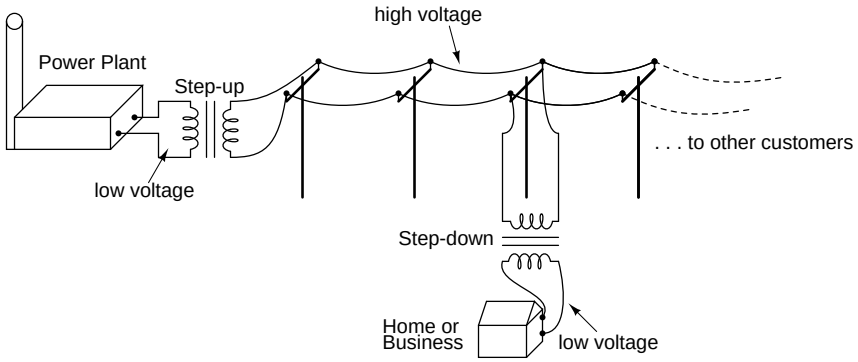


If the winding ratio is reversed so that the primary coil has less turns than the secondary coil, the transformer "steps up" the voltage from the source level to a higher level at the load:





The transformer's ability to step AC voltage up or down with ease gives AC an advantage unmatched by DC in the realm of power distribution. When transmitting electrical power over long distances, it is far more efficient to do so with stepped-up voltages and stepped-down currents (smaller-diameter wire with less resistive power losses), then step the voltage back down and the current back up for industry, business, or consumer use use.



Transformer technology has made long-range electric power distribution practical. Without the ability to efficiently step voltage up and down, it would be cost-prohibitive to construct power systems for anything but close-range (within a few miles at most) use.

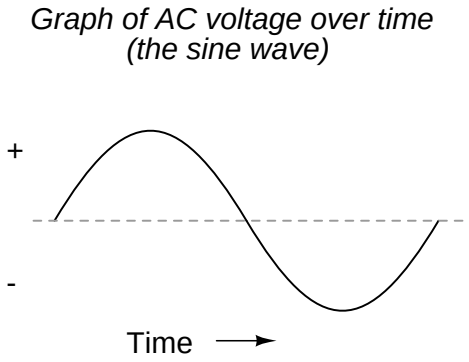
As useful as transformers are, they only work with AC, not DC. Because the phenomenon of mutual inductance relies on *changing* magnetic fields, and direct current (DC) can only produce steady magnetic fields, transformers simply will not work with direct current. Of course, direct

current may be interrupted (pulsed) through the primary winding of a transformer to create a changing magnetic field (as is done in automotive ignition systems to produce high-voltage spark plug power from a low-voltage DC battery), but pulsed DC is not that different from AC. Perhaps more than any other reason, this is why AC finds such widespread application in power systems.

- **REVIEW:**
- DC stands for "Direct Current," meaning voltage or current that maintains constant polarity or direction, respectively, over time.
- AC stands for "Alternating Current," meaning voltage or current that changes polarity or direction, respectively, over time.
- AC electromechanical generators, known as *alternators*, are of simpler construction than DC electromechanical generators.
- AC and DC motor design follows respective generator design principles very closely.
- A *transformer* is a pair of mutually-inductive coils used to convey AC power from one coil to the other. Often, the number of turns in each coil is set to create a voltage increase or decrease from the powered (primary) coil to the unpowered (secondary) coil.
- Secondary voltage = Primary voltage (secondary turns / primary turns)
- Secondary current = Primary current (primary turns / secondary turns)

1.2 AC waveforms

When an alternator produces AC voltage, the voltage switches polarity over time, but does so in a very particular manner. When graphed over time, the "wave" traced by this voltage of alternating polarity from an alternator takes on a distinct shape, known as a *sine wave*:

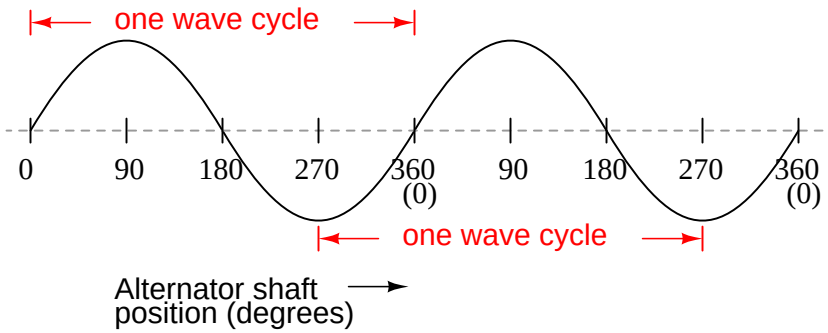


In the voltage plot from an electromechanical alternator, the change from one polarity to the other is a smooth one, the voltage level changing most rapidly at the zero ("crossover") point and most slowly at its peak. If we were to graph the trigonometric function of "sine" over a horizontal range of 0 to 360 degrees, we would find the exact same pattern:

```
Angle             Sine(angle)
in degrees
0  ..............  0.0000  -- zero
15 ..............  0.2588
30 ..............  0.5000
45 ..............  0.7071
60 ..............  0.8660
75 ..............  0.9659
90 ..............  1.0000  -- positive peak
105 .............  0.9659
120 .............  0.8660
135 .............  0.7071
150 .............  0.5000
165 .............  0.2588
180 .............  0.0000  -- zero
195 ............. -0.2588
210 ............. -0.5000
225 ............. -0.7071
240 ............. -0.8660
255 ............. -0.9659
270 ............. -1.0000  -- negative peak
285 ............. -0.9659
300 ............. -0.8660
315 ............. -0.7071
330 ............. -0.5000
345 ............. -0.2588
360 .............  0.0000  -- zero
```

The reason why an electromechanical alternator outputs sine-wave AC is due to the physics of its operation. The voltage produced by the stationary coils by the motion of the rotating magnet is proportional to the rate at which the magnetic flux is changing perpendicular to the coils (Faraday's Law of Electromagnetic Induction). That rate is greatest when the magnet poles are closest to the coils, and least when the magnet poles are furthest away from the coils. Mathematically, the rate of magnetic flux change due to a rotating magnet follows that of a sine function, so the voltage produced by the coils follows that same function.

If we were to follow the changing voltage produced by a coil in an alternator from any point on the sine wave graph to that point when the wave shape begins to repeat itself, we would have marked exactly one *cycle* of that wave. This is most easily shown by spanning the distance between identical peaks, but may be measured between any corresponding points on the graph. The degree marks on the horizontal axis of the graph represent the domain of the trigonometric sine function, and also the angular position of our simple two-pole alternator shaft as it rotates:



Since the horizontal axis of this graph can mark the passage of time as well as shaft position in degrees, the dimension marked for one cycle is often measured in a unit of time, most often seconds or fractions of a second. When expressed as a measurement, this is often called the *period* of a wave. The period of a wave in degrees is *always* 360, but the amount of time one period occupies depends on the rate voltage oscillates back and forth.

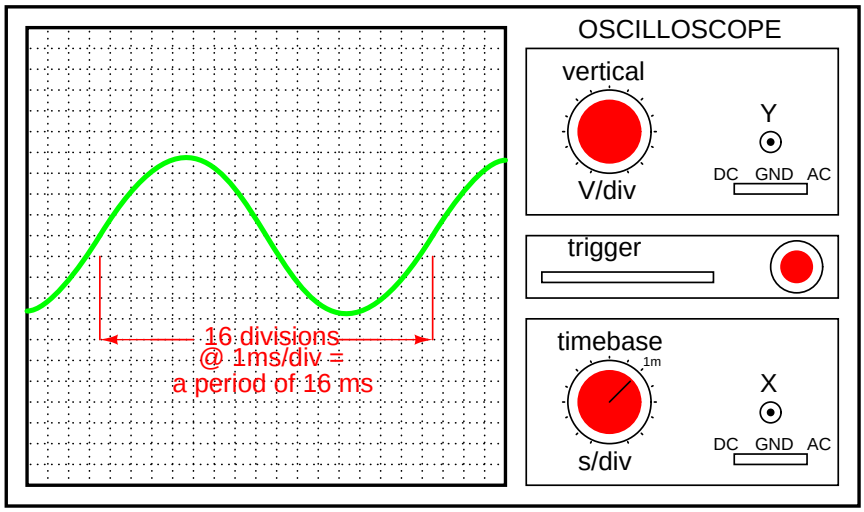
A more popular measure for describing the alternating rate of an AC voltage or current wave than *period* is the rate of that back-and-forth oscillation. This is called *frequency*. The modern unit for frequency is the Hertz (abbreviated Hz), which represents the number of wave cycles completed during one second of time. In the United States of America, the standard power-line frequency is 60 Hz, meaning that the AC voltage oscillates at a rate of 60 complete back-and-forth cycles every second. In Europe, where the power system frequency is 50 Hz, the AC voltage only completes 50 cycles every second. A radio station transmitter broadcasting at a frequency of 100 MHz generates an AC voltage oscillating at a rate of 100 *million* cycles every second.

Prior to the canonization of the Hertz unit, frequency was simply expressed as "cycles per second." Older meters and electronic equipment often bore frequency units of "CPS" (Cycles Per Second) instead of Hz. Many people believe the change from self-explanatory units like CPS to Hertz constitutes a step backward in clarity. A similar change occurred when the unit of "Celsius" replaced that of "Centigrade" for metric temperature measurement. The name Centigrade was based on a 100-count ("Centi-") scale ("-grade") representing the melting and boiling points of H_2O, respectively. The name Celsius, on the other hand, gives no hint as to the unit's origin or meaning.

Period and frequency are mathematical reciprocals of one another. That is to say, if a wave has a period of 10 seconds, its frequency will be 0.1 Hz, or 1/10 of a cycle per second:

$$\text{Frequency in Hertz} = \frac{1}{\text{Period in seconds}}$$

An instrument called an *oscilloscope* is used to display a changing voltage over time on a graphical screen. You may be familiar with the appearance of an *ECG* or *EKG* (electrocardiograph) machine, used by physicians to graph the oscillations of a patient's heart over time. The ECG is a special-purpose oscilloscope expressly designed for medical use. General-purpose oscilloscopes have the ability to display voltage from virtually any voltage source, plotted as a graph with time as the independent variable. The relationship between period and frequency is very useful to know when displaying an AC voltage or current waveform on an oscilloscope screen. By measuring the period of the wave on the horizontal axis of the oscilloscope screen and reciprocating that time value (in seconds), you can determine the frequency in Hertz.



$$\text{Frequency} = \frac{1}{\text{period}} = \frac{1}{16 \text{ ms}} = 62.5 \text{ Hz}$$

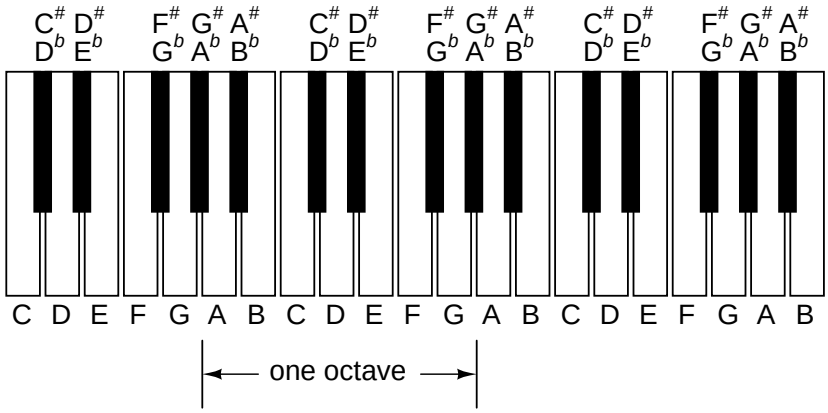
Voltage and current are by no means the only physical variables subject to variation over time. Much more common to our everyday experience is *sound*, which is nothing more than the alternating compression and decompression (pressure waves) of air molecules, interpreted by our ears as a physical sensation. Because alternating current is a wave phenomenon, it shares many of the properties of other wave phenomena, like sound. For this reason, sound (especially structured music) provides an excellent analogy for relating AC concepts.

In musical terms, frequency is equivalent to *pitch*. Low-pitch notes such as those produced by a tuba or bassoon consist of air molecule vibrations that are relatively slow (low frequency). High-pitch notes such as those produced by a flute or whistle consist of the same type of vibrations in the air, only vibrating at a much faster rate (higher frequency). Here is a table showing the actual frequencies for a range of common musical notes:

Note	Musical designation	Frequency (in hertz)
A	A_1	220.00
A sharp (or B flat)	$A^{\#}$ or B^{b}	233.08
B	B_1	246.94
C (middle)	C	261.63
C sharp (or D flat)	$C^{\#}$ or D^{b}	277.18
D	D	293.66
D sharp (or E flat)	$D^{\#}$ or E^{b}	311.13
E	E	329.63
F	F	349.23
F sharp (or G flat)	$F^{\#}$ or G^{b}	369.99
G	G	392.00
G sharp (or A flat)	$G^{\#}$ or A^{b}	415.30
A	A	440.00
A sharp (or B flat)	$A^{\#}$ or B^{b}	466.16
B	B	493.88
C	C^{1}	523.25

Astute observers will notice that all notes on the table bearing the same letter designation are related by a frequency ratio of 2:1. For example, the first frequency shown (designated with the letter "A") is 220 Hz. The next highest "A" note has a frequency of 440 Hz – exactly twice as many sound wave cycles per second. The same 2:1 ratio holds true for the first A sharp (233.08 Hz) and the next A sharp (466.16 Hz), and for all note pairs found in the table.

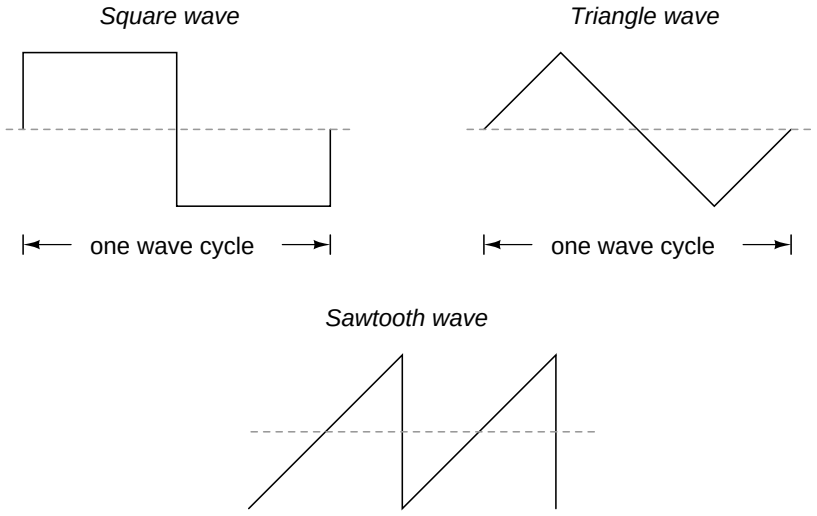
Audibly, two notes whose frequencies are exactly double each other sound remarkably similar. This similarity in sound is musically recognized, the shortest span on a musical scale separating such note pairs being called an *octave*. Following this rule, the next highest "A" note (one octave above 440 Hz) will be 880 Hz, the next lowest "A" (one octave below 220 Hz) will be 110 Hz. A view of a piano keyboard helps to put this scale into perspective:



As you can see, one octave is equal to *eight* white keys' worth of distance on a piano keyboard.

The familiar musical mnemonic (doe-ray-mee-fah-so-lah-tee-doe) – yes, the same pattern immortalized in the whimsical Rodgers and Hammerstein song sung in The Sound of Music – covers one octave from C to C.

While electromechanical alternators and many other physical phenomena naturally produce sine waves, this is not the only kind of alternating wave in existence. Other "waveforms" of AC are commonly produced within electronic circuitry. Here are but a few sample waveforms and their common designations:



These waveforms are by no means the only kinds of waveforms in existence. They're simply a few that are common enough to have been given distinct names. Even in circuits that are supposed to manifest "pure" sine, square, triangle, or sawtooth voltage/current waveforms, the real-life result is often a distorted version of the intended waveshape. Some waveforms are so complex that they defy classification as a particular "type" (including waveforms associated with many kinds of musical instruments). Generally speaking, any waveshape bearing close resemblance to a perfect sine wave is termed *sinusoidal*, anything different being labeled as *non-sinusoidal*. Being that the waveform of an AC voltage or current is crucial to its impact in a circuit, we need to be aware of the fact that AC waves come in a variety of shapes.

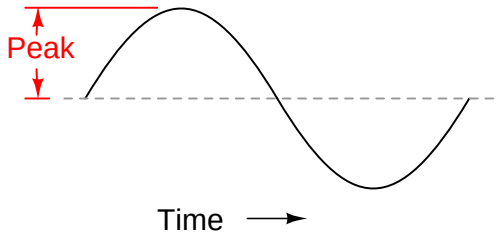
- **REVIEW:**
- AC produced by an electromechanical alternator follows the graphical shape of a sine wave.
- One *cycle* of a wave is one complete evolution of its shape until the point that it is ready to repeat itself.
- The *period* of a wave is the amount of time it takes to complete one cycle.
- *Frequency* is the number of complete cycles that a wave completes in a given amount of time. Usually measured in Hertz (Hz), 1 Hz being equal to one complete wave cycle per second.
- Frequency = 1/(period in seconds)

1.3 Measurements of AC magnitude

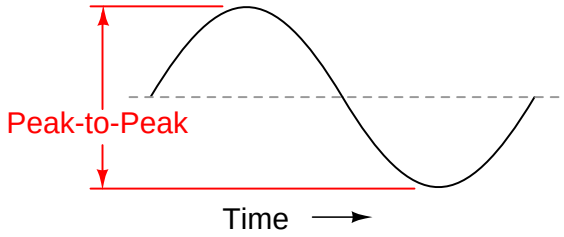
So far we know that AC voltage alternates in polarity and AC current alternates in direction. We also know that AC can alternate in a variety of different ways, and by tracing the alternation over time we can plot it as a "waveform." We can measure the rate of alternation by measuring the time it takes for a wave to evolve before it repeats itself (the "period"), and express this as cycles per unit time, or "frequency." In music, frequency is the same as *pitch*, which is the essential property distinguishing one note from another.

However, we encounter a measurement problem if we try to express how large or small an AC quantity is. With DC, where quantities of voltage and current are generally stable, we have little trouble expressing how much voltage or current we have in any part of a circuit. But how do you grant a single measurement of magnitude to something that is constantly changing?

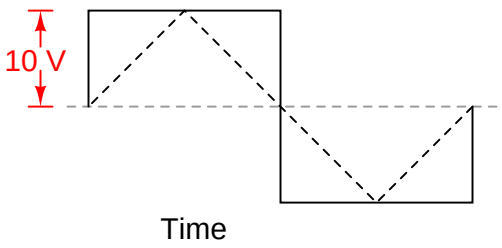
One way to express the intensity, or magnitude (also called the *amplitude*), of an AC quantity is to measure its peak height on a waveform graph. This is known as the *peak* or *crest* value of an AC waveform:

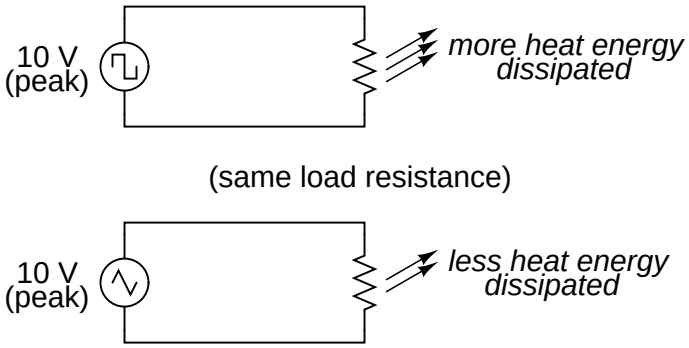


Another way is to measure the total height between opposite peaks. This is known as the *peak-to-peak* (P-P) value of an AC waveform:

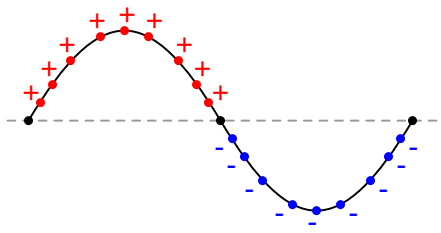


Unfortunately, either one of these expressions of waveform amplitude can be misleading when comparing two different types of waves. For example, a square wave peaking at 10 volts is obviously a greater amount of voltage for a greater amount of time than a triangle wave peaking at 10 volts. The effects of these two AC voltages powering a load would be quite different:



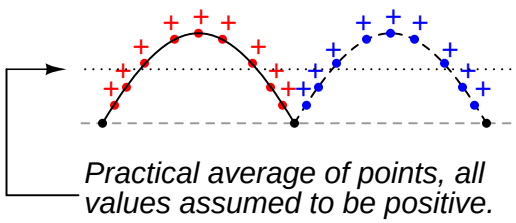


One way of expressing the amplitude of different waveshapes in a more equivalent fashion is to mathematically average the values of all the points on a waveform's graph to a single, aggregate number. This amplitude measure is known simply as the *average* value of the waveform. If we average all the points on the waveform algebraically (that is, to consider their *sign*, either positive or negative), the average value for most waveforms is technically zero, because all the positive points cancel out all the negative points over a full cycle:



True average value of all points (considering their signs) is ***zero****!*

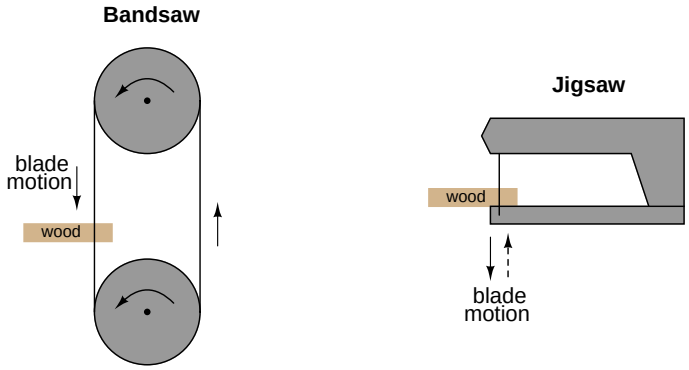
This, of course, will be true for any waveform having equal-area portions above and below the "zero" line of a plot. However, as a *practical* measure of a waveform's aggregate value, "average" is usually defined as the mathematical mean of all the points' *absolute values* over a cycle. In other words, we calculate the practical average value of the waveform by considering all points on the wave as positive quantities, as if the waveform looked like this:



Polarity-insensitive mechanical meter movements (meters designed to respond equally to the positive and negative half-cycles of an alternating voltage or current) register in proportion to the waveform's (practical) average value, because the inertia of the pointer against the tension of the spring naturally averages the force produced by the varying voltage/current values over time. Conversely, polarity-sensitive meter movements vibrate uselessly if exposed to AC voltage or current, their needles oscillating rapidly about the zero mark, indicating the true (algebraic) average value of zero for a symmetrical waveform. When the "average" value of a waveform is referenced in this text, it will be assumed that the "practical" definition of average is intended unless otherwise specified.

Another method of deriving an aggregate value for waveform amplitude is based on the waveform's ability to do useful work when applied to a load resistance. Unfortunately, an AC measurement based on work performed by a waveform is not the same as that waveform's "average" value, because the *power* dissipated by a given load (work performed per unit time) is not directly proportional to the magnitude of either the voltage or current impressed upon it. Rather, power is proportional to the *square* of the voltage or current applied to a resistance ($P = E^2/R$, and $P = I^2R$). Although the mathematics of such an amplitude measurement might not be straightforward, the utility of it is.

Consider a bandsaw and a jigsaw, two pieces of modern woodworking equipment. Both types of saws cut with a thin, toothed, motor-powered metal blade to cut wood. But while the bandsaw uses a continuous motion of the blade to cut, the jigsaw uses a back-and-forth motion. The comparison of alternating current (AC) to direct current (DC) may be likened to the comparison of these two saw types:

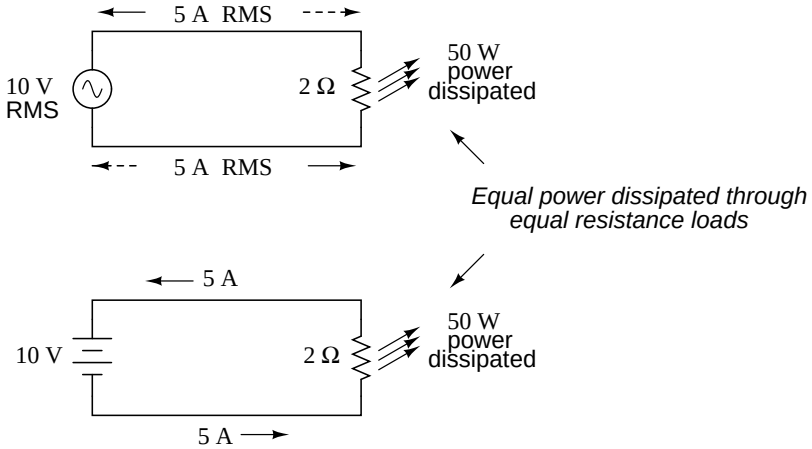


(analogous to DC) *(analogous to AC)*

The problem of trying to describe the changing quantities of AC voltage or current in a single, aggregate measurement is also present in this saw analogy: how might we express the speed of a jigsaw blade? A bandsaw blade moves with a constant speed, similar to the way DC voltage pushes

or DC current moves with a constant magnitude. A jigsaw blade, on the other hand, moves back and forth, its blade speed constantly changing. What is more, the back-and-forth motion of any two jigsaws may not be of the same type, depending on the mechanical design of the saws. One jigsaw might move its blade with a sine-wave motion, while another with a triangle-wave motion. To rate a jigsaw based on its *peak* blade speed would be quite misleading when comparing one jigsaw to another (or a jigsaw with a bandsaw!). Despite the fact that these different saws move their blades in different manners, they are equal in one respect: they all cut wood, and a quantitative comparison of this common function can serve as a common basis for which to rate blade speed.

Picture a jigsaw and bandsaw side-by-side, equipped with identical blades (same tooth pitch, angle, etc.), equally capable of cutting the same thickness of the same type of wood at the same rate. We might say that the two saws were equivalent or equal in their cutting capacity. Might this comparison be used to assign a "bandsaw equivalent" blade speed to the jigsaw's back-and-forth blade motion; to relate the wood-cutting effectiveness of one to the other? This is the general idea used to assign a "DC equivalent" measurement to any AC voltage or current: whatever magnitude of DC voltage or current would produce the same amount of heat energy dissipation through an equal resistance:



In the two circuits above, we have the same amount of load resistance (2 Ω) dissipating the same amount of power in the form of heat (50 watts), one powered by AC and the other by DC. Because the AC voltage source pictured above is equivalent (in terms of power delivered to a load) to a 10 volt DC battery, we would call this a "10 volt" AC source. More specifically, we would denote its voltage value as being 10 volts *RMS*. The qualifier "RMS" stands for *Root Mean Square*, the algorithm used to obtain the DC equivalent value from points on a graph (essentially, the procedure consists of squaring all the positive and negative points on a waveform graph, averaging those squared values, then taking the square root of that average to obtain the final answer). Sometimes the alternative terms *equivalent* or *DC equivalent* are used instead of "RMS," but the quantity and principle are both the same.

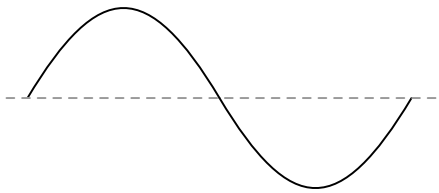
RMS amplitude measurement is the best way to relate AC quantities to DC quantities, or other AC quantities of differing waveform shapes, when dealing with measurements of electric power. For other considerations, peak or peak-to-peak measurements may be the best to employ. For instance, when determining the proper size of wire (ampacity) to conduct electric power from a source to a load, RMS current measurement is the best to use, because the principal concern with current

is overheating of the wire, which is a function of power dissipation caused by current through the resistance of the wire. However, when rating insulators for service in high-voltage AC applications, peak voltage measurements are the most appropriate, because the principal concern here is insulator "flashover" caused by brief spikes of voltage, irrespective of time.

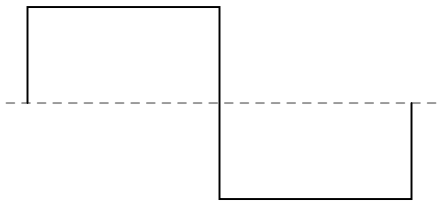
Peak and peak-to-peak measurements are best performed with an oscilloscope, which can capture the crests of the waveform with a high degree of accuracy due to the fast action of the cathode-ray-tube in response to changes in voltage. For RMS measurements, analog meter movements (D'Arsonval, Weston, iron vane, electrodynamometer) will work so long as they have been calibrated in RMS figures. Because the mechanical inertia and dampening effects of an electromechanical meter movement makes the deflection of the needle naturally proportional to the *average* value of the AC, not the true RMS value, analog meters must be specifically calibrated (or mis-calibrated, depending on how you look at it) to indicate voltage or current in RMS units. The accuracy of this calibration depends on an assumed waveshape, usually a sine wave.

Electronic meters specifically designed for RMS measurement are best for the task. Some instrument manufacturers have designed ingenious methods for determining the RMS value of any waveform. One such manufacturer produces "True-RMS" meters with a tiny resistive heating element powered by a voltage proportional to that being measured. The heating effect of that resistance element is measured thermally to give a true RMS value with no mathematical calculations whatsoever, just the laws of physics in action in fulfillment of the definition of RMS. The accuracy of this type of RMS measurement is independent of waveshape.

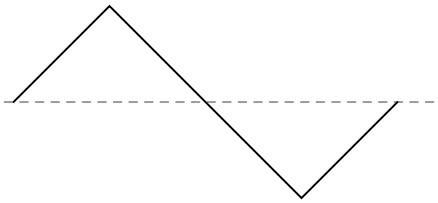
For "pure" waveforms, simple conversion coefficients exist for equating Peak, Peak-to-Peak, Average (practical, not algebraic), and RMS measurements to one another:



RMS = 0.707 (Peak)
AVG = 0.637 (Peak)
P-P = 2 (Peak)



RMS = Peak
AVG = Peak
P-P = 2 (Peak)

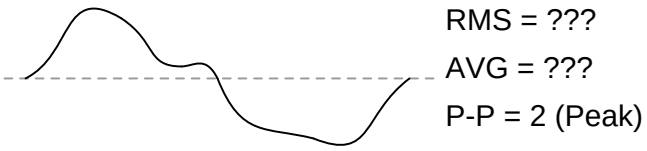


RMS = 0.577 (Peak)
AVG = 0.5 (Peak)
P-P = 2 (Peak)

In addition to RMS, average, peak (crest), and peak-to-peak measures of an AC waveform, there

are ratios expressing the proportionality between some of these fundamental measurements. The *crest factor* of an AC waveform, for instance, is the ratio of its peak (crest) value divided by its RMS value. The *form factor* of an AC waveform is the ratio of its peak value divided by its average value. Square-shaped waveforms always have crest and form factors equal to 1, since the peak is the same as the RMS and average values. Sinusoidal waveforms have crest factors of 1.414 (the square root of 2) and form factors of 1.571 ($\pi/2$). Triangle- and sawtooth-shaped waveforms have crest values of 1.732 (the square root of 3) and form factors of 2.

Bear in mind that the conversion constants shown here for peak, RMS, and average amplitudes of sine waves, square waves, and triangle waves hold true only for *pure* forms of these waveshapes. The RMS and average values of distorted waveshapes are not related by the same ratios:



This is a very important concept to understand when using an analog meter movement to measure AC voltage or current. An analog movement, calibrated to indicate sine-wave RMS amplitude, will only be accurate when measuring pure sine waves. If the waveform of the voltage or current being measured is anything but a pure sine wave, the indication given by the meter will not be the true RMS value of the waveform, because the degree of needle deflection in an analog meter movement is proportional to the *average* value of the waveform, not the RMS. RMS meter calibration is obtained by "skewing" the span of the meter so that it displays a small multiple of the average value, which will be equal to be the RMS value for a particular waveshape and *a particular waveshape only.*

Since the sine-wave shape is most common in electrical measurements, it is the waveshape assumed for analog meter calibration, and the small multiple used in the calibration of the meter is 1.1107 (the form factor $\pi/2$ divided by the crest factor 1.414: the ratio of RMS divided by average for a sinusoidal waveform). Any waveshape other than a pure sine wave will have a different ratio of RMS and average values, and thus a meter calibrated for sine-wave voltage or current will not indicate true RMS when reading a non-sinusoidal wave. Bear in mind that this limitation applies only to simple, analog AC meters not employing "True-RMS" technology.

- **REVIEW:**
- The *amplitude* of an AC waveform is its height as depicted on a graph over time. An amplitude measurement can take the form of peak, peak-to-peak, average, or RMS quantity.
- *Peak* amplitude is the height of an AC waveform as measured from the zero mark to the highest positive or lowest negative point on a graph. Also known as the *crest* amplitude of a wave.
- *Peak-to-peak* amplitude is the total height of an AC waveform as measured from maximum positive to maximum negative peaks on a graph. Often abbreviated as "P-P".
- *Average* amplitude is the mathematical "mean" of all a waveform's points over the period of one cycle. Technically, the average amplitude of any waveform with equal-area portions above and below the "zero" line on a graph is zero. However, as a practical measure of amplitude, a waveform's average value is often calculated as the mathematical mean of all the points'

absolute values (taking all the negative values and considering them as positive). For a sine wave, the average value so calculated is approximately 0.637 of its peak value.

- "RMS" stands for *Root Mean Square*, and is a way of expressing an AC quantity of voltage or current in terms functionally equivalent to DC. For example, 10 volts AC RMS is the amount of voltage that would produce the same amount of heat dissipation across a resistor of given value as a 10 volt DC power supply. Also known as the "equivalent" or "DC equivalent" value of an AC voltage or current. For a sine wave, the RMS value is approximately 0.707 of its peak value.

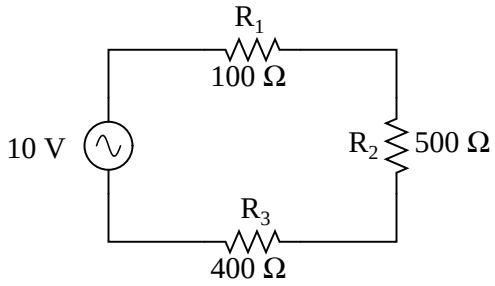
- The *crest factor* of an AC waveform is the ratio of its peak (crest) to its RMS value.

- The *form factor* of an AC waveform is the ratio of its peak (crest) value to its average value.

- Analog, electromechanical meter movements respond proportionally to the *average* value of an AC voltage or current. When RMS indication is desired, the meter's calibration must be "skewed" accordingly. This means that the accuracy of an electromechanical meter's RMS indication is dependent on the purity of the waveform: whether it is the exact same waveshape as the waveform used in calibrating.

1.4 Simple AC circuit calculations

Over the course of the next few chapters, you will learn that AC circuit measurements and calculations can get very complicated due to the complex nature of alternating current in circuits with inductance and capacitance. However, with simple circuits involving nothing more than an AC power source and resistance, the same laws and rules of DC apply simply and directly.



$$R_{total} = R_1 + R_2 + R_3$$

$$R_{total} = 1\ k\Omega$$

$$I_{total} = \frac{E_{total}}{R_{total}}$$

$$I_{total} = \frac{10\ V}{1\ k\Omega}$$

$$I_{total} = 10\ mA$$

$$E_{R1} = I_{total}R_1 \qquad E_{R2} = I_{total}R_2 \qquad E_{R3} = I_{total}R_3$$

$$E_{R1} = 1\ V \qquad E_{R2} = 5\ V \qquad E_{R3} = 4\ V$$

Series resistances still add, parallel resistances still diminish, and the Laws of Kirchhoff and Ohm still hold true. Actually, as we will discover later on, these rules and laws *always* hold true, it's just that we have to express the quantities of voltage, current, and opposition to current in more advanced mathematical forms. With purely resistive circuits, however, these complexities of AC are of no practical consequence, and so we can treat the numbers as though we were dealing with simple DC quantities.

Because all these mathematical relationships still hold true, we can make use of our familiar "table" method of organizing circuit values just as with DC:

	R_1	R_2	R_3	Total	
E	1	5	4	10	Volts
I	10m	10m	10m	10m	Amps
R	100	500	400	1k	Ohms

One major caveat needs to be given here: all measurements of AC voltage and current must be expressed in the same terms (peak, peak-to-peak, average, or RMS). If the source voltage is given in peak AC volts, then all currents and voltages subsequently calculated are cast in terms of peak units. If the source voltage is given in AC RMS volts, then all calculated currents and voltages are cast in AC RMS units as well. This holds true for *any* calculation based on Ohm's Laws, Kirchhoff's Laws, etc. Unless otherwise stated, all values of voltage and current in AC circuits are generally assumed to be RMS rather than peak, average, or peak-to-peak. In some areas of electronics, peak measurements are assumed, but in most applications (especially industrial electronics) the assumption is RMS.

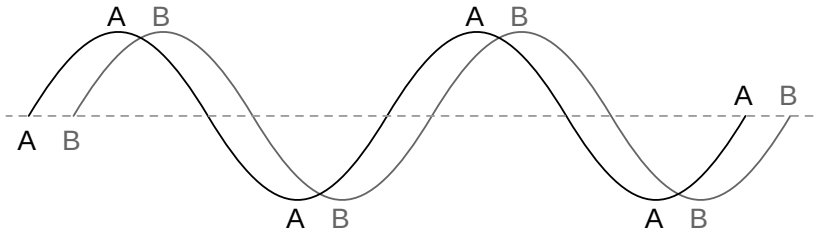
- **REVIEW:**
- All the old rules and laws of DC (Kirchhoff's Voltage and Current Laws, Ohm's Law) still hold

true for AC. However, with more complex circuits, we may need to represent the AC quantities in more complex form. More on this later, I promise!

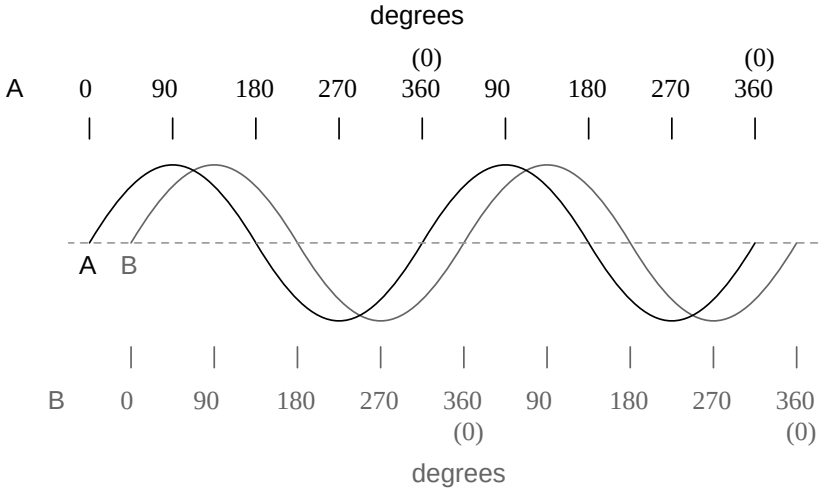
- The "table" method of organizing circuit values is still a valid analysis tool for AC circuits.

1.5 AC phase

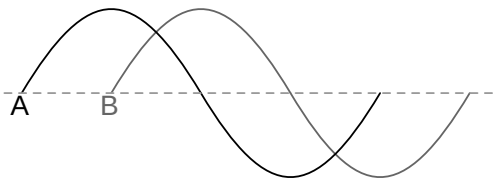
Things start to get complicated when we need to relate two or more AC voltages or currents that are out of step with each other. By "out of step," I mean that the two waveforms are not synchronized: that their peaks and zero points do not match up at the same points in time. The following graph illustrates an example of this:



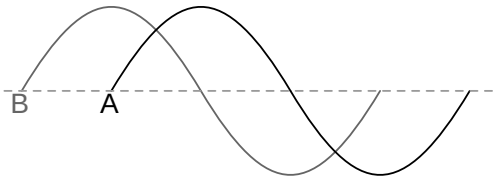
The two waves shown above (A versus B) are of the same amplitude and frequency, but they are out of step with each other. In technical terms, this is called a *phase shift*. Earlier we saw how we could plot a "sine wave" by calculating the trigonometric sine function for angles ranging from 0 to 360 degrees, a full circle. The starting point of a sine wave was zero amplitude at zero degrees, progressing to full positive amplitude at 90 degrees, zero at 180 degrees, full negative at 270 degrees, and back to the starting point of zero at 360 degrees. We can use this angle scale along the horizontal axis of our waveform plot to express just how far out of step one wave is with another:



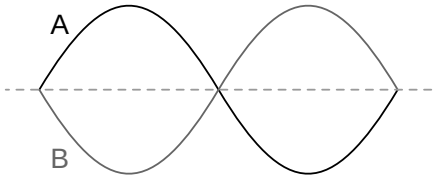
The shift between these two waveforms is about 45 degrees, the "A" wave being ahead of the "B" wave. A sampling of different phase shifts is given in the following graphs to better illustrate this concept:



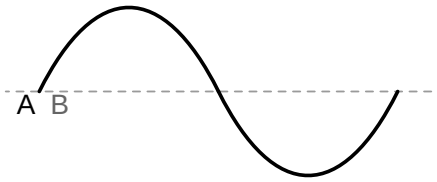
Phase shift = 90 degrees
A is ahead of B
(A "leads" B)



Phase shift = 90 degrees
B is ahead of A
(B "leads" A)



Phase shift = 180 degrees
A and B waveforms are
mirror-images of each other



Phase shift = 0 degrees
A and B waveforms are
in perfect step with each other

Because the waveforms in the above examples are at the same frequency, they will be out of step by the same angular amount at every point in time. For this reason, we can express phase shift for two or more waveforms of the same frequency as a constant quantity for the entire wave, and not just an expression of shift between any two particular points along the waves. That is, it is safe to say something like, "voltage 'A' is 45 degrees out of phase with voltage 'B'." Whichever waveform is ahead in its evolution is said to be *leading* and the one behind is said to be *lagging*.

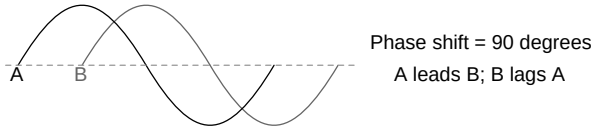
Phase shift, like voltage, is always a measurement relative between two things. There's really no such thing as a waveform with an *absolute* phase measurement because there's no known universal reference for phase. Typically in the analysis of AC circuits, the voltage waveform of the power supply is used as a reference for phase, that voltage stated as "xxx volts at 0 degrees." Any other AC voltage or current in that circuit will have its phase shift expressed in terms relative to that source voltage.

This is what makes AC circuit calculations more complicated than DC. When applying Ohm's Law and Kirchhoff's Laws, quantities of AC voltage and current must reflect phase shift as well as amplitude. Mathematical operations of addition, subtraction, multiplication, and division must operate on these quantities of phase shift as well as amplitude. Fortunately, there is a mathematical system of quantities called *complex numbers* ideally suited for this task of representing amplitude and phase.

Because the subject of complex numbers is so essential to the understanding of AC circuits, the

next chapter will be devoted to that subject alone.

- **REVIEW:**
- *Phase shift* is where two or more waveforms are out of step with each other.
- The amount of phase shift between two waves can be expressed in terms of degrees, as defined by the degree units on the horizontal axis of the waveform graph used in plotting the trigonometric sine function.
- A *leading* waveform is defined as one waveform that is ahead of another in its evolution. A *lagging* waveform is one that is behind another. Example:



- Calculations for AC circuit analysis must take into consideration both amplitude and phase shift of voltage and current waveforms to be completely accurate. This requires the use of a mathematical system called *complex numbers*.

1.6 Principles of radio

One of the more fascinating applications of electricity is in the generation of invisible ripples of energy called *radio waves*. The limited scope of this lesson on alternating current does not permit full exploration of the concept, some of the basic principles will be covered.

With Oersted's accidental discovery of electromagnetism, it was realized that electricity and magnetism were related to each other. When an electric current was passed through a conductor, a magnetic field was generated perpendicular to the axis of flow. Likewise, if a conductor was exposed to a change in magnetic flux perpendicular to the conductor, a voltage was produced along the length of that conductor. So far, scientists knew that electricity and magnetism always seemed to affect each other at right angles. However, a major discovery lay hidden just beneath this seemingly simple concept of related perpendicularity, and its unveiling was one of the pivotal moments in modern science.

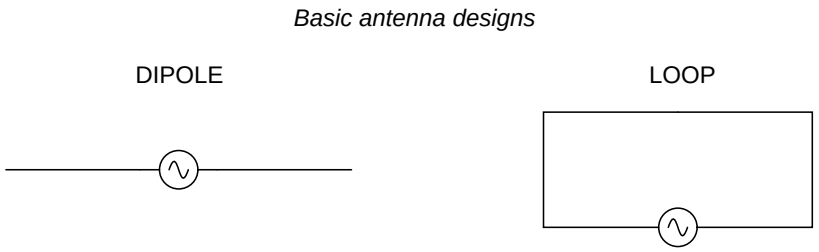
This breakthrough in physics is hard to overstate. The man responsible for this conceptual revolution was the Scottish physicist James Clerk Maxwell (1831-1879), who "unified" the study of electricity and magnetism in four relatively tidy equations. In essence, what he discovered was that electric and magnetic *fields* were intrinsically related to one another, with or without the presence of a conductive path for electrons to flow. Stated more formally, Maxwell's discovery was this:

A changing electric field produces a perpendicular magnetic field, and
A changing magnetic field produces a perpendicular electric field.

All of this can take place in open space, the alternating electric and magnetic fields supporting each other as they travel through space at the speed of light. This dynamic structure of electric and magnetic fields propagating through space is better known as an *electromagnetic wave*.

There are many kinds of natural radiative energy composed of electromagnetic waves. Even light is electromagnetic in nature. So are X-rays and "gamma" ray radiation. The only difference between these kinds of electromagnetic radiation is the frequency of their oscillation (alternation of the electric and magnetic fields back and forth in polarity). By using a source of AC voltage and a special device called an *antenna*, we can create electromagnetic waves (of a much lower frequency than that of light) with ease.

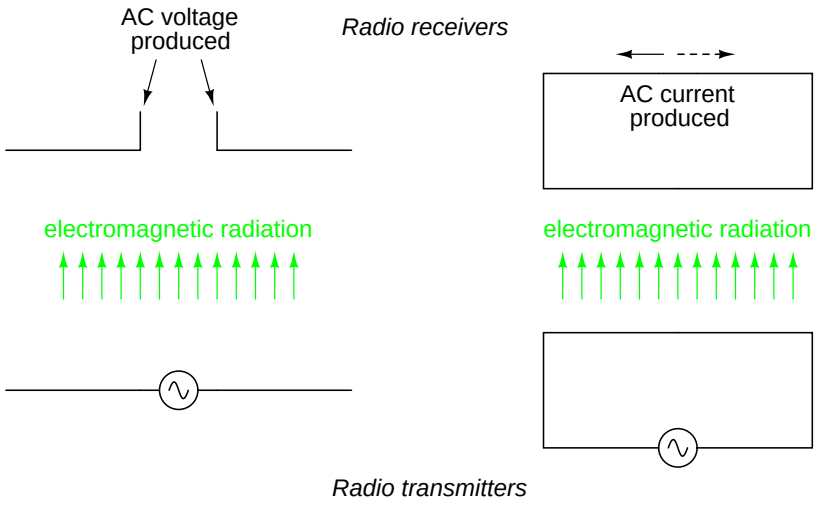
An antenna is nothing more than a device built to produce a dispersing electric or magnetic field. Two fundamental types of antennae are the *dipole* and the *loop*:



While the dipole looks like nothing more than an open circuit, and the loop a short circuit, these pieces of wire are effective radiators of electromagnetic fields when connected to AC sources of the proper frequency. The two open wires of the dipole act as a sort of capacitor (two conductors separated by a dielectric), with the electric field open to dispersal instead of being concentrated between two closely-spaced plates. The closed wire path of the loop antenna acts like an inductor with a large air core, again providing ample opportunity for the field to disperse away from the antenna instead of being concentrated and contained as in a normal inductor.

As the powered dipole radiates its changing electric field into space, a changing magnetic field is produced at right angles, thus sustaining the electric field further into space, and so on as the wave propagates at the speed of light. As the powered loop antenna radiates its changing magnetic field into space, a changing electric field is produced at right angles, with the same end-result of a continuous electromagnetic wave sent away from the antenna. Either antenna achieves the same basic task: the controlled production of an electromagnetic field.

When attached to a source of high-frequency AC power, an antenna acts as a *transmitting* device, converting AC voltage and current into electromagnetic wave energy. Antennas also have the ability to intercept electromagnetic waves and convert their energy into AC voltage and current. In this mode, an antenna acts as a *receiving* device:



While there is *much* more that may be said about antenna technology, this brief introduction is enough to give you the general idea of what's going on (and perhaps enough information to provoke a few experiments).

- **REVIEW:**
- James Maxwell discovered that changing electric fields produce perpendicular magnetic fields, and visa-versa, even in empty space.
- A twin set of electric and magnetic fields, oscillating at right angles to each other and traveling at the speed of light, constitutes an *electromagnetic wave*.
- An *antenna* is a device made of wire, designed to radiate a changing electric field or changing magnetic field when powered by a high-frequency AC source, or intercept an electromagnetic field and convert it to an AC voltage or current.
- The *dipole* antenna consists of two pieces of wire (not touching), primarily generating an electric field when energized, and secondarily producing a magnetic field in space.
- The *loop* antenna consists of a loop of wire, primarily generating a magnetic field when energized, and secondarily producing an electric field in space.

1.7 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Harvey Lew (February 7, 2004): Corrected typographical error: "circuit" should have been "circle".

Duane Damiano (February 25, 2003): Pointed out magnetic polarity error in DC generator illustration.

Mark D. Zarella (April 28, 2002): Suggestion for improving explanation of "average" waveform amplitude.

John Symonds (March 28, 2002): Suggestion for improving explanation of the unit "Hertz."

Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 2

COMPLEX NUMBERS

2.1 Introduction

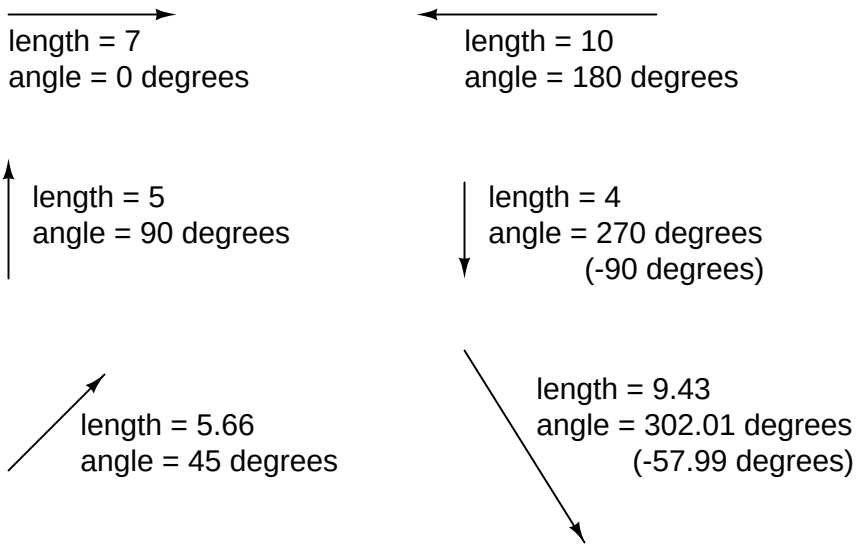
If I needed to describe the distance between two cities, I could provide an answer consisting of a single number in miles, kilometers, or some other unit of linear measurement. However, if I were to describe how to travel from one city to another, I would have to provide more information than just the distance between those two cities; I would also have to provide information about the *direction* to travel, as well.

The kind of information that expresses a single dimension, such as linear distance, is called a *scalar* quantity in mathematics. Scalar numbers are the kind of numbers you've used in most all of your mathematical applications so far. The voltage produced by a battery, for example, is a scalar quantity. So is the resistance of a piece of wire (ohms), or the current through it (amps).

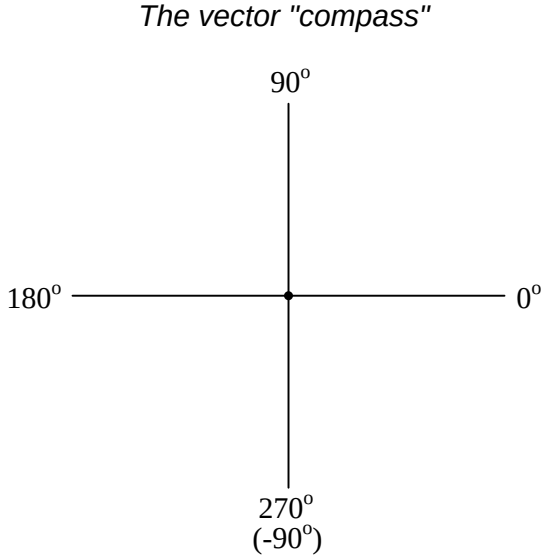
However, when we begin to analyze alternating current circuits, we find that quantities of voltage, current, and even resistance (called *impedance* in AC) are not the familiar one-dimensional quantities we're used to measuring in DC circuits. Rather, these quantities, because they're dynamic (alternating in direction and amplitude), possess other dimensions that must be taken into account. Frequency and phase shift are two of these dimensions that come into play. Even with relatively simple AC circuits, where we're only dealing with a single frequency, we still have the dimension of phase shift to contend with in addition to the amplitude.

In order to successfully analyze AC circuits, we need to work with mathematical objects and techniques capable of representing these multi-dimensional quantities. Here is where we need to abandon scalar numbers for something better suited: *complex numbers*. Just like the example of giving directions from one city to another, AC quantities in a single-frequency circuit have both amplitude (analogy: distance) and phase shift (analogy: direction). A complex number is a single mathematical quantity able to express these two dimensions of amplitude and phase shift at once.

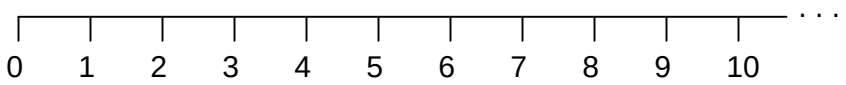
Complex numbers are easier to grasp when they're represented graphically. If I draw a line with a certain length (magnitude) and angle (direction), I have a graphic representation of a complex number which is commonly known in physics as a *vector*:



Like distances and directions on a map, there must be some common frame of reference for angle figures to have any meaning. In this case, directly right is considered to be 0^o, and angles are counted in a positive direction going counter-clockwise:

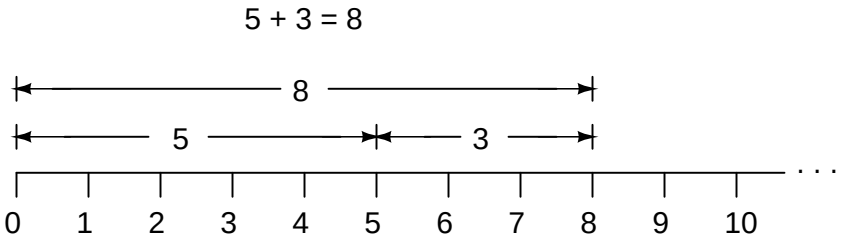


The idea of representing a number in graphical form is nothing new. We all learned this in grade school with the "number line:"

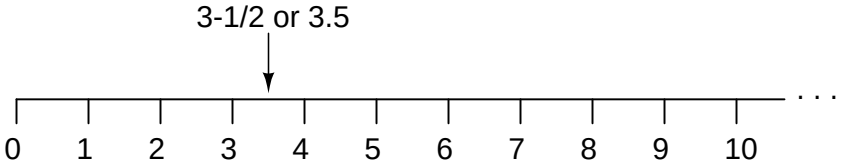


We even learned how addition and subtraction works by seeing how lengths (magnitudes) stacked

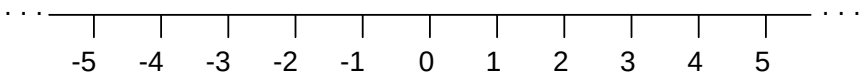
up to give a final answer:



Later, we learned that there were ways to designate the values *between* the whole numbers marked on the line. These were fractional or decimal quantities:



Later yet we learned that the number line could extend to the left of zero as well:



These fields of numbers (whole, integer, rational, irrational, real, etc.) learned in grade school share a common trait: they're all *one-dimensional*. The straightness of the number line illustrates this graphically. You can move up or down the number line, but all "motion" along that line is restricted to a single axis (horizontal). One-dimensional, scalar numbers are perfectly adequate for counting beads, representing weight, or measuring DC battery voltage, but they fall short of being able to represent something more complex like the distance *and* direction between two cities, or the amplitude *and* phase of an AC waveform. To represent these kinds of quantities, we need multidimensional representations. In other words, we need a number line that can point in different directions, and that's exactly what a vector is.

- **REVIEW:**

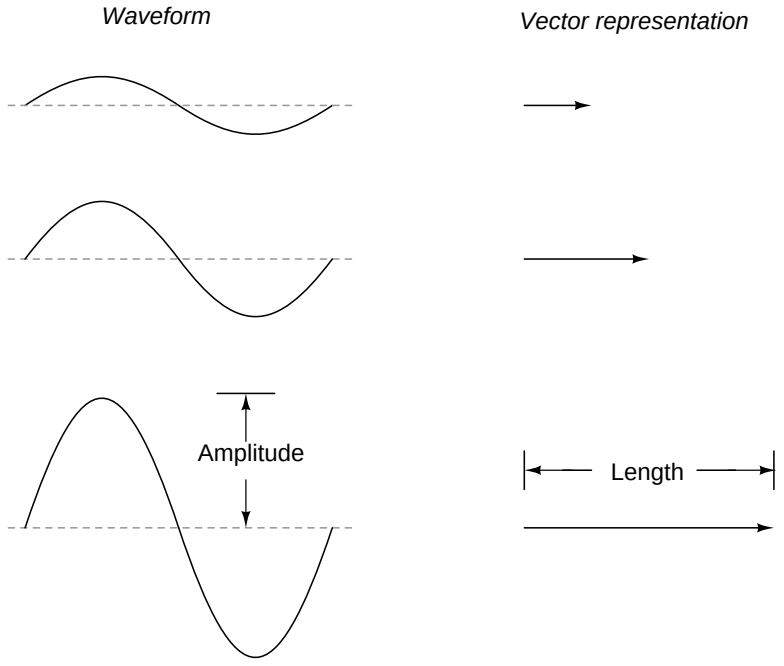
- A *scalar* number is the type of mathematical object that people are used to using in everyday life: a one-dimensional quantity like temperature, length, weight, etc.

- A *complex number* is a mathematical quantity representing two dimensions of magnitude and direction.

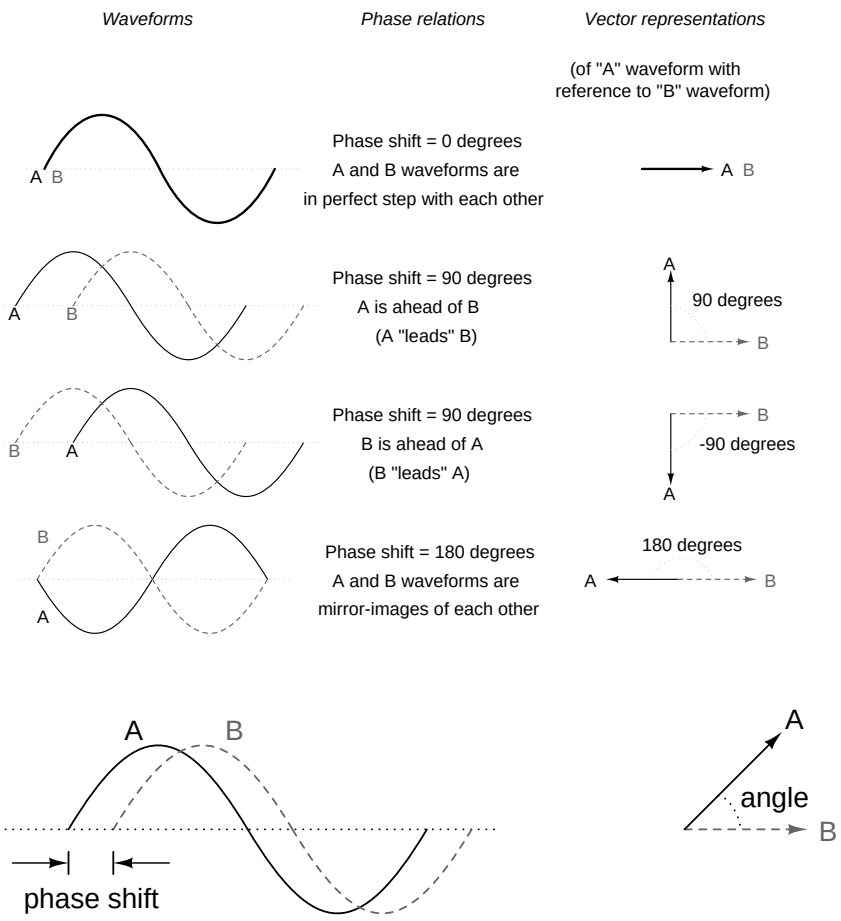
- A *vector* is a graphical representation of a complex number. It looks like an arrow, with a starting point, a tip, a definite length, and a definite direction. Sometimes the word *phasor* is used in electrical applications where the angle of the vector represents phase shift between waveforms.

2.2 Vectors and AC waveforms

Okay, so how exactly can we represent AC quantities of voltage or current in the form of a vector? The length of the vector represents the magnitude (or amplitude) of the waveform, like this:



The greater the amplitude of the waveform, the greater the length of its corresponding vector. The angle of the vector, however, represents the phase shift in degrees between the waveform in question and another waveform acting as a "reference" in time. Usually, when the phase of a waveform in a circuit is expressed, it is referenced to the power supply voltage waveform (arbitrarily stated to be "at" 0^o). Remember that phase is always a *relative* measurement between two waveforms rather than an absolute property.



The greater the phase shift in degrees between two waveforms, the greater the angle difference between the corresponding vectors. Being a relative measurement, like voltage, phase shift (vector angle) only has meaning in reference to some standard waveform. Generally this "reference" waveform is the main AC power supply voltage in the circuit. If there is more than one AC voltage source, then one of those sources is arbitrarily chosen to be the phase reference for all other measurements in the circuit.

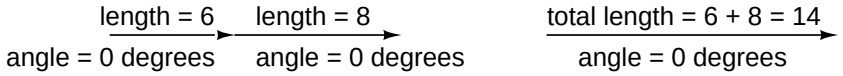
This concept of a reference point is not unlike that of the "ground" point in a circuit for the benefit of voltage reference. With a clearly defined point in the circuit declared to be "ground," it becomes possible to talk about voltage "on" or "at" single points in a circuit, being understood that those voltages (always relative between *two* points) are referenced to "ground." Correspondingly, with a clearly defined point of reference for phase it becomes possible to speak of voltages and currents in an AC circuit having definite phase angles. For example, if the current in an AC circuit is described as "24.3 milliamps at -64 degrees," it means that the current waveform has an amplitude of 24.3 mA, and it lags 64^o behind the reference waveform, usually assumed to be the main source voltage waveform.

- **REVIEW:**

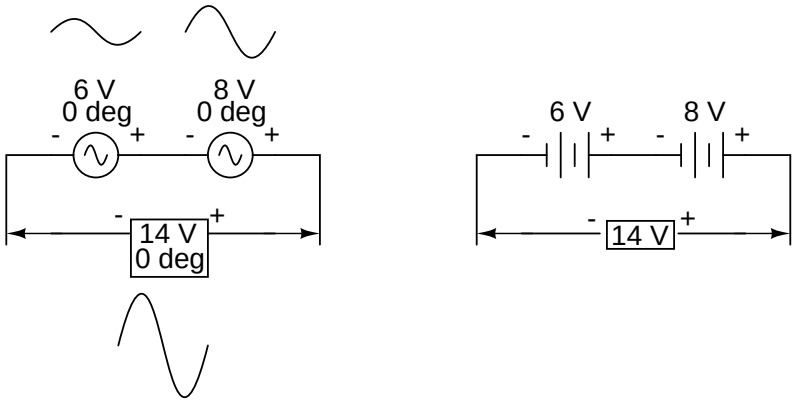
- When used to describe an AC quantity, the length of a vector represents the amplitude of the wave while the angle of a vector represents the phase angle of the wave relative to some other (reference) waveform.

2.3 Simple vector addition

Remember that vectors are mathematical objects just like numbers on a number line: they can be added, subtracted, multiplied, and divided. Addition is perhaps the easiest vector operation to visualize, so we'll begin with that. If vectors with common angles are added, their magnitudes (lengths) add up just like regular scalar quantities:

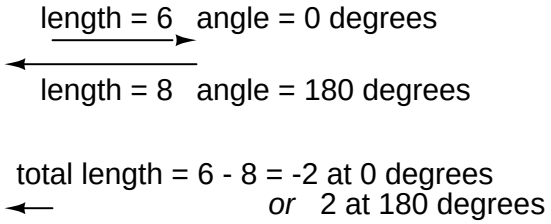


Similarly, if AC voltage sources with the same phase angle are connected together in series, their voltages add just as you might expect with DC batteries:

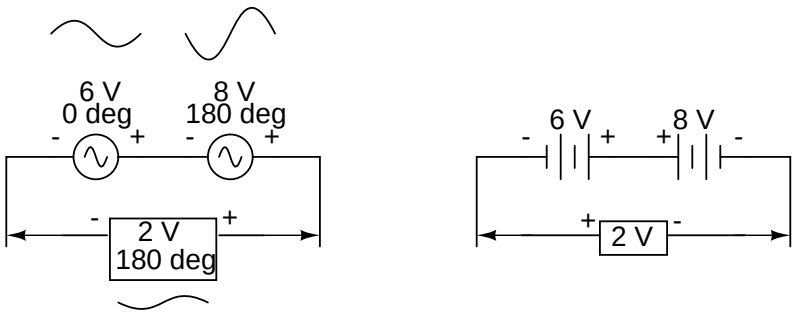


Please note the (+) and (-) polarity marks next to the leads of the two AC sources. Even though we know AC doesn't have "polarity" in the same sense that DC does, these marks are essential to knowing how to reference the given phase angles of the voltages. This will become more apparent in the next example.

If vectors directly opposing each other (180^o out of phase) are added together, their magnitudes (lengths) subtract just like positive and negative scalar quantities subtract when added:

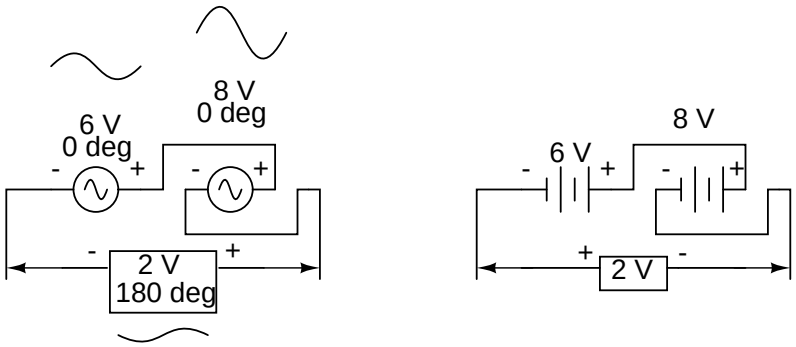


Similarly, if opposing AC voltage sources are connected in series, their voltages subtract as you might expect with DC batteries connected in an opposing fashion:

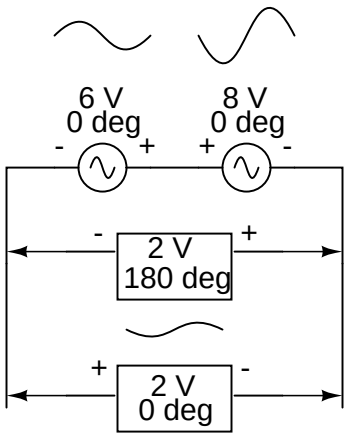


Determining whether or not these voltage sources are opposing each other requires an examination of their polarity markings *and* their phase angles. Notice how the polarity markings in the above diagram seem to indicate additive voltages (from left to right, we see - and + on the 6 volt source, - and + on the 8 volt source). Even though these polarity markings would normally indicate an *additive* effect in a DC circuit (the two voltages working together to produce a greater total voltage), in this AC circuit they're actually pushing in opposite directions because one of those voltages has a phase angle of 0^o and the other a phase angle of 180^o. The result, of course, is a total voltage of 2 volts.

We could have just as well shown the opposing voltages subtracting in series like this:

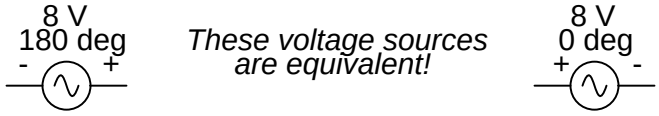


Note how the polarities appear to be opposed to each other now, due to the reversal of wire connections on the 8 volt source. Since both sources are described as having equal phase angles (0^o), they truly are opposed to one another, and the overall effect is the same as the former scenario with "additive" polarities and differing phase angles: a total voltage of only 2 volts.



Just as there are two ways to express the phase of the sources, there are two ways to express their resultant sum.

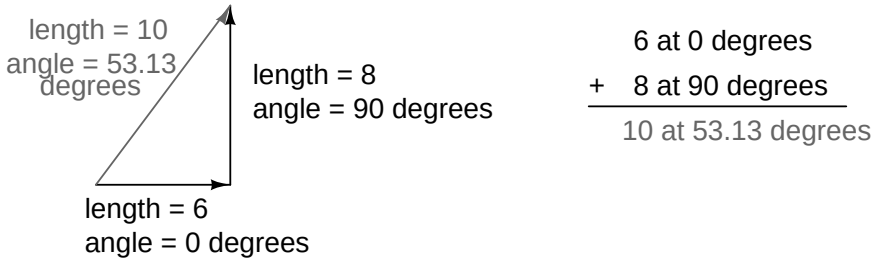
The resultant voltage can be expressed in two different ways: 2 volts at 180^o with the (-) symbol on the left and the (+) symbol on the right, or 2 volts at 0^o with the (+) symbol on the left and the (-) symbol on the right. A reversal of wires from an AC voltage source is the same as phase-shifting that source by 180^o.



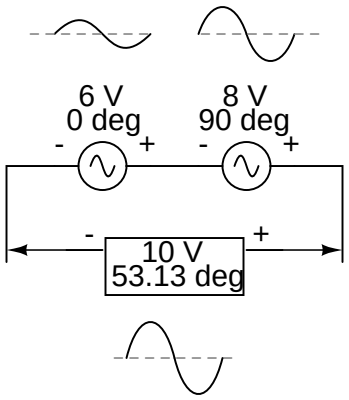
2.4 Complex vector addition

If vectors with uncommon angles are added, their magnitudes (lengths) add up quite differently than that of scalar magnitudes:

Vector addition



If two AC voltages – 90^o out of phase – are added together by being connected in series, their voltage magnitudes do not directly add or subtract as with scalar voltages in DC. Instead, these voltage quantities are complex quantities, and just like the above vectors, which add up in a trigonometric fashion, a 6 volt source at 0^o added to an 8 volt source at 90^o results in 10 volts at a phase angle of 53.13^o:



Compared to DC circuit analysis, this is very strange indeed. Note that it's possible to obtain voltmeter indications of 6 and 8 volts, respectively, across the two AC voltage sources, yet only read 10 volts for a total voltage!

There is no suitable DC analogy for what we're seeing here with two AC voltages slightly out of phase. DC voltages can only directly aid or directly oppose, with nothing in between. With AC, two voltages can be aiding or opposing one another *to any degree* between fully-aiding and fully-opposing, inclusive. Without the use of vector (complex number) notation to describe AC quantities, it would be *very* difficult to perform mathematical calculations for AC circuit analysis.

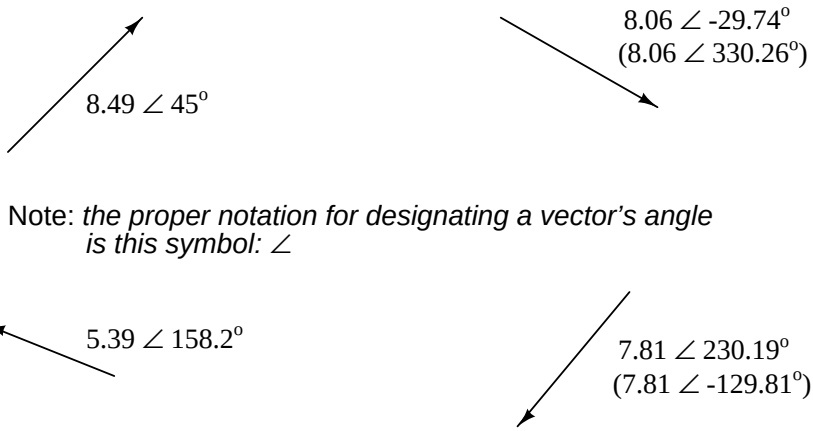
In the next section, we'll learn how to represent vector quantities in symbolic rather than graphical form. Vector and triangle diagrams suffice to illustrate the general concept, but more precise methods of symbolism must be used if any serious calculations are to be performed on these quantities.

- **REVIEW:**
- DC voltages can only either directly aid or directly oppose each other when connected in series. AC voltages may aid or oppose *to any degree* depending on the phase shift between them.

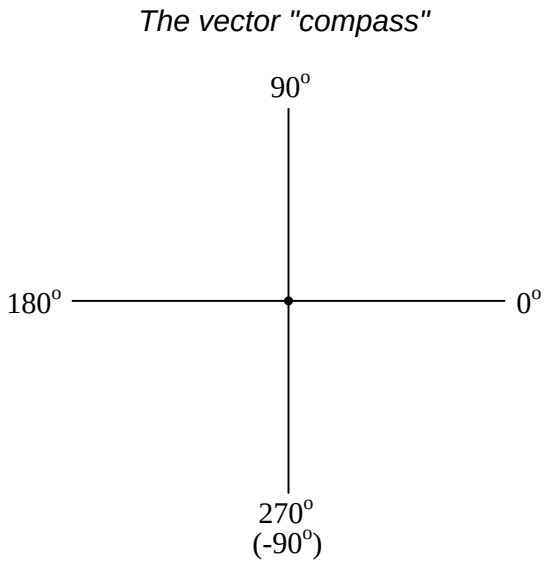
2.5 Polar and rectangular notation

In order to work with these complex numbers without drawing vectors, we first need some kind of standard mathematical notation. There are two basic forms of complex number notation: *polar* and *rectangular*.

Polar form is where a complex number is denoted by the *length* (otherwise known as the *magnitude*, *absolute value*, or *modulus*) and the *angle* of its vector (usually denoted by an angle symbol that looks like this: $\angle$). To use the map analogy, polar notation for the vector from New York City to San Diego would be something like "2400 miles, southwest." Here are two examples of vectors and their polar notations:



Standard orientation for vector angles in AC circuit calculations defines 0^o as being to the right (horizontal), making 90^o straight up, 180^o to the left, and 270^o straight down. Please note that vectors angled "down" can have angles represented in polar form as positive numbers in excess of 180, or negative numbers less than 180. For example, a vector angled $\angle$ 270^o (straight down) can also be said to have an angle of -90^o. The above vector on the right ($5.4 \angle 326^o$) can also be denoted as $5.4 \angle$ -34^o.

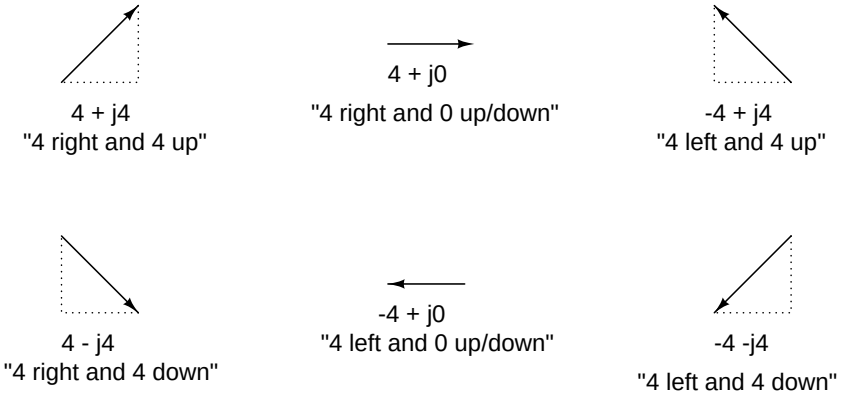


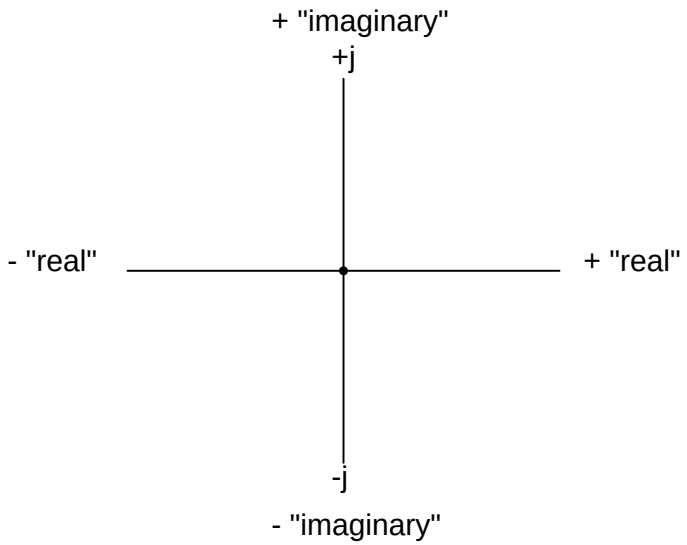
Rectangular form, on the other hand, is where a complex number is denoted by its respective horizontal and vertical components. In essence, the angled vector is taken to be the hypotenuse of a right triangle, described by the lengths of the adjacent and opposite sides. Rather than describing a vector's length and direction by denoting magnitude and angle, it is described in terms of "how far left/right" and "how far up/down."

These two dimensional figures (horizontal and vertical) are symbolized by two numerical figures. In order to distinguish the horizontal and vertical dimensions from each other, the vertical is prefixed

with a lower-case "i" (in pure mathematics) or "j" (in electronics). These lower-case letters do not represent a physical variable (such as instantaneous current, also symbolized by a lower-case letter "i"), but rather are mathematical *operators* used to distinguish the vector's vertical component from its horizontal component. As a complete complex number, the horizontal and vertical quantities are written as a sum:

In "rectangular" form, a vector's length and direction are denoted in terms of its horizontal and vertical span, the first number representing the horixontal ("real") and the second number (with the "j" prefix) representing the vertical ("imaginary") dimensions.

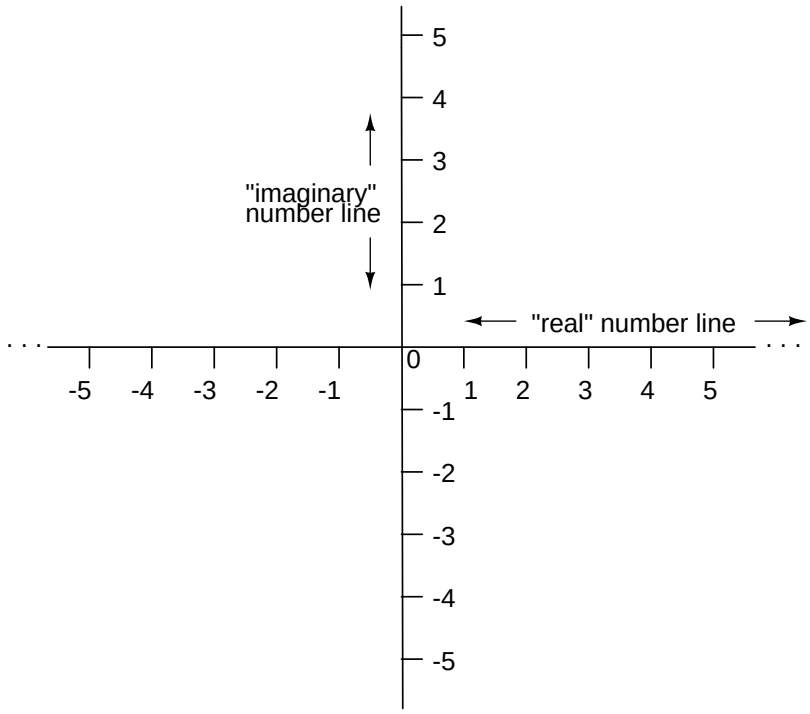




The horizontal component is referred to as the *real* component, since that dimension is compatible with normal, scalar ("real") numbers. The vertical component is referred to as the *imaginary* component, since that dimension lies in a different direction, totally alien to the scale of the real

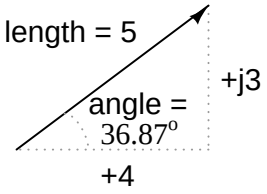
numbers.

The "real" axis of the graph corresponds to the familiar number line we saw earlier: the one with both positive and negative values on it. The "imaginary" axis of the graph corresponds to another number line situated at 90^o to the "real" one. Vectors being two-dimensional things, we must have a two-dimensional "map" upon which to express them, thus the two number lines perpendicular to each other:



Either method of notation is valid for complex numbers. The primary reason for having two methods of notation is for ease of longhand calculation, rectangular form lending itself to addition and subtraction, and polar form lending itself to multiplication and division.

Conversion between the two notational forms involves simple trigonometry. To convert from polar to rectangular, find the real component by multiplying the polar magnitude by the cosine of the angle, and the imaginary component by multiplying the polar magnitude by the sine of the angle. This may be understood more readily by drawing the quantities as sides of a right triangle, the hypotenuse of the triangle representing the vector itself (its length and angle with respect to the horizontal constituting the polar form), the horizontal and vertical sides representing the "real" and "imaginary" rectangular components, respectively:



$5 \angle 36.87^o$ **(polar form)**

(5)(cos 36.87^o) = 4 (real component)

(5)(sin 36.87^o) = 3 (imaginary component)

$4 + j3$ **(rectangular form)**

To convert from rectangular to polar, find the polar magnitude through the use of the Pythagorean Theorem (the polar magnitude is the hypotenuse of a right triangle, and the real and imaginary components are the adjacent and opposite sides, respectively), and the angle by taking the arctangent of the imaginary component divided by the real component:

$4 + j3$ **(rectangular form)**

$c = \sqrt{a^2 + b^2}$ (pythagorean theorem)

polar magnitude = $\sqrt{4^2 + 3^2}$

polar magnitude = 5

polar angle = $\arctan \frac{3}{4}$

polar angle = 36.87^o

$5 \angle 36.87^o$ **(polar form)**

- **REVIEW:**
- *Polar* notation denotes a complex number in terms of its vector's length and angular direction from the starting point. Example: fly 45 miles $\angle$ 203^o (West by Southwest).
- *Rectangular* notation denotes a complex number in terms of its horizontal and vertical dimensions. Example: drive 41 miles West, then turn and drive 18 miles South.

- In rectangular notation, the first quantity is the "real" component (horizontal dimension of vector) and the second quantity is the "imaginary" component (vertical dimension of vector). The imaginary component is preceded by a lower-case "j," sometimes called the *j operator*.

- Both polar and rectangular forms of notation for a complex number can be related graphically in the form of a right triangle, with the hypotenuse representing the vector itself (polar form: hypotenuse length = magnitude; angle with respect to horizontal side = angle), the horizontal side representing the rectangular "real" component, and the vertical side representing the rectangular "imaginary" component.

2.6 Complex number arithmetic

Since complex numbers are legitimate mathematical entities, just like scalar numbers, they can be added, subtracted, multiplied, divided, squared, inverted, and such, just like any other kind of number. Some scientific calculators are programmed to directly perform these operations on two or more complex numbers, but these operations can also be done "by hand." This section will show you how the basic operations are performed. It is *highly* recommended that you equip yourself with a scientific calculator capable of performing arithmetic functions easily on complex numbers. It will make your study of AC circuit much more pleasant than if you're forced to do all calculations the longer way.

Addition and subtraction with complex numbers in rectangular form is easy. For addition, simply add up the real components of the complex numbers to determine the real component of the sum, and add up the imaginary components of the complex numbers to determine the imaginary component of the sum:

2 + j5	175 - j34	-36 + j10
+ 4 - j3	+ 80 - j15	+ 20 + j82
6 + j2	**255 - j49**	**-16 + j92**

When subtracting complex numbers in rectangular form, simply subtract the real component of the second complex number from the real component of the first to arrive at the real component of the difference, and subtract the imaginary component of the second complex number from the imaginary component of the first to arrive the imaginary component of the difference:

2 + j5	175 - j34	-36 + j10
- 4 - j3	- 80 - j15	- 20 + j82
-2 + j8	**95 - j19**	**-56 - j72**

For longhand multiplication and division, polar is the favored notation to work with. When multiplying complex numbers in polar form, simply *multiply* the polar magnitudes of the complex numbers to determine the polar magnitude of the product, and *add* the angles of the complex numbers to determine the angle of the product:

$$(35 \angle 65^o)(10 \angle -12^o) = \mathbf{350 \angle 53^o}$$

$$(124 \angle 250^o)(11 \angle 100^o) = \mathbf{1364 \angle -10^o} \quad \textit{or} \quad \mathbf{1364 \angle 350^o}$$

$$(3 \angle 30^o)(5 \angle -30^o) = \mathbf{15 \angle 0^o}$$

Division of polar-form complex numbers is also easy: simply divide the polar magnitude of the first complex number by the polar magnitude of the second complex number to arrive at the polar magnitude of the quotient, and subtract the angle of the second complex number from the angle of the first complex number to arrive at the angle of the quotient:

$$\frac{35 \angle 65^o}{10 \angle -12^o} = \mathbf{3.5 \angle 77^o}$$

$$\frac{124 \angle 250^o}{11 \angle 100^o} = \mathbf{11.273 \angle 150^o}$$

$$\frac{3 \angle 30^o}{5 \angle -30^o} = \mathbf{0.6 \angle 60^o}$$

To obtain the reciprocal, or "invert" (1/x), a complex number, simply divide the number (in polar form) into a scalar value of 1, which is nothing more than a complex number with no imaginary component (angle = 0):

$$\frac{1}{35 \angle 65^o} = \frac{1 \angle 0^o}{35 \angle 65^o} = \mathbf{0.02857 \angle -65^o}$$

$$\frac{1}{10 \angle -12^o} = \frac{1 \angle 0^o}{10 \angle -12^o} = \mathbf{0.1 \angle 12^o}$$

$$\frac{1}{0.0032 \angle 10^o} = \frac{1 \angle 0^o}{0.0032 \angle 10^o} = \mathbf{312.5 \angle -10^o}$$

These are the basic operations you will need to know in order to manipulate complex numbers in the analysis of AC circuits. Operations with complex numbers are by no means limited just to addition, subtraction, multiplication, division, and inversion, however. Virtually any arithmetic operation that can be done with scalar numbers can be done with complex numbers, including powers, roots, solving simultaneous equations with complex coefficients, and even trigonometric functions (although this involves a whole new perspective in trigonometry called *hyperbolic functions* which is well beyond the scope of this discussion). Be sure that you're familiar with the basic arithmetic operations of addition, subtraction, multiplication, division, and inversion, and you'll have little trouble with AC circuit analysis.

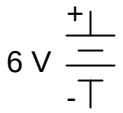
- **REVIEW:**
- To add complex numbers in rectangular form, add the real components and add the imaginary components. Subtraction is similar.
- To multiply complex numbers in polar form, multiply the magnitudes and add the angles. To divide, divide the magnitudes and subtract one angle from the other.

2.7 More on AC "polarity"

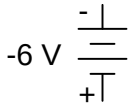
Complex numbers are useful for AC circuit analysis because they provide a convenient method of symbolically denoting phase shift between AC quantities like voltage and current. However, for most people the equivalence between abstract vectors and real circuit quantities is not an easy one to grasp. Earlier in this chapter we saw how AC voltage sources are given voltage figures in complex form (magnitude *and* phase angle), as well as polarity markings. Being that alternating current has no set "polarity" as direct current does, these polarity markings and their relationship to phase angle tends to be confusing. This section is written in the attempt to clarify some of these issues.

Voltage is an inherently *relative* quantity. When we measure a voltage, we have a choice in how we connect a voltmeter or other voltage-measuring instrument to the source of voltage, as there are two points between which the voltage exists, and two test leads on the instrument with which to make connection. In DC circuits, we denote the polarity of voltage sources and voltage drops explicitly, using "+" and "-" symbols, and use color-coded meter test leads (red and black). If a digital voltmeter indicates a negative DC voltage, we know that its test leads are connected "backward" to the voltage (red lead connected to the "-" and black lead to the "+").

Batteries have their polarity designated by way of intrinsic symbology: the short-line side of a battery is always the negative (-) side and the long-line side always the positive (+):



Although it would be mathematically correct to represent a battery's voltage as a negative figure with reversed polarity markings, it would be decidedly unconventional:



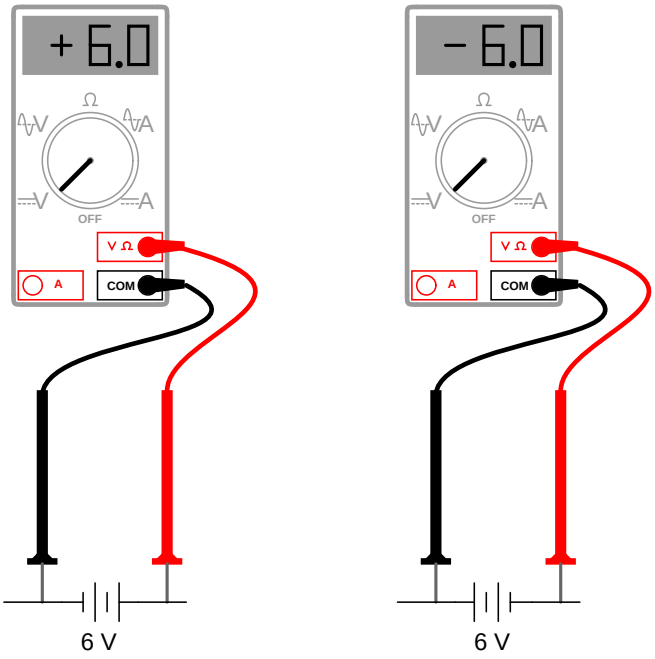
Interpreting such notation might be easier if the "+" and "-" polarity markings were viewed as reference points for voltmeter test leads, the "+" meaning "red" and the "-" meaning "black." A voltmeter connected to the above battery with red lead to the bottom terminal and black lead to the top terminal would indeed indicate a negative voltage (-6 volts). Actually, this form of notation and interpretation is not as unusual as you might think: it's commonly encountered in problems of DC network analysis where "+" and "-" polarity marks are initially drawn according to educated guess, and later interpreted as correct or "backward" according to the mathematical sign of the figure calculated.

In AC circuits, though, we don't deal with "negative" quantities of voltage. Instead, we describe to what degree one voltage aids or opposes another by *phase*: the time-shift between two waveforms. We never describe an AC voltage as being negative in sign, because the facility of polar notation allows for vectors pointing in an opposite direction. If one AC voltage directly opposes another AC voltage, we simply say that one is 180^o out of phase with the other.

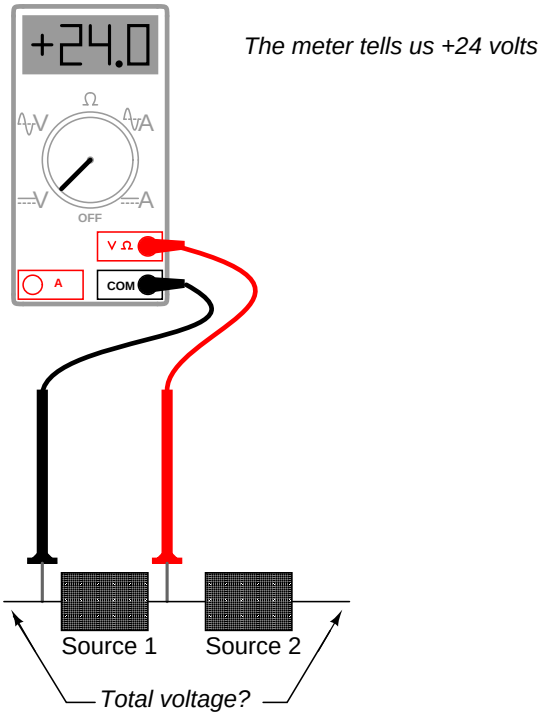
Still, voltage is relative between two points, and we have a choice in how we might connect a voltage-measuring instrument between those two points. The mathematical sign of a DC voltmeter's reading has meaning only in the context of its test lead connections: which terminal the red lead is touching, and which terminal the black lead is touching. Likewise, the phase angle of an AC voltage has meaning only in the context of knowing which of the two points is considered the "reference" point. Because of this fact, "+" and "-" polarity marks are often placed by the terminals of an AC voltage in schematic diagrams to give the stated phase angle a frame of reference.

Let's review these principles with some graphical aids. First, the principle of relating test lead connections to the mathematical sign of a DC voltmeter indication:

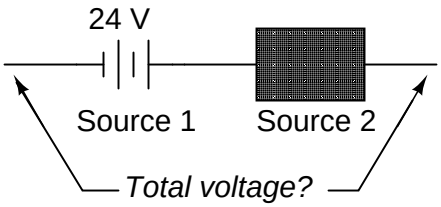
Test lead colors provide a frame of reference *for interpreting the sign (+ or -) of the meter's indication.*



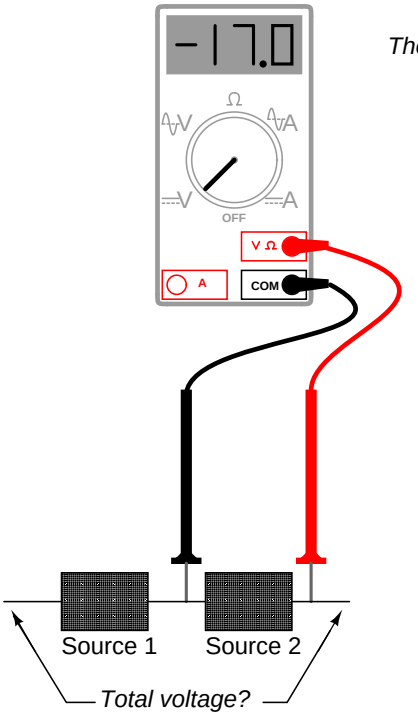
The mathematical sign of a digital DC voltmeter's display has meaning only in the context of its test lead connections. Consider the use of a DC voltmeter in determining whether or not two DC voltage sources are aiding or opposing each other, assuming that both sources are unlabeled as to their polarities. Using the voltmeter to measure across the first source:



This first measurement of +24 across the left-hand voltage source tells us that the black lead of the meter really is touching the negative side of voltage source #1, and the red lead of the meter really is touching the positive. Thus, we know source #1 is a battery facing in this orientation:

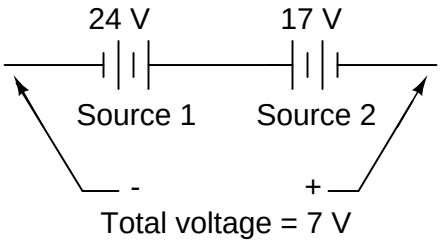


Measuring the other unknown voltage source:



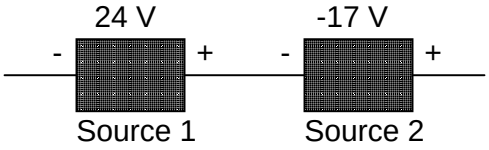
The meter tells us -17 volts

This second voltmeter reading, however, is a *negative* (-) 17 volts, which tells us that the black test lead is actually touching the positive side of voltage source #2, while the red test lead is actually touching the negative. Thus, we know that source #2 is a battery facing in the *opposite* direction:



It should be obvious to any experienced student of DC electricity that these two batteries are opposing one another. By definition, opposing voltages *subtract* from one another, so we subtract 17 volts from 24 volts to obtain the total voltage across the two: 7 volts.

We could, however, draw the two sources as nondescript boxes, labeled with the exact voltage figures obtained by the voltmeter, the polarity marks indicating voltmeter test lead placement:

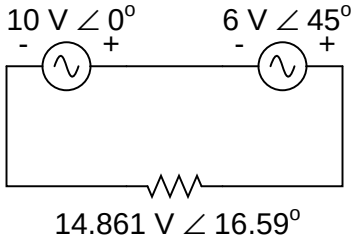


According to this diagram, the polarity marks (which indicate meter test lead placement) indicate

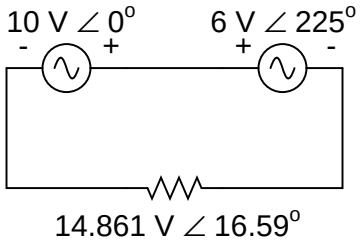
the sources *aiding* each other. By definition, aiding voltage sources *add* with one another to form the total voltage, so we add 24 volts to -17 volts to obtain 7 volts: still the correct answer. If we let the polarity markings guide our decision to either add or subtract voltage figures – whether those polarity markings represent the *true* polarity or just the meter test lead orientation – and include the mathematical signs of those voltage figures in our calculations, the result will always be correct. Again, the polarity markings serve as *frames of reference* to place the voltage figures' mathematical signs in proper context.

The same is true for AC voltages, except that *phase angle* substitutes for mathematical *sign*. In order to relate multiple AC voltages at different phase angles to each other, we need polarity markings to provide frames of reference for those voltages' phase angles.

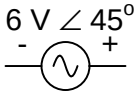
Take for example the following circuit:



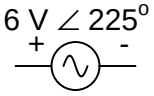
The polarity markings show these two voltage sources aiding each other, so to determine the total voltage across the resistor we must *add* the voltage figures of 10 V $\angle$ 0^o and 6 V $\angle$ 45^o together to obtain 14.861 V $\angle$ 16.59^o. However, it would be perfectly acceptable to represent the 6 volt source as 6 V $\angle$ 225^o, with a reversed set of polarity markings, and still arrive at the same total voltage:



6 V $\angle$ 45^o with negative on the left and positive on the right is exactly the same as 6 V $\angle$ 225^o with positive on the left and negative on the right: the reversal of polarity markings perfectly complements the addition of 180^o to the phase angle designation:



. . . is equivalent to . . .



Unlike DC voltage sources, whose symbols intrinsically define polarity by means of short and long lines, AC voltage symbols have no intrinsic polarity marking. Therefore, any polarity marks

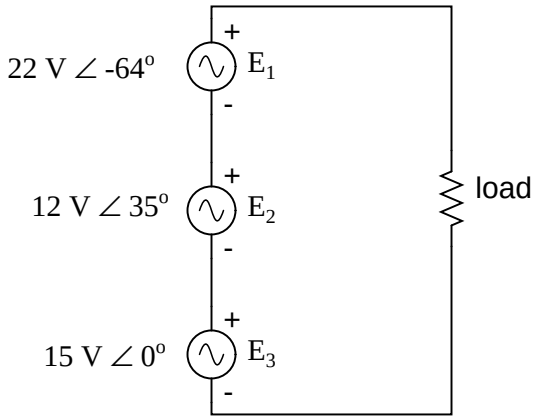
must by included as additional symbols on the diagram, and there is no one "correct" way in which to place them. They must, however, correlate with the given phase angle to represent the true phase relationship of that voltage with other voltages in the circuit.

- **REVIEW:**

- Polarity markings are sometimes given to AC voltages in circuit schematics in order to provide a frame of reference for their phase angles.

2.8 Some examples with AC circuits

Let's connect three AC voltage sources in series and use complex numbers to determine additive voltages. All the rules and laws learned in the study of DC circuits apply to AC circuits as well (Ohm's Law, Kirchhoff's Laws, network analysis methods), with the exception of power calculations (Joule's Law). The only qualification is that all variables *must* be expressed in complex form, taking into account phase as well as magnitude, and all voltages and currents must be of the same frequency (in order that their phase relationships remain constant).

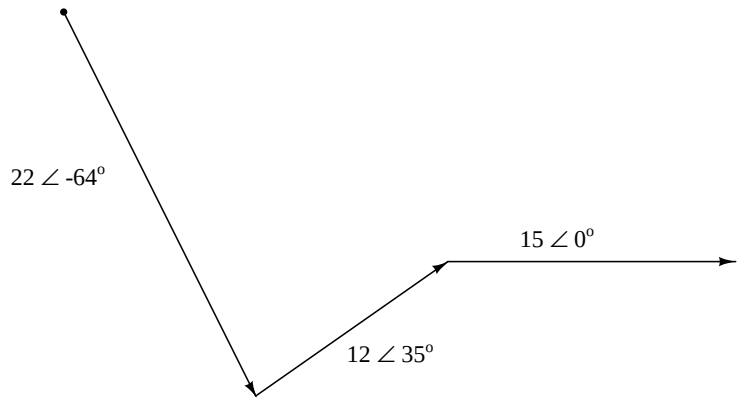


The polarity marks for all three voltage sources are oriented in such a way that their stated voltages should add to make the total voltage across the load resistor. Notice that although magnitude and phase angle is given for each AC voltage source, no frequency value is specified. If this is the case, it is assumed that all frequencies are equal, thus meeting our qualifications for applying DC rules to an AC circuit (all figures given in complex form, all of the same frequency). The setup of our equation to find total voltage appears as such:

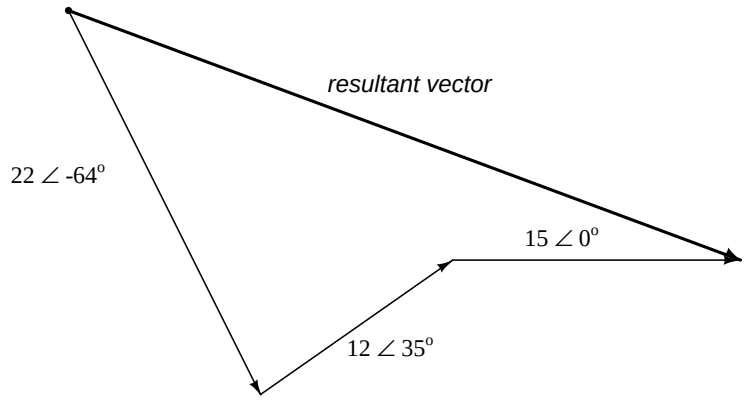
$$E_{total} = E_1 + E_2 + E_3$$

$$E_{total} = (22\ V \angle -64^o) + (12\ V \angle 35^o) + (15\ V \angle 0^o)$$

Graphically, the vectors add up in this manner:



The sum of these vectors will be a resultant vector originating at the starting point for the 22 volt vector (dot at upper-left of diagram) and terminating at the ending point for the 15 volt vector (arrow tip at the middle-right of the diagram):



In order to determine what the resultant vector's magnitude and angle are without resorting to graphic images, we can convert each one of these polar-form complex numbers into rectangular form and add. Remember, we're *adding* these figures together because the polarity marks for the three voltage sources are oriented in an additive manner:

$15 \text{ V} \angle 0^o = 15 + j0 \text{ V}$

$12 \text{ V} \angle 35^o = 9.8298 + j6.8829 \text{ V}$

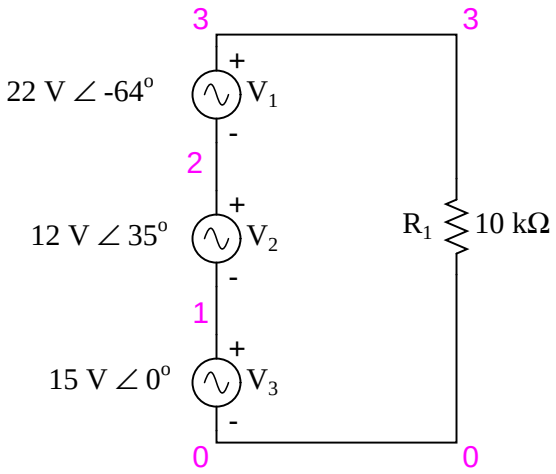
$22 \text{ V} \angle -64^o = 9.6442 - j19.7735 \text{ V}$

$$\begin{array}{rll} & 15 & + j0 \quad \text{V} \\ & 9.8298 & + j6.8829 \text{ V} \\ + & 9.6442 & - j19.7735 \text{ V} \\ \hline & \mathbf{34.4740} & \mathbf{- j12.8906 \text{ V}} \end{array}$$

In polar form, this equates to 36.8052 volts $\angle$ -20.5018^o. What this means in real terms is that the voltage measured across these three voltage sources will be 36.8052 volts, lagging the 15 volt (0^o phase reference) by 20.5018^o. A voltmeter connected across these points in a real circuit would only indicate the polar magnitude of the voltage (36.8052 volts), not the angle. An oscilloscope could be used to display two voltage waveforms and thus provide a phase shift measurement, but not a voltmeter. The same principle holds true for AC ammeters: they indicate the polar magnitude of the current, not the phase angle.

This is extremely important in relating calculated figures of voltage and current to real circuits. Although rectangular notation is convenient for addition and subtraction, and was indeed the final step in our sample problem here, it is not very applicable to practical measurements. Rectangular figures must be converted to polar figures (specifically polar *magnitude*) before they can be related to actual circuit measurements.

We can use SPICE to verify the accuracy of our results. In this test circuit, the 10 kΩ resistor value is quite arbitrary. It's there so that SPICE does not declare an open-circuit error and abort analysis. Also, the choice of frequencies for the simulation (60 Hz) is quite arbitrary, because resistors respond uniformly for all frequencies of AC voltage and current. There are other components (notably capacitors and inductors) which do not respond uniformly to different frequencies, but that is another subject!



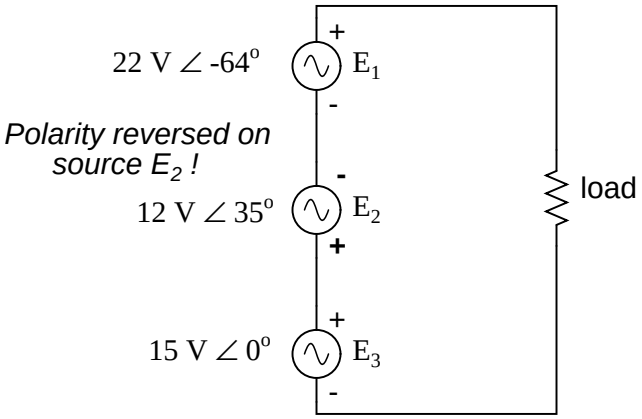

```
ac voltage addition
v1 1 0 ac 15 0 sin
v2 2 1 ac 12 35 sin
v3 3 2 ac 22 -64 sin
r1 3 0 10k
.ac lin 1 60 60            I'm using a frequency of 60 Hz
.print ac v(3,0) vp(3,0)   as a default value
.end
```

```
freq          v(3)        vp(3)
6.000E+01     3.681E+01  -2.050E+01
```

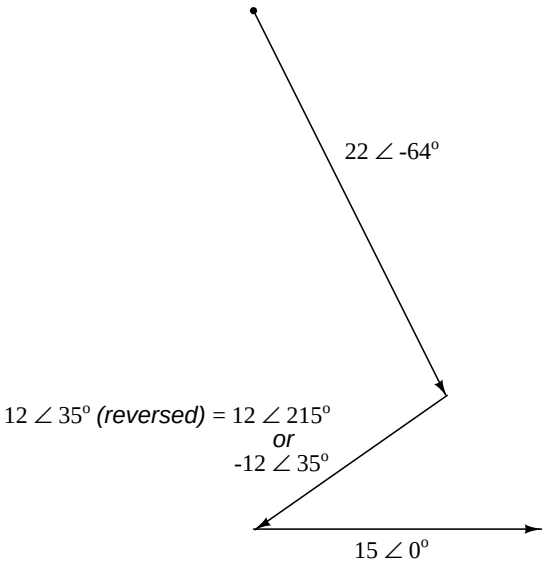
Sure enough, we get a total voltage of 36.81 volts ∠ -20.5^o (with reference to the 15 volt source, whose phase angle was arbitrarily stated at zero degrees so as to be the "reference" waveform).

At first glance, this is counter-intuitive. How is it possible to obtain a total voltage of just over 36 volts with 15 volt, 12 volt, and 22 volt supplies connected in series? With DC, this would be impossible, as voltage figures will either directly add or subtract, depending on polarity. But with AC, our "polarity" (phase shift) can vary anywhere in between full-aiding and full-opposing, and this allows for such paradoxical summing.

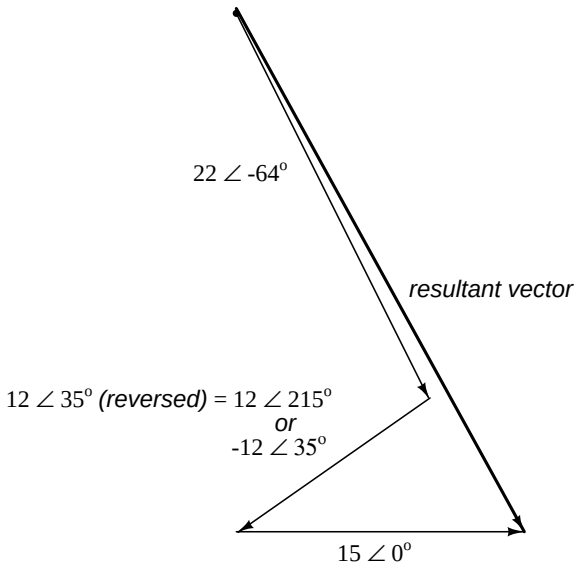
What if we took the same circuit and reversed one of the supply's connections? Its contribution to the total voltage would then be the opposite of what it was before:



Note how the 12 volt supply's phase angle is still referred to as 35^o, even though the leads have been reversed. Remember that the phase angle of any voltage drop is stated in reference to its noted polarity. Even though the angle is still written as 35^o, the vector will be drawn 180^o opposite of what it was before:



The resultant (sum) vector should begin at the upper-left point (origin of the 22 volt vector) and terminate at the right arrow tip of the 15 volt vector:



The connection reversal on the 12 volt supply can be represented in two different ways in polar form: by an addition of 180^o to its vector angle (making it 12 volts $\angle$ 215^o), or a reversal of sign on the magnitude (making it -12 volts $\angle$ 35^o). Either way, conversion to rectangular form yields the same result:

$$12 \text{ V} \angle 35^\circ \textit{ (reversed)} = 12 \text{ V} \angle 215^\circ = \textbf{-9.8298 - j6.8829 V}$$

or

$$-12 \text{ V} \angle 35^\circ = \textbf{-9.8298 - j6.8829 V}$$

The resulting addition of voltages in rectangular form, then:

$$\begin{array}{r} 15 \quad + \text{j}0 \quad \text{V} \\ -9.8298 - \text{j}6.8829 \text{ V} \\ + \quad 9.6442 \quad - \text{j}19.7735 \text{ V} \\ \hline \textbf{14.8143 - j26.6564 V} \end{array}$$

In polar form, this equates to 30.4964 V $\angle$ -60.9368^o. Once again, we will use SPICE to verify the results of our calculations:

```
ac voltage addition
v1 1 0 ac 15 0 sin
v2 1 2 ac 12 35 sin    Note the reversal of node numbers 2 and 1
v3 3 2 ac 22 -64 sin   to simulate the swapping of connections
r1 3 0 10k
.ac lin 1 60 60
.print ac v(3,0) vp(3,0)
.end
```

```
freq          v(3)        vp(3)
6.000E+01     3.050E+01  -6.094E+01
```

- **REVIEW:**
- All the laws and rules of DC circuits apply to AC circuits, with the exception of power calculations (Joule's Law), so long as all values are expressed and manipulated in complex form, and all voltages and currents are at the same frequency.
- When reversing the direction of a vector (equivalent to reversing the polarity of an AC voltage source in relation to other voltage sources), it can be expressed in either of two different ways: adding 180^o to the angle, or reversing the sign of the magnitude.
- Meter measurements in an AC circuit correspond to the *polar magnitudes* of calculated values. Rectangular expressions of complex quantities in an AC circuit have no direct, empirical equivalent, although they are convenient for performing addition and subtraction, as Kirchhoff's Voltage and Current Laws require.

2.9 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

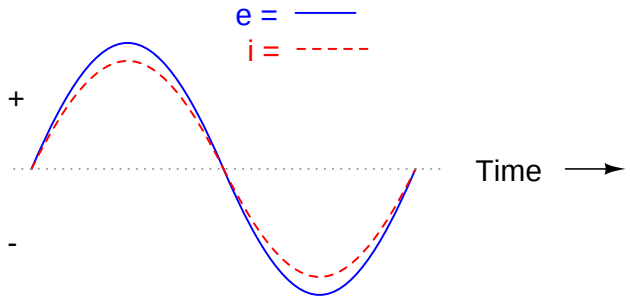
Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 3

REACTANCE AND IMPEDANCE – INDUCTIVE

3.1 AC resistor circuits

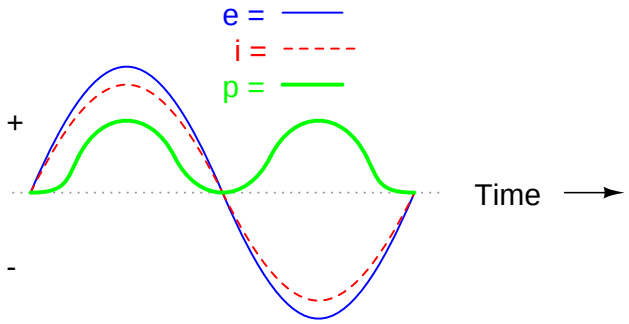
If we were to plot the current and voltage for a very simple AC circuit consisting of a source and a resistor, it would look something like this:



Because the resistor simply and directly resists the flow of electrons at all periods of time, the waveform for the voltage drop across the resistor is exactly in phase with the waveform for the current through it. We can look at any point in time along the horizontal axis of the plot and compare those values of current and voltage with each other (any "snapshot" look at the values of a wave are referred to as *instantaneous values*, meaning the values at that *instant* in time). When the instantaneous value for current is zero, the instantaneous voltage across the resistor is also zero. Likewise, at the moment in time where the current through the resistor is at its positive peak, the voltage across the resistor is also at its positive peak, and so on. At any given point in time along

the waves, Ohm's Law holds true for the instantaneous values of voltage and current.

We can also calculate the power dissipated by this resistor, and plot those values on the same graph:



Note that the power is never a negative value. When the current is positive (above the line), the voltage is also positive, resulting in a power (p=ie) of a positive value. Conversely, when the current is negative (below the line), the voltage is also negative, which results in a positive value for power (a negative number multiplied by a negative number equals a positive number). This consistent "polarity" of power tells us that the resistor is always dissipating power, taking it from the source and releasing it in the form of heat energy. Whether the current is positive or negative, a resistor still dissipates energy.

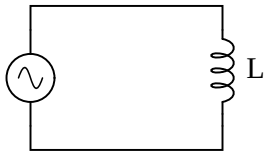
3.2 AC inductor circuits

Inductors do not behave the same as resistors. Whereas resistors simply oppose the flow of electrons through them (by dropping a voltage directly proportional to the current), inductors oppose *changes* in current through them, by dropping a voltage directly proportional to the *rate of change* of current. In accordance with *Lenz's Law*, this induced voltage is always of such a polarity as to try to maintain current at its present value. That is, if current is increasing in magnitude, the induced voltage will "push against" the electron flow; if current is decreasing, the polarity will reverse and "push with" the electron flow to oppose the decrease. This opposition to current change is called *reactance*, rather than resistance.

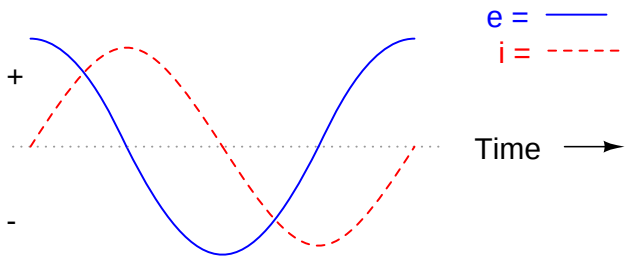
Expressed mathematically, the relationship between the voltage dropped across the inductor and rate of current change through the inductor is as such:

$$e = L\frac{di}{dt}$$

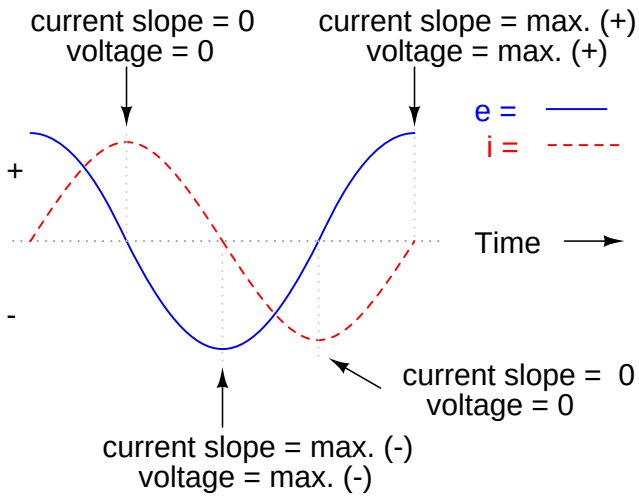
The expression *di/dt* is one from calculus, meaning the rate of change of instantaneous current (i) over time, in amps per second. The inductance (L) is in Henrys, and the instantaneous voltage (e), of course, is in volts. Sometimes you will find the rate of instantaneous voltage expressed as "v" instead of "e" (v = L di/dt), but it means the exact same thing. To show what happens with alternating current, let's analyze a simple inductor circuit:



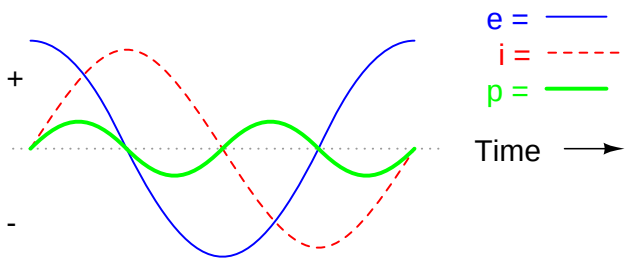
If we were to plot the current and voltage for this very simple circuit, it would look something like this:



Remember, the voltage dropped across an inductor is a reaction against the *change* in current through it. Therefore, the instantaneous voltage is zero whenever the instantaneous current is at a peak (zero change, or level slope, on the current sine wave), and the instantaneous voltage is at a peak wherever the instantaneous current is at maximum change (the points of steepest slope on the current wave, where it crosses the zero line). This results in a voltage wave that is 90^o out of phase with the current wave. Looking at the graph, the voltage wave seems to have a "head start" on the current wave; the voltage "leads" the current, and the current "lags" behind the voltage.



Things get even more interesting when we plot the power for this circuit:



Because instantaneous power is the product of the instantaneous voltage and the instantaneous current (p=ie), the power equals zero whenever the instantaneous current *or* voltage is zero. Whenever the instantaneous current and voltage are both positive (above the line), the power is positive. As with the resistor example, the power is also positive when the instantaneous current and voltage are both negative (below the line). However, because the current and voltage waves are 90^o out of phase, there are times when one is positive while the other is negative, resulting in equally frequent occurrences of *negative instantaneous power*.

But what does *negative* power mean? It means that the inductor is releasing power back to the circuit, while a positive power means that it is absorbing power from the circuit. Since the positive and negative power cycles are equal in magnitude and duration over time, the inductor releases just as much power back to the circuit as it absorbs over the span of a complete cycle. What this means in a practical sense is that the reactance of an inductor dissipates a net energy of zero, quite unlike the resistance of a resistor, which dissipates energy in the form of heat. Mind you, this is for perfect inductors only, which have no wire resistance.

An inductor's opposition to change in current translates to an opposition to alternating current in general, which is by definition always changing in instantaneous magnitude and direction. This opposition to alternating current is similar to resistance, but different in that it always results in a phase shift between current and voltage, and it dissipates zero power. Because of the differences, it has a different name: *reactance*. Reactance to AC is expressed in ohms, just like resistance is, except that its mathematical symbol is X instead of R. To be specific, reactance associate with an inductor is usually symbolized by the capital letter X with a letter L as a subscript, like this: X_L.

Since inductors drop voltage in proportion to the rate of current change, they will drop more voltage for faster-changing currents, and less voltage for slower-changing currents. What this means is that reactance in ohms for any inductor is directly proportional to the frequency of the alternating current. The exact formula for determining reactance is as follows:

$$X_L = 2\pi fL$$

If we expose a 10 mH inductor to frequencies of 60, 120, and 2500 Hz, it will manifest the following reactances:

For a 10 mH inductor:

Frequency (Hertz)	Reactance (Ohms)
60	3.7699
120	7.5398

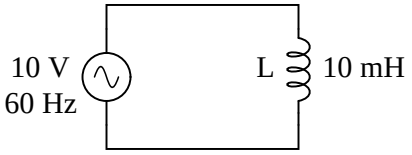
2500	157.0796

In the reactance equation, the term "2πf" (everything on the right-hand side except the L) has a special meaning unto itself. It is the number of radians per second that the alternating current is "rotating" at, if you imagine one cycle of AC to represent a full circle's rotation. A *radian* is a unit of angular measurement: there are 2π radians in one full circle, just as there are 360^o in a full circle. If the alternator producing the AC is a double-pole unit, it will produce one cycle for every full turn of shaft rotation, which is every 2π radians, or 360^o. If this constant of 2π is multiplied by frequency in Hertz (cycles per second), the result will be a figure in radians per second, known as the *angular velocity* of the AC system.

Angular velocity may be represented by the expression 2πf, or it may be represented by its own symbol, the lower-case Greek letter Omega, which appears similar to our Roman lower-case "w": ω. Thus, the reactance formula $X_L = 2\pi fL$ could also be written as $X_L = \omega L$.

It must be understood that this "angular velocity" is an expression of how rapidly the AC waveforms are cycling, a full cycle being equal to 2π radians. It is not necessarily representative of the actual shaft speed of the alternator producing the AC. If the alternator has more than two poles, the angular velocity will be a multiple of the shaft speed. For this reason, ω is sometimes expressed in units of *electrical* radians per second rather than (plain) radians per second, so as to distinguish it from mechanical motion.

Any way we express the angular velocity of the system, it is apparent that it is directly proportional to reactance in an inductor. As the frequency (or alternator shaft speed) is increased in an AC system, an inductor will offer greater opposition to the passage of current, and visa-versa. Alternating current in a simple inductive circuit is equal to the voltage (in volts) divided by the inductive reactance (in ohms), just as either alternating or direct current in a simple resistive circuit is equal to the voltage (in volts) divided by the resistance (in ohms). An example circuit is shown here:



(inductive reactance of 10 mH inductor at 60 Hz)

$X_L = 3.7699\ \Omega$

$$I = \frac{E}{X}$$

$$I = \frac{10\ V}{3.7699\ \Omega}$$

$I = 2.6526$ A

However, we need to keep in mind that voltage and current are not in phase here. As was shown earlier, the voltage has a phase shift of $+90^o$ with respect to the current. If we represent these phase angles of voltage and current mathematically in the form of complex numbers, we find that an inductor's opposition to current has a phase angle, too:

$$\text{Opposition} = \frac{\text{Voltage}}{\text{Current}}$$

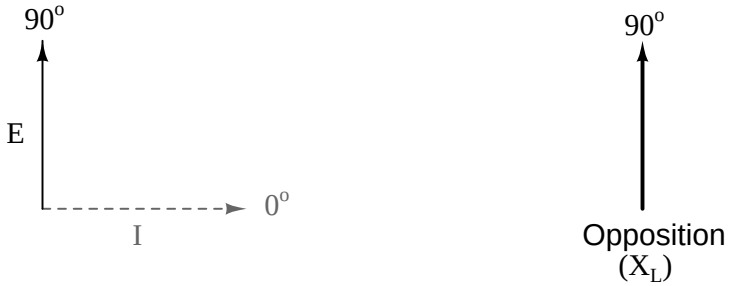
$$\text{Opposition} = \frac{10 \text{ V} \angle 90^o}{2.6526 \text{ A} \angle 0^o}$$

$$\text{Opposition} = 3.7699\ \Omega \angle 90^o$$

or

$$0 + j3.7699\ \Omega$$

For an inductor:



Mathematically, we say that the phase angle of an inductor's opposition to current is 90^o, meaning that an inductor's opposition to current is a positive imaginary quantity. This phase angle of reactive opposition to current becomes critically important in circuit analysis, especially for complex AC circuits where reactance and resistance interact. It will prove beneficial to represent *any* component's opposition to current in terms of complex numbers rather than scalar quantities of resistance and reactance.

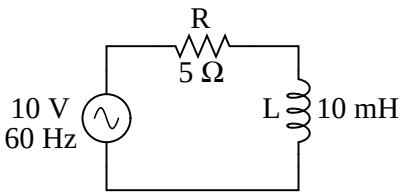
- **REVIEW:**
- *Inductive reactance* is the opposition that an inductor offers to alternating current due to its phase-shifted storage and release of energy in its magnetic field. Reactance is symbolized by the capital letter "X" and is measured in ohms just like resistance (R).
- Inductive reactance can be calculated using this formula: $X_L = 2\pi fL$
- The *angular velocity* of an AC circuit is another way of expressing its frequency, in units of electrical radians per second instead of cycles per second. It is symbolized by the lower-case Greek letter "omega," or ω.

- Inductive reactance *increases* with increasing frequency. In other words, the higher the frequency, the more it opposes the AC flow of electrons.

3.3 Series resistor-inductor circuits

In the previous section, we explored what would happen in simple resistor-only and inductor-only AC circuits. Now we will mix the two components together in series form and investigate the effects.

Take this circuit as an example to work with:



The resistor will offer 5 Ω of resistance to AC current regardless of frequency, while the inductor will offer 3.7699 Ω of reactance to AC current at 60 Hz. Because the resistor's resistance is a real number (5 Ω $\angle$ 0^o, or 5 + j0 Ω), and the inductor's reactance is an imaginary number (3.7699 Ω $\angle$ 90^o, or 0 + j3.7699 Ω), the combined effect of the two components will be an opposition to current equal to the complex sum of the two numbers. This combined opposition will be a vector combination of resistance and reactance. In order to express this opposition succinctly, we need a more comprehensive term for opposition to current than either resistance or reactance alone. This term is called *impedance*, its symbol is Z, and it is also expressed in the unit of ohms, just like resistance and reactance. In the above example, the total circuit impedance is:

$$Z_{total} = (5\ \Omega \text{ resistance}) + (3.7699\ \Omega \text{ inductive reactance})$$

$$Z_{total} = 5\ \Omega\ (R) \quad + \quad 3.7699\ \Omega\ (X_L)$$

$$Z_{total} = (5\ \Omega \angle 0^o) + (3.7699\ \Omega \angle 90^0)$$

or

$$(5 + j0\ \Omega) + (0 + j3.7699\ \Omega)$$

$$Z_{total} = 5 + j3.7699\ \Omega \qquad or \qquad 6.262\ \Omega \angle 37.016^o$$

Impedance is related to voltage and current just as you might expect, in a manner similar to resistance in Ohm's Law:

Ohm's Law for AC circuits:

$$\mathbf{E = IZ} \qquad \mathbf{I = \frac{E}{Z}} \qquad \mathbf{Z = \frac{E}{I}}$$

All quantities expressed in complex, not scalar, form

In fact, this is a far more comprehensive form of Ohm's Law than what was taught in DC electronics (E=IR), just as impedance is a far more comprehensive expression of opposition to the flow of electrons than resistance is. *Any* resistance and any reactance, separately or in combination (series/parallel), can be and should be represented as a single impedance in an AC circuit.

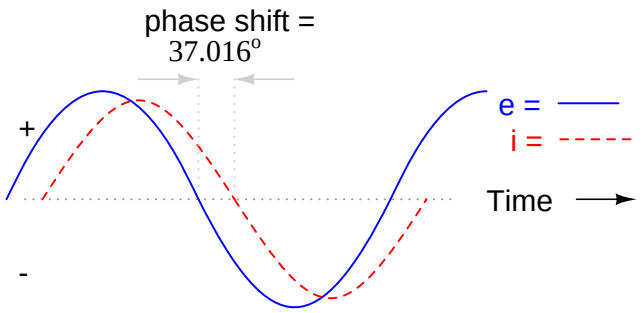
To calculate current in the above circuit, we first need to give a phase angle reference for the voltage source, which is generally assumed to be zero. (The phase angles of resistive and inductive impedance are *always* 0^o and $+90^o$, respectively, regardless of the given phase angles for voltage or current).

$$I = \frac{E}{Z}$$

$$I = \frac{10 \text{ V} \angle 0^o}{6.262\ \Omega \angle 37.016^o}$$

$$I = 1.597 \text{ A} \angle -37.016^o$$

As with the purely inductive circuit, the current wave lags behind the voltage wave (of the source), although this time the lag is not as great: only 37.016^o as opposed to a full 90^o as was the case in the purely inductive circuit.



For the resistor and the inductor, the phase relationships between voltage and current haven't changed. Across voltage across the resistor is in phase (0^o shift) with the current through it; and the voltage across the inductor is $+90^o$ out of phase with the current going through it. We can verify this mathematically:

$$E = IZ$$

$$E_R = I_R Z_R$$

$$E_R = (1.597 \text{ A} \angle -37.016^o)(5\ \Omega \angle 0^o)$$

$$E_R = 7.9847 \text{ V} \angle -37.016^o$$

Notice that the phase angle of E_R is equal to the phase angle of the current.

The voltage across the resistor has the exact same phase angle as the current through it, telling us that E and I are in phase (for the resistor only).

$$E = IZ$$

$$E_L = I_L Z_L$$

$$E_L = (1.597 \text{ A} \angle -37.016^o)(3.7699\ \Omega \angle 90^o)$$

$$E_L = 6.0203 \text{ V} \angle 52.984^o$$

Notice that the phase angle of E_L is exactly 90^o more than the phase angle of the current.

The voltage across the inductor has a phase angle of 52.984^o, while the current through the inductor has a phase angle of -37.016^o, a difference of exactly 90^o between the two. This tells us that E and I are still 90^o out of phase (for the inductor only).

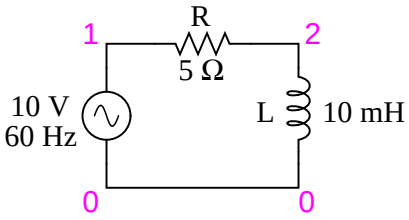
We can also mathematically prove that these complex values add together to make the total voltage, just as Kirchhoff's Voltage Law would predict:

$$E_{total} = E_R + E_L$$

$$E_{total} = (7.9847 \text{ V} \angle -37.016^o) + (6.0203 \text{ V} \angle 52.984^o)$$

$$E_{total} = 10 \text{ V} \angle 0^o$$

Let's check the validity of our calculations with SPICE:



```
ac r-l circuit
v1 1 0 ac 10 sin
r1 1 2 5
l1 2 0 10m
.ac lin 1 60 60
.print ac v(1,2) v(2,0) i(v1)
.print ac vp(1,2) vp(2,0) ip(v1)
.end

freq          v(1,2)      v(2)        i(v1)
6.000E+01     7.985E+00   6.020E+00   1.597E+00
freq          vp(1,2)     vp(2)       ip(v1)
6.000E+01    -3.702E+01   5.298E+01   1.430E+02
```

Interpreted SPICE results

$E_R = 7.985 \text{ V} \angle -37.02^o$

$E_L = 6.020 \text{ V} \angle 52.98^o$

$I = 1.597 \text{ A} \angle 143.0^o$

Note that just as with DC circuits, SPICE outputs current figures as though they were negative (180^o out of phase) with the supply voltage. Instead of a phase angle of -37.016^o, we get a current phase angle of 143^o (-37^o + 180^o). This is merely an idiosyncrasy of SPICE and does not represent anything significant in the circuit simulation itself. Note how both the resistor and inductor voltage phase readings match our calculations (-37.02^o and 52.98^o, respectively), just as we expected them to.

With all these figures to keep track of for even such a simple circuit as this, it would be beneficial for us to use the "table" method. Applying a table to this simple series resistor-inductor circuit would proceed as such. First, draw up a table for E/I/Z figures and insert all component values in these terms (in other words, don't insert actual resistance or inductance values in Ohms and Henrys, respectively, into the table; rather, convert them into complex figures of impedance and write those in):

	R	L	Total	
E			**10 + j0** **10** $\angle 0^o$	Volts
I				Amps
Z	**5 + j0** **5** $\angle 0^o$	**0 + j3.7699** **3.7699** $\angle 90^o$		Ohms

Although it isn't necessary, I find it helpful to write *both* the rectangular and polar forms of each quantity in the table. If you are using a calculator that has the ability to perform complex

arithmetic without the need for conversion between rectangular and polar forms, then this extra documentation is completely unnecessary. However, if you are forced to perform complex arithmetic "longhand" (addition and subtraction in rectangular form, and multiplication and division in polar form), writing each quantity in both forms will be useful indeed.

Now that our "given" figures are inserted into their respective locations in the table, we can proceed just as with DC: determine the total impedance from the individual impedances. Since this is a series circuit, we know that opposition to electron flow (resistance *or* impedance) adds to form the total opposition:

	R	L	Total	
E			$10 + j0$ $10 \angle 0^o$	Volts
I				Amps
Z	$5 + j0$ $5 \angle 0^o$	$0 + j3.7699$ $3.7699 \angle 90^o$	**5 + j3.7699** **6.262** $\angle$ 37.016^o	Ohms

Rule of series circuits

$Z_{total} = Z_R + Z_L$

Now that we know total voltage and total impedance, we can apply Ohm's Law (I=E/Z) to determine total current:

	R	L	Total	
E			$10 + j0$ $10 \angle 0^o$	Volts
I			**1.2751 - j0.9614** **1.597** $\angle$ -37.016^o	Amps
Z	$5 + j0$ $5 \angle 0^o$	$0 + j3.7699$ $3.7699 \angle 90^o$	$5 + j3.7699$ $6.262 \angle 37.016^o$	Ohms

Ohm's Law

$I = \frac{E}{Z}$

Just as with DC, the total current in a series AC circuit is shared equally by all components. This is still true because in a series circuit there is only a single path for electrons to flow, therefore the rate of their flow must uniform throughout. Consequently, we can transfer the figures for current into the columns for the resistor and inductor alike:

	R	L	Total	
E			10 + j0 $10 \angle 0^o$	Volts
I	**1.2751 - j0.9614** **1.597** $\angle$ -37.016^o	**1.2751 - j0.9614** **1.597** $\angle$ -37.016^o	1.2751 - j0.9614 1.597 $\angle$ -37.016^o	Amps
Z	5 + j0 $5 \angle 0^o$	0 + j3.7699 3.7699 $\angle 90^o$	5 + j3.7699 6.262 $\angle$ 37.016^o	Ohms

Rule of series circuits:

$I_{total} = I_R = I_L$

Now all that's left to figure is the voltage drop across the resistor and inductor, respectively. This is done through the use of Ohm's Law (E=IZ), applied vertically in each column of the table:

	R	L	Total	
E	**6.3756 - j4.8071** **7.9847** $\angle$ -37.016^o	**3.6244 + j4.8071** **6.0203** $\angle$ 52.984^o	10 + j0 $10 \angle 0^o$	Volts
I	1.2751 - j0.9614 1.597 $\angle$ -37.016^o	1.2751 - j0.9614 1.597 $\angle$ -37.016^o	1.2751 - j0.9614 1.597 $\angle$ -37.016^o	Amps
Z	5 + j0 $5 \angle 0^o$	0 + j3.7699 3.7699 $\angle 90^o$	5 + j3.7699 6.262 $\angle$ 37.016^o	Ohms

Ohm's Law E = IZ

Ohm's Law E = IZ

And with that, our table is complete. The exact same rules we applied in the analysis of DC circuits apply to AC circuits as well, with the caveat that all quantities must be represented and calculated in complex rather than scalar form. So long as phase shift is properly represented in our calculations, there is no fundamental difference in how we approach basic AC circuit analysis versus DC.

Now is a good time to review the relationship between these calculated figures and readings given by actual instrument measurements of voltage and current. The figures here that directly relate to real-life measurements are those in *polar notation*, not rectangular! In other words, if you were to connect a voltmeter across the resistor in this circuit, it would indicate **7.9847** volts, not 6.3756 (real rectangular) or 4.8071 (imaginary rectangular) volts. To describe this in graphical terms, measurement instruments simply tell you how long the vector is for that particular quantity (voltage or current).

Rectangular notation, while convenient for arithmetical addition and subtraction, is a more abstract form of notation than polar in relation to real-world measurements. As I stated before, I will indicate both polar and rectangular forms of each quantity in my AC circuit tables simply for convenience of mathematical calculation. This is not absolutely necessary, but may be helpful for those following along without the benefit of an advanced calculator. If we were restrict ourselves to

the use of only one form of notation, the best choice would be polar, because it is the only one that can be directly correlated to real measurements.

- **REVIEW:**

- *Impedance* is the total measure of opposition to electric current and is the complex (vector) sum of ("real") resistance and ("imaginary") reactance. It is symbolized by the letter "Z" and measured in ohms, just like resistance (R) and reactance (X).

- Impedances (Z) are managed just like resistances (R) in series circuit analysis: series impedances add to form the total impedance. Just be sure to perform all calculations in complex (not scalar) form! $Z_{Total} = Z_1 + Z_2 + \ldots Z_n$

- A purely resistive impedance will always have a phase angle of exactly 0^o ($Z_R = R\ \Omega\ \angle\ 0^o$).

- A purely inductive impedance will always have a phase angle of exactly $+90^o$ ($Z_L = X_L\ \Omega\ \angle\ 90^o$).

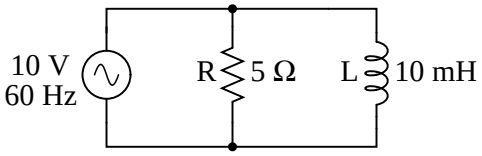
- Ohm's Law for AC circuits: E = IZ ; I = E/Z ; Z = E/I

- When resistors and inductors are mixed together in circuits, the total impedance will have a phase angle somewhere between 0^o and $+90^o$. The circuit current will have a phase angle somewhere between 0^o and -90^o.

- Series AC circuits exhibit the same fundamental properties as series DC circuits: current is uniform throughout the circuit, voltage drops add to form the total voltage, and impedances add to form the total impedance.

3.4 Parallel resistor-inductor circuits

Let's take the same components for our series example circuit and connect them in parallel:



Because the power source has the same frequency as the series example circuit, and the resistor and inductor both have the same values of resistance and inductance, respectively, they must also have the same values of impedance. So, we can begin our analysis table with the same "given" values:

	R	L	Total	
E			**10 + j0** **10** $\angle$ 0°	Volts
I				Amps
Z	**5 + j0** **5** $\angle$ 0°	**0 + j3.7699** **3.7699** $\angle$ 90°		Ohms

The only difference in our analysis technique this time is that we will apply the rules of parallel circuits instead of the rules for series circuits. The approach is fundamentally the same as for DC. We know that voltage is shared uniformly by all components in a parallel circuit, so we can transfer the figure of total voltage (10 volts $\angle$ 0°) to all components columns:

	R	L	Total	
E	10 + j0 10 $\angle$ 0°	10 + j0 10 $\angle$ 0°	10 + j0 10 $\angle$ 0°	Volts
I				Amps
Z	5 + j0 5 $\angle$ 0°	0 + j3.7699 3.7699 $\angle$ 90°		Ohms

Rule of parallel circuits:

$E_{total} = E_R = E_L$

Now we can apply Ohm's Law (I=E/Z) vertically to two columns of the table, calculating current through the resistor and current through the inductor:

	R	L	Total	
E	10 + j0 10 $\angle$ 0°	10 + j0 10 $\angle$ 0°	10 + j0 10 $\angle$ 0°	Volts
I	**2 + j0** **2** $\angle$ 0°	**0 - j2.6526** **2.6526** $\angle$ -90°		Amps
Z	5 + j0 5 $\angle$ 0°	0 + j3.7699 3.7699 $\angle$ 90°		Ohms

Ohm's Law $I = \frac{E}{Z}$ *Ohm's Law* $I = \frac{E}{Z}$

Just as with DC circuits, branch currents in a parallel AC circuit add to form the total current (Kirchhoff's Current Law still holds true for AC as it did for DC):

	R	L	Total	
E	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	Volts
I	2 + j0 $2 \angle 0^o$	0 - j2.6526 $2.6526 \angle -90^o$	**2 - j2.6526** **3.3221** $\angle$ -52.984^o	Amps
Z	5 + j0 $5 \angle 0^o$	0 + j3.7699 $3.7699 \angle 90^o$		Ohms

Rule of parallel circuits:

$I_{total} = I_R + I_L$

Finally, total impedance can be calculated by using Ohm's Law (Z=E/I) vertically in the "Total" column. Incidentally, parallel impedance can also be calculated by using a reciprocal formula identical to that used in calculating parallel resistances.

$$Z_{parallel} = \frac{1}{\frac{1}{Z_1} + \frac{1}{Z_2} + \ldots \frac{1}{Z_n}}$$

The only problem with using this formula is that it typically involves a lot of calculator keystrokes to carry out. And if you're determined to run through a formula like this "longhand," be prepared for a very large amount of work! But, just as with DC circuits, we often have multiple options in calculating the quantities in our analysis tables, and this example is no different. No matter which way you calculate total impedance (Ohm's Law or the reciprocal formula), you will arrive at the same figure:

	R	L	Total	
E	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	Volts
I	2 + j0 $2 \angle 0^o$	0 - j2.6526 $2.6526 \angle -90^o$	2 - j2.6526 $3.322 \angle -52.984^o$	Amps
Z	5 + j0 $5 \angle 0^o$	0 + j3.7699 $3.7699 \angle 90^o$	**1.8122 + j2.4035** **3.0102** $\angle$ 52.984^o	Ohms

Ohm's Law $Z = \frac{E}{I}$ *or* *Rule of parallel circuits:* $Z_{total} = \frac{1}{\frac{1}{Z_R} + \frac{1}{Z_L}}$

- **REVIEW:**
- Impedances (Z) are managed just like resistances (R) in parallel circuit analysis: parallel impedances diminish to form the total impedance, using the reciprocal formula. Just be sure

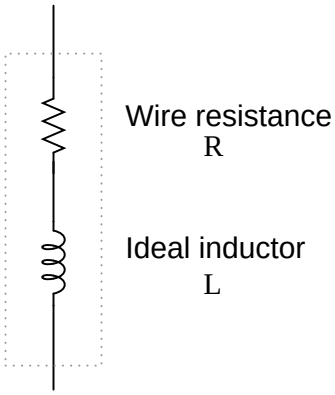
to perform all calculations in complex (not scalar) form! $Z_{Total} = 1/(1/Z_1 + 1/Z_2 + \ldots 1/Z_n)$

- Ohm's Law for AC circuits: E = IZ ; I = E/Z ; Z = E/I
- When resistors and inductors are mixed together in parallel circuits (just as in series circuits), the total impedance will have a phase angle somewhere between 0^o and $+90^o$. The circuit current will have a phase angle somewhere between 0^o and -90^o.
- Parallel AC circuits exhibit the same fundamental properties as parallel DC circuits: voltage is uniform throughout the circuit, branch currents add to form the total current, and impedances diminish (through the reciprocal formula) to form the total impedance.

3.5 Inductor quirks

In an ideal case, an inductor acts as a purely reactive device. That is, its opposition to AC current is strictly based on inductive reaction to changes in current, and not electron friction as is the case with resistive components. However, inductors are not quite so pure in their reactive behavior. To begin with, they're made of wire, and we know that all wire possesses some measurable amount of resistance (unless it's superconducting wire). This built-in resistance acts as though it were connected in series with the perfect inductance of the coil, like this:

Equivalent circuit for a real inductor



Consequently, the impedance of any real inductor will always be a complex combination of resistance and inductive reactance.

Compounding this problem is something called the *skin effect*, which is AC's tendency to flow through the outer areas of a conductor's cross-section rather than through the middle. When electrons flow in a single direction (DC), they use the entire cross-sectional area of the conductor to move. Electrons switching directions of flow, on the other hand, tend to avoid travel through the very middle of a conductor, limiting the effective cross-sectional area available. The skin effect becomes more pronounced as frequency increases.

Also, the alternating magnetic field of an inductor energized with AC may radiate off into space as part of an electromagnetic wave, especially if the AC is of high frequency. This radiated energy

does not return to the inductor, and so it manifests itself as resistance (power dissipation) in the circuit.

Added to the resistive losses of wire and radiation, there are other effects at work in iron-core inductors which manifest themselves as additional resistance between the leads. When an inductor is energized with AC, the alternating magnetic fields produced tend to induce circulating currents within the iron core known as *eddy currents*. These electric currents in the iron core have to overcome the electrical resistance offered by the iron, which is not as good a conductor as copper. Eddy current losses are primarily counteracted by dividing the iron core up into many thin sheets (laminations), each one separated from the other by a thin layer of electrically insulating varnish. With the cross-section of the core divided up into many electrically isolated sections, current cannot circulate within that cross-sectional area and there will be no (or very little) resistive losses from that effect.

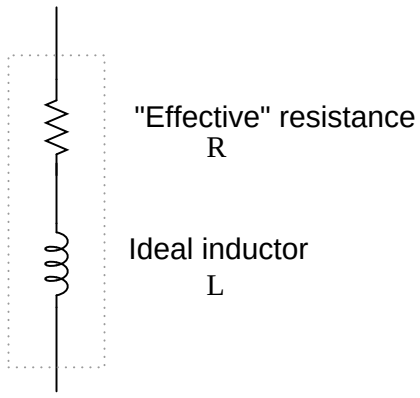
As you might have expected, eddy current losses in metallic inductor cores manifest themselves in the form of heat. The effect is more pronounced at higher frequencies, and can be so extreme that it is sometimes exploited in manufacturing processes to heat metal objects! In fact, this process of "inductive heating" is often used in high-purity metal foundry operations, where metallic elements and alloys must be heated in a vacuum environment to avoid contamination by air, and thus where standard combustion heating technology would be useless. It is a "non-contact" technology, the heated substance not having to touch the coil(s) producing the magnetic field.

In high-frequency service, eddy currents can even develop within the cross-section of the wire itself, contributing to additional resistive effects. To counteract this tendency, special wire made of very fine, individually insulated strands called *Litz wire* (short for *Litzendraht*) can be used. The insulation separating strands from each other prevent eddy currents from circulating through the whole wire's cross-sectional area.

Additionally, any magnetic hysteresis that needs to be overcome with every reversal of the inductor's magnetic field constitutes an expenditure of energy that manifests itself as resistance in the circuit. Some core materials (such as ferrite) are particularly notorious for their hysteretic effect. Counteracting this effect is best done by means of proper core material selection and limits on the peak magnetic field intensity generated with each cycle.

Altogether, the stray resistive properties of a real inductor (wire resistance, radiation losses, eddy currents, and hysteresis losses) are expressed under the single term of "effective resistance:"

Equivalent circuit for a real inductor



It is worthy to note that the skin effect and radiation losses apply just as well to straight lengths

of wire in an AC circuit as they do a coiled wire. Usually their combined effect is too small to notice, but at radio frequencies they can be quite large. A radio transmitter antenna, for example, is designed with the express purpose of dissipating the greatest amount of energy in the form of electromagnetic radiation.

Effective resistance in an inductor can be a serious consideration for the AC circuit designer. To help quantify the relative amount of effective resistance in an inductor, another value exists called the *Q factor*, or "quality factor" which is calculated as follows:

$$Q = \frac{X_L}{R}$$

The symbol "Q" has nothing to do with electric charge (coulombs), which tends to be confusing. For some reason, the Powers That Be decided to use the same letter of the alphabet to denote a totally different quantity.

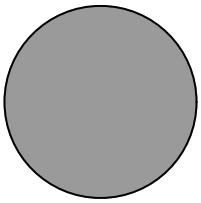
The higher the value for "Q," the "purer" the inductor is. Because it's so easy to add additional resistance if needed, a high-Q inductor is better than a low-Q inductor for design purposes. An ideal inductor would have a Q of infinity, with zero effective resistance.

Because inductive reactance (X) varies with frequency, so will Q. However, since the resistive effects of inductors (wire skin effect, radiation losses, eddy current, and hysteresis) also vary with frequency, Q does not vary proportionally with reactance. In order for a Q value to have precise meaning, it must be specified at a particular test frequency.

Stray resistance isn't the only inductor quirk we need to be aware of. Due to the fact that the multiple turns of wire comprising inductors are separated from each other by an insulating gap (air, varnish, or some other kind of electrical insulation), we have the potential for capacitance to develop between turns. AC capacitance will be explored in the next chapter, but it suffices to say at this point that it behaves very differently from AC inductance, and therefore further "taints" the reactive purity of real inductors.

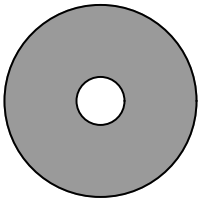
3.6 More on the "skin effect"

As previously mentioned, the skin effect is where alternating current tends to avoid travel through the center of a solid conductor, limiting itself to conduction near the surface. This effectively limits the cross-sectional conductor area available to carry alternating electron flow, increasing the resistance of that conductor above what it would normally be for direct current:



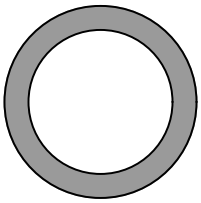
Cross-sectional area of a round conductor available for conducting DC current

"DC resistance"



Cross-sectional area of the same conductor available for conducting low-frequency AC

"AC resistance"



Cross-sectional area of the same conductor available for conducting high-frequency AC

"AC resistance"

The electrical resistance of the conductor with all its cross-sectional area in use is known as the "DC resistance," the "AC resistance" of the same conductor referring to a higher figure resulting from the skin effect. As you can see, at high frequencies the AC current avoids travel through most of the conductor's cross-sectional area. For the purpose of conducting current, the wire might as well be hollow!

In some radio applications (antennas, most notably) this effect is exploited. Since radio-frequency ("RF") AC currents wouldn't travel through the middle of a conductor anyway, why not just use hollow metal rods instead of solid metal wires and save both weight and cost? Most antenna structures and RF power conductors are made of hollow metal tubes for this reason.

In the following photograph you can see some large inductors used in a 50 kW radio transmitting circuit. The inductors are hollow copper tubes coated with silver, for excellent conductivity at the "skin" of the tube:

The degree to which frequency affects the effective resistance of a solid wire conductor is impacted by the gauge of that wire. As a rule, large-gauge wires exhibit a more pronounced skin effect (change in resistance from DC) than small-gauge wires at any given frequency. The equation for approximating skin effect at high frequencies (greater than 1 MHz) is as follows:

$$R_{AC} = (R_{DC})(k)\sqrt{f}$$

Where,

R_{AC} = AC resistance at given frequency "f"

R_{DC} = Resistance at zero frequency (DC)

k = Wire gage factor (see table below)

f = Frequency of AC in MHz (MegaHertz)

The following table gives approximate values of "k" factor for various round wire sizes:

Gage size	k factor
4/0	124.5
2/0	99.0

```
1/0 ---------- 88.0
2 ------------ 69.8
4 ------------ 55.5
6 ------------ 47.9
8 ------------ 34.8
10 ----------- 27.6
14 ----------- 17.6
18 ----------- 10.9
22 ----------- 6.86
```

For example, a length of number 10-gauge wire with a DC end-to-end resistance of 25 Ω would have an AC (effective) resistance of 2.182 kΩ at a frequency of 10 MHz:

$$R_{AC} = (R_{DC})(k)\sqrt{f}$$

$$R_{AC} = (25\ \Omega)(27.6)\ \sqrt{10}$$

$$R_{AC} = 2.182\ k\Omega$$

Please remember that this figure is *not* impedance, and it does *not* consider any reactive effects, inductive or capacitive. This is simply an estimated figure of pure resistance for the conductor (that opposition to the AC flow of electrons which *does* dissipate power in the form of heat), corrected for skin effect. Reactance, and the combined effects of reactance and resistance (impedance), are entirely different matters.

3.7 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Jim Palmer (June 2001): Identified and offered correction for typographical error in complex number calculation.

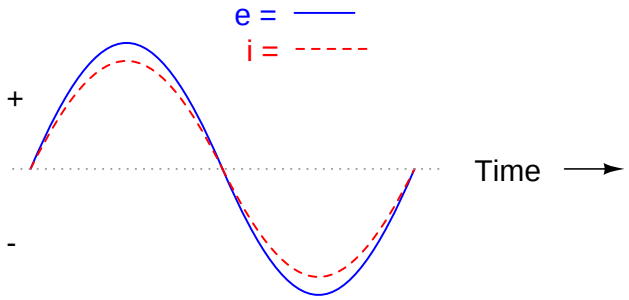
Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 4

REACTANCE AND IMPEDANCE – CAPACITIVE

4.1 AC resistor circuits

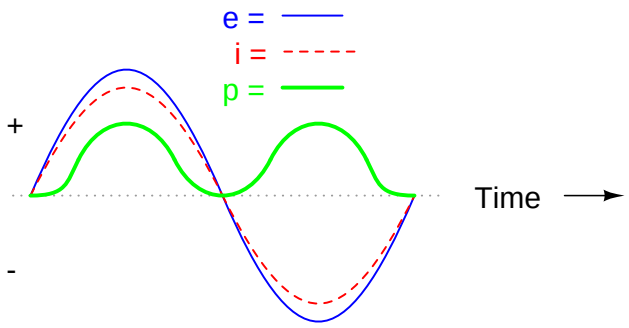
If we were to plot the current and voltage for a very simple AC circuit consisting of a source and a resistor, it would look something like this:



Because the resistor allows an amount of current directly proportional to the voltage across it at all periods of time, the waveform for the current is exactly in phase with the waveform for the voltage. We can look at any point in time along the horizontal axis of the plot and compare those values of current and voltage with each other (any "snapshot" look at the values of a wave are referred to as *instantaneous values*, meaning the values at that *instant* in time). When the instantaneous value for voltage is zero, the instantaneous current through the resistor is also zero. Likewise, at the moment in time where the voltage across the resistor is at its positive peak, the current through the resistor is also at its positive peak, and so on. At any given point in time along the waves, Ohm's Law holds

true for the instantaneous values of voltage and current.

We can also calculate the power dissipated by this resistor, and plot those values on the same graph:



Note that the power is never a negative value. When the current is positive (above the line), the voltage is also positive, resulting in a power (p=ie) of a positive value. Conversely, when the current is negative (below the line), the voltage is also negative, which results in a positive value for power (a negative number multiplied by a negative number equals a positive number). This consistent "polarity" of power tells us that the resistor is always dissipating power, taking it from the source and releasing it in the form of heat energy. Whether the current is positive or negative, a resistor still dissipates energy.

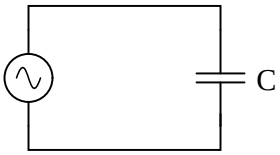
4.2 AC capacitor circuits

Capacitors do not behave the same as resistors. Whereas resistors allow a flow of electrons through them directly proportional to the voltage drop, capacitors oppose *changes* in voltage by drawing or supplying current as they charge or discharge to the new voltage level. The flow of electrons "through" a capacitor is directly proportional to the *rate of change* of voltage across the capacitor. This opposition to voltage change is another form of *reactance*, but one that is precisely opposite to the kind exhibited by inductors.

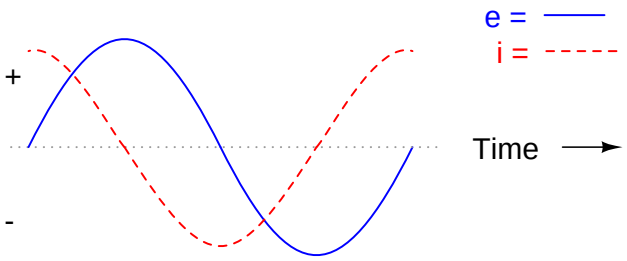
Expressed mathematically, the relationship between the current "through" the capacitor and rate of voltage change across the capacitor is as such:

$$i = C \frac{de}{dt}$$

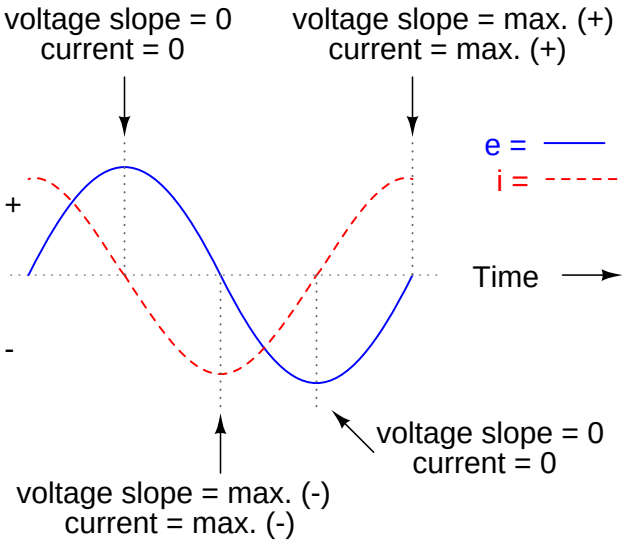
The expression *de/dt* is one from calculus, meaning the rate of change of instantaneous voltage (e) over time, in volts per second. The capacitance (C) is in Farads, and the instantaneous current (i), of course, is in amps. Sometimes you will find the rate of instantaneous voltage change over time expressed as dv/dt instead of de/dt: using the lower-case letter "v" instead or "e" to represent voltage, but it means the exact same thing. To show what happens with alternating current, let's analyze a simple capacitor circuit:



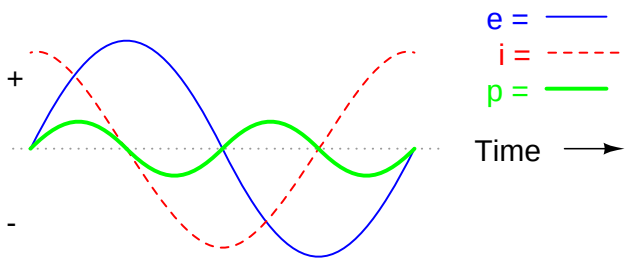
If we were to plot the current and voltage for this very simple circuit, it would look something like this:



Remember, the current through a capacitor is a reaction against the *change* in voltage across it. Therefore, the instantaneous current is zero whenever the instantaneous voltage is at a peak (zero change, or level slope, on the voltage sine wave), and the instantaneous current is at a peak wherever the instantaneous voltage is at maximum change (the points of steepest slope on the voltage wave, where it crosses the zero line). This results in a voltage wave that is -90^o out of phase with the current wave. Looking at the graph, the current wave seems to have a "head start" on the voltage wave; the current "leads" the voltage, and the voltage "lags" behind the current.



As you might have guessed, the same unusual power wave that we saw with the simple inductor circuit is present in the simple capacitor circuit, too:



As with the simple inductor circuit, the 90 degree phase shift between voltage and current results in a power wave that alternates equally between positive and negative. This means that a capacitor does not dissipate power as it reacts against changes in voltage; it merely absorbs and releases power, alternately.

A capacitor's opposition to change in voltage translates to an opposition to alternating voltage in general, which is by definition always changing in instantaneous magnitude and direction. For any given magnitude of AC voltage at a given frequency, a capacitor of given size will "conduct" a certain magnitude of AC current. Just as the current through a resistor is a function of the voltage across the resistor and the resistance offered by the resistor, the AC current through a capacitor is a function of the AC voltage across it, and the *reactance* offered by the capacitor. As with inductors, the reactance of a capacitor is expressed in ohms and symbolized by the letter X (or X_C to be more specific).

Since capacitors "conduct" current in proportion to the rate of voltage change, they will pass more current for faster-changing voltages (as they charge and discharge to the same voltage peaks in less time), and less current for slower-changing voltages. What this means is that reactance in ohms for any capacitor is *inversely* proportional to the frequency of the alternating current:

$$X_C = \frac{1}{2\pi fC}$$

For a 100 uF capacitor:

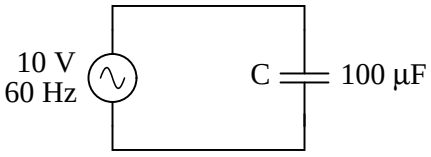
Frequency (Hertz)	Reactance (Ohms)
60	26.5258
120	13.2629
2500	0.6366

Please note that the relationship of capacitive reactance to frequency is exactly opposite from that of inductive reactance. Capacitive reactance (in ohms) decreases with increasing AC frequency. Conversely, inductive reactance (in ohms) increases with increasing AC frequency. Inductors oppose faster changing currents by producing greater voltage drops; capacitors oppose faster changing voltage drops by allowing greater currents.

As with inductors, the reactance equation's $2\pi f$ term may be replaced by the lower-case Greek letter Omega (ω), which is referred to as the *angular velocity* of the AC circuit. Thus, the equation

$X_C = 1/(2\pi fC)$ could also be written as $X_C = 1/(\omega C)$, with ω cast in units of *radians per second*.

Alternating current in a simple capacitive circuit is equal to the voltage (in volts) divided by the capacitive reactance (in ohms), just as either alternating or direct current in a simple resistive circuit is equal to the voltage (in volts) divided by the resistance (in ohms). The following circuit illustrates this mathematical relationship by example:



$$X_C = 26.5258\ \Omega$$

$$I = \frac{E}{X}$$

$$I = \frac{10\ V}{26.5258\ \Omega}$$

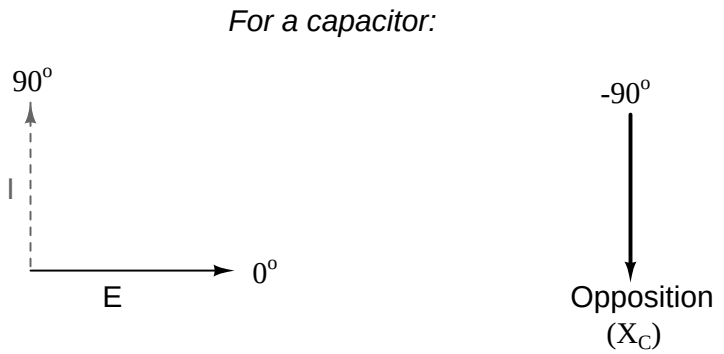
$$I = 0.3770\ A$$

However, we need to keep in mind that voltage and current are not in phase here. As was shown earlier, the current has a phase shift of $+90^o$ with respect to the voltage. If we represent these phase angles of voltage and current mathematically, we can calculate the phase angle of the inductor's reactive opposition to current.

$$\text{Opposition} = \frac{\text{Voltage}}{\text{Current}}$$

$$\text{Opposition} = \frac{10\ V \angle 0^o}{0.3770\ A \angle 90^o}$$

$$\text{Opposition} = 26.5258\ \Omega \angle -90^o$$



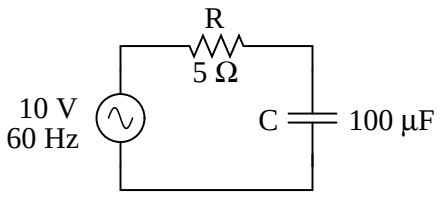
Mathematically, we say that the phase angle of a capacitor's opposition to current is -90^o, meaning that a capacitor's opposition to current is a negative imaginary quantity. This phase angle of reactive opposition to current becomes critically important in circuit analysis, especially for complex AC circuits where reactance and resistance interact. It will prove beneficial to represent *any* component's opposition to current in terms of complex numbers, and not just scalar quantities of resistance and reactance.

- **REVIEW:**
- *Capacitive reactance* is the opposition that a capacitor offers to alternating current due to its phase-shifted storage and release of energy in its electric field. Reactance is symbolized by the capital letter "X" and is measured in ohms just like resistance (R).
- Capacitive reactance can be calculated using this formula: $X_C = 1/(2\pi fC)$
- Capacitive reactance *decreases* with increasing frequency. In other words, the higher the frequency, the less it opposes (the more it "conducts") the AC flow of electrons.

4.3 Series resistor-capacitor circuits

In the last section, we learned what would happen in simple resistor-only and capacitor-only AC circuits. Now we will combine the two components together in series form and investigate the effects.

Take this circuit as an example to analyze:



The resistor will offer 5 Ω of resistance to AC current regardless of frequency, while the capacitor will offer 26.5258 Ω of reactance to AC current at 60 Hz. Because the resistor's resistance is a real number (5 Ω $\angle$ 0^o, or 5 + j0 Ω), and the capacitor's reactance is an imaginary number (26.5258 Ω $\angle$ -90^o, or 0 - j26.5258 Ω), the combined effect of the two components will be an opposition to current equal to the complex sum of the two numbers. The term for this complex opposition to current

is *impedance*, its symbol is Z, and it is also expressed in the unit of ohms, just like resistance and reactance. In the above example, the total circuit impedance is:

$$Z_{total} = (5\ \Omega \text{ resistance}) + (26.5258\ \Omega \text{ capacitive reactance})$$

$$Z_{total} = 5\ \Omega\ (R) + 26.5258\ \Omega\ (X_C)$$

$$Z_{total} = (5\ \Omega \angle 0^o) + (26.5258\ \Omega \angle -90^o)$$

or

$$(5 + j0\ \Omega) + (0 - j26.5258\ \Omega)$$

$$Z_{total} = 5 - j26.5258\ \Omega \qquad \textit{or} \qquad 26.993\ \Omega \angle -79.325^o$$

Impedance is related to voltage and current just as you might expect, in a manner similar to resistance in Ohm's Law:

Ohm's Law for AC circuits:

$$\mathbf{E = IZ} \qquad \mathbf{I = \frac{E}{Z}} \qquad \mathbf{Z = \frac{E}{I}}$$

All quantities expressed in complex, not scalar, form

In fact, this is a far more comprehensive form of Ohm's Law than what was taught in DC electronics (E=IR), just as impedance is a far more comprehensive expression of opposition to the flow of electrons than simple resistance is. Any resistance and any reactance, separately or in combination (series/parallel), can be and should be represented as a single impedance.

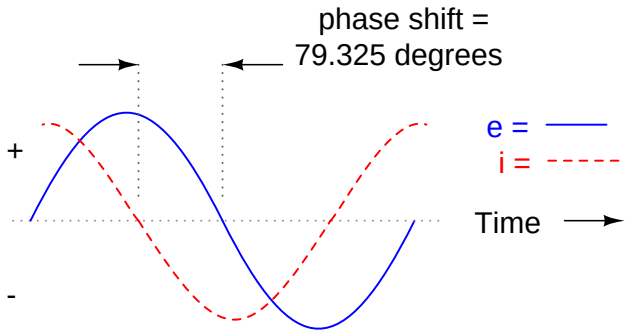
To calculate current in the above circuit, we first need to give a phase angle reference for the voltage source, which is generally assumed to be zero. (The phase angles of resistive and capacitive impedance are *always* 0^o and -90^o, respectively, regardless of the given phase angles for voltage or current).

$$I = \frac{E}{Z}$$

$$I = \frac{10\ V \angle 0^o}{26.933\ \Omega \angle -79.325^o}$$

$$I = 370.5\ mA \angle 79.325^o$$

As with the purely capacitive circuit, the current wave is leading the voltage wave (of the source), although this time the difference is 79.325^o instead of a full 90^o.



As we learned in the AC inductance chapter, the "table" method of organizing circuit quantities is a very useful tool for AC analysis just as it is for DC analysis. Let's place out known figures for this series circuit into a table and continue the analysis using this tool:

	R	C	Total	
E			10 + j0 10 ∠ 0°	Volts
I			68.623m + j364.06m 370.5m ∠ 79.325°	Amps
Z	5 + j0 5 ∠ 0°	0 - j26.5258 26.5258 ∠ -90°	5 - j26.5258 26.993 ∠ -79.325°	Ohms

Current in a series circuit is shared equally by all components, so the figures placed in the "Total" column for current can be distributed to all other columns as well:

	R	C	Total	
E			10 + j0 10 ∠ 0°	Volts
I	**68.623m + j364.06m** **370.5m** ∠ 79.325°	**68.623m + j364.06m** **370.5m** ∠ 79.325°	68.623m + j364.06m 370.5m ∠ 79.325°	Amps
Z	5 + j0 5 ∠ 0°	0 - j26.5258 26.5258 ∠ -90°	5 - j26.5258 26.993 ∠ -79.325°	Ohms

Rule of series circuits:

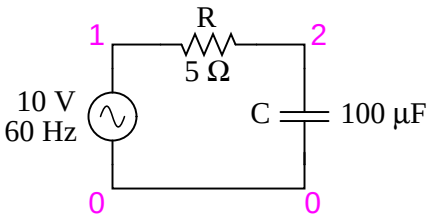
$I_{total} = I_R = I_C$

Continuing with our analysis, we can apply Ohm's Law (E=IR) vertically to determine voltage across the resistor and capacitor:

	R	C	Total	
E	**343.11m + j1.8203** **1.8523** ∠ 79.325°	**9.6569 - j1.8203** **9.8269** ∠ -10.675°	10 + j0 10 ∠ 0°	Volts
I	68.623m + j364.06m 370.5m ∠ 79.325°	68.623m + j364.06m 370.5m ∠ 79.325°	68.623m + j364.06m 370.5m ∠ 79.325°	Amps
Z	5 + j0 5 ∠ 0°	0 - j26.5258 26.5258 ∠ -90°	5 - j26.5258 26.993 ∠ -79.325°	Ohms
	Ohm's Law E = IZ	*Ohm's Law* E = IZ		

Notice how the voltage across the resistor has the exact same phase angle as the current through it, telling us that E and I are in phase (for the resistor only). The voltage across the capacitor has a phase angle of -10.675°, exactly 90° *less* than the phase angle of the circuit current. This tells us that the capacitor's voltage and current are still 90° out of phase with each other.

Let's check our calculations with SPICE:




```
ac r-c circuit
v1 1 0 ac 10 sin
r1 1 2 5
c1 2 0 100u
.ac lin 1 60 60
.print ac v(1,2) v(2,0) i(v1)
.print ac vp(1,2) vp(2,0) ip(v1)
.end
```

```
freq          v(1,2)      v(2)         i(v1)
6.000E+01     1.852E+00   9.827E+00    3.705E-01
freq          vp(1,2)     vp(2)        ip(v1)
6.000E+01     7.933E+01  -1.067E+01   -1.007E+02
```

Interpreted SPICE results

$E_R = 1.852 \text{ V} \angle 79.33^o$

$E_C = 9.827 \text{ V} \angle -10.67^o$

$I = 370.5 \text{ mA} \angle -100.7^o$

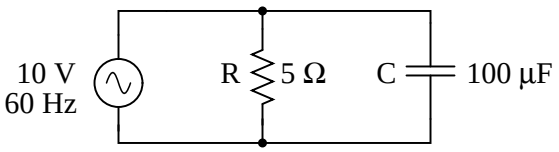
Once again, SPICE confusingly prints the current phase angle at a value equal to the real phase angle plus 180^o (or minus 180^o). However, it's a simple matter to correct this figure and check to see if our work is correct. In this case, the -100.7^o output by SPICE for current phase angle equates to a positive 79.3^o, which does correspond to our previously calculated figure of 79.325^o.

Again, it must be emphasized that the calculated figures corresponding to real-life voltage and current measurements are those in *polar* form, not rectangular form! For example, if we were to actually build this series resistor-capacitor circuit and measure voltage across the resistor, our voltmeter would indicate **1.8523** volts, not 343.11 millivolts (real rectangular) or 1.8203 volts (imaginary rectangular). Real instruments connected to real circuits provide indications corresponding to the vector length (magnitude) of the calculated figures. While the rectangular form of complex number notation is useful for performing addition and subtraction, it is a more abstract form of notation than polar, which alone has direct correspondence to true measurements.

- **REVIEW:**
- *Impedance* is the total measure of opposition to electric current and is the complex (vector) sum of ("real") resistance and ("imaginary") reactance.
- Impedances (Z) are managed just like resistances (R) in series circuit analysis: series impedances add to form the total impedance. Just be sure to perform all calculations in complex (not scalar) form! $Z_{Total} = Z_1 + Z_2 + \ldots Z_n$
- Please note that impedances always add in series, regardless of what type of components comprise the impedances. That is, resistive impedance, inductive impedance, and capacitive impedance are to be treated the same way mathematically.
- A purely resistive impedance will always have a phase angle of exactly 0^o ($Z_R = R\ \Omega \angle 0^o$).
- A purely capacitive impedance will always have a phase angle of exactly -90^o ($Z_C = X_C\ \Omega \angle -90^o$).
- Ohm's Law for AC circuits: $E = IZ$; $I = E/Z$; $Z = E/I$
- When resistors and capacitors are mixed together in circuits, the total impedance will have a phase angle somewhere between 0^o and -90^o.
- Series AC circuits exhibit the same fundamental properties as series DC circuits: current is uniform throughout the circuit, voltage drops add to form the total voltage, and impedances add to form the total impedance.

4.4 Parallel resistor-capacitor circuits

Using the same value components in our series example circuit, we will connect them in parallel and see what happens:



Because the power source has the same frequency as the series example circuit, and the resistor and capacitor both have the same values of resistance and capacitance, respectively, they must also have the same values of impedance. So, we can begin our analysis table with the same "given" values:

	R	C	Total	
E			10 + j0 10 ∠ 0°	Volts
I				Amps
Z	5 + j0 5 ∠ 0°	0 - j26.5258 26.5258 ∠ -90°		Ohms

This being a parallel circuit now, we know that voltage is shared equally by all components, so we can place the figure for total voltage (10 volts ∠ 0^o) in all the columns:

	R	C	Total	
E	**10 + j0** **10** ∠ 0°	**10 + j0** **10** ∠ 0°	10 + j0 10 ∠ 0°	Volts
I				Amps
Z	5 + j0 5 ∠ 0°	0 - j26.5258 26.5258 ∠ -90°		Ohms

Rule of parallel circuits:

$E_{total} = E_R = E_C$

Now we can apply Ohm's Law (I=E/Z) vertically to two columns in the table, calculating current through the resistor and current through the capacitor:

	R	C	Total	
E	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	Volts
I	**2 + j0** **2** $\angle 0^o$	**0 + j376.99m** **376.99m** $\angle 90^o$		Amps
Z	5 + j0 $5 \angle 0^o$	0 - j26.5258 $26.5258 \angle -90^o$		Ohms

Ohm's Law $I = \frac{E}{Z}$ *Ohm's Law* $I = \frac{E}{Z}$

Just as with DC circuits, branch currents in a parallel AC circuit add up to form the total current (Kirchhoff's Current Law again):

	R	C	Total	
E	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	10 + j0 $10 \angle 0^o$	Volts
I	2 + j0 $2 \angle 0^o$	0 + j376.99m 376.99m $\angle 90^o$	**2 + j376.99m** **2.0352** $\angle 10.675^o$	Amps
Z	5 + j0 $5 \angle 0^o$	0 - j26.5258 $26.5258 \angle -90^o$		Ohms

Rule of parallel circuits: $I_{total} = I_R + I_C$

Finally, total impedance can be calculated by using Ohm's Law (Z=E/I) vertically in the "Total" column. As we saw in the AC inductance chapter, parallel impedance can also be calculated by using a reciprocal formula identical to that used in calculating parallel resistances. It is noteworthy to mention that this parallel impedance rule holds true regardless of the kind of impedances placed in parallel. In other words, it doesn't matter if we're calculating a circuit composed of parallel resistors, parallel inductors, parallel capacitors, or some combination thereof: in the form of impedances (Z), all the terms are common and can be applied uniformly to the same formula. Once again, the parallel impedance formula looks like this:

$$Z_{parallel} = \frac{1}{\frac{1}{Z_1} + \frac{1}{Z_2} + \ldots \frac{1}{Z_n}}$$

The only drawback to using this equation is the significant amount of work required to work it out, especially without the assistance of a calculator capable of manipulating complex quantities. Regardless of how we calculate total impedance for our parallel circuit (either Ohm's Law or the reciprocal formula), we will arrive at the same figure:

	R	C	Total	
E	10 + j0 10 ∠ 0°	10 + j0 10 ∠ 0°	10 + j0 10 ∠ 0°	Volts
I	2 + j0 2 ∠ 0°	0 + j376.99m 376.99m ∠ 90°	2 + j376.99m 2.0352 ∠ 10.675°	Amps
Z	5 + j0 5 ∠ 0°	0 - j26.5258 26.5258 ∠ -90°	**4.8284 - j910.14m** **4.9135** ∠ -10.675°	Ohms

Ohm's Law $Z = \frac{E}{I}$ *or* *Rule of parallel circuits:* $Z_{total} = \frac{1}{\frac{1}{Z_R} + \frac{1}{Z_C}}$

- **REVIEW:**
- Impedances (Z) are managed just like resistances (R) in parallel circuit analysis: parallel impedances diminish to form the total impedance, using the reciprocal formula. Just be sure to perform all calculations in complex (not scalar) form! $Z_{Total} = 1/(1/Z_1 + 1/Z_2 + \ldots 1/Z_n)$
- Ohm's Law for AC circuits: E = IZ ; I = E/Z ; Z = E/I
- When resistors and capacitors are mixed together in parallel circuits (just as in series circuits), the total impedance will have a phase angle somewhere between 0^o and -90^o. The circuit current will have a phase angle somewhere between 0^o and $+90^o$.
- Parallel AC circuits exhibit the same fundamental properties as parallel DC circuits: voltage is uniform throughout the circuit, branch currents add to form the total current, and impedances diminish (through the reciprocal formula) to form the total impedance.

4.5 Capacitor quirks

As with inductors, the ideal capacitor is a purely reactive device, containing absolutely zero resistive (power dissipative) effects. In the real world, of course, nothing is so perfect. However, capacitors have the virtue of generally being *purer* reactive components than inductors. It is a lot easier to design and construct a capacitor with low internal series resistance than it is to do the same with an inductor. The practical result of this is that real capacitors typically have impedance phase angles more closely approaching 90^o (actually, -90^o) than inductors. Consequently, they will tend to dissipate less power than an equivalent inductor.

Capacitors also tend to be smaller and lighter weight than their equivalent inductor counterparts, and since their electric fields are almost totally contained between their plates (unlike inductors, whose magnetic fields naturally tend to extend beyond the dimensions of the core), they are less prone to transmitting or receiving electromagnetic "noise" to/from other components. For these

reasons, circuit designers tend to favor capacitors over inductors wherever a design permits either alternative.

Capacitors with significant resistive effects are said to be *lossy*, in reference to their tendency to dissipate ("lose") power like a resistor. The source of capacitor loss is usually the dielectric material rather than any wire resistance, as wire length in a capacitor is very minimal.

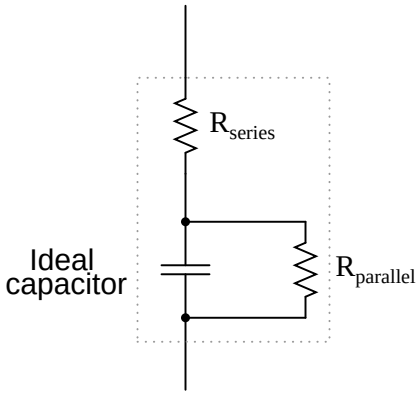
Dielectric materials tend to react to changing electric fields by producing heat. This heating effect represents a loss in power, and is equivalent to resistance in the circuit. The effect is more pronounced at higher frequencies and in fact can be so extreme that it is sometimes exploited in manufacturing processes to heat insulating materials like plastic! The plastic object to be heated is placed between two metal plates, connected to a source of high-frequency AC voltage. Temperature is controlled by varying the voltage or frequency of the source, and the plates never have to contact the object being heated.

This effect is undesirable for capacitors where we expect the component to behave as a purely *reactive* circuit element. One of the ways to mitigate the effect of dielectric "loss" is to choose a dielectric material less susceptible to the effect. Not all dielectric materials are equally "lossy." A relative scale of dielectric loss from least to greatest is given here:

```
Vacuum --------------- (Low Loss)
Air
Polystyrene
Mica
Glass
Low-K ceramic
Plastic film (Mylar)
Paper
High-K ceramic
Aluminum oxide
Tantalum pentoxide --- (High Loss)
```

Dielectric resistivity manifests itself both as a series and a parallel resistance with the pure capacitance:

Equivalent circuit for a real capacitor



Fortunately, these stray resistances are usually of modest impact (low series resistance and high parallel resistance), much less significant than the stray resistances present in an average inductor.

Electrolytic capacitors, known for their relatively high capacitance and low working voltage, are also known for their notorious lossiness, due to both the characteristics of the microscopically thin dielectric film and the electrolyte paste. Unless specially made for AC service, electrolytic capacitors should never be used with AC unless it is mixed (biased) with a constant DC voltage preventing the capacitor from ever being subjected to reverse voltage. Even then, their resistive characteristics may be too severe a shortcoming for the application anyway.

4.6 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 5

REACTANCE AND IMPEDANCE – R, L, AND C

5.1 Review of R, X, and Z

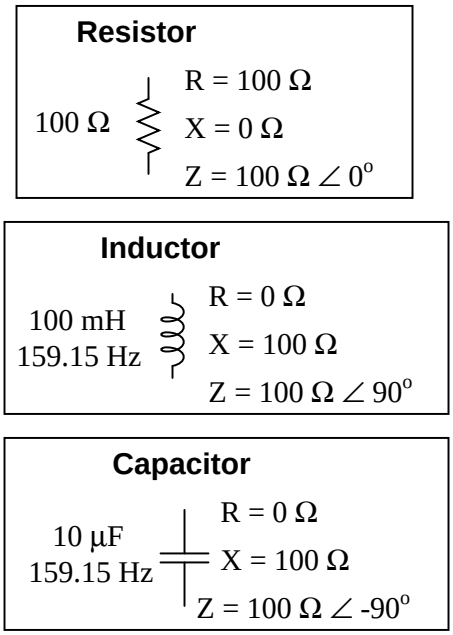
Before we begin to explore the effects of resistors, inductors, and capacitors connected together in the same AC circuits, let's briefly review some basic terms and facts.

Resistance is essentially *friction* against the motion of electrons. It is present in all conductors to some extent (except *super*conductors!), most notably in resistors. When alternating current goes through a resistance, a voltage drop is produced that is in-phase with the current. Resistance is mathematically symbolized by the letter "R" and is measured in the unit of ohms (Ω).

Reactance is essentially *inertia* against the motion of electrons. It is present anywhere electric or magnetic fields are developed in proportion to applied voltage or current, respectively; but most notably in capacitors and inductors. When alternating current goes through a pure reactance, a voltage drop is produced that is 90^o out of phase with the current. Reactance is mathematically symbolized by the letter "X" and is measured in the unit of ohms (Ω).

Impedance is a comprehensive expression of any and all forms of opposition to electron flow, including both resistance and reactance. It is present in all circuits, and in all components. When alternating current goes through an impedance, a voltage drop is produced that is somewhere between 0^o and 90^o out of phase with the current. Impedance is mathematically symbolized by the letter "Z" and is measured in the unit of ohms (Ω), in complex form.

Perfect resistors possess resistance, but not reactance. Perfect inductors and perfect capacitors possess reactance but no resistance. All components possess impedance, and because of this universal quality, it makes sense to translate all component values (resistance, inductance, capacitance) into common terms of impedance as the first step in analyzing an AC circuit.



The impedance phase angle for any component is the phase shift between voltage across that component and current through that component. For a perfect resistor, the voltage drop and current are *always* in phase with each other, and so the impedance angle of a resistor is said to be 0^o. For an perfect inductor, voltage drop always leads current by 90^o, and so an inductor's impedance phase angle is said to be $+90^o$. For a perfect capacitor, voltage drop always lags current by 90^o, and so a capacitor's impedance phase angle is said to be -90^o.

Impedances in AC behave analogously to resistances in DC circuits: they add in series, and they diminish in parallel. A revised version of Ohm's Law, based on impedance rather than resistance, looks like this:

Ohm's Law for AC circuits:

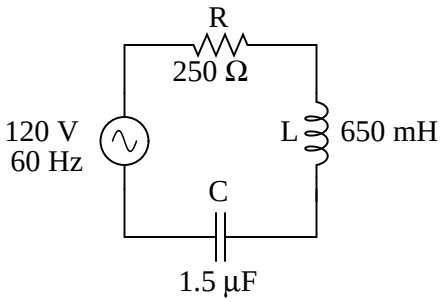
$$\mathbf{E = IZ} \qquad \mathbf{I = \frac{E}{Z}} \qquad \mathbf{Z = \frac{E}{I}}$$

All quantities expressed in complex, not scalar, form

Kirchhoff's Laws and all network analysis methods and theorems are true for AC circuits as well, so long as quantities are represented in complex rather than scalar form. While this qualified equivalence may be arithmetically challenging, it is conceptually simple and elegant. The only real difference between DC and AC circuit calculations is in regard to *power*. Because reactance doesn't dissipate power as resistance does, the concept of power in AC circuits is radically different from that of DC circuits. More on this subject in a later chapter!

5.2 Series R, L, and C

Let's take the following example circuit and analyze it:



The first step is to determine the reactances (in ohms) for the inductor and the capacitor.

$$X_L = 2\pi fL$$

$$X_L = (2)(\pi)(60\text{ Hz})(650\text{ mH})$$

$$X_L = 245.04\ \Omega$$

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(\pi)(60\text{ Hz})(1.5\ \mu\text{F})}$$

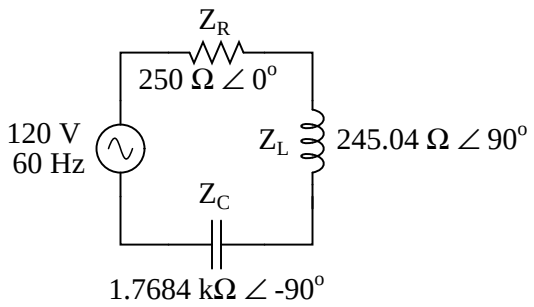
$$X_C = 1.7684\ \text{k}\Omega$$

The next step is to express all resistances and reactances in a mathematically common form: impedance. Remember that an inductive reactance translates into a positive imaginary impedance (or an impedance at $+90^o$), while a capacitive reactance translates into a negative imaginary impedance (impedance at -90^o). Resistance, of course, is still regarded as a purely "real" impedance (polar angle of 0^o):

$$Z_R = 250 + j0\ \Omega \quad or \quad 250\ \Omega \angle 0^o$$

$$Z_L = 0 + j245.04\ \Omega \quad or \quad 245.04\ \Omega \angle 90^o$$

$$Z_C = 0 - j1.7684\text{k}\ \Omega \quad or \quad 1.7684\ \text{k}\Omega \angle -90^o$$



Now, with all quantities of opposition to electric current expressed in a common, complex number format (as impedances, and not as resistances or reactances), they can be handled in the same way as plain resistances in a DC circuit. This is an ideal time to draw up an analysis table for this circuit and insert all the "given" figures (total voltage, and the impedances of the resistor, inductor, and capacitor).

	R	L	C	Total	
E				**120 + j0** **120** ∠ 0°	Volts
I					Amps
Z	**250 + j0** **250** ∠ 0°	**0 + j245.04** **254.04** ∠ 90°	**0 - j1.7684k** **1.7684k** ∠ -90°		Ohms

Unless otherwise specified, the source voltage will be our reference for phase shift, and so will be written at an angle of 0^o. Remember that there is no such thing as an "absolute" angle of phase shift for a voltage or current, since it's always a quantity relative to another waveform. Phase angles for impedance, however (like those of the resistor, inductor, and capacitor), are known absolutely, because the phase relationships between voltage and current at each component are absolutely defined.

Notice that I'm assuming a perfectly reactive inductor and capacitor, with impedance phase angles of exactly +90 and -90^o, respectively. Although real components won't be perfect in this regard, they should be fairly close. For simplicity, I'll assume perfectly reactive inductors and capacitors from now on in my example calculations except where noted otherwise.

Since the above example circuit is a series circuit, we know that the total circuit impedance is equal to the sum of the individuals, so:

$$Z_{total} = Z_R + Z_L + Z_C$$

$$Z_{total} = (250 + j0\ \Omega) + (0 + j245.04\ \Omega) + (0 - j1.7684k\ \Omega)$$

$$Z_{total} = 250 - j1.5233k\ \Omega \quad or \quad 1.5437\ k\Omega \angle -80.680^o$$

Inserting this figure for total impedance into our table:

	R	L	C	Total	
E				120 + j0 120 ∠ 0°	Volts
I					Amps
Z	250 + j0 250 ∠ 0°	0 + j245.04 254.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	**250 - j1.5233k** **1.5437k** ∠ -80.680°	Ohms

Rule of series circuits:
$Z_{total} = Z_R + Z_L + Z_C$

We can now apply Ohm's Law (I=E/R) vertically in the "Total" column to find total current for this series circuit:

	R	L	C	Total	
E				120 + j0 120 ∠ 0°	Volts
I				**12.589m + 76.708m** **77.734m** ∠ 80.680°	Amps
Z	250 + j0 250 ∠ 0°	0 + j245.04 254.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	250 - j1.5233k 1.5437k ∠ -80.680°	Ohms

Ohm's Law
$I = \frac{E}{Z}$

Being a series circuit, current must be equal through all components. Thus, we can take the figure obtained for total current and distribute it to each of the other columns:

	R	L	C	Total	
E				120 + j0 120 ∠ 0°	Volts
I	**12.589m + 76.708m** **77.734m** ∠ 80.680°	**12.589m + 76.708m** **77.734m** ∠ 80.680°	**12.589m + 76.708m** **77.734m** ∠ 80.680°	12.589m + 76.708m 77.734m ∠ 80.680°	Amps
Z	250 + j0 250 ∠ 0°	0 + j245.04 254.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	250 - j1.5233k 1.5437k ∠ -80.680°	Ohms

Rule of series circuits:
$I_{total} = I_R = I_L = I_C$

Now we're prepared to apply Ohm's Law (E=IZ) to each of the individual component columns in the table, to determine voltage drops:

	R	L	C	Total	
E	**3.1472 + j19.177** **19.434** $\angle$ 80.680^o	**-18.797 + j3.0848** **19.048** $\angle$ 170.68^o	**135.65 - j22.262** **137.46** $\angle$ -9.3199^o	120 + j0 120 $\angle$ 0^o	Volts
I	12.589m + 76.708m 77.734m $\angle$ 80.680^o	12.589m + 76.708m 77.734m $\angle$ 80.680^o	12.589m + 76.708m 77.734m $\angle$ 80.680^o	12.589m + 76.708m 77.734m $\angle$ 80.680^o	Amps
Z	250 + j0 250 $\angle$ 0^o	0 + j245.04 254.04 $\angle$ 90^o	0 - j1.7684k 1.7684k $\angle$ -90^o	250 - j1.5233k 1.5437k $\angle$ -80.680^o	Ohms
	Ohm's Law E = IZ	*Ohm's Law* E = IZ	*Ohm's Law* E = IZ		

Notice something strange here: although our supply voltage is only 120 volts, the voltage across the capacitor is 137.46 volts! How can this be? The answer lies in the interaction between the inductive and capacitive reactances. Expressed as impedances, we can see that the inductor opposes current in a manner precisely opposite that of the capacitor. Expressed in rectangular form, the inductor's impedance has a positive imaginary term and the capacitor has a negative imaginary term. When these two contrary impedances are added (in series), they tend to cancel each other out! Although they're still *added together* to produce a sum, that sum is actually *less* than either of the individual (capacitive or inductive) impedances alone. It is analogous to adding together a positive and a negative (scalar) number: the sum is a quantity less than either one's individual absolute value.

If the total impedance in a series circuit with both inductive and capacitive elements is less than the impedance of either element separately, then the total current in that circuit must be *greater* than what it would be with only the inductive or only the capacitive elements there. With this abnormally high current through each of the components, voltages greater than the source voltage may be obtained across some of the individual components! Further consequences of inductors' and capacitors' opposite reactances in the same circuit will be explored in the next chapter.

Once you've mastered the technique of reducing all component values to impedances (Z), analyzing any AC circuit is only about as difficult as analyzing any DC circuit, except that the quantities dealt with are vector instead of scalar. With the exception of equations dealing with power (P), equations in AC circuits are the same as those in DC circuits, using impedances (Z) instead of resistances (R). Ohm's Law (E=IZ) still holds true, and so do Kirchhoff's Voltage and Current Laws.

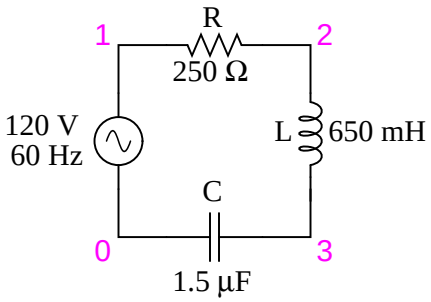
To demonstrate Kirchhoff's Voltage Law in an AC circuit, we can look at the answers we derived for component voltage drops in the last circuit. KVL tells us that the algebraic sum of the voltage drops across the resistor, inductor, and capacitor should equal the applied voltage from the source. Even though this may not look like it is true at first sight, a bit of complex number addition proves otherwise:

$E_R + E_L + E_C$ *should equal* E_{total}

$$\begin{array}{rrl} & 3.1472 + j19.177 \text{ V} & E_R \\ & -18.797 + j3.0848 \text{ V} & E_L \\ + & 135.65 - j22.262 \text{ V} & E_C \\ \hline & 120 + j0 \quad \text{V} & E_{total} \end{array}$$

Aside from a bit of rounding error, the sum of these voltage drops does equal 120 volts. Performed on a calculator (preserving all digits), the answer you will receive should be *exactly* 120 + j0 volts.

We can also use SPICE to verify our figures for this circuit:




```
ac r-l-c circuit
v1 1 0 ac 120 sin
r1 1 2 250
l1 2 3 650m
c1 3 0 1.5u
.ac lin 1 60 60
.print ac v(1,2) v(2,3) v(3,0) i(v1)
.print ac vp(1,2) vp(2,3) vp(3,0) ip(v1)
.end
```

```
freq          v(1,2)      v(2,3)      v(3)        i(v1)
6.000E+01     1.943E+01   1.905E+01   1.375E+02   7.773E-02
freq          vp(1,2)     vp(2,3)     vp(3)       ip(v1)
6.000E+01     8.068E+01   1.707E+02  -9.320E+00  -9.932E+01
```

Interpreted SPICE results

$E_R = 19.43 \text{ V} \angle 80.68^o$

$E_L = 19.05 \text{ V} \angle 170.7^o$

$E_C = 137.5 \text{ V} \angle -9.320^o$

$I = 77.73 \text{ mA} \angle -99.32^o$ (*actual phase angle* = 80.68^o)

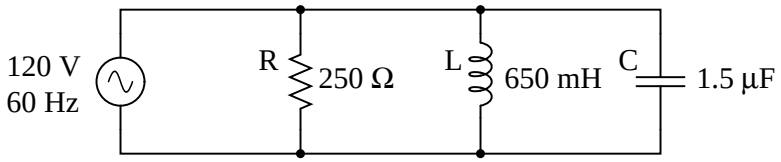
The SPICE simulation shows our hand-calculated results to be accurate.

As you can see, there is little difference between AC circuit analysis and DC circuit analysis, except that all quantities of voltage, current, and resistance (actually, *impedance*) must be handled in complex rather than scalar form so as to account for phase angle. This is good, since it means all you've learned about DC electric circuits applies to what you're learning here. The only exception to this consistency is the calculation of power, which is so unique that it deserves a chapter devoted to that subject alone.

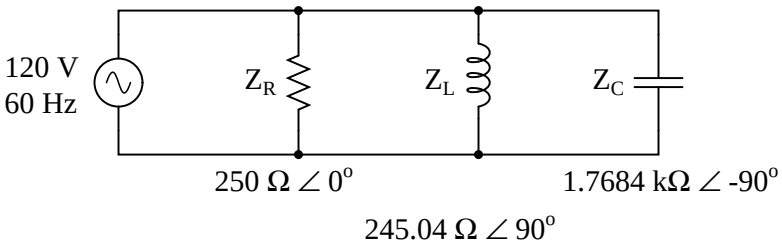
- **REVIEW:**
- Impedances of any kind add in series: $Z_{Total} = Z_1 + Z_2 + \ldots Z_n$
- Although impedances add in series, the total impedance for a circuit containing both inductance and capacitance may be less than one or more of the individual impedances, because series inductive and capacitive impedances tend to cancel each other out. This may lead to voltage drops across components exceeding the supply voltage!
- All rules and laws of DC circuits apply to AC circuits, so long as values are expressed in complex form rather than scalar. The only exception to this principle is the calculation of *power*, which is very different for AC.

5.3 Parallel R, L, and C

We can take the same components from the series circuit and rearrange them into a parallel configuration for an easy example circuit:



The fact that these components are connected in parallel instead of series now has absolutely no effect on their individual impedances. So long as the power supply is the same frequency as before, the inductive and capacitive reactances will not have changed at all:



With all component values expressed as impedances (Z), we can set up an analysis table and proceed as in the last example problem, except this time following the rules of parallel circuits instead of series:

	R	L	C	Total	
E				**120 + j0** **120** ∠ 0°	Volts
I					Amps
Z	**250 + j0** **250** ∠ 0°	**0 + j245.04** **254.04** ∠ 90°	**0 - j1.7684k** **1.7684k** ∠ -90°		Ohms

Knowing that voltage is shared equally by all components in a parallel circuit, we can transfer the figure for total voltage to all component columns in the table:

	R	L	C	Total	
E	**120 + j0** **120** ∠ 0°	**120 + j0** **120** ∠ 0°	**120 + j0** **120** ∠ 0°	120 + j0 120 ∠ 0°	Volts
I					Amps
Z	250 + j0 250 ∠ 0°	0 + j245.04 254.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°		Ohms

Rule of parallel circuits:

$E_{total} = E_R = E_L = E_C$

Now, we can apply Ohm's Law (I=E/Z) vertically in each column to determine current through each component:

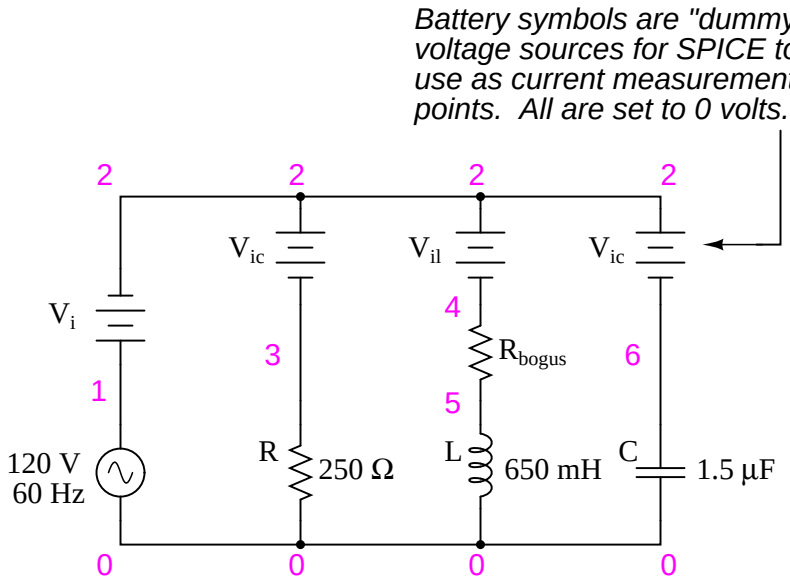
	R	L	C	Total	
E	120 + j0 120 ∠ 0°	120 + j0 120 ∠ 0°	120 + j0 120 ∠ 0°	120 + j0 120 ∠ 0°	Volts
I	**480m + j0** **480** ∠ 0°	**0 - j489.71m** **489.71m** ∠ -90°	**0 + j67.858m** **67.858m** ∠ 90°		Amps
Z	250 + j0 250 ∠ 0°	0 + j245.04 254.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°		Ohms

Ohm's Law $I = \frac{E}{Z}$ | *Ohm's Law* $I = \frac{E}{Z}$ | *Ohm's Law* $I = \frac{E}{Z}$

There are two strategies for calculating total current and total impedance. First, we could calculate total impedance from all the individual impedances in parallel ($Z_{Total} = 1/(1/Z_R + 1/Z_L + 1/Z_C)$), and then calculate total current by dividing source voltage by total impedance (I=E/Z). However, working through the parallel impedance equation with complex numbers is no easy task, with all the reciprocations (1/Z). This is especially true if you're unfortunate enough not to have a calculator that handles complex numbers and are forced to do it all by hand (reciprocate the individual impedances in polar form, then convert them all to rectangular form for addition, then convert back to polar form for the final inversion, then invert). The second way to calculate total current and total impedance is to add up all the branch currents to arrive at total current (total current in a parallel circuit – AC or DC – is equal to the sum of the branch currents), then use Ohm's Law to determine total impedance from total voltage and total current (Z=E/I).

	R	L	C	Total	
E	120 + j0 120 ∠ 0°	120 + j0 120 ∠ 0°	120 + j0 120 ∠ 0°	120 + j0 120 ∠ 0°	Volts
I	480m + j0 480 ∠ 0°	0 - j489.71m 489.71m ∠ -90°	0 + j67.858m 67.858m ∠ 90°	**480m - j421.85m** **639.03m** ∠ -41.311°	Amps
Z	250 + j0 250 ∠ 0°	0 + j245.04 254.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	**141.05 + j123.96** **187.79** ∠ 41.311°	Ohms

Either method, performed properly, will provide the correct answers. Let's try analyzing this circuit with SPICE and see what happens:




```
ac r-l-c circuit
v1 1 0 ac 120 sin
vi 1 2 ac 0
vir 2 3 ac 0
vil 2 4 ac 0
```

```
rbogus 4 5 1e-12
vic 2 6 ac 0
r1 3 0 250
l1 5 0 650m
c1 6 0 1.5u
.ac lin 1 60 60
.print ac i(vi) i(vir) i(vil) i(vic)
.print ac ip(vi) ip(vir) ip(vil) ip(vic)
.end


freq          i(vi)       i(vir)      i(vil)      i(vic)
6.000E+01     6.390E-01   4.800E-01   4.897E-01   6.786E-02
freq          ip(vi)      ip(vir)     ip(vil)     ip(vic)
6.000E+01    -4.131E+01   0.000E+00  -9.000E+01   9.000E+01
```

Interpreted SPICE results

I_{total} = 639.0 mA $\angle$ -41.31^o

I_R = 480 mA $\angle$ 0^o

I_L = 489.7 mA $\angle$ -90^o

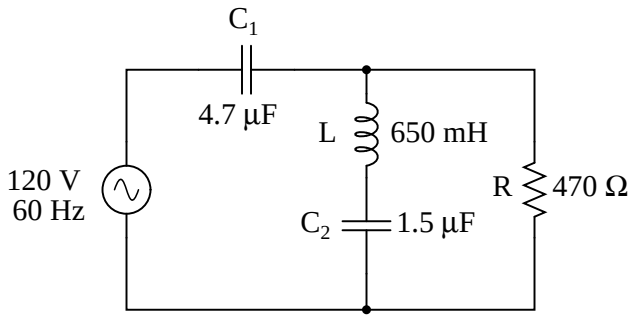
I_C = 67.86 mA $\angle$ 90^o

It took a little bit of trickery to get SPICE working as we would like on this circuit (installing "dummy" voltage sources in each branch to obtain current figures and installing the "dummy" resistor in the inductor branch to prevent a direct inductor-to-voltage source loop, which SPICE cannot tolerate), but we did get the proper readings. Even more than that, by installing the dummy voltage sources (current meters) in the proper directions, we were able to avoid that idiosyncrasy of SPICE of printing current figures 180^o out of phase. This way, our current phase readings came out to exactly match our hand calculations.

5.4 Series-parallel R, L, and C

Now that we've seen how series and parallel AC circuit analysis is not fundamentally different than DC circuit analysis, it should come as no surprise that series-parallel analysis would be the same as well, just using complex numbers instead of scalar to represent voltage, current, and impedance.

Take this series-parallel circuit for example:



The first order of business, as usual, is to determine values of impedance (Z) for all components based on the frequency of the AC power source. To do this, we need to first determine values of reactance (X) for all inductors and capacitors, then convert reactance (X) and resistance (R) figures into proper impedance (Z) form:

Reactances and Resistances:

$X_{C1} = \frac{1}{2\pi f C_1}$ $X_{C1} = \frac{1}{(2)(\pi)(60\text{ Hz})(4.7\ \mu\text{F})}$ $X_{C1} = 564.38\ \Omega$	$X_L = 2\pi f L$ $X_L = (2)(\pi)(60\text{ Hz})(650\text{ mH})$ $X_L = 245.04\ \Omega$
$X_{C2} = \frac{1}{2\pi f C_2}$ $X_{C2} = \frac{1}{(2)(\pi)(60\text{ Hz})(1.5\ \mu\text{F})}$ $X_{C2} = 1.7684\ \text{k}\Omega$	$R = 470\ \Omega$

$Z_{C1} = 0 - j564.38\ \Omega$ *or* $564.38\ \Omega \angle -90^o$

$Z_L = 0 + j245.04\ \Omega$ *or* $245.04\ \Omega \angle 90^o$

$Z_{C2} = 0 - j1.7684\text{k}\ \Omega$ *or* $1.7684\ \text{k}\Omega \angle -90^o$

$Z_R = 470 + j0\ \Omega$ *or* $470\ \Omega \angle 0^o$

Now we can set up the initial values in our table:

	C_1	L	C_2	R	Total	
E					120 + j0 120 ∠ 0°	Volts
I						Amps
Z	0 - j564.38 564.38 ∠ -90°	0 + j245.04 245.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	470 + j0 470 ∠ 0°		Ohms

Being a series-parallel *combination* circuit, we must reduce it to a total impedance in more than one step. The first step is to combine L and C_2 as a series combination of impedances, by adding their impedances together. Then, that impedance will be combined in parallel with the impedance of the resistor, to arrive at another combination of impedances. Finally, that quantity will be added to the impedance of C_1 to arrive at the total impedance.

In order that our table may follow all these steps, it will be necessary to add additional columns to it so that each step may be represented. Adding more columns horizontally to the table shown above would be impractical for formatting reasons, so I will place a new row of columns underneath, each column designated by its respective component combination:

	$L -- C_2$	$R // (L -- C_2)$	*Total* $C_1 -- [R // (L -- C_2)]$	
E				Volts
I				Amps
Z				Ohms

Calculating these new (combination) impedances will require complex addition for series combinations, and the "reciprocal" formula for complex impedances in parallel. This time, there is no avoidance of the reciprocal formula: the required figures can be arrived at no other way!

	$L -- C_2$	$R // (L -- C_2)$	*Total* $C_1 -- [R // (L -- C_2)]$	
E			120 + j0 120 ∠ 0°	Volts
I				Amps
Z	**0 - j1.5233k** **1.5233k** ∠ -90°	**429.15 - j132.41** **449.11** ∠ -17.147°	**429.15 - j696.79** **818.34** ∠ -58.371°	Ohms

Rule of series circuits:
$Z_{L--C2} = Z_L + Z_{C2}$

Rule of parallel circuits:
$$Z_{R//(L--C2)} = \frac{1}{\frac{1}{Z_R} + \frac{1}{Z_{L--C2}}}$$

Rule of series circuits:
$Z_{total} = Z_{C1} + Z_{R//(L--C2)}$

Seeing as how our second table contains a column for "Total," we can safely discard that column from the first table. This gives us one table with four columns and another table with three columns.

Now that we know the total impedance (818.34 Ω ∠ -58.371^o) and the total voltage (120 volts ∠ 0^o), we can apply Ohm's Law (I=E/Z) vertically in the "Total" column to arrive at a figure for total current:

	$L -- C_2$	$R // (L -- C_2)$	*Total* $C_1 -- [R // (L -- C_2)]$	
E			120 + j0 120 ∠ 0^o	Volts
I			**76.899m + j124.86m** **146.64m** ∠ 58.371^o	Amps
Z	0 - j1.5233k 1.5233k ∠ -90^o	429.15 - j132.41 449.11 ∠ -17.147^o	429.15 - j696.79 818.34 ∠ -58.371^o	Ohms

Ohm's Law

$$I = \frac{E}{Z}$$

At this point we ask ourselves the question: are there any components or component combinations which share either the total voltage or the total current? In this case, both C_1 and the parallel combination $R//(L--C_2)$ share the same (total) current, since the total impedance is composed of the two sets of impedances in series. Thus, we can transfer the figure for total current into both columns:

	C_1	L	C_2	R	
E					Volts
I	**76.899m + j124.86m** **146.64m** ∠ 58.371^o				Amps
Z	0 - j564.38 564.38 ∠ -90^o	0 + j245.04 245.04 ∠ 90^o	0 - j1.7684k 1.7684k ∠ -90^o	470 + j0 470 ∠ 0^o	Ohms

Rule of series circuits:

$I_{total} = I_{C1} = I_{R//(L--C2)}$

	$L -- C_2$	$R // (L -- C_2)$	*Total* $C_1 -- [R // (L -- C_2)]$	
E			120 + j0 120 ∠ 0^o	Volts
I		**76.899m + j124.86m** **146.64m** ∠ 58.371^o	76.899m + j124.86m 146.64m ∠ 58.371^o	Amps
Z	0 - j1.5233k 1.5233k ∠ -90^o	429.15 - j132.41 449.11 ∠ -17.147^o	429.15 - j696.79 818.34 ∠ -58.371^o	Ohms

Rule of series circuits:

$I_{total} = I_{C1} = I_{R//(L--C2)}$

Now, we can calculate voltage drops across C_1 and the series-parallel combination of $R//(L--C_2)$ using Ohm's Law (E=IZ) vertically in those table columns:

	C_1	L	C_2	R	
E	**70.467 - j43.400** **82.760** $\angle$ -31.629^o				Volts
I	76.899m + j124.86m 146.64m $\angle$ 58.371^o				Amps
Z	0 - j564.38 564.38 $\angle$ -90^o	0 + j245.04 245.04 $\angle$ 90^o	0 - j1.7684k 1.7684k $\angle$ -90^o	470 + j0 470 $\angle$ 0^o	Ohms

Ohm's Law
E = IZ

	L -- C_2	R // (L -- C_2)	*Total* C_1 -- [R // (L -- C_2)]	
E		**49.533 + j43.400** **65.857** $\angle$ 41.225^o	120 + j0 120 $\angle$ 0^o	Volts
I		76.899m + j124.86m 146.64m $\angle$ 58.371^o	76.899m + j124.86m 146.64m $\angle$ 58.371^o	Amps
Z	0 - j1.5233k 1.5233k $\angle$ -90^o	429.15 - j132.41 449.11 $\angle$ -17.147^o	429.15 - j696.79 818.34 $\angle$ -58.371^o	Ohms

Ohm's Law
E = IZ

A quick double-check of our work at this point would be to see whether or not the voltage drops across C_1 and the series-parallel combination of R//(L−−C_2) indeed add up to the total. According to Kirchhoff's Voltage Law, they should!

E_{total} *should be equal to* $E_{C1} + E_{R//(L--C2)}$

$$\begin{array}{r} 70.467 - j43.400 \text{ V} \\ + \ 49.533 + j43.400 \text{ V} \\ \hline 120 + j0 \text{ V} \end{array} \quad \longleftarrow \textit{ Indeed, it is!}$$

That last step was merely a precaution. In a problem with as many steps as this one has, there is much opportunity for error. Occasional cross-checks like that one can save a person a lot of work and unnecessary frustration by identifying problems prior to the final step of the problem.

After having solved for voltage drops across C_1 and the combination R//(L−−C_2), we again ask ourselves the question: what other components share the same voltage or current? In this case, the resistor (R) and the combination of the inductor and the second capacitor (L−−C_2) share the same voltage, because those sets of impedances are in parallel with each other. Therefore, we can transfer the voltage figure just solved for into the columns for R and L−−C_2:

	C_1	L	C_2	R	
E	70.467 - j43.400 82.760 ∠ -31.629°			**49.533 + j43.400** **65.857** ∠ 41.225°	Volts
I	76.899m + j124.86m 146.64m ∠ 58.371°				Amps
Z	0 - j564.38 564.38 ∠ -90°	0 + j245.04 245.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	470 + j0 470 ∠ 0°	Ohms

Rule of parallel circuits:

$E_{R//(L--C2)} = E_R = E_{L--C2}$

	$L -- C_2$	$R // (L -- C_2)$	*Total* $C_1 -- [R // (L -- C_2)]$	
E	**49.533 + j43.400** **65.857** ∠ 41.225°	49.533 + j43.400 65.857 ∠ 41.225°	120 + j0 120 ∠ 0°	Volts
I		76.899m + j124.86m 146.64m ∠ 58.371°	76.899m + j124.86m 146.64m ∠ 58.371°	Amps
Z	0 - j1.5233k 1.5233k ∠ -90°	429.15 - j132.41 449.11 ∠ -17.147°	429.15 - j696.79 818.34 ∠ -58.371°	Ohms

Rule of parallel circuits:

$E_{R//(L--C2)} = E_R = E_{L--C2}$

Now we're all set for calculating current through the resistor and through the series combination $L--C_2$. All we need to do is apply Ohm's Law (I=E/Z) vertically in both of those columns:

	C_1	L	C_2	R	
E	70.467 - j43.400 82.760 ∠ -31.629°			49.533 + j43.400 65.857 ∠ 41.225°	Volts
I	76.899m + j124.86m 146.64m ∠ 58.371°			**105.39m + j92.341m** **140.12m** ∠ 41.225°	Amps
Z	0 - j564.38 564.38 ∠ -90°	0 + j245.04 245.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	470 + j0 470 ∠ 0°	Ohms

Ohm's Law

$$I = \frac{E}{Z}$$

	L -- C_2	R // (L -- C_2)	*Total* C_1 -- [R // (L -- C_2)]	
E	49.533 + j43.400 65.857 ∠ 41.225°	49.533 + j43.400 65.857 ∠ 41.225°	120 + j0 120 ∠ 0°	Volts
I	**-28.490m + j32.516m** **43.232m** ∠ 131.22°	76.899m + j124.86m 146.64m ∠ 58.371°	76.899m + j124.86m 146.64m ∠ 58.371°	Amps
Z	0 - j1.5233k 1.5233k ∠ -90°	429.15 - j132.41 449.11 ∠ -17.147°	429.15 - j696.79 818.34 ∠ -58.371°	Ohms

Ohm's Law

$$I = \frac{E}{Z}$$

Another quick double-check of our work at this point would be to see if the current figures for L−−C_2 and R add up to the total current. According to Kirchhoff's Current Law, they should:

$I_{R//(L--C2)}$ *should be equal to* $I_R + I_{(L--C2)}$

$$\begin{array}{r} 105.39\text{m} + j92.341\text{m} \\ + \quad -28.490\text{m} + j32.516\text{m} \\ \hline 76.899\text{m} + j124.86\text{m} \end{array} \longleftarrow \textit{Indeed, it is!}$$

Since the L and C_2 are connected in series, and since we know the current through their series combination impedance, we can distribute that current figure to the L and C_2 columns following the rule of series circuits whereby series components share the same current:

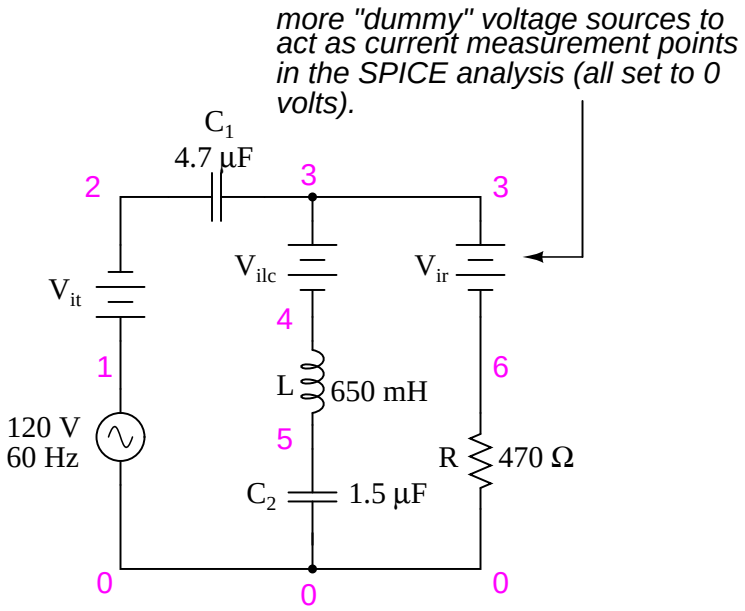
	C_1	L	C_2	R	
E	70.467 - j43.400 82.760 ∠ -31.629°			49.533 + j43.400 65.857 ∠ 41.225°	Volts
I	76.899m + j124.86m 146.64m ∠ 58.371°	**-28.490m + j32.516m** **43.232m** ∠ 131.22°	**-28.490m + j32.516m** **43.232m** ∠ 131.22°	105.39m + j92.341m 140.12m ∠ 41.225°	Amps
Z	0 - j564.38 564.38 ∠ -90°	0 + j245.04 245.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	470 + j0 470 ∠ 0°	Ohms

Rule of series circuits:
$I_{L--C2} = I_L = I_{C2}$

With one last step (actually, two calculations), we can complete our analysis table for this circuit. With impedance and current figures in place for L and C_2, all we have to do is apply Ohm's Law (E=IZ) vertically in those two columns to calculate voltage drops.

	C_1	L	C_2	R	
E	70.467 - j43.400 82.760 ∠ -31.629°	**-7.968 - j6.981** **10.594** ∠ 221.22°	**57.501 + j50.382** **76.451** ∠ 41.225	49.533 + j43.400 65.857 ∠ 41.225°	Volts
I	76.899m + j124.86m 146.64m ∠ 58.371°	-28.490m + j32.516m 43.232m ∠ 131.22°	-28.490m + j32.516m 43.232m ∠ 131.22°	105.39m + j92.341m 140.12m ∠ 41.225°	Amps
Z	0 - j564.38 564.38 ∠ -90°	0 + j245.04 245.04 ∠ 90°	0 - j1.7684k 1.7684k ∠ -90°	470 + j0 470 ∠ 0°	Ohms
		Ohm's Law E = IZ	*Ohm's Law* E = IZ		

Now, let's turn to SPICE for a computer verification of our work:




```
ac series-parallel r-l-c circuit
v1 1 0 ac 120 sin
vit 1 2 ac 0
vilc 3 4 ac 0
vir 3 6 ac 0
c1 2 3 4.7u
l 4 5 650m
c2 5 0 1.5u
r 6 0 470
.ac lin 1 60 60
.print ac v(2,3) vp(2,3) i(vit) ip(vit)
.print ac v(4,5) vp(4,5) i(vilc) ip(vilc)
.print ac v(5,0) vp(5,0) i(vilc) ip(vilc)
```

```
.print ac v(6,0) vp(6,0) i(vir) ip(vir)
.end

freq          v(2,3)      vp(2,3)     i(vit)      ip(vit)     C1
6.000E+01     8.276E+01  -3.163E+01   1.466E-01   5.837E+01

freq          v(4,5)      vp(4,5)     i(vilc)     ip(vilc)    L
6.000E+01     1.059E+01  -1.388E+02   4.323E-02   1.312E+02

freq          v(5)        vp(5)       i(vilc)     ip(vilc)    C2
6.000E+01     7.645E+01   4.122E+01   4.323E-02   1.312E+02

freq          v(6)        vp(6)       i(vir)      ip(vir)     R
6.000E+01     6.586E+01   4.122E+01   1.401E-01   4.122E+01
```

Each line of the SPICE output listing gives the voltage, voltage phase angle, current, and current phase angle for C_1, L, C_2, and R, in that order. As you can see, these figures do concur with our hand-calculated figures in the circuit analysis table.

As daunting a task as series-parallel AC circuit analysis may appear, it must be emphasized that there is nothing really new going on here besides the use of complex numbers. Ohm's Law (in its new form of E=IZ) still holds true, as do the voltage and current Laws of Kirchhoff. While there is more potential for human error in carrying out the necessary complex number calculations, the basic principles and techniques of series-parallel circuit reduction are exactly the same.

- **REVIEW:**
- Analysis of series-parallel AC circuits is much the same as series-parallel DC circuits. The only substantive difference is that all figures and calculations are in complex (not scalar) form.
- It is important to remember that before series-parallel reduction (simplification) can begin, you must determine the impedance (Z) of every resistor, inductor, and capacitor. That way, all component values will be expressed in common terms (Z) instead of an incompatible mix of resistance (R), inductance (L), and capacitance (C).

5.5 Susceptance and Admittance

In the study of DC circuits, the student of electricity comes across a term meaning the opposite of resistance: *conductance*. It is a useful term when exploring the mathematical formula for parallel resistances: $R_{parallel} = 1 \ / \ (1/R_1 + 1/R_2 + \ . \ . \ . \ 1/R_n)$. Unlike resistance, which diminishes as more parallel components are included in the circuit, conductance simply adds. Mathematically, conductance is the reciprocal of resistance, and each 1/R term in the "parallel resistance formula" is actually a conductance.

Whereas the term "resistance" denotes the amount of opposition to flowing electrons in a circuit, "conductance" represents the ease of which electrons may flow. Resistance is the measure of how much a circuit *resists* current, while conductance is the measure of how much a circuit *conducts* current. Conductance used to be measured in the unit of *mhos*, or "ohms" spelled backward. Now,

the proper unit of measurement is *Siemens*. When symbolized in a mathematical formula, the proper letter to use for conductance is "G".

Reactive components such as inductors and capacitors oppose the flow of electrons with respect to time, rather than with a constant, unchanging friction as resistors do. We call this time-based opposition, *reactance*, and like resistance we also measure it in the unit of *ohms*.

As conductance is the complement of resistance, there is also a complementary expression of reactance, called *susceptance*. Mathematically, it is equal to 1/X, the reciprocal of reactance. Like conductance, it used to be measured in the unit of mhos, but now is measured in Siemens. Its mathematical symbol is "B", unfortunately the same symbol used to represent magnetic flux density.

The terms "reactance" and "susceptance" have a certain linguistic logic to them, just like resistance and conductance. While reactance is the measure of how much a circuit *reacts* against change in current over time, susceptance is the measure of how much a circuit is *susceptible* to conducting a changing current.

If one were tasked with determining the total effect of several parallel-connected, pure reactances, one could convert each reactance (X) to a susceptance (B), then add susceptances rather than diminish reactances: $\mathrm{X}_{parallel} = 1/(1/\mathrm{X}_1 + 1/\mathrm{X}_2 + \ldots 1/\mathrm{X}_n)$. Like conductances (G), susceptances (B) add in parallel and diminish in series. Also like conductance, susceptance is a scalar quantity.

When resistive and reactive components are interconnected, their combined effects can no longer be analyzed with scalar quantities of resistance (R) and reactance (X). Likewise, figures of conductance (G) and susceptance (B) are most useful in circuits where the two types of opposition are not mixed, i.e. either a purely resistive (conductive) circuit, or a purely reactive (susceptive) circuit. In order to express and quantify the effects of mixed resistive and reactive components, we had to have a new term: *impedance*, measured in ohms and symbolized by the letter "Z".

To be consistent, we need a complementary measure representing the reciprocal of impedance. The name for this measure is *admittance*. Admittance is measured in (guess what?) the unit of Siemens, and its symbol is "Y". Like impedance, admittance is a complex quantity rather than scalar. Again, we see a certain logic to the naming of this new term: while impedance is a measure of how much alternating current is *impeded* in a circuit, admittance is a measure of how much current is *admitted*.

Given a scientific calculator capable of handling complex number arithmetic in both polar and rectangular forms, you may never have to work with figures of susceptance (B) or admittance (Y). Be aware, though, of their existence and their meanings.

5.6 Summary

With the notable exception of calculations for power (P), all AC circuit calculations are based on the same general principles as calculations for DC circuits. The only significant difference is that fact that AC calculations use complex quantities while DC calculations use scalar quantities. Ohm's Law, Kirchhoff's Laws, and even the network theorems learned in DC still hold true for AC when voltage, current, and impedance are all expressed with complex numbers. The same troubleshooting strategies applied toward DC circuits also hold for AC, although AC can certainly be more difficult to work with due to phase angles which aren't registered by a handheld multimeter.

Power is another subject altogether, and will be covered in its own chapter in this book. Because power in a reactive circuit is both absorbed and released – not just dissipated as it is with resistors – its mathematical handling requires a more direct application of trigonometry to solve.

When faced with analyzing an AC circuit, the first step in analysis is to convert all resistor, inductor, and capacitor component values into impedances (Z), based on the frequency of the power source. After that, proceed with the same steps and strategies learned for analyzing DC circuits, using the "new" form of Ohm's Law: E=IZ ; I=E/Z ; and Z=E/I

Remember that only the calculated figures expressed in *polar* form apply directly to empirical measurements of voltage and current. Rectangular notation is merely a useful tool for us to add and subtract complex quantities together. Polar notation, where the magnitude (length of vector) directly relates to the magnitude of the voltage or current measured, and the angle directly relates to the phase shift in degrees, is the most practical way to express complex quantities for circuit analysis.

5.7 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

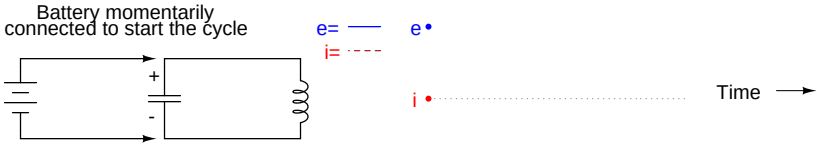
Chapter 6

RESONANCE

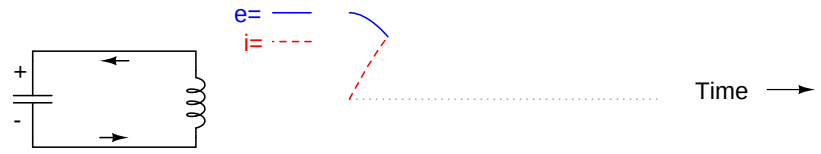
6.1 An electric pendulum

Capacitors store energy in the form of an electric field, and electrically manifest that stored energy as a potential: *static voltage*. Inductors store energy in the form of a magnetic field, and electrically manifest that stored energy as a kinetic motion of electrons: *current*. Capacitors and inductors are flip-sides of the same reactive coin, storing and releasing energy in complementary modes. When these two types of reactive components are directly connected together, their complementary tendencies to store energy will produce an unusual result.

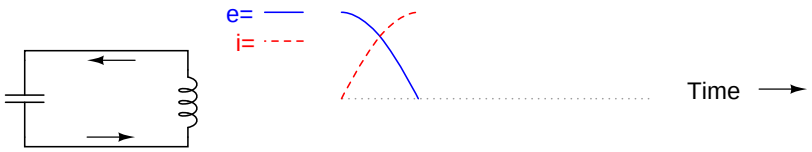
If either the capacitor or inductor starts out in a charged state, the two components will exchange energy between them, back and forth, creating their own AC voltage and current cycles. If we assume that both components are subjected to a sudden application of voltage (say, from a momentarily connected battery), the capacitor will very quickly charge and the inductor will oppose change in current, leaving the capacitor in the charged state and the inductor in the discharged state:



The capacitor will begin to discharge, its voltage decreasing. Meanwhile, the inductor will begin to build up a "charge" in the form of a magnetic field as current increases in the circuit:

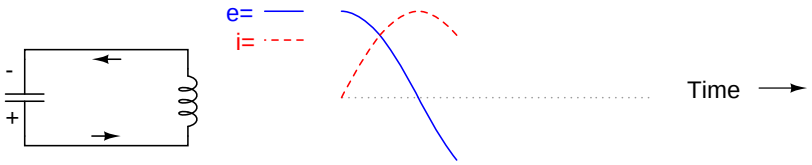


The inductor, still charging, will keep electrons flowing in the circuit until the capacitor has been completely discharged, leaving zero voltage across it:



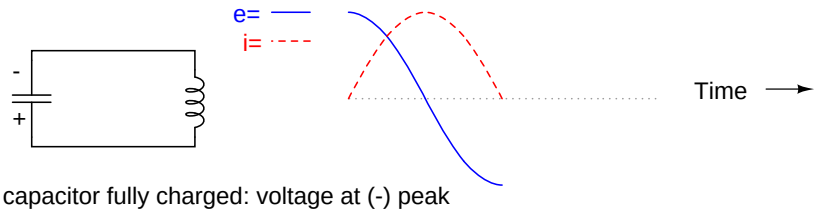
capacitor fully discharged: zero voltage
inductor fully charged: maximum current

The inductor will maintain current flow even with no voltage applied. In fact, it will generate a voltage (like a battery) in order to keep current in the same direction. The capacitor, being the recipient of this current, will begin to accumulate a charge in the opposite polarity as before:



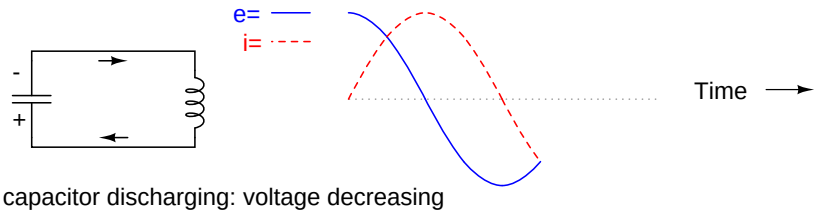
capacitor charging: voltage increasing (in opposite polarity)
inductor discharging: current decreasing

When the inductor is finally depleted of its energy reserve and the electrons come to a halt, the capacitor will have reached full (voltage) charge in the opposite polarity as when it started:



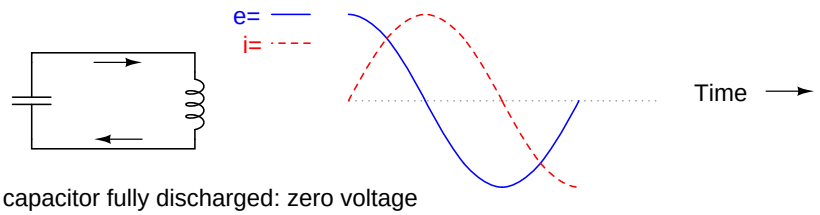
capacitor fully charged: voltage at (-) peak
inductor fully discharged: zero current

Now we're at a condition very similar to where we started: the capacitor at full charge and zero current in the circuit. The capacitor, as before, will begin to discharge through the inductor, causing an increase in current (in the opposite direction as before) and a decrease in voltage as it depletes its own energy reserve:



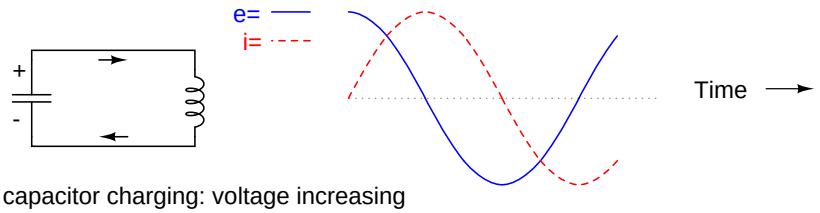
capacitor discharging: voltage decreasing
inductor charging: current increasing

Eventually the capacitor will discharge to zero volts, leaving the inductor fully charged with full current through it:



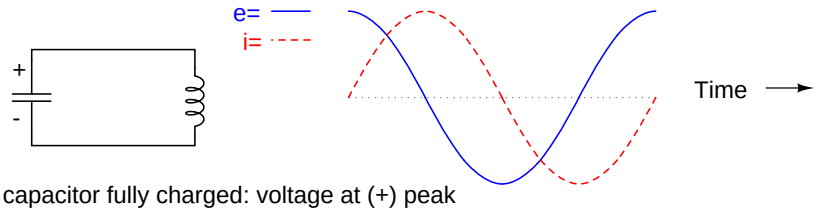
capacitor fully discharged: zero voltage
inductor fully charged: current at (-) peak

The inductor, desiring to maintain current in the same direction, will act like a source again, generating a voltage like a battery to continue the flow. In doing so, the capacitor will begin to charge up and the current will decrease in magnitude:



capacitor charging: voltage increasing
inductor discharging: current decreasing

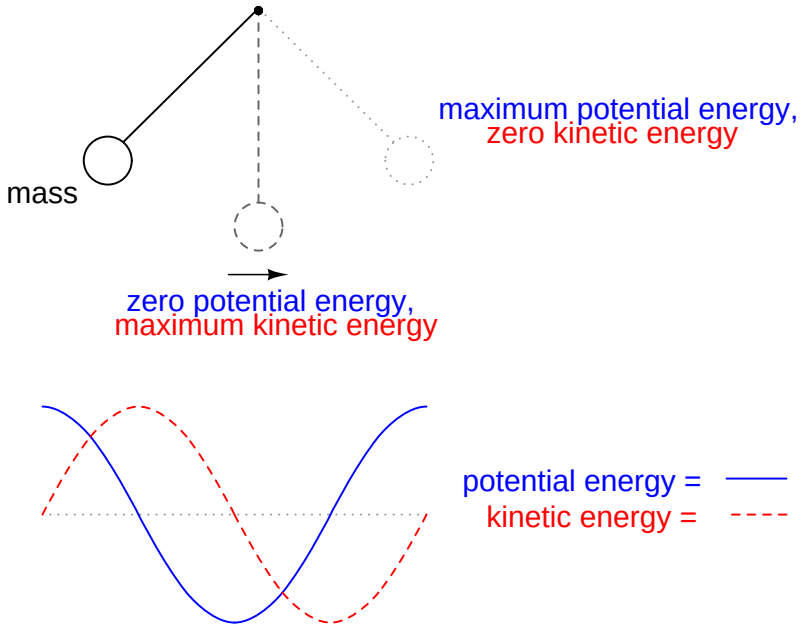
Eventually the capacitor will become fully charged again as the inductor expends all of its energy reserves trying to maintain current. The voltage will once again be at its positive peak and the current at zero. This completes one full cycle of the energy exchange between the capacitor and inductor:



capacitor fully charged: voltage at (+) peak
inductor fully discharged: zero current

This oscillation will continue with steadily decreasing amplitude due to power losses from stray resistances in the circuit, until the process stops altogether. Overall, this behavior is akin to that of a pendulum: as the pendulum mass swings back and forth, there is a transformation of energy taking place from kinetic (motion) to potential (height), in a similar fashion to the way energy is transferred in the capacitor/inductor circuit back and forth in the alternating forms of current (kinetic motion of electrons) and voltage (potential electric energy).

At the peak height of each swing of a pendulum, the mass briefly stops and switches directions. It is at this point that potential energy (height) is at a maximum and kinetic energy (motion) is at zero. As the mass swings back the other way, it passes quickly through a point where the string is pointed straight down. At this point, potential energy (height) is at zero and kinetic energy (motion) is at maximum. Like the circuit, a pendulum's back-and-forth oscillation will continue with a steadily dampened amplitude, the result of air friction (resistance) dissipating energy. Also like the circuit, the pendulum's position and velocity measurements trace two sine waves (90 degrees out of phase) over time:



In physics, this kind of natural sine-wave oscillation for a mechanical system is called *Simple Harmonic Motion* (often abbreviated as "SHM"). The same underlying principles govern both the oscillation of a capacitor/inductor circuit and the action of a pendulum, hence the similarity in effect. It is an interesting property of any pendulum that its periodic time is governed by the length of the string holding the mass, and not the weight of the mass itself. That is why a pendulum will keep swinging at the same frequency as the oscillations decrease in amplitude. The oscillation rate is independent of the *amount* of energy stored in it.

The same is true for the capacitor/inductor circuit. The rate of oscillation is strictly dependent on the sizes of the capacitor and inductor, not on the amount of voltage (or current) at each respective peak in the waves. The ability for such a circuit to store energy in the form of oscillating voltage and current has earned it the name *tank circuit*. Its property of maintaining a single, natural frequency regardless of how much or little energy is actually being stored in it gives it special significance in electric circuit design.

However, this tendency to oscillate, or *resonate*, at a particular frequency is not limited to circuits exclusively designed for that purpose. In fact, nearly any AC circuit with a combination of capacitance and inductance (commonly called an "LC circuit") will tend to manifest unusual effects when the AC power source frequency approaches that natural frequency. This is true regardless of the circuit's intended purpose.

If the power supply frequency for a circuit exactly matches the natural frequency of the circuit's LC combination, the circuit is said to be in a state of *resonance*. The unusual effects will reach maximum in this condition of resonance. For this reason, we need to be able to predict what the resonant frequency will be for various combinations of L and C, and be aware of what the effects of resonance are.

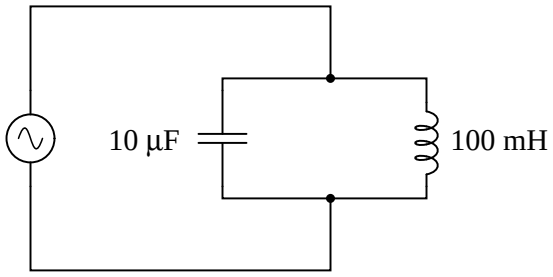
- **REVIEW:**

- A capacitor and inductor directly connected together form something called a *tank circuit*, which oscillates (or *resonates*) at one particular frequency. At that frequency, energy is alternately shuffled between the capacitor and the inductor in the form of alternating voltage and current 90 degrees out of phase with each other.

- When the power supply frequency for an AC circuit exactly matches that circuit's natural oscillation frequency as set by the L and C components, a condition of *resonance* will have been reached.

6.2 Simple parallel (tank circuit) resonance

A condition of resonance will be experienced in a tank circuit when the reactances of the capacitor and inductor are equal to each other. Because inductive reactance increases with increasing frequency and capacitive reactance decreases with increasing frequency, there will only be one frequency where these two reactances will be equal.



In the above circuit, we have a 10 μF capacitor and a 100 mH inductor. Since we know the equations for determining the reactance of each at a given frequency, and we're looking for that point where the two reactances are equal to each other, we can set the two reactance formulae equal to each other and solve for frequency algebraically:

$$X_L = 2\pi fL \qquad X_C = \frac{1}{2\pi fC}$$

. . . setting the two equal to each other, representing a condition of equal reactance (resonance) . . .

$$2\pi fL = \frac{1}{2\pi fC}$$

Multiplying both sides by f eliminates the f term in the denominator of the fraction . . .

$$2\pi f^2 L = \frac{1}{2\pi C}$$

Dividing both sides by 2πL leaves f^2 by itself on the left-hand side of the equation . . .

$$f^2 = \frac{1}{2\pi 2\pi LC}$$

Taking the square root of both sides of the equation leaves f by itself on the left side . . .

$$f = \frac{\sqrt{1}}{\sqrt{2\pi 2\pi LC}}$$

. . . simplifying . . .

$$\boxed{f = \frac{1}{2\pi \sqrt{LC}}}$$

So there we have it: a formula to tell us the resonant frequency of a tank circuit, given the values of inductance (L) in Henrys and capacitance (C) in Farads. Plugging in the values of L and C in our example circuit, we arrive at a resonant frequency of 159.155 Hz.

What happens at resonance is quite interesting. With capacitive and inductive reactances equal to each other, the total impedance increases to infinity, meaning that the tank circuit draws no current from the AC power source! We can calculate the individual impedances of the 10 μF capacitor and the 100 mH inductor and work through the parallel impedance formula to demonstrate this mathematically:

$$X_L = 2\pi fL$$

$$X_L = (2)(\pi)(159.155 \text{ Hz})(100 \text{ mH})$$

$$X_L = 100\ \Omega$$

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(\pi)(159.155 \text{ Hz})(10\ \mu\text{F})}$$

$$X_C = 100\ \Omega$$

As you might have guessed, I chose these component values to give resonance impedances that were easy to work with (100 Ω even). Now, we use the parallel impedance formula to see what happens to total Z:

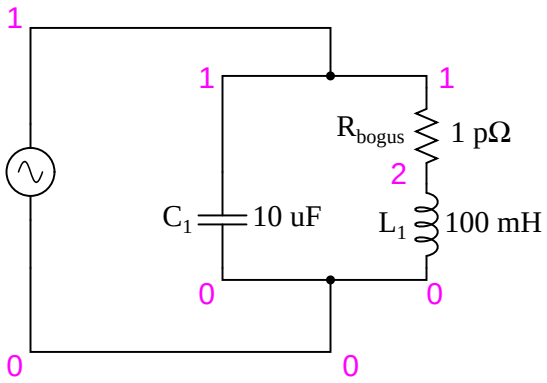
$$Z_{parallel} = \frac{1}{\frac{1}{Z_L} + \frac{1}{Z_C}}$$

$$Z_{parallel} = \frac{1}{\frac{1}{100\ \Omega \angle 90^o} + \frac{1}{100\ \Omega \angle -90^o}}$$

$$Z_{parallel} = \frac{1}{0.01 \angle -90^o + 0.01 \angle 90^o}$$

$$Z_{parallel} = \frac{1}{0} \quad \textit{Undefined!}$$

We can't divide any number by zero and arrive at a meaningful result, but we can say that the result approaches a value of *infinity* as the two parallel impedances get closer to each other. What this means in practical terms is that, the total impedance of a tank circuit is infinite (behaving as an *open circuit*) at resonance. We can plot the consequences of this over a wide power supply frequency range with a short SPICE simulation:




```
tank circuit frequency sweep
v1 1 0 ac 1 sin
c1 1 0 10u
* rbogus is necessary to eliminate a direct loop
* between v1 and l1, which SPICE can't handle
rbogus 1 2 1e-12
l1 2 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

```
freq       i(v1)        3.162E-04   1.000E-03   3.162E-03  1.0E-02
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  9.632E-03 .       .           .           .          *
1.053E+02  8.506E-03 .       .           .           .        * .
1.105E+02  7.455E-03 .       .           .           .      *   .
1.158E+02  6.470E-03 .       .           .           .     *    .
1.211E+02  5.542E-03 .       .           .           .   *      .
1.263E+02  4.663E-03 .       .           .           . *        .
1.316E+02  3.828E-03 .       .           .           .*         .
1.368E+02  3.033E-03 .       .           .          *.          .
1.421E+02  2.271E-03 .       .           .       *   .          .
1.474E+02  1.540E-03 .       .           .    *      .          .
1.526E+02  8.373E-04 .       .         * .           .          .
1.579E+02  1.590E-04 . *     .           .           .          .
1.632E+02  4.969E-04 .       .    *      .           .          .
1.684E+02  1.132E-03 .       .           . *         .          .
1.737E+02  1.749E-03 .       .           .    *      .          .
1.789E+02  2.350E-03 .       .           .       *   .          .
1.842E+02  2.934E-03 .       .           .          *.          .
1.895E+02  3.505E-03 .       .           .           .*         .
1.947E+02  4.063E-03 .       .           .           .  *       .
2.000E+02  4.609E-03 .       .           .           .    *     .
```

- -

The 1 pico-ohm (1 pΩ) resistor is placed in this SPICE analysis to overcome a limitation of SPICE: namely, that it cannot analyze a circuit containing a direct inductor-voltage source loop. A very low resistance value was chosen so as to have minimal effect on circuit behavior.

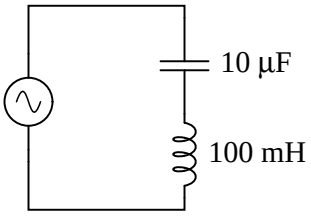
This SPICE simulation plots circuit current over a frequency range of 100 to 200 Hz in twenty even steps (100 and 200 Hz inclusive). Current magnitude on the graph increases from left to right, while frequency increases from top to bottom. The current in this circuit takes a sharp dip around the analysis point of 157.9 Hz, which is the closest analysis point to our predicted resonance frequency of 159.155 Hz. It is at this point that total current from the power source falls to zero.

Incidentally, the graph output produced by this SPICE computer analysis is more generally known as a *Bode plot*. Such graphs plot amplitude or phase shift on one axis and frequency on the other. The steepness of a Bode plot curve characterizes a circuit's "frequency response," or how sensitive it is to changes in frequency.

- **REVIEW:**
- Resonance occurs when capacitive and inductive reactances are equal to each other.
- For a tank circuit with no resistance (R), resonant frequency can be calculated with the following formula:
- $$f_{resonant} = \frac{1}{2\pi\sqrt{LC}}$$
- The total impedance of a parallel LC circuit approaches infinity as the power supply frequency approaches resonance.
- A *Bode plot* is a graph plotting waveform amplitude or phase on one axis and frequency on the other.

6.3 Simple series resonance

A similar effect happens in series inductive/capacitive circuits. When a state of resonance is reached (capacitive and inductive reactances equal), the two impedances cancel each other out and the total impedance drops to zero!



At 159.155 Hz:

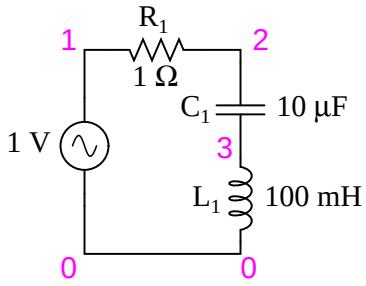
$Z_L = 0 + j100\ \Omega$ $Z_C = 0 - j100\ \Omega$

$Z_{series} = Z_L + Z_C$

$Z_{series} = (0 + j100\ \Omega) + (0 - j100\ \Omega)$

$Z_{series} = 0\ \Omega$

With the total series impedance equal to 0 Ω at the resonant frequency of 159.155 Hz, the result is a *short circuit* across the AC power source at resonance. In the circuit drawn above, this would not be good. I'll add a small resistor in series along with the capacitor and the inductor to keep the maximum circuit current somewhat limited, and perform another SPICE analysis over the same range of frequencies:




```
series lc circuit
v1 1 0 ac 1 sin
r1 1 2 1
c1 2 3 10u
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

```
freq       i(v1)           3.162E-02     1.000E-01  3.162E-01  1.0
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  1.038E-02 *             .             .          .    .
1.053E+02  1.176E-02 . *           .             .          .    .
1.105E+02  1.341E-02 .   *         .             .          .    .
1.158E+02  1.545E-02 .    *        .             .          .    .
1.211E+02  1.804E-02 .      *      .             .          .    .
1.263E+02  2.144E-02 .        *    .             .          .    .
1.316E+02  2.611E-02 .           * .             .          .    .
1.368E+02  3.296E-02 .             .*            .          .    .
```

```
1.421E+02  4.399E-02 .             .   *         .         .    .
1.474E+02  6.478E-02 .             .         *   .         .    .
1.526E+02  1.186E-01 .             .             . *       .    .
1.579E+02  5.324E-01 .             .             .         . *  .
1.632E+02  1.973E-01 .             .             .   *     .    .
1.684E+02  8.797E-02 .             .           * .         .    .
1.737E+02  5.707E-02 .             .      *      .         .    .
1.789E+02  4.252E-02 .             .   *         .         .    .
1.842E+02  3.406E-02 .            .*             .         .    .
1.895E+02  2.852E-02 .           *.              .         .    .
1.947E+02  2.461E-02 .         *  .              .         .    .
2.000E+02  2.169E-02 .       *    .              .         .    .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

As before, circuit current amplitude increases from left to right, while frequency increases from top to bottom. The peak is still seen to be at the plotted frequency point of 157.9 Hz, the closest analyzed point to our predicted resonance point of 159.155 Hz. This would suggest that our resonant frequency formula holds as true for simple series LC circuits as it does for simple parallel LC circuits, which is the case:

$$f_{resonant} = \frac{1}{2\pi\sqrt{LC}}$$

A word of caution is in order with series LC resonant circuits: because of the high currents which may be present in a series LC circuit at resonance, it is possible to produce dangerously high voltage drops across the capacitor and the inductor, as each component possesses significant impedance. We can edit the SPICE netlist in the above example to include a plot of voltage across the capacitor and inductor to demonstrate what happens:

```
series lc circuit
v1 1 0 ac 1 sin
r1 1 2 1
c1 2 3 10u
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1) v(2,3) v(3)
.end

legend:
*: i(v1)
+: v(2,3)
=: v(3)
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
(*)---------- 1.000E-02     3.162E-02     1.000E-01    0.3162  1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
(+)---------- 1.000E+00     3.162E+00     1.000E+01    31.62 100
```

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
(=)----------- 1.000E-01     1.000E+00     1.000E+01     100 1000
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
freq        i(v1)
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  1.038E-02 *     +    = .          .           .     .
1.053E+02  1.176E-02 . *    +    =.          .           .     .
1.105E+02  1.341E-02 .   *   +    =          .           .     .
1.158E+02  1.545E-02 .    *   +   .=         .           .     .
1.211E+02  1.804E-02 .      *   + . =        .           .     .
1.263E+02  2.144E-02 .        *  +.  =       .           .     .
1.316E+02  2.611E-02 .           *+    =     .           .     .
1.368E+02  3.296E-02 .             .*+   =    .           .     .
1.421E+02  4.399E-02 .             .   *+  =  .           .     .
1.474E+02  6.478E-02 .             .        *+=           .     .
1.526E+02  1.186E-01 .             .          .=*+        .     .
1.579E+02  5.324E-01 .             .          .       =   .   x .
1.632E+02  1.973E-01 .             .          .   = x     .     .
1.684E+02  8.797E-02 .             .        x =           .     .
1.737E+02  5.707E-02 .             .     +* = .           .     .
1.789E+02  4.252E-02 .             . + *  =   .           .     .
1.842E+02  3.406E-02 .           +.*   =      .           .     .
1.895E+02  2.852E-02 .         + *.   =       .           .     .
1.947E+02  2.461E-02 .      +  *  .   =       .           .     .
2.000E+02  2.169E-02 .     + *    .  =        .           .     .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

According to SPICE, voltage across the capacitor and inductor (plotted with "+" and "=" symbols, respectively) reach a peak somewhere between 100 and 1000 volts (marked by the "x" where the graphs overlap)! This is quite impressive for a power supply that only generates 1 volt. Needless to say, caution is in order when experimenting with circuits such as this.

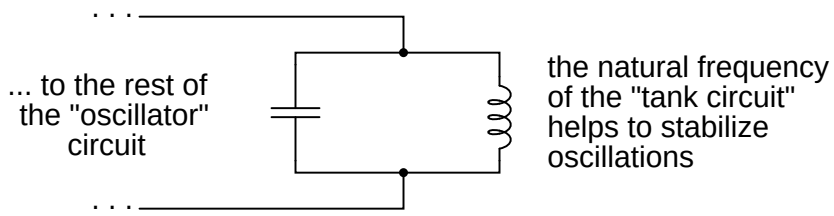
- **REVIEW:**
- The total impedance of a series LC circuit approaches zero as the power supply frequency approaches resonance.
- The same formula for determining resonant frequency in a simple tank circuit applies to simple series circuits as well.
- Extremely high voltages can be formed across the individual components of series LC circuits at resonance, due to high current flows and substantial individual component impedances.

6.4 Applications of resonance

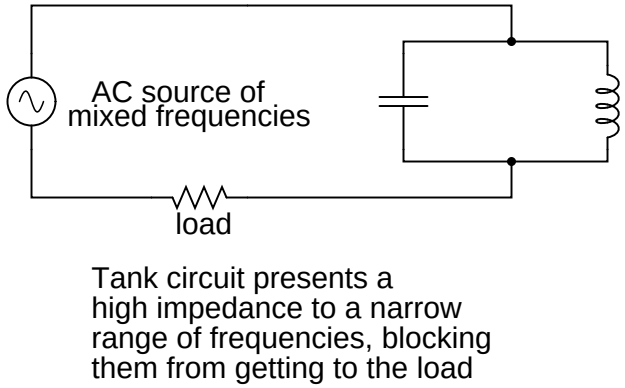
So far, the phenomenon of resonance appears to be a useless curiosity, or at most a nuisance to be avoided (especially if series resonance makes for a short-circuit across our AC voltage source!).

However, this is not the case. Resonance is a very valuable property of reactive AC circuits, employed in a variety of applications.

One use for resonance is to establish a condition of stable frequency in circuits designed to produce AC signals. Usually, a parallel (tank) circuit is used for this purpose, with the capacitor and inductor directly connected together, exchanging energy between each other. Just as a pendulum can be used to stabilize the frequency of a clock mechanism's oscillations, so can a tank circuit be used to stabilize the electrical frequency of an AC *oscillator* circuit. As was noted before, the frequency set by the tank circuit is solely dependent upon the values of L and C, and not on the magnitudes of voltage or current present in the oscillations:



Another use for resonance is in applications where the effects of greatly increased or decreased impedance at a particular frequency is desired. A resonant circuit can be used to "block" (present high impedance toward) a frequency or range of frequencies, thus acting as a sort of frequency "filter" to strain certain frequencies out of a mix of others. In fact, these particular circuits are called *filters*, and their design constitutes a discipline of study all by itself:



In essence, this is how analog radio receiver tuner circuits work to filter, or select, one station frequency out of the mix of different radio station frequency signals intercepted by the antenna.

- **REVIEW:**
- Resonance can be employed to maintain AC circuit oscillations at a constant frequency, just as a pendulum can be used to maintain constant oscillation speed in a timekeeping mechanism.
- Resonance can be exploited for its impedance properties: either dramatically increasing or decreasing impedance for certain frequencies. Circuits designed to screen certain frequencies out of a mix of different frequencies are called *filters*.

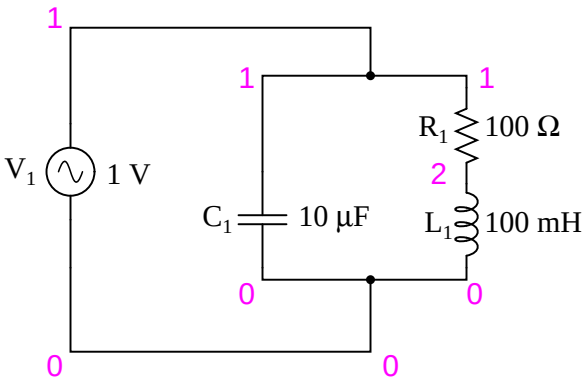
6.5 Resonance in series-parallel circuits

In simple reactive circuits with little or no resistance, the effects of radically altered impedance will manifest at the resonance frequency predicted by the equation given earlier. In a parallel (tank) LC circuit, this means infinite impedance at resonance. In a series LC circuit, it means zero impedance at resonance:

$$f_{resonant} = \frac{1}{2\pi\sqrt{LC}}$$

However, as soon as significant levels of resistance are introduced into most LC circuits, this simple calculation for resonance becomes invalid. We'll take a look at several LC circuits with added resistance, using the same values for capacitance and inductance as before: 10 μF and 100 mH, respectively. According to our simple equation, the resonant frequency should be 159.155 Hz. Watch, though, where current reaches maximum or minimum in the following SPICE analyses:

Parallel LC with resistance in series with L

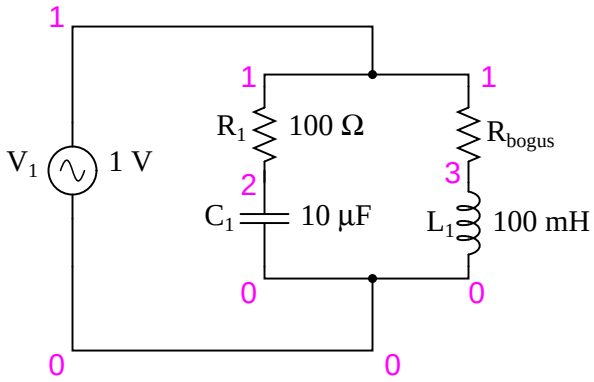



```
resonant circuit
v1 1 0 ac 1 sin
c1 1 0 10u
r1 1 2 100
l1 2 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

```
freq       i(v1)           7.079E-03     7.943E-03     8.913E-03
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  7.387E-03 .             .    *        .             .
1.053E+02  7.242E-03 .             .  *          .             .
1.105E+02  7.115E-03 .             .*            .             .
1.158E+02  7.007E-03 .            *.             .             .
```

```
1.211E+02  6.921E-03 .          *  .             .             .
1.263E+02  6.859E-03 .         *   .             .             .
1.316E+02  6.823E-03 .         *   .             .             .
1.368E+02  6.813E-03 .        *    .             .             .
1.421E+02  6.830E-03 .         *   .             .             .
1.474E+02  6.874E-03 .         *   .             .             .
1.526E+02  6.946E-03 .           * .             .             .
1.579E+02  7.044E-03 .            *.             .             .
1.632E+02  7.167E-03 .             .*            .             .
1.684E+02  7.315E-03 .             .   *         .             .
1.737E+02  7.485E-03 .             .      *      .             .
1.789E+02  7.676E-03 .             .         *   .             .
1.842E+02  7.886E-03 .             .            *.             .
1.895E+02  8.114E-03 .             .             .  *          .
1.947E+02  8.358E-03 .             .             .     *       .
2.000E+02  8.616E-03 .             .             .        *    .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Minimum current at 136.8 Hz instead of 159.2 Hz!
```

Parallel LC with resistance in series with C



Here, an extra resistor (R_{bogus}) is necessary to prevent SPICE from encountering trouble in analysis. SPICE can't handle an inductor connected directly in parallel with any voltage source or any other inductor, so the addition of a series resistor is necessary to "break up" the voltage source/inductor loop that would otherwise be formed. This resistor is chosen to be a *very* low value for minimum impact on the circuit's behavior.

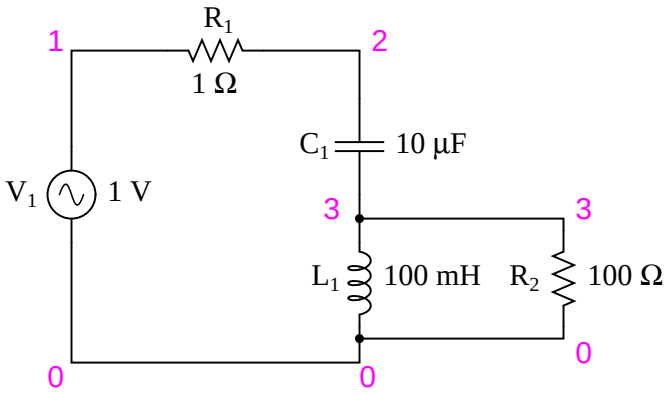
```
resonant circuit
v1 1 0 ac 1 sin
r1 1 2 100
c1 2 0 10u
rbogus 1 3 1e-12
l1 3 0 100m
```

```
.ac lin 20 100 400
.plot ac i(v1)
.end
```

```
freq       i(v1)            7.943E-03     1.000E-02     1.259E-02
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  1.176E-02 .             .             .         *   .
1.158E+02  9.635E-03 .             .           * .             .
1.316E+02  8.257E-03 .             . *           .             .
1.474E+02  7.430E-03 .         *   .             .             .
1.632E+02  6.998E-03 .     *       .             .             .
1.789E+02  6.835E-03 .    *        .             .             .
1.947E+02  6.839E-03 .    *        .             .             .
2.105E+02  6.941E-03 .     *       .             .             .
2.263E+02  7.093E-03 .      *      .             .             .
2.421E+02  7.268E-03 .        *    .             .             .
2.579E+02  7.449E-03 .         *   .             .             .
2.737E+02  7.626E-03 .           * .             .             .
2.895E+02  7.794E-03 .            *.             .             .
3.053E+02  7.951E-03 .             *             .             .
3.211E+02  8.096E-03 .             .*            .             .
3.368E+02  8.230E-03 .             . *           .             .
3.526E+02  8.352E-03 .             .  *          .             .
3.684E+02  8.464E-03 .             .   *         .             .
3.842E+02  8.567E-03 .             .    *        .             .
4.000E+02  8.660E-03 .             .    *        .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Minimum current at roughly 180 Hz instead of 159.2 Hz!
```

Switching our attention to series LC circuits, we experiment with placing significant resistances in parallel with either L or C. In the following series circuit examples, a 1 Ω resistor (R_1) is placed in series with the inductor and capacitor to limit total current at resonance. The "extra" resistance inserted to influence resonant frequency effects is the 100 Ω resistor, R_2:

Series LC with resistance in parallel with L



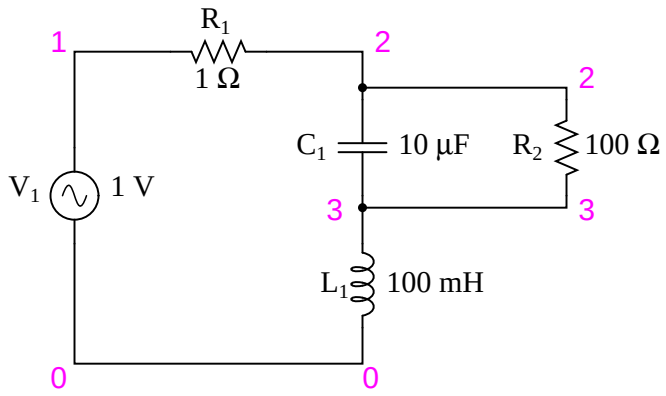

```
resonant circuit
v1 1 0 ac 1 sin
r1 1 2 1
c1 2 3 10u
l1 3 0 100m
r2 3 0 100
.ac lin 20 100 400
.plot ac i(v1)
.end
```

```
freq       i(v1)            1.000E-02     1.259E-02     1.585E-02
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  8.488E-03 .   *         .             .             .
1.158E+02  1.034E-02 .            . *           .             .
1.316E+02  1.204E-02 .            .           *  .             .
1.474E+02  1.336E-02 .            .              .   *         .
1.632E+02  1.415E-02 .            .              .      *      .
1.789E+02  1.447E-02 .            .              .       *     .
1.947E+02  1.445E-02 .            .              .       *     .
2.105E+02  1.424E-02 .            .              .      *      .
2.263E+02  1.393E-02 .            .              .     *       .
2.421E+02  1.360E-02 .            .              .    *        .
2.579E+02  1.327E-02 .            .              .  *          .
2.737E+02  1.296E-02 .            .              . *           .
2.895E+02  1.269E-02 .            .             *              .
3.053E+02  1.244E-02 .            .            *.              .
3.211E+02  1.222E-02 .            .           * .              .
3.368E+02  1.202E-02 .            .          *  .              .
3.526E+02  1.185E-02 .            .         *   .              .
3.684E+02  1.169E-02 .            .        *    .              .
```

```
3.842E+02  1.155E-02 .             .        *    .             .
4.000E+02  1.143E-02 .             .       *     .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Maximum current at roughly 178.9 Hz instead of 159.2 Hz!
```

And finally, a series LC circuit with the significant resistance in parallel with the capacitor:

Series LC with resistance in parallel with C




```
resonant circuit
v1 1 0 ac 1 sin
r1 1 2 1
c1 2 3 10u
r2 2 3 100
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end

freq      i(v1)
freq       i(v1)           1.259E-02     1.413E-02     1.585E-02
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  1.336E-02 .             .      *      .             .
1.053E+02  1.363E-02 .             .         *   .             .
1.105E+02  1.387E-02 .             .           * .             .
1.158E+02  1.408E-02 .             .             *             .
1.211E+02  1.426E-02 .             .             .*            .
1.263E+02  1.439E-02 .             .             . *           .
1.316E+02  1.447E-02 .             .             .  *          .
1.368E+02  1.450E-02 .             .             .  *          .
1.421E+02  1.447E-02 .             .             .  *          .
1.474E+02  1.438E-02 .             .             . *           .
```

```
1.526E+02  1.424E-02 .             .             .*            .
1.579E+02  1.405E-02 .             .            *.             .
1.632E+02  1.382E-02 .             .          *  .             .
1.684E+02  1.355E-02 .             .        *    .             .
1.737E+02  1.325E-02 .             .     *       .             .
1.789E+02  1.293E-02 .             .  *          .             .
1.842E+02  1.259E-02 .             *             .             .
1.895E+02  1.225E-02 .          *  .             .             .
1.947E+02  1.190E-02 .      *      .             .             .
2.000E+02  1.155E-02 .  *          .             .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Maximum current at 136.8 Hz instead of 159.2 Hz!
```

The tendency for added resistance to skew the point at which impedance reaches a maximum or minimum in an LC circuit is called *antiresonance*. The astute observer will notice a pattern between the four SPICE examples given above, in terms of how resistance affects the resonant peak of a circuit:

- **Parallel ("tank") LC circuit:**
- R in series with L: resonant frequency shifted *down*
- R in series with C: resonant frequency shifted *up*

- **Series LC circuit:**
- R in parallel with L: resonant frequency shifted *up*
- R in parallel with C: resonant frequency shifted *down*

Again, this illustrates the complementary nature of capacitors and inductors: how resistance in series with one creates an antiresonance effect equivalent to resistance in parallel with the other. If you look even closer to the four SPICE examples given, you'll see that the frequencies are shifted by the *same amount*, and that the shape of the complementary graphs are mirror-images of each other!

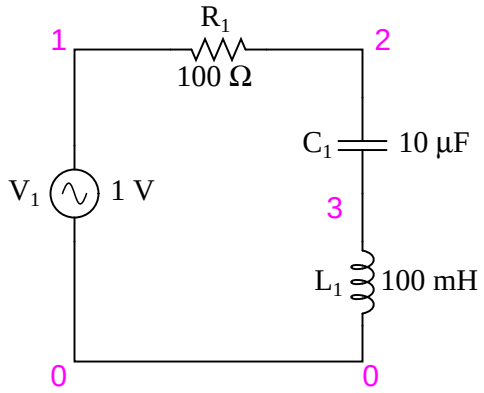
Antiresonance is an effect that resonant circuit designers must be aware of. The equations for determining antiresonance "shift" are complex, and will not be covered in this brief lesson. It should suffice the beginning student of electronics to understand that the effect exists, and what its general tendencies are.

Added resistance in an LC circuit is no academic matter. While it is possible to manufacture capacitors with negligible unwanted resistances, inductors are typically plagued with substantial amounts of resistance due to the long lengths of wire used in their construction. What is more, the resistance of wire tends to increase as frequency goes up, due to a strange phenomenon known as the *skin effect* where AC current tends to be excluded from travel through the very center of a wire, thereby reducing the wire's effective cross-sectional area. Thus, inductors not only have resistance, but *changing, frequency-dependent* resistance at that.

As if the resistance of an inductor's wire weren't enough to cause problems, we also have to contend with the "core losses" of iron-core inductors, which manifest themselves as added resistance in the circuit. Since iron is a conductor of electricity as well as a conductor of magnetic flux, changing flux produced by alternating current through the coil will tend to induce electric currents in the core itself (*eddy currents*). This effect can be thought of as though the iron core of the transformer were a sort of secondary transformer coil powering a resistive load: the less-than-perfect conductivity of the iron metal. This effects can be minimized with laminated cores, good core design and high-grade materials, but never completely eliminated.

One notable exception to the rule of circuit resistance causing a resonant frequency shift is the case of series resistor-inductor-capacitor ("RLC") circuits. So long as *all* components are connected in series with each other, the resonant frequency of the circuit will be unaffected by the resistance:

Series LC with resistance in series




```
series rlc circuit
v1 1 0 ac 1 sin
r1 1 2 100
c1 2 3 10u
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

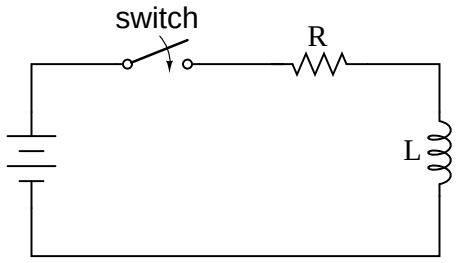
```
freq       i(v1)          7.943E-03     8.913E-03     1.000E-02
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  7.202E-03 *         .             .             .
1.053E+02  7.617E-03 .     *   .             .             .
1.105E+02  8.017E-03 .         .*            .             .
1.158E+02  8.396E-03 .         .      *      .             .
1.211E+02  8.747E-03 .         .           * .             .
1.263E+02  9.063E-03 .         .             . *           .
1.316E+02  9.339E-03 .         .             .     *       .
```

```
1.368E+02  9.570E-03 .         .             .         *    .
1.421E+02  9.752E-03 .         .             .           *  .
1.474E+02  9.883E-03 .         .             .             *.
1.526E+02  9.965E-03 .         .             .              .
1.579E+02  9.999E-03 .         .             .              *
1.632E+02  9.988E-03 .         .             .              *
1.684E+02  9.936E-03 .         .             .             *.
1.737E+02  9.850E-03 .         .             .            * .
1.789E+02  9.735E-03 .         .             .           *  .
1.842E+02  9.595E-03 .         .             .         *    .
1.895E+02  9.437E-03 .         .             .       *      .
1.947E+02  9.265E-03 .         .             .     *        .
2.000E+02  9.082E-03 .         .             . *            .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Maximum current at 159.2 Hz once again!
```

Note that the peak of the current graph has not changed from the earlier series LC circuit (the one with the 1 Ω token resistance in it), even though the resistance is now 100 times greater. The only thing that has changed is the "sharpness" of the curve. Obviously, this circuit does not resonate as strongly as one with less series resistance (it is said to be "less selective"), but at least it has the same natural frequency!

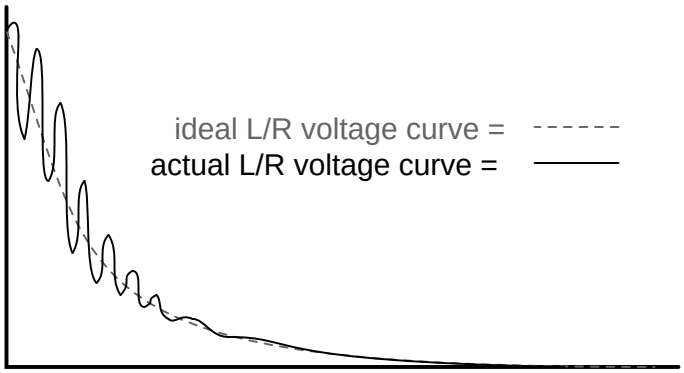
It is noteworthy that antiresonance has the effect of dampening the oscillations of free-running LC circuits such as tank circuits. In the beginning of this chapter we saw how a capacitor and inductor connected directly together would act something like a pendulum, exchanging voltage and current peaks just like a pendulum exchanges kinetic and potential energy. In a perfect tank circuit (no resistance), this oscillation would continue forever, just as a frictionless pendulum would continue to swing at its resonant frequency forever. But frictionless machines are difficult to find in the real world, and so are lossless tank circuits. Energy lost through resistance (or inductor core losses or radiated electromagnetic waves or . . .) in a tank circuit will cause the oscillations to decay in amplitude until they are no more. If enough energy losses are present in a tank circuit, it will fail to resonate at all.

Antiresonance's dampening effect is more than just a curiosity: it can be used quite effectively to eliminate *unwanted* oscillations in circuits containing stray inductances and/or capacitances, as almost all circuits do. Take note of the following L/R time delay circuit:



The idea of this circuit is simple: to "charge" the inductor when the switch is closed. The rate of inductor charging will be set by the ratio L/R, which is the time constant of the circuit in seconds. However, if you were to build such a circuit, you might find unexpected oscillations (AC) of voltage

across the inductor when the switch is closed. Why is this? There's no capacitor in the circuit, so how can we have resonant oscillation with just an inductor, resistor, and battery?



All inductors contain a certain amount of stray capacitance due to turn-to-turn and turn-to-core insulation gaps. Also, the placement of circuit conductors may create stray capacitance. While clean circuit layout is important in eliminating much of this stray capacitance, there will always be some that you cannot eliminate. If this causes resonant problems (unwanted AC oscillations), added resistance may be a way to combat it. If resistor R is large enough, it will cause a condition of antiresonance, dissipating enough energy to prohibit the inductance and stray capacitance from sustaining oscillations for very long.

Interestingly enough, the principle of employing resistance to eliminate unwanted resonance is one frequently used in the design of mechanical systems, where any moving object with mass is a potential resonator. A very common application of this is the use of shock absorbers in automobiles. Without shock absorbers, cars would bounce wildly at their resonant frequency after hitting any bump in the road. The shock absorber's job is to introduce a strong antiresonant effect by dissipating energy hydraulically (in the same way that a resistor dissipates energy electrically).

- **REVIEW:**
- Added resistance to an LC circuit can cause a condition known as *antiresonance*, where the peak impedance effects happen at frequencies other than that which gives equal capacitive and inductive reactances.
- Unwanted resistances inherent in real-world inductors can contribute greatly to conditions of antiresonance. One source of such resistance is the *skin effect*, caused by the exclusion of AC current from the center of conductors. Another source is that of *core losses* in iron-core inductors.
- In a simple series LC circuit containing resistance (an "RLC" circuit), resistance does *not* produce antiresonance. Resonance still occurs when capacitive and inductive reactances are equal.

6.6 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 7

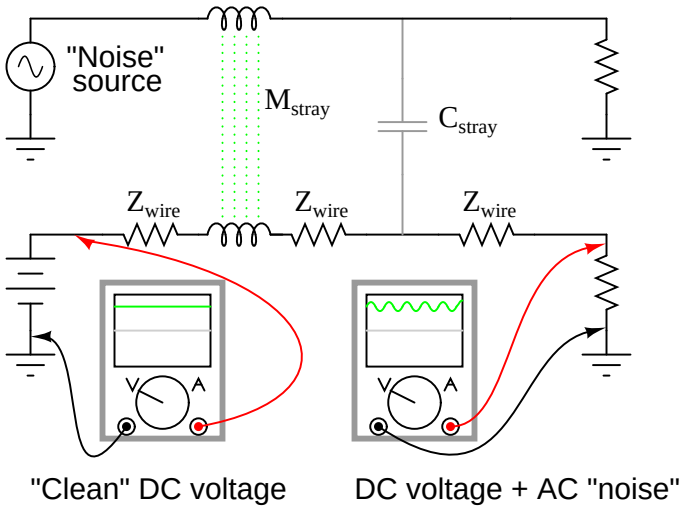
MIXED-FREQUENCY AC SIGNALS

7.1 Introduction

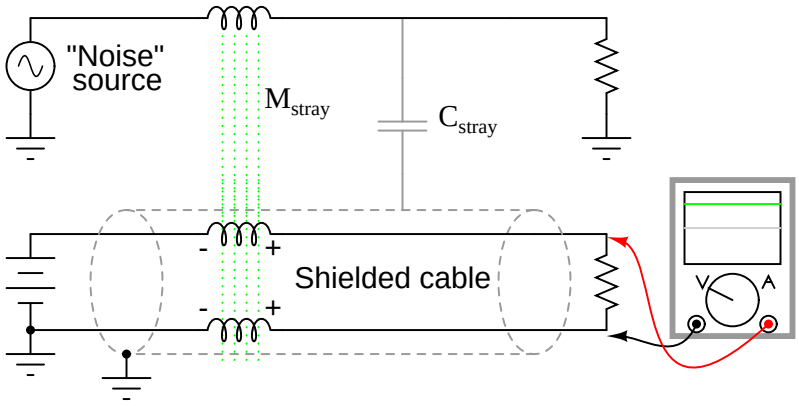
In our study of AC circuits thus far, we've explored circuits powered by a single-frequency sine voltage waveform. In many applications of electronics, though, single-frequency signals are the exception rather than the rule. Quite often we may encounter circuits where multiple frequencies of voltage coexist simultaneously. Also, circuit waveforms may be something other than sine-wave shaped, in which case we call them *non-sinusoidal waveforms*.

Additionally, we may encounter situations where DC is mixed with AC: where a waveform is superimposed on a steady (DC) signal. The result of such a mix is a signal varying in intensity, but never changing polarity, or changing polarity asymmetrically (spending more time positive than negative, for example). Since DC does not alternate as AC does, its "frequency" is said to be zero, and any signal containing DC along with a signal of varying intensity (AC) may be rightly called a mixed-frequency signal as well. In any of these cases where there is a mix of frequencies in the same circuit, analysis is more complex than what we've seen up to this point.

Sometimes mixed-frequency voltage and current signals are created accidentally. This may be the result of unintended connections between circuits – called *coupling* – made possible by stray capacitance and/or inductance between the conductors of those circuits. A classic example of coupling phenomenon is seen frequently in industry where DC signal wiring is placed in close proximity to AC power wiring. The nearby presence of high AC voltages and currents may cause "foreign" voltages to be impressed upon the length of the signal wiring. Stray capacitance formed by the electrical insulation separating power conductors from signal conductors may cause voltage (with respect to earth ground) from the power conductors to be impressed upon the signal conductors, while stray inductance formed by parallel runs of wire in conduit may cause current from the power conductors to electromagnetically induce voltage along the signal conductors. The result is a mix of DC and AC at the signal load. The following schematic shows how an AC "noise" source may "couple" to a DC circuit through mutual inductance (M_{stray}) and capacitance (C_{stray}) along the length of the conductors.

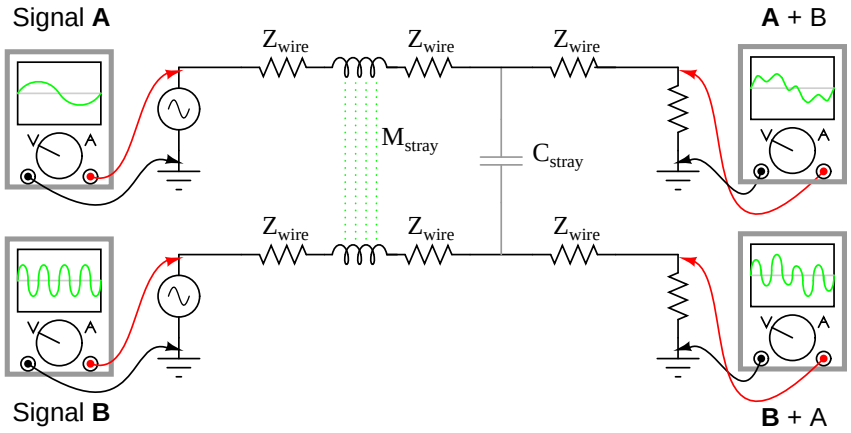


When stray AC voltages from a "noise" source mix with DC signals conducted along signal wiring, the results are usually undesirable. For this reason, power wiring and low-level signal wiring should *always* be routed through separated, dedicated metal conduit, and signals should be conducted via 2-conductor "twisted pair" cable rather than through a single wire and ground connection:

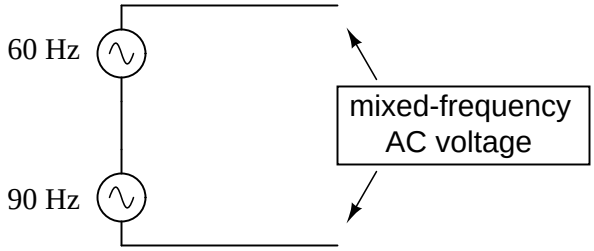


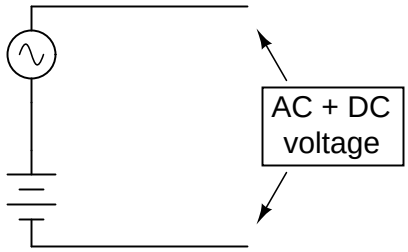
The grounded cable shield – a wire braid or metal foil wrapped around the two insulated conductors – isolates both conductors from electrostatic (capacitive) coupling by blocking any external electric fields, while the parallal proximity of the two conductors effectively cancels any electromagnetic (mutually inductive) coupling because any induced noise voltage will be approximately equal in magnitude and opposite in phase along both conductors, canceling each other at the receiving end for a net (differential) noise voltage of almost zero. Polarity marks placed near each inductive portion of signal conductor length shows how the induced voltages are phased in such a way as to cancel one another.

Coupling may also occur between two sets of conductors carrying AC signals, in which case both signals may become "mixed" with each other:



Coupling is but one example of how signals of different frequencies may become mixed. Whether it be AC mixed with DC, or two AC signals mixing with each other, signal coupling via stray inductance and capacitance is usually accidental and undesired. In other cases, mixed-frequency signals are the result of intentional design or they may be an intrinsic quality of a signal. It is generally quite easy to create mixed-frequency signal sources. Perhaps the easiest way is to simply connect voltage sources in series:





Some computer communications networks operate on the principle of superimposing high-frequency voltage signals along 60 Hz power-line conductors, so as to convey computer data along existing lengths of power cabling. This technique has been used for years in electric power distribution networks to communicate load data along high-voltage power lines. Certainly these are examples of mixed-frequency AC voltages, under conditions that are deliberately established.

In some cases, mixed-frequency signals may be produced by a single voltage source. Such is the

case with microphones, which convert audio-frequency air pressure waves into corresponding voltage waveforms. The particular mix of frequencies in the voltage signal output by the microphone is dependent on the sound being reproduced. If the sound waves consist of a single, pure note or tone, the voltage waveform will likewise be a sine wave at a single frequency. If the sound wave is a chord or other harmony of several notes, the resulting voltage waveform produced by the microphone will consist of those frequencies mixed together. Very few natural sounds consist of single, pure sine wave vibrations but rather are a mix of different frequency vibrations at different amplitudes.

Musical *chords* are produced by blending one frequency with other frequencies of particular fractional multiples of the first. However, investigating a little further, we find that even a single piano note (produced by a plucked string) consists of one predominant frequency mixed with several other frequencies, each frequency a whole-number multiple of the first (called *harmonics*, while the first frequency is called the *fundamental*). An illustration of these terms is shown below with a fundamental frequency of 1000 Hz (an arbitrary figure chosen for this example), each of the frequency multiples appropriately labeled:

```
FOR A "BASE" FREQUENCY OF 1000 Hz:
Frequency (Hz)           Term
-----------------------------------------
1000 --------- 1st harmonic, or fundamental
2000 --------- 2nd harmonic
3000 --------- 3rd harmonic
4000 --------- 4th harmonic
5000 --------- 5th harmonic
6000 --------- 6th harmonic
7000 --------- 7th harmonic
ad infinitum
```

Sometimes the term "overtone" is used to describe the a harmonic frequency produced by a musical instrument. The "first" overtone is the first harmonic frequency *greater than* the fundamental. If we had an instrument producing the entire range of harmonic frequencies shown in the table above, the first overtone would be 2000 Hz (the 2nd harmonic), while the second overtone would be 3000 Hz (the 3rd harmonic), etc. However, this application of the term "overtone" is specific to particular instruments.

It so happens that certain instruments are incapable of producing certain types of harmonic frequencies. For example, an instrument made from a tube that is open on one end and closed on the other (such as a bottle, which produces sound when air is blown across the opening) is incapable of producing even-numbered harmonics. Such an instrument set up to produce a fundamental frequency of 1000 Hz would also produce frequencies of 3000 Hz, 5000 Hz, 7000 Hz, etc, but would *not* produce 2000 Hz, 4000 Hz, 6000 Hz, or any other even-multiple frequencies of the fundamental. As such, we would say that the first overtone (the first frequency greater than the fundamental) in such an instrument would be 3000 Hz (the 3rd harmonic), while the second overtone would be 5000 Hz (the 5th harmonic), and so on.

A pure sine wave (single frequency), being entirely devoid of any harmonics, sounds very "flat" and "featureless" to the human ear. Most musical instruments are incapable of producing sounds this simple. What gives each instrument its distinctive tone is the same phenomenon that gives each

person a distinctive voice: the unique blending of harmonic waveforms with each fundamental note, described by the physics of motion for each unique object producing the sound.

Brass instruments do not possess the same "harmonic content" as woodwind instruments, and neither produce the same harmonic content as stringed instruments. A distinctive blend of frequencies is what gives a musical instrument its characteristic tone. As anyone who has played guitar can tell you, steel strings have a different sound than nylon strings. Also, the tone produced by a guitar string changes depending on where along its length it is plucked. These differences in tone, as well, are a result of different harmonic content produced by differences in the mechanical vibrations of an instrument's parts. All these instruments produce harmonic frequencies (whole-number multiples of the fundamental frequency) when a single note is played, but the relative amplitudes of those harmonic frequencies are different for different instruments. In musical terms, the measure of a tone's harmonic content is called *timbre* or *color*.

Musical tones become even more complex when the resonating element of an instrument is a two-dimensional surface rather than a one-dimensional string. Instruments based on the vibration of a string (guitar, piano, banjo, lute, dulcimer, etc.) or of a column of air in a tube (trumpet, flute, clarinet, tuba, pipe organ, etc.) tend to produce sounds composed of a single frequency (the "fundamental") and a mix of harmonics. Instruments based on the vibration of a flat plate (steel drums, and some types of bells), however, produce a much broader range of frequencies, not limited to whole-number multiples of the fundamental. The result is a distinctive tone that some people find acoustically offensive.

As you can see, music provides a rich field of study for mixed frequencies and their effects. Later sections of this chapter will refer to musical instruments as sources of waveforms for analysis in more detail.

- **REVIEW:**

- A *sinusoidal* waveform is one shaped exactly like a sine wave.

- A *non-sinusoidal* waveform can be anything from a distorted sine-wave shape to something completely different like a square wave.

- Mixed-frequency waveforms can be accidently created, purposely created, or simply exist out of necessity. Most musical tones, for instance, are not composed of a single frequency sine-wave, but are rich blends of different frequencies.

- When multiple sine waveforms are mixed together (as is often the case in music), the lowest frequency sine-wave is called the *fundamental*, and the other sine-waves whose frequencies are whole-number multiples of the fundamental wave are called *harmonics*.

- An *overtone* is a harmonic produced by a particular device. The "first" overtone is the first frequency greater than the fundamental, while the "second" overtone is the next greater frequency produced. Successive overtones may or may not correspond to incremental harmonics, depending on the device producing the mixed frequencies. Some devices and systems do not permit the establishment of certain harmonics, and so their overtones would only include some (not all) harmonic frequencies.

7.2 Square wave signals

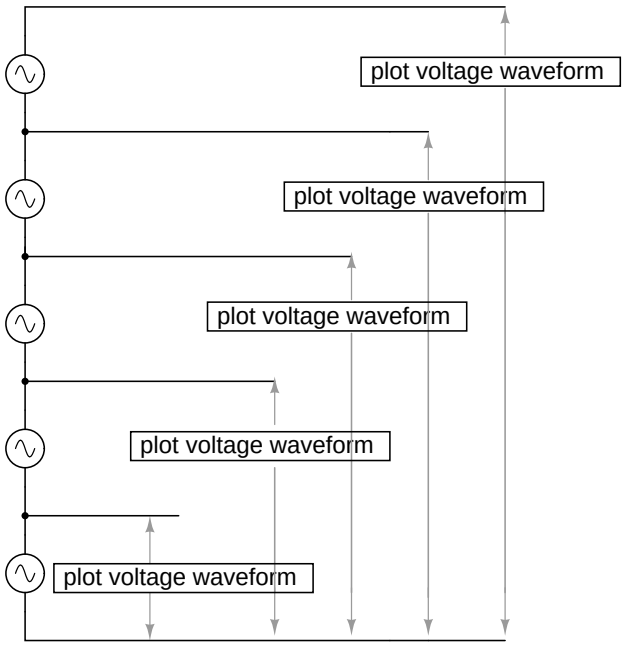
It has been found that *any* repeating, non-sinusoidal waveform can be equated to a combination of DC voltage, sine waves, and/or cosine waves (sine waves with a 90 degree phase shift) at various amplitudes and frequencies. This is true no matter how strange or convoluted the waveform in question may be. So long as it repeats itself regularly over time, it is reducible to this series of sinusoidal waves. In particular, it has been found that square waves are mathematically equivalent to the sum of a sine wave at that same frequency, plus an infinite series of odd-multiple frequency sine waves at diminishing amplitude:

1 V (peak) repeating square wave at 50 Hz is equivalent to:

$$\left(\frac{4}{\pi}\right)(\text{1 V peak sine wave at 50 Hz})$$
$$+ \left(\frac{4}{\pi}\right)(\text{1/3 V peak sine wave at 150 Hz})$$
$$+ \left(\frac{4}{\pi}\right)(\text{1/5 V peak sine wave at 250 Hz})$$
$$+ \left(\frac{4}{\pi}\right)(\text{1/7 V peak sine wave at 350 Hz})$$
$$+ \left(\frac{4}{\pi}\right)(\text{1/9 V peak sine wave at 450 Hz})$$
$$+ \textit{ . . . ad infinitum . . .}$$

This truth about waveforms at first may seem too strange to believe. However, if a square wave is actually an infinite series of sine wave harmonics added together, it stands to reason that we should be able to prove this by adding together several sine wave harmonics to produce a close approximation of a square wave. This reasoning is not only sound, but easily demonstrated with SPICE.

The circuit we'll be simulating is nothing more than several sine wave AC voltage sources of the proper amplitudes and frequencies connected together in series. We'll use SPICE to plot the voltage waveforms across successive additions of voltage sources, like this:



In this particular SPICE simulation, I've summed the 1st, 3rd, 5th, 7th, and 9th harmonic voltage sources in series for a total of five AC voltage sources. The fundamental frequency is 50 Hz and each harmonic is, of course, an integer multiple of that frequency. The amplitude (voltage) figures are not random numbers; rather, they have been arrived at through the equations shown in the frequency series (the fraction $4/\pi$ multiplied by 1, 1/3, 1/5, 1/7, etc. for each of the increasing odd harmonics).

```
building a squarewave
v1 1 0 sin (0 1.27324 50 0 0)       1st harmonic (50 Hz)
v3 2 1 sin (0 424.413m 150 0 0)     3rd harmonic
v5 3 2 sin (0 254.648m 250 0 0)     5th harmonic
v7 4 3 sin (0 181.891m 350 0 0)     7th harmonic
v9 5 4 sin (0 141.471m 450 0 0)     9th harmonic
r1 5 0 10k
.tran 1m 20m
.plot tran v(1,0)    Plot 1st harmonic
.plot tran v(2,0)    Plot 1st + 3rd harmonics
.plot tran v(3,0)    Plot 1st + 3rd + 5th harmonics
.plot tran v(4,0)    Plot 1st + 3rd + 5th + 7th harmonics
.plot tran v(5,0)    Plot 1st + . . . + 9th harmonics
.end
```

I'll narrate the analysis step by step from here, explaining what it is we're looking at. In this first plot, we see the fundamental-frequency sine-wave of 50 Hz by itself. It is nothing but a pure sine shape, with no additional harmonic content. This is the kind of waveform produced by an ideal AC power source:

```
time       v(1)          -1.000E+00    0.000E+00    1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .       .            *            .       .
1.000E-03  3.915E-01 .       .            .    *       .       .
2.000E-03  7.414E-01 .       .            .         *  .       .
3.000E-03  1.020E+00 .       .            .            *       .
4.000E-03  1.199E+00 .       .            .            .  *    .
5.000E-03  1.261E+00 .       .            .            .   *   .
6.000E-03  1.199E+00 .       .            .            .  *    .
7.000E-03  1.020E+00 .       .            .            *       .
8.000E-03  7.405E-01 .       .            .         *  .       .
9.000E-03  3.890E-01 .       .            .    *       .       .
1.000E-02 -5.819E-04 .       .            *            .       .
1.100E-02 -3.901E-01 .       .        *   .            .       .
1.200E-02 -7.414E-01 .       .   *        .            .       .
1.300E-02 -1.020E+00 .       *            .            .       .
1.400E-02 -1.199E+00 .    *  .            .            .       .
1.500E-02 -1.261E+00 .   *   .            .            .       .
1.600E-02 -1.199E+00 .    *  .            .            .       .
1.700E-02 -1.020E+00 .       *            .            .       .
1.800E-02 -7.405E-01 .       .   *        .            .       .
1.900E-02 -3.890E-01 .       .        *   .            .       .
2.000E-02  5.819E-04 .       .            *            .       .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Next, we see what happens when this clean and simple waveform is combined with the third harmonic (three times 50 Hz, or 150 Hz). Suddenly, it doesn't look like a clean sine wave any more:

```
time       v(2)          -1.000E+00    0.000E+00    1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .       .            *            .       .
1.000E-03  7.199E-01 .       .            .         *  .       .
2.000E-03  1.108E+00 .       .            .            . *     .
3.000E-03  1.135E+00 .       .            .            . *     .
4.000E-03  9.672E-01 .       .            .            *       .
5.000E-03  8.731E-01 .       .            .          * .       .
6.000E-03  9.751E-01 .       .            .            *       .
7.000E-03  1.144E+00 .       .            .            . *     .
8.000E-03  1.111E+00 .       .            .            . *     .
9.000E-03  6.995E-01 .       .            .         *  .       .
1.000E-02 -5.697E-03 .       .            *            .       .
1.100E-02 -7.066E-01 .       .   *        .            .       .
1.200E-02 -1.108E+00 .     * .            .            .       .
1.300E-02 -1.135E+00 .     * .            .            .       .
1.400E-02 -9.672E-01 .       *            .            .       .
1.500E-02 -8.731E-01 .       . *          .            .       .
```

```
1.600E-02 -9.751E-01 .       *             .             .       .
1.700E-02 -1.144E+00 .     * .             .             .       .
1.800E-02 -1.111E+00 .     * .             .             .       .
1.900E-02 -6.995E-01 .       .   *         .             .       .
2.000E-02  5.697E-03 .       .             *             .       .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

The rise and fall times between positive and negative cycles are much steeper now, and the crests of the wave are closer to becoming flat like a squarewave. Watch what happens as we add the next odd harmonic frequency:

```
time      v(3)
time       v(3)          -1.000E+00    0.000E+00     1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .       .             *             .       .
1.000E-03  9.436E-01 .       .             .            *.       .
2.000E-03  1.095E+00 .       .             .             .*      .
3.000E-03  9.388E-01 .       .             .            *.       .
4.000E-03  9.807E-01 .       .             .             *       .
5.000E-03  1.069E+00 .       .             .             .*      .
6.000E-03  9.616E-01 .       .             .            *.       .
7.000E-03  9.479E-01 .       .             .            *.       .
8.000E-03  1.124E+00 .       .             .             . *     .
9.000E-03  8.957E-01 .       .             .            *.       .
1.000E-02 -1.925E-02 .       .             *             .       .
1.100E-02 -9.029E-01 .       .*            .             .       .
1.200E-02 -1.095E+00 .      *.             .             .       .
1.300E-02 -9.388E-01 .       .*            .             .       .
1.400E-02 -9.807E-01 .       *             .             .       .
1.500E-02 -1.069E+00 .      *.             .             .       .
1.600E-02 -9.616E-01 .       .*            .             .       .
1.700E-02 -9.479E-01 .       .*            .             .       .
1.800E-02 -1.124E+00 .     * .             .             .       .
1.900E-02 -8.957E-01 .       .*            .             .       .
2.000E-02  1.925E-02 .       .             *             .       .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

The most noticeable change here is how the crests of the wave have flattened even more. There are more several dips and crests at each end of the wave, but those dips and crests are smaller in amplitude than they were before. Watch again as we add the next odd harmonic waveform to the mix:

```
time       v(4)           -1.000E+00    0.000E+00    1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .       .             *             .       .
1.000E-03  1.055E+00 .       .             .            .*      .
```

```
2.000E-03  9.861E-01 .       .             .             *       .
3.000E-03  9.952E-01 .       .             .             *       .
4.000E-03  1.023E+00 .       .             .             *       .
5.000E-03  9.631E-01 .       .             .            *.       .
6.000E-03  1.044E+00 .       .             .             .*      .
7.000E-03  9.572E-01 .       .             .            *.       .
8.000E-03  1.031E+00 .       .             .             *       .
9.000E-03  9.962E-01 .       .             .             *       .
1.000E-02 -4.396E-02 .       .            *.             .       .
1.100E-02 -9.743E-01 .       *             .             .       .
1.200E-02 -9.861E-01 .       *             .             .       .
1.300E-02 -9.952E-01 .       *             .             .       .
1.400E-02 -1.023E+00 .       *             .             .       .
1.500E-02 -9.631E-01 .       .*            .             .       .
1.600E-02 -1.044E+00 .      *.             .             .       .
1.700E-02 -9.572E-01 .       .*            .             .       .
1.800E-02 -1.031E+00 .       *             .             .       .
1.900E-02 -9.962E-01 .       *             .             .       .
2.000E-02  4.396E-02 .       .             .*            .       .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Here we can see the wave becoming flatter at each peak. Finally, adding the 9th harmonic, the fifth sine wave voltage source in our circuit, we obtain this result:

```
time       v(5)          -1.000E+00    0.000E+00    1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .       .             *             .       .
1.000E-03  1.079E+00 .       .             .             .*      .
2.000E-03  9.845E-01 .       .             .             *       .
3.000E-03  1.017E+00 .       .             .             *       .
4.000E-03  9.835E-01 .       .             .             *       .
5.000E-03  1.017E+00 .       .             .             *       .
6.000E-03  9.814E-01 .       .             .             *       .
7.000E-03  1.023E+00 .       .             .             *       .
8.000E-03  9.691E-01 .       .             .             *       .
9.000E-03  1.048E+00 .       .             .             .*      .
1.000E-02 -8.103E-02 .       .            *.             .       .
1.100E-02 -9.557E-01 .       .*            .             .       .
1.200E-02 -9.845E-01 .       *             .             .       .
1.300E-02 -1.017E+00 .       *             .             .       .
1.400E-02 -9.835E-01 .       *             .             .       .
1.500E-02 -1.017E+00 .       *             .             .       .
1.600E-02 -9.814E-01 .       *             .             .       .
1.700E-02 -1.023E+00 .       *             .             .       .
1.800E-02 -9.691E-01 .       *             .             .       .
1.900E-02 -1.048E+00 .      *.             .             .       .
```

```
2.000E-02  8.103E-02 .       .             .*           .       .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

The end result of adding the first five odd harmonic waveforms together (all at the proper amplitudes, of course) is a close approximation of a square wave. The point in doing this is to illustrate how we can build a square wave up from multiple sine waves at different frequencies, to prove that a pure square wave is actually equivalent to a *series* of sine waves. When a square wave AC voltage is applied to a circuit with reactive components (capacitors and inductors), those components react as if they were being exposed to several sine wave voltages of different frequencies, which in fact they are.

The fact that repeating, non-sinusoidal waves are equivalent to a definite series of additive DC voltage, sine waves, and/or cosine waves is a consequence of how waves work: a fundamental property of all wave-related phenomena, electrical or otherwise. The mathematical process of reducing a non-sinusoidal wave into these constituent frequencies is called *Fourier analysis*, the details of which are well beyond the scope of this text. However, computer algorithms have been created to perform this analysis at high speeds on real waveforms, and its application in AC power quality and signal analysis is widespread.

SPICE has the ability to sample a waveform and reduce it into its constituent sine wave harmonics by way of a *Fourier Transform* algorithm, outputting the frequency analysis as a table of numbers. Let's try this on a square wave, which we already know is composed of odd-harmonic sine waves:

```
squarewave analysis netlist
v1 1 0 pulse (-1 1 0 .1m .1m 10m 20m)
r1 1 0 10k
.tran 1m 40m
.plot tran v(1,0)
.four 50 v(1,0)
.end
```

The *pulse* option in the netlist line describing voltage source `v1` instructs SPICE to simulate a square-shaped "pulse" waveform, in this case one that is symmetrical (equal time for each half-cycle) and has a peak amplitude of 1 volt. First we'll plot the square wave to be analyzed:

```
time       v(1)     -1         -0.5          0          0.5     1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00 -1.000E+00 *         .           .           .         .
1.000E-03  1.000E+00 .         .           .           .         *
2.000E-03  1.000E+00 .         .           .           .         *
3.000E-03  1.000E+00 .         .           .           .         *
4.000E-03  1.000E+00 .         .           .           .         *
5.000E-03  1.000E+00 .         .           .           .         *
6.000E-03  1.000E+00 .         .           .           .         *
7.000E-03  1.000E+00 .         .           .           .         *
8.000E-03  1.000E+00 .         .           .           .         *
9.000E-03  1.000E+00 .         .           .           .         *
1.000E-02  1.000E+00 .         .           .           .         *
```

```
1.100E-02 -1.000E+00 *         .           .           .         .
1.200E-02 -1.000E+00 *         .           .           .         .
1.300E-02 -1.000E+00 *         .           .           .         .
1.400E-02 -1.000E+00 *         .           .           .         .
1.500E-02 -1.000E+00 *         .           .           .         .
1.600E-02 -1.000E+00 *         .           .           .         .
1.700E-02 -1.000E+00 *         .           .           .         .
1.800E-02 -1.000E+00 *         .           .           .         .
1.900E-02 -1.000E+00 *         .           .           .         .
2.000E-02 -1.000E+00 *         .           .           .         .
2.100E-02  1.000E+00 .         .           .           .         *
2.200E-02  1.000E+00 .         .           .           .         *
2.300E-02  1.000E+00 .         .           .           .         *
2.400E-02  1.000E+00 .         .           .           .         *
2.500E-02  1.000E+00 .         .           .           .         *
2.600E-02  1.000E+00 .         .           .           .         *
2.700E-02  1.000E+00 .         .           .           .         *
2.800E-02  1.000E+00 .         .           .           .         *
2.900E-02  1.000E+00 .         .           .           .         *
3.000E-02  1.000E+00 .         .           .           .         *
3.100E-02 -1.000E+00 *         .           .           .         .
3.200E-02 -1.000E+00 *         .           .           .         .
3.300E-02 -1.000E+00 *         .           .           .         .
3.400E-02 -1.000E+00 *         .           .           .         .
3.500E-02 -1.000E+00 *         .           .           .         .
3.600E-02 -1.000E+00 *         .           .           .         .
3.700E-02 -1.000E+00 *         .           .           .         .
3.800E-02 -1.000E+00 *         .           .           .         .
3.900E-02 -1.000E+00 *         .           .           .         .
4.000E-02 -1.000E+00 *         .           .           .         .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Next, we'll print the Fourier analysis generated by SPICE for this square wave:

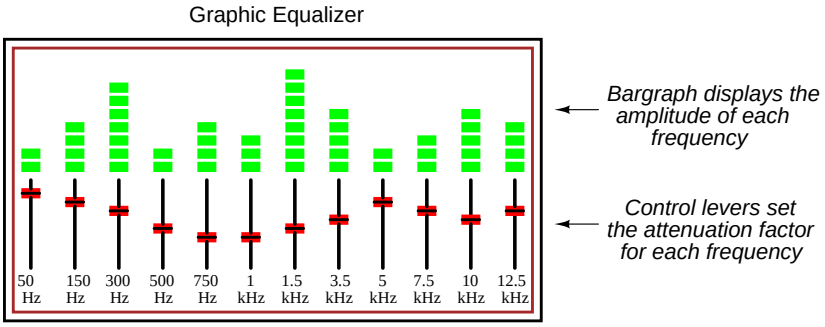
```
fourier components of transient response v(1)
dc component =  -2.439E-02
harmonic  frequency   fourier     normalized   phase   normalized
no          (hz)     component    component    (deg)   phase (deg)
1         5.000E+01   1.274E+00     1.000000    -2.195       0.000
2         1.000E+02   4.892E-02     0.038415   -94.390     -92.195
3         1.500E+02   4.253E-01     0.333987    -6.585      -4.390
4         2.000E+02   4.936E-02     0.038757   -98.780     -96.585
5         2.500E+02   2.562E-01     0.201179   -10.976      -8.780
6         3.000E+02   5.010E-02     0.039337  -103.171    -100.976
7         3.500E+02   1.841E-01     0.144549   -15.366     -13.171
8         4.000E+02   5.116E-02     0.040175  -107.561    -105.366
```

```
9         4.500E+02   1.443E-01     0.113316   -19.756     -17.561
total harmonic distortion =     43.805747  percent
```

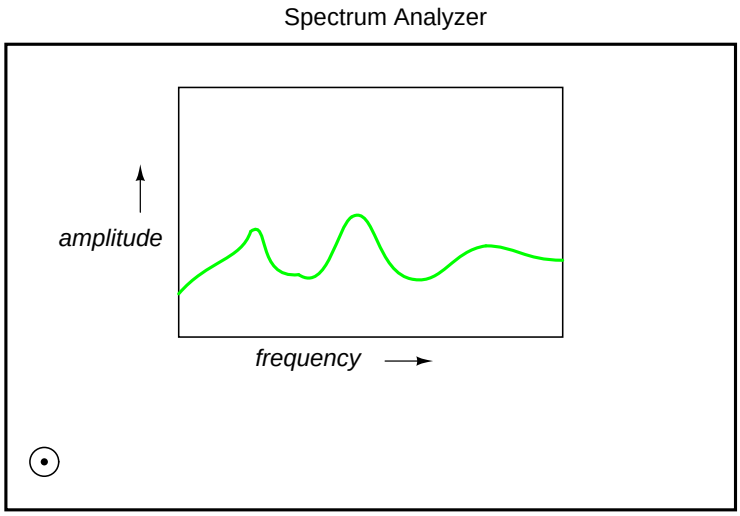
Here, SPICE has broken the waveform down into a spectrum of sinusoidal frequencies up to the ninth harmonic, plus a small DC voltage labelled `DC component`. I had to inform SPICE of the fundamental frequency (for a square wave with a 20 millisecond period, this frequency is 50 Hz), so it knew how to classify the harmonics. Note how small the figures are for all the even harmonics (2nd, 4th, 6th, 8th), and how the amplitudes of the odd harmonics diminish (1st is largest, 9th is smallest).

This same technique of "Fourier Transformation" is often used in computerized power instrumentation, sampling the AC waveform(s) and determining the harmonic content thereof. A common computer algorithm (sequence of program steps to perform a task) for this is the *Fast Fourier Transform* or *FFT* function. You need not be concerned with exactly how these computer routines work, but be aware of their existence and application.

This same mathematical technique used in SPICE to analyze the harmonic content of waves can be applied to the technical analysis of music: breaking up any particular sound into its constituent sine-wave frequencies. In fact, you may have already seen a device designed to do just that without realizing what it was! A *graphic equalizer* is a piece of high-fidelity stereo equipment that controls (and sometimes displays) the nature of music's harmonic content. Equipped with several knobs or slide levers, the equalizer is able to selectively attenuate (reduce) the amplitude of certain frequencies present in music, to "customize" the sound for the listener's benefit. Typically, there will be a "bar graph" display next to each control lever, displaying the amplitude of each particular frequency.



A device built strictly to display – not control – the amplitudes of each frequency range for a mixed-frequency signal is typically called a *spectrum analyzer*. The design of spectrum analyzers may be as simple as a set of "filter" circuits (see the next chapter for details) designed to separate the different frequencies from each other, or as complex as a special-purpose digital computer running an FFT algorithm to mathematically split the signal into its harmonic components. Spectrum analyzers are often designed to analyze extremely high-frequency signals, such as those produced by radio transmitters and computer network hardware. In that form, they often have an appearance like that of an oscilloscope:



Like an oscilloscope, the spectrum analyzer uses a CRT (or a computer display mimicking a CRT) to display a plot of the signal. Unlike an oscilloscope, this plot is amplitude over *frequency* rather than amplitude over *time*. In essence, a frequency analyzer gives the operator a Bode plot of the signal: something an engineer might call a *frequency-domain* rather than a *time-domain* analysis.

The term "domain" is mathematical: a sophisticated word to describe the horizontal axis of a graph. Thus, an oscilloscope's plot of amplitude (vertical) over time (horizontal) is a "time-domain" analysis, whereas a spectrum analyzer's plot of amplitude (vertical) over frequency (horizontal) is a "frequency-domain" analysis. When we use SPICE to plot signal amplitude (either voltage or current amplitude) over a range of frequencies, we are performing *frequency-domain* analysis.

Please take note of how the Fourier analysis from the last SPICE simulation isn't "perfect." Ideally, the amplitudes of all the even harmonics should be absolutely zero, and so should the DC component. Again, this is not so much a quirk of SPICE as it is a property of waveforms in general. A waveform of infinite duration (infinite number of cycles) can be analyzed with absolute precision, but the less cycles available to the computer for analysis, the less precise the analysis. It is only when we have an equation describing a waveform in its entirety that Fourier analysis can reduce it to a definite series of sinusoidal waveforms. The fewer times that a wave cycles, the less certain its frequency is. Taking this concept to its logical extreme, a short pulse – a waveform that doesn't even complete a cycle – actually *has no frequency*, but rather acts as an infinite range of frequencies. This principle is common to *all* wave-based phenomena, not just AC voltages and currents.

Suffice it to say that the number of cycles and the certainty of a waveform's frequency component(s) are directly related. We could improve the precision of our analysis here by letting the wave oscillate on and on for many cycles, and the result would be a spectrum analysis more consistent with the ideal. In the following analysis, I've omitted the waveform plot for brevity's sake – it's just a really long square wave:

```
squarewave
v1 1 0 pulse (-1 1 0 .1m .1m 10m 20m)
r1 1 0 10k
.option limpts=1001
```

```
.tran 1m 1
.plot tran v(1,0)
.four 50 v(1,0)
.end

fourier components of transient response v(1)
dc component =   9.999E-03
harmonic  frequency    fourier    normalized   phase   normalized
no           (hz)     component   component    (deg)   phase (deg)
1         5.000E+01   1.273E+00     1.000000    -1.800       0.000
2         1.000E+02   1.999E-02     0.015704    86.382      88.182
3         1.500E+02   4.238E-01     0.332897    -5.400      -3.600
4         2.000E+02   1.997E-02     0.015688    82.764      84.564
5         2.500E+02   2.536E-01     0.199215    -9.000      -7.200
6         3.000E+02   1.994E-02     0.015663    79.146      80.946
7         3.500E+02   1.804E-01     0.141737   -12.600     -10.800
8         4.000E+02   1.989E-02     0.015627    75.529      77.329
9         4.500E+02   1.396E-01     0.109662   -16.199     -14.399
```

Notice how this analysis shows less of a DC component voltage and lower amplitudes for each of the even harmonic frequency sine waves, all because we let the computer sample more cycles of the wave. Again, the imprecision of the first analysis is not so much a flaw in SPICE as it is a fundamental property of waves and of signal analysis.

- **REVIEW:**
- Square waves are equivalent to a sine wave at the same (fundamental) frequency added to an infinite series of odd-multiple sine-wave harmonics at decreasing amplitudes.
- Computer algorithms exist which are able to sample waveshapes and determine their constituent sinusoidal components. The *Fourier Transform* algorithm (particularly the *Fast Fourier Transform*, or *FFT*) is commonly used in computer circuit simulation programs such as SPICE and in electronic metering equipment for determining power quality.

7.3 Other waveshapes

As strange as it may seem, *any* repeating, non-sinusoidal waveform is actually equivalent to a series of sinusoidal waveforms of different amplitudes and frequencies added together. Square waves are a very common and well-understood case, but not the only one.

Electronic power control devices such as transistors and silicon-controlled rectifiers (*SCRs*) often produce voltage and current waveforms that are essentially chopped-up versions of the otherwise "clean" (pure) sine-wave AC from the power supply. These devices have the ability to suddenly *change* their resistance with the application of a control signal voltage or current, thus "turning on" or "turning off" almost instantaneously, producing current waveforms bearing little resemblance to the source voltage waveform powering the circuit. These current waveforms then produce changes in

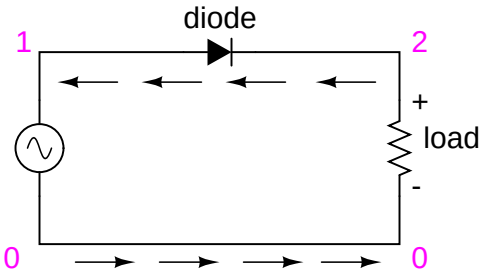
the voltage waveform to other circuit components, due to voltage drops created by the non-sinusoidal current through circuit impedances.

Circuit components that distort the normal sine-wave shape of AC voltage or current are called *nonlinear*. Nonlinear components such as SCRs find popular use in power electronics due to their ability to regulate large amounts of electrical power without dissipating much heat. While this is an advantage from the perspective of energy efficiency, the waveshape distortions they introduce can cause problems.

These non-sinusoidal waveforms, regardless of their actual shape, are equivalent to a series of sinusoidal waveforms of higher (harmonic) frequencies. If not taken into consideration by the circuit designer, these harmonic waveforms created by electronic switching components may cause erratic circuit behavior. It is becoming increasingly common in the electric power industry to observe overheating of transformers and motors due to distortions in the sine-wave shape of the AC power line voltage stemming from "switching" loads such as computers and high-efficiency lights. This is no theoretical exercise: it is very real and potentially very troublesome.

In this section, I will investigate a few of the more common waveshapes and show their harmonic components by way of Fourier analysis using SPICE.

One very common way harmonics are generated in an AC power system is when AC is converted, or "rectified" into DC. This is generally done with components called *diodes*, which only allow passage current in one direction. The simplest type of AC/DC rectification is *half-wave*, where a single diode blocks half of the AC current (over time) from passing through the load. Oddly enough, the conventional diode schematic symbol is drawn such that electrons flow *against* the direction of the symbol's arrowhead:



The diode only allows electron flow in a counter-clockwise direction.

```
halfwave rectifier
v1 1 0 sin(0 15 60 0 0)
rload 2 0 10k
d1 1 2 mod1
.model mod1 d
.tran .5m 17m
.plot tran v(1,0) v(2,0)
.four 60 v(1,0) v(2,0)
.end
```

```
legend:
*: v(1)
+: v(2)
time      v(1)
(*)----------      -20     -10           0          10          20
(+)----------       -5       0           5          10          15
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .       +           *           .           .
5.000E-04  2.806E+00 .       .   +       .   *       .           .
1.000E-03  5.483E+00 .       .           +       *   .           .
1.500E-03  7.929E+00 .       .           .     +    *.           .
2.000E-03  1.013E+01 .       .           .           +*          .
2.500E-03  1.198E+01 .       .           .           .  *+       .
3.000E-03  1.338E+01 .       .           .           .    *  +   .
3.500E-03  1.435E+01 .       .           .           .     *   + .
4.000E-03  1.476E+01 .       .           .           .      *   +.
4.500E-03  1.470E+01 .       .           .           .      *   +.
5.000E-03  1.406E+01 .       .           .           .     *   + .
5.500E-03  1.299E+01 .       .           .           .   *  +    .
6.000E-03  1.139E+01 .       .           .           . x         .
6.500E-03  9.455E+00 .       .           .         + *.          .
7.000E-03  7.113E+00 .       .           .   +     *  .          .
7.500E-03  4.591E+00 .       .         +  .     *     .          .
8.000E-03  1.841E+00 .       . +         .  *        .          .
8.500E-03 -9.177E-01 .       +           *.          .          .
9.000E-03 -3.689E+00 .       +      *    .           .          .
9.500E-03 -6.298E+00 .       +  *        .           .          .
1.000E-02 -8.701E+00 .       +*          .           .          .
1.050E-02 -1.079E+01 .      *+           .           .          .
1.100E-02 -1.249E+01 .    *  +           .           .          .
1.150E-02 -1.377E+01 .  *    +           .           .          .
1.200E-02 -1.453E+01 . *     +           .           .          .
1.250E-02 -1.482E+01 .*      +           .           .          .
1.300E-02 -1.452E+01 . *     +           .           .          .
1.350E-02 -1.378E+01 .  *    +           .           .          .
1.400E-02 -1.248E+01 .    *  +           .           .          .
1.450E-02 -1.081E+01 .      *+           .           .          .
1.500E-02 -8.681E+00 .       +*          .           .          .
1.550E-02 -6.321E+00 .       +  *        .           .          .
1.600E-02 -3.666E+00 .       +      *    .           .          .
1.650E-02 -9.432E-01 .       .   +      *.           .          .
1.700E-02  1.865E+00 .       .      +    .  *        .          .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

First, we'll see how SPICE analyzes the source waveform, a pure sine wave voltage:

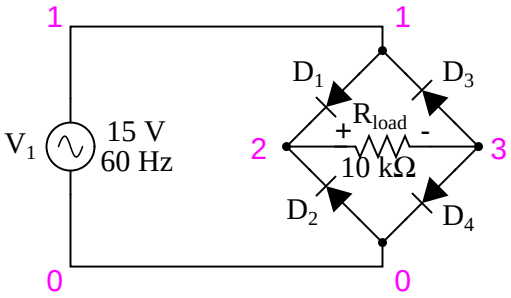
```
fourier components of transient response v(1)
dc component =   8.016E-04
harmonic   frequency  fourier    normalized   phase   normalized
no          (hz)      component   component    (deg)   phase (deg)
1        6.000E+01   1.482E+01     1.000000    -0.005       0.000
2        1.200E+02   2.492E-03     0.000168  -104.347    -104.342
3        1.800E+02   6.465E-04     0.000044   -86.663     -86.658
4        2.400E+02   1.132E-03     0.000076   -61.324     -61.319
5        3.000E+02   1.185E-03     0.000080   -70.091     -70.086
6        3.600E+02   1.092E-03     0.000074   -63.607     -63.602
7        4.200E+02   1.220E-03     0.000082   -56.288     -56.283
8        4.800E+02   1.354E-03     0.000091   -54.669     -54.664
9        5.400E+02   1.467E-03     0.000099   -52.660     -52.655
```

Notice the extremely small harmonic and DC components of this sinusoidal waveform. Ideally, there would be nothing but the fundamental frequency showing (being a perfect sine wave), but our Fourier analysis figures aren't perfect because SPICE doesn't have the luxury of sampling a waveform of infinite duration. Next, we'll compare this with the Fourier analysis of the half-wave "rectified" voltage across the load resistor:

```
fourier components of transient response v(2)
dc component =   4.456E+00
harmonic  frequency   fourier    normalized   phase    normalized
no          (hz)      component   component    (deg)    phase (deg)
1        6.000E+01   7.000E+00     1.000000    -0.195       0.000
2        1.200E+02   3.016E+00     0.430849   -89.765     -89.570
3        1.800E+02   1.206E-01     0.017223  -168.005    -167.810
4        2.400E+02   5.149E-01     0.073556   -87.295     -87.100
5        3.000E+02   6.382E-02     0.009117  -152.790    -152.595
6        3.600E+02   1.727E-01     0.024676   -79.362     -79.167
7        4.200E+02   4.492E-02     0.006417  -132.420    -132.224
8        4.800E+02   7.493E-02     0.010703   -61.479     -61.284
9        5.400E+02   4.051E-02     0.005787  -115.085    -114.889
```

Notice the relatively large even-multiple harmonics in this analysis. By cutting out half of our AC wave, we've introduced the equivalent of several higher-frequency sinusoidal (actually, cosine) waveforms into our circuit from the original, pure sine-wave. Also take note of the large DC component: 4.456 volts. Because our AC voltage waveform has been "rectified" (only allowed to push in one direction across the load rather than back-and-forth), it behaves a lot more like DC.

Another method of AC/DC conversion is called *full-wave*, which as you may have guessed utilizes the full cycle of AC power from the source, reversing the polarity of half the AC cycle to get electrons to flow through the load the same direction all the time. I won't bore you with details of exactly how this is done, but we can examine the waveform and its harmonic analysis through SPICE:




```
fullwave bridge rectifier
v1 1 0 sin(0 15 60 0 0)
rload 2 3 10k
d1 1 2 mod1
d2 0 2 mod1
d3 3 1 mod1
d4 3 0 mod1
.model mod1 d
.tran .5m 17m
.plot tran v(1,0) v(2,3)
.four 60 v(2,3)
.end

legend:
*: v(1)
+: v(2,3)
time       v(1)
(*)----------      -20          -10     0.000E+00     1.000E+01
(+)----------  0.000E+00     5.000E+00     1.000E+01     1.500E+01
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 +          .            *            .      .
5.000E-04  2.806E+00 .   +      .            .   *        .      .
1.000E-03  5.483E+00 .         +.            .      *     .      .
1.500E-03  7.929E+00 .          .    +       .         *  .      .
2.000E-03  1.013E+01 .          .         +  .            *      .
2.500E-03  1.198E+01 .          .            . +          . *    .
3.000E-03  1.338E+01 .          .            .     +      .   *  .
3.500E-03  1.435E+01 .          .            .       +    .     *.
4.000E-03  1.476E+01 .          .            .        +   .      *
4.500E-03  1.470E+01 .          .            .        +   .     *.
5.000E-03  1.406E+01 .          .            .       +    .    * .
5.500E-03  1.299E+01 .          .            .    +       .  *   .
6.000E-03  1.139E+01 .          .           +            .*     .
6.500E-03  9.455E+00 .          .       +    .           *.      .
7.000E-03  7.113E+00 .          . +          .        *   .      .
```

```
7.500E-03  4.591E+00 .      +   .            .     *      .      .
8.000E-03  1.841E+00 .   +      .            .  *         .      .
8.500E-03 -9.177E-01 . +        .           *.            .      .
9.000E-03 -3.689E+00 .    +     .       *    .            .      .
9.500E-03 -6.298E+00 .          +    *       .            .      .
1.000E-02 -8.701E+00 .          . *    +     .            .      .
1.050E-02 -1.079E+01 .         *.           +.            .      .
1.100E-02 -1.249E+01 .       *  .            .  +         .      .
1.150E-02 -1.377E+01 .     *    .            .      +     .      .
1.200E-02 -1.453E+01 .    *     .            .        +   .      .
1.250E-02 -1.482E+01 .   *      .            .         +  .      .
1.300E-02 -1.452E+01 .    *     .            .        +   .      .
1.350E-02 -1.378E+01 .     *    .            .      +     .      .
1.400E-02 -1.248E+01 .       *  .            .  +         .      .
1.450E-02 -1.081E+01 .         *.           +.            .      .
1.500E-02 -8.681E+00 .          . *    +     .            .      .
1.550E-02 -6.321E+00 .          +    *       .            .      .
1.600E-02 -3.666E+00 .     +    .       *    .            .      .
1.650E-02 -9.432E-01 .   +      .           *.            .      .
1.700E-02  1.865E+00 .    +     .            .  *         .      .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

fourier components of transient response v(2,3)
dc component =   8.273E+00
harmonic   frequency  fourier    normalized   phase   normalized
no         (hz)       component   component    (deg)   phase (deg)
1        6.000E+01   7.000E-02     1.000000   -93.519       0.000
2        1.200E+02   5.997E+00    85.669415   -90.230       3.289
3        1.800E+02   7.241E-02     1.034465   -93.787      -0.267
4        2.400E+02   1.013E+00    14.465161   -92.492       1.027
5        3.000E+02   7.364E-02     1.052023   -95.026      -1.507
6        3.600E+02   3.337E-01     4.767350  -100.271      -6.752
7        4.200E+02   7.496E-02     1.070827   -94.023      -0.504
8        4.800E+02   1.404E-01     2.006043  -118.839     -25.319
9        5.400E+02   7.457E-02     1.065240   -90.907       2.612
```

What a difference! According to SPICE's Fourier transform, we have a 2nd harmonic component to this waveform that's over 85 times the amplitude of the original AC source frequency! The DC component of this wave shows up as being 8.273 volts (almost twice what is was for the half-wave rectifier circuit) while the second harmonic is almost 6 volts in amplitude. Notice all the other harmonics further on down the table. The odd harmonics are actually stronger at some of the higher frequencies than they are at the lower frequencies, which is interesting.

As you can see, what may begin as a neat, simple AC sine-wave may end up as a complex mess of harmonics after passing through just a few electronic components. While the complex mathematics behind all this Fourier transformation is not necessary for the beginning student of electric circuits to understand, it is of the utmost importance to realize the principles at work and to grasp the practical

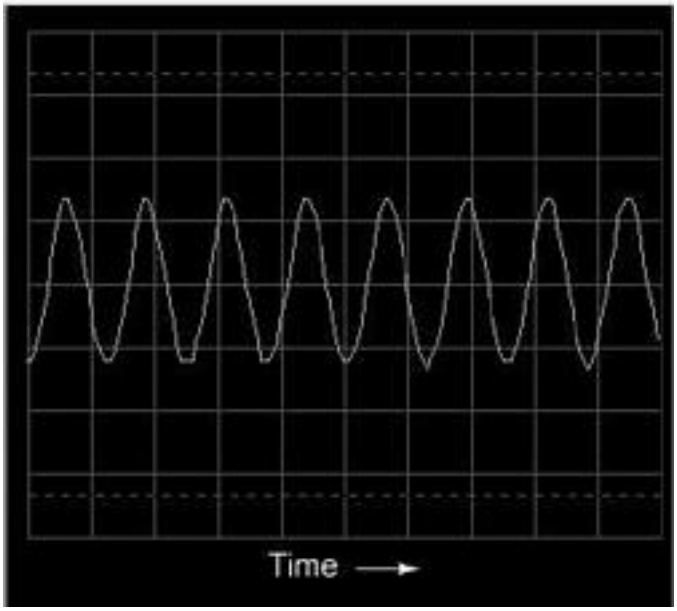
effects that harmonic signals may have on circuits. The practical effects of harmonic frequencies in circuits will be explored in the last section of this chapter, but before we do that we'll take a closer look at waveforms and their respective harmonics.

- **REVIEW:**
- *Any* waveform at all, so long as it is repetitive, can be reduced to a series of sinusoidal waveforms added together. Different waveshapes consist of different blends of sine-wave harmonics.
- Rectification of AC to DC is a very common source of harmonics within industrial power systems.

7.4 More on spectrum analysis

Computerized Fourier analysis, particularly in the form of the *FFT* algorithm, is a powerful tool for furthering our understanding of waveforms and their related spectral components. This same mathematical routine programmed into the SPICE simulator as the `.fourier` option is also programmed into a variety of electronic test instruments to perform real-time Fourier analysis on measured signals. This section is devoted to the use of such tools and the analysis of several different waveforms.

First we have a simple sine wave at a frequency of 523.25 Hz. This particular frequency value is a "C" pitch on a piano keyboard, one octave above "middle C". Actually, the signal measured for this demonstration was created by an electronic keyboard set to produce the tone of a panflute, the closest instrument "voice" I could find resembling a perfect sine wave. The plot below was taken from an oscilloscope display, showing signal amplitude (voltage) over time:

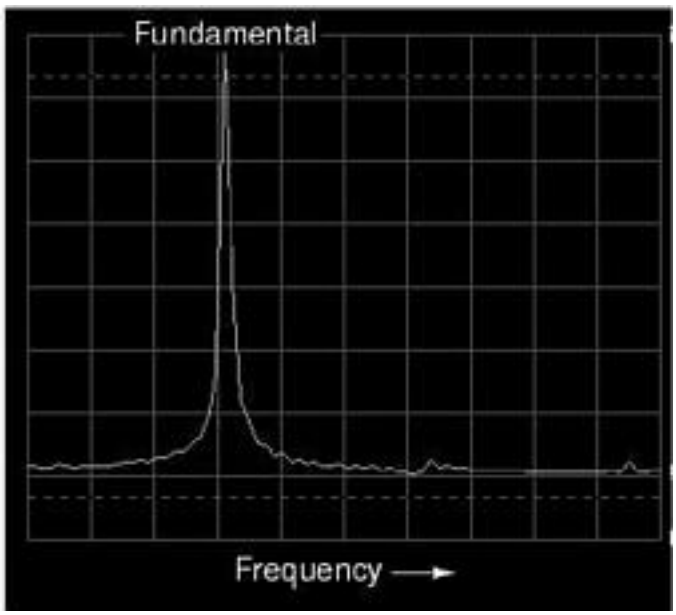


Viewed with an oscilloscope, a sine wave looks like a wavy curve traced horizontally on the screen. The horizontal axis of this oscilloscope display is marked with the word "Time" and an

arrow pointing in the direction of time's progression. The curve itself, of course, represents the cyclic increase and decrease of voltage over time.

Close observation reveals imperfections in the sine-wave shape. This, unfortunately, is a result of the specific equipment used to analyze the waveform. Characteristics like these due to quirks of the test equipment are technically known as *artifacts*: phenomena existing solely because of a peculiarity in the equipment used to perform the experiment.

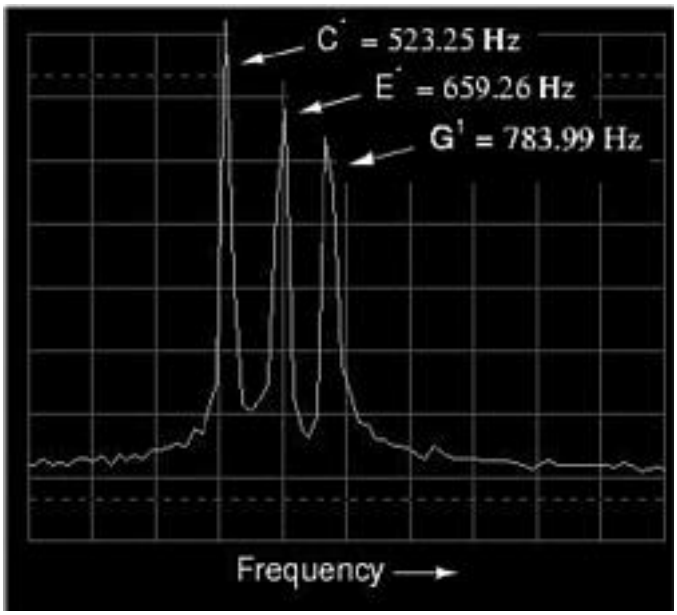
If we view this same AC voltage on a spectrum analyzer, the result is quite different:



As you can see, the horizontal axis of the display is marked with the word "Frequency," denoting the domain of this measurement. The single peak on the curve represents the predominance of a single frequency within the range of frequencies covered by the width of the display. If the scale of this analyzer instrument were marked with numbers, you would see that this peak occurs at 523.25 Hz. The height of the peak represents the signal amplitude (voltage).

If we mix three different sine-wave tones together on the electronic keyboard (C-E-G, a C-major chord) and measure the result, both the oscilloscope display and the spectrum analyzer display reflect this increased complexity:

The oscilloscope display (time-domain) shows a waveform with many more peaks and valleys than before, a direct result of the mixing of these three frequencies. As you will notice, some of these peaks are higher than the peaks of the original single-pitch waveform, while others are lower. This is a result of the three different waveforms alternately reinforcing and canceling each other as their respective phase shifts change in time.

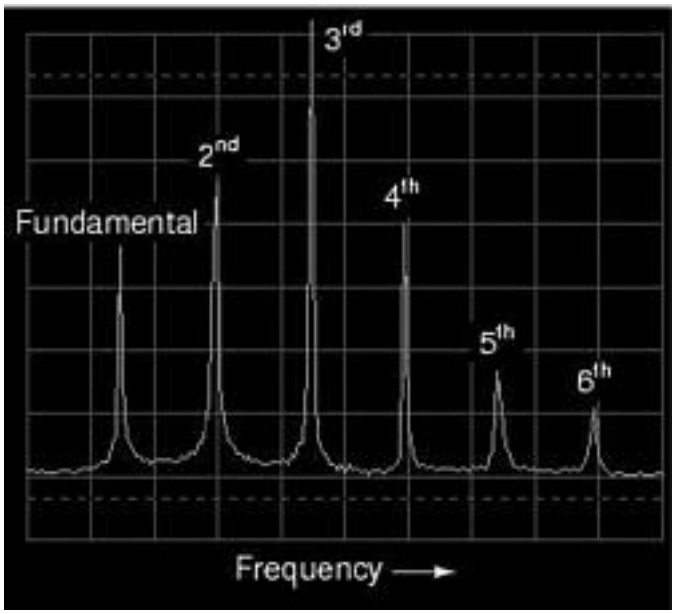


The spectrum display (frequency-domain) is much easier to interpret: each pitch is represented

by its own peak on the curve. The difference in height between these three peaks is another artifact of the test equipment: a consequence of limitations within the equipment used to generate and analyze these waveforms, and not a necessary characteristic of the musical chord itself.

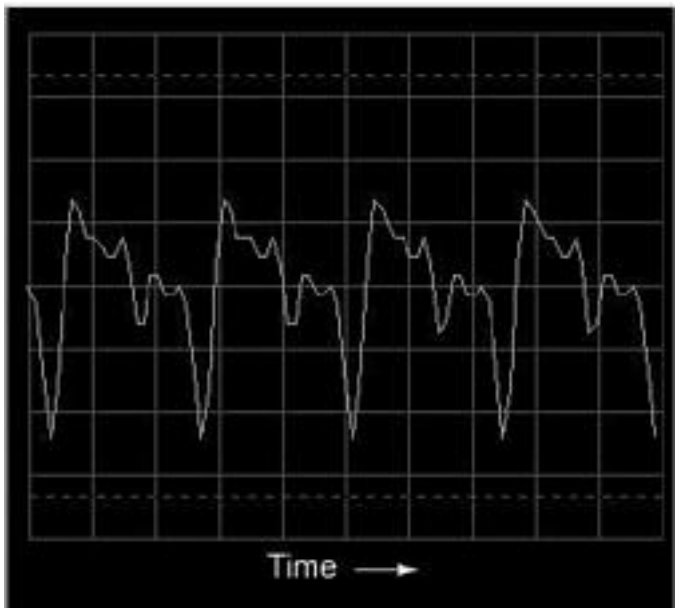
As was stated before, the device used to generate these waveforms is an electronic keyboard: a musical instrument designed to mimic the tones of many different instruments. The panflute "voice" was chosen for the first demonstrations because it most closely resembled a pure sine wave (a single frequency on the spectrum analyzer display). Other musical instrument "voices" are not as simple as this one, though. In fact, the unique tone produced by *any* instrument is a function of its waveshape (or spectrum of frequencies). For example, let's view the signal for a trumpet tone:

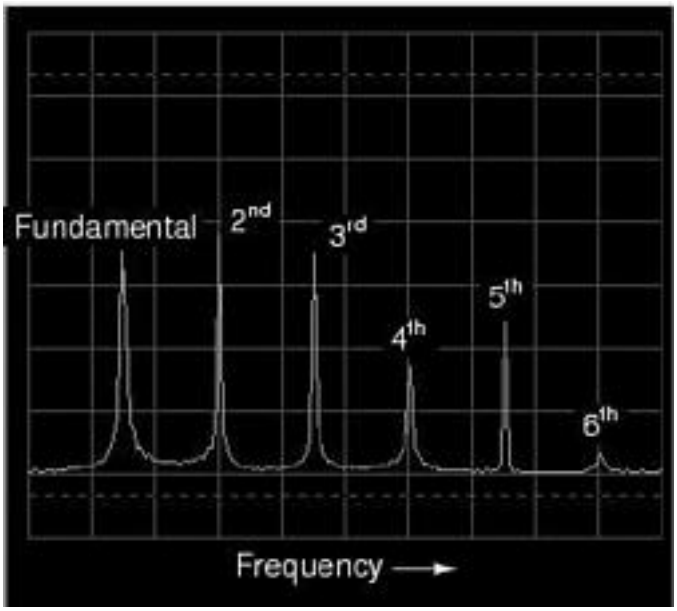
The fundamental frequency of this tone is the same as in the first panflute example: 523.25 Hz, one octave above "middle C." The waveform itself is far from a pure and simple sine-wave form. Knowing that any repeating, non-sinusoidal waveform is equivalent to a series of sinusoidal waveforms at different amplitudes and frequencies, we should expect to see multiple peaks on the spectrum analyzer display:



Indeed we do! The fundamental frequency component of 523.25 Hz is represented by the left-most peak, with each successive harmonic represented as its own peak along the width of the analyzer screen. The second harmonic is twice the frequency of the fundamental (1046.5 Hz), the third harmonic three times the fundamental (1569.75 Hz), and so on. This display only shows the first six harmonics, but there are many more comprising this complex tone.

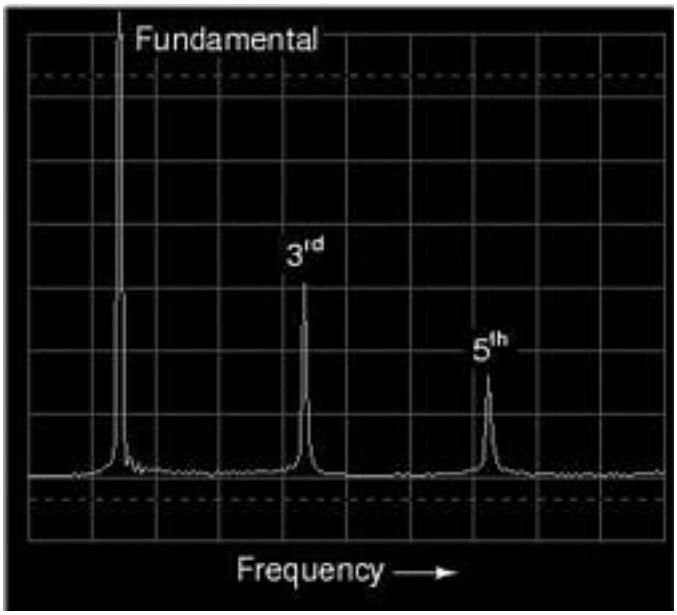
Trying a different instrument voice (the accordion) on the keyboard, we obtain a similarly complex oscilloscope (time-domain) plot and spectrum analyzer (frequency-domain) display:



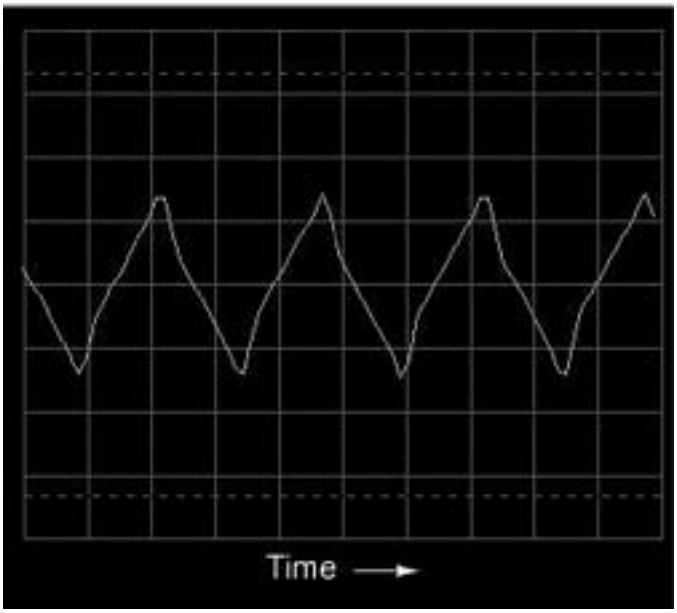


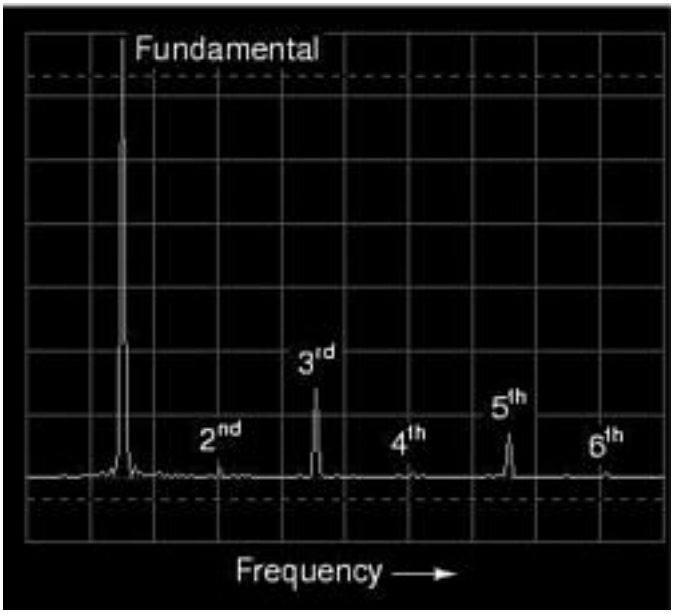
Note the differences in relative harmonic amplitudes (peak heights) on the spectrum displays for trumpet and accordion. Both instrument tones contain harmonics all the way from 1st (fundamental) to 6th (and beyond!), but the proportions aren't the same. Each instrument has a unique harmonic "signature" to its tone. Bear in mind that all this complexity is in reference to *a single note* played with these two instrument "voices." Multiple notes played on an accordion, for example, would create a much more complex mixture of frequencies than what is seen here.

The analytical power of the oscilloscope and spectrum analyzer permit us to derive general rules about waveforms and their harmonic spectra from real waveform examples. We already know that any deviation from a pure sine-wave results in the equivalent of a mixture of multiple sine-wave waveforms at different amplitudes and frequencies. However, close observation allows us to be more specific than this. Note, for example, the time- and frequency-domain plots for a waveform approximating a square wave:



According to the spectrum analysis, this waveform contains *no* even harmonics, only odd. Although this display doesn't show frequencies past the sixth harmonic, the pattern of odd-only harmonics in descending amplitude continues indefinitely. This should come as no surprise, as we've already seen with SPICE that a square wave is comprised of an infinitude of odd harmonics. The trumpet and accordion tones, however, contained *both* even and odd harmonics. This difference in harmonic content is noteworthy. Let's continue our investigation with an analysis of a triangle wave:

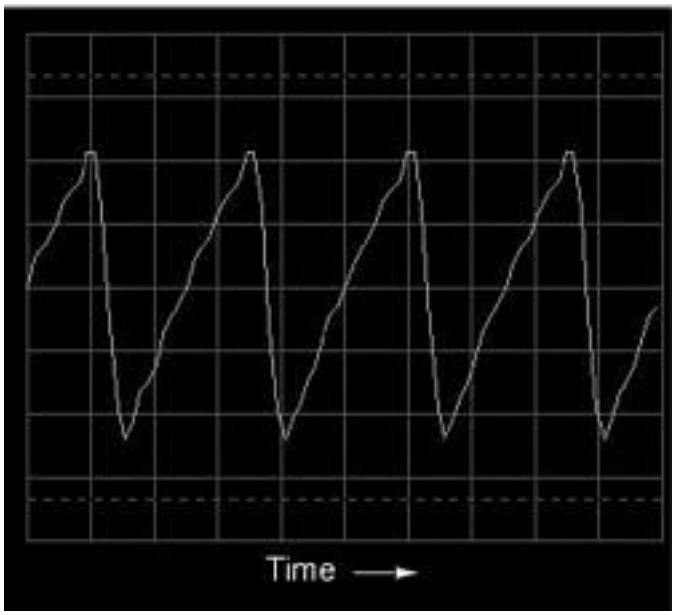




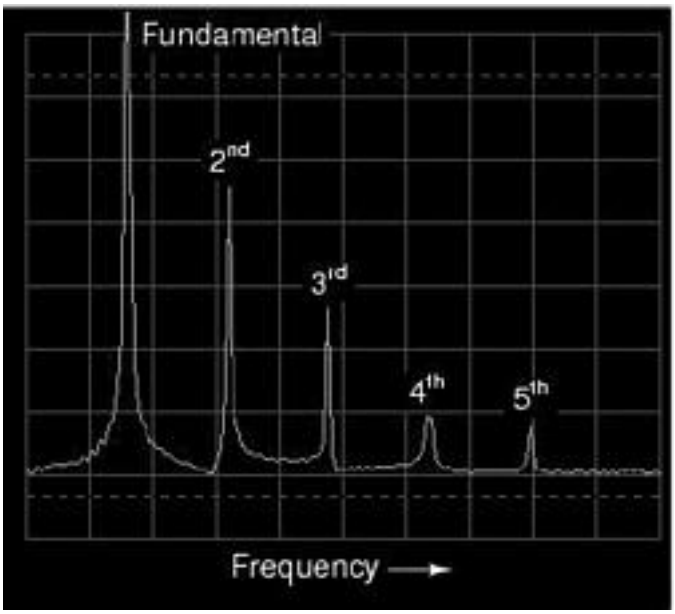
In this waveform there are practically no even harmonics: the only significant frequency peaks on the spectrum analyzer display belong to odd-numbered multiples of the fundamental frequency. Tiny peaks can be seen for the second, fourth, and sixth harmonics, but this is due to imperfections in this particular triangle waveshape (once again, artifacts of the test equipment used in this analysis). A perfect triangle waveshape produces no even harmonics, just like a perfect square wave. It should be obvious from inspection that the harmonic spectrum of the triangle wave is not identical to the

spectrum of the square wave: the respective harmonic peaks are of different heights. However, the two different waveforms are common in their lack of even harmonics.

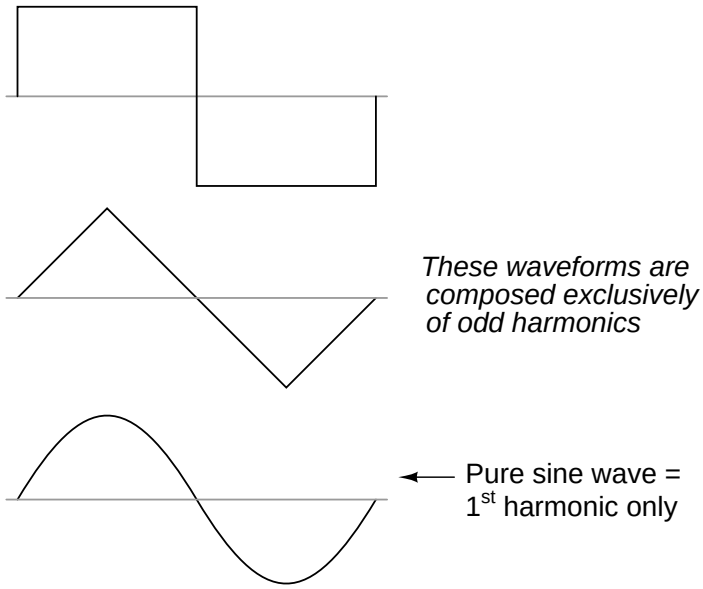
Let's examine another waveform, this one very similar to the triangle wave, except that its rise-time is not the same as its fall-time. Known as a *sawtooth wave*, its oscilloscope plot reveals it to be aptly named:



When the spectrum analysis of this waveform is plotted, we see a result that is quite different from that of the regular triangle wave, for this analysis shows the strong presence of even-numbered harmonics (second and fourth):

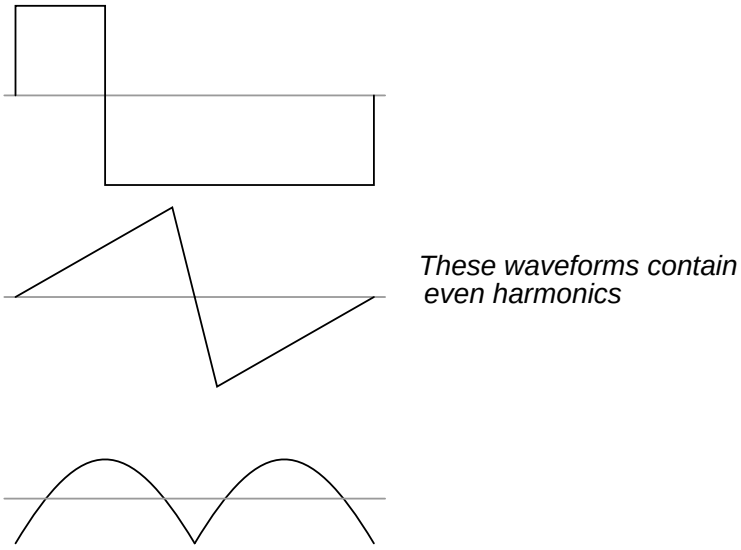


The distinction between a waveform having even harmonics versus no even harmonics resides in the difference between a triangle waveshape and a sawtooth waveshape. That difference is *symmetry* above and below the horizontal centerline of the wave. A waveform that is symmetrical above and below its centerline (the shape on both sides mirror each other precisely) will contain *no* even-numbered harmonics.

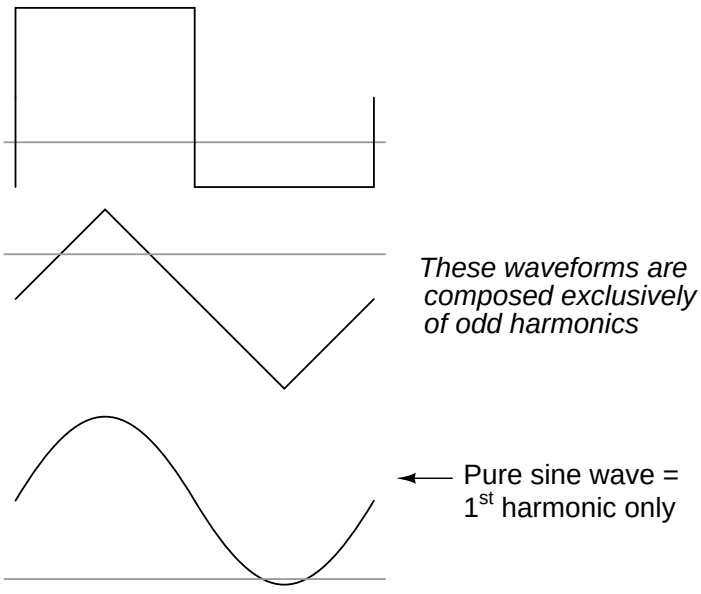


Square waves, triangle waves, and pure sine waves all exhibit this symmetry, and all are devoid of even harmonics. Waveforms like the trumpet tone, the accordion tone, and the sawtooth wave

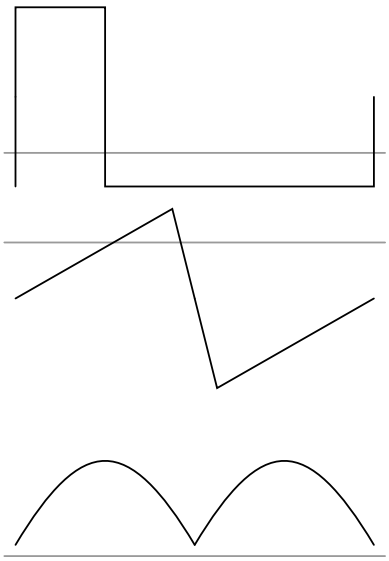
are unsymmetrical around their centerlines and therefore *do* contain even harmonics.



This principle of centerline symmetry should not be confused with symmetry around the *zero* line. In the examples shown, the horizontal centerline of the waveform happens to be zero volts on the time-domain graph, but this has nothing to do with harmonic content. This rule of harmonic content (even harmonics only with unsymmetrical waveforms) applies whether or not the waveform is shifted above or below zero volts with a "DC component." For further clarification, I will show the same sets of waveforms, shifted with DC voltage, and note that their harmonic contents are unchanged.



Again, the amount of DC voltage present in a waveform has nothing to do with that waveform's harmonic frequency content.



Why is this harmonic rule-of-thumb an important rule to know? It can help us comprehend the relationship between harmonics in AC circuits and specific circuit components. Since most sources of sine-wave distortion in AC power circuits tend to be symmetrical, even-numbered harmonics are rarely seen in those applications. This is good to know if you're a power system designer and are planning ahead for harmonic reduction: you only have to concern yourself with mitigating the odd harmonic frequencies, even harmonics being practically nonexistent. Also, if you happen to measure even harmonics in an AC circuit with a spectrum analyzer or frequency meter, you know that something in that circuit must be *unsymmetrically* distorting the sine-wave voltage or current, and that clue may be helpful in locating the source of a problem (look for components or conditions more likely to distort one half-cycle of the AC waveform more than the other).

Now that we have this rule to guide our interpretation of nonsinusoidal waveforms, it makes more sense that a waveform like that produced by a rectifier circuit should contain such strong even harmonics, there being no symmetry at all above and below center.

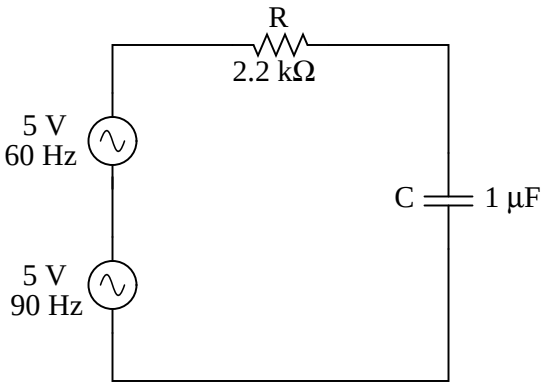
- **REVIEW:**
- Waveforms that are symmetrical above and below their horizontal centerlines contain no even-numbered harmonics.
- The amount of DC "bias" voltage present (a waveform's "DC component") has no impact on that wave's harmonic frequency content.

7.5 Circuit effects

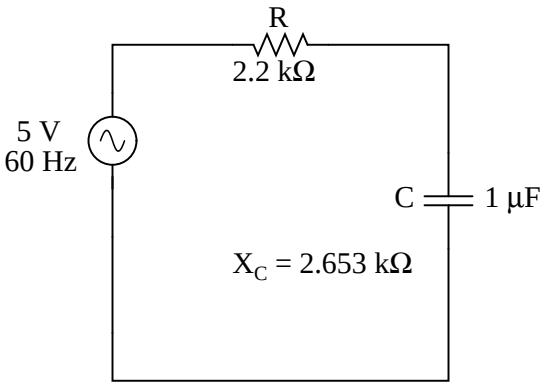
The principle of non-sinusoidal, repeating waveforms being equivalent to a series of sine waves at different frequencies is a fundamental property of waves in general and it has great practical import

in the study of AC circuits. It means that any time we have a waveform that isn't perfectly sine-wave-shaped, the circuit in question will react as though it's having an array of different frequency voltages imposed on it at once.

When an AC circuit is subjected to a source voltage consisting of a mixture of frequencies, the components in that circuit respond to each constituent frequency in a different way. Any reactive component such as a capacitor or an inductor will simultaneously present a unique amount of impedance to each and every frequency present in a circuit. Thankfully, the analysis of such circuits is made relatively easy by applying the *Superposition Theorem*, regarding the multiple-frequency source as a set of single-frequency voltage sources connected in series, and analyzing the circuit for one source at a time, summing the results at the end to determine the aggregate total:

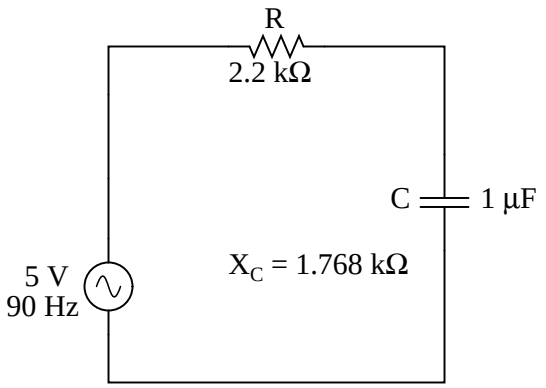


Analyzing circuit for 60 Hz source alone:



	R	C	Total	
E	2.0377 + j2.4569 3.1919 ∠ 50.328°	2.9623 - j2.4569 3.8486 ∠ -39.6716°	5 + j0 5 ∠ 0°	Volts
I	926.22µ + j1.1168m 1.4509m ∠ 50.328°	926.22µ + j1.1168m 1.4509m ∠ 50.328°	926.22µ + j1.1168m 1.4509m ∠ 50.328°	Amps
Z	2.2k + j0 2.2k ∠ 0°	0 - j2.653k 2.653k ∠ -90°	2.2k - j2.653k 3.446k ∠ -50.328°	Ohms

Analyzing the circuit for 90 Hz source alone:



	R	C	Total	
E	3.0375 + j2.4415 3.8971 ∠ 38.793°	1.9625 - j2.4415 3.1325 ∠ -51.207°	5 + j0 5 ∠ 0°	Volts
I	1.3807m + j1.1098m 1.7714m ∠ 38.793°	1.3807m + j1.1098m 1.7714m ∠ 38.793°	1.3807m + j1.1098m 1.7714m ∠ 38.793°	Amps
Z	2.2k + j0 2.2k ∠ 0°	0 - j1.768k 1.768k ∠ -90°	2.2k - j1.768k 2.823k ∠ -38.793°	Ohms

Superimposing the voltage drops across R and C, we get:

$$E_R = [3.1919 \text{ V} \angle 50.328^\circ \text{ (60 Hz)}] + [3.8971 \text{ V} \angle 38.793^\circ \text{ (90 Hz)}]$$

$$E_C = [3.8486 \text{ V} \angle -39.6716^\circ \text{ (60 Hz)}] + [3.1325 \text{ V} \angle -51.207^\circ \text{ (90 Hz)}]$$

Because the two voltages across each component are at different frequencies, we cannot consolidate them into a single voltage figure as we could if we were adding together two voltages of different amplitude and/or phase angle at the same frequency. Complex number notation give us the ability to represent waveform amplitude (polar magnitude) and phase angle (polar angle), but not frequency.

What we can tell from this application of the superposition theorem is that there will be a greater 60 Hz voltage dropped across the capacitor than a 90 Hz voltage. Just the opposite is true for the

resistor's voltage drop. This is worthy to note, especially in light of the fact that the two source voltages are equal. It is this kind of unequal circuit response to signals of differing frequency that will be our specific focus in the next chapter.

We can also apply the superposition theorem to the analysis of a circuit powered by a non-sinusoidal voltage, such as a square wave. If we know the Fourier series (multiple sine/cosine wave equivalent) of that wave, we can regard it as originating from a series-connected string of multiple sinusoidal voltage sources at the appropriate amplitudes, frequencies, and phase shifts. Needless to say, this can be a laborious task for some waveforms (an accurate square-wave Fourier Series is considered to be expressed out to the ninth harmonic, or five sine waves in all!), but it is possible. I mention this not to scare you, but to inform you of the potential complexity lurking behind seemingly simple waveforms. A real-life circuit will respond just the same to being powered by a square wave as being powered by an *infinite* series of sine waves of odd-multiple frequencies and diminishing amplitudes. This has been known to translate into unexpected circuit resonances, transformer and inductor core overheating due to eddy currents, electromagnetic noise over broad ranges of the frequency spectrum, and the like. Technicians and engineers need to be made aware of the potential effects of non-sinusoidal waveforms in reactive circuits.

Harmonics are known to manifest their effects in the form of electromagnetic radiation as well. Studies have been performed on the potential hazards of using portable computers aboard passenger aircraft, citing the fact that computers' high frequency square-wave "clock" voltage signals are capable of generating radio waves that could interfere with the operation of the aircraft's electronic navigation equipment. It's bad enough that typical microprocessor clock signal frequencies are within the range of aircraft radio frequency bands, but worse yet is the fact that the harmonic multiples of those fundamental frequencies span an even larger range, due to the fact that clock signal voltages are square-wave in shape and not sine-wave.

Electromagnetic "emissions" of this nature can be a problem in industrial applications, too, with harmonics abounding in very large quantities due to (nonlinear) electronic control of motor and electric furnace power. The fundamental power line frequency may only be 60 Hz, but those harmonic frequency multiples theoretically extend into infinitely high frequency ranges. Low frequency power line voltage and current doesn't radiate into space very well as electromagnetic energy, but high frequencies do.

Also, capacitive and inductive "coupling" caused by close-proximity conductors is usually more severe at high frequencies. Signal wiring nearby power wiring will tend to "pick up" harmonic interference from the power wiring to a far greater extent than pure sine-wave interference. This problem can manifest itself in industry when old motor controls are replaced with new, solid-state electronic motor controls providing greater energy efficiency. Suddenly there may be weird electrical noise being impressed upon signal wiring that never used to be there, because the old controls never generated harmonics, and those high-frequency harmonic voltages and currents tend to inductively and capacitively "couple" better to nearby conductors than any 60 Hz signals from the old controls used to.

- **REVIEW:**

- Any regular (repeating), non-sinusoidal waveform is equivalent to a particular series of sine/cosine waves of different frequencies, phases, and amplitudes, plus a DC offset voltage if necessary. The mathematical process for determining the sinusoidal waveform equivalent for any waveform is called *Fourier analysis*.

- Multiple-frequency voltage sources can be simulated for analysis by connecting several single-frequency voltage sources in series. Analysis of voltages and currents is accomplished by using the superposition theorem. NOTE: superimposed voltages and currents of different frequencies *cannot* be added together in complex number form, since complex numbers only account for amplitude and phase shift, not frequency!
- Harmonics can cause problems by impressing unwanted ("noise") voltage signals upon nearby circuits. These unwanted signals may come by way of capacitive coupling, inductive coupling, electromagnetic radiation, or a combination thereof.

7.6 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 8

FILTERS

8.1 What is a filter?

It is sometimes desirable to have circuits capable of selectively filtering one frequency or range of frequencies out of a mix of different frequencies in a circuit. A circuit designed to perform this frequency selection is called a *filter circuit*, or simply a *filter*. A common need for filter circuits is in high-performance stereo systems, where certain ranges of audio frequencies need to be amplified or suppressed for best sound quality and power efficiency. You may be familiar with *equalizers*, which allow the amplitudes of several frequency ranges to be adjusted to suit the listener's taste and acoustic properties of the listening area. You may also be familiar with *crossover networks*, which block certain ranges of frequencies from reaching speakers. A tweeter (high-frequency speaker) is inefficient at reproducing low-frequency signals such as drum beats, so a crossover circuit is connected between the tweeter and the stereo's output terminals to block low-frequency signals, only passing high-frequency signals to the speaker's connection terminals. This gives better audio system efficiency and thus better performance. Both equalizers and crossover networks are examples of filters, designed to accomplish filtering of certain frequencies.

Another practical application of filter circuits is in the "conditioning" of non-sinusoidal voltage waveforms in power circuits. Some electronic devices are sensitive to the presence of harmonics in the power supply voltage, and so require power conditioning for proper operation. If a distorted sine-wave voltage behaves like a series of harmonic waveforms added to the fundamental frequency, then it should be possible to construct a filter circuit that only allows the fundamental waveform frequency to pass through, blocking all (higher-frequency) harmonics.

We will be studying the design of several elementary filter circuits in this lesson. To reduce the load of math on the reader, I will make extensive use of SPICE as an analysis tool, displaying Bode plots (amplitude versus frequency) for the various kinds of filters. Bear in mind, though, that these circuits can be analyzed over several points of frequency by repeated series-parallel analysis, much like the previous example with two sources (60 and 90 Hz), if the student is willing to invest a lot of time working and re-working circuit calculations for each frequency.

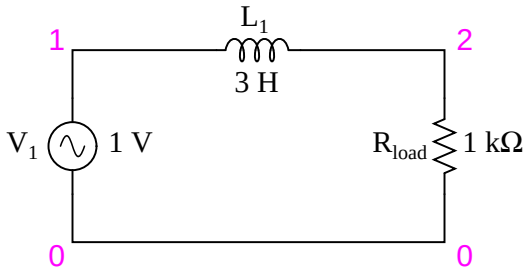
- **REVIEW:**
- A *filter* is an AC circuit that separates some frequencies from others in within mixed-frequency signals.

- Audio *equalizers* and *crossover networks* are two well-known applications of filter circuits.
- A *Bode plot* is a graph plotting waveform amplitude or phase on one axis and frequency on the other.

8.2 Low-pass filters

By definition, a low-pass filter is a circuit offering easy passage to low-frequency signals and difficult passage to high-frequency signals. There are two basic kinds of circuits capable of accomplishing this objective, and many variations of each one:

Inductive low-pass filter



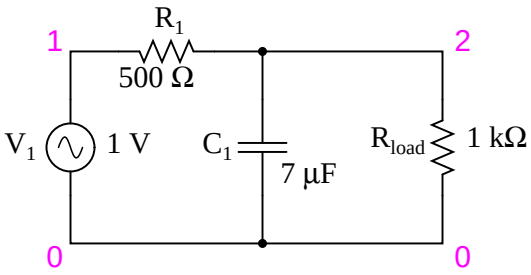
The inductor's impedance increases with increasing frequency. This high impedance in series tends to block high-frequency signals from getting to the load. This can be demonstrated with a SPICE analysis:

```
inductive lowpass filter
v1 1 0 ac 1 sin
l1 1 2 3
rload 2 0 1k
.ac lin 20 1 200
.plot ac v(2)
.end
```

```
freq       v(2)    0.2512      0.3981        0.631          1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+00  9.998E-01 .            .             .            *
1.147E+01  9.774E-01 .            .             .           *.
2.195E+01  9.240E-01 .            .             .          * .
3.242E+01  8.533E-01 .            .             .       *    .
4.289E+01  7.776E-01 .            .             .     *      .
5.337E+01  7.050E-01 .            .             .  *         .
6.384E+01  6.391E-01 .            .            *             .
7.432E+01  5.810E-01 .            .          *  .             .
8.479E+01  5.304E-01 .            .        *   .             .
9.526E+01  4.865E-01 .            .     *      .             .
```

```
1.057E+02  4.485E-01 .            .   *        .            .
1.162E+02  4.153E-01 .            .*           .            .
1.267E+02  3.863E-01 .           *.            .            .
1.372E+02  3.607E-01 .         *  .            .            .
1.476E+02  3.382E-01 .        *   .            .            .
1.581E+02  3.181E-01 .      *     .            .            .
1.686E+02  3.002E-01 .    *       .            .            .
1.791E+02  2.841E-01 .   *        .            .            .
1.895E+02  2.696E-01 . *          .            .            .
2.000E+02  2.564E-01 .*           .            .            .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Load voltage decreases with increasing frequency
```

Capacitive low-pass filter



The capacitor's impedance decreases with increasing frequency. This low impedance in parallel with the load resistance tends to short out high-frequency signals, dropping most of the voltage gets across series resistor R_1.

```
capacitive lowpass filter
v1 1 0 ac 1 sin
r1 1 2 500
c1 2 0 7u
rload 2 0 1k
.ac lin 20 30 150
.plot ac v(2)
.end
```

```
freq       v(2)            0.3162     0.3981     0.5012  0.631
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3.000E+01  6.102E-01 .          .          .          .        *.
3.632E+01  5.885E-01 .          .          .          .      *  .
4.263E+01  5.653E-01 .          .          .          .   *     .
4.895E+01  5.416E-01 .          .          .          .  *      .
5.526E+01  5.180E-01 .          .          .         .*         .
6.158E+01  4.948E-01 .          .          .        *.          .
6.789E+01  4.725E-01 .          .          .     *   .          .
```

```
7.421E+01  4.511E-01 .         .          .   *      .          .
8.053E+01  4.309E-01 .         .          . *        .          .
8.684E+01  4.118E-01 .         .          .*         .          .
9.316E+01  3.938E-01 .         .         *.          .          .
9.947E+01  3.770E-01 .         .       *  .          .          .
1.058E+02  3.613E-01 .         .    *     .          .          .
1.121E+02  3.465E-01 .         .  *       .          .          .
1.184E+02  3.327E-01 .         .*         .          .          .
1.247E+02  3.199E-01 .         *          .          .          .
1.311E+02  3.078E-01 .       * .          .          .          .
1.374E+02  2.965E-01 .    *    .          .          .          .
1.437E+02  2.859E-01 .  *      .          .          .          .
1.500E+02  2.760E-01 .*        .          .          .          .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Load voltage decreases with increasing frequency
```

The inductive low-pass filter is the pinnacle of simplicity, with only one component comprising the filter. The capacitive version of this filter is not that much more complex, with only a resistor and capacitor needed for operation. However, despite their increased complexity, capacitive filter designs are generally preferred over inductive because capacitors tend to be "purer" reactive components than inductors and therefore are more predictable in their behavior. By "pure" I mean that capacitors exhibit little resistive effects than inductors, making them almost 100% reactive. Inductors, on the other hand, typically exhibit significant dissipative (resistor-like) effects, both in the long lengths of wire used to make them, and in the magnetic losses of the core material. Capacitors also tend to participate less in "coupling" effects with other components (generate and/or receive interference from other components via mutual electric or magnetic fields) than inductors, and are less expensive.

However, the inductive low-pass filter is often preferred in AC-DC power supplies to filter out the AC "ripple" waveform created when AC is converted (rectified) into DC, passing only the pure DC component. The primary reason for this is the requirement of low filter resistance for the output of such a power supply. A capacitive low-pass filter requires an extra resistance in series with the source, whereas the inductive low-pass filter does not. In the design of a high-current circuit like a DC power supply where additional series resistance is undesirable, the inductive low-pass filter is the better design choice. On the other hand, if low weight and compact size are higher priorities than low internal supply resistance in a power supply design, the capacitive low-pass filter might make more sense.

All low-pass filters are rated at a certain *cutoff frequency*. That is, the frequency above which the output voltage falls below 70.7% of the input voltage. This cutoff percentage of 70.7 is not really arbitrary, all though it may seem so at first glance. In a simple capacitive/resistive low-pass filter, it is the frequency at which capacitive reactance in ohms equals resistance in ohms. In a simple capacitive low-pass filter (one resistor, one capacitor), the cutoff frequency is given as:

$$f_{cutoff} = \frac{1}{2\pi RC}$$

Inserting the values of R and C from the last SPICE simulation into this formula, we arrive at a cutoff frequency of 45.473 Hz. However, when we look at the plot generated by the SPICE

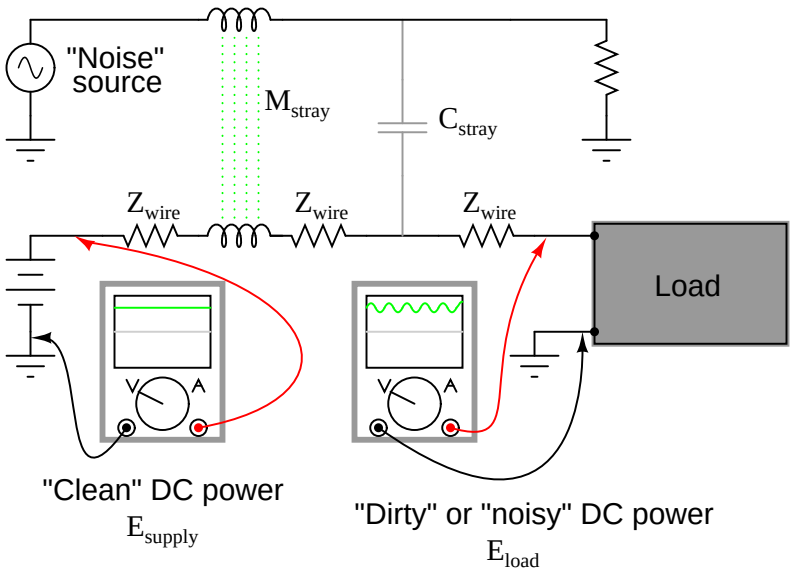
simulation, we see the load voltage well below 70.7% of the source voltage (1 volt) even at a frequency as low as 30 Hz, below the calculated cutoff point. What's wrong? The problem here is that the load resistance of 1 kΩ affects the frequency response of the filter, skewing it down from what the formula told us it would be. Without that load resistance in place, SPICE produces a Bode plot whose numbers make more sense:

```
capacitive lowpass filter
v1 1 0 ac 1 sin
r1 1 2 500
c1 2 0 7u
* note: no load resistor!
.ac lin 20 40 50
.plot ac v(2)
.end
```

```
freq       v(2)    0.6607      0.6918       0.7244       0.7586
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4.000E+01  7.508E-01 .           .             .          *  .
4.053E+01  7.465E-01 .           .             .        *    .
4.105E+01  7.423E-01 .           .             .      *      .
4.158E+01  7.380E-01 .           .             .     *       .
4.211E+01  7.338E-01 .           .             .   *         .
4.263E+01  7.295E-01 .           .             . *           .
4.316E+01  7.253E-01 .           .             *             .
4.368E+01  7.211E-01 .           .            *.             .
4.421E+01  7.170E-01 .           .          *  .             .
4.474E+01  7.129E-01 .           .        *    .             .
4.526E+01  7.087E-01 .           .      *      .             .
4.579E+01  7.046E-01 .           .     *       .             .
4.632E+01  7.006E-01 .           .   *         .             .
4.684E+01  6.965E-01 .           . *           .             .
4.737E+01  6.925E-01 .           *             .             .
4.789E+01  6.885E-01 .          *.             .             .
4.842E+01  6.846E-01 .        *  .             .             .
4.895E+01  6.806E-01 .      *    .             .             .
4.947E+01  6.767E-01 .    *      .             .             .
5.000E+01  6.728E-01 .   *       .             .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
At 45.26 Hz, the output voltage is above 70.7 percent;
At 45.79 Hz, the output voltage is below 70.7 percent;
It should be exactly 70.7% at 45.473 Hz!
```

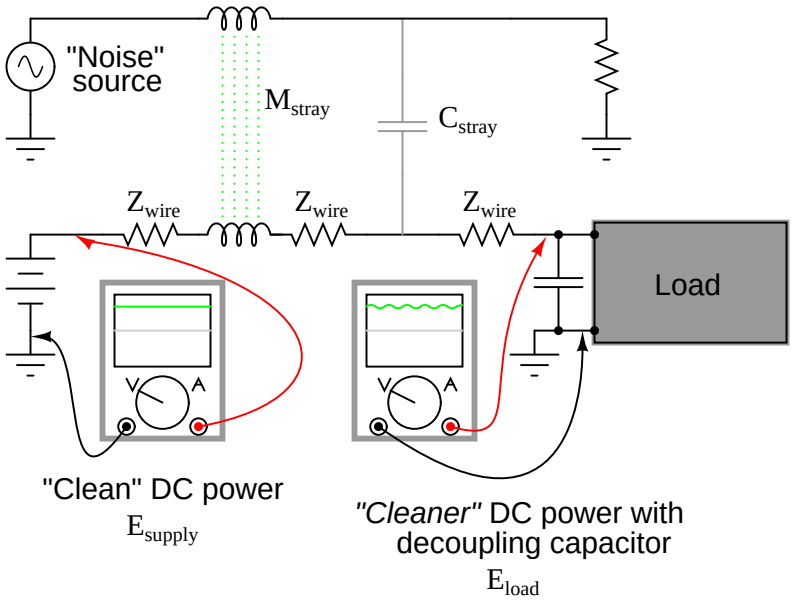
When dealing with filter circuits, it is always important to note that the response of the filter depends on the filter's component values *and* the impedance of the load. If a cutoff frequency equation fails to give consideration to load impedance, it assumes no load and will fail to give accurate results for a real-life filter conducting power to a load.

One frequent application of the capacitive low-pass filter principle is in the design of circuits having components or sections sensitive to electrical "noise." As mentioned at the beginning of the last chapter, sometimes AC signals can "couple" from one circuit to another via capacitance (C_{stray}) and/or mutual inductance (M_{stray}) between the two sets of conductors. A prime example of this is unwanted AC signals ("noise") becoming impressed on DC power lines supplying sensitive circuits:



The oscilloscope-meter on the left shows the "clean" power from the DC voltage source. After coupling with the AC noise source via stray mutual inductance and stray capacitance, though, the voltage as measured at the load terminals is now a mix of AC and DC, the AC being unwanted. Normally, one would expect E_{load} to be precisely identical to E_{source}, because the uninterrupted conductors connecting them should make the two sets of points electrically common. However, power conductor impedance allows the two voltages to differ, which means the noise magnitude can vary at different points in the DC system.

If we wish to prevent such "noise" from reaching the DC load, all we need to do is connect a low-pass filter near the load to block any coupled signals. In its simplest form, this is nothing more than a capacitor connected directly across the power terminals of the load, the capacitor behaving as a very low impedance to any AC noise, and shorting it out. Such a capacitor is called a *decoupling capacitor*:



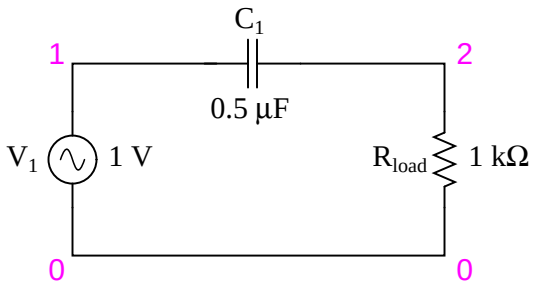
A cursory glance at a crowded printed-circuit board (PCB) will typically reveal decoupling capacitors scattered throughout, usually located as close as possible to the sensitive DC loads. Capacitor size is usually 0.1 μF or more, a minimum amount of capacitance needed to produce a low enough impedance to short out any noise. Greater capacitance will do a better job at filtering noise, but size and economics limit decoupling capacitors to meager values.

- **REVIEW:**
- A low-pass filter allows for easy passage of low-frequency signals from source to load, and difficult passage of high-frequency signals.
- Inductive low-pass filters insert an inductor in series with the load; capacitive low-pass filters insert a resistor in series and a capacitor in parallel with the load. The former filter design tries to "block" the unwanted frequency signal while the latter tries to short it out.
- The *cutoff frequency* for a low-pass filter is that frequency at which the output (load) voltage equals 70.7% of the input (source) voltage. Above the cutoff frequency, the output voltage is lower than 70.7% of the input, and visa-versa.

8.3 High-pass filters

A high-pass filter's task is just the opposite of a low-pass filter: to offer easy passage of a high-frequency signal and difficult passage to a low-frequency signal. As one might expect, the inductive and capacitive versions of the high-pass filter are just the opposite of their respective low-pass filter designs:

Capacitive high-pass filter

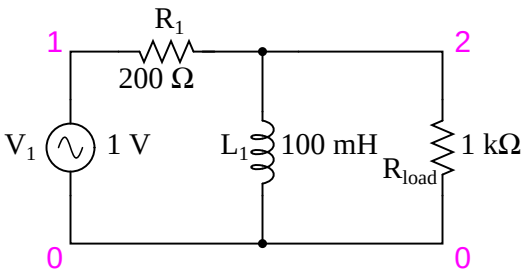


The capacitor's impedance increases with decreasing frequency. This high impedance in series tends to block low-frequency signals from getting to load.

```
capacitive highpass filter
v1 1 0 ac 1 sin
c1 1 2 0.5u
rload 2 0 1k
.ac lin 20 1 200
.plot ac v(2)
.end
```

```
freq       v(2)  1.000E-03     1.000E-02     1.000E-01  1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+00  3.142E-03 .      *      .             .             .
1.147E+01  3.602E-02 .             .       *     .             .
2.195E+01  6.879E-02 .             .           * .             .
3.242E+01  1.013E-01 .             .             *             .
4.289E+01  1.336E-01 .             .             . *           .
5.337E+01  1.654E-01 .             .             .  *          .
6.384E+01  1.966E-01 .             .             .   *         .
7.432E+01  2.274E-01 .             .             .    *        .
8.479E+01  2.574E-01 .             .             .     *       .
9.526E+01  2.867E-01 .             .             .     *       .
1.057E+02  3.152E-01 .             .             .      *      .
1.162E+02  3.429E-01 .             .             .      *      .
1.267E+02  3.698E-01 .             .             .       *     .
1.372E+02  3.957E-01 .             .             .       *     .
1.476E+02  4.207E-01 .             .             .        *    .
1.581E+02  4.448E-01 .             .             .        *    .
1.686E+02  4.680E-01 .             .             .        *    .
1.791E+02  4.903E-01 .             .             .         *   .
1.895E+02  5.116E-01 .             .             .         *   .
2.000E+02  5.320E-01 .             .             .         *   .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Load voltage increases with increasing frequency
```

Inductive high-pass filter



The inductor's impedance decreases with decreasing frequency. This low impedance in parallel tends to short out low-frequency signals from getting to the load resistor. As a consequence, most of the voltage gets dropped across series resistor R_1.

```
inductive highpass filter
v1 1 0 ac 1 sin
r1 1 2 200
l1 2 0 100m
rload 2 0 1k
.ac lin 20 1 200
.plot ac v(2)
.end
```

```
freq       v(2)   1.000E-03  1.000E-02   1.000E-01    1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+00  3.142E-03 .  *      .             .             .
1.147E+01  3.601E-02 .         .       *     .             .
2.195E+01  6.871E-02 .         .           * .             .
3.242E+01  1.011E-01 .         .             *             .
4.289E+01  1.330E-01 .         .             . *           .
5.337E+01  1.644E-01 .         .             .  *          .
6.384E+01  1.950E-01 .         .             .   *         .
7.432E+01  2.248E-01 .         .             .    *        .
8.479E+01  2.537E-01 .         .             .     *       .
9.526E+01  2.817E-01 .         .             .     *       .
1.057E+02  3.086E-01 .         .             .      *      .
1.162E+02  3.344E-01 .         .             .      *      .
1.267E+02  3.591E-01 .         .             .       *     .
1.372E+02  3.828E-01 .         .             .       *     .
1.476E+02  4.053E-01 .         .             .        *    .
1.581E+02  4.267E-01 .         .             .        *    .
1.686E+02  4.470E-01 .         .             .        *    .
1.791E+02  4.662E-01 .         .             .        *    .
1.895E+02  4.845E-01 .         .             .         *   .
2.000E+02  5.017E-01 .         .             .         *   .
```

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Load voltage increases with increasing frequency
```

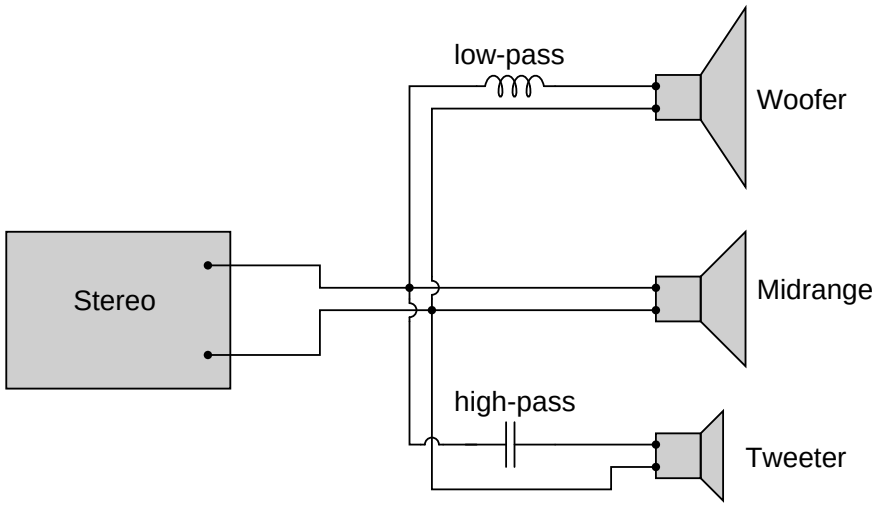
This time, the capacitive design is the simplest, requiring only one component above and beyond the load. And, again, the reactive purity of capacitors over inductors tends to favor their use in filter design, especially with high-pass filters where high frequencies commonly cause inductors to behave strangely due to the skin effect and electromagnetic core losses.

As with low-pass filters, high-pass filters have a rated *cutoff frequency*, above which the output voltage increases above 70.7% of the input voltage. Just as in the case of the capacitive low-pass filter circuit, the capacitive high-pass filter's cutoff frequency can be found with the same formula:

$$f_{cutoff} = \frac{1}{2\pi RC}$$

In the example circuit, there is no resistance other than the load resistor, so that is the value for R in the formula.

Using a stereo system as a practical example, a capacitor connected in series with the tweeter (treble) speaker will serve as a high-pass filter, imposing a high impedance to low-frequency bass signals, thereby preventing that power from being wasted on a speaker inefficient for reproducing such sounds. In like fashion, an inductor connected in series with the woofer (bass) speaker will serve as a low-pass filter for the low frequencies that particular speaker is designed to reproduce. In this simple example circuit, the midrange speaker is subjected to the full spectrum of frequencies from the stereo's output. More elaborate filter networks are sometimes used, but this should give you the general idea. Also bear in mind that I'm only showing you one channel (either left or right) on this stereo system. A real stereo would have six speakers: 2 woofers, 2 midranges, and 2 tweeters.



For better performance yet, we might like to have some kind of filter circuit capable of passing frequencies that are between low (bass) and high (treble) to the midrange speaker so that none of the low- or high-frequency signal power is wasted on a speaker incapable of efficiently reproducing those sounds. What we would be looking for is called a *band-pass* filter, which is the topic of the next section.

- **REVIEW:**

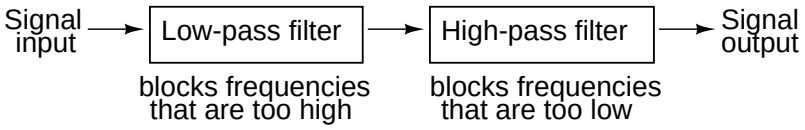
- A high-pass filter allows for easy passage of high-frequency signals from source to load, and difficult passage of low-frequency signals.

- Capacitive high-pass filters insert a capacitor in series with the load; inductive high-pass filters insert a resistor in series and an inductor in parallel with the load. The former filter design tries to "block" the unwanted frequency signal while the latter tries to short it out.

- The *cutoff frequency* for a high-pass filter is that frequency at which the output (load) voltage equals 70.7% of the input (source) voltage. Above the cutoff frequency, the output voltage is greater than 70.7% of the input, and visa-versa.

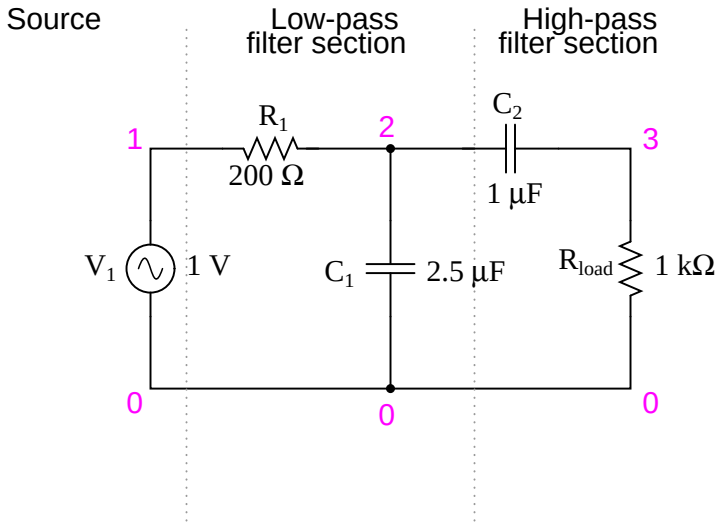
8.4 Band-pass filters

There are applications where a particular band, or spread, or frequencies need to be filtered from a wider range of mixed signals. Filter circuits can be designed to accomplish this task by combining the properties of low-pass and high-pass into a single filter. The result is called a *band-pass* filter. Creating a bandpass filter from a low-pass and high-pass filter can be illustrated using block diagrams:



What emerges from the series combination of these two filter circuits is a circuit that will only allow passage of those frequencies that are neither too high nor too low. Using real components, here is what a typical schematic might look like:

Capacitive band-pass filter

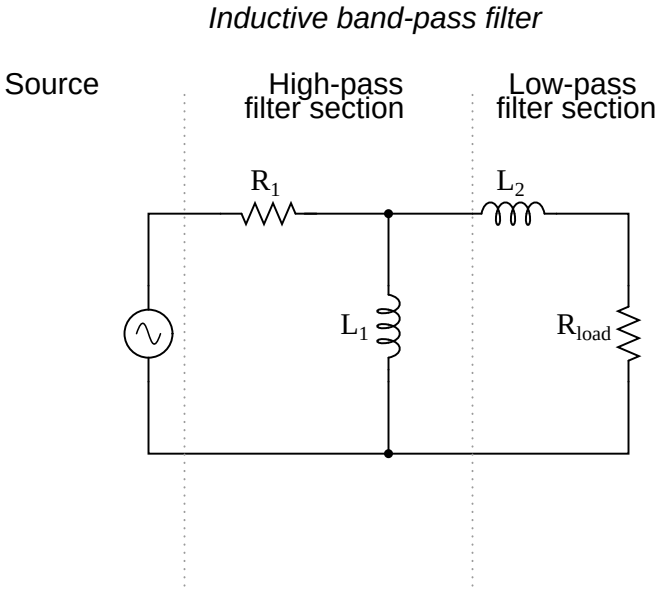



```
capacitive bandpass filter
v1 1 0 ac 1 sin
r1 1 2 200
c1 2 0 2.5u
c2 2 3 1u
rload 3 0 1k
.ac lin 20 100 500
.plot ac v(3)
.end
```

```
freq       v(3)   4.467E-01    5.012E-01    5.623E-01    6.310E-01
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  4.703E-01 .     *       .            .             .
1.211E+02  5.155E-01 .             .  *         .             .
1.421E+02  5.469E-01 .             .          *  .             .
1.632E+02  5.676E-01 .             .            .*            .
1.842E+02  5.801E-01 .             .            .   *         .
2.053E+02  5.865E-01 .             .            .    *        .
2.263E+02  5.882E-01 .             .            .    *        .
2.474E+02  5.864E-01 .             .            .    *        .
2.684E+02  5.820E-01 .             .            .   *         .
2.895E+02  5.755E-01 .             .            .  *          .
3.105E+02  5.676E-01 .             .            .*            .
3.316E+02  5.585E-01 .             .           *.             .
3.526E+02  5.487E-01 .             .          *  .             .
3.737E+02  5.384E-01 .             .        *    .             .
```

```
3.947E+02  5.277E-01 .             .     *       .             .
4.158E+02  5.169E-01 .             .   *         .             .
4.368E+02  5.060E-01 .             .*            .             .
4.579E+02  4.951E-01 .           *.              .             .
4.789E+02  4.843E-01 .         *   .             .             .
5.000E+02  4.736E-01 .      *      .             .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Load voltage peaks within narrow frequency range
```

Band-pass filters can also be constructed using inductors, but as mentioned before, the reactive "purity" of capacitors gives them a design advantage. If we were to design a bandpass filter using inductors, it might look something like this:



The fact that the high-pass section comes "first" in this design instead of the low-pass section makes no difference in its overall operation. It will still filter out all frequencies too high or too low.

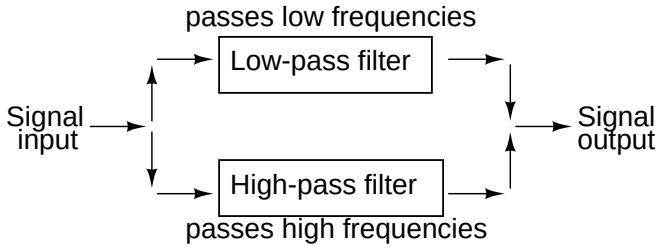
While the general idea of combining low-pass and high-pass filters together to make a bandpass filter is sound, it is not without certain limitations. Because this type of band-pass filter works by relying on either section to *block* unwanted frequencies, it can be difficult to design such a filter to allow unhindered passage within the desired frequency range. Both the low-pass and high-pass sections will always be blocking signals to some extent, and their combined effort makes for an attenuated (reduced amplitude) signal at best, even at the peak of the "pass-band" frequency range. Notice the curve peak on the previous SPICE analysis: the load voltage of this filter never rises above 0.59 volts, although the source voltage is a full volt. This signal attenuation becomes more pronounced if the filter is designed to be more selective (steeper curve, narrower band of passable frequencies).

There are other methods to achieve band-pass operation without sacrificing signal strength within the pass-band. We will discuss those methods a little later in this chapter.

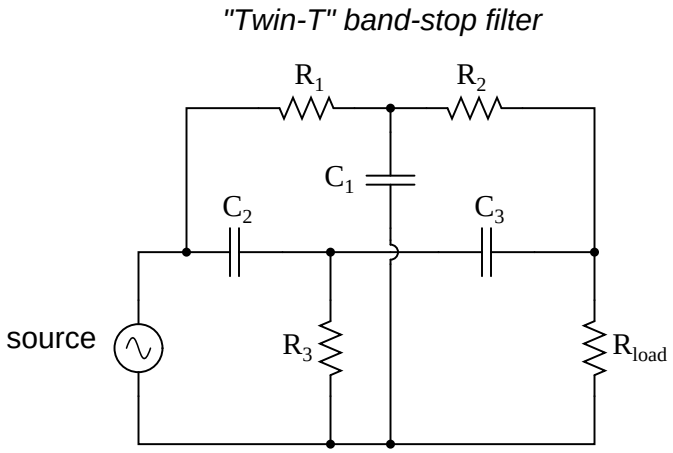
- **REVIEW:**
- A *band-pass* filter works to screen out frequencies that are too low or too high, giving easy passage only to frequencies within a certain range.
- Band-pass filters can be made by stacking a low-pass filter on the end of a high-pass filter, or visa-versa.
- "Attenuate" means to reduce or diminish in amplitude. When you turn down the volume control on your stereo, you are "attenuating" the signal being sent to the speakers.

8.5 Band-stop filters

Also called *band-elimination*, *band-reject*, or *notch* filters, this kind of filter passes all frequencies above and below a particular range set by the component values. Not surprisingly, it can be made out of a low-pass and a high-pass filter, just like the band-pass design, except that this time we connect the two filter sections in parallel with each other instead of in series.



Constructed using two capacitive filter sections, it looks something like this:



The low-pass filter section is comprised of R_1, R_2, and C_1 in a "T" configuration. The high-pass filter section is comprised of C_2, C_3, and R_3 in a "T' configuration as well. Together, this arrangement is commonly known as a "Twin-T" filter, giving sharp response when the component values are chosen in the following ratios:

Component value ratios for the "Twin-T" band-stop filter

$$R_1 = R_2 = 2(R_3)$$

$$C_2 = C_3 = (0.5)C_1$$

Given these component ratios, the frequency of maximum rejection (the "notch frequency") can be calculated as follows:

$$f_{notch} = \frac{1}{4\pi R_3 C_3}$$

The impressive band-stopping ability of this filter is illustrated by the following SPICE analysis:

```
twin-t bandstop filter
v1 1 0 ac 1 sin
r1 1 2 200
c1 2 0 2u
r2 2 3 200
c2 1 4 1u
r3 4 0 100
c3 4 3 1u
rload 3 0 1k
.ac lin 20 200 1.5k
.plot ac v(3)
.end
```

```
freq       v(3)  1.000E-02  3.162E-02  1.000E-01    3.162E-01
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
2.000E+02  5.400E-01 .          .           .              .      *.
2.684E+02  4.512E-01 .          .           .              .   *   .
3.368E+02  3.686E-01 .          .           .              . *     .
4.053E+02  2.946E-01 .          .           .            *.        .
4.737E+02  2.290E-01 .          .           .         *   .        .
5.421E+02  1.707E-01 .          .           .      *      .        .
6.105E+02  1.185E-01 .          .           . *           .        .
6.789E+02  7.134E-02 .          .       *   .             .        .
7.474E+02  2.832E-02 .        *.            .             .        .
8.158E+02  1.126E-02 .*         .            .             .        .
8.842E+02  4.796E-02 .          .    *       .             .        .
9.526E+02  8.222E-02 .          .         * .             .        .
1.021E+03  1.144E-01 .          .           . *           .        .
1.089E+03  1.447E-01 .          .           .   *         .        .
1.158E+03  1.734E-01 .          .           .      *      .        .
1.226E+03  2.007E-01 .          .           .       *     .        .
1.295E+03  2.267E-01 .          .           .         *   .        .
```

```
1.363E+03  2.515E-01 .         .           .          *  .       .
1.432E+03  2.752E-01 .         .           .           * .       .
1.500E+03  2.980E-01 .         .           .            *.       .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

- **REVIEW:**
- A *band-stop* filter works to screen out frequencies that are within a certain range, giving easy passage only to frequencies outside of that range. Also known as *band-elimination*, *band-reject*, or *notch* filters.
- Band-stop filters can be made by placing a low-pass filter in parallel with a high-pass filter. Commonly, both the low-pass and high-pass filter sections are of the "T" configuration, giving the name "Twin-T" to the band-stop combination.
- The frequency of maximum attenuation is called the *notch* frequency.

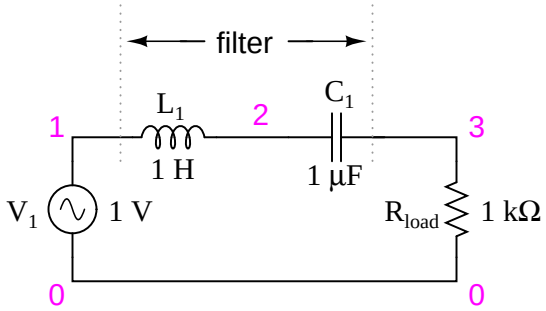
8.6 Resonant filters

So far, the filter designs we've concentrated on have employed *either* capacitors *or* inductors, but never both at the same time. We should know by now that combinations of L and C will tend to resonate, and this property can be exploited in designing band-pass and band-stop filter circuits.

Series LC circuits give minimum impedance at resonance, while parallel LC ("tank") circuits give maximum impedance at their resonant frequency. Knowing this, we have two basic strategies for designing either band-pass or band-stop filters.

For band-pass filters, the two basic resonant strategies are this: series LC to pass a signal, or parallel LC to short a signal. The two schemes will be contrasted and simulated here:

Series resonant band-pass filter



Series LC components pass signal at resonance, and block signals of any other frequencies from getting to the load.

```
series resonant bandpass filter
v1 1 0 ac 1 sin
l1 1 2 1
```

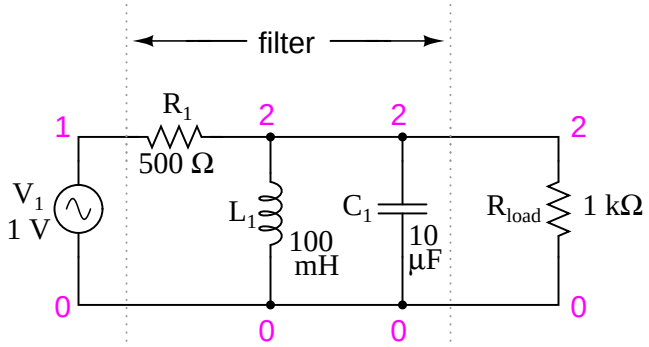
```
c1 2 3 1u
rload 3 0 1k
.ac lin 20 50 250
.plot ac v(3)
.end
```

```
freq       v(3)   2.512E-01   3.981E-01    6.310E-01    1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5.000E+01  3.291E-01 .     *     .             .             .
6.053E+01  4.063E-01 .           .*            .             .
7.105E+01  4.870E-01 .           .     *       .             .
8.158E+01  5.708E-01 .           .          *  .             .
9.211E+01  6.564E-01 .           .             .*            .
1.026E+02  7.411E-01 .           .             .    *        .
1.132E+02  8.210E-01 .           .             .       *     .
1.237E+02  8.910E-01 .           .             .         *   .
1.342E+02  9.460E-01 .           .             .           * .
1.447E+02  9.824E-01 .           .             .            *.
1.553E+02  9.988E-01 .           .             .             *
1.658E+02  9.967E-01 .           .             .             *
1.763E+02  9.796E-01 .           .             .            *.
1.868E+02  9.518E-01 .           .             .           * .
1.974E+02  9.174E-01 .           .             .          *  .
2.079E+02  8.797E-01 .           .             .         *   .
2.184E+02  8.408E-01 .           .             .        *    .
2.289E+02  8.026E-01 .           .             .      *      .
2.395E+02  7.657E-01 .           .             .     *       .
2.500E+02  7.307E-01 .           .             .   *         .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Load voltage peaks at resonant frequency (159.15 Hz)
```

A couple of points to note: see how there is virtually no signal attenuation within the "pass band" (the range of frequencies near the load voltage peak), unlike the band-pass filters made from capacitors or inductors alone. Also, since this filter works on the principle of series LC resonance, the resonant frequency of which is unaffected by circuit resistance, the value of the load resistor will not skew the peak frequency. However, different values for the load resistor *will* change the "steepness" of the Bode plot (the "selectivity" of the filter).

The other basic style of resonant band-pass filters employs a tank circuit (parallel LC combination) to short out signals too high or too low in frequency from getting to the load:

Parallel resonant band-pass filter



The tank circuit will have a lot of impedance at resonance, allowing the signal to get to the load with minimal attenuation. Under or over resonant frequency, however, the tank circuit will have a low impedance, shorting out the signal and dropping most of it across series resistor R_1.

```
parallel resonant bandpass filter
v1 1 0 ac 1 sin
r1 1 2 500
l1 2 0 100m
c1 2 0 10u
rload 2 0 1k
.ac lin 20 50 250
.plot ac v(2)
.end
```

```
freq       v(2)   3.162E-02   1.000E-01    3.162E-01     1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5.000E+01  6.933E-02 .       *   .             .             .
6.053E+01  8.814E-02 .         * .             .             .
7.105E+01  1.100E-01 .           .*            .             .
8.158E+01  1.361E-01 .           .   *         .             .
9.211E+01  1.684E-01 .           .     *       .             .
1.026E+02  2.096E-01 .           .        *    .             .
1.132E+02  2.640E-01 .           .           * .             .
1.237E+02  3.382E-01 .           .             .*            .
1.342E+02  4.392E-01 .           .             .   *         .
1.447E+02  5.630E-01 .           .             .      *      .
1.553E+02  6.578E-01 .           .             .        *    .
1.658E+02  6.432E-01 .           .             .        *    .
1.763E+02  5.503E-01 .           .             .      *      .
1.868E+02  4.543E-01 .           .             .   *         .
1.974E+02  3.792E-01 .           .             . *           .
2.079E+02  3.234E-01 .           .              *            .
2.184E+02  2.816E-01 .           .             *.            .
```

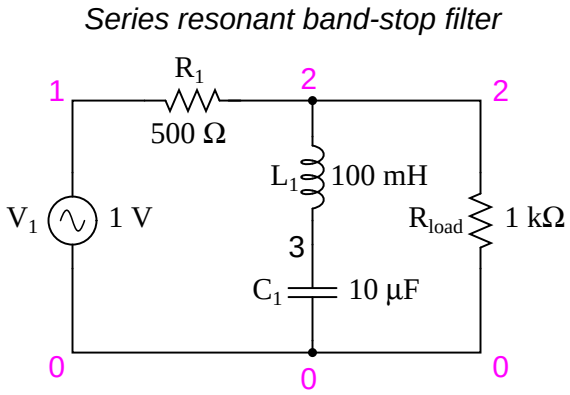
```
2.289E+02  2.495E-01 .           .          *  .             .
2.395E+02  2.242E-01 .           .         *   .             .
2.500E+02  2.038E-01 .           .        *    .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Load voltage peaks at resonant frequency (159.15 Hz)
```

Just like the low-pass and high-pass filter designs relying on a series resistance and a parallel "shorting" component to attenuate unwanted frequencies, this resonant circuit can never provide full input (source) voltage to the load. That series resistance will always be dropping some amount of voltage so long as there is a load resistance connected to the output of the filter.

It should be noted that this form of band-pass filter circuit is very popular in analog radio tuning circuitry, for selecting a particular radio frequency from the multitudes of frequencies available from the antenna. In most analog radio tuner circuits, the rotating dial for station selection moves a variable capacitor in a tank circuit.

The variable capacitor and air-core inductor shown in the above photograph of a simple radio comprise the main elements in the tank circuit filter used to discriminate one radio station's signal from another.

Just as we can use series and parallel LC resonant circuits to pass only those frequencies within a certain range, we can also use them to block frequencies within a certain range, creating a band-stop filter. Again, we have two major strategies to follow in doing this, to use either series or parallel resonance. First, we'll look at the series variety:



When the series LC combination reaches resonance, its very low impedance shorts out the signal, dropping it across resistor R_1 and preventing its passage on to the load.

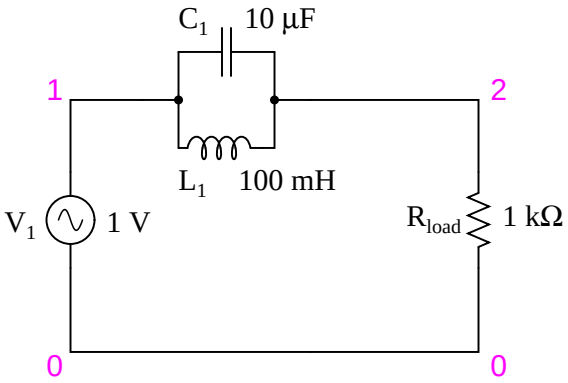
```
series resonant bandstop filter
v1 1 0 ac 1 sin
r1 1 2 500
l1 2 3 100m
c1 3 0 10u
rload 2 0 1k
.ac lin 20 70 230
.plot ac v(2)
.end
```

```
freq       v(2)   1.000E-03  1.000E-02   1.000E-01     1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7.000E+01  3.213E-01 .         .             .      *      .
7.842E+01  2.791E-01 .         .             .     *       .
8.684E+01  2.401E-01 .         .             .    *        .
9.526E+01  2.041E-01 .         .             .   *         .
1.037E+02  1.708E-01 .         .             .  *          .
1.121E+02  1.399E-01 .         .             . *           .
1.205E+02  1.111E-01 .         .             .*            .
1.289E+02  8.413E-02 .         .            *.             .
1.374E+02  5.887E-02 .         .          *  .             .
1.458E+02  3.508E-02 .         .       *     .             .
1.542E+02  1.262E-02 .         .*            .             .
1.626E+02  8.644E-03 .        *.             .             .
1.711E+02  2.884E-02 .         .     *       .             .
1.795E+02  4.805E-02 .         .         *   .             .
1.879E+02  6.638E-02 .         .           * .             .
1.963E+02  8.388E-02 .         .            *.             .
2.047E+02  1.006E-01 .         .             *             .
2.132E+02  1.167E-01 .         .             .*            .
```

```
2.216E+02  1.321E-01 .         .             . *           .
2.300E+02  1.469E-01 .         .             . *           .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Notch frequency = LC resonant frequency (159.15 Hz)
```

Next, we will examine the parallel resonant band-stop filter:

Parallel resonant band-stop filter



The parallel LC components present a high impedance at resonant frequency, thereby blocking the signal from the load at that frequency. Conversely, it passes signals to the load at any other frequencies.

```
parallel resonant bandstop filter
v1 1 0 ac 1 sin
l1 1 2 100m
c1 1 2 10u
rload 2 0 1k
.ac lin 20 100 200
.plot ac v(2)
.end

freq       v(2)  3.162E-02  1.000E-01    3.162E-01    1.000E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  9.947E-01 .         .             .             *     .
1.053E+02  9.932E-01 .         .             .             *     .
1.105E+02  9.911E-01 .         .             .             *     .
1.158E+02  9.883E-01 .         .             .             *     .
1.211E+02  9.841E-01 .         .             .             *     .
1.263E+02  9.778E-01 .         .             .             *     .
1.316E+02  9.675E-01 .         .             .             *     .
1.368E+02  9.497E-01 .         .             .            *.     .
1.421E+02  9.152E-01 .         .             .            *.     .
1.474E+02  8.388E-01 .         .             .           * .     .
```

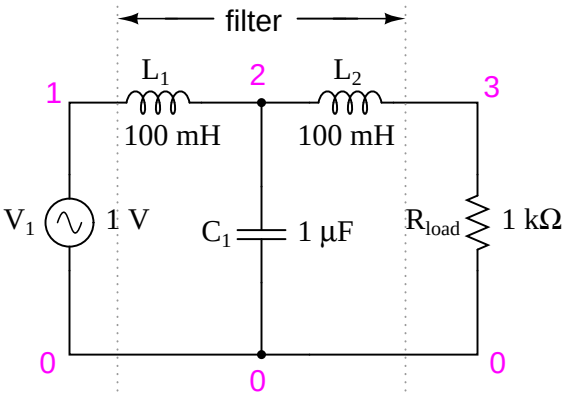
```
1.526E+02  6.420E-01 .         .              .        *    .     .
1.579E+02  1.570E-01 .         .    *         .             .     .
1.632E+02  4.450E-01 .         .              .   *         .     .
1.684E+02  7.496E-01 .         .              .        *    .     .
1.737E+02  8.682E-01 .         .              .           * .     .
1.789E+02  9.201E-01 .         .              .            *.     .
1.842E+02  9.465E-01 .         .              .            *.     .
1.895E+02  9.616E-01 .         .              .             *     .
1.947E+02  9.710E-01 .         .              .             *     .
2.000E+02  9.773E-01 .         .              .             *     .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Notch frequency = LC resonant frequency (159.15 Hz)
```

Once again, notice how the absence of a series resistor makes for minimum attenuation for all the desired (passed) signals. The amplitude at the notch frequency, on the other hand, is very low. In other words, this is a very "selective" filter.

In all these resonant filter designs, the selectivity depends greatly upon the "purity" of the inductance and capacitance used. If there is any stray resistance (especially likely in the inductor), this will diminish the filter's ability to finely discriminate frequencies, as well as introduce antiresonant effects that will skew the peak/notch frequency.

A word of caution to those designing low-pass and high-pass filters is in order at this point. After assessing the standard RC and LR low-pass and high-pass filter designs, it might occur to a student that a better, more effective design of low-pass or high-pass filter might be realized by combining capacitive and inductive elements together like this:

Capacitive/inductive low-pass filter



The inductors should block any high frequencies, while the capacitor should short out any high frequencies as well, both working together to allow only low frequency signals to reach the load.

At first, this seems to be a good strategy, and eliminates the need for a series resistance. However, the more insightful student will recognize that any combination of capacitors and inductors together in a circuit is likely to cause resonant effects to happen at a certain frequency. Resonance, as we have seen before, can cause strange things to happen. Let's plot a SPICE analysis and see what happens over a wide frequency range:

```
lc lowpass filter
v1 1 0 ac 1 sin
l1 1 2 100m
c1 2 0 1u
l2 2 3 100m
rload 3 0 1k
.ac lin 20 100 1k
.plot ac v(3)
.end
```

```
freq       v(3)  1.000E-01  3.162E-01    1.000E+00    3.162E+00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  1.033E+00 .         .             *             .
1.474E+02  1.074E+00 .         .             .*            .
1.947E+02  1.136E+00 .         .             . *           .
2.421E+02  1.228E+00 .         .             . *           .
2.895E+02  1.361E+00 .         .             .   *         .
3.368E+02  1.557E+00 .         .             .    *        .
3.842E+02  1.853E+00 .         .             .       *     .
4.316E+02  2.308E+00 .         .             .         *   .
4.789E+02  2.919E+00 .         .             .            *.
5.263E+02  3.185E+00 .         .             .             *
5.737E+02  2.553E+00 .         .             .          *  .
6.211E+02  1.802E+00 .         .             .      *      .
6.684E+02  1.298E+00 .         .             .  *          .
7.158E+02  9.778E-01 .         .             *             .
7.632E+02  7.650E-01 .         .          *  .             .
8.105E+02  6.165E-01 .         .       *     .             .
8.579E+02  5.084E-01 .         .     *       .             .
9.053E+02  4.268E-01 .         .   *         .             .
9.526E+02  3.635E-01 .         . *           .             .
1.000E+03  3.133E-01 .         *             .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

What was supposed to be a low-pass filter turns out to be a band-pass filter with a peak somewhere around 526 Hz! The capacitance and inductance in this filter circuit are attaining resonance at that point, creating a large voltage drop around C_1, which is seen at the load, regardless of L_2's attenuating influence. The output voltage to the load at this point actually exceeds the input (source) voltage! A little more reflection reveals that if L_1 and C_2 are at resonance, they will impose a very heavy (very low impedance) load on the AC source, which might not be good either. We'll run the same analysis again, only this time plotting C_1's voltage and the source current along with load voltage:

```
legend:
*: v(3)
+: v(2)
```

```
=: i(v1)
freq       v(3)
(*)----------  1.000E-01     3.162E-01     1.000E+00     3.162E+00
(+)----------  3.162E-01     1.000E+00     3.162E+00     1.000E+01
(=)----------  1.000E-03     3.162E-03     1.000E-02     3.162E-02
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  1.033E+00 . =           +             *             .
1.474E+02  1.074E+00 .   =         .+            .*            .
1.947E+02  1.136E+00 .     =       . +           . *           .
2.421E+02  1.228E+00 .       =     .  +          . *           .
2.895E+02  1.361E+00 .          = .   +          .   *         .
3.368E+02  1.557E+00 .            .=    +        .    *        .
3.842E+02  1.853E+00 .            .   =   +      .       *     .
4.316E+02  2.308E+00 .            .       =  +   .         *   .
4.789E+02  2.919E+00 .            .           = +            *.
5.263E+02  3.185E+00 .            .             .x            *
5.737E+02  2.553E+00 .            .           +=.           *  .
6.211E+02  1.802E+00 .            .       + =    .      *      .
6.684E+02  1.298E+00 .            .   +  =       .  *          .
7.158E+02  9.778E-01 .            .+  =          *             .
7.632E+02  7.650E-01 .          + . =         *  .             .
8.105E+02  6.165E-01 .        +   =        *     .             .
8.579E+02  5.084E-01 .      +     =.      *       .             .
9.053E+02  4.268E-01 .    +       = .   *         .             .
9.526E+02  3.635E-01 .   +     =    . *           .             .
1.000E+03  3.133E-01 . +       =    *             .             .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Sure enough, we see the voltage across C_1 and the source current spiking to a high point at the same frequency where the load voltage is maximum. If we were expecting this filter to provide a simple low-pass function, we might be disappointed by the results.

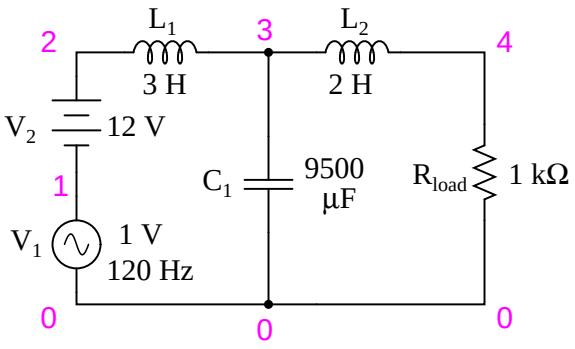
Despite this unintended resonance, low-pass filters made up of capacitors and inductors are frequently used as final stages in AC/DC power supplies to filter the unwanted AC "ripple" voltage out of the DC converted from AC. Why is this, if this particular filter design possesses a potentially troublesome resonant point?

The answer lies in the selection of filter component sizes and the frequencies encountered from an AC/DC converter (rectifier). What we're trying to do in an AC/DC power supply filter is separate DC voltage from a small amount of relatively high-frequency AC voltage. The filter inductors and capacitors are generally quite large (several Henrys for the inductors and thousands of μF for the capacitors is typical), making the filter's resonant frequency very, very low. DC of course, has a "frequency" of zero, so there's no way it can make an LC circuit resonate. The ripple voltage, on the other hand, is a non-sinusoidal AC voltage consisting of a fundamental frequency at least twice the frequency of the converted AC voltage, with harmonics many times that in addition. For plug-in-the-wall power supplies running on 60 Hz AC power (60 Hz United States; 50 Hz in Europe), the lowest frequency the filter will ever see is 120 Hz (100 Hz in Europe), which is well above its resonant point. Therefore, the potentially troublesome resonant point in a such a filter is completely

avoided.

The following SPICE analysis calculates the voltage output (AC and DC) for such a filter, with series DC and AC (120 Hz) voltage sources providing a rough approximation of the mixed-frequency output of an AC/DC converter.

AC/DC power supply filter




```
ac/dc power supply filter
v1 1 0 ac 1 sin
v2 2 1 dc
l1 2 3 3
c1 3 0 9500u
l2 3 4 2
rload 4 0 1k
.dc v2 12 12 1
.ac lin 1 120 120
.print dc v(4)
.print ac v(4)
.end

v2             v(4)
1.200E+01      1.200E+01   DC voltage at load = 12 volts

freq           v(4)
1.200E+02      3.412E-05   AC voltage at load = 34.12 microvolts
```

With a full 12 volts DC at the load and only 34.12 μV of AC left from the 1 volt AC source imposed across the load, this circuit design proves itself to be a very effective power supply filter.

The lesson learned here about resonant effects also applies to the design of high-pass filters using both capacitors and inductors. So long as the desired and undesired frequencies are well to either side of the resonant point, the filter will work okay. But if any signal of significant magnitude close to the resonant frequency is applied to the input of the filter, strange things will happen!

- **REVIEW:**

- Resonant combinations of capacitance and inductance can be employed to create very effective band-pass and band-stop filters without the need for added resistance in a circuit that would diminish the passage of desired frequencies.

- $$f_{resonant} = \frac{1}{2\pi \sqrt{LC}}$$

8.7 Summary

As lengthy as this chapter has been up to this point, it only begins to scratch the surface of filter design. A quick perusal of any advanced filter design textbook is sufficient to prove my point. The mathematics involved with component selection and frequency response prediction is daunting to say the least – well beyond the scope of the beginning electronics student. It has been my intent here to present the basic principles of filter design with as little math as possible, leaning on the power of the SPICE circuit analysis program to explore filter performance. The benefit of such computer simulation software cannot be understated, for the beginning student or for the working engineer.

Circuit simulation software empowers the student to explore circuit designs far beyond the reach of their math skills. With the ability to generate Bode plots and precise figures, an intuitive understanding of circuit concepts can be attained, which is something often lost when a student is burdened with the task of solving lengthy equations by hand. If you are not familiar with the use of SPICE or other circuit simulation programs, take the time to become so! It will be of great benefit to your study. To see SPICE analyses presented in this book is an aid to understanding circuits, but to actually set up and analyze your own circuit simulations is a much more engaging and worthwhile endeavor as a student.

8.8 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

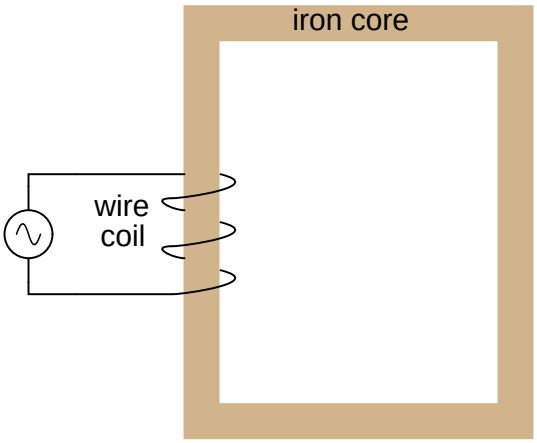
Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 9

TRANSFORMERS

9.1 Mutual inductance and basic operation

Suppose we were to wrap a coil of insulated wire around a loop of ferromagnetic material and energize this coil with an AC voltage source:



As an inductor, we would expect this iron-core coil to oppose the applied voltage with its inductive reactance, limiting current through the coil as predicted by the equations $X_L = 2\pi fL$ and I=E/X (or I=E/Z). For the purposes of this example, though, we need to take a more detailed look at the interactions of voltage, current, and magnetic flux in the device.

Kirchhoff's voltage law describes how the algebraic sum of all voltages in a loop must equal zero. In this example, we could apply this fundamental law of electricity to describe the respective voltages of the source and of the inductor coil. Here, as in any one-source, one-load circuit, the voltage dropped across the load must equal the voltage supplied by the source, assuming zero voltage dropped along the resistance of any connecting wires. In other words, the load (inductor coil) must produce an opposing voltage equal in magnitude to the source, in order that it may balance against the source voltage and produce an algebraic loop voltage sum of zero. From where does this opposing voltage arise? If the load were a resistor, the opposing voltage would originate from the "friction"

of electrons flowing through the resistance of the resistor. With a perfect inductor (no resistance in the coil wire), the opposing voltage comes from another mechanism: the *reaction* to a changing magnetic flux in the iron core.

Michael Faraday discovered the mathematical relationship between magnetic flux (Φ) and induced voltage with this equation:

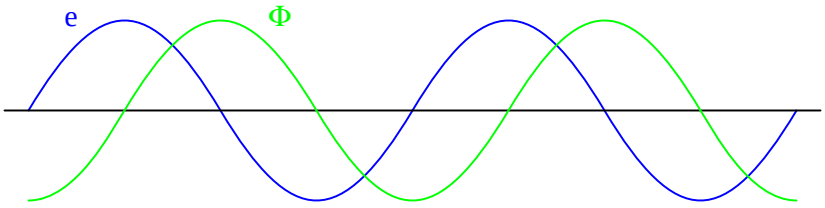
$$e = N \frac{d\Phi}{dt}$$

Where,

e = (Instantaneous) induced voltage in volts
N = Number of turns in wire coil (straight wire = 1)
Φ = Magnetic flux in Webers
t = Time in seconds

The instantaneous voltage (voltage dropped at any instant in time) across a wire coil is equal to the number of turns of that coil around the core (N) multiplied by the instantaneous rate-of-change in magnetic flux ($d\Phi/dt$) linking with the coil. Graphed, this shows itself as a set of sine waves (assuming a sinusoidal voltage source), the flux wave 90^o lagging behind the voltage wave:

e = voltage
Φ = magnetic flux



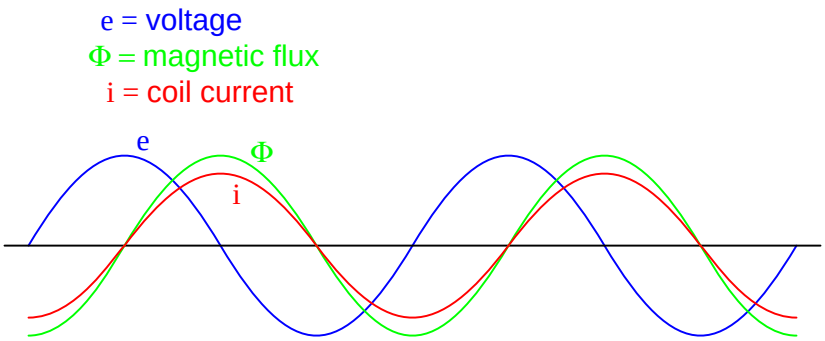
Magnetic flux through a ferromagnetic material is analogous to current through a conductor: it must be motivated by some force in order to occur. In electric circuits, this motivating force is voltage (a.k.a. electromotive force, or EMF). In magnetic "circuits," this motivating force is *magnetomotive force*, or *mmf*. Magnetomotive force (mmf) and magnetic flux (Φ) are related to each other by a property of magnetic materials known as *reluctance* (the latter quantity symbolized by a strange-looking letter "R"):

A comparison of "Ohm's Law" for electric and magnetic circuits:

$E = IR$	$mmf = \Phi\Re$
Electrical	Magnetic

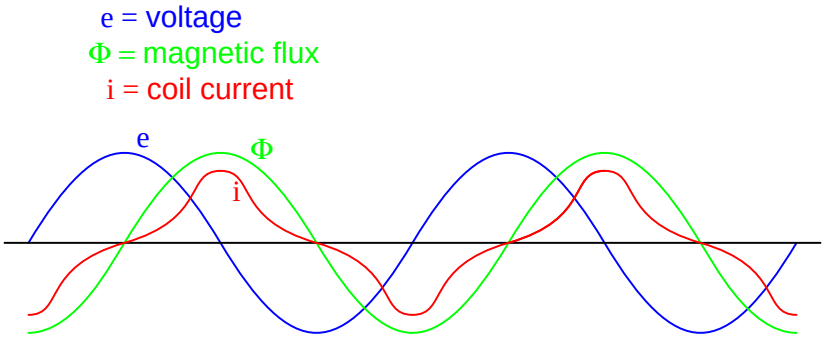
In our example, the mmf required to produce this changing magnetic flux (Φ) must be supplied by a changing current through the coil. Magnetomotive force generated by an electromagnet coil

is equal to the amount of current through that coil (in amps) multiplied by the number of turns of that coil around the core (the SI unit for mmf is the *amp-turn*). Because the mathematical relationship between magnetic flux and mmf is directly proportional, and because the mathematical relationship between mmf and current is also directly proportional (no rates-of-change present in either equation), the current through the coil will be in-phase with the flux wave:



This is why alternating current through an inductor lags the applied voltage waveform by 90°: because that is what is required to produce a changing magnetic flux whose rate-of-change produces an opposing voltage in-phase with the applied voltage. Due to its function in providing magnetizing force (mmf) for the core, this current is sometimes referred to as the *magnetizing current*.

It should be mentioned that the current through an iron-core inductor is not perfectly sinusoidal (sine-wave shaped), due to the nonlinear B/H magnetization curve of iron. In fact, if the inductor is cheaply built, using as little iron as possible, the magnetic flux density might reach high levels (approaching saturation), resulting in a magnetizing current waveform that looks something like this:



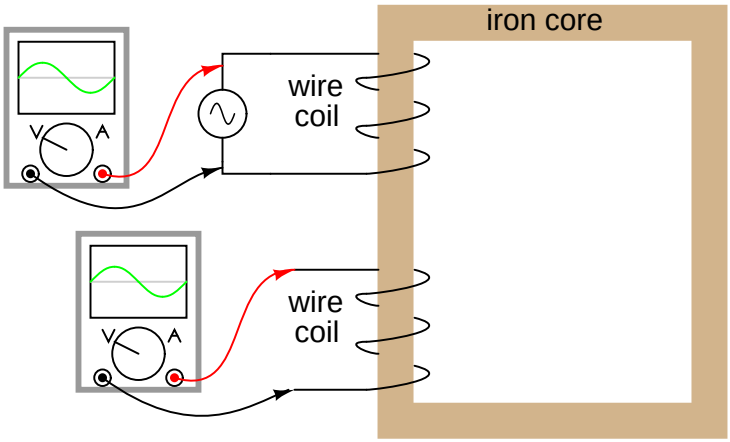
When a ferromagnetic material approaches magnetic flux saturation, disproportionately greater levels of magnetic field force (mmf) are required to deliver equal increases in magnetic field flux (Φ). Because mmf is proportional to current through the magnetizing coil (mmf = NI, where "N" is the number of turns of wire in the coil and "I" is the current through it), the large increases of mmf required to supply the needed increases in flux results in large increases in coil current. Thus, coil current increases dramatically at the peaks in order to maintain a flux waveform that isn't distorted, accounting for the bell-shaped half-cycles of the current waveform in the above plot.

The situation is further complicated by energy losses within the iron core. The effects of hysteresis and eddy currents conspire to further distort and complicate the current waveform, making it even

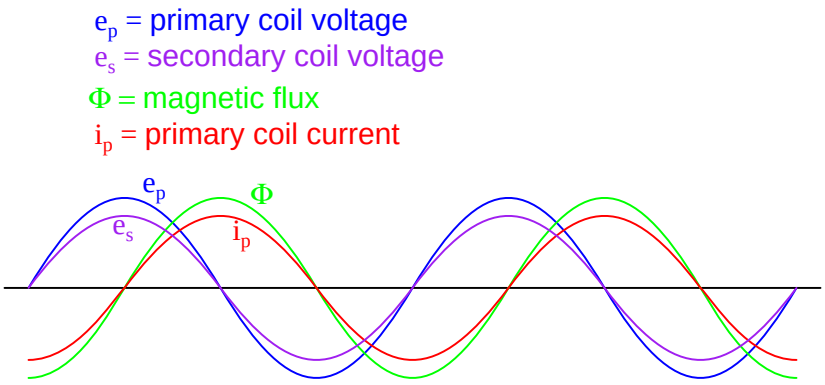
less sinusoidal and altering its phase to be lagging slightly less than 90^o behind the applied voltage waveform. This coil current resulting from the sum total of all magnetic effects in the core ($d\Phi/dt$ magnetization plus hysteresis losses, eddy current losses, etc.) is called the *exciting current*. The distortion of an iron-core inductor's exciting current may be minimized if it is designed for and operated at very low flux densities. Generally speaking, this requires a core with large cross-sectional area, which tends to make the inductor bulky and expensive. For the sake of simplicity, though, we'll assume that our example core is far from saturation and free from all losses, resulting in a perfectly sinusoidal exciting current.

As we've seen already in the inductors chapter, having a current waveform 90^o out of phase with the voltage waveform creates a condition where power is alternately absorbed and returned to the circuit by the inductor. If the inductor is perfect (no wire resistance, no magnetic core losses, etc.), it will dissipate zero power.

Let us now consider the same inductor device, except this time with a second coil wrapped around the same iron core. The first coil will be labeled the *primary* coil, while the second will be labeled the *secondary*:



If this secondary coil experiences the same magnetic flux change as the primary (which it should, assuming perfect containment of the magnetic flux through the common core), and has the same number of turns around the core, a voltage of equal magnitude and phase to the applied voltage will be induced along its length. In the following graph, the induced voltage waveform is drawn slightly smaller than the source voltage waveform simply to distinguish one from the other:



This effect is called *mutual inductance*: the induction of a voltage in one coil in response to a change in current in the other coil. Like normal (self-) inductance, it is measured in the unit of Henrys, but unlike normal inductance it is symbolized by the capital letter "M" rather than the letter "L":

Inductance

$$e = L \frac{di}{dt}$$

Mutual inductance

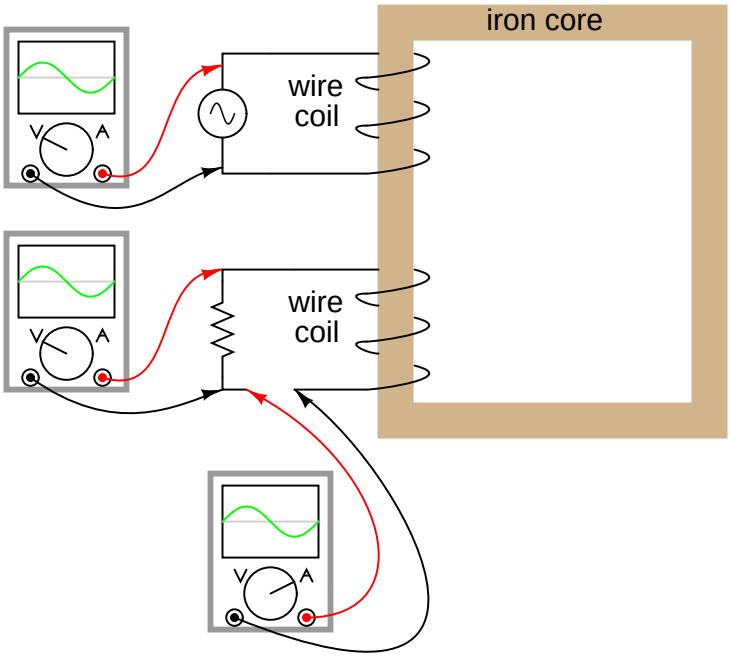
$$e_2 = M \frac{di_1}{dt}$$

Where,

e_2 = voltage induced in secondary coil

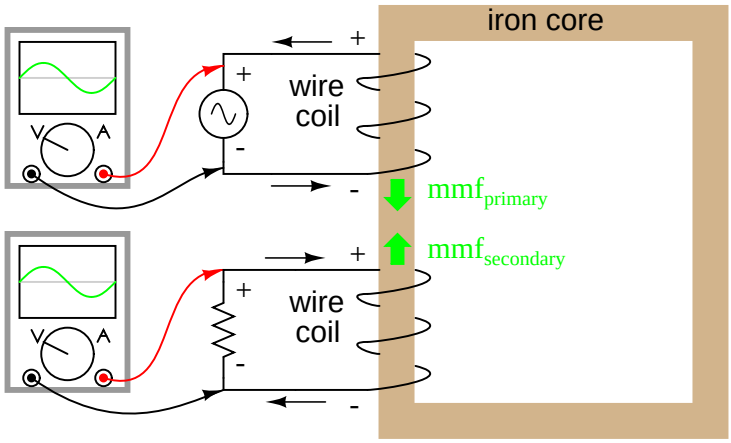
i_1 = current in primary coil

No current will exist in the secondary coil, since it is open-circuited. However, if we connect a load resistor to it, an alternating current will go through the coil, in phase with the induced voltage (because the voltage across a resistor and the current through it are *always* in phase with each other).



At first, one might expect this secondary coil current to cause additional magnetic flux in the core. In fact, it does not. If more flux were induced in the core, it would cause more voltage to be induced voltage in the primary coil (remember that e = dΦ/dt). This cannot happen, because the primary coil's induced voltage must remain at the same magnitude and phase in order to balance with the applied voltage, in accordance with Kirchhoff's voltage law. Consequently, the magnetic flux in the core cannot be affected by secondary coil current. However, what *does* change is the amount of mmf in the magnetic circuit.

Magnetomotive force is produced any time electrons move through a wire. Usually, this mmf is accompanied by magnetic flux, in accordance with the mmf=ΦR "magnetic Ohm's Law" equation. In this case, though, additional flux is not permitted, so the only way the secondary coil's mmf may exist is if a counteracting mmf is generated by the primary coil, of equal magnitude and opposite phase. Indeed, this is what happens, an alternating current forming in the primary coil – 180^o out of phase with the secondary coil's current – to generate this counteracting mmf and prevent additional core flux. Polarity marks and current direction arrows have been added to the illustration to clarify phase relations:



If you find this process a bit confusing, do not worry. Transformer dynamics is a complex subject. What is important to understand is this: when an AC voltage is applied to the primary coil, it creates a magnetic flux in the core, which induces AC voltage in the secondary coil in-phase with the source voltage. Any current drawn through the secondary coil to power a load induces a corresponding current in the primary coil, drawing current from the source.

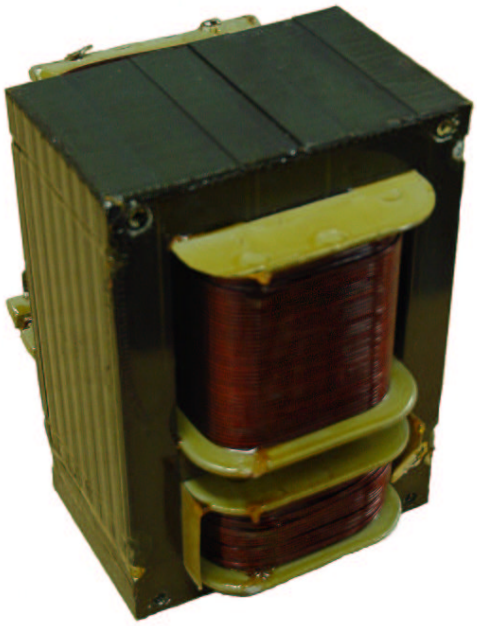
Notice how the primary coil is behaving as a load with respect to the AC voltage source, and how the secondary coil is behaving as a source with respect to the resistor. Rather than energy merely being alternately absorbed and returned the primary coil circuit, energy is now being *coupled* to the secondary coil where it is delivered to a dissipative (energy-consuming) load. As far as the source "knows," it's directly powering the resistor. Of course, there is also an additional primary coil current lagging the applied voltage by 90^o, just enough to magnetize the core to create the necessary voltage for balancing against the source (the *exciting current*).

We call this type of device a *transformer*, because it transforms electrical energy into magnetic energy, then back into electrical energy again. Because its operation depends on electromagnetic induction between two stationary coils and a magnetic flux of changing magnitude and "polarity," transformers are necessarily AC devices. Its schematic symbol looks like two inductors (coils) sharing the same magnetic core:

Transformer

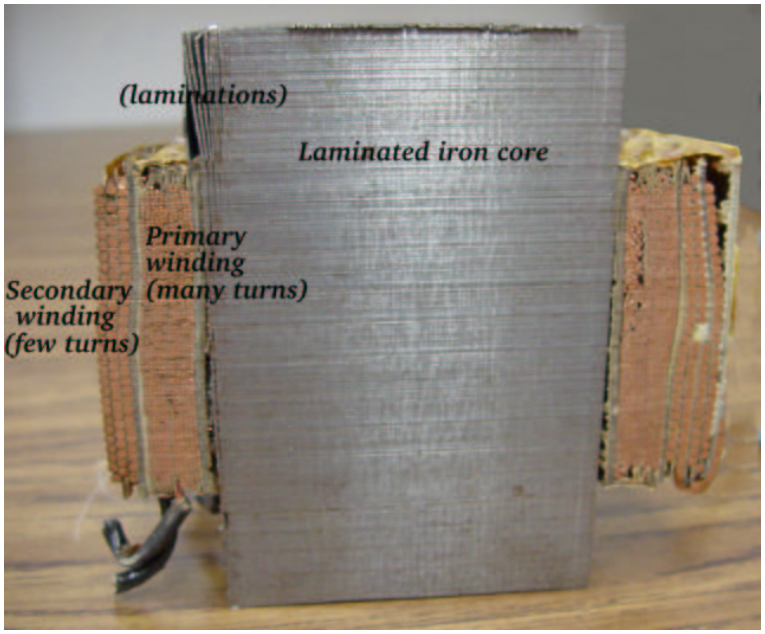
The two inductor coils are easily distinguished in the above symbol. The pair of vertical lines represent an iron core common to both inductors. While many transformers have ferromagnetic core materials, there are some that do not, their constituent inductors being magnetically linked together through the air.

The following photograph shows a power transformer of the type used in gas-discharge lighting. Here, the two inductor coils can be clearly seen, wound around an iron core. While most transformer designs enclose the coils and core in a metal frame for protection, this particular transformer is open for viewing and so serves its illustrative purpose well:

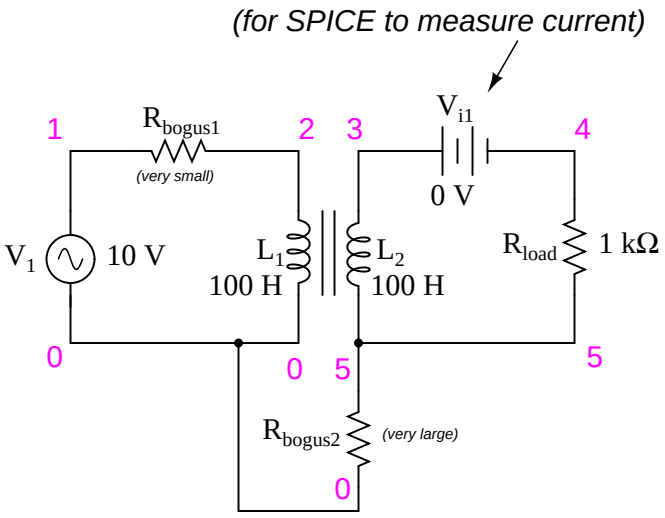


Both coils of wire can be seen here with copper-colored varnish insulation. The top coil is larger than the bottom coil, having a greater number of "turns" around the core. In transformers, the inductor coils are often referred to as *windings*, in reference to the manufacturing process where wire is *wound* around the core material. As modeled in our initial example, the powered inductor of a transformer is called the *primary* winding, while the unpowered coil is called the *secondary* winding.

In the next photograph, a transformer is shown cut in half, exposing the cross-section of the iron core as well as both windings. Like the transformer shown previously, this unit also utilizes primary and secondary windings of differing turn counts. The wire gauge can also be seen to differ between primary and secondary windings. The reason for this disparity in wire gauge will be made clear in the next section of this chapter. Additionally, the iron core can be seen in this photograph to be made of many thin sheets (laminations) rather than a solid piece. The reason for this will also be explained in a later section of this chapter.



It is easy to demonstrate simple transformer action using SPICE, setting up the primary and secondary windings of the simulated transformer as a pair of "mutual" inductors. The coefficient of magnetic field coupling is given at the end of the "`k`" line in the SPICE circuit description, this example being set very nearly at perfection (1.000). This coefficient describes how closely "linked" the two inductors are, magnetically. The better these two inductors are magnetically coupled, the more efficient the energy transfer between them should be.




```
transformer
v1 1 0 ac 10 sin
```

```
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
l1 2 0 100
l2 3 5 100
** This line tells SPICE that the two inductors
** l1 and l2 are magnetically "linked" together
k l1 l2 0.999
vi1 3 4 ac 0
rload 4 5 1k
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end
```

Note: the R_{bogus} resistors are required to satisfy certain quirks of SPICE. The first breaks the otherwise continuous loop between the voltage source and L_1 which would not be permitted by SPICE. The second provides a path to ground (node 0) from the secondary circuit, necessary because SPICE cannot function with any ungrounded circuits.

```
freq          v(2)        i(v1)
6.000E+01     1.000E+01   9.975E-03   Primary winding

freq          v(3,5)      i(vi1)
6.000E+01     9.962E+00   9.962E-03   Secondary winding
```

Note that with equal inductances for both windings (100 Henrys each), the AC voltages and currents are nearly equal for the two. The difference between primary and secondary currents is the magnetizing current spoken of earlier: the 90^o lagging current necessary to magnetize the core. As is seen here, it is usually very small compared to primary current induced by the load, and so the primary and secondary currents are almost equal. What you are seeing here is quite typical of transformer efficiency. Anything less than 95% efficiency is considered poor for modern power transformer designs, and this transfer of power occurs with no moving parts or other components subject to wear.

If we decrease the load resistance so as to draw more current with the same amount of voltage, we see that the current through the primary winding increases in response. Even though the AC power source is not directly connected to the load resistance (rather, it is electromagnetically "coupled"), the amount of current drawn from the source will be almost the same as the amount of current that would be drawn if the load were directly connected to the source. Take a close look at the next two SPICE simulations, showing what happens with different values of load resistors:

```
transformer
v1 1 0 ac 10 sin
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
l1 2 0 100
l2 3 5 100
```

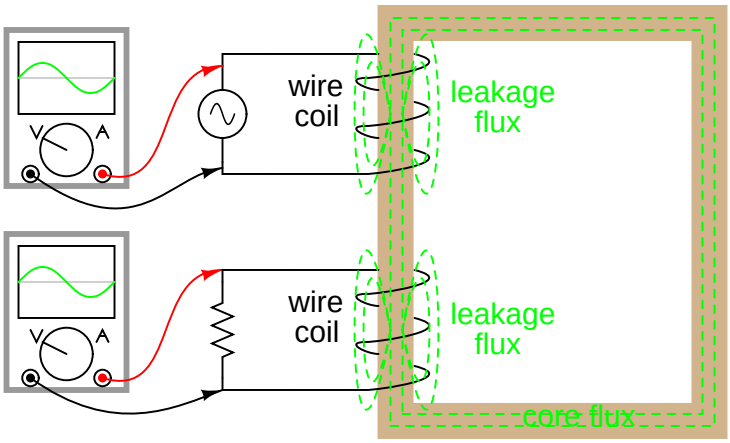
```
k l1 l2 0.999
vi1 3 4 ac 0
** Note load resistance value of 200 ohms
rload 4 5 200
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

freq          v(2)        i(v1)
6.000E+01     1.000E+01   4.679E-02

freq          v(3,5)      i(vi1)
6.000E+01     9.348E+00   4.674E-02
```

Notice how the primary current closely follows the secondary current. In our first simulation, both currents were approximately 10 mA, but now they are both around 47 mA. In this second simulation, the two currents are closer to equality, because the magnetizing current remains the same as before while the load current has increased. Note also how the secondary voltage has decreased some with the heavier (greater current) load. Let's try another simulation with an even lower value of load resistance (15 Ω):

```
transformer
v1 1 0 ac 10 sin
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
l1 2 0 100
l2 3 5 100
k l1 l2 0.999
vi1 3 4 ac 0
rload 4 5 15
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

freq          v(2)        i(v1)
6.000E+01     1.000E+01   1.301E-01

freq          v(3,5)      i(vi1)
6.000E+01     1.950E+00   1.300E-01
```

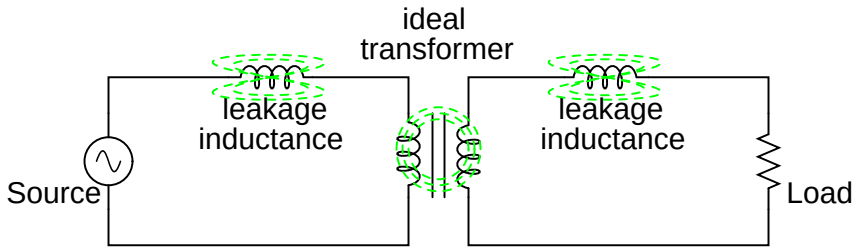
Our load current is now 0.13 amps, or 130 mA, which is substantially higher than the last time. The primary current is very close to being the same, but notice how the secondary voltage has fallen well below the primary voltage (1.95 volts versus 10 volts at the primary). The reason for this is an imperfection in our transformer design: because the primary and secondary inductances aren't

perfectly linked (a `k` factor of 0.999 instead of 1.000) there is "stray" or "*leakage*" inductance. In other words, some of the magnetic field isn't linking with the secondary coil, and thus cannot couple energy to it:



Consequently, this "leakage" flux merely stores and returns energy to the source circuit via self-inductance, effectively acting as a series impedance in both primary and secondary circuits. Voltage gets dropped across this series impedance, resulting in a reduced load voltage: voltage across the load "sags" as load current increases.

Equivalent transformer circuit with leakage inductance shown



If we change the transformer design to have better magnetic coupling between the primary and secondary coils, the figures for voltage between primary and secondary windings will be much closer to equality again:

```
transformer
v1 1 0 ac 10 sin
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
l1 2 0 100
l2 3 5 100
** Coupling factor = 0.99999 instead of 0.999
k l1 l2 0.99999
```

```
vi1 3 4 ac 0
rload 4 5 15
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

freq          v(2)        i(v1)
6.000E+01     1.000E+01   6.658E-01

freq          v(3,5)      i(vi1)
6.000E+01     9.987E+00   6.658E-01
```

Here we see that our secondary voltage is back to being equal with the primary, and the secondary current is equal to the primary current as well. Unfortunately, building a real transformer with coupling this complete is very difficult. A compromise solution is to design both primary and secondary coils with less inductance, the strategy being that less inductance overall leads to less "leakage" inductance to cause trouble, for any given degree of magnetic coupling inefficiency. This results in a load voltage that is closer to ideal with the same (heavy) load and the same coupling factor:

```
transformer
v1 1 0 ac 10 sin
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
** inductance = 1 henry instead of 100 henrys
l1 2 0 1
l2 3 5 1
k l1 l2 0.999
vi1 3 4 ac 0
rload 4 5 15
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

freq          v(2)        i(v1)
6.000E+01     1.000E+01   6.664E-01

freq          v(3,5)      i(vi1)
6.000E+01     9.977E+00   6.652E-01
```

Simply by using primary and secondary coils of less inductance, the load voltage for this heavy load has been brought back up to nearly ideal levels (9.977 volts). At this point, one might ask, "If less inductance is all that's needed to achieve near-ideal performance under heavy load, then why worry about coupling efficiency at all? If it's impossible to build a transformer with perfect

coupling, but easy to design coils with low inductance, then why not just build all transformers with low-inductance coils and have excellent efficiency even with poor magnetic coupling?"

The answer to this question is found in another simulation: the same low-inductance transformer, but this time with a lighter load (1 kΩ instead of 15 Ω):

```
transformer
v1 1 0 ac 10 sin
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
l1 2 0 1
l2 3 5 1
k l1 l2 0.999
vi1 3 4 ac 0
rload 4 5 1k
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

freq          v(2)        i(v1)
6.000E+01     1.000E+01   2.835E-02

freq          v(3,5)      i(vi1)
6.000E+01     9.990E+00   9.990E-03
```

With lower winding inductances, the primary and secondary voltages are closer to being equal, but the primary and secondary currents are not. In this particular case, the primary current is 28.35 mA while the secondary current is only 9.990 mA: almost three times as much current in the primary as the secondary. Why is this? With less inductance in the primary winding, there is less inductive reactance, and consequently a much larger magnetizing current. A substantial amount of the current through the primary winding merely works to magnetize the core rather than *transfer* useful energy to the secondary winding and load.

An ideal transformer with identical primary and secondary windings would manifest equal voltage and current in both sets of windings for any load condition. In a perfect world, transformers would transfer electrical power from primary to secondary as smoothly as though the load were directly connected to the primary power source, with no transformer there at all. However, you can see this ideal goal can only be met if there is *perfect* coupling of magnetic flux between primary and secondary windings. Being that this is impossible to achieve, transformers must be designed to operate within certain expected ranges of voltages and loads in order to perform as close to ideal as possible. For now, the most important thing to keep in mind is a transformer's basic operating principle: the transfer of power from the primary to the secondary circuit via electromagnetic coupling.

- **REVIEW:**
- *Mutual inductance* is where the magnetic flux of two or more inductors are "linked" so that voltage is induced in one coil proportional to the rate-of-change of current in another.

- A *transformer* is a device made of two or more inductors, one of which is powered by AC, inducing an AC voltage across the second inductor. If the second inductor is connected to a load, power will be electromagnetically coupled from the first inductor's power source to that load.
- The powered inductor in a transformer is called the *primary winding*. The unpowered inductor in a transformer is called the *secondary winding*.
- Magnetic flux in the core (Φ) lags 90^o behind the source voltage waveform. The current drawn by the primary coil from the source to produce this flux is called the *magnetizing current*, and it also lags the supply voltage by 90^o.
- Total primary current in an unloaded transformer is called the *exciting current*, and is comprised of magnetizing current plus any additional current necessary to overcome core losses. It is never perfectly sinusoidal in a real transformer, but may be made more so if the transformer is designed and operated so that magnetic flux density is kept to a minimum.
- Core flux induces a voltage in any coil wrapped around the core. The induces voltage(s) are ideally in phase with the primary winding source voltage and share the same waveshape.
- Any current drawn through the secondary winding by a load will be "reflected" to the primary winding and drawn from the voltage source, as if the source were directly powering a similar load.

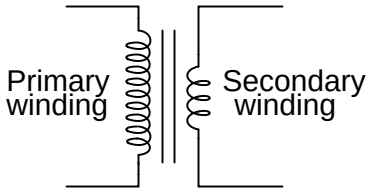
9.2 Step-up and step-down transformers

So far, we've observed simulations of transformers where the primary and secondary windings were of identical inductance, giving approximately equal voltage and current levels in both circuits. Equality of voltage and current between the primary and secondary sides of a transformer, however, is not the norm for all transformers. If the inductances of the two windings are not equal, something interesting happens:

```
transformer
v1 1 0 ac 10 sin
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
l1 2 0 10000
l2 3 5 100
k l1 l2 0.999
vi1 3 4 ac 0
rload 4 5 1k
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

freq          v(2)        i(v1)
```

```
6.000E+01     1.000E+01   9.975E-05    Primary winding



freq          v(3,5)      i(vi1)
6.000E+01     9.962E-01   9.962E-04    Secondary winding
```

Notice how the secondary voltage is approximately ten times less than the primary voltage (0.9962 volts compared to 10 volts), while the secondary current is approximately ten times greater (0.9962 mA compared to 0.09975 mA). What we have here is a device that steps voltage *down* by a factor of ten and current *up* by a factor of ten:

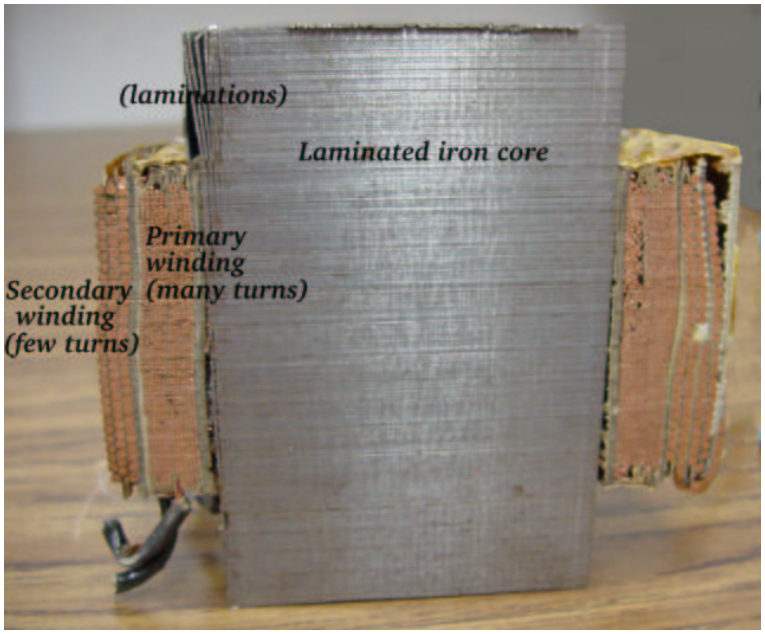


10:1 primary:secondary voltage ratio

1:10 primary:secondary current ratio

This is a very useful device, indeed. With it, we can easily multiply or divide voltage and current in AC circuits. Indeed, the transformer has made long-distance transmission of electric power a practical reality, as AC voltage can be "stepped up" and current "stepped down" for reduced wire resistance power losses along power lines connecting generating stations with loads. At either end (both the generator and at the loads), voltage levels are reduced by transformers for safer operation and less expensive equipment. A transformer that increases voltage from primary to secondary (more secondary winding turns than primary winding turns) is called a *step-up* transformer. Conversely, a transformer designed to do just the opposite is called a *step-down* transformer.

Let's re-examine a photograph shown in the previous section:



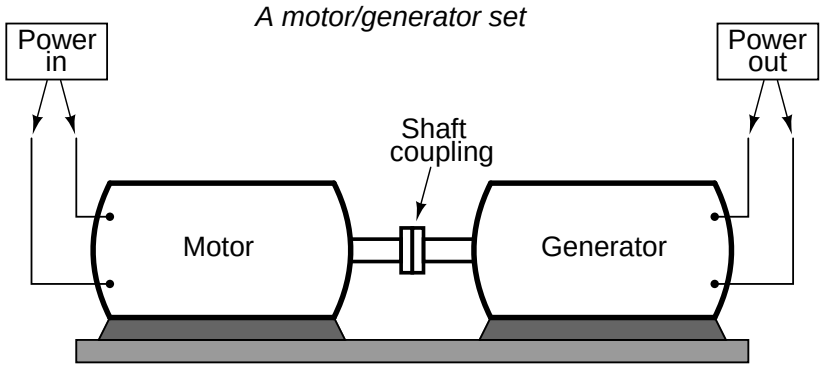
This is a step-down transformer, as evidenced by the high turn count of the primary winding and the low turn count of the secondary. As a step-down unit, this transformer converts high-voltage, low-current power into low-voltage, high-current power. The larger-gauge wire used in the secondary winding is necessary due to the increase in current. The primary winding, which doesn't have to conduct as much current, may be made of smaller-gauge wire.

In case you were wondering, it *is* possible to operate either of these transformer types backwards (powering the secondary winding with an AC source and letting the primary winding power a load) to perform the opposite function: a step-up can function as a step-down and visa-versa. However, as we saw in the first section of this chapter, efficient operation of a transformer requires that the individual winding inductances be engineered for specific operating ranges of voltage and current, so if a transformer is to be used "backwards" like this it must be employed within the original design parameters of voltage and current for each winding, lest it prove to be inefficient (or lest it be *damaged* by excessive voltage or current!).

Transformers are often constructed in such a way that it is not obvious which wires lead to the primary winding and which lead to the secondary. One convention used in the electric power industry to help alleviate confusion is the use of "H" designations for the higher-voltage winding (the primary winding in a step-down unit; the secondary winding in a step-up) and "X" designations for the lower-voltage winding. Therefore, a simple power transformer will have wires labeled "H_1", "H_2", "X_1", and "X_2". There is usually significance to the numbering of the wires (H_1 versus H_2, etc.), which we'll explore a little later in this chapter.

The fact that voltage and current get "stepped" in opposite directions (one up, the other down) makes perfect sense when you recall that power is equal to voltage times current, and realize that transformers cannot *produce* power, only convert it. Any device that could output more power than it took in would violate the *Law of Energy Conservation* in physics, namely that energy cannot be created or destroyed, only converted. As with the first transformer example we looked at, power transfer efficiency is very good from the primary to the secondary sides of the device.

The practical significance of this is made more apparent when an alternative is considered: before the advent of efficient transformers, voltage/current level conversion could only be achieved through the use of motor/generator sets. A drawing of a motor/generator set reveals the basic principle involved:



In such a machine, a motor is mechanically coupled to a generator, the generator designed to produce the desired levels of voltage and current at the rotating speed of the motor. While both motors and generators are fairly efficient devices, the use of both in this fashion compounds their inefficiencies so that the overall efficiency is in the range of 90% or less. Furthermore, because motor/generator sets obviously require moving parts, mechanical wear and balance are factors influencing both service life and performance. Transformers, on the other hand, are able to convert levels of AC voltage and current at very high efficiencies with no moving parts, making possible the widespread distribution and use of electric power we take for granted.

In all fairness it should be noted that motor/generator sets have not necessarily been obsoleted by transformers for *all* applications. While transformers are clearly superior over motor/generator sets for AC voltage and current level conversion, they cannot convert one frequency of AC power to another, or (by themselves) convert DC to AC or visa-versa. Motor/generator sets can do all these things with relative simplicity, albeit with the limitations of efficiency and mechanical factors already described. Motor/generator sets also have the unique property of kinetic energy storage: that is, if the motor's power supply is momentarily interrupted for any reason, its angular momentum (the inertia of that rotating mass) will maintain rotation of the generator for a short duration, thus isolating any loads powered by the generator from "glitches" in the main power system.

Looking closely at the numbers in the SPICE analysis, we should see a correspondence between the transformer's ratio and the two inductances. Notice how the primary inductor (l1) has 100 times more inductance than the secondary inductor (10000 H versus 100 H), and that the measured voltage step-down ratio was 10 to 1. The winding with more inductance will have higher voltage and less current than the other. Since the two inductors are wound around the same core material in the transformer (for the most efficient magnetic coupling between the two), the parameters affecting inductance for the two coils are equal except for the number of turns in each coil. If we take another look at our inductance formula, we see that inductance is proportional to the *square* of the number of coil turns:

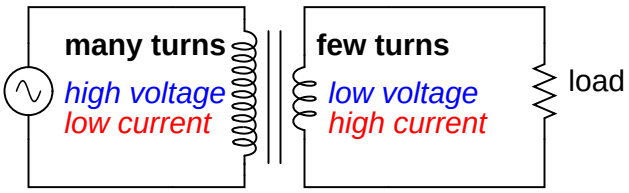
$$L = \frac{N^2 \mu A}{l}$$

Where,

L = Inductance of coil in Henrys
N = Number of turns in wire coil (straight wire = 1)
μ = Permeability of core material (absolute, not relative)
A = Area of coil in square meters
l = Average length of coil in meters

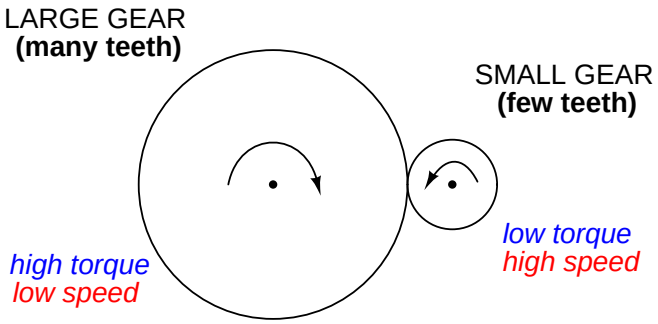
So, it should be apparent that our two inductors in the last SPICE transformer example circuit – with inductance ratios of 100:1 – should have coil turn ratios of 10:1, because 10 squared equals 100. This works out to be the same ratio we found between primary and secondary voltages and currents (10:1), so we can say as a rule that the voltage and current transformation ratio is equal to the ratio of winding turns between primary and secondary.

Step-down transformer



The step-up/step-down effect of coil turn ratios in a transformer is analogous to gear tooth ratios in mechanical gear systems, transforming values of speed and torque in much the same way:

Torque reduction geartrain



Step-up and step-down transformers for power distribution purposes can be gigantic in proportion to the power transformers previously shown, some units standing as tall as a home. The following photograph shows a substation transformer standing about twelve feet tall:

- **REVIEW:**
- Transformers "step up" or "step down" voltage according to the ratios of primary to secondary wire turns.

$$\text{Voltage transformation ratio} = \frac{N_{secondary}}{N_{primary}}$$

$$\text{Current transformation ratio} = \frac{N_{primary}}{N_{secondary}}$$

Where,

N = number of turns in winding

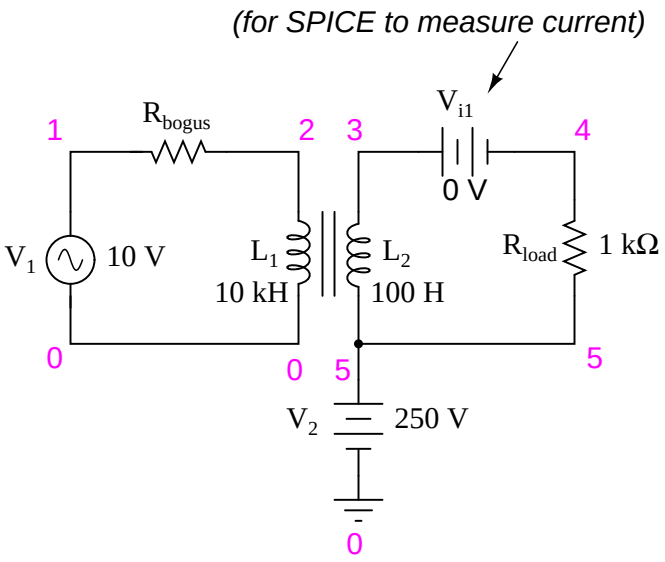
- A transformer designed to increase voltage from primary to secondary is called a *step-up* transformer. A transformer designed to reduce voltage from primary to secondary is called a *step-down* transformer.
- The transformation ratio of a transformer will be equal to the square root of its primary to secondary inductance (L) ratio.

$$\text{Voltage transformation ratio} = \sqrt{\frac{L_{secondary}}{L_{primary}}}$$

9.3 Electrical isolation

Aside from the ability to easily convert between different levels of voltage and current in AC and DC circuits, transformers also provide an extremely useful feature called *isolation*, which is the ability

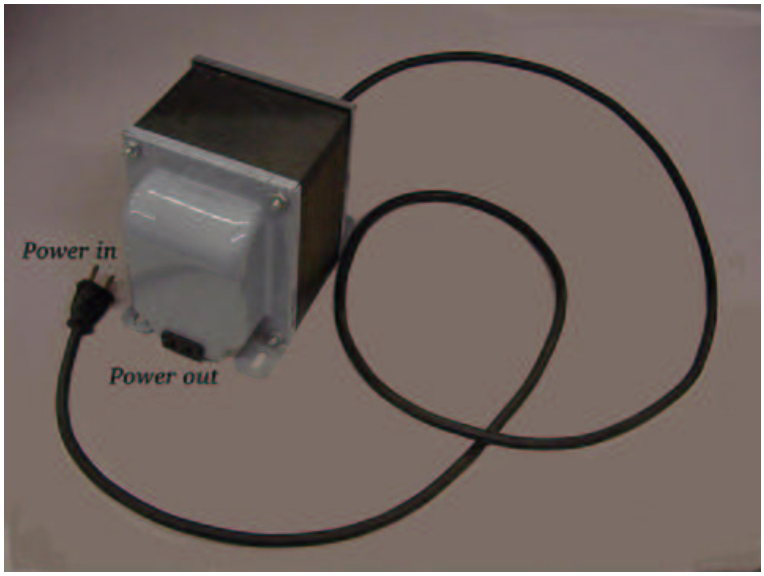
to couple one circuit to another without the use of direct wire connections. We can demonstrate an application of this effect with another SPICE simulation: this time showing "ground" connections for the two circuits, imposing a high DC voltage between one circuit and ground through the use of an additional voltage source:




```
v1 1 0 ac 10 sin
rbogus1 1 2 1e-12
v2 5 0 dc 250
l1 2 0 10000
l2 3 5 100
k l1 l2 0.999
vi1 3 4 ac 0
rload 4 5 1k
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

DC voltages referenced to ground (node 0):
(1)    0.0000    (2)    0.0000    (3)  250.0000
(4)  250.0000    (5)  250.0000

AC voltages:
freq          v(2)        i(v1)
6.000E+01     1.000E+01   9.975E-05    Primary winding

freq          v(3,5)      i(vi1)
6.000E+01     9.962E-01   9.962E-04    Secondary winding
```

SPICE shows the 250 volts DC being impressed upon the secondary circuit elements with respect to ground, but as you can see there is no effect on the primary circuit (zero DC voltage) at nodes 1 and 2, and the transformation of AC power from primary to secondary circuits remains the same as before. The impressed voltage in this example is often called a *common-mode* voltage because it is seen at more than one point in the circuit with reference to the common point of ground. The transformer isolates the common-mode voltage so that it is not impressed upon the primary circuit at all, but rather isolated to the secondary side. For the record, it does not matter that the common-mode voltage is DC, either. It could be AC, even at a different frequency, and the transformer would isolate it from the primary circuit all the same.

There are applications where electrical isolation is needed between two AC circuit without any transformation of voltage or current levels. In these instances, transformers called *isolation transformers* having 1:1 transformation ratios are used. A benchtop isolation transformer is shown in the following photograph:

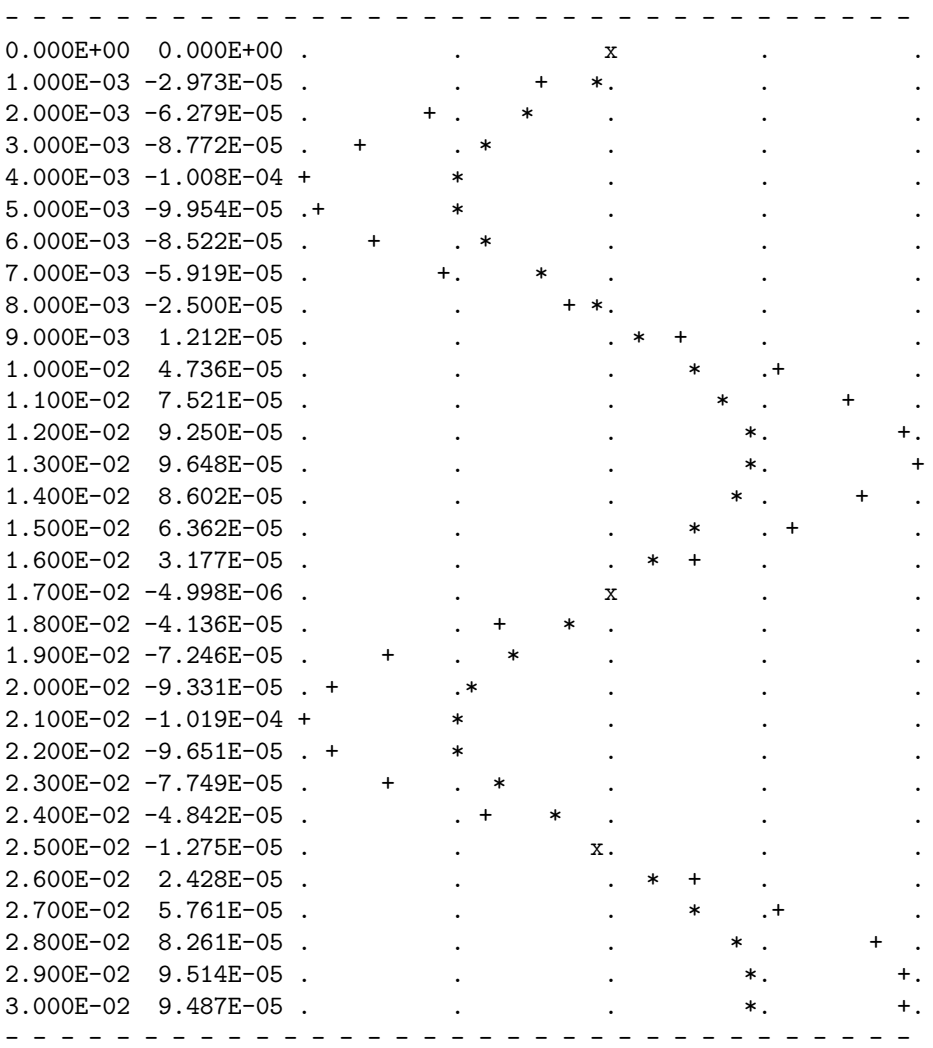


- **REVIEW:**
- By being able to transfer power from one circuit to another without the use of interconnecting conductors between the two circuits, transformers provide the useful feature of *electrical isolation*.
- Transformers designed to provide electrical isolation without stepping voltage and current either up or down are called *isolation transformers*.

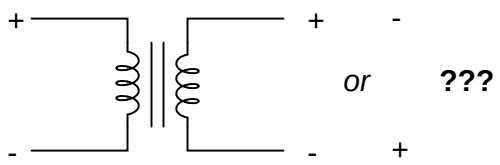
9.4 Phasing

Since transformers are essentially AC devices, we need to be aware of the phase relationships between the primary and secondary circuits. Using our SPICE example from before, we can plot the waveshapes for the primary and secondary circuits and see the phase relations for ourselves:

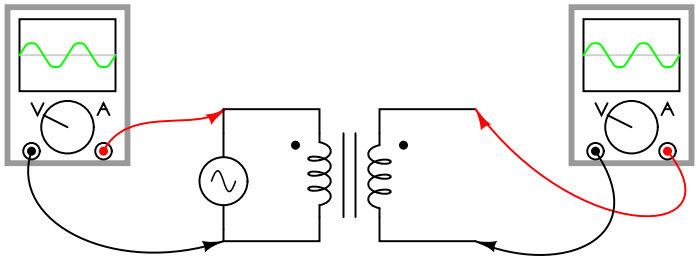
```
legend:
*: v(2)     Primary voltage
+: v(3,5)   Secondary voltage
time       v(2)
(*)-----------     -10         -5          0          5         10
(+)-----------     -10         -5          0          5         10
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .          .          x          .          .
1.000E-03  3.675E+00 .          .          .    + *   .          .
2.000E-03  6.803E+00 .          .          .          . + *      .
3.000E-03  9.008E+00 .          .          .          .      **  .
4.000E-03  9.955E+00 .          .          .          .          x
5.000E-03  9.450E+00 .          .          .          .        *+.
6.000E-03  7.672E+00 .          .          .          .    * +   .
7.000E-03  4.804E+00 .          .          .         *.+         .
8.000E-03  1.245E+00 .          .          .  *  +    .          .
9.000E-03 -2.474E+00 .          .     * +  .          .          .
1.000E-02 -5.864E+00 .         *+          .          .          .
1.100E-02 -8.390E+00 .    *+    .          .          .          .
1.200E-02 -9.779E+00 .x         .          .          .          .
1.300E-02 -9.798E+00 +*         .          .          .          .
1.400E-02 -8.390E+00 .   **     .          .          .          .
1.500E-02 -5.854E+00 .       + *.          .          .          .
1.600E-02 -2.479E+00 .          .    + *   .          .          .
1.700E-02  1.246E+00 .          .          .+ *       .          .
1.800E-02  4.795E+00 .          .          .       + *.          .
1.900E-02  7.686E+00 .          .          .          .   + *    .
2.000E-02  9.451E+00 .          .          .          .        x .
2.100E-02  9.937E+00 .          .          .          .          x
2.200E-02  9.025E+00 .          .          .          .       *+ .
2.300E-02  6.802E+00 .          .          .          .  *+      .
2.400E-02  3.667E+00 .          .          .      * + .          .
2.500E-02 -1.487E-03 .          .          * +        .          .
2.600E-02 -3.658E+00 .          .   * +    .          .          .
2.700E-02 -6.814E+00 .      * + .          .          .          .
2.800E-02 -9.026E+00 .  *+      .          .          .          .
2.900E-02 -9.917E+00 *+         .          .          .          .
3.000E-02 -9.511E+00 .x         .          .          .          .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

legend:
*: i(v1)   Primary current
+: i(vi1)  Secondary current
time      i(v1)
(*)---------- -2.000E-04   -1.000E-04      0   1.000E-04  2.000E-04
(+)---------- -1.000E-03   -5.000E-04      0   5.000E-04  1.000E-03
```



It would appear that both voltage and current for the two transformer windings are in phase with each other, at least for our resistive load. This is simple enough, but it would be nice to know *which way* we should connect a transformer in order to ensure the proper phase relationships be kept. After all, a transformer is nothing more than a set of magnetically-linked inductors, and inductors don't usually come with polarity markings of any kind. If we were to look at an unmarked transformer, we would have no way of knowing which way to hook it up to a circuit to get in-phase (or 180^o out-of-phase) voltage and current:



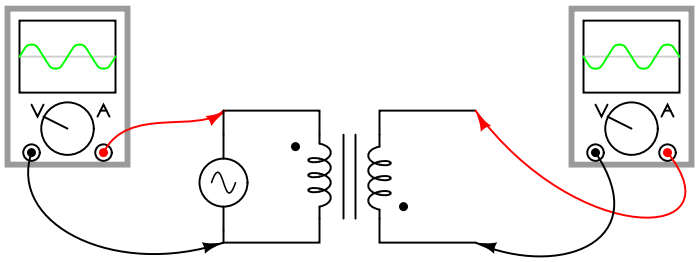
Since this is a practical concern, transformer manufacturers have come up with a sort of polarity marking standard to denote phase relationships. It is called the *dot convention*, and is nothing more than a dot placed next to each corresponding leg of a transformer winding:



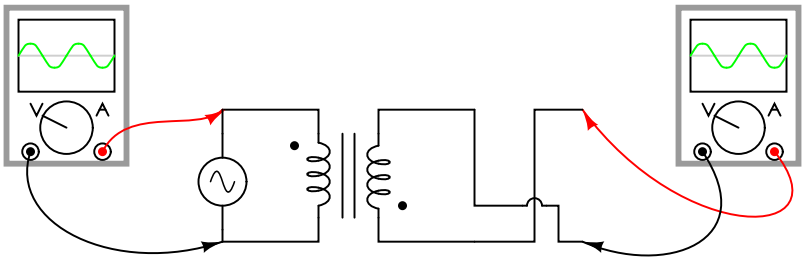
Typically, the transformer will come with some kind of schematic diagram labeling the wire leads for primary and secondary windings. On the diagram will be a pair of dots similar to what is seen above. Sometimes dots will be omitted, but when "H" and "X" labels are used to label transformer winding wires, the subscript numbers are supposed to represent winding polarity. The "1" wires (H_1 and X_1) represent where the polarity-marking dots would normally be placed.

The similar placement of these dots next to the top ends of the primary and secondary windings tells us that whatever instantaneous voltage polarity seen across the primary winding will be the same as that across the secondary winding. In other words, the phase shift from primary to secondary will be zero degrees.

On the other hand, if the dots on each winding of the transformer do *not* match up, the phase shift will be 180^o between primary and secondary, like this:



Of course, the dot convention only tells you which end of each winding is which, relative to the other winding(s). If you want to reverse the phase relationship yourself, all you have to do is swap the winding connections like this:



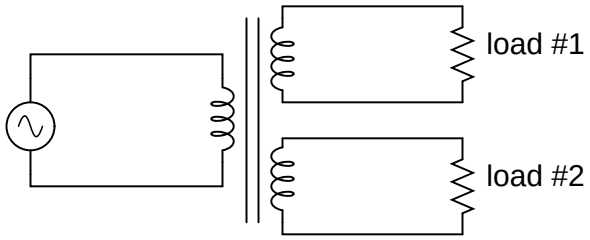
- **REVIEW:**

- The phase relationships for voltage and current between primary and secondary circuits of a transformer are direct: ideally, zero phase shift.

- The *dot convention* is a type of polarity marking for transformer windings showing which end of the winding is which, relative to the other windings.

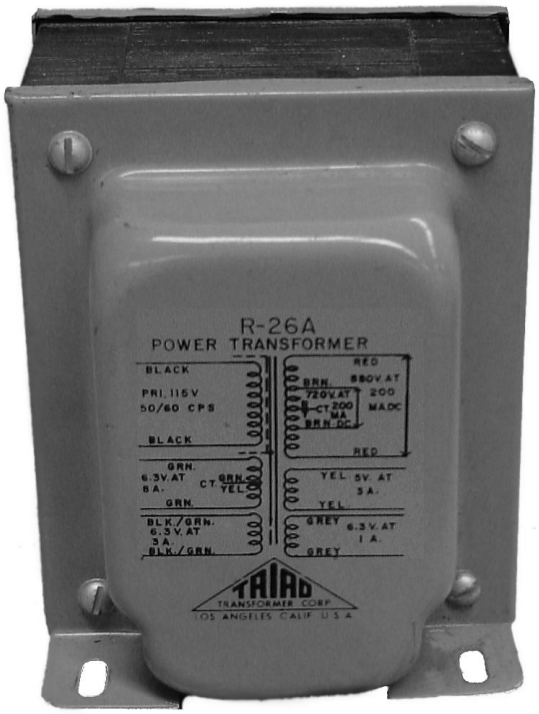
9.5 Winding configurations

Transformers are very versatile devices. The basic concept of energy transfer between mutual inductors is useful enough between a single primary and single secondary coil, but transformers don't have to be made with just two sets of windings. Consider this transformer circuit:



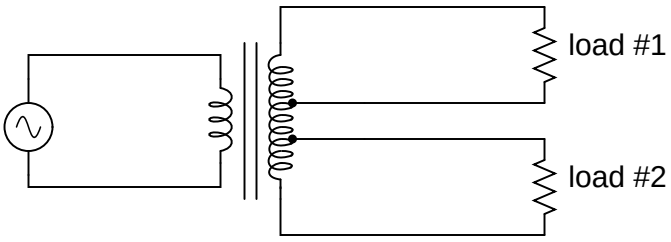
Here, three inductor coils share a common magnetic core, magnetically "coupling" or "linking" them together. The relationship of winding turn ratios and voltage ratios seen with a single pair of mutual inductors still holds true here for multiple pairs of coils. It is entirely possible to assemble a transformer such as the one above (one primary winding, two secondary windings) in which one secondary winding is a step-down and the other is a step-up. In fact, this design of transformer was quite common in vacuum tube power supply circuits, which were required to supply low voltage for the tubes' filaments (typically 6 or 12 volts) and high voltage for the tubes' plates (several hundred volts) from a nominal primary voltage of 110 volts AC. Not only are voltages and currents of completely different magnitudes possible with such a transformer, but all circuits are electrically isolated from one another.

A photograph of a multiple-winding transformer is shown here:

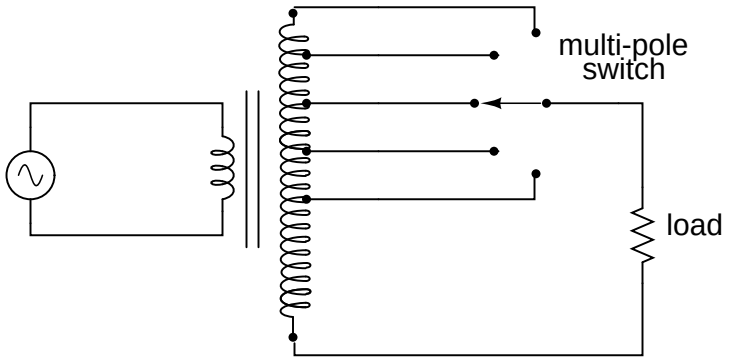


This particular transformer is intended to provide both high and low voltages necessary in an electronic system using vacuum tubes. Low voltage is required to power the filaments of vacuum tubes, while high voltage is required to create the potential difference between the plate and cathode elements of each tube. One transformer with multiple windings suffices elegantly to provide all the necessary voltage levels from a single 115 V source. The wires for this transformer (15 of them!) are not shown in the photograph, being hidden from view.

If electrical isolation between secondary circuits is not of great importance, a similar effect can be obtained by "tapping" a single secondary winding at multiple points along its length, like this:

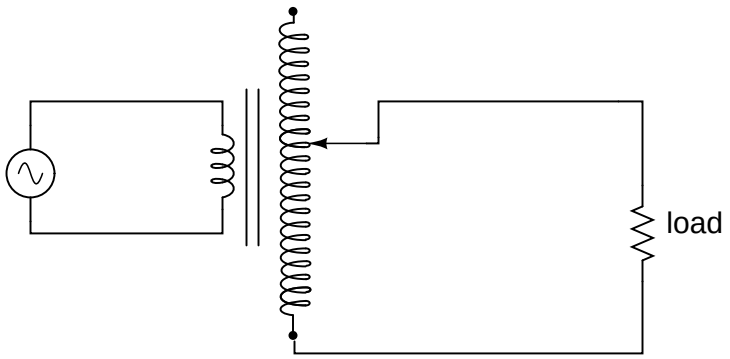


A tap is nothing more than a wire connection made at some point on a winding between the very ends. Not surprisingly, the winding turn/voltage magnitude relationship of a normal transformer holds true for all tapped segments of windings. This fact can be exploited to produce a transformer capable of multiple ratios:



Carrying the concept of winding taps further, we end up with a "variable transformer," where a sliding contact is moved along the length of an exposed secondary winding, able to connect with it at any point along its length. The effect is equivalent to having a winding tap at every turn of the winding, and a switch with poles at every tap position:

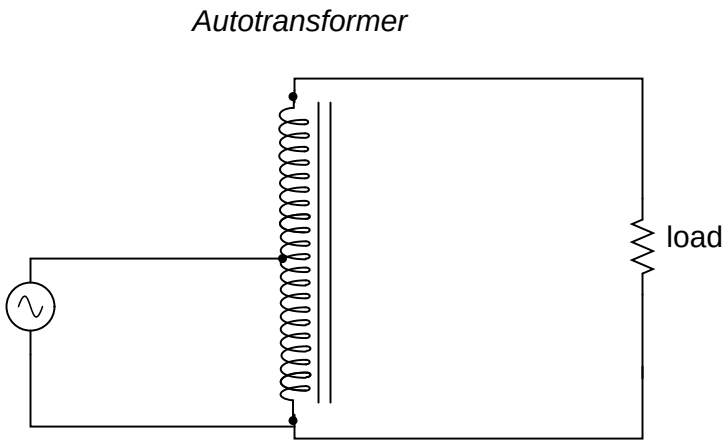
Variable transformer



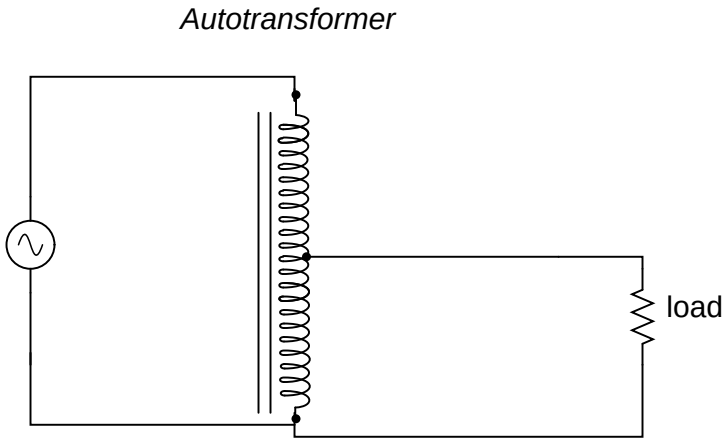
One consumer application of the variable transformer is in speed controls for model train sets, especially the train sets of the 1950's and 1960's. These transformers were essentially step-down units, the highest voltage obtainable from the secondary winding being substantially less than the primary voltage of 110 to 120 volts AC. The variable-sweep contact provided a simple means of voltage control with little wasted power, much more efficient than control using a variable resistor!

Moving-slide contacts are too impractical to be used in large industrial power transformer designs, but multi-pole switches and winding taps are common for voltage adjustment. Adjustments need to be made periodically in power systems to accommodate changes in loads over months or years in time, and these switching circuits provide a convenient means. Typically, such "tap switches" are not engineered to handle full-load current, but must be actuated only when the transformer has been de-energized (no power).

Seeing as how we can tap any transformer winding to obtain the equivalent of several windings (albeit with loss of electrical isolation between them), it makes sense that it should be possible to forego electrical isolation altogether and build a transformer from a single winding. Indeed this is possible, and the resulting device is called an *autotransformer*:

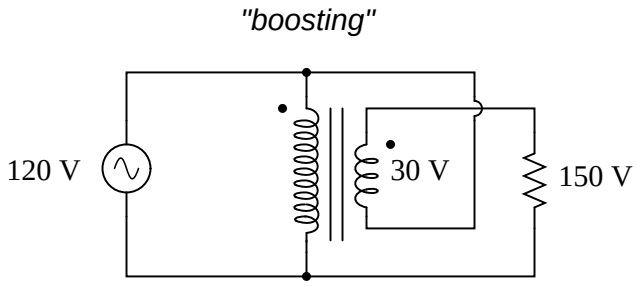


The autotransformer depicted above performs a voltage step-up function. A step-down autotransformer would look something like this:

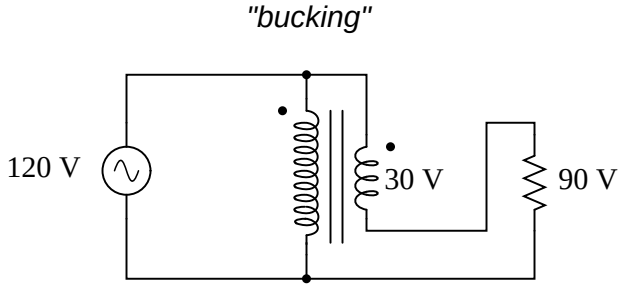


Autotransformers find popular use in applications requiring a slight boost or reduction in voltage to a load. The alternative with a normal (isolated) transformer would be to either have just the right primary/secondary winding ratio made for the job or use a step-down configuration with the secondary winding connected in series-aiding ("boosting") or series-opposing ("bucking") fashion. Primary, secondary, and load voltages are given to illustrate how this would work.

First, the "boosting" configuration. Here, the secondary coil's polarity is oriented so that its voltage directly adds to the primary voltage:

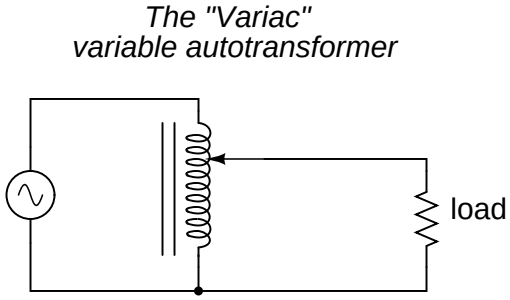


Next, the "bucking" configuration. Here, the secondary coil's polarity is oriented so that its voltage directly subtracts from the primary voltage:



The prime advantage of an autotransformer is that the same boosting or bucking function is obtained with only a single winding, making it cheaper and lighter to manufacture than a regular (isolating) transformer having both primary and secondary windings.

Like regular transformers, autotransformer windings can be tapped to provide variations in ratio. Additionally, they can be made continuously variable with a sliding contact to tap the winding at any point along its length. The latter configuration is popular enough to have earned itself its own name: the *Variac*.



Small variacs for benchtop use are popular pieces of equipment for the electronics experimenter, being able to step household AC voltage down (or sometimes up as well) with a wide, fine range of control by a simple twist of a knob.

- **REVIEW:**
- Transformers can be equipped with more than just a single primary and single secondary winding pair. This allows for multiple step-up and/or step-down ratios in the same device.

- Transformer windings can also be "tapped:" that is, intersected at many points to segment a single winding into sections.
- Variable transformers can be made by providing a movable arm that sweeps across the length of a winding, making contact with the winding at any point along its length. The winding, of course, has to be bare (no insulation) in the area where the arm sweeps.
- An autotransformer is a single, tapped inductor coil used to step up or step down voltage like a transformer, except without providing electrical isolation.
- A *Variac* is a variable autotransformer.

9.6 Voltage regulation

As we saw in a few SPICE analyses earlier in this chapter, the output voltage of a transformer varies some with varying load resistances, even with a constant voltage input. The degree of variance is affected by the primary and secondary winding inductances, among other factors, not the least of which includes winding resistance and the degree of mutual inductance (magnetic coupling) between the primary and secondary windings. For power transformer applications, where the transformer is seen by the load (ideally) as a constant source of voltage, it is good to have the secondary voltage vary as little as possible for wide variances in load current.

The measure of how well a power transformer maintains constant secondary voltage over a range of load currents is called the transformer's *voltage regulation*. It can be calculated from the following formula:

$$\text{Regulation percentage} = \frac{E_{\text{no-load}} - E_{\text{full-load}}}{E_{\text{full-load}}} (100\%)$$

"Full-load" means the point at which the transformer is operating at maximum permissible secondary current. This operating point will be determined primarily by the winding wire size (ampacity) and the method of transformer cooling. Taking our first SPICE transformer simulation as an example, let's compare the output voltage with a 1 kΩ load versus a 200 Ω load (assuming that the 200 Ω load will be our "full load" condition). Recall if you will that our constant primary voltage was 10.00 volts AC:

```
freq          v(3,5)      i(vi1)
6.000E+01     9.962E+00   9.962E-03    Output with 1k ohm load

freq          v(3,5)      i(vi1)
6.000E+01     9.348E+00   4.674E-02    Output with 200 ohm load
```

Notice how the output voltage decreases as the load gets heavier (more current). Now let's take that same transformer circuit and place a load resistance of extremely high magnitude across the secondary winding to simulate a "no-load" condition:

```
transformer
v1 1 0 ac 10 sin
```

```
rbogus1 1 2 1e-12
rbogus2 5 0 9e12
l1 2 0 100
l2 3 5 100
k l1 l2 0.999
vi1 3 4 ac 0
rload 4 5 9e12
.ac lin 1 60 60
.print ac v(2,0) i(v1)
.print ac v(3,5) i(vi1)
.end

freq          v(2)        i(v1)
6.000E+01     1.000E+01   2.653E-04

freq          v(3,5)      i(vi1)
6.000E+01     9.990E+00   1.110E-12   Output with (almost) no load
```

So, we see that our output (secondary) voltage spans a range of 9.990 volts at (virtually) no load and 9.348 volts at the point we decided to call "full load." Calculating voltage regulation with these figures, we get:

$$\text{Regulation percentage} = \frac{\text{9.990 V - 9.348 V}}{\text{9.990 V}} \text{ (100\%)}$$

$$\text{Regulation percentage} = 6.4264\ \%$$

Incidentally, this would be considered rather poor (or "loose") regulation for a power transformer. Powering a simple resistive load like this, a good power transformer should exhibit a regulation percentage of less than 3%. Inductive loads tend to create a condition of worse voltage regulation, so this analysis with purely resistive loads was a "best-case" condition.

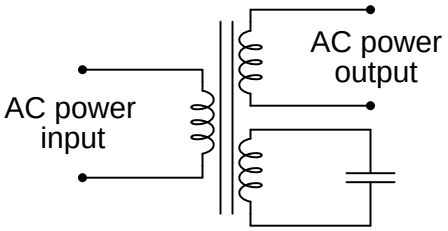
There are some applications, however, where poor regulation is actually desired. One such case is in discharge lighting, where a step-up transformer is required to initially generate a high voltage (necessary to "ignite" the lamps), then the voltage is expected to drop off once the lamp begins to draw current. This is because discharge lamps' voltage requirements tend to be much lower after a current has been established through the arc path. In this case, a step-up transformer with poor voltage regulation suffices nicely for the task of conditioning power to the lamp.

Another application is in current control for AC arc welders, which are nothing more than step-down transformers supplying low-voltage, high-current power for the welding process. A high voltage is desired to assist in "striking" the arc (getting it started), but like the discharge lamp, an arc doesn't require as much voltage to sustain itself once the air has been heated to the point of ionization. Thus, a decrease of secondary voltage under high load current would be a good thing. Some arc welder designs provide arc current adjustment by means of a movable iron core in the transformer, cranked in or out of the winding assembly by the operator. Moving the iron slug away from the windings reduces the strength of magnetic coupling between the windings, which diminishes no-load secondary voltage *and* makes for poorer voltage regulation.

No exposition on transformer regulation could be called complete without mention of an unusual device called a *ferroresonant transformer*. "Ferroresonance" is a phenomenon associated with the behavior of iron cores while operating near a point of magnetic saturation (where the core is so strongly magnetized that further increases in winding current results in little or no increase in magnetic flux).

While being somewhat difficult to describe without going deep into electromagnetic theory, the ferroresonant transformer is a power transformer engineered to operate in a condition of persistent core saturation. That is, its iron core is "stuffed full" of magnetic lines of flux for a large portion of the AC cycle so that variations in supply voltage (primary winding current) have little effect on the core's magnetic flux density, which means the secondary winding outputs a nearly constant voltage despite significant variations in supply (primary winding) voltage. Normally, core saturation in a transformer results in distortion of the sinewave shape, and the ferroresonant transformer is no exception. To combat this side effect, ferroresonant transformers have an auxiliary secondary winding paralleled with one or more capacitors, forming a resonant circuit tuned to the power supply frequency. This "tank circuit" serves as a filter to reject harmonics created by the core saturation, and provides the added benefit of storing energy in the form of AC oscillations, which is available for sustaining output winding voltage for brief periods of input voltage loss (milliseconds' worth of time, but certainly better than nothing).

Ferroresonant transformer



Resonant LC circuit

In addition to blocking harmonics created by the saturated core, this resonant circuit also "filters out" harmonic frequencies generated by nonlinear (switching) loads in the secondary winding circuit and any harmonics present in the source voltage, providing "clean" power to the load.

Ferroresonant transformers offer several features useful in AC power conditioning: constant output voltage given substantial variations in input voltage, harmonic filtering between the power source and the load, and the ability to "ride through" brief losses in power by keeping a reserve of energy in its resonant tank circuit. These transformers are also highly tolerant of excessive loading and transient (momentary) voltage surges. They are so tolerant, in fact, that some may be briefly paralleled with unsynchronized AC power sources, allowing a load to be switched from one source of power to another in a "make-before-break" fashion with no interruption of power on the secondary side!

Unfortunately, these devices have equally noteworthy disadvantages: they waste a lot of energy (due to hysteresis losses in the saturated core), generating *significant* heat in the process, and are intolerant of frequency variations, which means they don't work very well when powered by small engine-driven generators having poor speed regulation. Voltages produced in the resonant winding/capacitor circuit tend to be very high, necessitating expensive capacitors and presenting the service technician with very dangerous working voltages. Some applications, though, may prioritize

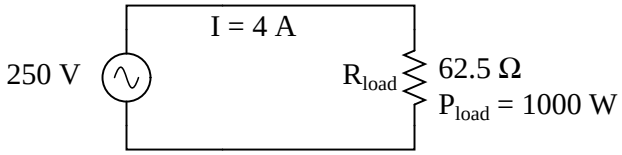
the ferroresonant transformer's advantages over its disadvantages. Semiconductor circuits exist to "condition" AC power as an alternative to ferroresonant devices, but none can compete with this transformer in terms of sheer simplicity.

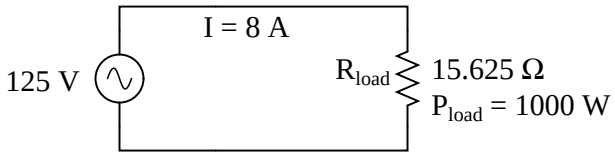
- **REVIEW:**
- *Voltage regulation* is the measure of how well a power transformer can maintain constant secondary voltage given a constant primary voltage and wide variance in load current. The lower the percentage (closer to zero), the more stable the secondary voltage and the better the regulation it will provide.
- A *ferroresonant* transformer is a special transformer designed to regulate voltage at a stable level despite wide variation in input voltage.

9.7 Special transformers and applications

Because transformers can step voltage and current to different levels, and because power is transferred equivalently between primary and secondary windings, they can be used to "convert" the impedance of a load to a different level. That last phrase deserves some explanation, so let's investigate what it means.

The purpose of a load (usually) is to do something productive with the power it dissipates. In the case of a resistive heating element, the practical purpose for the power dissipated is to heat something up. Loads are engineered to safely dissipate a certain maximum amount of power, but two loads of equal power rating are not necessarily identical. Consider these two 1000 watt resistive heating elements:

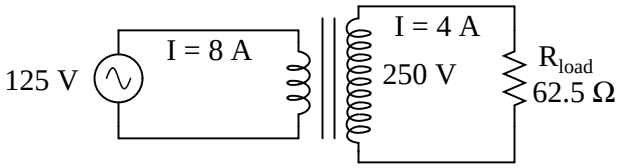




Both heaters dissipate exactly 1000 watts of power, but they do so at different voltage and current levels (either 250 volts and 4 amps, or 125 volts and 8 amps). Using Ohm's Law to determine the necessary resistance of these heating elements (R=E/I), we arrive at figures of 62.5 Ω and 15.625 Ω, respectively. If these are AC loads, we might refer to their opposition to current in terms of impedance rather than plain resistance, although in this case that's all they're composed of (no reactance). The 250 volt heater would be said to be a higher impedance load than the 125 volt heater.

If we desired to operate the 250 volt heater element directly on a 125 volt power system, we would end up being disappointed. With 62.5 Ω of impedance (resistance), the current would only be 2 amps (I=E/R; 125/62.5), and the power dissipation would only be 250 watts (P=IE; 125 x 2), or one-quarter of its rated power. The impedance of the heater and the voltage of our source would be mismatched, and we couldn't obtain the full rated power dissipation from the heater.

All hope is not lost, though. With a step-up transformer, we could operate the 250 volt heater element on the 125 volt power system like this:



1000 watts dissipation at the load resistor !

The ratio of the transformer's windings provides the voltage step-up *and* current step-down we need for the otherwise mismatched load to operate properly on this system. Take a close look at the primary circuit figures: 125 volts at 8 amps. As far as the power supply "knows," it's powering a 15.625 Ω (R=E/I) load at 125 volts, not a 62.5 Ω load! The voltage and current figures for the primary winding are indicative of 15.625 Ω load impedance, not the actual 62.5 Ω of the load itself. In other words, not only has our step-up transformer transformed voltage and current, but it has transformed *impedance* as well.

The transformation ratio of impedance is the square of the voltage/current transformation ratio, the same as the winding inductance ratio:

$$\text{Voltage transformation ratio} = \frac{N_{secondary}}{N_{primary}}$$

$$\text{Current transformation ratio} = \frac{N_{primary}}{N_{secondary}}$$

$$\text{Impedance transformation ratio} = \left(\frac{N_{secondary}}{N_{primary}} \right)^2$$

$$\text{Inductance ratio} = \left(\frac{N_{secondary}}{N_{primary}} \right)^2$$

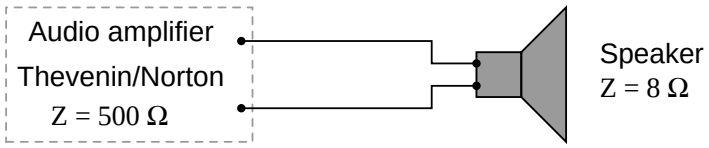
Where,

N = number of turns in winding

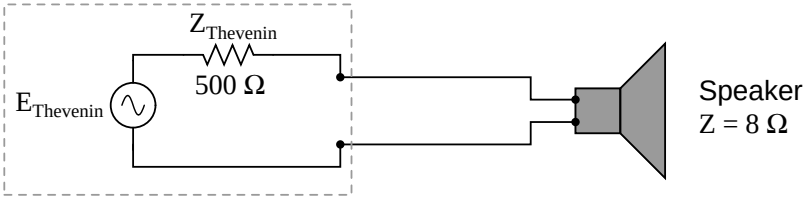
This concurs with our example of the 2:1 step-up transformer and the impedance ratio of 62.5 Ω to 15.625 Ω (a 4:1 ratio, which is 2:1 squared). Impedance transformation is a highly useful ability of transformers, for it allows a load to dissipate its full rated power even if the power system is not

at the proper voltage to directly do so.

Recall from our study of network analysis the *Maximum Power Transfer Theorem*, which states that the maximum amount of power will be dissipated by a load resistance when that load resistance is equal to the Thevenin/Norton resistance of the network supplying the power. Substitute the word "impedance" for "resistance" in that definition and you have the AC version of that Theorem. If we're trying to obtain theoretical maximum power dissipation from a load, we must be able to properly match the load impedance and source (Thevenin/Norton) impedance together. This is generally more of a concern in specialized electric circuits such as radio transmitter/antenna and audio amplifier/speaker systems. Let's take an audio amplifier system and see how it works:

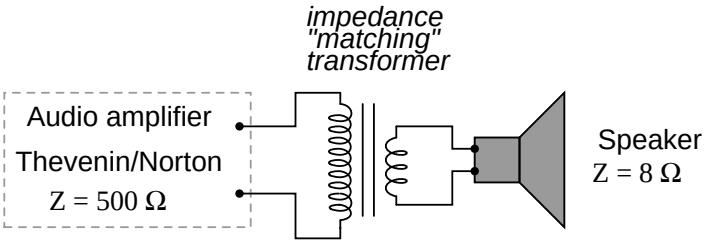


. . . equivalent to . . .



With an internal impedance of 500 Ω, the amplifier can only deliver full power to a load (speaker) also having 500 Ω of impedance. Such a load would drop higher voltage and draw less current than an 8 Ω speaker dissipating the same amount of power. If an 8 Ω speaker were connected directly to the 500 Ω amplifier as shown, the *impedance mismatch* would result in very poor (low peak power) performance. Additionally, the amplifier would tend to dissipate more than its fair share of power in the form of heat trying to drive the low impedance speaker.

To make this system work better, we can use a transformer to match these mismatched impedances. Since we're going from a high impedance (high voltage, low current) supply to a low impedance (low voltage, high current) load, we'll need to use a step-down transformer:



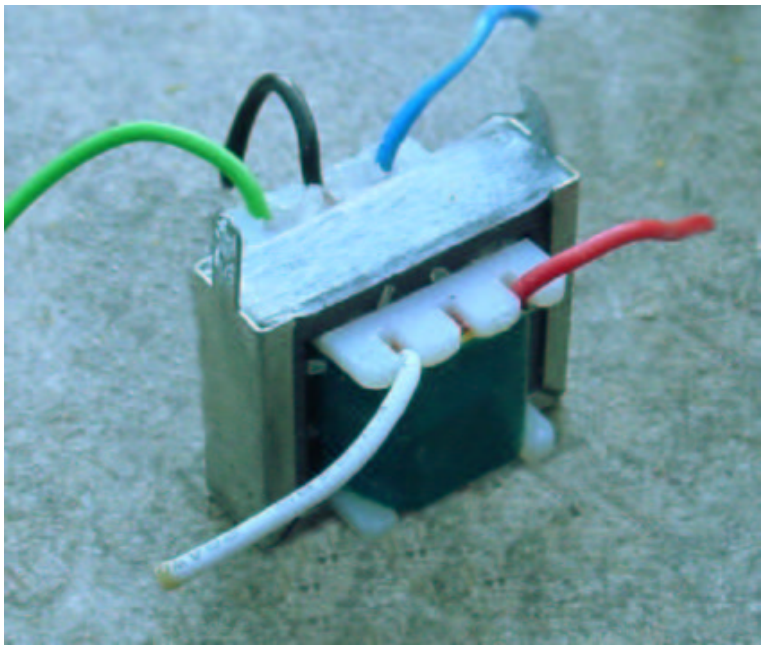
impedance ratio = 500 : 8
winding ratio = 7.906 : 1

To obtain an impedance transformation ratio of 500:8, we would need a winding ratio equal to the square root of 500:8 (the square root of 62.5:1, or 7.906:1). With such a transformer in place, the speaker will load the amplifier to just the right degree, drawing power at the correct voltage and current levels to satisfy the Maximum Power Transfer Theorem and make for the most efficient power delivery to the load. The use of a transformer in this capacity is called *impedance matching*.

Anyone who has ridden a multi-speed bicycle can intuitively understand the principle of impedance matching. A human's legs will produce maximum power when spinning the bicycle crank at a particular speed (about 60 to 90 revolution per minute). Above or below that rotational speed, human leg muscles are less efficient at generating power. The purpose of the bicycle's "gears" is to impedance-match the rider's legs to the riding conditions so that they always spin the crank at the optimum speed.

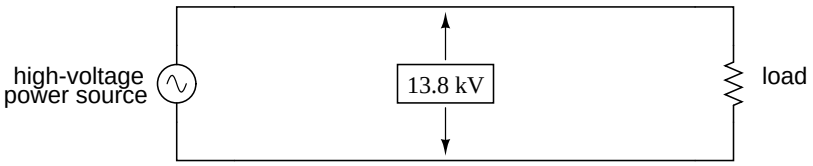
If the rider attempts to start moving while the bicycle is shifted into its "top" gear, he or she will find it very difficult to get moving. Is it because the rider is weak? No, it's because the high step-up ratio of the bicycle's chain and sprockets in that top gear presents a mismatch between the conditions (lots of inertia to overcome) and their legs (needing to spin at 60-90 RPM for maximum power output). On the other hand, selecting a gear that is too low will enable the rider to get moving immediately, but limit the top speed they will be able to attain. Again, is the lack of speed an indication of weakness in the bicyclist's legs? No, it's because the lower speed ratio of the selected gear creates another type of mismatch between the conditions (low load) and the rider's legs (losing power if spinning faster than 90 RPM). It is much the same with electric power sources and loads: there must be an impedance match for maximum system efficiency. In AC circuits, transformers perform the same matching function as the sprockets and chain ("gears") on a bicycle to match otherwise mismatched sources and loads.

Impedance matching transformers are not fundamentally different from any other type of transformer in construction or appearance. A small impedance-matching transformer (about two centimeters in width) for audio-frequency applications is shown in the following photograph:

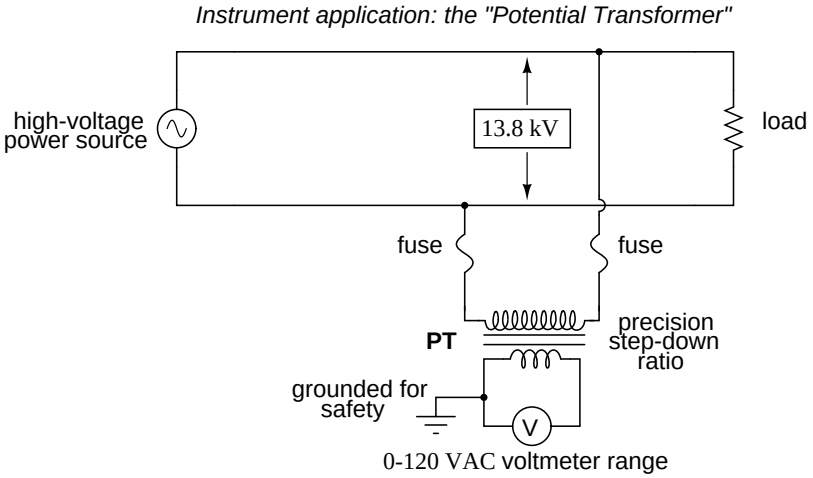


Another impedance-matching transformer can be seen on this printed circuit board, in the upper right corner, to the immediate left of resistors R_2 and R_1. It is labeled "T1":

Transformers can also be used in electrical instrumentation systems. Due to transformers' ability to step up or step down voltage and current, and the electrical isolation they provide, they can serve as a way of connecting electrical instrumentation to high-voltage, high current power systems. Suppose we wanted to accurately measure the voltage of a 13.8 kV power system (a very common power distribution voltage in American industry):



Designing, installing, and maintaining a voltmeter capable of directly measuring 13,800 volts AC would be no easy task. The safety hazard alone of bringing 13.8 kV conductors into an instrument panel would be severe, not to mention the design of the voltmeter itself. However, by using a precision step-down transformer, we can reduce the 13.8 kV down to a safe level of voltage at a constant ratio, and isolate it from the instrument connections, adding an additional level of safety to the metering system:

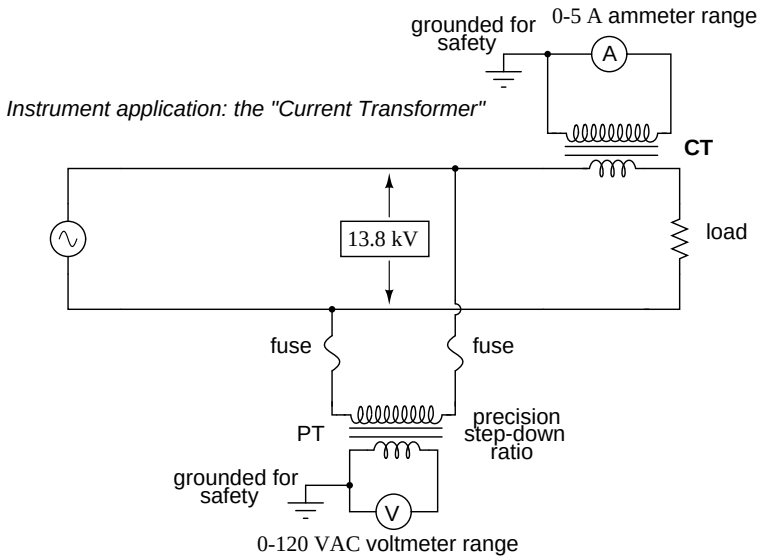


Now the voltmeter reads a precise fraction, or ratio, of the actual system voltage, its scale set to read as though it were measuring the voltage directly. The transformer keeps the instrument voltage at a safe level and electrically isolates it from the power system, so there is no direct connection between the power lines and the instrument or instrument wiring. When used in this capacity, the transformer is called a *Potential Transformer*, or simply *PT*.

Potential transformers are designed to provide as accurate a voltage step-down ratio as possible. To aid in precise voltage regulation, loading is kept to a minimum: the voltmeter is made to have high input impedance so as to draw as little current from the PT as possible. As you can see, a fuse has been connected in series with the PTs primary winding, for safety and ease of disconnecting the PT from the circuit.

A standard secondary voltage for a PT is 120 volts AC, for full-rated power line voltage. The standard voltmeter range to accompany a PT is 150 volts, full-scale. PTs with custom winding ratios can be manufactured to suit any application. This lends itself well to industry standardization of the actual voltmeter instruments themselves, since the PT will be sized to step the system voltage down to this standard instrument level.

Following the same line of thinking, we can use a transformer to step down current through a power line so that we are able to safely and easily measure high system currents with inexpensive ammeters. Of course, such a transformer would be connected in series with the power line, like this:



Note that while the PT is a step-down device, the *Current Transformer* (or *CT*) is a step-up device (with respect to voltage), which is what is needed to step *down* the power line current. Quite often, CTs are built as donut-shaped devices through which the power line conductor is run, the power line itself acting as a single-turn primary winding:

Some CTs are made to hinge open, allowing insertion around a power conductor without disturbing the conductor at all. The industry standard secondary current for a CT is a range of 0 to 5 amps AC. Like PTs, CTs can be made with custom winding ratios to fit almost any application. Because their "full load" secondary current is 5 amps, CT ratios are usually described in terms of full-load primary amps to 5 amps, like this:

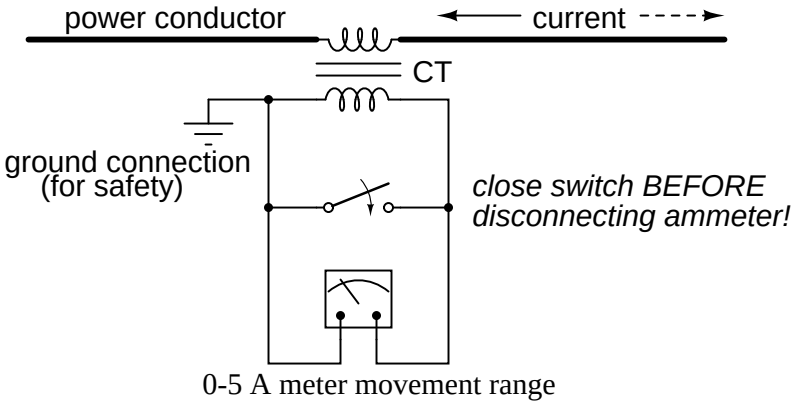
600 : 5 ratio *(for measuring up to 600 A line current)*

100 : 5 ratio *(for measuring up to 100 A line current)*

1k : 5 ratio *(for measuring up to 1000 A line current)*

The "donut" CT shown in the photograph has a ratio of 50:5. That is, when the conductor through the center of the torus is carrying 50 amps of current (AC), there will be 5 amps of current induced in the CT's winding.

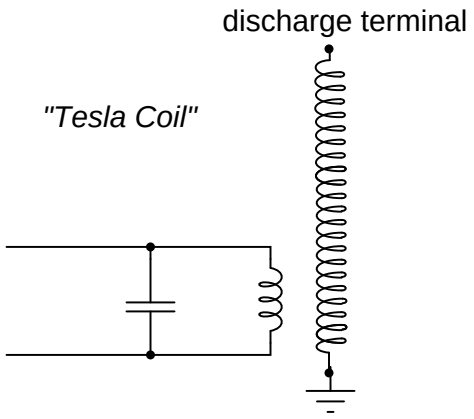
Because CTs are designed to be powering ammeters, which are low-impedance loads, and they are wound as voltage step-up transformers, they should never, *ever* be operated with an open-circuited secondary winding. Failure to heed this warning will result in the CT producing extremely high secondary voltages, dangerous to equipment and personnel alike. To facilitate maintenance of ammeter instrumentation, short-circuiting switches are often installed in parallel with the CT's secondary winding, to be closed whenever the ammeter is removed for service:



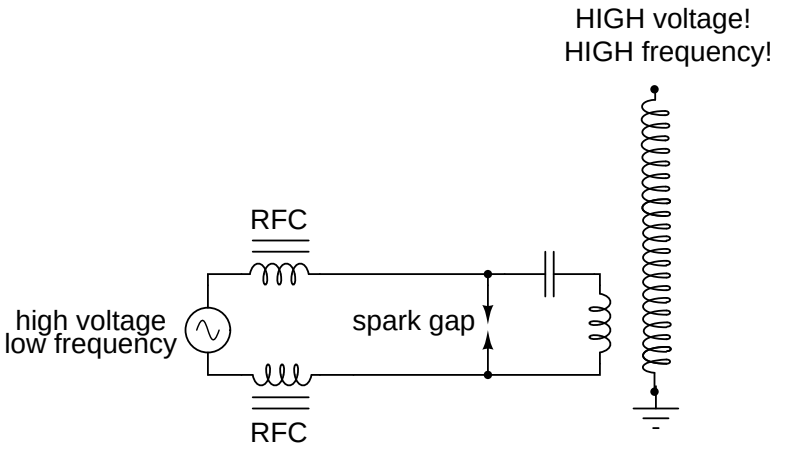
Though it may seem strange to *intentionally* short-circuit a power system component, it is perfectly proper and quite necessary when working with current transformers.

Another kind of special transformer, seen often in radio-frequency circuits, is the *air core* transformer. True to its name, an air core transformer has its windings wrapped around a nonmagnetic form, usually a hollow tube of some material. The degree of coupling (mutual inductance) between windings in such a transformer is many times less than that of an equivalent iron-core transformer, but the undesirable characteristics of a ferromagnetic core (eddy current losses, hysteresis, saturation, etc.) are completely eliminated. It is in high-frequency applications that these effects of iron cores are most problematic.

One notable example of air-core transformer is the *Tesla Coil*, named after the Serbian electrical genius Nikola Tesla, who was also the inventor of the rotating magnetic field AC motor, polyphase AC power systems, and many elements of radio technology. The Tesla Coil is a resonant, high-frequency step-up transformer used to produce extremely high voltages. One of Tesla's dreams was to employ his coil technology to distribute electric power without the need for wires, simply broadcasting it in the form of radio waves which could be received and conducted to loads by means of antennas. The basic schematic for a Tesla Coil looks like this:



The capacitor, in conjunction with the transformer's primary winding, forms a tank circuit. The secondary winding is wound in close proximity to the primary, usually around the same nonmagnetic form. Several options exist for "exciting" the primary circuit, the simplest being a high-voltage, low-frequency AC source and spark gap:



The purpose of the high-voltage, low-frequency AC power source is to "charge" the primary tank circuit. When the spark gap fires, its low impedance acts to complete the capacitor/primary coil tank circuit, allowing it to oscillate at its resonant frequency. The "RFC" inductors are "Radio Frequency Chokes," which act as high impedances to prevent the AC source from interfering with the oscillating tank circuit.

The secondary side of the Tesla coil transformer is also a tank circuit, relying on the parasitic (stray) capacitance existing between the discharge terminal and earth ground to complement the secondary winding's inductance. For optimum operation, this secondary tank circuit is tuned to the same resonant frequency as the primary circuit, with energy exchanged not only between capacitors and inductors during resonant oscillation, but also back-and-forth between primary and secondary windings. The visual results are spectacular:

Tesla Coils find application primarily as novelty devices, showing up in high school science fairs, basement workshops, and the occasional low budget science-fiction movie.

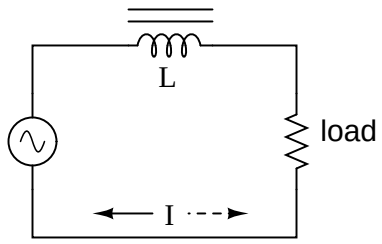
It should be noted that Tesla coils can be extremely dangerous devices. Burns caused by radio-frequency ("RF") current, like all electrical burns, can be very deep, unlike skin burns caused by contact with hot objects or flames. Although the high-frequency discharge of a Tesla coil has the curious property of being beyond the "shock perception" frequency of the human nervous system, this does not mean Tesla coils cannot hurt or even kill you! I strongly advise seeking the assistance of an experienced Tesla coil experimenter if you would embark on building one yourself.

So far, we've explored the transformer as a device for converting different levels of voltage, current, and even impedance from one circuit to another. Now we'll take a look at it as a completely different kind of device: one that allows a small electrical signal to exert *control* over a much larger quantity of electrical power. In this mode, a transformer acts as an *amplifier*.

The device I'm referring to is called a *saturable-core reactor*, or simply *saturable reactor*. Actually, it is not really a transformer at all, but rather a special kind of inductor whose inductance can be varied by the application of a DC current through a second winding wound around the same iron core. Like the ferroresonant transformer, the saturable reactor relies on the principle of magnetic saturation. When a material such as iron is completely saturated (that is, all its magnetic domains are lined up with the applied magnetizing force), additional increases in current through the magnetizing winding will not result in further increases of magnetic flux.

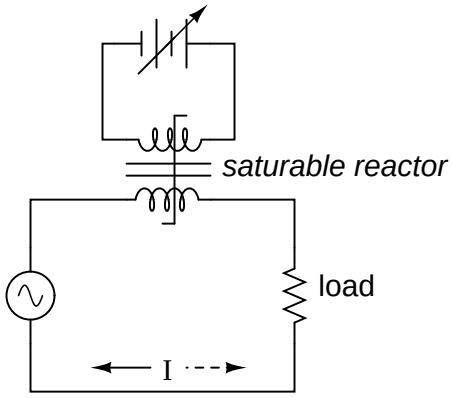
Now, inductance is the measure of how well an inductor opposes changes in current by developing a voltage in an opposing direction. The ability of an inductor to generate this opposing voltage is directly connected with the change in magnetic flux inside the inductor resulting from the change in current, and the number of winding turns in the inductor. If an inductor has a saturated core, no further magnetic flux will result from further increases in current, and so there will be no voltage induced in opposition to the change in current. In other words, an inductor loses its inductance (ability to oppose changes in current) when its core becomes magnetically saturated.

If an inductor's inductance changes, then its reactance (and impedance) to AC current changes as well. In a circuit with a constant voltage source, this will result in a change in current:



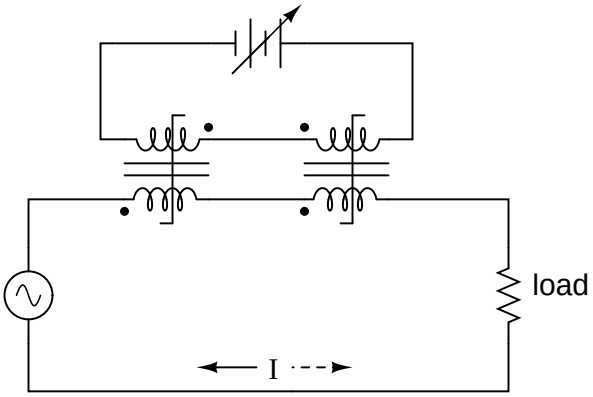
If L changes in inductance, Z_L will correspondingly change, thus changing the circuit current.

A saturable reactor capitalizes on this effect by forcing the core into a state of saturation with a strong magnetic field generated by current through another winding. The reactor's "power" winding is the one carrying the AC load current, and the "control" winding is one carrying a DC current strong enough to drive the core into saturation:



The strange-looking transformer symbol shown in the above schematic represents a saturable-core reactor, the upper winding being the DC control winding and the lower being the "power" winding through which the controlled AC current goes. Increased DC control current produces more magnetic flux in the reactor core, driving it closer to a condition of saturation, thus decreasing the power winding's inductance, decreasing its impedance, and increasing current to the load. Thus, the DC control current is able to exert *control* over the AC current delivered to the load.

The circuit shown would work, but it would not work very well. The first problem is the natural transformer action of the saturable reactor: AC current through the power winding will induce a voltage in the control winding, which may cause trouble for the DC power source. Also, saturable reactors tend to regulate AC power only in one direction: in one half of the AC cycle, the mmf's from both windings add; in the other half, they subtract. Thus, the core will have more flux in it during one half of the AC cycle than the other, and will saturate first in that cycle half, passing load current more easily in one direction than the other. Fortunately, both problems can be overcome with a little ingenuity:



Notice the placement of the phasing dots on the two reactors: the power windings are "in phase" while the control windings are "out of phase." If both reactors are identical, any voltage induced in the control windings by load current through the power windings will cancel out to zero at the battery terminals, thus eliminating the first problem mentioned. Furthermore, since the DC control current through both reactors produces magnetic fluxes in different directions through the reactor cores, one reactor will saturate more in one cycle of the AC power while the other reactor will saturate more in the other, thus equalizing the control action through each half-cycle so that the AC power is "throttled" symmetrically. This phasing of control windings can be accomplished with two separate reactors as shown, or in a single reactor design with intelligent layout of the windings and core.

Saturable reactor technology has even been miniaturized to the circuit-board level in compact packages more generally known as *magnetic amplifiers*. I personally find this to be fascinating: the effect of amplification (one electrical signal controlling another), normally requiring the use of physically fragile vacuum tubes or electrically "fragile" semiconductor devices, can be realized in a device both physically and electrically rugged. Magnetic amplifiers do have disadvantages over their more fragile counterparts, namely size, weight, nonlinearity, and bandwidth (frequency response), but their utter simplicity still commands a certain degree of appreciation, if not practical application.

Saturable-core reactors are less commonly known as "saturable-core inductors" or *transductors*.

- **REVIEW:**
- Transformers can be used to transform impedance as well as voltage and current. When this is done to improve power transfer to a load, it is called *impedance matching*.
- A *Potential Transformer* (PT) is a special instrument transformer designed to provide a precise voltage step-down ratio for voltmeters measuring high power system voltages.

- A *Current Transformer* (CT) is another special instrument transformer designed to step down the current through a power line to a safe level for an ammeter to measure.

- An *air-core* transformer is one lacking a ferromagnetic core.

- A *Tesla Coil* is a resonant, air-core, step-up transformer designed to produce very high AC voltages at high frequency.

- A *saturable reactor* is a special type of inductor, the inductance of which can be controlled by the DC current through a second winding around the same core. With enough DC current, the magnetic core can be saturated, decreasing the inductance of the power winding in a controlled fashion.

9.8 Practical considerations

9.8.1 Power capacity

As has already been observed, transformers must be well designed in order to achieve acceptable power coupling, tight voltage regulation, and low exciting current distortion. Also, transformers must be designed to carry the expected values of primary and secondary winding current without any trouble. This means the winding conductors must be made of the proper gauge wire to avoid any heating problems. An ideal transformer would have perfect coupling (no leakage inductance), perfect voltage regulation, perfectly sinusoidal exciting current, no hysteresis or eddy current losses, and wire thick enough to handle any amount of current. Unfortunately, the ideal transformer would have to be infinitely large and heavy to meet these design goals. Thus, in the business of *practical* transformer design, compromises must be made.

Additionally, winding conductor insulation is a concern where high voltages are encountered, as they often are in step-up and step-down power distribution transformers. Not only do the windings have to be well insulated from the iron core, but each winding has to be sufficiently insulated from the other in order to maintain electrical isolation between windings.

Respecting these limitations, transformers are rated for certain levels of primary and secondary winding voltage and current, though the current rating is usually derived from a volt-amp (VA) rating assigned to the transformer. For example, take a step-down transformer with a primary voltage rating of 120 volts, a secondary voltage rating of 48 volts, and a VA rating of 1 kVA (1000 VA). The maximum winding currents can be determined as such:

$$\frac{1000 \text{ VA}}{120 \text{ V}} = 8.333 \text{ A} \quad \textit{(maximum primary winding current)}$$

$$\frac{1000 \text{ VA}}{48 \text{ V}} = 20.833 \text{ A} \quad \textit{(maximum secondary winding current)}$$

Sometimes windings will bear current ratings in amps, but this is typically seen on small transformers. Large transformers are almost always rated in terms of winding voltage and VA or kVA.

9.8.2 Energy losses

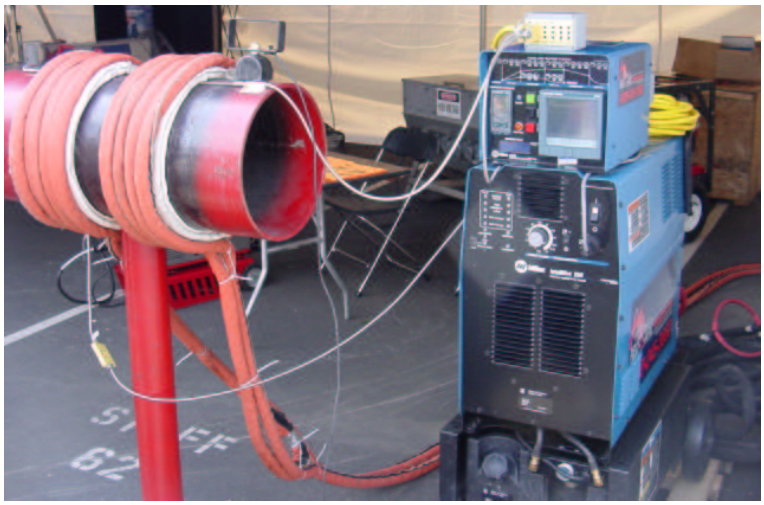
When transformers transfer power, they do so with a minimum of loss. As it was stated earlier, modern power transformer designs typically exceed 95% efficiency. It is good to know where some of this lost power goes, however, and what causes it to be lost.

There is, of course, power lost due to resistance of the wire windings. Unless superconducting wires are used, there will always be power dissipated in the form of heat through the resistance of current-carrying conductors. Because transformers require such long lengths of wire, this loss can be a significant factor. Increasing the gauge of the winding wire is one way to minimize this loss, but only with substantial increases in cost, size, and weight.

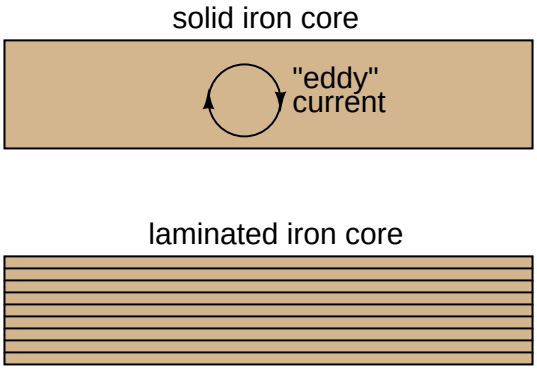
Resistive losses aside, the bulk of transformer power loss is due to magnetic effects in the core. Perhaps the most significant of these "core losses" is *eddy-current loss*, which is resistive power dissipation due to the passage of induced currents through the iron of the core. Because iron is a conductor of electricity as well as being an excellent "conductor" of magnetic flux, there will be currents induced in the iron just as there are currents induced in the secondary windings from the alternating magnetic field. These induced currents – as described by the perpendicularity clause of Faraday's Law – tend to circulate through the cross-section of the core perpendicularly to the primary winding turns. Their circular motion gives them their unusual name: like eddies in a stream of water that circulate rather than move in straight lines.

Iron is a fair conductor of electricity, but not as good as the copper or aluminum from which wire windings are typically made. Consequently, these "eddy currents" must overcome significant electrical resistance as they circulate through the core. In overcoming the resistance offered by the iron, they dissipate power in the form of heat. Hence, we have a source of inefficiency in the transformer that is difficult to eliminate.

This phenomenon is so pronounced that it is often exploited as a means of heating ferrous (iron-containing) materials. The following photograph shows an "induction heating" unit raising the temperature of a large pipe section. Loops of wire covered by high-temperature insulation encircle the pipe's circumference, inducing eddy currents within the pipe wall by electromagnetic induction. In order to maximize the eddy current effect, high-frequency alternating current is used rather than power line frequency (60 Hz). The box units at the right of the picture produce the high-frequency AC and control the amount of current in the wires to stabilize the pipe temperature at a pre-determined "set-point."



The main strategy in mitigating these wasteful eddy currents in transformer cores is to form the iron core in sheets, each sheet covered with an insulating varnish so that the core is divided up into thin slices. The result is very little width in the core for eddy currents to circulate in:



Laminated cores like the one shown here are standard in almost all low-frequency transformers. Recall from the photograph of the transformer cut in half that the iron core was composed of many thin sheets rather than one solid piece. Eddy current losses increase with frequency, so transformers designed to run on higher-frequency power (such as 400 Hz, used in many military and aircraft applications) must use thinner laminations to keep the losses down to a respectable minimum. This has the undesirable effect of increasing the manufacturing cost of the transformer.

Another, similar technique for minimizing eddy current losses which works better for high-frequency applications is to make the core out of iron powder instead of thin iron sheets. Like the lamination sheets, these granules of iron are individually coated in an electrically insulating material, which makes the core nonconductive except for within the width of each granule. Powdered iron cores are often found in transformers handling radio-frequency currents.

Another "core loss" is that of magnetic *hysteresis*. All ferromagnetic materials tend to retain some degree of magnetization after exposure to an external magnetic field. This tendency to stay magnetized is called "hysteresis," and it takes a certain investment in energy to overcome this

opposition to change every time the magnetic field produced by the primary winding changes polarity (twice per AC cycle). This type of loss can be mitigated through good core material selection (choosing a core alloy with low hysteresis, as evidenced by a "thin" B/H hysteresis curve), and designing the core for minimum flux density (large cross-sectional area).

Transformer energy losses tend to worsen with increasing frequency. The skin effect within winding conductors reduces the available cross-sectional area for electron flow, thereby increasing effective resistance as the frequency goes up and creating more power lost through resistive dissipation. Magnetic core losses are also exaggerated with higher frequencies, eddy currents and hysteresis effects becoming more severe. For this reason, transformers of significant size are designed to operate efficiently in a limited range of frequencies. In most power distribution systems where the line frequency is very stable, one would think excessive frequency would never pose a problem. Unfortunately it does, in the form of harmonics created by nonlinear loads.

As we've seen in earlier chapters, nonsinusoidal waveforms are equivalent to additive series of multiple sinusoidal waveforms at different amplitudes and frequencies. In power systems, these other frequencies are whole-number multiples of the fundamental (line) frequency, meaning that they will always be higher, not lower, than the design frequency of the transformer. In significant measure, they can cause severe transformer overheating. Power transformers can be engineered to handle certain levels of power system harmonics, and this capability is sometimes denoted with a "K factor" rating.

9.8.3 Stray capacitance and inductance

Aside from power ratings and power losses, transformers often harbor other undesirable limitations which circuit designers must be made aware of. Like their simpler counterparts – inductors – transformers exhibit capacitance due to the insulation dielectric between conductors: from winding to winding, turn to turn (in a single winding), and winding to core. Usually this capacitance is of no concern in a power application, but small signal applications (especially those of high frequency) may not tolerate this quirk well. Also, the effect of having capacitance along with the windings' designed inductance gives transformers the ability to *resonate* at a particular frequency, definitely a design concern in signal applications where the applied frequency may reach this point (usually the resonant frequency of a power transformer is well beyond the frequency of the AC power it was designed to operate on).

Flux containment (making sure a transformer's magnetic flux doesn't escape so as to interfere with another device, and making sure other devices' magnetic flux is shielded from the transformer core) is another concern shared both by inductors and transformers.

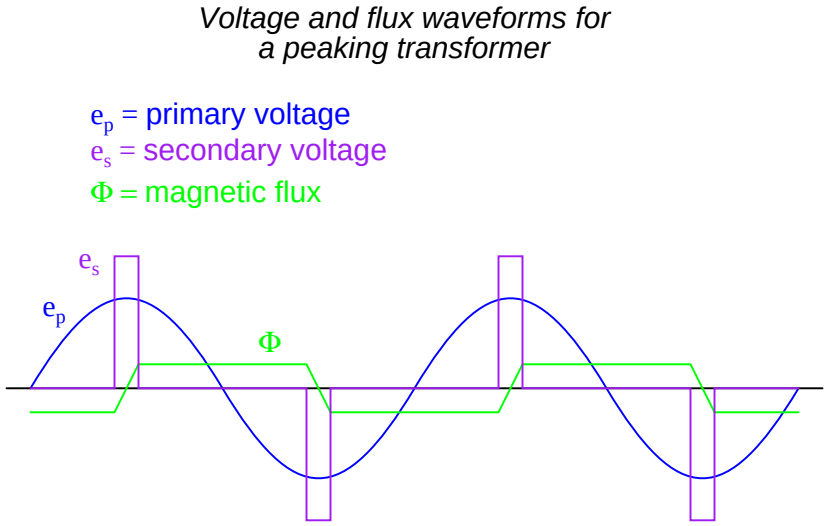
Closely related to the issue of flux containment is leakage inductance. We've already seen the detrimental effects of leakage inductance on voltage regulation with SPICE simulations early in this chapter. Because leakage inductance is equivalent to an inductance connected in series with the transformer's winding, it manifests itself as a series impedance with the load. Thus, the more current drawn by the load, the less voltage available at the secondary winding terminals. Usually, good voltage regulation is desired in transformer design, but there are exceptional applications. As was stated before, discharge lighting circuits require a step-up transformer with "loose" (poor) voltage regulation to ensure reduced voltage after the establishment of an arc through the lamp. One way to meet this design criterion is to engineer the transformer with flux leakage paths for magnetic flux to bypass the secondary winding(s). The resulting leakage flux will produce leakage inductance, which will in turn produce the poor regulation needed for discharge lighting.

9.8.4 Core saturation

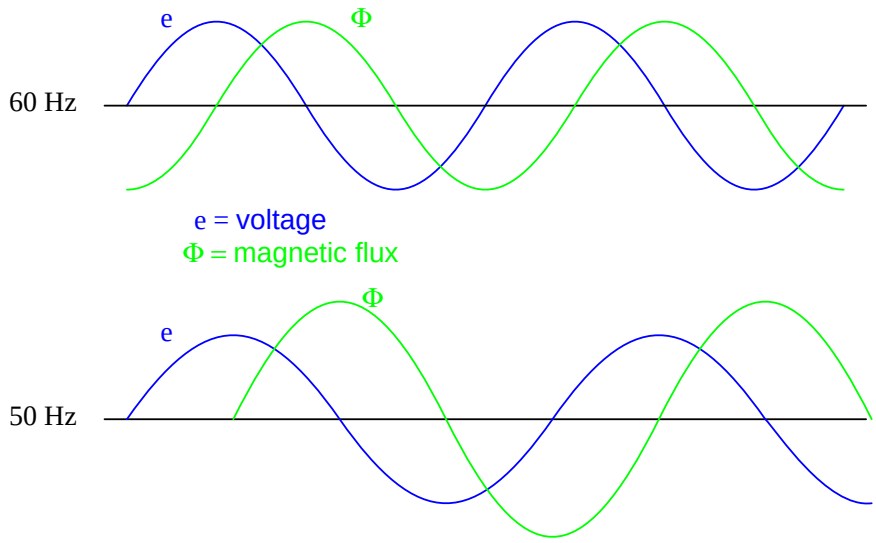
Transformers are also constrained in their performance by the magnetic flux limitations of the core. For ferromagnetic core transformers, we must be mindful of the saturation limits of the core. Remember that ferromagnetic materials cannot support infinite magnetic flux densities: they tend to "saturate" at a certain level (dictated by the material and core dimensions), meaning that further increases in magnetic field force (mmf) do not result in proportional increases in magnetic field flux (Φ).

When a transformer's primary winding is overloaded from excessive applied voltage, the core flux may reach saturation levels during peak moments of the AC sinewave cycle. If this happens, the voltage induced in the secondary winding will no longer match the wave-shape as the voltage powering the primary coil. In other words, the overloaded transformer will *distort* the waveshape from primary to secondary windings, creating harmonics in the secondary winding's output. As we discussed before, harmonic content in AC power systems typically causes problems.

Special transformers known as *peaking transformers* exploit this principle to produce brief voltage pulses near the peaks of the source voltage waveform. The core is designed to saturate quickly and sharply, at voltage levels well below peak. This results in a severely cropped sine-wave flux waveform, and secondary voltage pulses only when the flux is changing (below saturation levels):



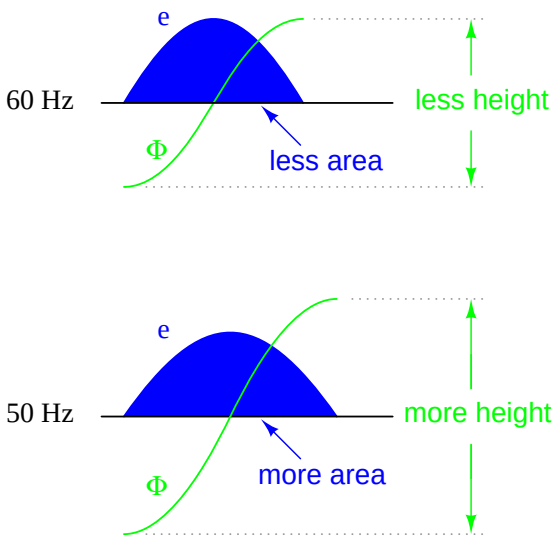
Another cause of abnormal transformer core saturation is operation at frequencies lower than normal. For example, if a power transformer designed to operate at 60 Hz is forced to operate at 50 Hz instead, the flux must reach greater peak levels than before in order to produce the same opposing voltage needed to balance against the source voltage. This is true even if the source voltage is the same as before.



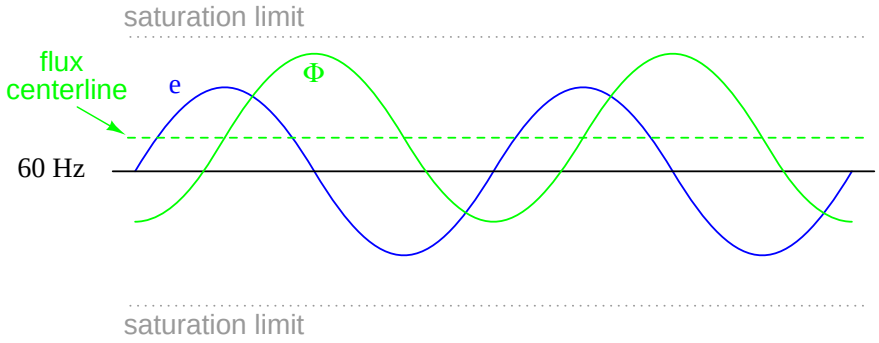
Since instantaneous winding voltage is proportional to the instantaneous magnetic flux's *rate of change* in a transformer, a voltage waveform reaching the same peak value, but taking a longer amount of time to complete each half-cycle, demands that the flux maintain the same rate of change as before, but for longer periods of time. Thus, if the flux has to climb at the same rate as before, but for longer periods of time, it will climb to a greater peak value.

Mathematically, this is another example of calculus in action. Because the voltage is proportional to the flux's rate-of-change, we say that the voltage waveform is the *derivative* of the flux waveform, "derivative" being that calculus operation defining one mathematical function (waveform) in terms of the rate-of-change of another. If we take the opposite perspective, though, and relate the original waveform to its derivative, we may call the original waveform the *integral* of the derivative waveform. In this case, the voltage waveform is the derivative of the flux waveform, and the flux waveform is the integral of the voltage waveform.

The integral of any mathematical function is proportional to the area accumulated underneath the curve of that function. Since each half-cycle of the 50 Hz waveform accumulates more area between it and the zero line of the graph than the 60 Hz waveform will – and we know that the magnetic flux is the integral of the voltage – the flux will attain higher values:



Yet another cause of transformer saturation is the presence of DC current in the primary winding. Any amount of DC voltage dropped across the primary winding of a transformer will cause additional magnetic flux in the core. This additional flux "bias" or "offset" will push the alternating flux waveform closer to saturation in one half-cycle than the other:



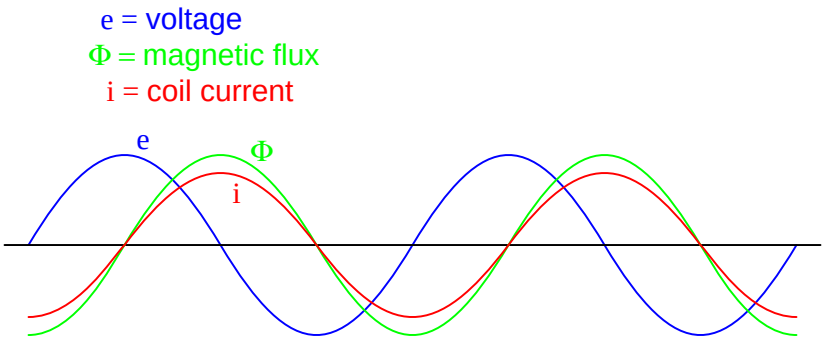
For most transformers, core saturation is a very undesirable effect, and it is avoided through good design: engineering the windings and core so that magnetic flux densities remain well below the saturation levels. This ensures that the relationship between mmf and Φ is more linear throughout the flux cycle, which is good because it makes for less distortion in the magnetization current waveform. Also, engineering the core for low flux densities provides a safe margin between the normal flux peaks and the core saturation limits to accommodate occasional, abnormal conditions such as frequency variation and DC offset.

9.8.5 Inrush current

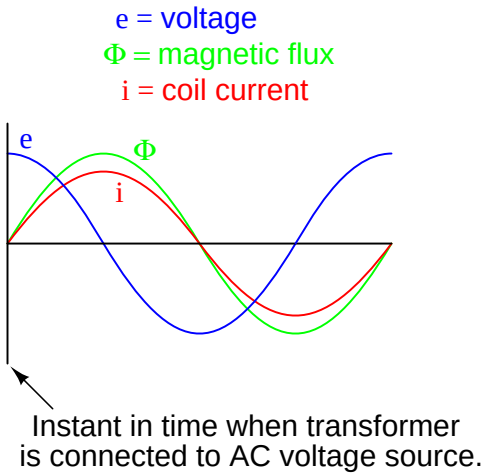
When a transformer is initially connected to a source of AC voltage, there may be a substantial surge of current through the primary winding called *inrush current*. This is analogous to the inrush

current exhibited by an electric motor that is started up by sudden connection to a power source, although transformer inrush is caused by a different phenomenon.

We know that the rate of change of instantaneous flux in a transformer core is proportional to the instantaneous voltage drop across the primary winding. Or, as stated before, the voltage waveform is the derivative of the flux waveform, and the flux waveform is the integral of the voltage waveform. In a continuously-operating transformer, these two waveforms are phase-shifted by 90^o. Since flux (Φ) is proportional to the magnetomotive force (mmf) in the core, and the mmf is proportional to winding current, the current waveform will be in-phase with the flux waveform, and both will be lagging the voltage waveform by 90^o:



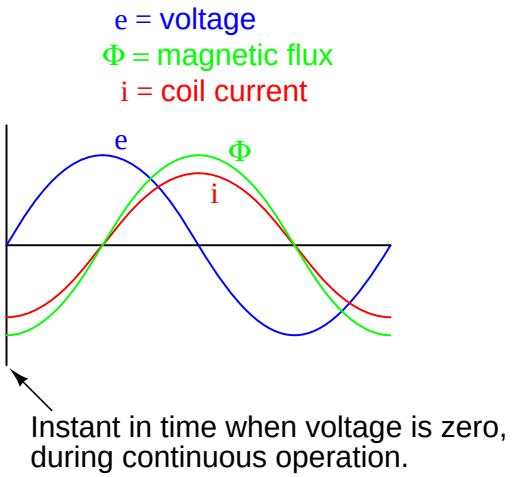
Let us suppose that the primary winding of a transformer is suddenly connected to an AC voltage source at the exact moment in time when the instantaneous voltage is at its positive peak value. In order for the transformer to create an opposing voltage drop to balance against this applied source voltage, a magnetic flux of rapidly increasing value must be generated. The result is that winding current increases rapidly, but actually no more rapidly than under normal conditions:



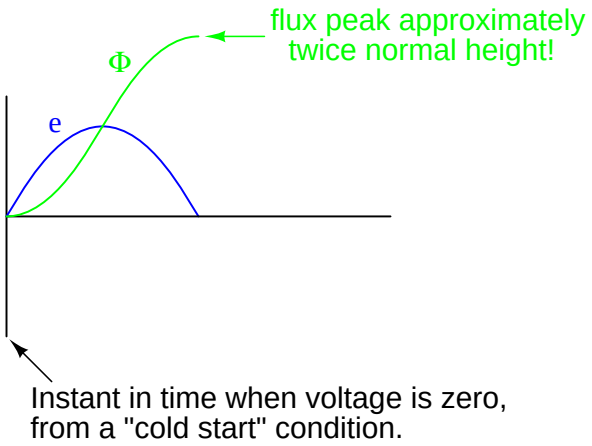
Both core flux and coil current start from zero and build up to the same peak values experienced during continuous operation. Thus, there is no "surge" or "inrush" or current in this scenario.

Alternatively, let us consider what happens if the transformer's connection to the AC voltage source occurs at the exact moment in time when the instantaneous voltage is at zero. During

continuous operation (when the transformer has been powered for quite some time), this is the point in time where both flux and winding current are at their negative peaks, experiencing zero rate-of-change ($d\Phi/dt = 0$ and $di/dt = 0$). As the voltage builds to its positive peak, the flux and current waveforms build to their maximum positive rates-of-change, and on upward to their positive peaks as the voltage descends to a level of zero:

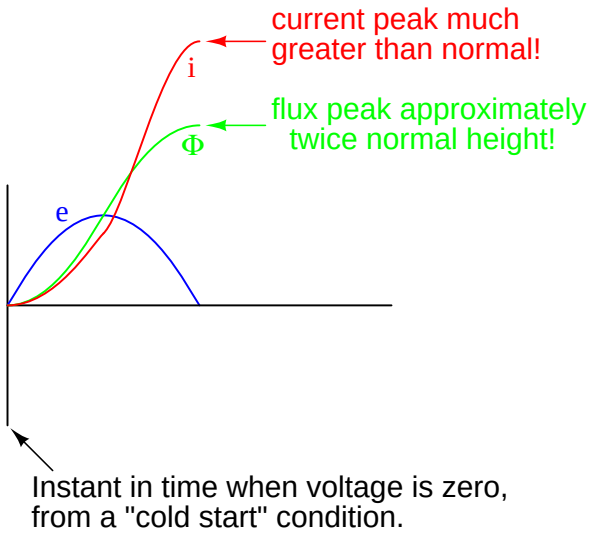


A significant difference exists, however, between continuous-mode operation and the sudden starting condition assumed in this scenario: during continuous operation, the flux and current levels were at their negative peaks when voltage was at its zero point; in a transformer that has been sitting idle, however, both magnetic flux and winding current should start at *zero*. When the magnetic flux increases in response to a rising voltage, it will increase from zero upwards, not from a previously negative (magnetized) condition as we would normally have in a transformer that's been powered for awhile. Thus, in a transformer that's just "starting," the flux will reach approximately twice its normal peak magnitude as it "integrates" the area under the voltage waveform's first half-cycle:



In an ideal transformer, the magnetizing current would rise to approximately twice its normal peak value as well, generating the necessary mmf to create this higher-than-normal flux. However,

most transformers aren't designed with enough of a margin between normal flux peaks and the saturation limits to avoid saturating in a condition like this, and so the core will almost certainly saturate during this first half-cycle of voltage. During saturation, disproportionate amounts of mmf are needed to generate magnetic flux. This means that winding current, which creates the mmf to cause flux in the core, will disproportionately rise to a value *easily exceeding* twice its normal peak:

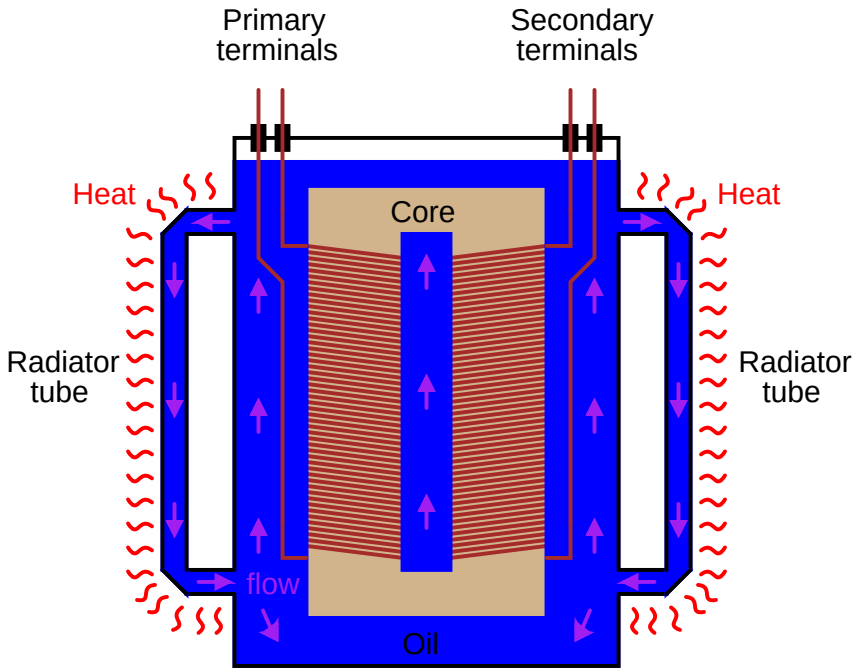


This is the mechanism causing inrush current in a transformer's primary winding when connected to an AC voltage source. As you can see, the magnitude of the inrush current strongly depends on the exact time that electrical connection to the source is made. If the transformer happens to have some residual magnetism in its core at the moment of connection to the source, the inrush could be even more severe. Because of this, transformer overcurrent protection devices are usually of the "slow-acting" variety, so as to tolerate current surges such as this without opening the circuit.

9.8.6 Heat and Noise

In addition to unwanted electrical effects, transformers may also exhibit undesirable physical effects, the most notable being the production of heat and noise. Noise is primarily a nuisance effect, but heat is a potentially serious problem because winding insulation will be damaged if allowed to overheat. Heating may be minimized by good design, ensuring that the core does not approach saturation levels, that eddy currents are minimized, and that the windings are not overloaded or operated too close to maximum ampacity.

Large power transformers have their core and windings submerged in an oil bath to transfer heat and muffle noise, and also to displace moisture which would otherwise compromise the integrity of the winding insulation. Heat-dissipating "radiator" tubes on the outside of the transformer case provide a convective oil flow path to transfer heat from the transformer's core to ambient air:



Oil-less, or "dry," transformers are often rated in terms of maximum operating temperature "rise" (temperature increase beyond ambient) according to a letter-class system: A, B, F, or H. These letter codes are arranged in order of lowest heat tolerance to highest:

- **Class A:** No more than 55^o Celsius winding temperature rise, at 40^o Celsius (maximum) ambient air temperature.
- **Class B:** No more than 80^o Celsius winding temperature rise, at 40^o Celsius (maximum)ambient air temperature.
- **Class F:** No more than 115^o Celsius winding temperature rise, at 40^o Celsius (maximum)ambient air temperature.
- **Class H:** No more than 150^o Celsius winding temperature rise, at 40^o Celsius (maximum)ambient air temperature.

Audible noise is an effect primarily originating from the phenomenon of *magnetostriction*: the slight change of length exhibited by a ferromagnetic object when magnetized. The familiar "hum" heard around large power transformers is the sound of the iron core expanding and contracting at 120 Hz (twice the system frequency, which is 60 Hz in the United States) – one cycle of core contraction and expansion for every peak of the magnetic flux waveform – plus noise created by mechanical forces between primary and secondary windings. Again, maintaining low magnetic flux levels in the core is the key to minimizing this effect, which explains why ferroresonant transformers – which must operate in saturation for a large portion of the current waveform – operate both hot and noisy.

Another noise-producing phenomenon in power transformers is the physical reaction force between primary and secondary windings when heavily loaded. If the secondary winding is open-circuited, there will be no current through it, and consequently no magneto-motive force (mmf) produced by it. However, when the secondary is "loaded" (current supplied to a load), the winding generates an mmf, which becomes counteracted by a "reflected" mmf in the primary winding to prevent core flux levels from changing. These opposing mmf's generated between primary and secondary windings as a result of secondary (load) current produce a repulsive, physical force between the windings which will tend to make them vibrate. Transformer designers have to consider these physical forces in the construction of the winding coils, to ensure there is adequate mechanical support to handle the stresses. Under heavy load conditions, though, these stresses may be great enough to cause audible noise to emanate from the transformer.

- **REVIEW:**
- Power transformers are limited in the amount of power they can transfer from primary to secondary winding(s). Large units are typically rated in VA (volt-amps) or kVA (kilo volt-amps).
- Resistance in transformer windings contributes to inefficiency, as current will dissipate heat, wasting energy.
- Magnetic effects in a transformer's iron core also contribute to inefficiency. Among the effects are *eddy currents* (circulating induction currents in the iron core) and *hysteresis* (power lost due to overcoming the tendency of iron to magnetize in a particular direction).
- Increased frequency results in increased power losses within a power transformer. The presence of harmonics in a power system is a source of frequencies significantly higher than normal, which may cause overheating in large transformers.
- Both transformers and inductors harbor certain unavoidable amounts of capacitance due to wire insulation (dielectric) separating winding turns from the iron core and from each other. This capacitance can be significant enough to give the transformer a natural *resonant frequency*, which can be problematic in signal applications.
- *Leakage inductance* is caused by magnetic flux not being 100% coupled between windings in a transformer. Any flux not involved with *transferring* energy from one winding to another will store and release energy, which is how (self-) inductance works. Leakage inductance tends to worsen a transformer's voltage regulation (secondary voltage "sags" more for a given amount of load current).
- Magnetic *saturation* of a transformer core may be caused by excessive primary voltage, operation at too low of a frequency, and/or by the presence of a DC current in any of the windings. Saturation may be minimized or avoided by conservative design, which provides an adequate margin of safety between peak magnetic flux density values and the saturation limits of the core.
- Transformers often experience significant *inrush currents* when initially connected to an AC voltage source. Inrush current is most severe when connection to the AC source is made at the moment instantaneous source voltage is zero.

- Noise is a common phenomenon exhibited by transformers – especially power transformers – and is primarily caused by *magnetostriction* of the core. Physical forces causing winding vibration may also generate noise under conditions of heavy (high current) secondary winding load.

9.9 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

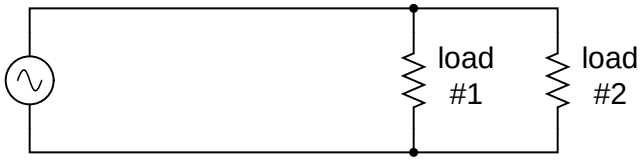
Bart Anderson (January 2004): Corrected conceptual errors regarding Tesla coil operation and safety.

Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 10

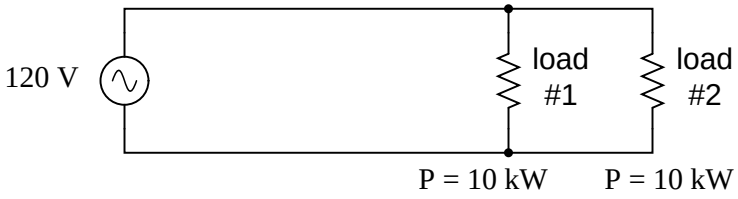
POLYPHASE AC CIRCUITS

10.1 Single-phase power systems



Depicted above is a very simple AC circuit. If the load resistor's power dissipation were substantial, we might call this a "power circuit" or "power system" instead of regarding it as just a regular circuit. The distinction between a "power circuit" and a "regular circuit" may seem arbitrary, but the practical concerns are definitely not.

One such concern is the size and cost of wiring necessary to deliver power from the AC source to the load. Normally, we do not give much thought to this type of concern if we're merely analyzing a circuit for the sake of learning about the laws of electricity. However, in the real world it can be a major concern. If we give the source in the above circuit a voltage value and also give power dissipation values to the two load resistors, we can determine the wiring needs for this particular circuit:



$$I = \frac{P}{E}$$

$$I = \frac{10\ \text{kW}}{120\ \text{V}}$$

$I = 83.33$ A *(for each load resistor)*

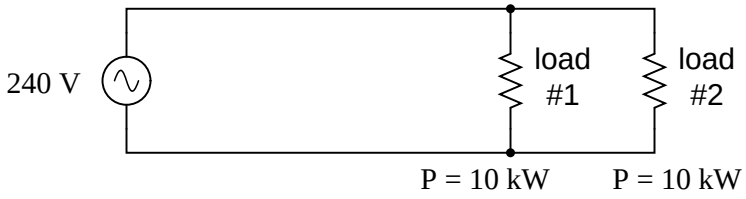
$I_{total} = I_{load\#1} + I_{load\#2}$ $\qquad$ $P_{total} = (10\ \text{kW}) + (10\ \text{kW})$

$I_{total} = (83.33\ \text{A}) + (83.33\ \text{A})$ $\qquad$ $P_{total} = 20\ \text{kW}$

$I_{total} = 166.67$ A

83.33 amps for each load resistor adds up to 166.66 amps total circuit current. This is no small amount of current, and would necessitate copper wire conductors of at least 1/0 gage. Such wire is well over 1/4 inch in diameter, weighing over 300 pounds per thousand feet. Bear in mind that copper is not cheap either! It would be in our best interest to find ways to minimize such costs if we were designing a power system with long conductor lengths.

One way to do this would be to increase the voltage of the power source and use loads built to dissipate 10 kW each at this higher voltage. The loads, of course, would have to have greater resistance values to dissipate the same power as before (10 kW each) at a greater voltage than before. The advantage would be less current required, permitting the use of smaller, lighter, and cheaper wire:



$$I = \frac{P}{E}$$

$$I = \frac{10\ \text{kW}}{240\ \text{V}}$$

$I = 41.67$ A *(for each load resistor)*

$I_{total} = I_{load\#1} + I_{load\#2}$ $P_{total} = (10\ \text{kW}) + (10\ \text{kW})$

$I_{total} = (41.67\ \text{A}) + (41.67\ \text{A})$ $P_{total} = 20$ kW

$I_{total} = 83.33$ A

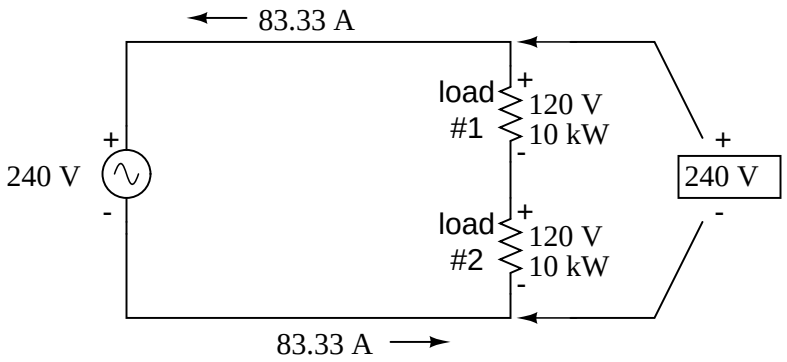
Now our *total* circuit current is 83.33 amps, half of what it was before. We can now use number 4 gage wire, which weighs less than half of what 1/0 gage wire does per unit length. This is a considerable reduction in system cost with no degradation in performance. This is why power distribution system designers elect to transmit electric power using very high voltages (many thousands of volts): to capitalize on the savings realized by the use of smaller, lighter, cheaper wire.

However, this solution is not without disadvantages. Another practical concern with power circuits is the danger of electric shock from high voltages. Again, this is not usually the sort of thing we concentrate on while learning about the laws of electricity, but it is a very valid concern in the real world, especially when large amounts of power are being dealt with. The gain in efficiency realized by stepping up the circuit voltage presents us with increased danger of electric shock. Power distribution companies tackle this problem by stringing their power lines along high poles or towers, and insulating the lines from the supporting structures with large, porcelain insulators.

At the point of use (the electric power customer), there is still the issue of what voltage to use for powering loads. High voltage gives greater system efficiency by means of reduced conductor current, but it might not always be practical to keep power wiring out of reach at the point of use the way it can be elevated out of reach in distribution systems. This tradeoff between efficiency and danger is one that European power system designers have decided to risk, all their households and appliances operating at a nominal voltage of 240 volts instead of 120 volts as it is in North America. That is why tourists from America visiting Europe must carry small step-down transformers for their portable appliances, to step the 240 VAC (volts AC) power down to a more suitable 120 VAC.

Is there any way to realize the advantages of both increased efficiency and reduced safety hazard at the same time? One solution would be to install step-down transformers at the end-point of power use, just as the American tourist must do while in Europe. However, this would be expensive and inconvenient for anything but very small loads (where the transformers can be built cheaply) or very large loads (where the expense of thick copper wires would exceed the expense of a transformer).

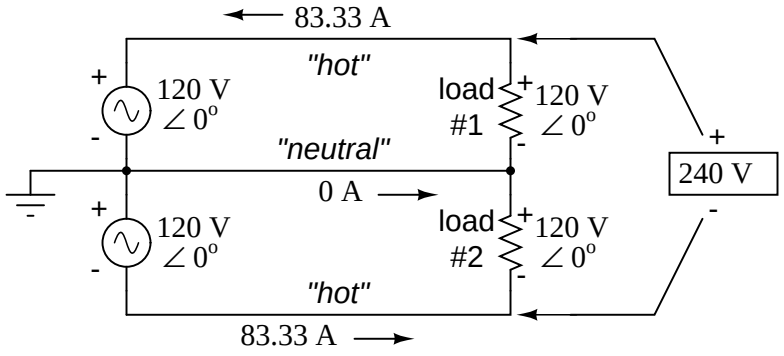
An alternative solution would be to use a higher voltage supply to provide power to two lower voltage loads in series. This approach combines the efficiency of a high-voltage system with the safety of a low-voltage system:



Notice the polarity markings (+ and -) for each voltage shown, as well as the unidirectional arrows for current. For the most part, I've avoided labeling "polarities" in the AC circuits we've been analyzing, even though the notation is valid to provide a frame of reference for phase. In later sections of this chapter, phase relationships will become very important, so I'm introducing this notation early on in the chapter for your familiarity.

The current through each load is the same as it was in the simple 120 volt circuit, but the currents are not additive because the loads are in series rather than parallel. The voltage across each load is only 120 volts, not 240, so the safety factor is better. Mind you, we still have a full 240 volts across the power system wires, but *each load* is operating at a reduced voltage. If anyone is going to get shocked, the odds are that it will be from coming into contact with the conductors of a particular load rather than from contact across the main wires of a power system.

There's only one disadvantage to this design: the consequences of one load failing open, or being turned off (assuming each load has a series on/off switch to interrupt current) are not good. Being a series circuit, if either load were to open, current would stop in the other load as well. For this reason, we need to modify the design a bit:



$E_{total} = (120\ V \angle 0^o) + (120\ V \angle 0^o)$

$E_{total} = 240\ V \angle 0^o$

$$I = \frac{P}{E}$$

$$I = \frac{10\ kW}{120\ V}$$

$P_{total} = (10\ kW) + (10\ kW)$

$P_{total} = 20\ kW$

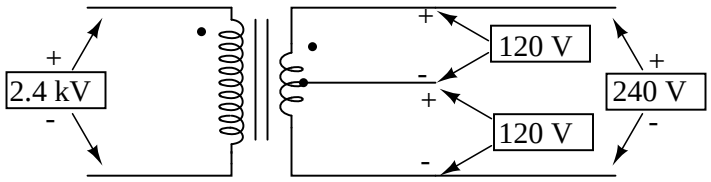
$I = 83.33\ A$ *(for each load resistor)*

Instead of a single 240 volt power supply, we use two 120 volt supplies (in phase with each other!) in series to produce 240 volts, then run a third wire to the connection point between the loads to handle the eventuality of one load opening. This is called a *split-phase* power system. Three smaller wires are still cheaper than the two wires needed with the simple parallel design, so we're still ahead on efficiency. The astute observer will note that the neutral wire only has to carry the *difference* of current between the two loads back to the source. In the above case, with perfectly "balanced" loads consuming equal amounts of power, the neutral wire carries zero current.

Notice how the neutral wire is connected to earth ground at the power supply end. This is a common feature in power systems containing "neutral" wires, since grounding the neutral wire ensures the least possible voltage at any given time between any "hot" wire and earth ground.

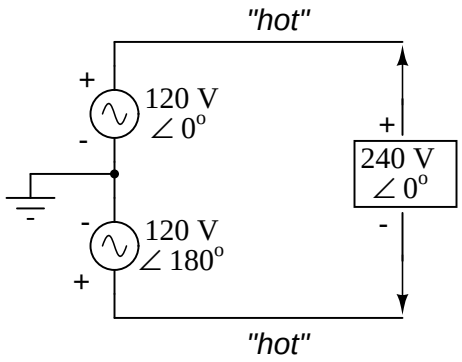
An essential component to a split-phase power system is the dual AC voltage source. Fortunately, designing and building one is not difficult. Since most AC systems receive their power from a step-down transformer anyway (stepping voltage down from high distribution levels to a user-level voltage like 120 or 240), that transformer can be built with a center-tapped secondary winding:

Step-down transformer with center-tapped secondary winding



If the AC power comes directly from a generator (alternator), the coils can be similarly center-tapped for the same effect. The extra expense to include a center-tap connection in a transformer or alternator winding is minimal.

Here is where the (+) and (-) polarity markings really become important. This notation is often used to reference the phasings of *multiple* AC voltage sources, so it is clear whether they are aiding ("boosting") each other or opposing ("bucking") each other. If not for these polarity markings, phase relations between multiple AC sources might be very confusing. Note that the split-phase sources in the schematic (each one 120 volts $\angle\ 0^o$), with polarity marks (+) to (-) just like series-aiding batteries can alternatively be represented as such:



To mathematically calculate voltage between "hot" wires, we must *subtract* voltages, because their polarity marks show them to be opposed to each other:

Polar	*Rectangular*
$120 \angle 0^o$	120 + j0 V
- $120 \angle 180^o$	- (-120 + j0) V
$240 \angle 0^o$	240 + j0 V

If we mark the two sources' common connection point (the neutral wire) with the same polarity mark (-), we must express their relative phase shifts as being 180^o apart. Otherwise, we'd be denoting two voltage sources in direct opposition with each other, which would give 0 volts between the two "hot" conductors. Why am I taking the time to elaborate on polarity marks and phase angles? It will make more sense in the next section!

Power systems in American households and light industry are most often of the split-phase variety, providing so-called 120/240 VAC power. The term "split-phase" merely refers to the split-voltage supply in such a system. In a more general sense, this kind of AC power supply is called *single phase* because both voltage waveforms are in phase, or in step, with each other.

The term "single phase" is a counterpoint to another kind of power system called "polyphase" which we are about to investigate in detail. Apologies for the long introduction leading up to the title-topic of this chapter. The advantages of polyphase power systems are more obvious if one first has a good understanding of single phase systems.

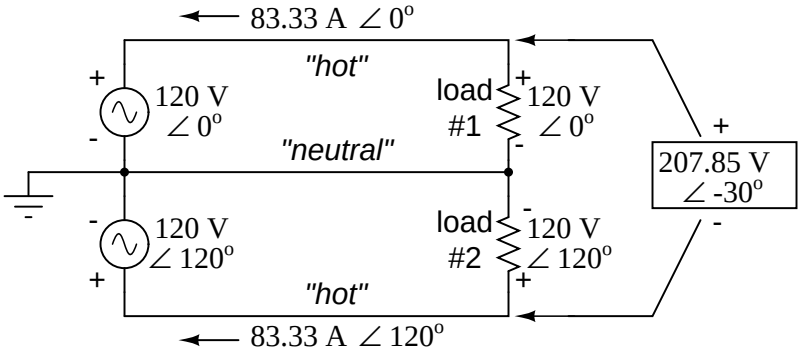
- **REVIEW:**
- *Single phase* power systems are defined by having an AC source with only one voltage waveform.
- A *split-phase* power system is one with multiple (in-phase) AC voltage sources connected in series, delivering power to loads at more than one voltage, with more than two wires. They are used primarily to achieve balance between system efficiency (low conductor currents) and safety (low load voltages).
- Split-phase AC sources can be easily created by center-tapping the coil windings of transformers or alternators.

10.2 Three-phase power systems

Split-phase power systems achieve their high conductor efficiency *and* low safety risk by splitting up the total voltage into lesser parts and powering multiple loads at those lesser voltages, while drawing currents at levels typical of a full-voltage system. This technique, by the way, works just as well for DC power systems as it does for single-phase AC systems. Such systems are usually referred to as *three-wire* systems rather than *split-phase* because "phase" is a concept restricted to AC.

But we know from our experience with vectors and complex numbers that AC voltages don't always add up as we think they would if they are out of phase with each other. This principle, applied to power systems, can be put to use to make power systems with even greater conductor efficiencies and lower shock hazard than with split-phase.

Suppose that we had two sources of AC voltage connected in series just like the split-phase system we saw before, except that each voltage source was 120^o out of phase with the other:



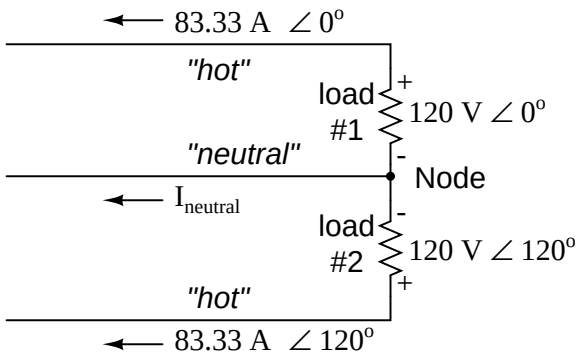
Since each voltage source is 120 volts, and each load resistor is connected directly in parallel with its respective source, the voltage across each load *must* be 120 volts as well. Given load currents of 83.33 amps, each load must still be dissipating 10 kilowatts of power. However, voltage between the two "hot" wires is not 240 volts (120 $\angle$ 0^o - 120 $\angle$ 180^o) because the phase difference between the two sources is not 180^o. Instead, the voltage is:

$$E_{total} = (120 \text{ V} \angle 0^o) - (120 \text{ V} \angle 120^o)$$

$$E_{total} = 207.85 \text{ V} \angle -30^o$$

Nominally, we say that the voltage between "hot" conductors is 208 volts (rounding up), and thus the power system voltage is designated as 120/208.

If we calculate the current through the "neutral" conductor, we find that it is *not* zero, even with balanced load resistances. Kirchhoff's Current Law tells us that the currents entering and exiting the node between the two loads must be zero:



$$-I_{load\#1} - I_{load\#2} - I_{neutral} = 0$$

$$-I_{neutral} = I_{load\#1} + I_{load\#2}$$

$$I_{neutral} = -I_{load\#1} - I_{load\#2}$$

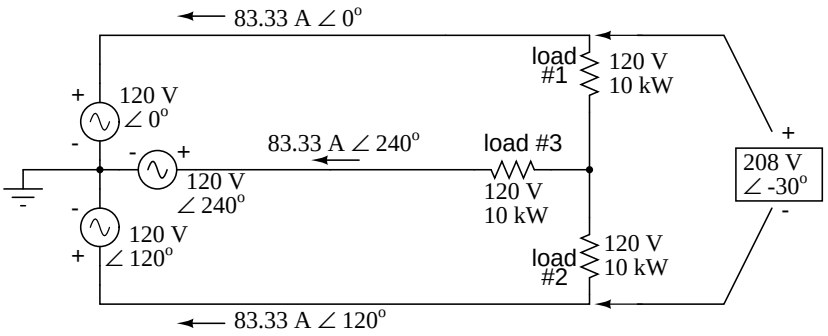
$$I_{neutral} = -\,(83.33\ A \angle 0^{o}) - (83.33\ A \angle 120^{0})$$

$$I_{neutral} = 83.33\ A \angle 240^{o} \quad or \quad 83.33\ A \angle -120^{o}$$

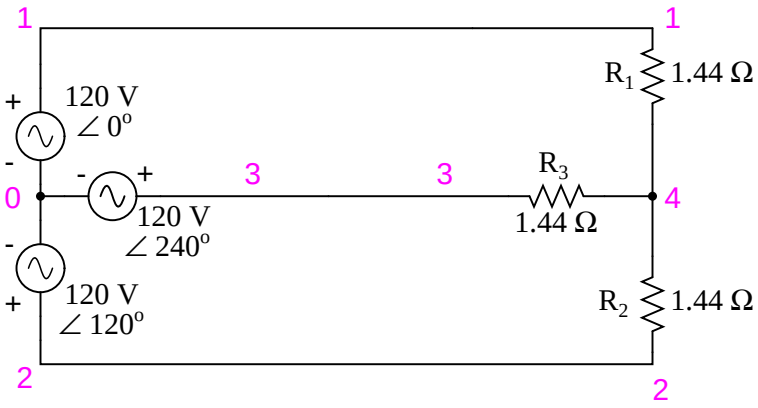
So, we find that the "neutral" wire is carrying a full 83.33 amps, just like each "hot" wire.

Note that we are still conveying 20 kW of total power to the two loads, with each load's "hot" wire carrying 83.33 amps as before. With the same amount of current through each "hot" wire, we must use the same gage copper conductors, so we haven't reduced system cost over the split-phase 120/240 system. However, we have realized a gain in safety, because the overall voltage between the two "hot" conductors is 32 volts lower than it was in the split-phase system (208 volts instead of 240 volts).

The fact that the neutral wire is carrying 83.33 amps of current raises an interesting possibility: since it's carrying current anyway, why not use that third wire as another "hot" conductor, powering another load resistor with a third 120 volt source having a phase angle of 240^{o}? That way, we could transmit *more* power (another 10 kW) without having to add any more conductors. Let's see how this might look:



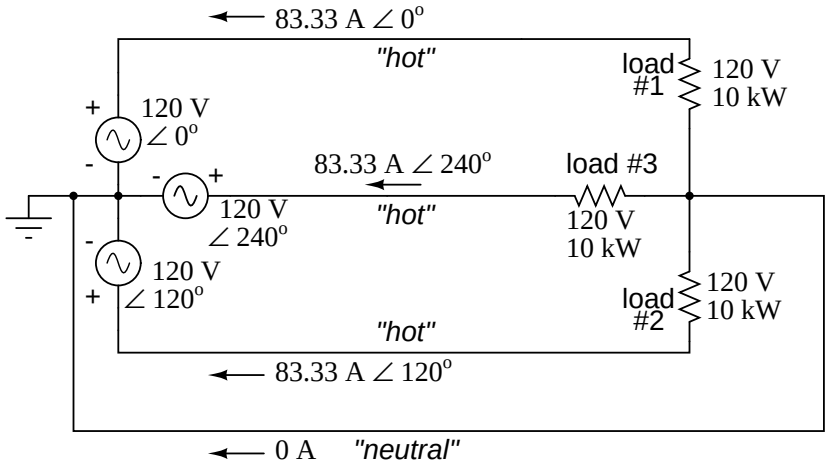
A full mathematical analysis of all the voltages and currents in this circuit would necessitate the use of a network theorem, the easiest being the Superposition Theorem. I'll spare you the long, drawn-out calculations because you should be able to intuitively understand that the three voltage sources at three different phase angles will deliver 120 volts each to a balanced triad of load resistors. For proof of this, we can use SPICE to do the math for us:




```
120/208 polyphase power system
v1 1 0 ac 120 0 sin
v2 2 0 ac 120 120 sin
v3 3 0 ac 120 240 sin
r1 1 4 1.44
r2 2 4 1.44
r3 3 4 1.44
.ac lin 1 60 60
.print ac v(1,4) v(2,4) v(3,4)
.print ac v(1,2) v(2,3) v(3,1)
.print ac i(v1) i(v2) i(v3)
.end

VOLTAGE ACROSS EACH LOAD
freq        v(1,4)      v(2,4)      v(3,4)
```

```
6.000E+01   1.200E+02   1.200E+02   1.200E+02

VOLTAGE BETWEEN "HOT" CONDUCTORS
freq        v(1,2)      v(2,3)      v(3,1)
6.000E+01   2.078E+02   2.078E+02   2.078E+02

CURRENT THROUGH EACH VOLTAGE SOURCE
freq        i(v1)       i(v2)       i(v3)
6.000E+01   8.333E+01   8.333E+01   8.333E+01
```

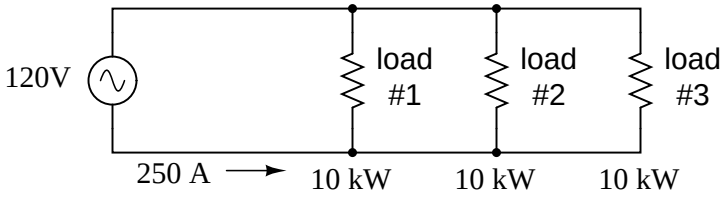
Sure enough, we get 120 volts across each load resistor, with (approximately) 208 volts between any two "hot" conductors and conductor currents equal to 83.33 amps. At that current and voltage, each load will be dissipating 10 kW of power. Notice that this circuit has no "neutral" conductor to ensure stable voltage to all loads if one should open. What we have here is a situation similar to our split-phase power circuit with no "neutral" conductor: if one load should happen to fail open, the voltage drops across the remaining load(s) will change. To ensure load voltage stability in the even of another load opening, we need a neutral wire to connect the source node and load node together:



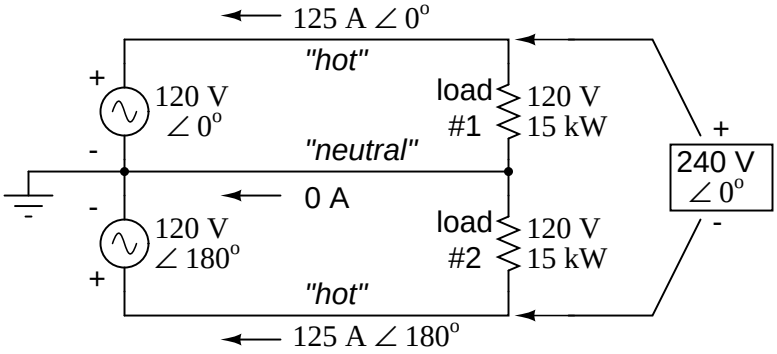
So long as the loads remain balanced (equal resistance, equal currents), the neutral wire will not have to carry any current at all. It is there just in case one or more load resistors should fail open (or be shut off through a disconnecting switch).

This circuit we've been analyzing with three voltage sources is called a *polyphase* circuit. The prefix "poly" simply means "more than one," as in "*poly*theism" (belief in more than one deity), *poly*gon" (a geometrical shape made of multiple line segments: for example, *pentagon* and *hexagon*), and "*poly*atomic" (a substance composed of multiple types of atoms). Since the voltage sources are all at different phase angles (in this case, three different phase angles), this is a "*poly*phase" circuit. More specifically, it is a *three-phase circuit*, the kind used predominantly in large power distribution systems.

Let's survey the advantages of a three-phase power system over a single-phase system of equivalent load voltage and power capacity. A single-phase system with three loads connected directly in parallel would have a very high total current (83.33 times 3, or 250 amps:



This would necessitate 3/0 gage copper wire (*very* large!), at about 510 pounds per thousand feet, and with a considerable price tag attached. If the distance from source to load was 1000 feet, we would need over a half-ton of copper wire to do the job. On the other hand, we could build a split-phase system with two 15 kW, 120 volt loads:

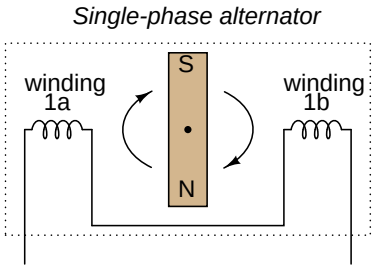


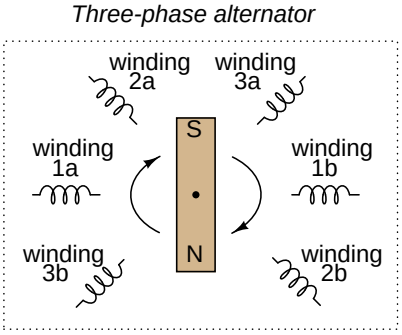
Our current is half of what it was with the simple parallel circuit, which is a great improvement. We could get away with using number 2 gage copper wire at a total mass of about 600 pounds, figuring about 200 pounds per thousand feet with three runs of 1000 feet each between source and loads. However, we also have to consider the increased safety hazard of having 240 volts present in the system, even though each load only receives 120 volts. Overall, there is greater potential for dangerous electric shock to occur.

When we contrast these two examples against our three-phase system, the advantages are quite clear. First, the conductor currents are quite a bit less (83.33 amps versus 125 or 250 amps), permitting the use of much thinner and lighter wire. We can use number 4 gage wire at about 125 pounds per thousand feet, which will total 500 pounds (four runs of 1000 feet each) for our example circuit. This represents a significant cost savings over the split-phase system, with the additional benefit that the maximum voltage in the system is lower (208 versus 240).

One question remains to be answered: how in the world do we get three AC voltage sources whose phase angles are exactly 120^o apart? Obviously we can't center-tap a transformer or alternator winding like we did in the split-phase system, since that can only give us voltage waveforms that are either in phase or 180^o out of phase. Perhaps we could figure out some way to use capacitors and inductors to create phase shifts of 120^o, but then those phase shifts would depend on the phase angles of our load impedances as well (substituting a capacitive or inductive load for a resistive load would change everything!).

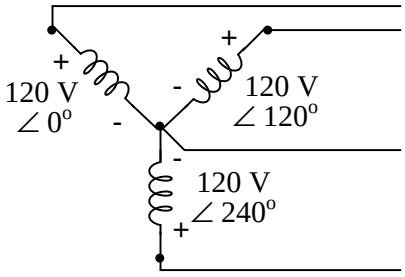
The best way to get the phase shifts we're looking for is to generate it at the source: construct the AC generator (alternator) providing the power in such a way that the rotating magnetic field passes by three sets of wire windings, each set spaced 120^o apart around the circumference of the machine:





Together, the six "pole" windings of a three-phase alternator are connected to comprise three winding pairs, each pair producing AC voltage with a phase angle 120^o shifted from either of the other two winding pairs. The interconnections between pairs of windings (as shown for the single-phase alternator: the jumper wire between windings 1a and 1b) have been omitted from the three-phase alternator drawing for simplicity.

In our example circuit, we showed the three voltage sources connected together in a "Y" configuration (sometimes called the "star" configuration), with one lead of each source tied to a common point (the node where we attached the "neutral" conductor). The common way to depict this connection scheme is to draw the windings in the shape of a "Y" like this:



The "Y" configuration is not the only option open to us, but it is probably the easiest to understand at first. More to come on this subject later in the chapter.

- **REVIEW:**

- A *single-phase* power system is one where there is only one AC voltage source (one source voltage waveform).

- A *split-phase* power system is one where there are two voltage sources, 180^o phase-shifted from each other, powering a two series-connected loads. The advantage of this is the ability to have lower conductor currents while maintaining low load voltages for safety reasons.

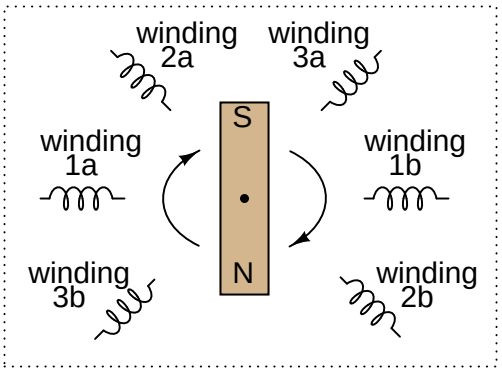
- A *polyphase* power system uses multiple voltage sources at different phase angles from each other (many "phases" of voltage waveforms at work). A polyphase power system can deliver more power at less voltage with smaller-gage conductors than single- or split-phase systems.

- The phase-shifted voltage sources necessary for a polyphase power system are created in alternators with multiple sets of wire windings. These winding sets are spaced around the circumference of the rotor's rotation at the desired angle(s).

10.3 Phase rotation

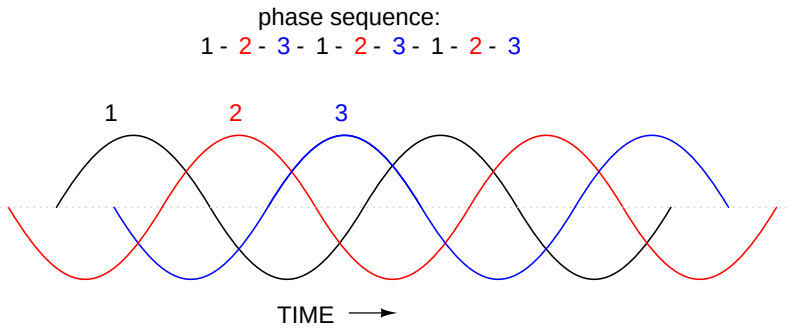
Let's take the three-phase alternator design laid out earlier and watch what happens as the magnet rotates:

Three-phase alternator

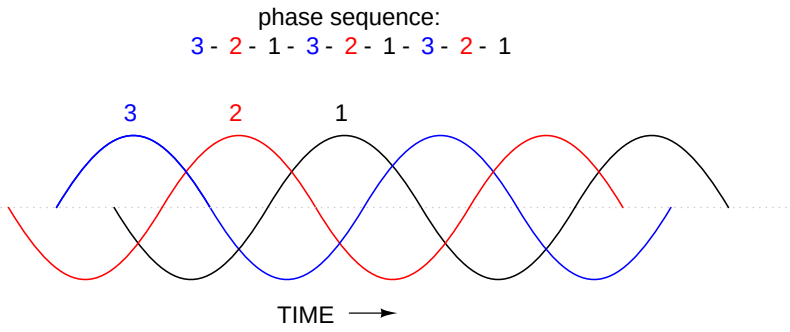


The phase angle shift of 120^o is a function of the actual rotational angle shift of the three pairs of windings. If the magnet is rotating clockwise, winding 3 will generate its peak instantaneous voltage exactly 120^o (of alternator shaft rotation) after winding 2, which will hits its peak 120^o after winding 1. The magnet passes by each pole pair at different positions in the rotational movement of the shaft. Where we decide to place the windings will dictate the amount of phase shift between the windings' AC voltage waveforms. If we make winding 1 our "reference" voltage source for phase angle (0^o), then winding 2 will have a phase angle of -120^o (120^o lagging, or 240^o leading) and winding 3 an angle of -240^o (or 120^o leading).

This sequence of phase shifts has a definite order. For clockwise rotation of the shaft, the order is 1-2-3 (winding 1 peaks first, them winding 2, then winding 3). This order keeps repeating itself as long as we continue to rotate the alternator's shaft:

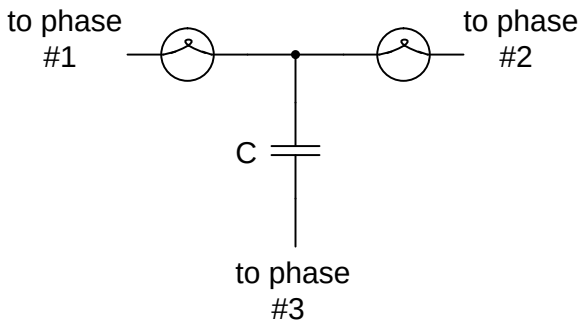


However, if we *reverse* the rotation of the alternator's shaft (turn it counter-clockwise), the magnet will pass by the pole pairs in the opposite sequence. Instead of 1-2-3, we'll have 3-2-1. Now, winding 2's waveform will be *leading* 120^o ahead of 1 instead of lagging, and 3 will be another 120^o ahead of 2:

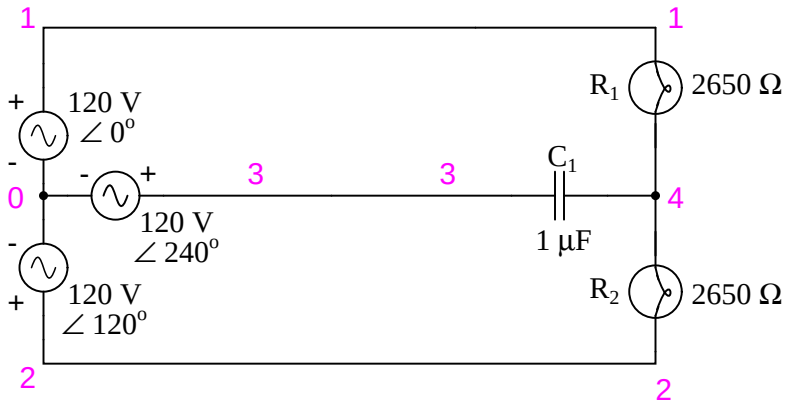


The order of voltage waveform sequences in a polyphase system is called *phase rotation* or *phase sequence*. If we're using a polyphase voltage source to power resistive loads, phase rotation will make no difference at all. Whether 1-2-3 or 3-2-1, the voltage and current magnitudes will all be the same. There are some applications of three-phase power, as we will see shortly, that depend on having phase rotation being one way or the other. Since voltmeters and ammeters would be useless in telling us what the phase rotation of an operating power system is, we need to have some other kind of instrument capable of doing the job.

One ingenious circuit design uses a capacitor to introduce a phase shift between voltage and current, which is then used to detect the sequence by way of comparison between the brightness of two indicator lamps:



The two lamps are of equal filament resistance and wattage. The capacitor is sized to have approximately the same amount of reactance at system frequency as each lamp's resistance. If the capacitor were to be replaced by a resistor of equal value to the lamps' resistance, the two lamps would glow at equal brightness, the circuit being balanced. However, the capacitor introduces a phase shift between voltage and current in the third leg of the circuit equal to 90^o. This phase shift, greater than 0^o but less than 120^o, skews the voltage and current values across the two lamps according to their phase shifts relative to phase 3. The following SPICE analysis demonstrates what will happen:




```
phase rotation detector -- sequence = v1-v2-v3
v1 1 0 ac 120 0 sin
v2 2 0 ac 120 120 sin
v3 3 0 ac 120 240 sin
r1 1 4 2650
r2 2 4 2650
c1 3 4 1u
.ac lin 1 60 60
.print ac v(1,4) v(2,4) v(3,4)
.end

freq          v(1,4)      v(2,4)      v(3,4)
6.000E+01     4.810E+01   1.795E+02   1.610E+02
```

The resulting phase shift from the capacitor causes the voltage across phase 1 lamp (between nodes 1 and 4) to fall to 48.1 volts and the voltage across phase 2 lamp (between nodes 2 and 4) to rise to 179.5 volts, making the first lamp dim and the second lamp bright. Just the opposite will happen if the phase sequence is reversed:

```
phase rotation detector -- sequence = v3-v2-v1
v1 1 0 ac 120 240 sin
v2 2 0 ac 120 120 sin
v3 3 0 ac 120 0 sin
r1 1 4 2650
r2 2 4 2650
c1 3 4 1u
.ac lin 1 60 60
.print ac v(1,4) v(2,4) v(3,4)
.end
```

```
freq          v(1,4)      v(2,4)      v(3,4)
6.000E+01     1.795E+02   4.810E+01   1.610E+02
```

Here, the first lamp receives 179.5 volts while the second receives only 48.1 volts.

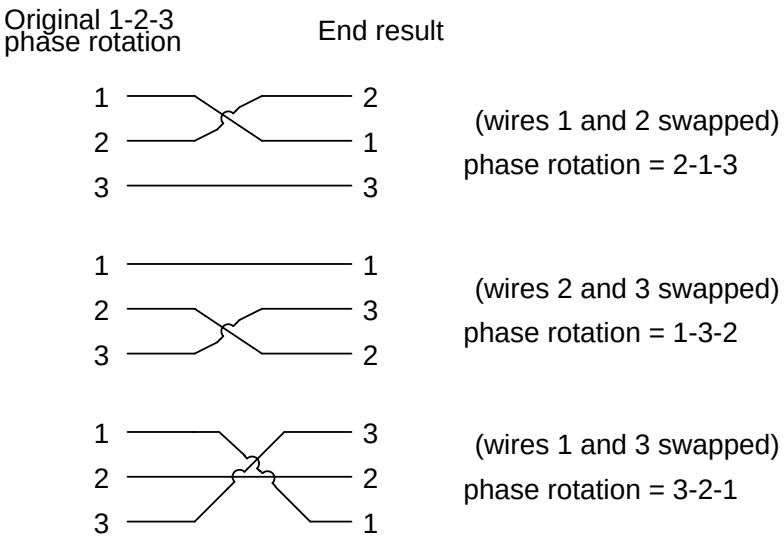
We've investigated how phase rotation is produced (the order in which pole pairs get passed by the alternator's rotating magnet) and how it can be changed by reversing the alternator's shaft rotation. However, reversal of the alternator's shaft rotation is not usually an option open to an end-user of electrical power supplied by a nationwide grid ("the" alternator actually being the combined total of all alternators in all power plants feeding the grid). There is a *much* easier way to reverse phase sequence than reversing alternator rotation: just exchange any two of the three "hot" wires going to a three-phase load.

This trick makes more sense if we take another look at a running phase sequence of a three-phase voltage source:

```
1-2-3 rotation:  1-2-3-1-2-3-1-2-3-1-2-3-1-2-3 . . .
3-2-1 rotation:  3-2-1-3-2-1-3-2-1-3-2-1-3-2-1 . . .
```

What is commonly designated as a "1-2-3" phase rotation could just as well be called "2-3-1" or "3-1-2," going from left to right in the number string above. Likewise, the opposite rotation (3-2-1) could just as easily be called "2-1-3" or "1-3-2."

Starting out with a phase rotation of 3-2-1, we can try all the possibilities for swapping any two of the wires at a time and see what happens to the resulting sequence:



No matter which pair of "hot" wires out of the three we choose to swap, the phase rotation ends up being reversed (1-2-3 gets changed to 2-1-3, 1-3-2 or 3-2-1, all equivalent).

- **REVIEW:**

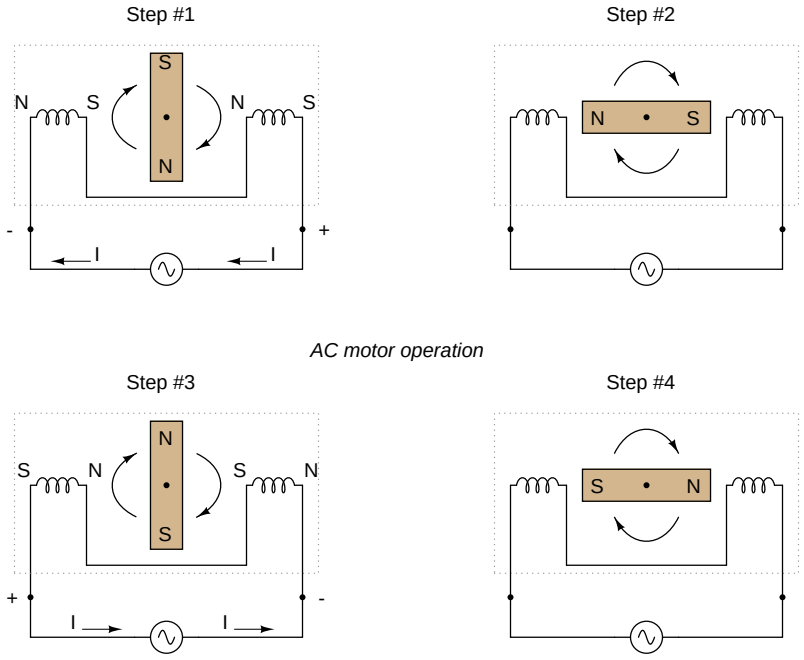
- *Phase rotation*, or *phase sequence*, is the order in which the voltage waveforms of a polyphase AC source reach their respective peaks. For a three-phase system, there are only two possible phase sequences: 1-2-3 and 3-2-1, corresponding to the two possible directions of alternator rotation.

- Phase rotation has no impact on resistive loads, but it will have impact on unbalanced reactive loads, as shown in the operation of a phase rotation detector circuit.

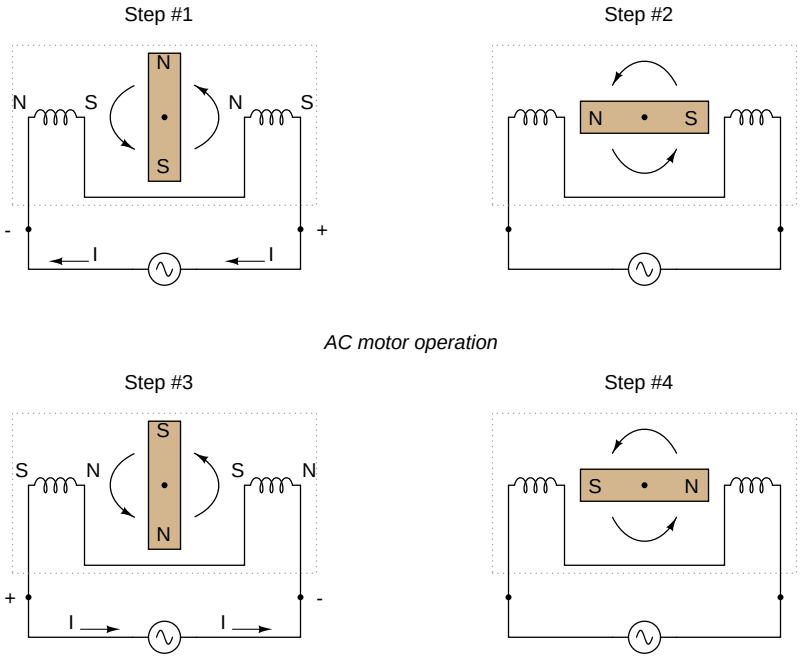
- Phase rotation can be reversed by swapping any two of the three "hot" leads supplying three-phase power to a three-phase load.

10.4 Polyphase motor design

Perhaps the most important benefit of polyphase AC power over single-phase is the design and operation of AC motors. As we studied in the first chapter of this book, some types of AC motors are virtually identical in construction to their alternator (generator) counterparts, consisting of stationary wire windings and a rotating magnet assembly. (Other AC motor designs are not quite this simple, but we will leave those details to another lesson).

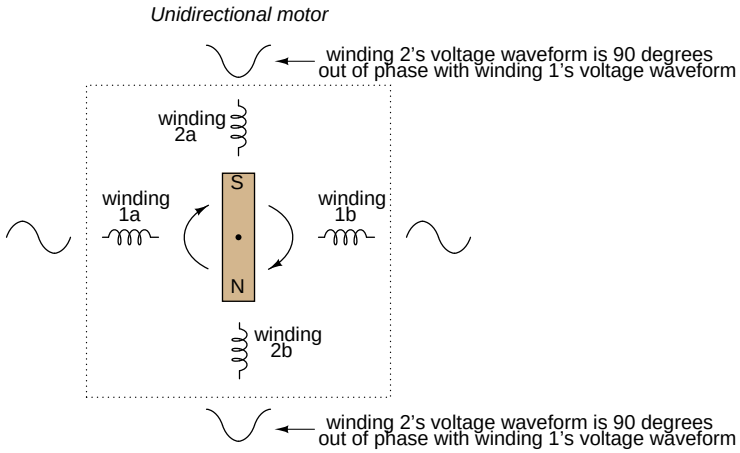


If the rotating magnet is able to keep up with the frequency of the alternating current energizing the electromagnet windings (coils), it will continue to be pulled around clockwise. However, clockwise is not the only valid direction for this motor's shaft to spin. It could just as easily be powered in a counter-clockwise direction by the same AC voltage waveform:

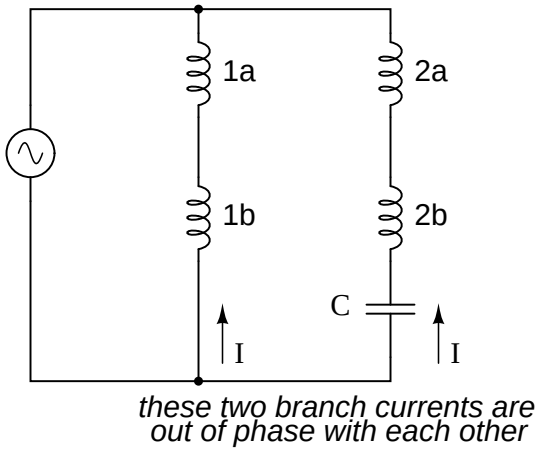


Notice that with the exact same sequence of polarity cycles (voltage, current, and magnetic poles produced by the coils), the magnetic rotor can spin in either direction. This is a common trait of all single-phase AC "induction" and "synchronous" motors: they have no normal or "correct" direction of rotation. The natural question should arise at this point: how can the motor get started in the intended direction if it can run either way just as well? The answer is that these motors need a little help getting started. Once helped to spin in a particular direction. they will continue to spin that way as long as AC power is maintained to the windings.

Where that "help" comes from for a single-phase AC motor to get going in one direction can vary. Usually, it comes from an additional set of windings positioned differently from the main set, and energized with an AC voltage that is out of phase with the main power:



These supplementary coils are typically connected in series with a capacitor to introduce a phase shift in current between the two sets of windings:



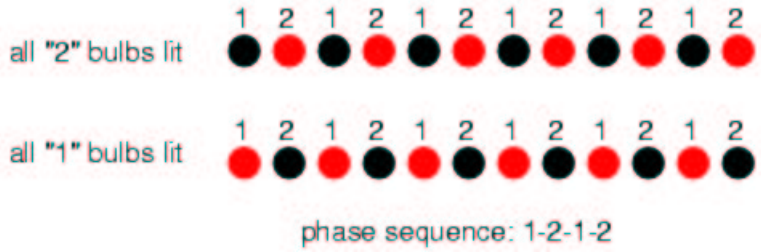
That phase shift creates magnetic fields from coils 2a and 2b that are equally out of step with the fields from coils 1a and 1b. The result is a set of magnetic fields with a definite phase rotation. It is this phase rotation that pulls the rotating magnet around in a definite direction.

Polyphase AC motors require no such trickery to spin in a definite direction. Because their supply voltage waveforms already have a definite rotation sequence, so do the respective magnetic fields generated by the motor's stationary windings. In fact, the combination of all three phase winding sets working together creates what is often called a *rotating magnetic field*. It was this concept of a rotating magnetic field that inspired Nikola Tesla to design the world's first polyphase electrical systems (simply to make simpler, more efficient motors). The line current and safety advantages of polyphase power over single phase power were discovered later.

What can be a confusing concept is made much clearer through analogy. Have you ever seen a row of blinking light bulbs such as the kind used in Christmas decorations? Some strings appear to "move" in a definite direction as the bulbs alternately glow and darken in sequence. Other strings just blink on and off with no apparent motion. What makes the difference between the two types

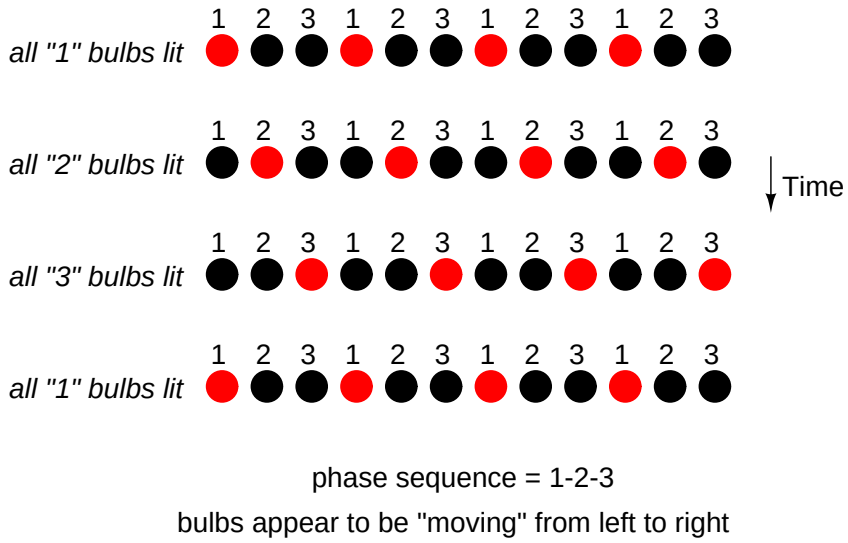
of bulb strings? Answer: phase shift!

Examine a string of lights where every other bulb is lit at any given time:

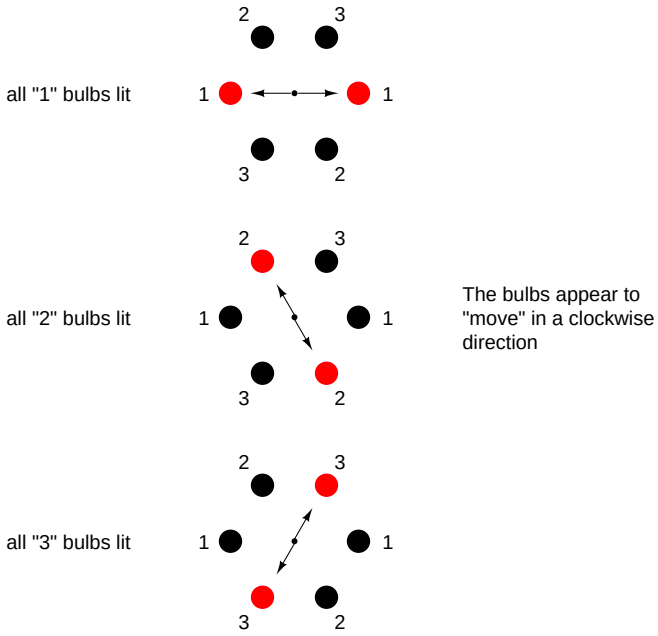


When all of the "1" bulbs are lit, the "2" bulbs are dark, and visa-versa. With this blinking sequence, there is no definite "motion" to the bulbs' light. Your eyes could follow a "motion" from left to right just as easily as from right to left. Technically, the "1" and "2" bulb blinking sequences are 180^o out of phase (exactly opposite each other). This is analogous to the single-phase AC motor, which can run just as easily in either direction, but which cannot start on its own because its magnetic field alternation lacks a definite "rotation."

Now let's examine a string of lights where there are three sets of bulbs to be sequenced instead of just two, and these three sets are equally out of phase with each other:



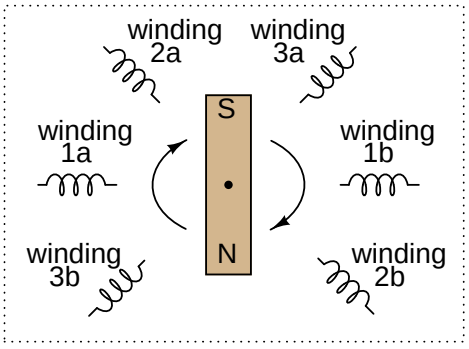
If the lighting sequence is 1-2-3 (the sequence shown), the bulbs will appear to "move" from left to right. Now imagine this blinking string of bulbs arranged into a circle:



Now the lights appear to be "moving" in a clockwise direction because they are arranged around a circle instead of a straight line. It should come as no surprise that the appearance of motion will reverse if the phase sequence of the bulbs is reversed.

The blinking pattern will either appear to move clockwise or counter-clockwise depending on the phase sequence. This is analogous to a three-phase AC motor with three sets of windings energized by voltage sources of three different phase shifts:

Three-phase AC motor



A phase sequence of 1-2-3 will spin the magnet in a clockwise direction. A phase sequence of 3-2-1 will spin the magnet in a counter-clockwise direction.

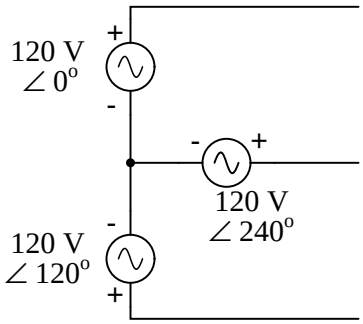
With phase shifts of less than 180^o we get true rotation of the magnetic field. With single-phase motors, the rotating magnetic field necessary for self-starting must to be created by way of capacitive

phase shift. With polyphase motors, the necessary phase shifts are there already. Plus, the direction of shaft rotation for polyphase motors is very easily reversed: just swap any two "hot" wires going to the motor, and it will run in the opposite direction!

- **REVIEW:**
- AC "induction" and "synchronous" motors work by having a rotating magnet follow the alternating magnetic fields produced by stationary wire windings.
- Single-phase AC motors of this type need help to get started spinning in a particular direction.
- By introducing a phase shift of less than 180^o to the magnetic fields in such a motor, a definite direction of shaft rotation can be established.
- Single-phase induction motors often use an auxiliary winding connected in series with a capacitor to create the necessary phase shift.
- Polyphase motors don't need such measures; their direction of rotation is fixed by the phase sequence of the voltage they're powered by.
- Swapping any two "hot" wires on a polyphase AC motor will reverse its phase sequence, thus reversing its shaft rotation.

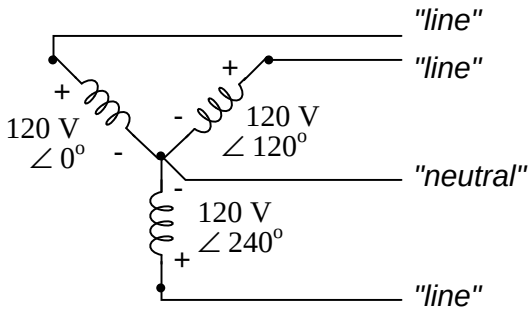
10.5 Three-phase Y and Δ configurations

Initially we explored the idea of three-phase power systems by connecting three voltage sources together in what is commonly known as the "Y" (or "star") configuration. This configuration of voltage sources is characterized by a common connection point joining one side of each source:



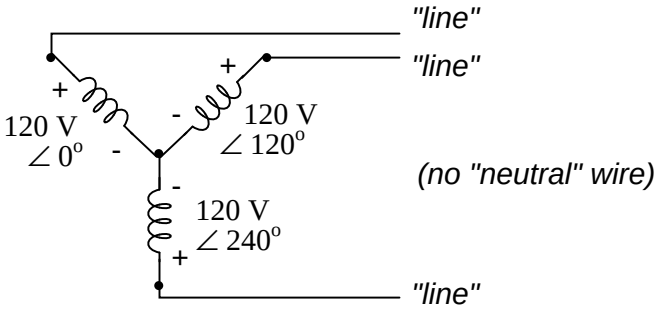
If we draw a circuit showing each voltage source to be a coil of wire (alternator or transformer winding) and do some slight rearranging, the "Y" configuration becomes more obvious:

3-phase, 4-wire "Y" connection



The three conductors leading away from the voltage sources (windings) toward a load are typically called *lines*, while the windings themselves are typically called *phases*. In a Y-connected system, there may or may not be a neutral wire attached at the junction point in the middle, although it certainly helps alleviate potential problems should one element of a three-phase load fail open, as discussed earlier:

3-phase, 3-wire "Y" connection



When we measure voltage and current in three-phase systems, we need to be specific as to *where* we're measuring. *Line voltage* refers to the amount of voltage measured between any two line conductors in a balanced three-phase system. With the above circuit, the line voltage is roughly 208 volts. *Phase voltage* refers to the voltage measured across any one component (source winding or load impedance) in a balanced three-phase source or load. For the circuit shown above, the phase voltage is 120 volts. The terms *line current* and *phase current* follow the same logic: the former referring to current through any one line conductor, and the latter to current through any one component.

Y-connected sources and loads always have line voltages greater than phase voltages, and line currents equal to phase currents. If the Y-connected source or load is balanced, the line voltage will be equal to the phase voltage times the square root of 3:

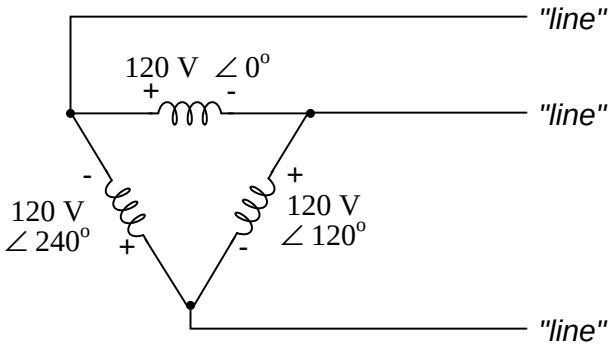
For "Y" circuits:

$E_{line} = \sqrt{3}\ E_{phase}$

$I_{line} = I_{phase}$

However, the "Y" configuration is not the only valid one for connecting three-phase voltage source or load elements together. Another configuration is known as the "Delta," for its geometric resemblance to the Greek letter of the same name (Δ). Take close notice of the polarity for each winding in the drawing below:

3-phase, 3-wire Δ connection



At first glance it seems as though three voltage sources like this would create a short-circuit, electrons flowing around the triangle with nothing but the internal impedance of the windings to hold them back. Due to the phase angles of these three voltage sources, however, this is not the case.

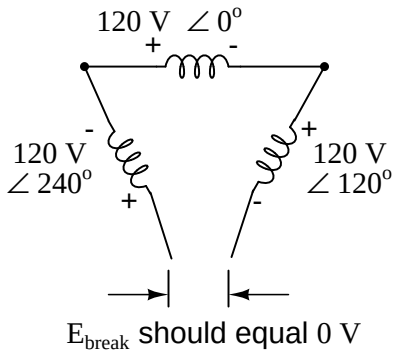
One quick check of this is to use Kirchhoff's Voltage Law to see if the three voltages around the loop add up to zero. If they do, then there will be no voltage available to push current around and around that loop, and consequently there will be no circulating current. Starting with the top winding and progressing counter-clockwise, our KVL expression looks something like this:

$$(120\ \text{V} \angle 0^\circ) + (120\ \text{V} \angle 240^\circ) + (120\ \text{V} \angle 120^\circ)$$

Does it all equal 0?

Yes!

Indeed, if we add these three vector quantities together, they do add up to zero. Another way to verify the fact that these three voltage sources can be connected together in a loop without resulting in circulating currents is to open up the loop at one junction point and calculate voltage across the break:



Starting with the right winding (120 V $\angle$ 120^o) and progressing counter-clockwise, our KVL equation looks like this:

$$(120\ \text{V} \angle 120^\text{o}) + (120 \angle 0^\text{o}) + (120\ \text{V} \angle 240^\text{o}) + E_{break} = 0$$

$$0 + E_{break} = 0$$

$$E_{break} = 0$$

Sure enough, there will be zero voltage across the break, telling us that no current will circulate within the triangular loop of windings when that connection is made complete.

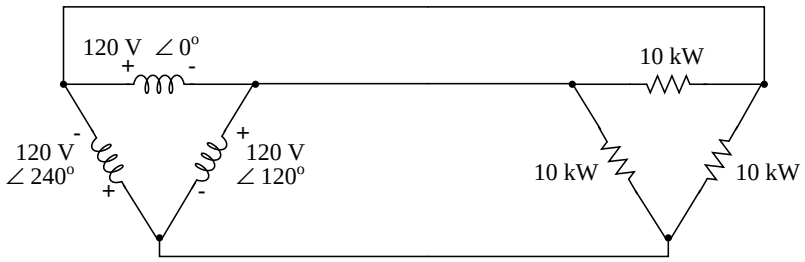
Having established that a Δ-connected three-phase voltage source will not burn itself to a crisp due to circulating currents, we turn to its practical use as a source of power in three-phase circuits. Because each pair of line conductors is connected directly across a single winding in a Δ circuit, the line voltage will be equal to the phase voltage. Conversely, because each line conductor attaches at a node between two windings, the line current will be the vector sum of the two joining phase currents. Not surprisingly, the resulting equations for a Δ configuration are as follows:

For Δ ("delta") circuits:

$$E_{line} = E_{phase}$$

$$I_{line} = \sqrt{3}\ I_{phase}$$

Let's see how this works in an example circuit:



With each load resistance receiving 120 volts from its respective phase winding at the source, the current in each phase of this circuit will be 83.33 amps:

$$I = \frac{P}{E}$$

$$I = \frac{10\ \text{kW}}{120\ \text{V}}$$

$I = 83.33$ A *(for each load resistor and source winding)*

$$I_{line} = \sqrt{3}\ I_{phase}$$

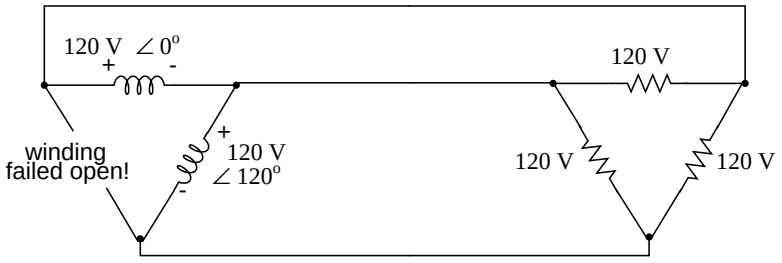
$$I_{line} = \sqrt{3}\ (83.33\ \text{A})$$

$$I_{line} = 144.34\ \text{A}$$

So, the each line current in this three-phase power system is equal to 144.34 amps, substantially more than the line currents in the Y-connected system we looked at earlier. One might wonder if we've lost all the advantages of three-phase power here, given the fact that we have such greater conductor currents, necessitating thicker, more costly wire. The answer is no. Although this circuit would require three number 1 gage copper conductors (at 1000 feet of distance between source and load this equates to a little over 750 pounds of copper for the whole system), it is still less than the 1000+ pounds of copper required for a single-phase system delivering the same power (30 kW) at the same voltage (120 volts conductor-to-conductor).

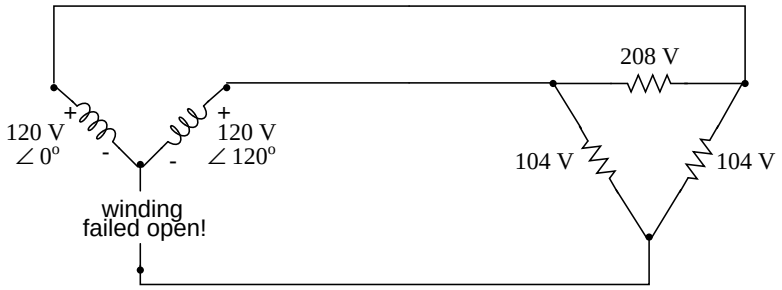
One distinct advantage of a Δ-connected system is its lack of a neutral wire. With a Y-connected system, a neutral wire was needed in case one of the phase loads were to fail open (or be turned off), in order to keep the phase voltages at the load from changing. This is not necessary (or even possible!) in a Δ-connected circuit. With each load phase element directly connected across a respective source phase winding, the phase voltage will be constant regardless of open failures in the load elements.

Perhaps the greatest advantage of the Δ-connected source is its fault tolerance. It is possible for one of the windings in a Δ-connected three-phase source to fail open without affecting load voltage or current!

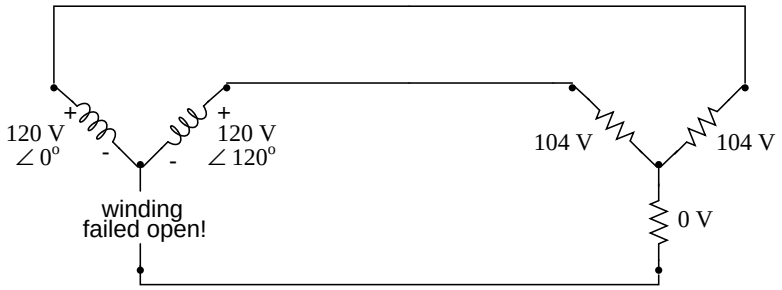


Even with the winding failure in the source, line voltage will still be 120 V and load phase voltage will still be 120 V. The only difference will be extra current in the remaining functioning source windings.

The only consequence of a source winding failing open for a Δ-connected source is increased phase current in the remaining windings. Compare this fault tolerance with a Y-connected system suffering an open source winding:



With a Δ-connected load, two of the resistances suffer reduced voltage while one remains at the original line voltage, 208. A Y-connected load suffers an even worse fate with the same winding failure in a Y-connected source:



In this case, two load resistances suffer reduced voltage while the third loses supply voltage completely! For this reason, Δ-connected sources are preferred for reliability. However, if dual voltages are needed (e.g. 120/208) or preferred for lower line currents, Y-connected systems are the configuration of choice.

- **REVIEW:**

- The conductors connected to the three points of a three-phase source or load are called *lines*.
- The three components comprising a three-phase source or load are called *phases*.
- *Line voltage* is the voltage measured between any two lines in a three-phase circuit.
- *Phase voltage* is the voltage measured across a single component in a three-phase source or load.
- *Line current* is the current through any one line between a three-phase source and load.
- *Phase current* is the current through any one component comprising a three-phase source or load.
- In balanced "Y" circuits, line voltage is equal to phase voltage times the square root of 3, while line current is equal to phase current.

  For "Y" circuits:

  $E_{line} = \sqrt{3}\ E_{phase}$

- $I_{line} = I_{phase}$
- In balanced Δ circuits, line voltage is equal to phase voltage, while line current is equal to phase current times the square root of 3.

  For Δ ("delta") circuits:

  $E_{line} = E_{phase}$

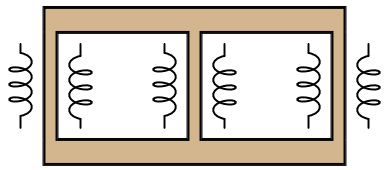
- $I_{line} = \sqrt{3}\ I_{phase}$
- Δ-connected three-phase voltage sources give greater reliability in the event of winding failure than Y-connected sources. However, Y-connected sources can deliver the same amount of power with less line current than Δ-connected sources.

10.6 Three-phase transformer circuits

Since three-phase is used so often for power distribution systems, it makes sense that we would need three-phase transformers to be able to step voltages up or down. This is only partially true, as regular single-phase transformers can be ganged together to transform power between two three-phase systems in a variety of configurations, eliminating the requirement for a special three-phase transformer. However, special three-phase transformers are built for those tasks, and are able to perform with less material requirement, less size, and less weight from their modular counterparts.

A three-phase transformer is made of three sets of primary and secondary windings, each set wound around one leg of an iron core assembly. Essentially it looks like three single-phase transformers sharing a joined core:

Three-phase transformer core



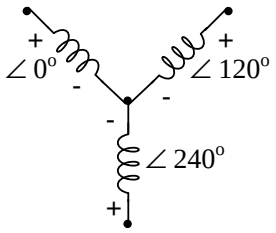
Those sets of primary and secondary windings will be connected in either Δ or Y configurations to form a complete unit. The various combinations of ways that these windings can be connected together in will be the focus of this section.

Whether the winding sets share a common core assembly or each winding pair is a separate transformer, the winding connection options are the same:

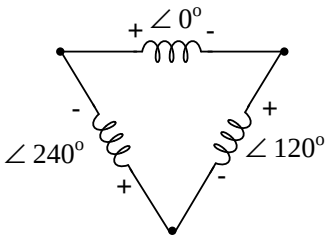
- **Primary - Secondary**
- Y - Y
- Y - Δ
- Δ - Y
- Δ - Δ

The reasons for choosing a Y or Δ configuration for transformer winding connections are the same as for any other three-phase application: Y connections provide the opportunity for multiple voltages, while Δ connections enjoy a higher level of reliability (if one winding fails open, the other two can still maintain full line voltages to the load).

Probably the most important aspect of connecting three sets of primary and secondary windings together to form a three-phase transformer bank is attention to proper winding phasing (the dots used to denote "polarity" of windings). Remember the proper phase relationships between the phase windings of Δ and Y:

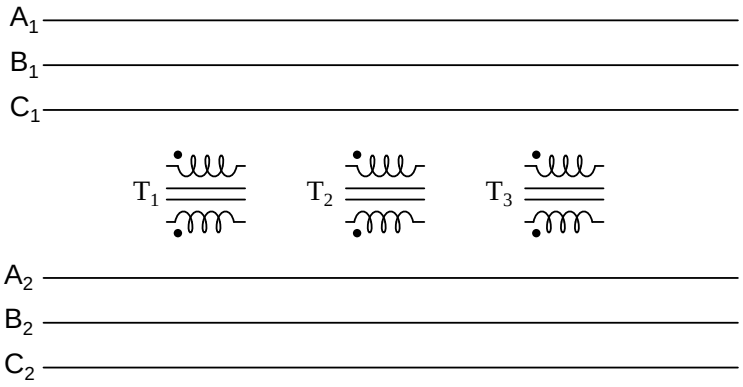


With these phase angles, the center point of the Y must tie either all "-" or all "+" winding ends together.

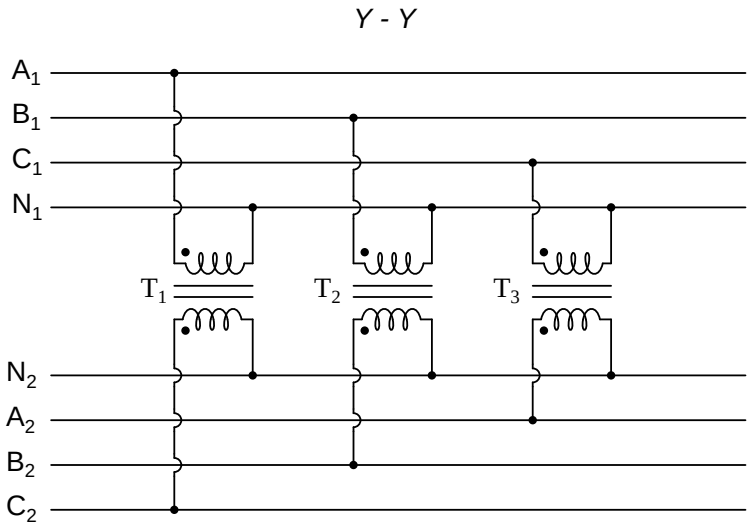


With these phase angles, the winding polarities must stack together in a complementary manner (+ to -).

Getting this phasing correct when the windings aren't shown in regular Y or Δ configuration can be tricky. Let me illustrate:

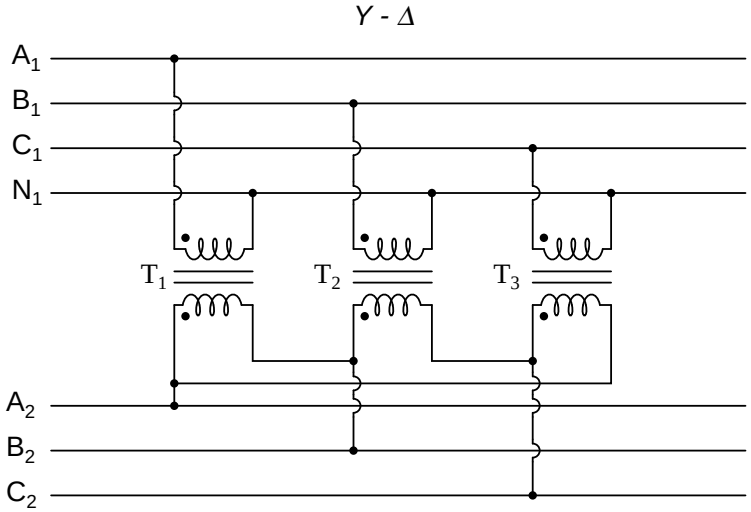


Three individual transformers are to be connected together to transform power from one three-phase system to another. First, I'll show the wiring connections for a Y-Y configuration:



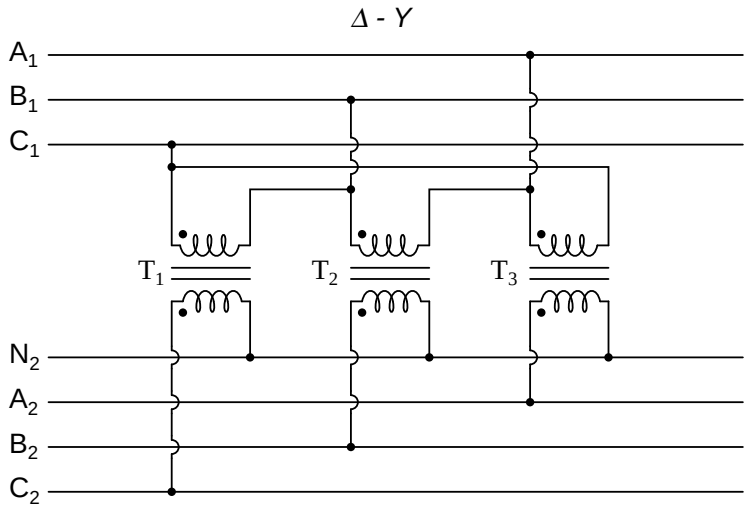
Note how all the winding ends marked with dots are connected to their respective phases A, B, and C, while the non-dot ends are connected together to form the centers of each "Y". Having both primary and secondary winding sets connected in "Y" formations allows for the use of neutral conductors (N_1 and N_2) in each power system.

Now, we'll take a look at a Y-Δ configuration:



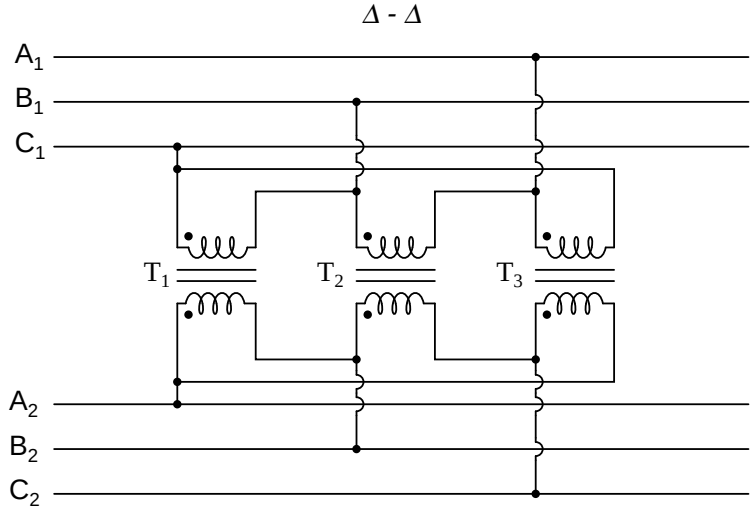
Note how the secondary windings (bottom set) are connected in a chain, the "dot"' side of one winding connected to the "non-dot" side of the next, forming the Δ loop. At every connection point between pairs of windings, a connection is made to a line of the second power system (A, B, and C).

Now, let's examine a Δ-Y system:



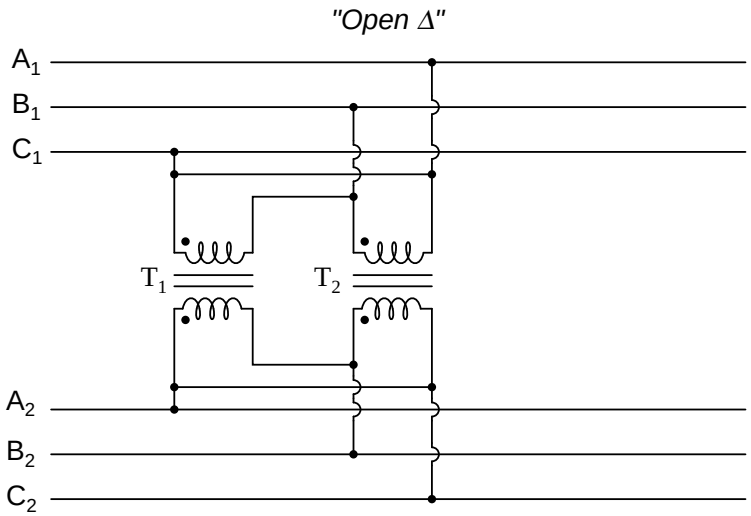
Such a configuration would allow for the provision of multiple voltages (line-to-line or line-to-neutral) in the second power system, from a source power system having no neutral.

And finally, we turn to the Δ-Δ configuration:



When there is no need for a neutral conductor in the secondary power system, Δ-Δ connection schemes are preferred because of the inherent reliability of the Δ configuration.

Considering that a Δ configuration can operate satisfactorily missing one winding, some power system designers choose to create a three-phase transformer bank with only two transformers, representing a Δ-Δ configuration with a missing winding in both the primary and secondary sides:



This configuration is called "V" or "Open-Δ." Of course, each of the two transformers have to be oversized to handle the same amount of power as three in a standard Δ configuration, but the overall size, weight, and cost advantages are often worth it. Bear in mind, however, that with one winding set missing from the Δ shape, this system no longer provides the fault tolerance of a normal Δ-Δ system. If one of the two transformers were to fail, the load voltage and current would definitely be affected.

The following photograph shows a bank of step-up transformers at the Grand Coulee hydroelectric dam in Washington state. Several transformers (green in color) may be seen from this vantage point, and they are grouped in threes: three transformers per hydroelectric generator, wired together in some form of three-phase configuration. The photograph doesn't reveal the primary winding connections, but it appears the secondaries are connected in a Y configuration, being that there is only one large high-voltage insulator protruding from each transformer. This suggests the other side of each transformer's secondary winding is at or near ground potential, which could only be true in a Y system. The building to the left is the powerhouse, where the generators and turbines are housed. On the right, the sloping concrete wall is the downstream face of the dam:

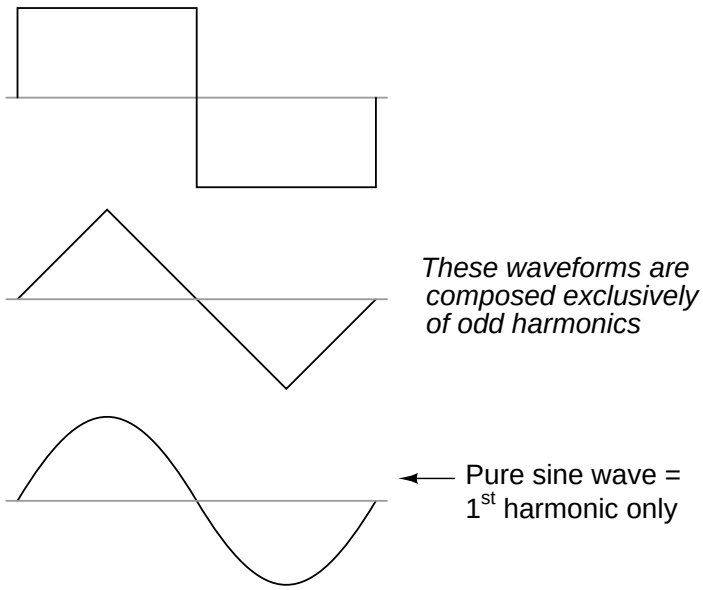
10.7 Harmonics in polyphase power systems

In the chapter on mixed-frequency signals, we explored the concept of *harmonics* in AC systems: frequencies that are integer multiples of the fundamental source frequency. With AC power systems where the source voltage waveform coming from an AC generator (alternator) is supposed to be a single-frequency sine wave, undistorted, there should be no harmonic content . . . ideally.

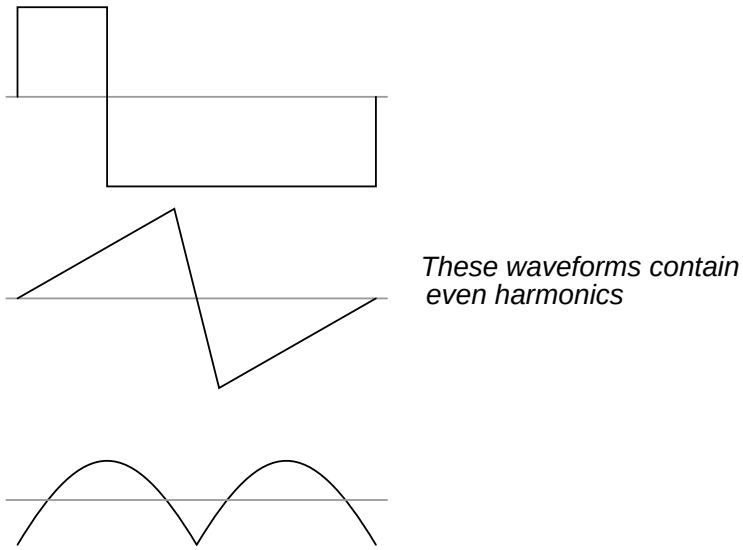
This would be true were it not for *nonlinear components*. Nonlinear components draw current disproportionately with respect to the source voltage, causing non-sinusoidal current waveforms. Examples of nonlinear components include gas-discharge lamps, semiconductor power-control devices (diodes, transistors, SCRs, TRIACs), transformers (primary winding magnetization current is usually non-sinusoidal due to the B/H saturation curve of the core), and electric motors (again, when magnetic fields within the motor's core operate near saturation levels). Even incandescent lamps generate slightly nonsinusoidal currents, as the filament resistance changes throughout the cycle due to rapid fluctuations in temperature. As we learned in the mixed-frequency chapter, *any* distortion of an otherwise sine-wave shaped waveform constitutes the presence of harmonic frequencies.

When the nonsinusoidal waveform in question is symmetrical above and below its average centerline, the harmonic frequencies will be odd integer multiples of the fundamental source frequency only, with no even integer multiples. Most nonlinear loads produce current waveforms like this, and so even-numbered harmonics (2nd, 4th, 6th, 8th, 10th, 12th, etc.) are absent or only minimally present in most AC power systems.

Examples of symmetrical waveforms – odd harmonics only:



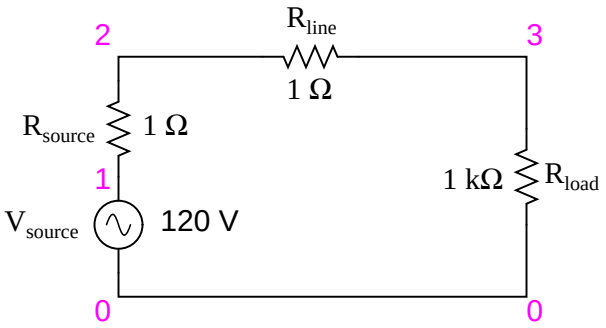
Examples of nonsymmetrical waveforms – even harmonics present:



Even though half of the possible harmonic frequencies are eliminated by the typically symmetrical distortion of nonlinear loads, the odd harmonics can still cause problems. Some of these problems are general to all power systems, single-phase or otherwise. Transformer overheating due to eddy current losses, for example, can occur in *any* AC power system where there is significant harmonic content. However, there are some problems caused by harmonic currents that are specific to polyphase power systems, and it is these problems to which this section is specifically devoted.

It is helpful to be able to simulate nonlinear loads in SPICE so as to avoid a lot of complex

mathematics and obtain a more intuitive understanding of harmonic effects. First, we'll begin our simulation with a very simple AC circuit: a single sine-wave voltage source with a purely linear load and all associated resistances:



The R_{source} and R_{line} resistances in this circuit do more than just mimic the real world: they also provide convenient shunt resistances for measuring currents in the SPICE simulation: by reading voltage across a 1 Ω resistance, you obtain a direct indication of current through it, since E = IR.

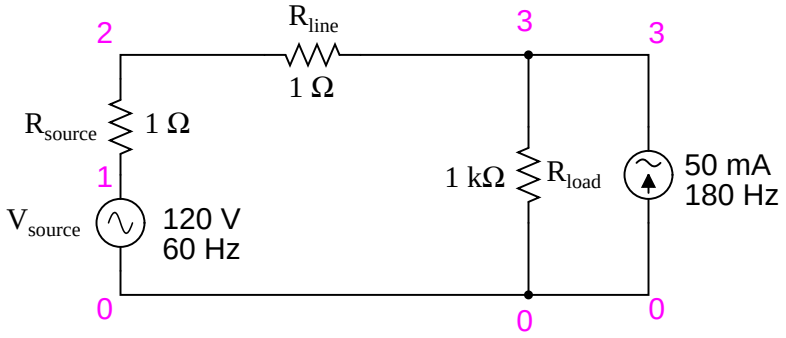
A SPICE simulation of this circuit with Fourier analysis on the voltage measured across R_{line} should show us the harmonic content of this circuit's line current. Being completely linear in nature, we should expect no harmonics other than the 1st (fundamental) of 60 Hz, assuming a 60 Hz source:

```
linear load simulation
vsource 1 0 sin(0 120 60 0 0)
rsource 1 2 1
rline 2 3 1
rload 3 0 1k
.options itl5=0
.tran 0.5m 30m 0 1u
.plot tran v(2,3)
.four 60 v(2,3)
.end

fourier components of transient response v(2,3)
dc component =   4.028E-12
harmonic frequency   fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.198E-01     1.000000   -72.000       0.000
2      1.200E+02   5.793E-12     0.000000    51.122     123.122
3      1.800E+02   7.407E-12     0.000000   -34.624      37.376
4      2.400E+02   9.056E-12     0.000000     4.267      76.267
5      3.000E+02   1.651E-11     0.000000   -83.461     -11.461
6      3.600E+02   3.931E-11     0.000000    36.399     108.399
7      4.200E+02   2.338E-11     0.000000   -41.343      30.657
8      4.800E+02   4.716E-11     0.000000    53.324     125.324
9      5.400E+02   3.453E-11     0.000000    21.691      93.691
total harmonic distortion =      0.000000  percent
```

A `.plot` command appears in the SPICE netlist, and normally this would result in a sine-wave graph output. In this case, however, I've purposely omitted the waveform display for brevity's sake – the `.plot` command is in the netlist simply to satisfy a quirk of SPICE's Fourier transform function.

No discrete Fourier transform is perfect, and so we see very small harmonic currents indicated (in the pico-amp range!) for all frequencies up to the 9th harmonic, which is as far as SPICE goes in performing Fourier analysis. We show 0.1198 amps (1.198E-01) for the "fourier component" of the 1st harmonic, or the fundamental frequency, which is our expected load current: about 120 mA, given a source voltage of 120 volts and a load resistance of 1 kΩ.

Next, I'd like to simulate a nonlinear load so as to generate harmonic currents. This can be done in two fundamentally different ways. One way is to design a load using nonlinear components such as diodes or other semiconductor devices which as easy to simulate with SPICE. Another is to add some AC current sources in parallel with the load resistor. The latter method is often preferred by engineers for simulating harmonics, since current sources of known value lend themselves better to mathematical network analysis than components with highly complex response characteristics. Since we're letting SPICE do all the math work, the complexity of a semiconductor component would cause no trouble for us, but since current sources can be fine-tuned to produce any arbitrary amount of current (a convenient feature), I'll choose the latter approach:




```
Nonlinear load simulation
vsource 1 0 sin(0 120 60 0 0)
rsource 1 2 1
rline 2 3 1
rload 3 0 1k
i3har 3 0 sin(0 50m 180 0 0)
.options itl5=0
.tran 0.5m 30m 0 1u
.plot tran v(2,3)
.four 60 v(2,3)
.end
```

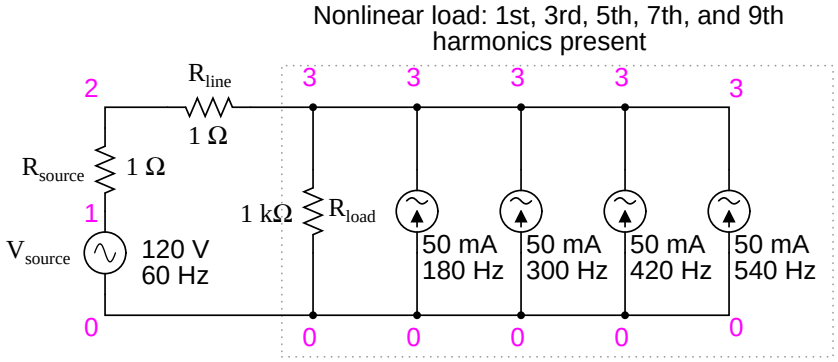
In this circuit, we have a current source of 50 mA magnitude and a frequency of 180 Hz, which is three times the source frequency of 60 Hz. Connected in parallel with the 1 kΩ load resistor, its current will add with the resistor's to make a nonsinusoidal total line current. I'll show the waveform plot here just so you can see the effects of this 3rd-harmonic current on the total current, which would ordinarily be a plain sine wave:

```
time        v(2,3)
- - - - - - - - - - -  - - - - - - - - - - - - - - - - - - - - - - - -
0.000E+00  0.000E+00 .        .             *             .        .
5.000E-04  4.918E-02 .        .             .      *      .        .
1.000E-03  8.924E-02 .        .             .           * .        .
1.500E-03  1.137E-01 .        .             .             . *      .
2.000E-03  1.204E-01 .        .             .             .  *     .
2.500E-03  1.123E-01 .        .             .             . *      .
3.000E-03  9.595E-02 .        .             .            *.        .
3.500E-03  7.962E-02 .        .             .          *  .        .
4.000E-03  7.051E-02 .        .             .         *   .        .
4.500E-03  7.242E-02 .        .             .         *   .        .
5.000E-03  8.457E-02 .        .             .           * .        .
5.500E-03  1.018E-01 .        .             .             *        .
6.000E-03  1.163E-01 .        .             .             . *      .
6.500E-03  1.201E-01 .        .             .             .  *     .
7.000E-03  1.075E-01 .        .             .             .*       .
7.500E-03  7.738E-02 .        .             .          *  .        .
8.000E-03  3.338E-02 .        .             .    *        .        .
8.500E-03 -1.687E-02 .        .           * .             .        .
9.000E-03 -6.394E-02 .        .    *        .             .        .
9.500E-03 -9.932E-02 .        *             .             .        .
1.000E-02 -1.179E-01 .      * .             .             .        .
1.050E-02 -1.191E-01 .     *  .             .             .        .
1.100E-02 -1.074E-01 .       *.             .             .        .
1.150E-02 -9.010E-02 .        .*            .             .        .
1.200E-02 -7.551E-02 .        .  *          .             .        .
1.250E-02 -6.986E-02 .        .   *         .             .        .
1.300E-02 -7.551E-02 .        .  *          .             .        .
1.350E-02 -9.010E-02 .        .*            .             .        .
1.400E-02 -1.074E-01 .       *.             .             .        .
1.450E-02 -1.191E-01 .     *  .             .             .        .
1.500E-02 -1.179E-01 .      * .             .             .        .
1.550E-02 -9.932E-02 .        *             .             .        .
1.600E-02 -6.394E-02 .        .    *        .             .        .
1.650E-02 -1.687E-02 .        .           * .             .        .
- - - - - - - - - - -  - - - - - - - - - - - - - - - - - - - - - - - -

fourier components of transient response v(2,3)
dc component =   1.349E-11
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.198E-01     1.000000   -72.000       0.000
2      1.200E+02   1.609E-11     0.000000    67.570     139.570
3      1.800E+02   4.990E-02     0.416667   144.000     216.000
4      2.400E+02   1.074E-10     0.000000  -169.546     -97.546
```

```
5       3.000E+02   3.871E-11     0.000000   169.582     241.582
6       3.600E+02   5.736E-11     0.000000   140.845     212.845
7       4.200E+02   8.407E-11     0.000000   177.071     249.071
8       4.800E+02   1.329E-10     0.000000   156.772     228.772
9       5.400E+02   2.619E-10     0.000000   160.498     232.498
total harmonic distortion =     41.666663  percent
```

In the Fourier analysis, the mixed frequencies are unmixed and presented separately. Here we see the same 0.1198 amps of 60 Hz (fundamental) current as we did in the first simulation, but appearing in the 3rd harmonic row we see 49.9 mA: our 50 mA, 180 Hz current source at work. Why don't we see the entire 50 mA through the line? Because that current source is connected across the 1 kΩ load resistor, so some of its current is shunted through the load and never goes through the line back to the source. It's an inevitable consequence of this type of simulation, where one part of the load is "normal" (a resistor) and the other part is imitated by a current source.

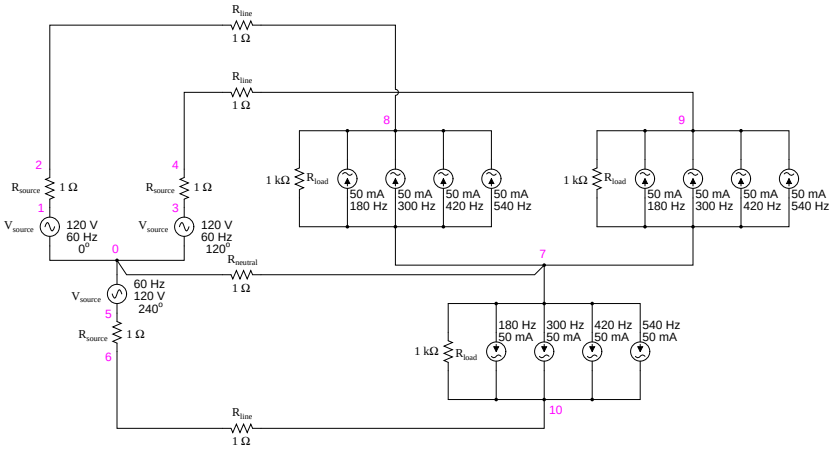
If we were to add more current sources to the "load," we would see further distortion of the line current waveform from the ideal sine-wave shape, and each of those harmonic currents would appear in the Fourier analysis breakdown:




```
Nonlinear load simulation
vsource 1 0 sin(0 120 60 0 0)
rsource 1 2 1
rline 2 3 1
rload 3 0 1k
i3har 3 0 sin(0 50m 180 0 0)
i5har 3 0 sin(0 50m 300 0 0)
i7har 3 0 sin(0 50m 420 0 0)
i9har 3 0 sin(0 50m 540 0 0)
.options itl5=0
.tran 0.5m 30m 0 1u
.plot tran v(2,3)
.four 60 v(2,3)
.end

fourier components of transient response v(2,3)
```

```
dc component =   6.299E-11
harmonic  frequency  fourier    normalized    phase    normalized
no          (hz)     component   component    (deg)    phase (deg)
1      6.000E+01   1.198E-01     1.000000   -72.000       0.000
2      1.200E+02   1.900E-09     0.000000   -93.908     -21.908
3      1.800E+02   4.990E-02     0.416667   144.000     216.000
4      2.400E+02   5.469E-09     0.000000  -116.873     -44.873
5      3.000E+02   4.990E-02     0.416667     0.000      72.000
6      3.600E+02   6.271E-09     0.000000    85.062     157.062
7      4.200E+02   4.990E-02     0.416666  -144.000     -72.000
8      4.800E+02   2.742E-09     0.000000   -38.781      33.219
9      5.400E+02   4.990E-02     0.416666    72.000     144.000
total harmonic distortion =     83.333296  percent
```

As you can see from the Fourier analysis, every harmonic current source is equally represented in the line current, at 49.9 mA each. So far, this is just a single-phase power system simulation. Things get more interesting when we make it a three-phase simulation. Two Fourier analyses will be performed: one for the voltage across a line resistor, and one for the voltage across the neutral resistor. As before, reading voltages across fixed resistances of 1 Ω each gives direct indications of current through those resistors:




```
Y-Y source/load 4-wire system with harmonics
*
* phase1 voltage source and r (120 v /_ 0 deg)
vsource1 1 0 sin(0 120 60 0 0)
rsource1 1 2 1
*
* phase2 voltage source and r (120 v /_ 120 deg)
vsource2 3 0 sin(0 120 60 5.55555m 0)
rsource2 3 4 1
*
```

```
* phase3 voltage source and r (120 v /_ 240 deg)
vsource3 5 0 sin(0 120 60 11.1111m 0)
rsource3 5 6 1
*
* line and neutral wire resistances
rline1 2 8 1
rline2 4 9 1
rline3 6 10 1
rneutral 0 7 1
*
* phase 1 of load
rload1 8 7 1k
i3har1 8 7 sin(0 50m 180 0 0)
i5har1 8 7 sin(0 50m 300 0 0)
i7har1 8 7 sin(0 50m 420 0 0)
i9har1 8 7 sin(0 50m 540 0 0)
*
* phase 2 of load
rload2 9 7 1k
i3har2 9 7 sin(0 50m 180 5.55555m 0)
i5har2 9 7 sin(0 50m 300 5.55555m 0)
i7har2 9 7 sin(0 50m 420 5.55555m 0)
i9har2 9 7 sin(0 50m 540 5.55555m 0)
*
* phase 3 of load
rload3 10 7 1k
i3har3 10 7 sin(0 50m 180 11.1111m 0)
i5har3 10 7 sin(0 50m 300 11.1111m 0)
i7har3 10 7 sin(0 50m 420 11.1111m 0)
i9har3 10 7 sin(0 50m 540 11.1111m 0)
*
* analysis stuff
.options itl5=0
.tran 0.5m 100m 12m 1u
.plot tran v(2,8)
.four 60 v(2,8)
.plot tran v(0,7)
.four 60 v(0,7)
.end
```

Fourier analysis of line current:

```
fourier components of transient response v(2,8)
dc component =  -6.404E-12
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.198E-01     1.000000     0.000       0.000
```

```
2      1.200E+02   2.218E-10     0.000000   172.985     172.985
3      1.800E+02   4.975E-02     0.415423     0.000       0.000
4      2.400E+02   4.236E-10     0.000000   166.990     166.990
5      3.000E+02   4.990E-02     0.416667     0.000       0.000
6      3.600E+02   1.877E-10     0.000000  -147.146    -147.146
7      4.200E+02   4.990E-02     0.416666     0.000       0.000
8      4.800E+02   2.784E-10     0.000000  -148.811    -148.811
9      5.400E+02   4.975E-02     0.415422     0.000       0.000
total harmonic distortion =     83.209009  percent
```

Fourier analysis of neutral current:

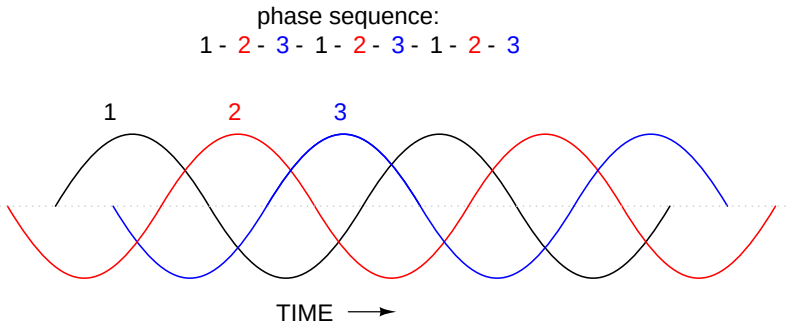
```
fourier components of transient response v(0,7)
dc component =   1.819E-10
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   4.337E-07     1.000000    60.018       0.000
2      1.200E+02   1.869E-10     0.000431    91.206      31.188
3      1.800E+02   1.493E-01  344147.7638  -180.000    -240.018
4      2.400E+02   1.257E-09     0.002898   -21.103     -81.121
5      3.000E+02   9.023E-07     2.080596   119.981      59.963
6      3.600E+02   3.396E-10     0.000783    15.882     -44.136
7      4.200E+02   1.264E-06     2.913955    59.993      -0.025
8      4.800E+02   5.975E-10     0.001378    35.584     -24.434
9      5.400E+02   1.493E-01  344147.4889  -179.999    -240.017
```

This is a balanced Y-Y power system, each phase identical to the single-phase AC system simulated earlier. Consequently, it should come as no surprise that the Fourier analysis for line current in one phase of the 3-phase system is nearly identical to the Fourier analysis for line current in the single-phase system: a fundamental (60 Hz) line current of 0.1198 amps, and odd harmonic currents of approximately 50 mA each.

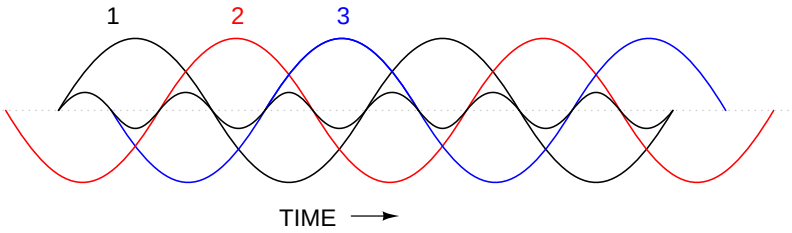
What should be surprising here is the analysis for the neutral conductor's current, as determined by the voltage drop across the $R_{neutral}$ resistor between SPICE nodes 0 and 7. In a balanced 3-phase Y load, we would expect the neutral current to be zero. Each phase current – which by itself would go through the neutral wire back to the supplying phase on the source Y – should cancel each other in regard to the neutral conductor because they're all the same magnitude and all shifted 120^o apart. In a system with no harmonic currents, this *is* what happens, leaving zero current through the neutral conductor. However, we cannot say the same for *harmonic* currents in the same system.

Note that the fundamental frequency (60 Hz, or the 1st harmonic) current is virtually absent from the neutral conductor. Our Fourier analysis shows only 0.4337 μA of 1st harmonic when reading voltage across $R_{neutral}$. The same may be said about the 5th and 7th harmonics, both of those currents having negligible magnitude. In contrast, the 3rd and 9th harmonics are strongly represented within the neutral conductor, with 149.3 mA (1.493E-01 volts across 1 Ω) each! This is very nearly 150 mA, or three times the current sources' values, individually. With three sources per harmonic frequency in the load, it appears our 3rd and 9th harmonic currents in each phase are *adding* to form the neutral current.

This is exactly what's happening, though it might not be apparent why this is so. The key to understanding this is made clear in a time-domain graph of phase currents. Examine this plot of balanced phase currents over time, with a phase sequence of 1-2-3:

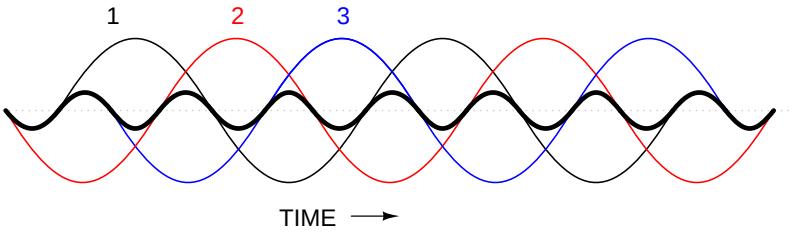


With the three fundamental waveforms equally shifted across the time axis of the graph, it is easy to see how they would cancel each other to give a resultant current of zero in the neutral conductor. Let's consider, though, what a 3rd harmonic waveform for phase 1 would look like superimposed on the graph:



Observe how this harmonic waveform has the same phase relationship to the 2nd and 3rd fundamental waveforms as it does with the 1st: in each positive half-cycle of *any* of the fundamental waveforms, you will find exactly two positive half-cycles and one negative half-cycle of the harmonic waveform. What this means is that the 3rd-harmonic waveforms of three 120^o phase-shifted fundamental-frequency waveforms are actually *in phase* with each other. The phase shift figure of 120^o generally assumed in three-phase AC systems applies only to the fundamental frequencies, not to their harmonic multiples!

If we were to plot all three 3rd-harmonic waveforms on the same graph, we would see them precisely overlap and appear as a single, unified waveform (shown here in bold):



For the more mathematically inclined, this principle may be expressed symbolically. Suppose that **A** represents one waveform and **B** another, both at the same frequency, but shifted 120^o from

each other in terms of phase. Let's call the 3rd harmonic of each waveform **A'** and **B'**, respectively. The phase shift between **A'** and **B'** is not 120^o (that is the phase shift between **A** and **B**), but 3 times that, because the **A'** and **B'** waveforms alternate three times as fast as **A** and **B**. The shift between waveforms is only accurately expressed in terms of *phase angle* when the same angular velocity is assumed. When relating waveforms of different frequency, the most accurate way to represent phase shift is in terms of *time*; and the *time-shift* between **A'** and **B'** is equivalent to 120^o at a frequency three times lower, or 360^o at the frequency of **A'** and **B'**. A phase shift of 360^o is the same as a phase shift of 0^o, which is to say no phase shift at all. Thus, **A'** and **B'** must be in phase with each other:

Phase sequence = ***A-B-C***

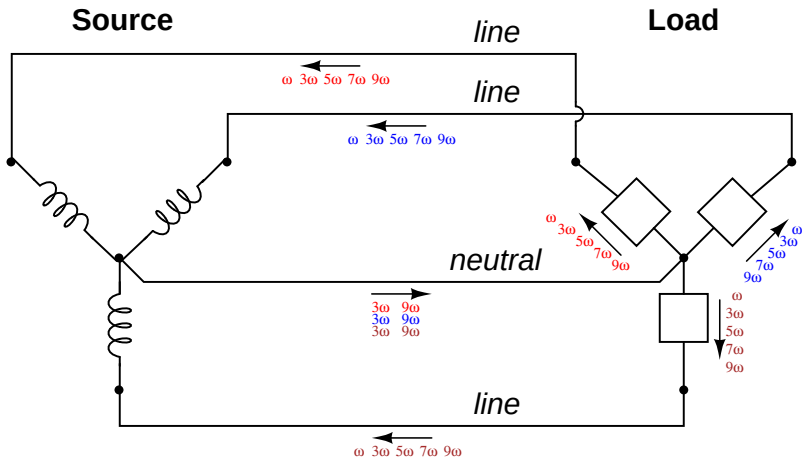
Fundamental	**A** 0^o	**B** 120^o	**C** 240^o
3rd harmonic	**A'** 3×0^o (0^o)	**B'** 3×120^o $(360^o = 0^o)$	**C'** 3×240^o $(720^o = 0^o)$

This characteristic of the 3rd harmonic in a three-phase system also holds true for any integer multiples of the 3rd harmonic. So, not only are the 3rd harmonic waveforms of each fundamental waveform in phase with each other, but so are the 6th harmonics, the 9th harmonics, the 12th harmonics, the 15th harmonics, the 18th harmonics, the 21st harmonics, and so on. Since only odd harmonics appear in systems where waveform distortion is symmetrical about the centerline – and most nonlinear loads create symmetrical distortion – even-numbered multiples of the 3rd harmonic (6th, 12th, 18th, etc.) are generally not significant, leaving only the odd-numbered multiples (3rd, 9th, 21st, etc.) to significantly contribute to neutral currents.

In polyphase power systems with some number of phases other than three, this effect occurs with harmonics of the same multiple. For instance, the harmonic currents that add in the neutral conductor of a star-connected 4-phase system where the phase shift between fundamental waveforms is 90^o would be the 4th, 8th, 12th, 16th, 20th, and so on.

Due to their abundance and significance in three-phase power systems, the 3rd harmonic and its multiples have their own special name: *triplen harmonics*. All triplen harmonics add with each other in the neutral conductor of a 4-wire Y-connected load. In power systems containing substantial nonlinear loading, the triplen harmonic currents may be of great enough magnitude to cause neutral conductors to overheat. This is very problematic, as other safety concerns prohibit neutral conductors from having overcurrent protection, and thus there is no provision for automatic interruption of these high currents.

The following illustration shows how triplen harmonic currents created at the load add within the neutral conductor. The symbol "ω" is used to represent angular velocity, and is mathematically equivalent to 2πf. So, "ω" represents the fundamental frequency, "3ω " represents the 3rd harmonic, "5ω" represents the 5th harmonic, and so on:



Triplen harmonic currents add in neutral conductor

In an effort to mitigate these additive triplen currents, one might be tempted to remove the neutral wire entirely. If there is no neutral wire in which triplen currents can flow together, then they won't, right? Unfortunately, doing so just causes a different problem: the load's "Y" center-point will no longer be at the same potential as the source's, meaning that each phase of the load will receive a different voltage than what is produced by the source. We'll re-run the last SPICE simulation without the 1 Ω $R_{neutral}$ resistor and see what happens:

```
Y-Y source/load (no neutral) with harmonics
*
* phase1 voltage source and r (120 v /_ 0 deg)
vsource1 1 0 sin(0 120 60 0 0)
rsource1 1 2 1
*
* phase2 voltage source and r (120 v /_ 120 deg)
vsource2 3 0 sin(0 120 60 5.55555m 0)
rsource2 3 4 1
*
* phase3 voltage source and r (120 v /_ 240 deg)
vsource3 5 0 sin(0 120 60 11.1111m 0)
rsource3 5 6 1
*
* line resistances
rline1 2 8 1
rline2 4 9 1
rline3 6 10 1
*
* phase 1 of load
rload1 8 7 1k
i3har1 8 7 sin(0 50m 180 0 0)
i5har1 8 7 sin(0 50m 300 0 0)
```

```
i7har1 8 7 sin(0 50m 420 0 0)
i9har1 8 7 sin(0 50m 540 0 0)
*
* phase 2 of load
rload2 9 7 1k
i3har2 9 7 sin(0 50m 180 5.55555m 0)
i5har2 9 7 sin(0 50m 300 5.55555m 0)
i7har2 9 7 sin(0 50m 420 5.55555m 0)
i9har2 9 7 sin(0 50m 540 5.55555m 0)
*
* phase 3 of load
rload3 10 7 1k
i3har3 10 7 sin(0 50m 180 11.1111m 0)
i5har3 10 7 sin(0 50m 300 11.1111m 0)
i7har3 10 7 sin(0 50m 420 11.1111m 0)
i9har3 10 7 sin(0 50m 540 11.1111m 0)
*
* analysis stuff
.options itl5=0
.tran 0.5m 100m 12m 1u
.plot tran v(2,8)
.four 60 v(2,8)
.plot tran v(0,7)
.four 60 v(0,7)
.plot tran v(8,7)
.four 60 v(8,7)
.end
```

Fourier analysis of line current:

```
fourier components of transient response v(2,8)
dc component =   5.423E-11
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.198E-01     1.000000     0.000       0.000
2      1.200E+02   2.388E-10     0.000000   158.016     158.016
3      1.800E+02   3.136E-07     0.000003   -90.009     -90.009
4      2.400E+02   5.963E-11     0.000000  -111.510    -111.510
5      3.000E+02   4.990E-02     0.416665     0.000       0.000
6      3.600E+02   8.606E-11     0.000000  -124.565    -124.565
7      4.200E+02   4.990E-02     0.416668     0.000       0.000
8      4.800E+02   8.126E-11     0.000000  -159.638    -159.638
9      5.400E+02   9.406E-07     0.000008   -90.005     -90.005
total harmonic distortion =     58.925539  percent
```

Fourier analysis of voltage between the two "Y" center-points:

```
fourier components of transient response v(0,7)
```

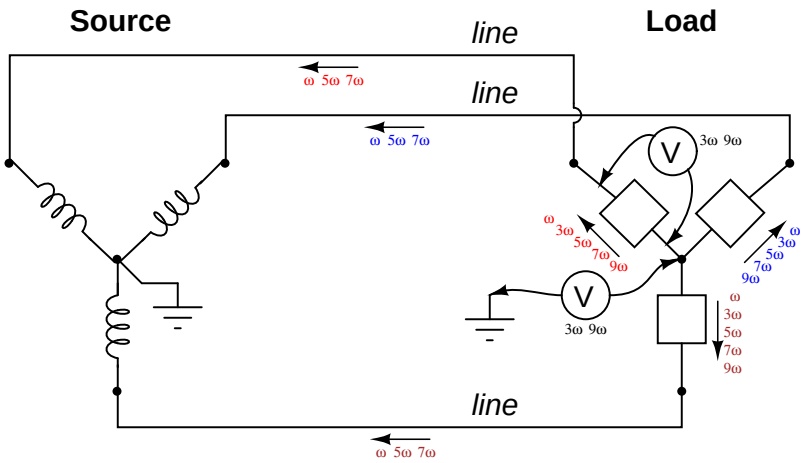
```
dc component =   6.093E-08
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.453E-04     1.000000    60.018       0.000
2      1.200E+02   6.263E-08     0.000431    91.206      31.188
3      1.800E+02   5.000E+01  344147.7879  -180.000    -240.018
4      2.400E+02   4.210E-07     0.002898   -21.103     -81.121
5      3.000E+02   3.023E-04     2.080596   119.981      59.963
6      3.600E+02   1.138E-07     0.000783    15.882     -44.136
7      4.200E+02   4.234E-04     2.913955    59.993      -0.025
8      4.800E+02   2.001E-07     0.001378    35.584     -24.434
9      5.400E+02   5.000E+01  344147.4728  -179.999    -240.017
total harmonic distortion =  ************  percent
```

Fourier analysis of load phase voltage:

```
fourier components of transient response v(8,7)
dc component =   6.070E-08
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.198E+02     1.000000     0.000       0.000
2      1.200E+02   6.231E-08     0.000000    90.473      90.473
3      1.800E+02   5.000E+01     0.417500  -180.000    -180.000
4      2.400E+02   4.278E-07     0.000000   -19.747     -19.747
5      3.000E+02   9.995E-02     0.000835   179.850     179.850
6      3.600E+02   1.023E-07     0.000000    13.485      13.485
7      4.200E+02   9.959E-02     0.000832   179.790     179.789
8      4.800E+02   1.991E-07     0.000000    35.462      35.462
9      5.400E+02   5.000E+01     0.417499  -179.999    -179.999
total harmonic distortion =     59.043467  percent
```

Strange things are happening, indeed. First, we see that the triplen harmonic currents (3rd and 9th) all but disappear in the lines connecting load to source. The 5th and 7th harmonic currents are present at their normal levels (approximately 50 mA), but the 3rd and 9th harmonic currents are of negligible magnitude. Second, we see that there is substantial harmonic voltage between the two "Y" center-points, between which the neutral conductor used to connect. According to SPICE, there is 50 volts of both 3rd and 9th harmonic frequency between these two points, which is definitely not normal in a linear (no harmonics), balanced Y system. Finally, the voltage as measured across one of the load's phases (between nodes 8 and 7 in the SPICE analysis) likewise shows strong triplen harmonic voltages of 50 volts each.

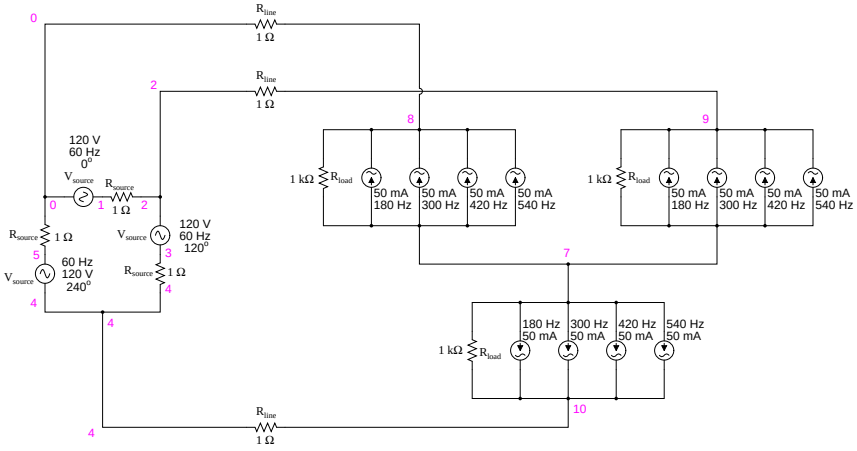
The following illustration is a graphical summary of the aforementioned effects:



Triplen voltages appear between Y centers.
Triplen voltages appear across load phases.
Non-triplen currents appear in line conductors.

In summary, removal of the neutral conductor leads to a "hot" center-point on the load "Y", and also to harmonic load phase voltages of equal magnitude, all comprised of triplen frequencies. In the previous simulation where we had a 4-wire, Y-connected system, the undesirable effect from harmonics was excessive neutral *current*, but at least each phase of the load received voltage nearly free of harmonics.

Since removing the neutral wire didn't seem to work in eliminating the problems caused by harmonics, perhaps switching to a Δ configuration will. Let's try a Δ source instead of a Y, keeping the load in its present Y configuration, and see what happens. The measured parameters will be line current (voltage across R_{line}, nodes 0 and 8), load phase voltage (nodes 8 and 7), and source phase current (voltage across R_{source}, nodes 1 and 2):




```
Delta-Y source/load with harmonics
*
```

```
* phase1 voltage source and r (120 v /_ 0 deg)
vsource1 1 0 sin(0 207.846 60 0 0)
rsource1 1 2 1
*
* phase2 voltage source and r (120 v /_ 120 deg)
vsource2 3 2 sin(0 207.846 60 5.55555m 0)
rsource2 3 4 1
*
* phase3 voltage source and r (120 v /_ 240 deg)
vsource3 5 4 sin(0 207.846 60 11.1111m 0)
rsource3 5 0 1
*
* line resistances
rline1 0 8 1
rline2 2 9 1
rline3 4 10 1
*
* phase 1 of load
rload1 8 7 1k
i3har1 8 7 sin(0 50m 180 9.72222m 0)
i5har1 8 7 sin(0 50m 300 9.72222m 0)
i7har1 8 7 sin(0 50m 420 9.72222m 0)
i9har1 8 7 sin(0 50m 540 9.72222m 0)
*
* phase 2 of load
rload2 9 7 1k
i3har2 9 7 sin(0 50m 180 15.2777m 0)
i5har2 9 7 sin(0 50m 300 15.2777m 0)
i7har2 9 7 sin(0 50m 420 15.2777m 0)
i9har2 9 7 sin(0 50m 540 15.2777m 0)
*
* phase 3 of load
rload3 10 7 1k
i3har3 10 7 sin(0 50m 180 4.16666m 0)
i5har3 10 7 sin(0 50m 300 4.16666m 0)
i7har3 10 7 sin(0 50m 420 4.16666m 0)
i9har3 10 7 sin(0 50m 540 4.16666m 0)
*
* analysis stuff
.options itl5=0
.tran 0.5m 100m 16m 1u
.plot tran v(0,8) v(8,7) v(1,2)
.four 60 v(0,8) v(8,7) v(1,2)
.end
```

Note: the following paragraph is for those curious readers who follow every detail of my SPICE

netlists. If you just want to find out what happens in the circuit, skip this paragraph! When simulating circuits having AC sources of differing frequency and differing phase, the only way to do it in SPICE is to set up the sources with a *delay time* or *phase offset* specified in seconds. Thus, the 0^o source has these five specifying figures: "(0 207.846 60 0 0)", which means 0 volts DC offset, 207.846 volts peak amplitude (120 times the square root of three, to ensure the load phase voltages remain at 120 volts each), 60 Hz, 0 time delay, and 0 damping factor. The 120^o phase-shifted source has these figures: "(0 207.846 60 5.55555m 0)", all the same as the first except for the time delay factor of 5.55555 milliseconds, or 1/3 of the full period of 16.6667 milliseconds for a 60 Hz waveform. The 240^o source must be time-delayed twice that amount, equivalent to a fraction of 240/360 of 16.6667 milliseconds, or 11.1111 milliseconds. This is for the Δ-connected source. The Y-connected load, on the other hand, requires a different set of time-delay figures for its harmonic current sources, because the phase voltages in a Y load are not in phase with the phase voltages of a Δ source. If Δ source voltages V_{AC}, V_{BA}, and V_{CB} are referenced at 0^o, 120^o, and 240^o, respectively, then "Y" load voltages V_A, V_B, and V_C will have phase angles of -30^o, 90^o, and 210^o, respectively. This is an intrinsic property of all Δ-Y circuits and not a quirk of SPICE. Therefore, when I specified the delay times for the harmonic sources, I had to set them at 15.2777 milliseconds (-30^o, or +330^o), 4.16666 milliseconds (90^o), and 9.72222 milliseconds (210^o). One final note: when delaying AC sources in SPICE, they don't "turn on" until their delay time has elapsed, which means any mathematical analysis up to that point in time will be in error. Consequently, I set the `.tran` transient analysis line to hold off analysis until 16 milliseconds after start, which gives all sources in the netlist time to engage before any analysis takes place.

The result of this analysis is almost as disappointing as the last. Line currents remain unchanged (the only substantial harmonic content being the 5th and 7th harmonics), and load phase voltages remain unchanged as well, with a full 50 volts of triplen harmonic (3rd and 9th) frequencies across each load component. Source phase current is a fraction of the line current, which should come as no surprise. Both 5th and 7th harmonics are represented there, with negligible triplen harmonics:

Fourier analysis of line current:

```
fourier components of transient response v(0,8)
dc component =  -6.850E-11
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.198E-01     1.000000   150.000       0.000
2      1.200E+02   2.491E-11     0.000000   159.723       9.722
3      1.800E+02   1.506E-06     0.000013     0.005    -149.996
4      2.400E+02   2.033E-11     0.000000    52.772     -97.228
5      3.000E+02   4.994E-02     0.416682    30.002    -119.998
6      3.600E+02   1.234E-11     0.000000    57.802     -92.198
7      4.200E+02   4.993E-02     0.416644   -29.998    -179.998
8      4.800E+02   8.024E-11     0.000000  -174.200    -324.200
9      5.400E+02   4.518E-06     0.000038  -179.995    -329.995
total harmonic distortion =     58.925038  percent
```

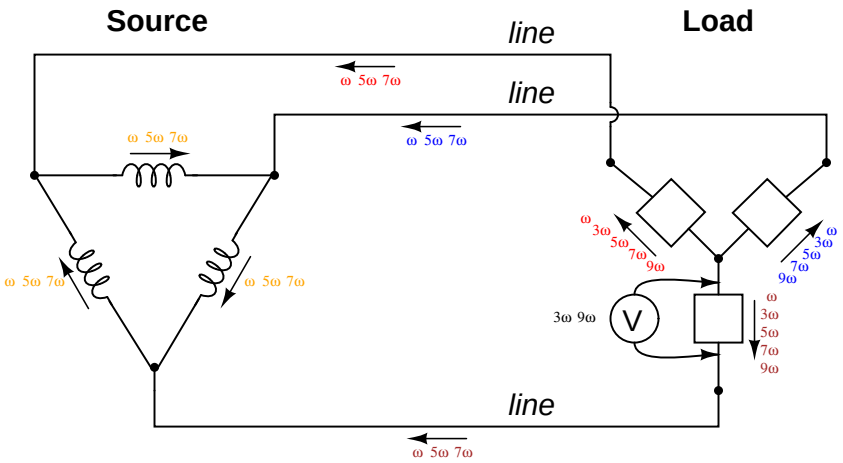
Fourier analysis of load phase voltage:

```
fourier components of transient response v(8,7)
dc component =   1.259E-08
```

```
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.198E+02     1.000000   150.000       0.000
2      1.200E+02   1.941E-07     0.000000    49.693    -100.307
3      1.800E+02   5.000E+01     0.417222   -89.998    -239.998
4      2.400E+02   1.519E-07     0.000000    66.397     -83.603
5      3.000E+02   6.466E-02     0.000540  -151.112    -301.112
6      3.600E+02   2.433E-07     0.000000    68.162     -81.838
7      4.200E+02   6.931E-02     0.000578   148.548      -1.453
8      4.800E+02   2.398E-07     0.000000  -174.897    -324.897
9      5.400E+02   5.000E+01     0.417221    90.006     -59.995
total harmonic distortion =     59.004109  percent
```

Fourier analysis of source phase current:

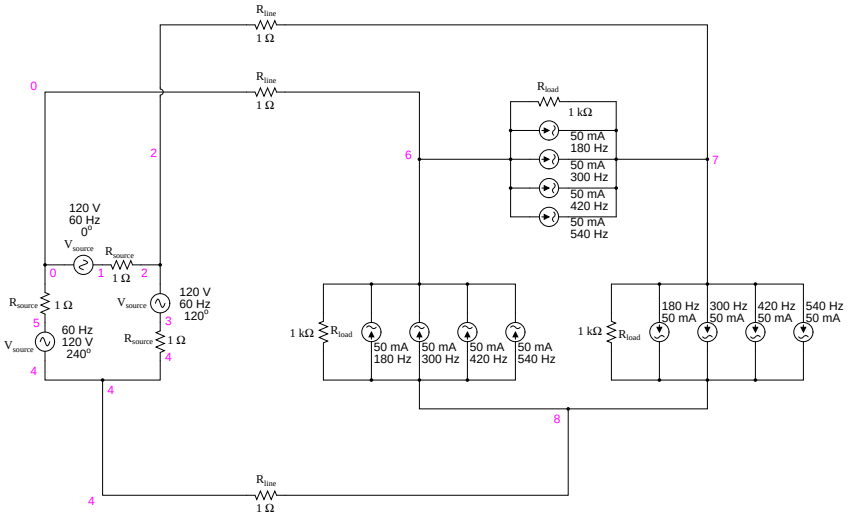
```
fourier components of transient response v(1,2)
dc component =   3.564E-11
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   6.906E-02     1.000000    -0.181       0.000
2      1.200E+02   1.525E-11     0.000000  -156.674    -156.493
3      1.800E+02   1.422E-06     0.000021  -179.996    -179.815
4      2.400E+02   2.949E-11     0.000000  -110.570    -110.390
5      3.000E+02   2.883E-02     0.417440  -179.996    -179.815
6      3.600E+02   2.324E-11     0.000000   -91.926     -91.745
7      4.200E+02   2.883E-02     0.417398  -179.994    -179.813
8      4.800E+02   4.140E-11     0.000000   -39.875     -39.694
9      5.400E+02   4.267E-06     0.000062     0.006       0.186
total harmonic distortion =     59.031969  percent
```



Triplen voltages appear across load phases.
Non-triplen currents appear in line conductors
and in source phase windings.

Really, the only advantage of the Δ-Y configuration from the standpoint of harmonics is that there is no longer a center-point at the load posing a shock hazard. Otherwise, the load components receive the same harmonically-rich voltages and the lines see the same currents as in a three-wire Y system.

If we were to reconfigure the system into a Δ-Δ arrangement, that should guarantee that each load component receives non-harmonic voltage, since each load phase would be directly connected in parallel with each source phase. The complete lack of any neutral wires or "center points" in a Δ-Δ system prevents strange voltages or additive currents from occurring. It would seem to be the ideal solution. Let's simulate and observe, analyzing line current, load phase voltage, and source phase current:



```
Delta-Delta source/load with harmonics
*
* phase1 voltage source and r (120 v /_ 0 deg)
vsource1 1 0 sin(0 120 60 0 0)
rsource1 1 2 1
*
* phase2 voltage source and r (120 v /_ 120 deg)
vsource2 3 2 sin(0 120 60 5.55555m 0)
rsource2 3 4 1
*
* phase3 voltage source and r (120 v /_ 240 deg)
vsource3 5 4 sin(0 120 60 11.1111m 0)
rsource3 5 0 1
*
* line resistances
rline1 0 6 1
rline2 2 7 1
rline3 4 8 1
*
* phase 1 of load
rload1 7 6 1k
i3har1 7 6 sin(0 50m 180 0 0)
i5har1 7 6 sin(0 50m 300 0 0)
i7har1 7 6 sin(0 50m 420 0 0)
i9har1 7 6 sin(0 50m 540 0 0)
*
* phase 2 of load
rload2 8 7 1k
i3har2 8 7 sin(0 50m 180 5.55555m 0)
i5har2 8 7 sin(0 50m 300 5.55555m 0)
i7har2 8 7 sin(0 50m 420 5.55555m 0)
i9har2 8 7 sin(0 50m 540 5.55555m 0)
*
* phase 3 of load
rload3 6 8 1k
i3har3 6 8 sin(0 50m 180 11.1111m 0)
i5har3 6 8 sin(0 50m 300 11.1111m 0)
i7har3 6 8 sin(0 50m 420 11.1111m 0)
i9har3 6 8 sin(0 50m 540 11.1111m 0)
*
* analysis stuff
.options itl5=0
.tran 0.5m 100m 16m 1u
.plot tran v(0,6) v(7,6) v(2,1) i(3har1)
.four 60 v(0,6) v(7,6) v(2,1)
.end
```

Fourier analysis of line current:

```
fourier components of transient response v(0,6)
dc component =  -6.007E-11
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   2.070E-01     1.000000   150.000       0.000
2      1.200E+02   5.480E-11     0.000000   156.666       6.666
3      1.800E+02   6.257E-07     0.000003    89.990     -60.010
4      2.400E+02   4.911E-11     0.000000     8.187    -141.813
5      3.000E+02   8.626E-02     0.416664  -149.999    -300.000
6      3.600E+02   1.089E-10     0.000000   -31.997    -181.997
7      4.200E+02   8.626E-02     0.416669   150.001       0.001
8      4.800E+02   1.578E-10     0.000000   -63.940    -213.940
9      5.400E+02   1.877E-06     0.000009    89.987     -60.013
total harmonic distortion =     58.925538  percent
```

Fourier analysis of load phase voltage:

```
fourier components of transient response v(7,6)
dc component =  -5.680E-10
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.195E+02     1.000000     0.000       0.000
2      1.200E+02   1.039E-09     0.000000   144.749     144.749
3      1.800E+02   1.251E-06     0.000000    89.974      89.974
4      2.400E+02   4.215E-10     0.000000    36.127      36.127
5      3.000E+02   1.992E-01     0.001667  -180.000    -180.000
6      3.600E+02   2.499E-09     0.000000    -4.760      -4.760
7      4.200E+02   1.992E-01     0.001667  -180.000    -180.000
8      4.800E+02   2.951E-09     0.000000  -151.385    -151.385
9      5.400E+02   3.752E-06     0.000000    89.905      89.905
total harmonic distortion =      0.235702  percent
```

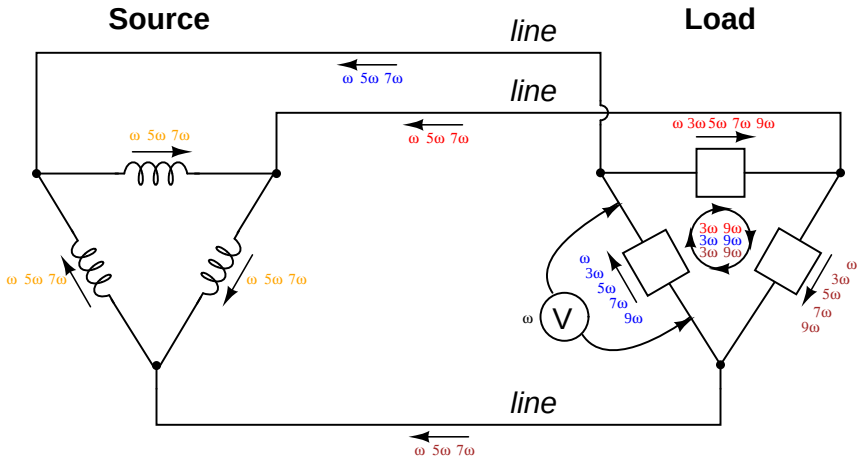
Fourier analysis of source phase current:

```
fourier components of transient response v(2,1)
dc component =  -1.923E-12
harmonic  frequency  fourier    normalized    phase    normalized
no         (hz)     component    component    (deg)    phase (deg)
1      6.000E+01   1.194E-01     1.000000   179.940       0.000
2      1.200E+02   2.569E-11     0.000000   133.491     -46.449
3      1.800E+02   3.129E-07     0.000003    89.985     -89.955
4      2.400E+02   2.657E-11     0.000000    23.368    -156.571
5      3.000E+02   4.980E-02     0.416918  -180.000    -359.939
6      3.600E+02   4.595E-11     0.000000   -22.475    -202.415
7      4.200E+02   4.980E-02     0.416921  -180.000    -359.939
8      4.800E+02   7.385E-11     0.000000   -63.759    -243.699
9      5.400E+02   9.385E-07     0.000008    89.991     -89.949
```

```
total harmonic distortion =     58.961298  percent
```

As predicted earlier, the load phase voltage is almost a pure sine-wave, with negligible harmonic content, thanks to the direct connection with the source phases in a Δ-Δ system. But what happened to the triplen harmonics? The 3rd and 9th harmonic frequencies don't appear in any substantial amount in the line current, nor in the load phase voltage, nor in the source phase current! We know that triplen currents exist, because the 3rd and 9th harmonic current sources are intentionally placed in the phases of the load, but where did those currents go?

Remember that the triplen harmonics of 120^o phase-shifted fundamental frequencies are in phase with each other. Note the directions that the arrows of the current sources within the load phases are pointing, and think about what would happen if the 3rd and 9th harmonic sources were DC sources instead. What we would have is current *circulating within the loop formed by the* Δ*-connected phases*. This is where the triplen harmonic currents have gone: they stay within the Δ of the load, never reaching the line conductors or the windings of the source. These results may be graphically summarized as such:



Load phases receive undistorted sine-wave voltage.
Triplen currents are confined to circulate within load phases.
Non-triplen currents appear in line conductors
and in source phase windings.

This is a major benefit of the Δ-Δ system configuration: triplen harmonic currents remain confined in whatever set of components create them, and do not "spread" to other parts of the system.

- **REVIEW:**
- *Nonlinear* components are those that draw a non-sinusoidal (non-sine-wave) current waveform when energized by a sinusoidal (sine-wave) voltage. Since any distortion of an originally pure sine-wave constitutes harmonic frequencies, we can say that nonlinear components generate harmonic currents.
- When the sine-wave distortion is symmetrical above and below the average centerline of the waveform, the only harmonics present will be *odd-numbered*, not even-numbered.

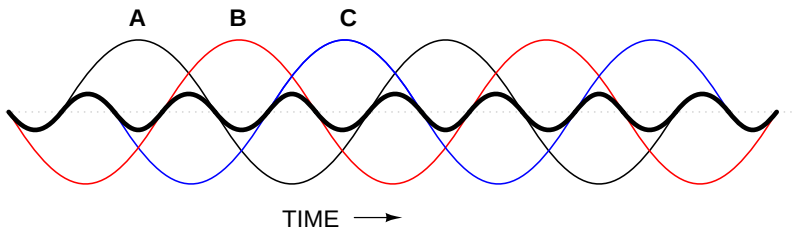
- The 3rd harmonic, and integer multiples of it (6th, 9th, 12th, 15th) are known as *triplen* harmonics. They are in phase with each other, despite the fact that their respective fundamental waveforms are 120^o out of phase with each other.

- In a 4-wire Y-Y system, triplen harmonic currents add within the neutral conductor.

- Triplen harmonic currents in a Δ-connected set of components circulate within the loop formed by the Δ.

10.8 Harmonic phase sequences

In the last section, we saw how the 3rd harmonic and all of its integer multiples (collectively called *triplen* harmonics) generated by 120^o phase-shifted fundamental waveforms are actually in phase with each other. In a 60 Hz three-phase power system, where phases **A**, **B**, and **C** are 120^o apart, the third-harmonic multiples of those frequencies (180 Hz) fall perfectly into phase with each other. This can be thought of in graphical terms, and/or in mathematical terms:



Phase sequence = ***A-B-C***

Fundamental	**A** 0^o	**B** 120^o	**C** 240^o
3rd harmonic	**A'** 3 x 0^o (0^o)	**B'** 3 x 120^o ($360^o = 0^o$)	**C'** 3 x 240^o ($720^o = 0^o$)

If we extend the mathematical table to include higher odd-numbered harmonics, we will notice an interesting pattern develop with regard to the rotation or sequence of the harmonic frequencies:

Fundamental	**A** 0^o	**B** 120^o	**C** 240^o	**A-B-C**
3rd harmonic	**A'** 3×0^o (0^o)	**B'** 3×120^o ($360^o = 0^o$)	**C'** 3×240^o ($720^o = 0^o$)	*no rotation*
5th harmonic	**A''** 5×0^o (0^o)	**B''** 5×120^o ($600^o = 720^o - 120^o$) (-120^o)	**C''** 5×240^o ($1200^o = 1440^o - 240^o$) (-240^o)	**C-B-A**
7th harmonic	**A'''** 7×0^o (0^o)	**B'''** 7×120^o ($840^o = 720^o + 120^o$) (120^o)	**C'''** 7×240^o ($1680^o = 1440^o + 240^o$) (240^o)	**A-B-C**
9th harmonic	**A''''** 9×0^o (0^o)	**B''''** 9×120^o ($1080^o = 0^o$)	**C''''** 9×240^o ($2160^o = 0^o$)	*no rotation*

Harmonics such as the 7th, which "rotate" with the same sequence as the fundamental, are called *positive sequence*. Harmonics such as the 5th, which "rotate" in the opposite sequence as the fundamental, are called *negative sequence*. Triplen harmonics (3rd and 9th shown in this table) which don't "rotate" at all because they're in phase with each other, are called *zero sequence*.

This pattern of positive-zero-negative-positive continues indefinitely for all odd-numbered harmonics, lending itself to expression in a table like this:

Rotation sequences according to harmonic number

+	1st	7th	13th	19th	← Rotates with fundamental
0	3rd	9th	15th	21st	← Does not rotate
-	5th	11th	17th	23rd	← Rotates against fundamental

Sequence especially matters when we're dealing with AC motors, since the mechanical rotation of the rotor depends on the torque produced by the sequential "rotation" of the applied 3-phase power. Positive-sequence frequencies work to push the rotor in the proper direction, whereas negative-sequence frequencies actually work *against* the direction of the rotor's rotation. Zero-sequence frequencies neither contribute to nor detract from the rotor's torque. An excess of negative-sequence harmonics (5th, 11th, 17th, and/or 23rd) in the power supplied to a three-phase AC motor will result in a degradation of performance and possible overheating. Since the higher-order harmonics tend to be attenuated more by system inductances and magnetic core losses, and generally originate with less amplitude anyway, the primary harmonic of concern is the 5th, which is 300 Hz in 60 Hz power systems and 250 Hz in 50 Hz power systems.

10.9 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Ed Beroset (May 6, 2002): Suggested better ways to illustrate the meaning of the prefix "poly-".

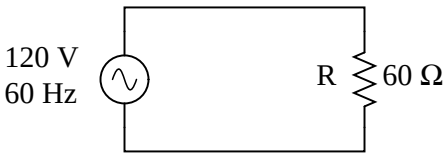
Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 11

POWER FACTOR

11.1 Power in resistive and reactive AC circuits

Consider a circuit for a single-phase AC power system, where a 120 volt, 60 Hz AC voltage source is delivering power to a resistive load:



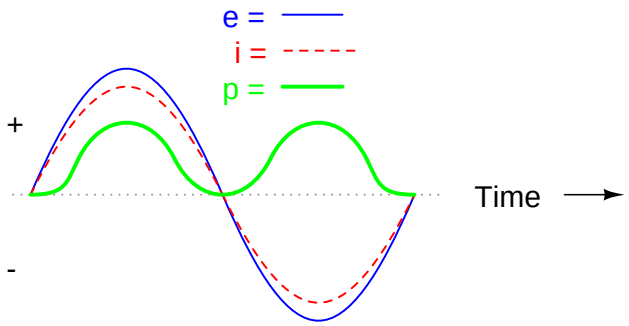
$Z_R = 60 + j0\ \Omega \quad or \quad 60\ \Omega \angle 0^o$

$$I = \frac{E}{Z}$$

$$I = \frac{120\ V}{60\ \Omega}$$

$I = 2\ A$

In this example, the current to the load would be 2 amps, RMS. The power dissipated at the load would be 240 watts. Because this load is purely resistive (no reactance), the current is in phase with the voltage, and calculations look similar to that in an equivalent DC circuit. If we were to plot the voltage, current, and power waveforms for this circuit, it would look like this:

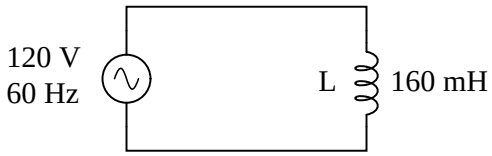


Note that the waveform for power is always positive, never negative for this resistive circuit. This means that power is always being dissipated by the resistive load, and never returned to the source as it is with reactive loads. If the source were a mechanical generator, it would take 240 watts worth of mechanical energy (about 1/3 horsepower) to turn the shaft.

Also note that the waveform for power is not at the same frequency as the voltage or current! Rather, its frequency is *double* that of either the voltage or current waveforms. This different frequency prohibits our expression of power in an AC circuit using the same complex (rectangular or polar) notation as used for voltage, current, and impedance, because this form of mathematical symbolism implies unchanging phase relationships. When frequencies are not the same, phase relationships constantly change.

As strange as it may seem, the best way to proceed with AC power calculations is to use *scalar* notation, and to handle any relevant phase relationships with trigonometry.

For comparison, let's consider a simple AC circuit with a purely reactive load:



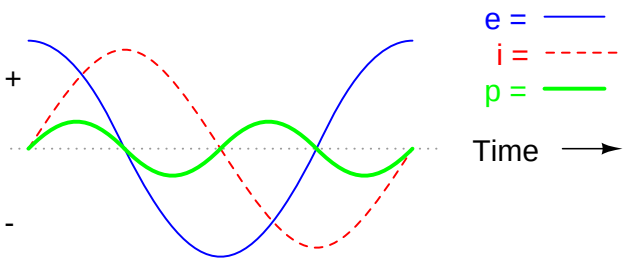
$X_L = 60.319\ \Omega$

$Z_L = 0 + j60.319\ \Omega \quad or \quad 60.319\ \Omega \angle 90^o$

$$I = \frac{E}{Z}$$

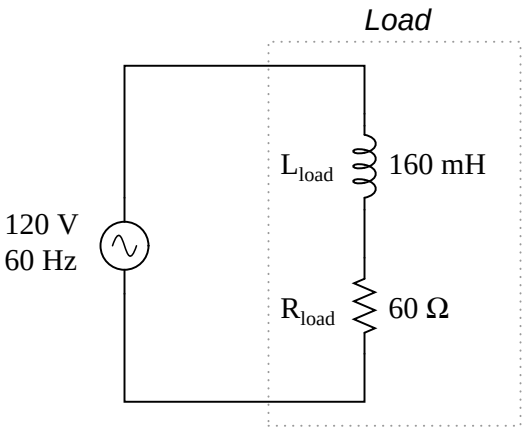
$$I = \frac{120\ V}{60.319\ \Omega}$$

$I = 1.989$ A



Note that the power alternates equally between cycles of positive and negative. This means that power is being alternately absorbed from and returned to the source. If the source were a mechanical generator, it would take (practically) no net mechanical energy to turn the shaft, because no power would be used by the load. The generator shaft would be easy to spin, and the inductor would not become warm as a resistor would.

Now, let's consider an AC circuit with a load consisting of both inductance and resistance:



$X_L = 60.319\ \Omega$

$Z_L = 0 + j60.319\ \Omega \quad or \quad 60.319\ \Omega \angle 90^o$

$Z_R = 60 + j0\ \Omega \quad or \quad 60\ \Omega \angle 0^o$

$Z_{total} = 60 + j60.319\ \Omega \quad or \quad 85.078\ \Omega \angle 45.152^o$

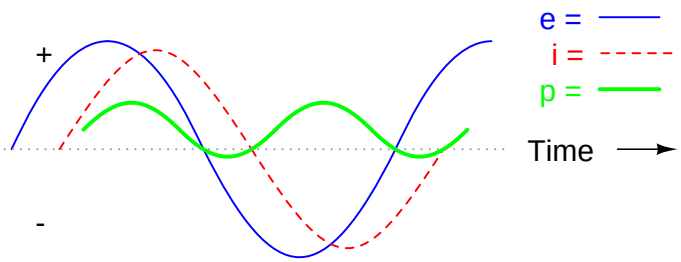
$$I = \frac{E}{Z}$$

$$I = \frac{120\ V}{85.078\ \Omega}$$

$I = 1.410\ A$

At a frequency of 60 Hz, the 160 millihenrys of inductance gives us 60.319 Ω of inductive reactance. This reactance combines with the 60 Ω of resistance to form a total load impedance of 60 + j60.319 Ω, or 85.078 Ω ∠ 45.152°. If we're not concerned with phase angles (which we're not at this point), we may calculate current in the circuit by taking the polar magnitude of the voltage source (120 volts) and dividing it my the polar magnitude of the impedance (85.078 Ω). With a power supply voltage of 120 volts RMS, our load current is 1.410 amps. This is the figure an RMS ammeter would indicate if connected in series with the resistor and inductor.

We already know that reactive components dissipate zero power, as they equally absorb power from, and return power to, the rest of the circuit. Therefore, any inductive reactance in this load will likewise dissipate zero power. The only thing left to dissipate power here is the resistive portion of the load impedance. If we look at the waveform plot of voltage, current, and total power for this circuit, we see how this combination works:



As with any reactive circuit, the power alternates between positive and negative instantaneous values over time. In a purely reactive circuit that alternation between positive and negative power is equally divided, resulting in a net power dissipation of zero. However, in circuits with mixed resistance and reactance like this one, the power waveform will still alternate between positive and negative, but the amount of positive power will exceed the amount of negative power. In other words, the combined inductive/resistive load will consume more power than it returns back to the

source.

Looking at the waveform plot for power, it should be evident that the wave spends more time on the positive side of the center line than on the negative, indicating that there is more power absorbed by the load than there is returned to the circuit. What little returning of power that occurs is due to the reactance; the imbalance of positive versus negative power is due to the resistance as it dissipates energy outside of the circuit (usually in the form of heat). If the source were a mechanical generator, the amount of mechanical energy needed to turn the shaft would be the amount of power averaged between the positive and negative power cycles.

Mathematically representing power in an AC circuit is a challenge, because the power wave isn't at the same frequency as voltage or current. Furthermore, the phase angle for power means something quite different from the phase angle for either voltage or current. Whereas the angle for voltage or current represents a relative *shift in timing* between two waves, the phase angle for power represents a *ratio* between power dissipated and power returned. Because of this way in which AC power differs from AC voltage or current, it is actually easier to arrive at figures for power by calculating with *scalar* quantities of voltage, current, resistance, and reactance than it is to try to derive it from *vector*, or *complex* quantities of voltage, current, and impedance that we've worked with so far.

- **REVIEW:**
- In a purely resistive circuit, all circuit power is dissipated by the resistor(s). Voltage and current are in phase with each other.
- In a purely reactive circuit, no circuit power is dissipated by the load(s). Rather, power is alternately absorbed from and returned to the AC source. Voltage and current are 90^o out of phase with each other.
- In a circuit consisting of resistance and reactance mixed, there will be more power dissipated by the load(s) than returned, but some power will definitely be dissipated and some will merely be absorbed and returned. Voltage and current in such a circuit will be out of phase by a value somewhere between 0^o and 90^o.

11.2 True, Reactive, and Apparent power

We know that reactive loads such as inductors and capacitors dissipate zero power, yet the fact that they drop voltage and draw current gives the deceptive impression that they actually *do* dissipate power. This "phantom power" is called *reactive power*, and it is measured in a unit called *Volt-Amps-Reactive* (VAR), rather than watts. The mathematical symbol for reactive power is (unfortunately) the capital letter Q. The actual amount of power being used, or dissipated, in a circuit is called *true power*, and it is measured in watts (symbolized by the capital letter P, as always). The combination of reactive power and true power is called *apparent power*, and it is the product of a circuit's voltage and current, without reference to phase angle. Apparent power is measured in the unit of *Volt-Amps* (VA) and is symbolized by the capital letter S.

As a rule, true power is a function of a circuit's dissipative elements, usually resistances (R). Reactive power is a function of a circuit's reactance (X). Apparent power is a function of a circuit's total impedance (Z). Since we're dealing with scalar quantities for power calculation, any complex

starting quantities such as voltage, current, and impedance must be represented by their *polar magnitudes*, not by real or imaginary rectangular components. For instance, if I'm calculating true power from current and resistance, I must use the polar magnitude for current, and not merely the "real" or "imaginary" portion of the current. If I'm calculating apparent power from voltage and impedance, both of these formerly complex quantities must be reduced to their polar magnitudes for the scalar arithmetic.

There are several power equations relating the three types of power to resistance, reactance, and impedance (all using scalar quantities):

P = true power $P = I^2R$ $P = \frac{E^2}{R}$

Measured in units of ***Watts***

Q = reactive power $Q = I^2X$ $Q = \frac{E^2}{X}$

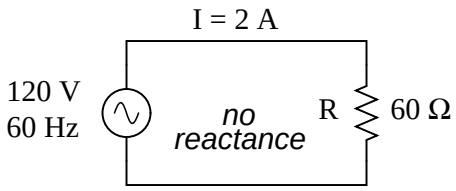
Measured in units of ***Volt-Amps-Reactive (VAR)***

S = apparent power $S = I^2Z$ $S = \frac{E^2}{Z}$ $S = IE$

Measured in units of ***Volt-Amps (VA)***

Please note that there are two equations each for the calculation of true and reactive power. There are three equations available for the calculation of apparent power, P=IE being useful *only* for that purpose. Examine the following circuits and see how these three types of power interrelate:

Resistive load only:

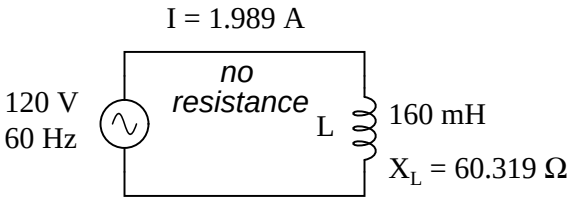


P = true power = I^2R = 240 W

Q = reactive power = I^2X = 0 VAR

S = apparent power = I^2Z = 240 VA

Reactive load only:

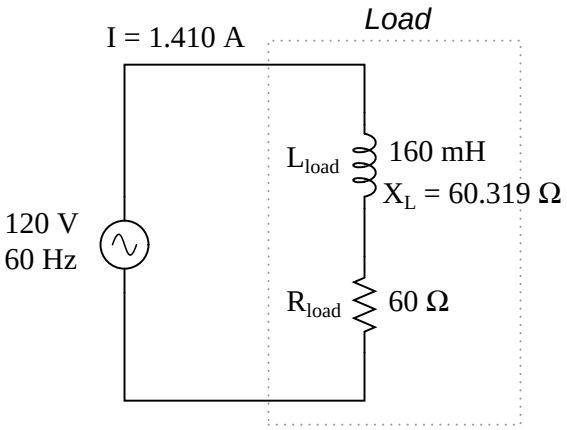


P = true power = I^2R = 0 W

Q = reactive power = I^2X = 238.73 VAR

S = apparent power = I^2Z = 238.73 VA

Resistive/reactive load:

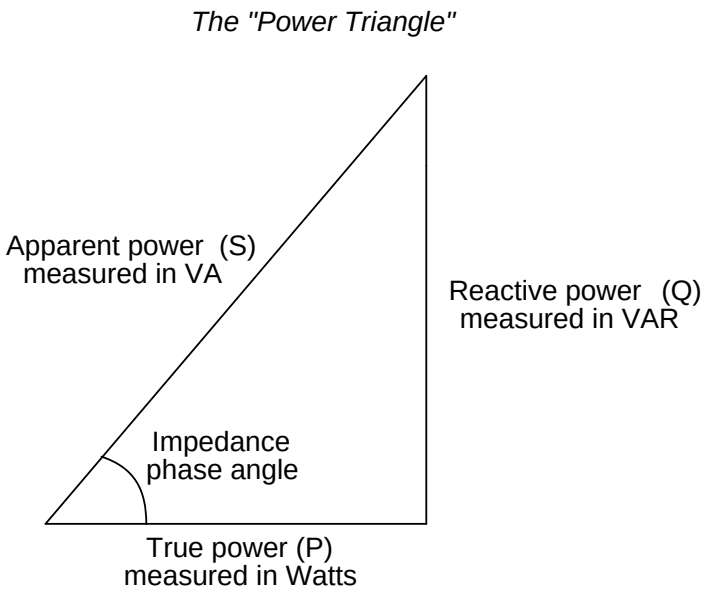


P = true power = I^2R = 119.365 W

Q = reactive power = I^2X = 119.998 VAR

S = apparent power = I^2Z = 169.256 VA

These three types of power – true, reactive, and apparent – relate to one another in trigonometric form. We call this the *power triangle*:



Using the laws of trigonometry, we can solve for the length of any side (amount of any type of power), given the lengths of the other two sides, or the length of one side and an angle.

- **REVIEW:**
- Power dissipated by a load is referred to as *true power*. True power is symbolized by the letter P and is measured in the unit of Watts (W).
- Power merely absorbed and returned in load due to its reactive properties is referred to as *reactive power*. Reactive power is symbolized by the letter Q and is measured in the unit of Volt-Amps-Reactive (VAR).
- Total power in an AC circuit, both dissipated and absorbed/returned is referred to as *apparent power*. Apparent power is symbolized by the letter S and is measured in the unit of Volt-Amps (VA).
- These three types of power are trigonometrically related to one another. In a right triangle, P = adjacent length, Q = opposite length, and S = hypotenuse length. The opposite angle is equal to the circuit's impedance (Z) phase angle.

11.3 Calculating power factor

As was mentioned before, the angle of this "power triangle" graphically indicates the ratio between the amount of dissipated (or *consumed*) power and the amount of absorbed/returned power. It also happens to be the same angle as that of the circuit's impedance in polar form. When expressed as a fraction, this ratio between true power and apparent power is called the *power factor* for this circuit. Because true power and apparent power form the adjacent and hypotenuse sides of a right

triangle, respectively, the power factor ratio is also equal to the cosine of that phase angle. Using values from the last example circuit:

$$\text{Power factor} = \frac{\text{True power}}{\text{Apparent power}}$$

$$\text{Power factor} = \frac{119.365 \text{ W}}{169.256 \text{ VA}}$$

$$\text{Power factor} = 0.705$$

$$\cos 45.152^{\text{o}} = 0.705$$

It should be noted that power factor, like all ratio measurements, is a *unitless* quantity.

For the purely resistive circuit, the power factor is 1 (perfect), because the reactive power equals zero. Here, the power triangle would look like a horizontal line, because the opposite (reactive power) side would have zero length.

For the purely inductive circuit, the power factor is zero, because true power equals zero. Here, the power triangle would look like a vertical line, because the adjacent (true power) side would have zero length.

The same could be said for a purely capacitive circuit. If there are no dissipative (resistive) components in the circuit, then the true power must be equal to zero, making any power in the circuit purely reactive. The power triangle for a purely capacitive circuit would again be a vertical line (pointing down instead of up as it was for the purely inductive circuit).

Power factor can be an important aspect to consider in an AC circuit, because any power factor less than 1 means that the circuit's wiring has to carry more current than what would be necessary with zero reactance in the circuit to deliver the same amount of (true) power to the resistive load. If our last example circuit had been purely resistive, we would have been able to deliver a full 169.256 watts to the load with the same 1.410 amps of current, rather than the mere 119.365 watts that it is presently dissipating with that same current quantity. The poor power factor makes for an inefficient power delivery system.

Poor power factor can be corrected, paradoxically, by adding another load to the circuit drawing an equal and opposite amount of reactive power, to cancel out the effects of the load's inductive reactance. Inductive reactance can only be canceled by capacitive reactance, so we have to add a *capacitor* in parallel to our example circuit as the additional load. The effect of these two opposing reactances in parallel is to bring the circuit's total impedance equal to its total resistance (to make the impedance phase angle equal, or at least closer, to zero).

Since we know that the (uncorrected) reactive power is 119.998 VAR (inductive), we need to calculate the correct capacitor size to produce the same quantity of (capacitive) reactive power. Since this capacitor will be directly in parallel with the source (of known voltage), we'll use the power formula which starts from voltage and reactance:

$$Q = \frac{E^2}{X}$$

. . . solving for X . . .

$$X = \frac{E^2}{Q}$$

$$X = \frac{(120\ \text{V})^2}{119.998\ \text{VAR}}$$

$$X = 120.002\ \Omega$$

$$X_C = \frac{1}{2\pi fC}$$

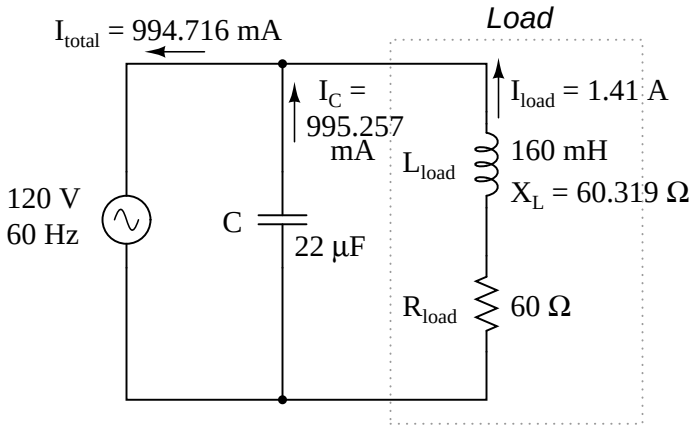
. . . solving for C . . .

$$C = \frac{1}{2\pi f X_C}$$

$$C = \frac{1}{2\pi(60\ \text{Hz})(120.002\ \Omega)}$$

$$C = 22.105\ \mu\text{F}$$

Let's use a rounded capacitor value of 22 μF and see what happens to our circuit:



$Z_{total} = Z_C \,//\, (Z_L \text{ -- } Z_R)$

$Z_{total} = (120.57\ \Omega \angle -90^o) \,//\, (60.319\ \Omega \angle 90^o \text{ -- } 60\ \Omega \angle 0^o)$

$Z_{total} = 120.64 - j573.58m\ \Omega \quad \textit{or} \quad 120.64\ \Omega \angle 0.2724^o$

$P = \text{true power} = I^2R = 119.365\ W$

$S = \text{apparent power} = I^2Z = 119.366\ VA$

The power factor for the circuit, overall, has been substantially improved. The main current has been decreased from 1.41 amps to 994.7 milliamps, while the power dissipated at the load resistor remains unchanged at 119.365 watts. The power factor is much closer to being 1:

$$\text{Power factor} = \frac{\text{True power}}{\text{Apparent power}}$$

$$\text{Power factor} = \frac{119.365\ W}{119.366\ VA}$$

$$\text{Power factor} = 0.9999887$$

$$\text{Impedance (polar) angle} = 0.272^o$$

$$\cos 0.272^o = 0.9999887$$

Since the impedance angle is still a positive number, we know that the circuit, overall, is still more inductive than it is capacitive. If our power factor correction efforts had been perfectly on-target, we would have arrived at an impedance angle of exactly zero, or purely resistive. If we had added too large of a capacitor in parallel, we would have ended up with an impedance angle that was negative, indicating that the circuit was more capacitive than inductive.

It should be noted that too much capacitance in an AC circuit will result in a low power factor just as well as too much inductance. You must be careful not to over-correct when adding capacitance to an AC circuit. You must also be *very* careful to use the proper capacitors for the job (rated adequately for power system voltages and the occasional voltage spike from lightning strikes, for continuous AC service, and capable of handling the expected levels of current).

If a circuit is predominantly inductive, we say that its power factor is *lagging* (because the current wave for the circuit lags behind the applied voltage wave). Conversely, if a circuit is predominantly capacitive, we say that its power factor is *leading*. Thus, our example circuit started out with a power factor of 0.705 lagging, and was corrected to a power factor of 0.999 lagging.

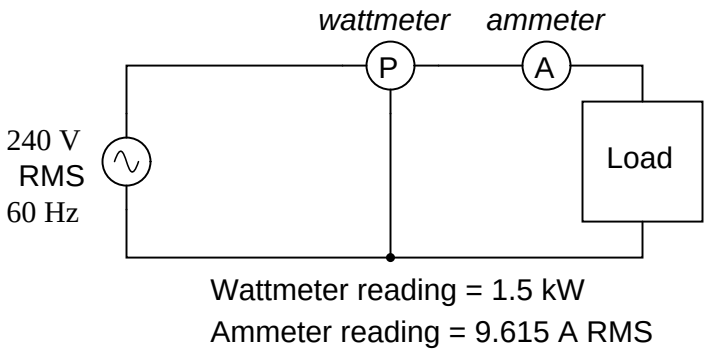
- **REVIEW:**
- Poor power factor in an AC circuit may be "corrected," or re-established at a value close to 1, by adding a parallel reactance opposite the effect of the load's reactance. If the load's

reactance is inductive in nature (which is almost always will be), parallel *capacitance* is what is needed to correct poor power factor.

11.4 Practical power factor correction

When the need arises to correct for poor power factor in an AC power system, you probably won't have the luxury of knowing the load's exact inductance in henrys to use for your calculations. You may be fortunate enough to have an instrument called a *power factor meter* to tell you what the power factor is (a number between 0 and 1), and the apparent power (which can be figured by taking a voltmeter reading in volts and multiplying by an ammeter reading in amps). In less favorable circumstances you may have to use an oscilloscope to compare voltage and current waveforms, measuring phase shift in *degrees* and calculating power factor by the cosine of that phase shift.

Most likely, you will have access to a wattmeter for measuring true power, whose reading you can compare against a calculation of apparent power (from multiplying total voltage and total current measurements). From the values of true and apparent power, you can determine reactive power and power factor. Let's do an example problem to see how this works:



First, we need to calculate the apparent power in kVA. We can do this by multiplying load voltage by load current:

$$S = IE$$

$$S = (9.615 \text{ A})(240 \text{ V})$$

$$S = 2.308 \text{ kVA}$$

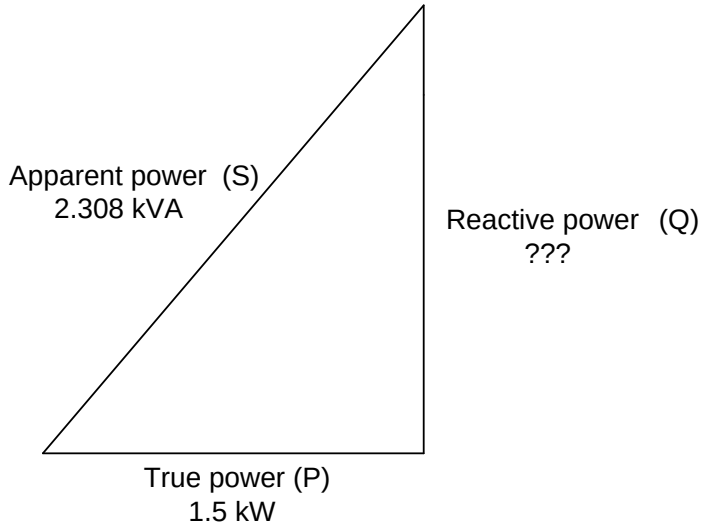
As we can see, 2.308 kVA is a much larger figure than 1.5 kW, which tells us that the power factor in this circuit is rather poor (substantially less than 1). Now, we figure the power factor of this load by dividing the true power by the apparent power:

$$\text{Power factor} = \frac{P}{S}$$

$$\text{Power factor} = \frac{1.5 \text{ kW}}{2.308 \text{ kVA}}$$

$$\text{Power factor} = 0.65$$

Using this value for power factor, we can draw a power triangle, and from that determine the reactive power of this load:



To determine the unknown (reactive power) triangle quantity, we use the Pythagorean Theorem "backwards," given the length of the hypotenuse (apparent power) and the length of the adjacent side (true power):

$$\text{Reactive power} = \sqrt{(\text{Apparent power})^2 - (\text{True power})^2}$$

$$Q = 1.754 \text{ kVAR}$$

If this load is an electric motor, or most any other industrial AC load, it will have a lagging (inductive) power factor, which means that we'll have to correct for it with a *capacitor* of appropriate size, wired in parallel. Now that we know the amount of reactive power (1.754 kVAR), we can calculate the size of capacitor needed to counteract its effects:

$$Q = \frac{E^2}{X}$$

. . . solving for X . . .

$$X = \frac{E^2}{Q}$$

$$X = \frac{(240)^2}{1.754 \text{ kVAR}}$$

$$X = 32.845\ \Omega$$

$$X_C = \frac{1}{2\pi fC}$$

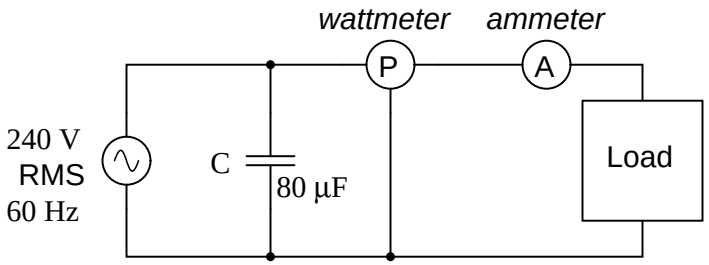
. . . solving for C . . .

$$C = \frac{1}{2\pi f X_C}$$

$$C = \frac{1}{2\pi(60 \text{ Hz})(32.845\ \Omega)}$$

$$C = 80.761\ \mu\text{F}$$

Rounding this answer off to 80 μF, we can place that size of capacitor in the circuit and calculate the results:

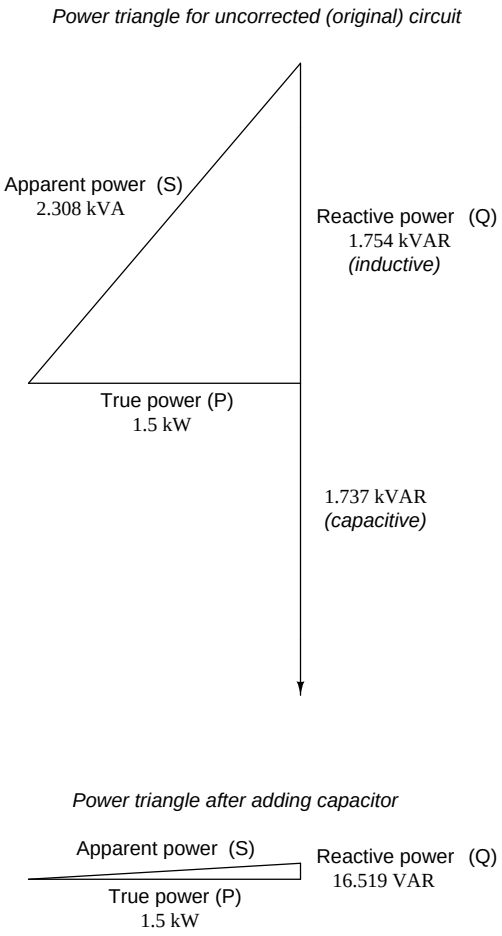


An 80 μF capacitor will have a capacitive reactance of 33.157 Ω, giving a current of 7.238 amps, and a corresponding reactive power of 1.737 kVAR (for the capacitor *only*). Since the capacitor's current is 180^o out of phase from the the load's inductive contribution to current draw, the capacitor's reactive power will directly subtract from the load's reactive power, resulting in:

Inductive kVAR - Capacitive kVAR = Total kVAR

1.754 kVAR - 1.737 kVAR = 16.519 VAR

This correction, of course, will not change the amount of true power consumed by the load, but it will result in a substantial reduction of apparent power, and of the total current drawn from the 240 Volt source:



The new apparent power can be found from the true and new reactive power values, using the standard form of the Pythagorean Theorem:

$$\text{Apparent power} = \sqrt{(\text{Reactive power})^2 + (\text{True power})^2}$$

$$\text{Apparent power} = 1.50009 \text{ kVA}$$

This gives a corrected power factor of (1.5kW / 1.5009 kVA), or 0.99994, and a new total current of (1.50009 kVA / 240 Volts), or 6.25 amps, a substantial improvement over the uncorrected value of 9.615 amps! This lower total current will translate to less heat losses in the circuit wiring, meaning greater system efficiency (less power wasted).

11.5 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

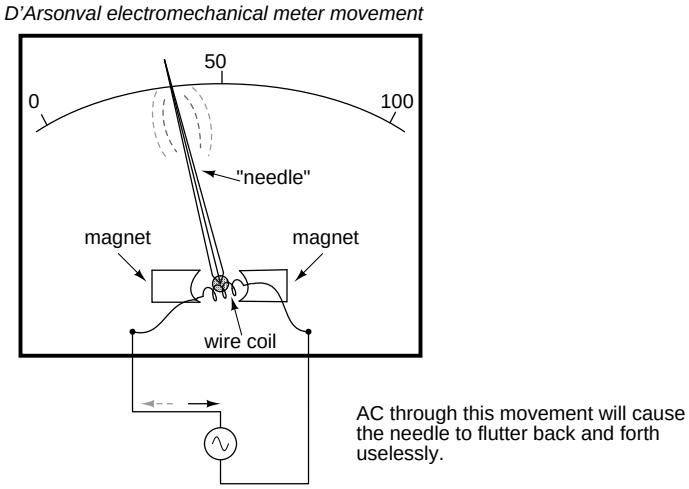
Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 12

AC METERING CIRCUITS

12.1 AC voltmeters and ammeters

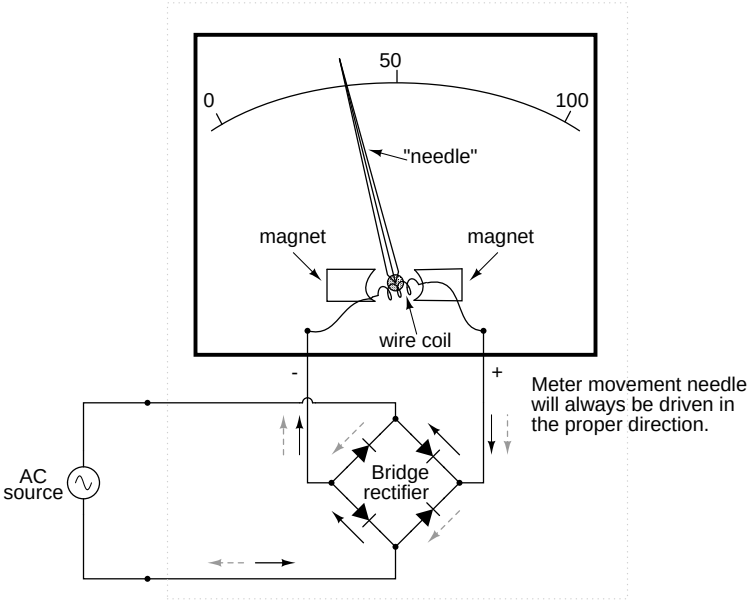
AC electromechanical meter movements come in two basic arrangements: those based on DC movement designs, and those engineered specifically for AC use. Permanent-magnet moving coil (PMMC) meter movements will not work correctly if directly connected to alternating current, because the direction of needle movement will change with each half-cycle of the AC. Permanent-magnet meter movements, like permanent-magnet motors, are devices whose motion depends on the polarity of the applied voltage (or, you can think of it in terms of the direction of the current).



In order to use a DC-style meter movement such as the D'Arsonval design, the alternating current must be *rectified* into DC. This is most easily accomplished through the use of devices called *diodes*. We saw diodes used in an example circuit demonstrating the creation of harmonic frequencies from a distorted (or rectified) sine wave. Without going into elaborate detail over how and why diodes work as they do, just remember that they each act like a one-way valve for electrons to flow: acting as a conductor for one polarity and an insulator for another. Oddly enough, the arrowhead in each

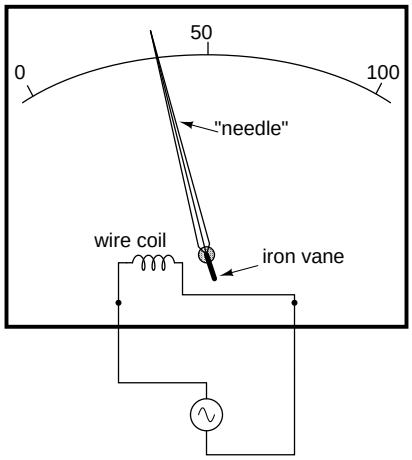
diode symbol points *against* the permitted direction of electron flow rather than with it as one might expect. Arranged in a bridge, four diodes will serve to steer AC through the meter movement in a constant direction throughout all portions of the AC cycle:

"Rectified" AC meter movement



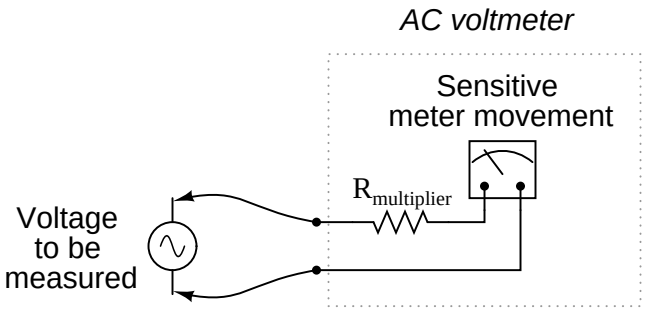
Another strategy for a practical AC meter movement is to redesign the movement without the inherent polarity sensitivity of the DC types. This means avoiding the use of permanent magnets. Probably the simplest design is to use a nonmagnetized iron vane to move the needle against spring tension, the vane being attracted toward a stationary coil of wire energized by the AC quantity to be measured.

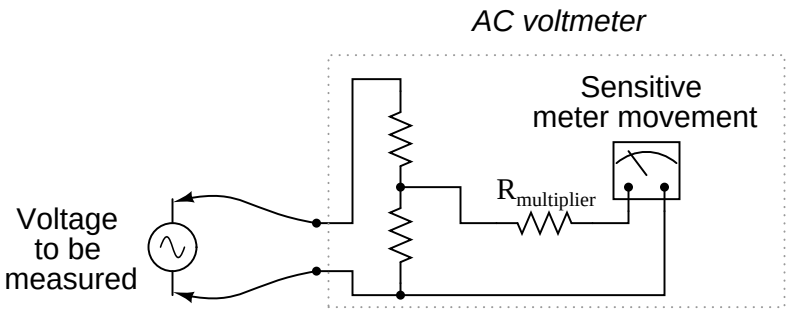
Iron-vane electromechanical meter movement



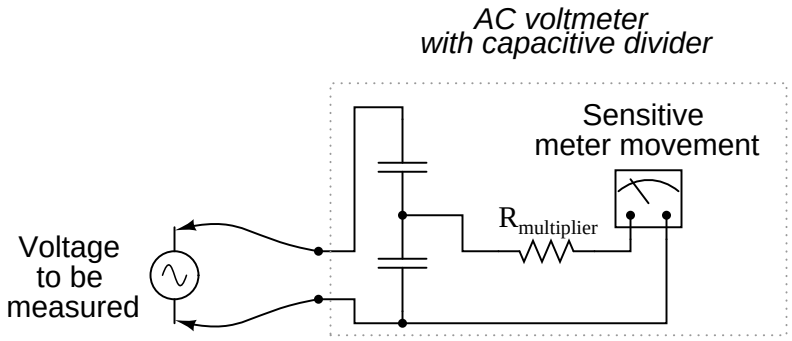
Electrostatic attraction between two metal plates separated by an air gap is an alternative mechanism for generating a needle-moving force proportional to applied voltage. This works just as well for AC as it does for DC, or should I say, just as poorly! The forces involved are very small, much smaller than the magnetic attraction between an energized coil and an iron vane, and as such these "electrostatic" meter movements tend to be fragile and easily disturbed by physical movement. But, for some high-voltage AC applications, the electrostatic movement is an elegant technology. If nothing else, this technology possesses the advantage of extremely high input impedance, meaning that no current need be drawn from the circuit under test. Also, electrostatic meter movements are capable of measuring very high voltages without need for range resistors or other, external apparatus.

When a sensitive meter movement needs to be re-ranged to function as an AC voltmeter, series-connected "multiplier" resistors and/or resistive voltage dividers may be employed just as in DC meter design:

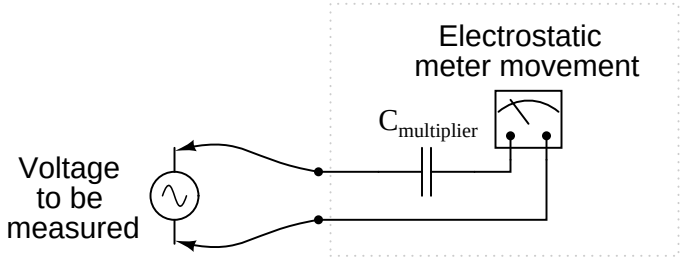




Capacitors may be used instead of resistors, though, to make voltmeter divider circuits. This strategy has the advantage of being non-dissipative (no true power consumed and no heat produced):

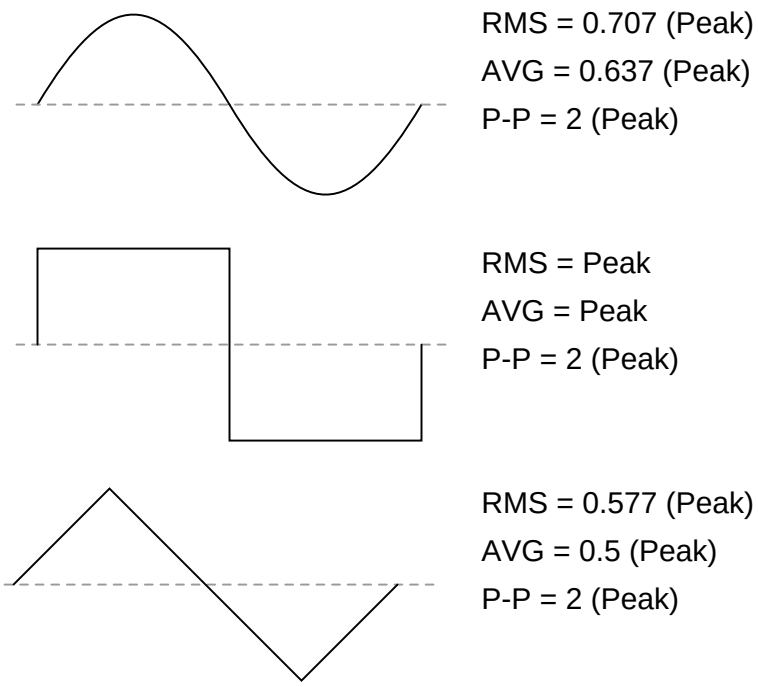


If the meter movement is electrostatic, and thus inherently capacitive in nature, a single "multiplier" capacitor may be connected in series to give it a greater voltage measuring range, just as a series-connected multiplier resistor gives a moving-coil (inherently resistive) meter movement a greater voltage range:



The Cathode Ray Tube (CRT) mentioned in the DC metering chapter is ideally suited for measuring AC voltages, especially if the electron beam is swept side-to-side across the screen of the tube while the measured AC voltage drives the beam up and down. A graphical representation of the AC wave shape and not just a measurement of magnitude can easily be had with such a device. However, CRT's have the disadvantages of weight, size, significant power consumption, and fragility (being made of evacuated glass) working against them. For these reasons, electromechanical AC meter movements still have a place in practical usage.

With some of the advantages and disadvantages of these meter movement technologies having been discussed already, there is another factor crucially important for the designer and user of AC metering instruments to be aware of. This is the issue of RMS measurement. As we already know, AC measurements are often cast in a scale of DC power equivalence, called *RMS* (**R**oot-**M**ean-**S**quare) for the sake of meaningful comparisons with DC and with other AC waveforms of varying shape. None of the meter movement technologies so far discussed inherently measure the RMS value of an AC quantity. Meter movements relying on the motion of a mechanical needle ("rectified" D'Arsonval, iron-vane, and electrostatic) all tend to mechanically average the instantaneous values into an overall average value for the waveform. This average value is not necessarily the same as RMS, although many times it is mistaken as such. Average and RMS values rate against each other as such for these three common waveform shapes:



Since RMS seems to be the kind of measurement most people are interested in obtaining with an instrument, and electromechanical meter movements naturally deliver *average* measurements rather than RMS, what are AC meter designers to do? Cheat, of course! Typically the assumption is made that the waveform shape to be measured is going to be sine (by far the most common, especially for power systems), and then the meter movement scale is altered by the appropriate multiplication factor. For sine waves we see that RMS is equal to 0.707 times the peak value while Average is 0.637 times the peak, so we can divide one figure by the other to obtain an average-to-RMS conversion factor of 1.109:

$$\frac{0.707}{0.637} = 1.1099$$

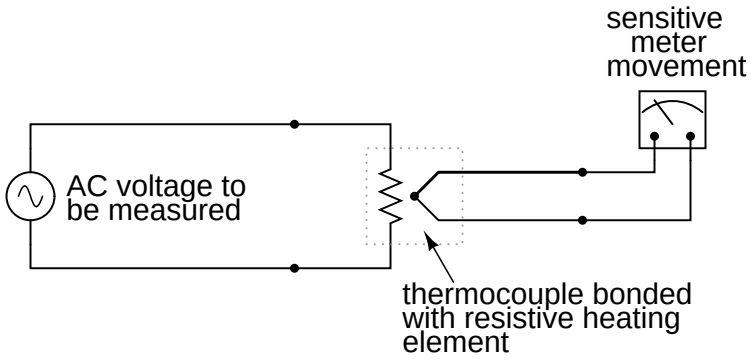
In other words, the meter movement will be calibrated to indicate approximately 1.11 times higher than it would ordinarily (naturally) indicate with no special accommodations. It must be stressed that this "cheat" only works well when the meter is used to measure pure sine wave sources. Note that for triangle waves, the ratio between RMS and Average is not the same as for sine waves:

$$\frac{0.577}{0.5} = 1.154$$

With square waves, the RMS and Average values are identical! An AC meter calibrated to accurately read RMS voltage or current on a pure sine wave will *not* give the proper value while indicating the magnitude of anything other than a perfect sine wave. This includes triangle waves, square waves, or any kind of distorted sine wave. With harmonics becoming an ever-present phenomenon in large AC power systems, this matter of accurate RMS measurement is no small matter.

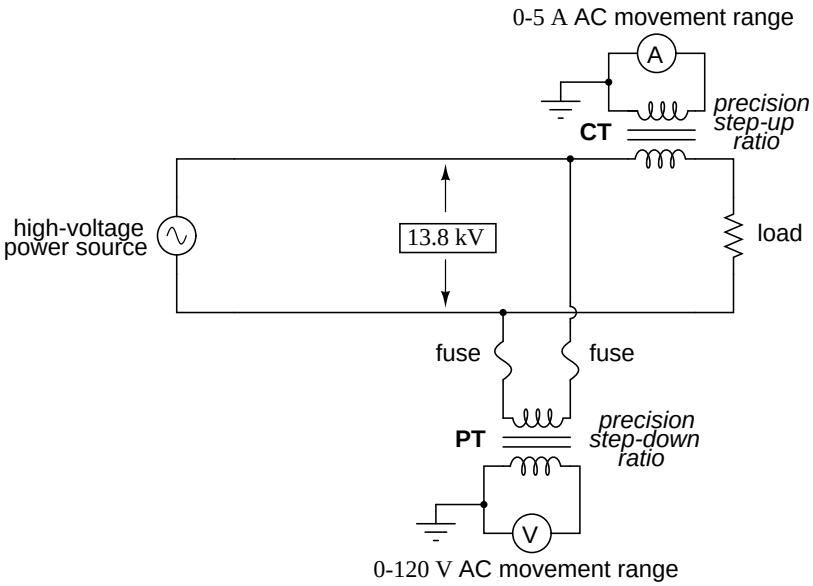
The astute reader will note that I have omitted the CRT "movement" from the RMS/Average discussion. This is because a CRT with its practically weightless electron beam "movement" displays the Peak (or Peak-to-Peak if you wish) of an AC waveform rather than Average or RMS. Still, a similar problem arises: how do you determine the RMS value of a waveform from it? Conversion factors between Peak and RMS only hold so long as the waveform falls neatly into a known category of shape (sine, triangle, and square are the only examples with Peak/RMS/Average conversion factors given here!).

One answer is to design the meter movement around the very definition of RMS: the effective heating value of an AC voltage/current as it powers a resistive load. Suppose that the AC source to be measured is connected across a resistor of known value, and the heat output of that resistor is measured with a device like a thermocouple. This would provide a far more direct measurement means of RMS than any conversion factor could, for it will work with ANY waveform shape whatsoever:

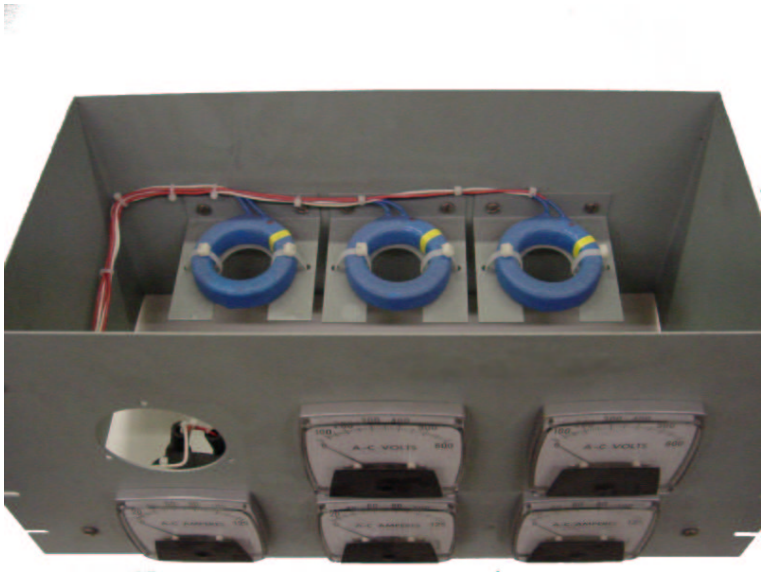


While the device shown above is somewhat crude and would suffer from unique engineering problems of its own, the concept illustrated is very sound. The resistor converts the AC voltage or current quantity into a thermal (heat) quantity, effectively squaring the values in real-time. The system's mass works to average these values by the principle of thermal inertia, and then the meter scale itself is calibrated to give an indication based on the square-root of the thermal measurement: perfect Root-Mean-Square indication all in one device! In fact, one major instrument manufacturer has implemented this technique into its high-end line of handheld electronic multimeters for "true-RMS" capability.

Calibrating AC voltmeters and ammeters for different full-scale ranges of operation is much the same as with DC instruments: series "multiplier" resistors are used to give voltmeter movements higher range, and parallel "shunt" resistors are used to allow ammeter movements to measure currents beyond their natural range. However, we are not limited to these techniques as we were with DC: because we can to use transformers with AC, meter ranges can be electromagnetically rather than resistively "stepped up" or "stepped down," sometimes far beyond what resistors would have practically allowed for. Potential Transformers (PT's) and Current Transformers (CT's) are precision instrument devices manufactured to produce very precise ratios of transformation between primary and secondary windings. They can allow small, simple AC meter movements to indicate extremely high voltages and currents in power systems with accuracy and complete electrical isolation (something multiplier and shunt resistors could never do):



Shown here is a voltage and current meter panel from a three-phase AC system. The three "donut" current transformers (CTs) can be seen in the rear of the panel. Three AC ammeters (rated 5 amps full-scale deflection each) on the front of the panel indicate current through each conductor going through a CT. As this panel has been removed from service, there are no current-carrying conductors threaded through the center of the CT "donuts" anymore:



Because of the expense (and often large size) of instrument transformers, they are not used to scale AC meters for any applications other than high voltage and high current. For scaling a milliamp or microamp movement to a range of 120 volts or 5 amps, normal precision resistors (multipliers

and shunts) are used, just as with DC.

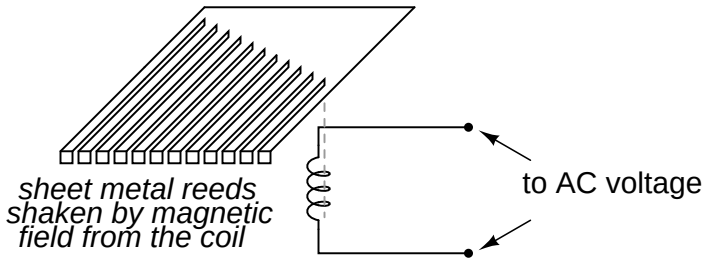
- **REVIEW:**
- Polarized (DC) meter movements must use devices called *diodes* to be able to indicate AC quantities.
- Electromechanical meter movements, whether electromagnetic or electrostatic, naturally provide the *average* value of a measured AC quantity. These instruments may be ranged to indicate RMS value, but only if the shape of the AC waveform is precisely known beforehand!
- So-called *true RMS* meters use different technology to provide indications representing the actual RMS (rather than skewed average or peak) of an AC waveform.

12.2 Frequency and phase measurement

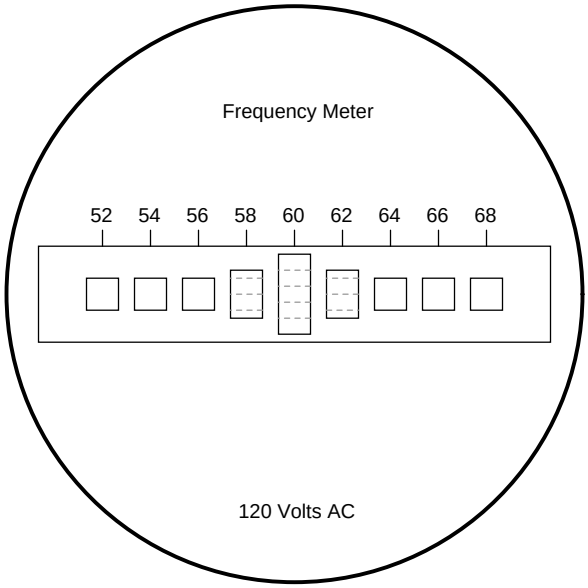
An important electrical quantity with no equivalent in DC circuits is *frequency*. Frequency measurement is very important in many applications of alternating current, especially in AC power systems designed to run efficiently at one frequency and one frequency only. If the AC is being generated by an electromechanical alternator, the frequency will be directly proportional to the shaft speed of the machine, and frequency could be measured simply by measuring the speed of the shaft. If frequency needs to be measured at some distance from the alternator, though, other means of measurement will be necessary.

One simple but crude method of frequency measurement in power systems utilizes the principle of mechanical resonance. Every physical object possessing the property of elasticity (springiness) has an inherent frequency at which it will prefer to vibrate. The tuning fork is a great example of this: strike it once and it will continue to vibrate at a tone specific to its length. Longer tuning forks have lower resonant frequencies: their tones will be lower on the musical scale than shorter forks.

Imagine a row of progressively-sized tuning forks arranged side-by-side. They are all mounted on a common base, and that base is vibrated at the frequency of the measured AC voltage (or current) by means of an electromagnet. Whichever tuning fork is closest in resonant frequency to the frequency of that vibration will tend to shake the most (or the loudest). If the forks' tines were flimsy enough, we could see the relative motion of each by the length of the blur we would see as we inspected each one from an end-view perspective. Well, make a collection of "tuning forks" out of a strip of sheet metal cut in a pattern akin to a rake, and you have the *vibrating reed* frequency meter:

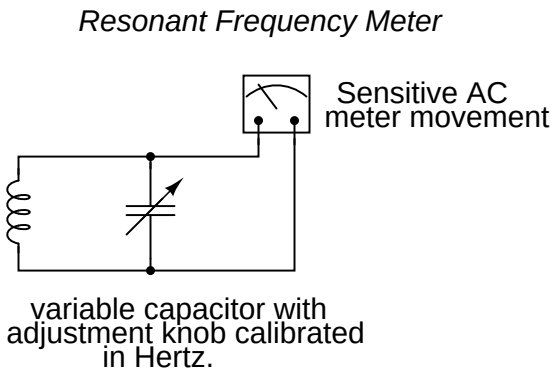


The user of this meter views the ends of all those unequal length reeds as they are collectively shaken at the frequency of the applied AC voltage to the coil. The one closest in resonant frequency to the applied AC will vibrate the most, looking something like this:



Vibrating reed meters, obviously, are not precision instruments, but they are very simple and therefore easy to manufacture to be rugged. They are often found on small engine-driven generator sets for the purpose of setting engine speed so that the frequency is somewhat close to 60 (50 in Europe) Hertz.

While reed-type meters are imprecise, their operational principle is not. In lieu of mechanical resonance, we may substitute electrical resonance and design a frequency meter using an inductor and capacitor in the form of a tank circuit (parallel inductor and capacitor). One or both components are made adjustable, and a meter is placed in the circuit to indicate maximum amplitude of voltage across the two components. The adjustment knob(s) are calibrated to show resonant frequency for any given setting, and the frequency is read from them after the device has been adjusted for maximum indication on the meter. Essentially, this is a tunable filter circuit which is adjusted and then read in a manner similar to a bridge circuit (which must be balanced for a "null" condition and then read).



This technique is a popular one for amateur radio operators (or at least it was before the advent of inexpensive digital frequency instruments called *counters*), especially because it doesn't require direct connection to the circuit. So long as the inductor and/or capacitor can intercept enough stray field (magnetic or electric, respectively) from the circuit under test to cause the meter to indicate, it will work.

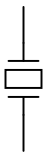
In frequency as in other types of electrical measurement, the most accurate means of measurement are usually those where an unknown quantity is compared against a known *standard*, the basic instrument doing nothing more than indicating when the two quantities are equal to each other. This is the basic principle behind the DC (Wheatstone) bridge circuit and it is a sound metrological principle applied throughout the sciences. If we have access to an accurate frequency standard (a source of AC voltage holding very precisely to a single frequency), then measurement of any unknown frequency by comparison should be relatively easy.

For that frequency standard, we turn our attention back to the tuning fork, or at least a more modern variation of it called the *quartz crystal*. Quartz is a naturally occurring mineral possessing a very interesting property called *piezoelectricity*. Piezoelectric materials produce a voltage across their length when physically stressed, and will physically deform when an external voltage is applied across their lengths. This deformation is very, very slight in most cases, but it does exist.

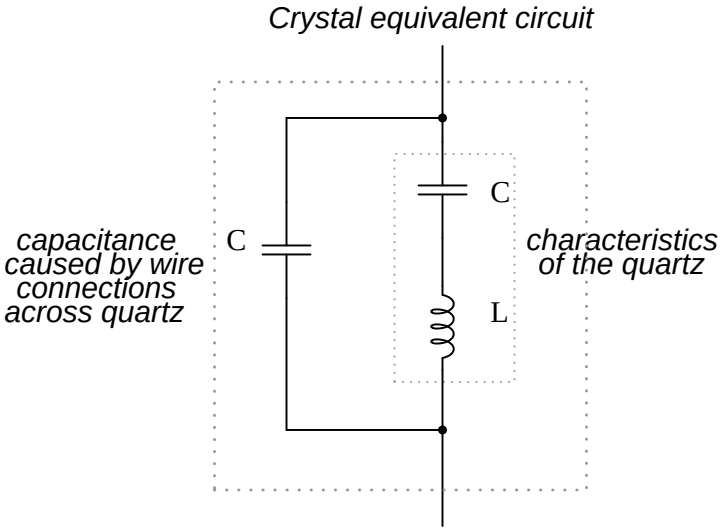
Quartz rock is elastic (springy) within that small range of bending which an external voltage would produce, which means that it will have a mechanical resonant frequency of its own capable of being manifested as an electrical voltage signal. In other words, if a chip of quartz is struck, it will "ring" with its own unique frequency determined by the length of the chip, and that resonant oscillation will produce an equivalent voltage across multiple points of the quartz chip which can be tapped into by wires fixed to the surface of the chip. In reciprocal manner, the quartz chip will tend to vibrate most when it is "excited" by an applied AC voltage at precisely the right frequency, just like the reeds on a vibrating-reed frequency meter.

Chips of quartz rock can be precisely cut for desired resonant frequencies, and that chip mounted securely inside a protective shell with wires extending for connection to an external electric circuit. When packaged as such, the resulting device is simply called a *crystal* (or sometimes "*xtal*"), and its schematic symbol looks like this:

crystal or *xtal*



Electrically, that quartz chip is equivalent to a series LC resonant circuit. The dielectric properties of quartz contribute an additional capacitive element to the equivalent circuit, and in the end it looks something like this:



The "capacitance" and "inductance" shown in series are merely electrical equivalents of the quartz's mechanical resonance properties: they do not exist as discrete components within the crystal. The capacitance shown in parallel due to the wire connections across the dielectric (insulating) quartz body is real, and it has an effect on the resonant response of the whole system. A full discussion on crystal dynamics is not necessary here, but what needs to be understood about crystals is this resonant circuit equivalence and how it can be exploited within an oscillator circuit to achieve an output voltage with a stable, known frequency.

Crystals, as resonant elements, typically have much higher "Q" (*quality*) values than tank circuits built from inductors and capacitors, principally due to the relative absence of stray resistance, making their resonant frequencies very definite and precise. Because the resonant frequency is solely dependent on the physical properties of quartz (a very stable substance, mechanically), the resonant frequency variation over time with a quartz crystal is very, very low. This is how *quartz movement* watches obtain their high accuracy: by means of an electronic oscillator stabilized by the resonant action of a quartz crystal.

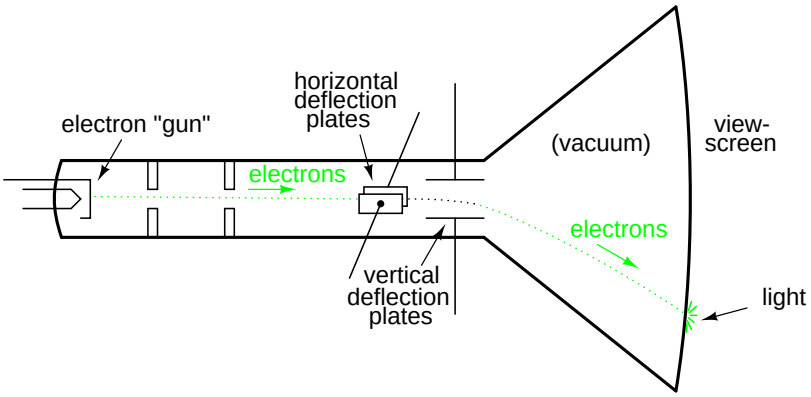
For laboratory applications, though, even greater frequency stability may be desired. To achieve this, the crystal in question may be placed in a temperature stabilized environment (usually an oven), thus eliminating frequency errors due to thermal expansion and contraction of the quartz.

For the ultimate in a frequency standard though, nothing discovered thus far surpasses the

accuracy of a single resonating atom. This is the principle of the so-called *atomic clock*, which uses an atom of mercury (or cesium) suspended in a vacuum, excited by outside energy to resonate at its own unique frequency. The resulting frequency is detected as a radio-wave signal and that forms the basis for the most accurate clocks known to humanity. National standards laboratories around the world maintain a few of these hyper-accurate clocks, and broadcast frequency signals based on those atoms' vibrations for scientists and technicians to tune in and use for frequency calibration purposes.

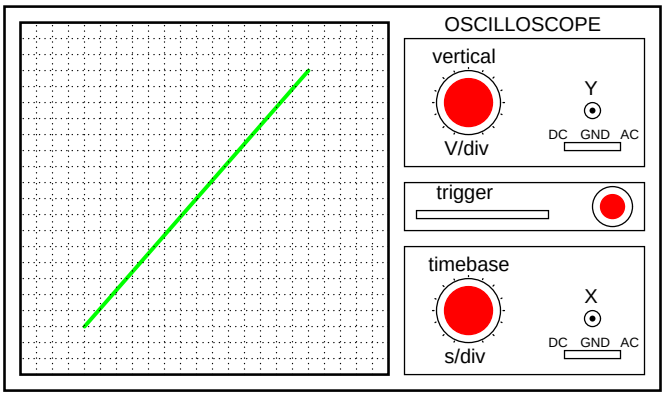
Now we get to the practical part: once we have a *source* of accurate frequency, how do we compare that against an unknown frequency to obtain a measurement? One way is to use a CRT as a frequency-comparison device. Cathode Ray Tubes typically have means of deflecting the electron beam in the horizontal as well as the vertical axis. If metal plates are used to electrostatically deflect the electrons, there will be a pair of plates to the left and right of the beam as well as a pair of plates above and below the beam.

CRT with horizontal and vertical deflection plates



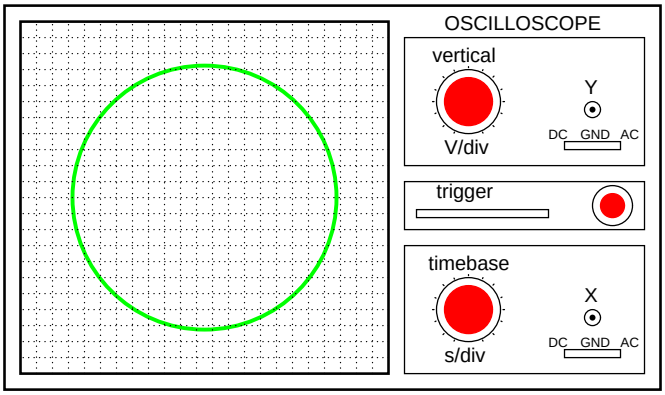
If we allow one AC signal to deflect the beam up and down (connect that AC voltage source to the "vertical" deflection plates) and another AC signal to deflect the beam left and right (using the other pair of deflection plates), patterns will be produced on the screen of the CRT indicative of the *ratio* of these two AC frequencies. These patterns are called *Lissajous figures* and are a common means of comparative frequency measurement in electronics.

If the two frequencies are the same, we will obtain a simple figure on the screen of the CRT, the shape of that figure being dependent upon the phase shift between the two AC signals. Here is a sampling of Lissajous figures for two sine-wave signals of equal frequency, shown as they would appear on the face of an oscilloscope (an AC voltage-measuring instrument using a CRT as its "movement"). The first picture is of the Lissajous figure formed by two AC voltages perfectly in phase with each other:



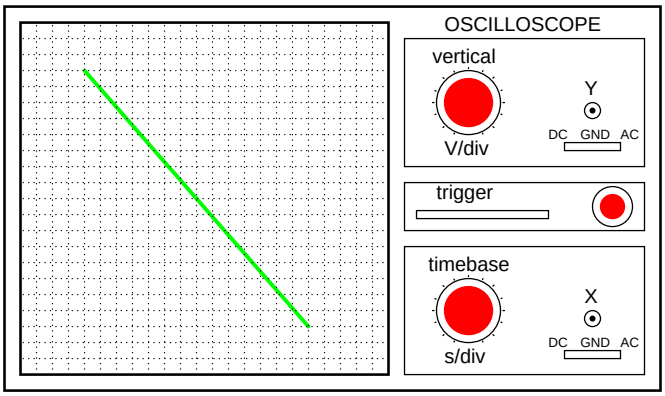
Lissajous figure: same frequency, 0 degrees phase shift

If the two AC voltages are not in phase with each other, a straight line will not be formed. Rather, the Lissajous figure will take on the appearance of an oval, becoming perfectly circular if the phase shift is exactly 90^o between the two signals, and if their amplitudes are equal:



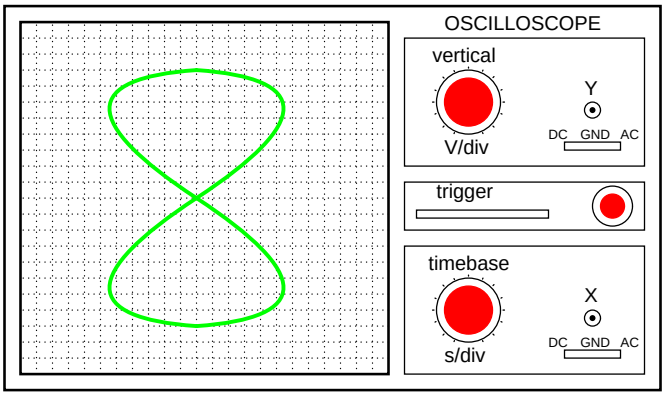
Lissajous figure: same frequency, 90 or 270 degrees phase shift

Finally, if the two AC signals are directly opposing one another in phase (180^o shift), we will end up with a line again, only this time it will be oriented in the opposite direction:



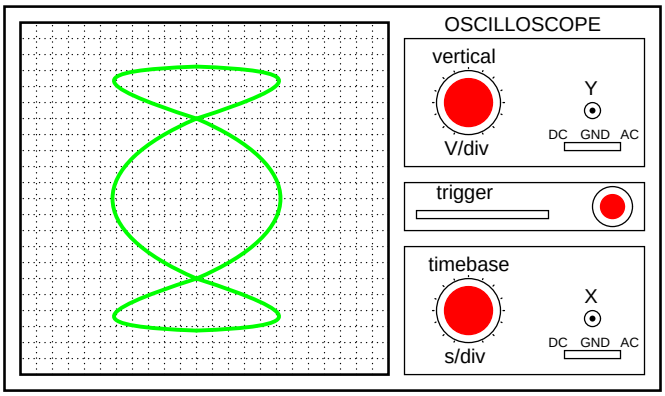
Lissajous figure: same frequency, 180 degrees phase shift

When we are faced with signal frequencies that are not the same, Lissajous figures get quite a bit more complex. Consider the following examples and their given vertical/horizontal frequency ratios:



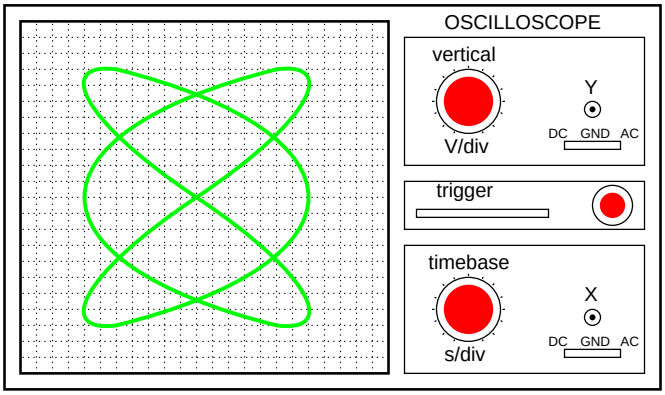
Lissajous figure: Horizontal frequency is twice that of vertical

The more complex the ratio between horizontal and vertical frequencies, the more complex the Lissajous figure. Consider the following illustration of a 3:1 frequency ratio between horizontal and vertical:



Lissajous figure: Horizontal frequency is three times that of vertical

. . . and a 3:2 frequency ratio (horizontal = 3, vertical = 2):



Lissajous figure: Horizontal/Vertical frequency ratio is 3:2

In cases where the frequencies of the two AC signals are not exactly a simple ratio of each other (but close), the Lissajous figure will appear to "move," slowly changing orientation as the phase angle between the two waveforms rolls between 0^o and 180^o. If the two frequencies are locked in an exact integer ratio between each other, the Lissajous figure will be stable on the viewscreen of the CRT.

The physics of Lissajous figures limits their usefulness as a frequency-comparison technique to cases where the frequency ratios are simple integer values (1:1, 1:2, 1:3, 2:3, 3:4, etc.). Despite this limitation, Lissajous figures are a popular means of frequency comparison wherever an accessible frequency standard (signal generator) exists.

- **REVIEW:**
- Some frequency meters work on the principle of mechanical resonance, indicating frequency by relative oscillation among a set of uniquely tuned "reeds" shaken at the measured frequency.
- Other frequency meters use electric resonant circuits (LC tank circuits, usually) to indicate frequency. One or both components is made to be adjustable, with an accurately calibrated

adjustment knob, and a sensitive meter is read for maximum voltage or current at the point of resonance.

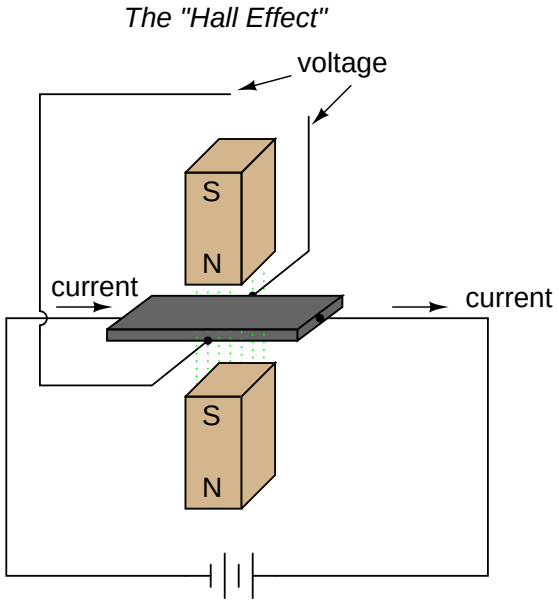
- Frequency can be measured in a comparative fashion, as is the case when using a CRT to generate *Lissajous figures*. Reference frequency signals can be made with a high degree of accuracy by oscillator circuits using quartz crystals as resonant devices. For ultra precision, atomic clock signal standards (based on the resonant frequencies of individual atoms) can be used.

12.3 Power measurement

Power measurement in AC circuits can be quite a bit more complex than with DC circuits for the simple reason that phase shift makes complicates the matter beyond multiplying voltage by current figures obtained with meters. What is needed is an instrument able to determine the product (multiplication) of *instantaneous* voltage and current. Fortunately, the common electrodynamometer movement with its stationary and moving coil does a fine job of this.

Three phase power measurement can be accomplished using two dynamometer movements with a common shaft linking the two moving coils together so that a single pointer registers power on a meter movement scale. This, obviously, makes for a rather expensive and complex movement mechanism, but it is a workable solution.

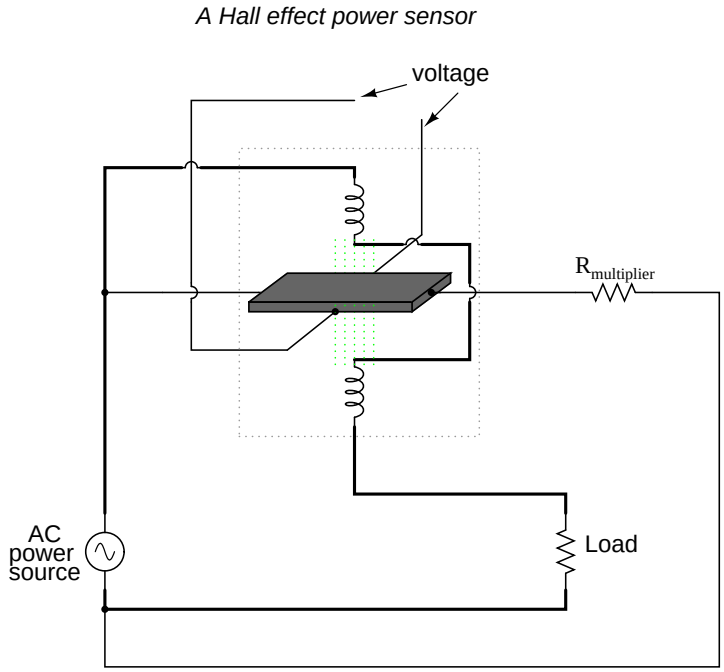
An ingenious method of deriving an electronic power meter (one that generates an electric signal representing power in the system rather than merely move a pointer) is based on the Hall effect. The Hall effect is an unusual effect first noticed by E. H. Hall in 1879, whereby a voltage is generated along the width of a current-carrying conductor exposed to a perpendicular magnetic field:



The voltage generated across the width of the flat, rectangular conductor is directly proportional to both the magnitude of the current through it and the strength of the magnetic field. Mathemati-

cally, it is a product (multiplication) of these two variables. The amount of "Hall Voltage" produced for any given set of conditions also depends on the type of material used for the flat, rectangular conductor. It has been found that specially prepared "semiconductor" materials produce a greater Hall voltage than do metals, and so modern Hall Effect devices are made of these.

It makes sense then that if we were to build a device using a Hall-effect sensor where the current through the conductor was pushed by AC voltage from an external circuit and the magnetic field was set up by a pair or wire coils energized by the current of the AC power circuit, the Hall voltage would be in direct proportion to the multiple of circuit current and voltage. Having no mass to move (unlike an electromechanical movement), this device is able to provide *instantaneous* power measurement:



Not only will the output voltage of the Hall effect device be the representation of instantaneous power at any point in time, but it will also be a DC signal! This is because the Hall voltage polarity is dependent upon *both* the polarity of the magnetic field and the direction of current through the conductor. If both current direction and magnetic field polarity reverses – as it would ever half-cycle of the AC power – the output voltage polarity will stay the same.

If voltage and current in the power circuit are 90^o out of phase (a power factor of zero, meaning *no* real power delivered to the load), the alternate peaks of Hall device current and magnetic field will never coincide with each other: when one is at its peak, the other will be zero. At those points in time, the Hall output voltage will likewise be zero, being the product (multiplication) of current and magnetic field strength. Between those points in time, the Hall output voltage will fluctuate equally between positive and negative, generating a signal corresponding to the instantaneous absorption and release of power through the reactive load. The net DC output voltage will be zero, indicating zero true power in the circuit.

Any phase shift between voltage and current in the power circuit less than 90^o will result in

a Hall output voltage that oscillates between positive and negative, but spends more time positive than negative. Consequently there will be a net DC output voltage. Conditioned through a low-pass filter circuit, this net DC voltage can be separated from the AC mixed with it, the final output signal registered on a sensitive DC meter movement.

Often it is useful to have a meter to totalize power usage over a period of time rather than instantaneously. The output of such a meter can be set in units of Joules, or total energy consumed, since *power* is a measure of work being done *per* unit time. Or, more commonly, the output of the meter can be set in units of Watt-Hours.

Mechanical means for measuring Watt-Hours are usually centered around the concept of the motor: build an AC motor that spins at a rate of speed proportional to the instantaneous power in a circuit, then have that motor turn an "odometer" style counting mechanism to keep a running total of energy consumed. The "motor" used in these meters has a rotor made of a thin aluminum disk, with the rotating magnetic field established by sets of coils energized by line voltage and load current so that the rotational speed of the disk is dependent on both voltage and current.

12.4 Power quality measurement

It used to be with large AC power systems that "power quality" was an unheard-of concept, aside from power factor. Almost all loads were of the "linear" variety, meaning that they did not distort the shape of the voltage sine wave, or cause non-sinusoidal currents to flow in the circuit. This is not true anymore. Loads controlled by "nonlinear" electronic components are becoming more prevalent in both home and industry, meaning that the voltages and currents in the power system(s) feeding these loads are rich in harmonics: what should be nice, clean sine-wave voltages and currents are becoming highly distorted, which is equivalent to the presence of an infinite series of high-frequency sine waves at multiples of the fundamental power line frequency.

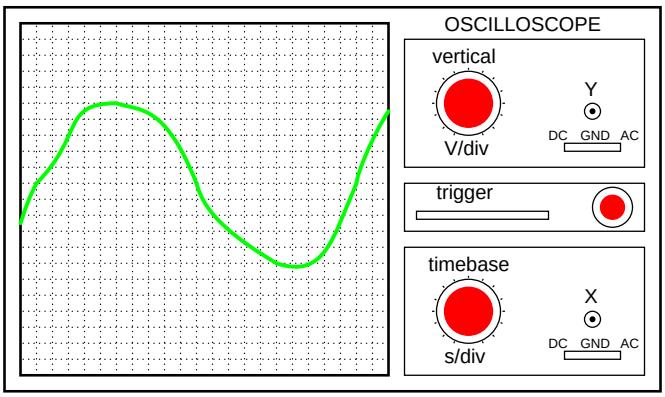
Excessive harmonics in an AC power system can overheat transformers, cause exceedingly high neutral conductor currents in three-phase systems, create electromagnetic "noise" in the form of radio emissions that can interfere with sensitive electronic equipment, reduce electric motor horsepower output, and can be difficult to pinpoint. With problems like these plaguing power systems, engineers and technicians require ways to precisely detect and measure these conditions.

Power Quality is the general term given to represent an AC power system's freedom from harmonic content. A "power quality" meter is one that gives some form of harmonic content indication.

A simple way for a technician to determine power quality in their system without sophisticated equipment is to compare voltage readings between two accurate voltmeters measuring the same system voltage: one meter being an "averaging" type of unit (such as an electromechanical movement meter) and the other being a "true-RMS" type of unit (such as a high-quality digital meter). Remember that "averaging" type meters are calibrated so that their scales indicate volts RMS, *based on the assumption that the AC voltage being measured is sinusoidal.* If the voltage is anything but sinewave-shaped, the averaging meter will *not* register the proper value, whereas the true-RMS meter always will, regardless of waveshape. The rule of thumb here is this: the greater the disparity between the two meters, the worse the power quality is, and the greater its harmonic content. A power system with good quality power should generate equal voltage readings between the two meters, to within the rated error tolerance of the two instruments.

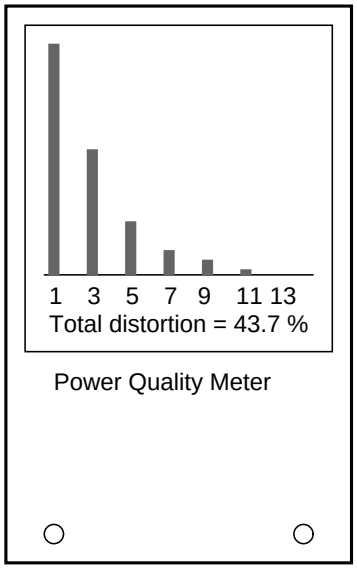
Another qualitative measurement of power quality is the oscilloscope test: connect an oscilloscope (CRT) to the AC voltage and observe the shape of the wave. Anything other than a clean sine wave

could be an indication of trouble:



This is a moderately ugly "sine" wave. Definite harmonic content here!

Still, if quantitative analysis (definite, numerical figures) is necessary, there is no substitute for an instrument specifically designed for that purpose. Such an instrument is called a *power quality meter* and is sometimes better known in electronic circles as a low-frequency *spectrum analyzer*. What this instrument does is provide a graphical representation on a CRT or digital display screen of the AC voltage's frequency "spectrum." Just as a prism splits a beam of white light into its constituent color components (how much red, orange, yellow, green, and blue is in that light), the spectrum analyzer splits a mixed-frequency signal into its constituent frequencies, and displays the result in the form of a histogram:



Each number on the horizontal scale of this meter represents a harmonic of the fundamental frequency. For American power systems, the "1" represents 60 Hz (the 1st harmonic, or *fundamental*), the "3" for 180 Hz (the 3rd harmonic), the "5" for 300 Hz (the 5th harmonic), and so on. The black rectangles represent the relative magnitudes of each of these harmonic components in the measured

AC voltage. A pure, 60 Hz sine wave would show only a tall black bar over the "1" with no black bars showing at all over the other frequency markers on the scale, because a pure sine wave has no harmonic content.

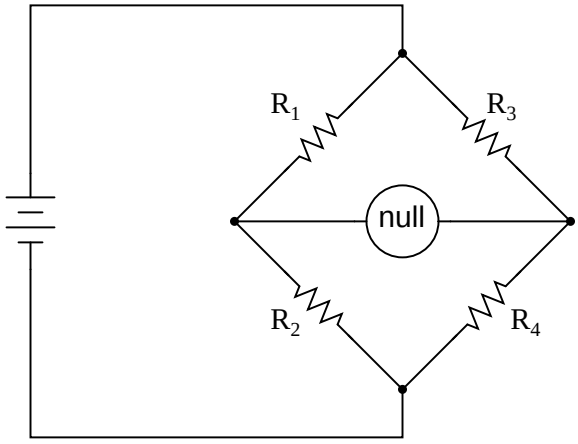
Power quality meters such as this might be better referred to as *overtone* meters, because they are designed to display only those frequencies known to be generated by the power system. In three-phase AC power systems (predominant for large power applications), even-numbered harmonics tend to be canceled out, and so only harmonics existing in significant measure are the odd-numbered.

Meters like these are very useful in the hands of a skilled technician, because different types of nonlinear loads tend to generate different spectrum "signatures" which can clue the troubleshooter to the source of the problem. These meters work by very quickly sampling the AC voltage at many different points along the waveform shape, digitizing those points of information, and using a microprocessor (small computer) to perform numerical Fourier analysis (the *Fast Fourier Transform* or "*FFT*" algorithm) on those data points to arrive at harmonic frequency magnitudes. The process is not much unlike what the SPICE program tells a computer to do when performing a Fourier analysis on a simulated circuit voltage or current waveform.

12.5 AC bridge circuits

As we saw with DC measurement circuits, the circuit configuration known as a *bridge* can be a very useful way to measure unknown values of resistance. This is true with AC as well, and we can apply the very same principle to the accurate measurement of unknown impedances.

To review, the bridge circuit works as a pair of two-component voltage dividers connected across the same source voltage, with a *null-detector* meter movement connected between them to indicate a condition of "balance" at zero volts:



Any one of the four resistors in the above bridge can be the resistor of unknown value, and its value can be determined by a ratio of the other three, which are "calibrated," or whose resistances are known to a precise degree. When the bridge is in a balanced condition (zero voltage as indicated by the null detector), the ratio works out to be this:

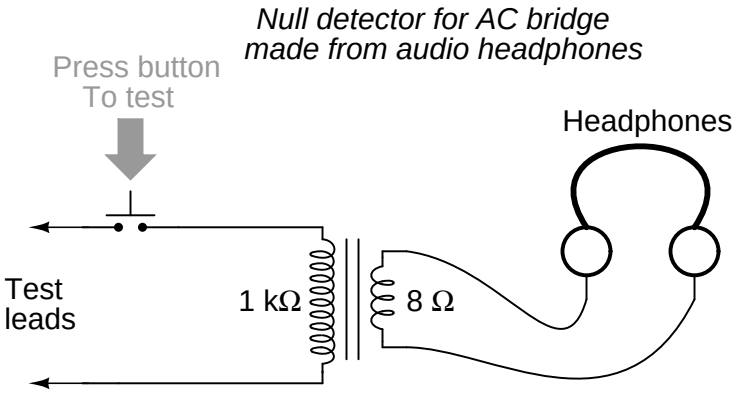
In a condition of ***balance****:*

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}$$

One of the advantages of using a bridge circuit to measure resistance is that the voltage of the power source is irrelevant. Practically speaking, the higher the supply voltage, the easier it is to detect a condition of imbalance between the four resistors with the null detector, and thus the more sensitive it will be. A greater supply voltage leads to the possibility of increased measurement precision. However, there will be no fundamental error introduced as a result of a lesser or greater power supply voltage unlike other types of resistance measurement schemes.

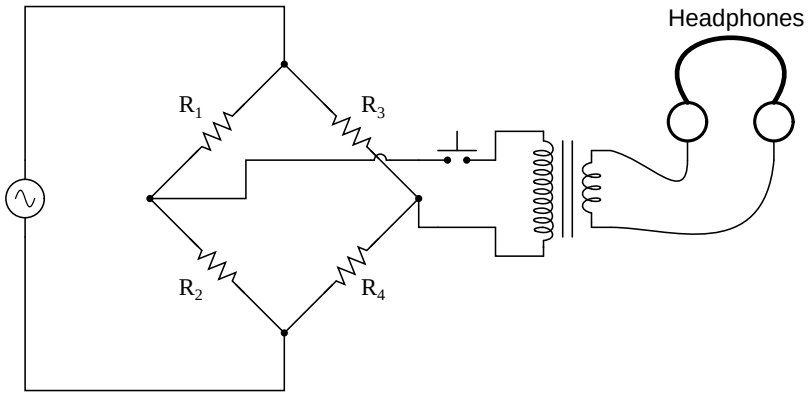
Impedance bridges work the same, only the balance equation is with *complex* quantities, as both magnitude and phase across the components of the two dividers must be equal in order for the null detector to indicate "zero." The null detector, of course, must be a device capable of detecting very small AC voltages. An oscilloscope is often used for this, although very sensitive electromechanical meter movements and even headphones (small speakers) may be used if the source frequency is within audio range.

One way to maximize the effectiveness of audio headphones as a null detector is to connect them to the signal source through an impedance-matching transformer. Headphone speakers are typically low-impedance units (8 Ω), requiring substantial current to drive, and so a step-down transformer helps "match" low-current signals to the impedance of the headphone speakers. An audio output transformer works well for this purpose:



Using a pair of headphones that completely surround the ears (the "closed-cup" type), I've been able to detect currents of less than 0.1 μA with this simple detector circuit. Roughly equal performance was obtained using two different step-down transformers: a small power transformer (120/6 volt ratio), and an audio output transformer (1000:8 ohm impedance ratio). With the pushbutton switch in place to interrupt current, this circuit is usable for detecting signals from DC to over 2 MHz: even if the frequency is far above or below the audio range, a "click" will be heard from the headphones each time the switch is pressed and released.

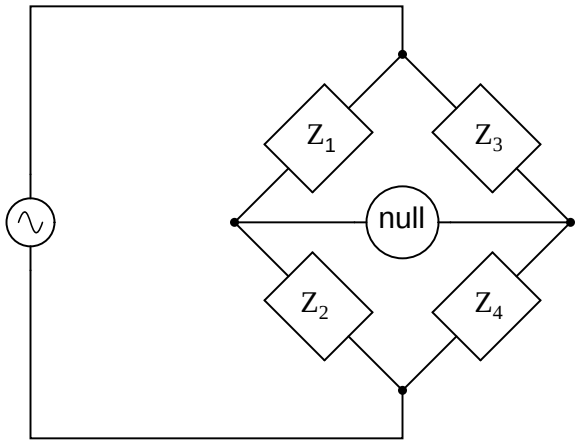
Connected to a resistive bridge, the whole circuit looks like this:



Listening to the headphones as one or more of the resistor "arms" of the bridge is adjusted, a condition of balance will be realized when the headphones fail to produce "clicks" (or tones, if the bridge's power source frequency is within audio range) as the switch is actuated.

When describing general AC bridges, where *impedances* and not just resistances must be in proper ratio for balance, it is sometimes helpful to draw the respective bridge legs in the form of box-shaped components, each one with a certain impedance:

General impedance bridge



A box with a "Z" written inside is the symbol for any nonspecific impedance.

For this general form of AC bridge to balance, the impedance ratios of each branch must be equal:

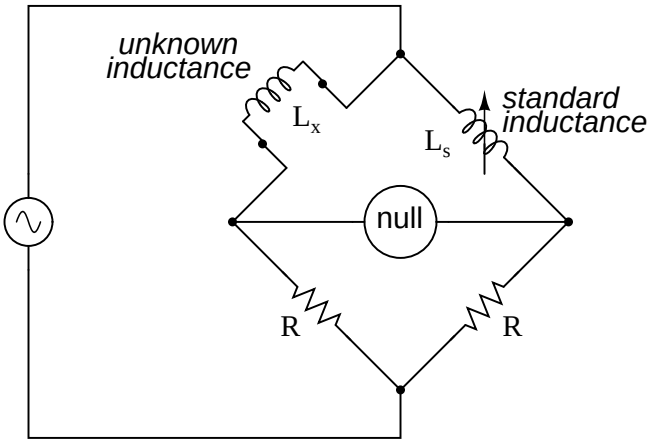
$$\frac{Z_1}{Z_2} = \frac{Z_3}{Z_4}$$

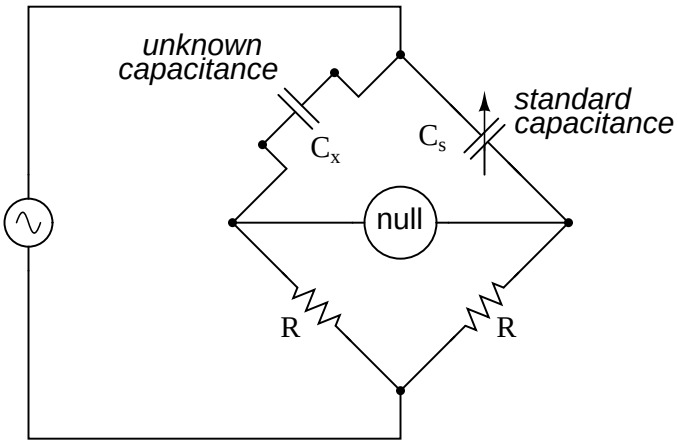
Again, it must be stressed that the impedance quantities in the above equation *must* be complex,

accounting for both magnitude and phase angle. It is insufficient that the impedance magnitudes alone be balanced; without phase angles in balance as well, there will still be voltage across the terminals of the null detector and the bridge will not be balanced.

Bridge circuits can be constructed to measure just about any device value desired, be it capacitance, inductance, resistance, or even "Q." As always in bridge measurement circuits, the unknown quantity is always "balanced" against a known standard, obtained from a high-quality, calibrated component that can be adjusted in value until the null detector device indicates a condition of balance. Depending on how the bridge is set up, the unknown component's value may be determined directly from the setting of the calibrated standard, or derived from that standard through a mathematical formula.

A couple of simple bridge circuits are shown below, one for inductance and one for capacitance:



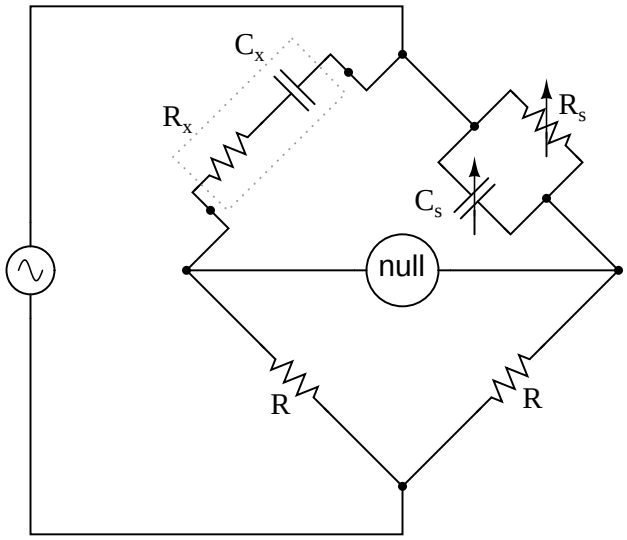


Simple "symmetrical" bridges such as these are so named because they exhibit symmetry (mirror-image similarity) from left to right. The two bridge circuits shown above are balanced by adjusting the calibrated reactive component (L_s or C_s). They are a bit simplified from their real-life counterparts, as practical symmetrical bridge circuits often have a calibrated, variable resistor in series

or parallel with the reactive component to balance out stray resistance in the unknown component. But, in the hypothetical world of perfect components, these simple bridge circuits do just fine to illustrate the basic concept.

An example of a little extra complexity added to compensate for real-world effects can be found in the so-called *Wien bridge*, which uses a parallel capacitor-resistor standard impedance to balance out an unknown series capacitor-resistor combination. All capacitors have some amount of internal resistance, be it literal or equivalent (in the form of dielectric heating losses) which tend to spoil their otherwise perfectly reactive natures. This internal resistance may be of interest to measure, and so the Wien bridge attempts to do so by providing a balancing impedance that isn't "pure" either:

The Wien bridge

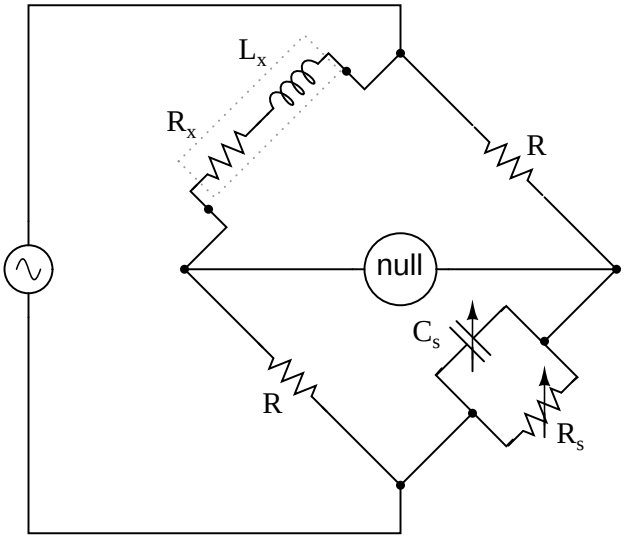


Being that there are two standard components to be adjusted (a resistor and a capacitor) this bridge will take a little more time to balance than the others we've seen so far. The combined effect of R_s and C_s is to alter the magnitude and phase angle until the bridge achieves a condition of balance. Once that balance is achieved, the settings of R_s and C_s can be read from their calibrated knobs, the parallel impedance of the two determined mathematically, and the unknown capacitance and resistance determined mathematically from the balance equation ($Z_1/Z_2 = Z_3/Z_4$).

It is assumed in the operation of the Wien bridge that the standard capacitor has negligible internal resistance, or at least that resistance is already known so that it can be factored into the balance equation. Wien bridges are useful for determining the values of "lossy" capacitor designs like electrolytics, where the internal resistance is relatively high. They are also used as frequency meters, because the balance of the bridge is frequency-dependent. When used in this fashion, the capacitors are made fixed (and usually of equal value) and the top two resistors are made variable and are adjusted by means of the same knob.

An interesting variation on this theme is found in the next bridge circuit, used to precisely measure inductances.

The Maxwell-Wien bridge



This ingenious bridge circuit is known as the *Maxwell-Wien bridge* (sometimes known plainly as the *Maxwell bridge*), and is used to measure unknown inductances in terms of calibrated resistance and capacitance. Calibration-grade inductors are more difficult to manufacture than capacitors of similar precision, and so the use of a simple "symmetrical" inductance bridge is not always practical. Because the phase shifts of inductors and capacitors are exactly opposite each other, a capacitive impedance can balance out an inductive impedance if they are located in opposite legs of a bridge, as they are here.

Another advantage of using a Maxwell bridge to measure inductance rather than a symmetrical inductance bridge is the elimination of measurement error due to mutual inductance between two inductors. Magnetic fields can be difficult to shield, and even a small amount of coupling between coils in a bridge can introduce substantial errors in certain conditions. With no second inductor to react with in the Maxwell bridge, this problem is eliminated.

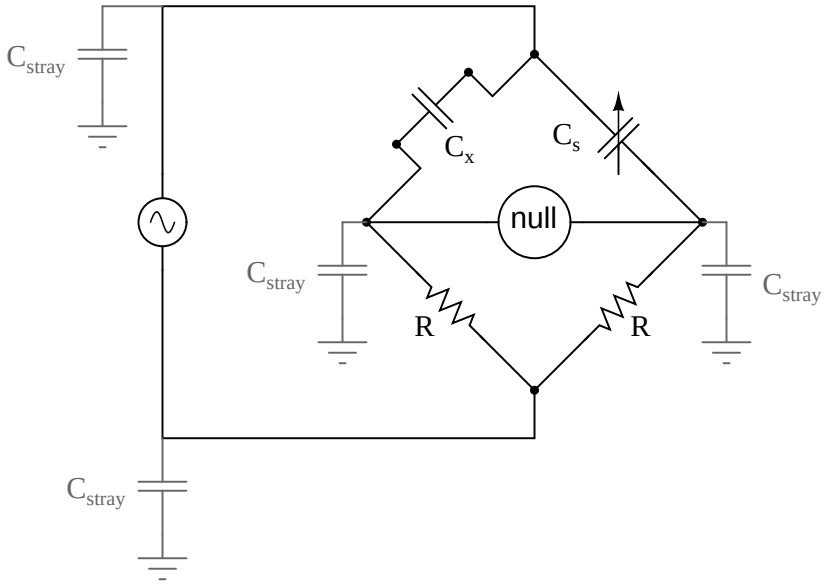
For easiest operation, the standard capacitor (C_s) and the resistor in parallel with it (R_s) are made variable, and both must be adjusted to achieve balance. However, the bridge can be made to work if the capacitor is fixed (non-variable) and more than one resistor made variable (at least the resistor in parallel with the capacitor, and one of the other two). However, in the latter configuration it takes more trial-and-error adjustment to achieve balance, as the different variable resistors interact in balancing magnitude and phase.

Unlike the plain Wien bridge, the balance of the Maxwell-Wien bridge is independent of source frequency, and in some cases this bridge can be made to balance in the presence of mixed frequencies from the AC voltage source, the limiting factor being the inductor's stability over a wide frequency range.

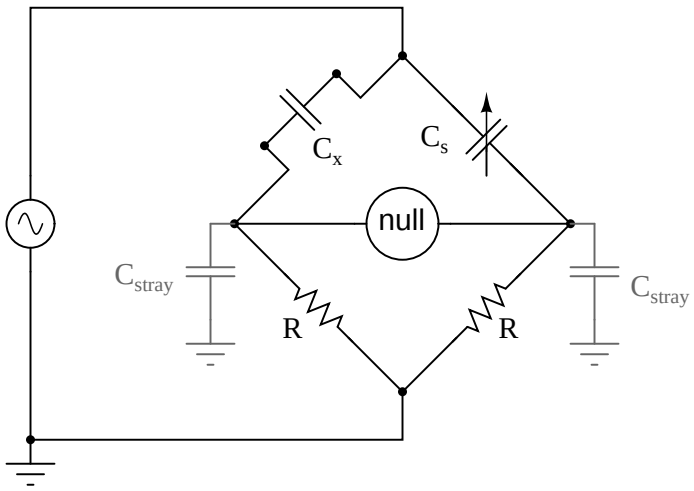
There are more variations beyond these designs, but a full discussion is not warranted here. General-purpose impedance bridge circuits are manufactured which can be switched into more than one configuration for maximum flexibility of use.

A potential problem in sensitive AC bridge circuits is that of stray capacitance between either

end of the null detector unit and ground (earth) potential. Because capacitances can "conduct" alternating current by charging and discharging, they form stray current paths to the AC voltage source which may affect bridge balance:

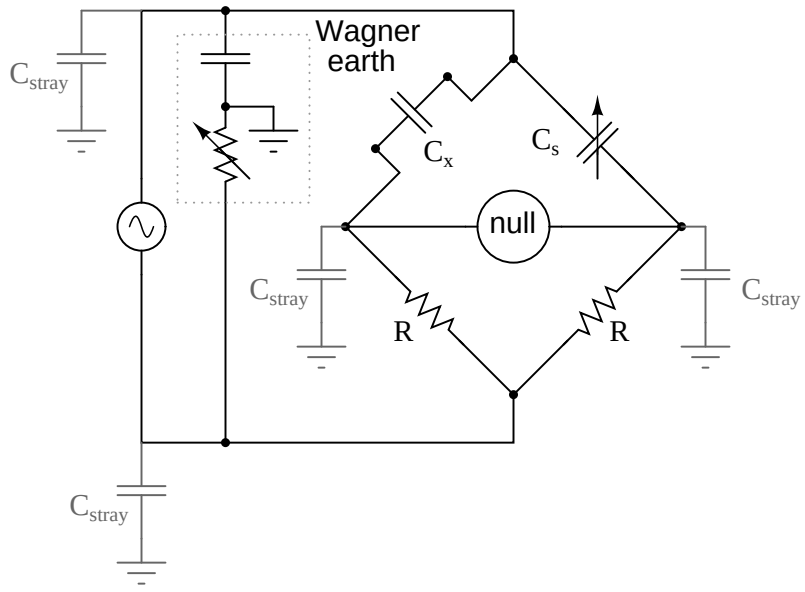


The problem is worsened if the AC voltage source is firmly grounded at one end, the total stray impedance for leakage currents made far less and any leakage currents through these stray capacitances made greater as a result:



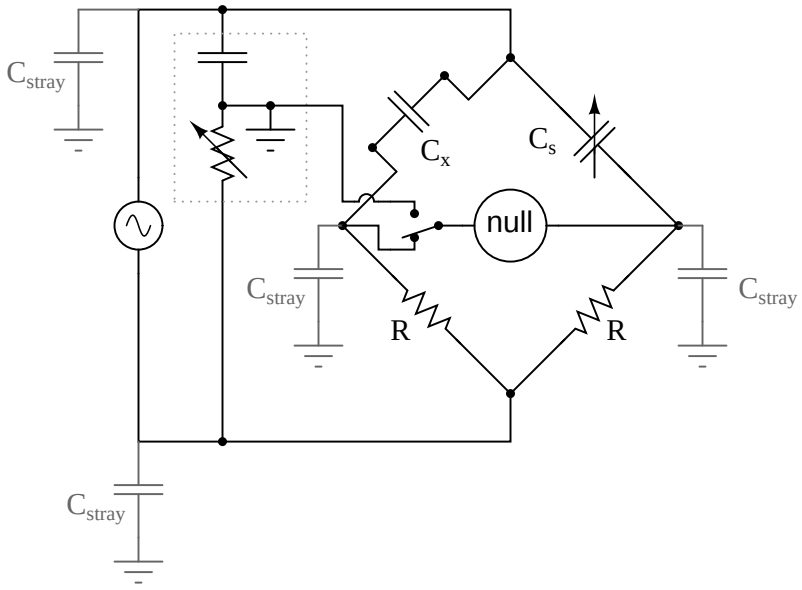
One way of greatly reducing this effect is to keep the null detector at ground potential, so there will be no AC voltage between it and the ground, and thus no current through stray capacitances. However, directly connecting the null detector to ground is not an option, as it would create a *direct* current path for stray currents, which would be worse than any capacitive path. Instead, a special

voltage divider circuit called a *Wagner ground* or *Wagner earth* may be used to maintain the null detector at ground potential without the need for a direct connection to the null detector.



The Wagner earth circuit is nothing more than a voltage divider, designed to have the voltage ratio and phase shift as each side of the bridge. Because the midpoint of the Wagner divider is directly grounded, any other divider circuit (including either side of the bridge) having the same voltage proportions and phases as the Wagner divider, and powered by the same AC voltage source, will be at ground potential as well. Thus, the Wagner earth divider forces the null detector to be at ground potential, without a direct connection between the detector and ground.

There is often a provision made in the null detector connection to confirm proper setting of the Wagner earth divider circuit: a two-position switch, so that one end of the null detector may be connected to either the bridge or the Wagner earth. When the null detector registers zero signal in both switch positions, the bridge is not only guaranteed to be balanced, but the null detector is also guaranteed to be at zero potential with respect to ground, thus eliminating any errors due to leakage currents through stray detector-to-ground capacitances:



- **REVIEW:**

- AC bridge circuits work on the same basic principle as DC bridge circuits: that a balanced ratio of impedances (rather than resistances) will result in a "balanced" condition as indicated by the null-detector device.

- Null detectors for AC bridges may be sensitive electromechanical meter movements, oscilloscopes (CRT's), headphones (amplified or unamplified), or any other device capable of registering very small AC voltage levels. Like DC null detectors, its only required point of calibration accuracy is at zero.

- AC bridge circuits can be of the "symmetrical" type where an unknown impedance is balanced by a standard impedance of similar type on the same side (top or bottom) of the bridge. Or, they can be "nonsymmetrical," using parallel impedances to balance series impedances, or even capacitances balancing out inductances.

- AC bridge circuits often have more than one adjustment, since both impedance magnitude *and* phase angle must be properly matched to balance.

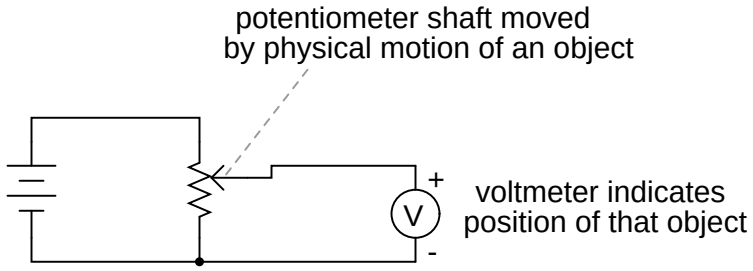
- Some impedance bridge circuits are frequency-sensitive while others are not. The frequency-sensitive types may be used as frequency measurement devices if all component values are accurately known.

- A *Wagner earth* or *Wagner ground* is a voltage divider circuit added to AC bridges to help reduce errors due to stray capacitance coupling the null detector to ground.

12.6 AC instrumentation transducers

Just as devices have been made to measure certain physical quantities and repeat that information in the form of DC electrical signals (thermocouples, strain gauges, pH probes, etc.), special devices have been made that do the same with AC.

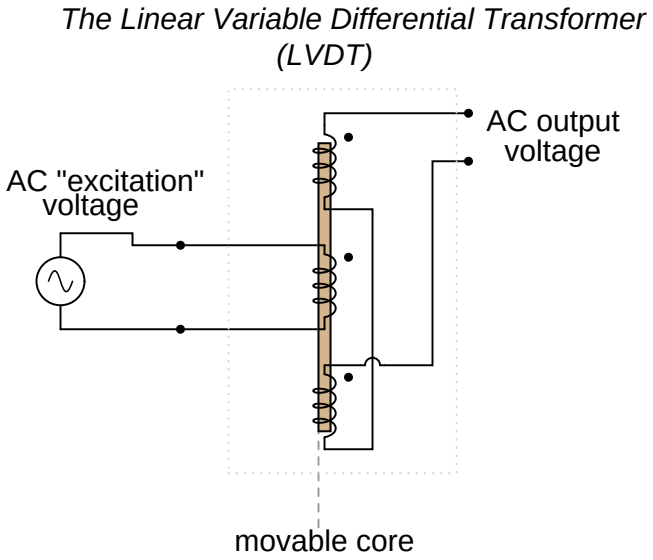
It is often necessary to be able to detect and transmit the physical position of mechanical parts via electrical signals. This is especially true in the fields of automated machine tool control and robotics. A simple and easy way to do this is with a potentiometer:



However, potentiometers have their own unique problems. For one, they rely on physical contact between the "wiper" and the resistance strip, which means they suffer the effects of physical wear over time. As potentiometers wear, their proportional output versus shaft position becomes less and less certain. You might have already experienced this effect when adjusting the volume control on an old radio: when twisting the knob, you might hear "scratching" sounds coming out of the speakers. Those noises are the result of poor wiper contact in the volume control potentiometer.

Also, this physical contact between wiper and strip creates the possibility of arcing (sparking) between the two as the wiper is moved. With most potentiometer circuits, the current is so low that wiper arcing is negligible, but it is a possibility to be considered. If the potentiometer is to be operated in an environment where combustible vapor or dust is present, this potential for arcing translates into a potential for an explosion!

Using AC instead of DC, we are able to completely avoid sliding contact between parts if we use a *variable transformer* instead of a potentiometer. Devices made for this purpose are called LVDT's, which stands for **L**inear **V**ariable **D**ifferential **T**ransformers. The design of an LVDT looks like this:



Obviously, this device is a *transformer*: it has a single primary winding powered by an external source of AC voltage, and two secondary windings connected in series-bucking fashion. It is *variable* because the core is free to move between the windings. It is *differential* because of the way the two secondary windings are connected. Being arranged to oppose each other (180^o out of phase) means that the output of this device will be the *difference* between the voltage output of the two secondary windings. When the core is centered and both windings are outputting the same voltage, the net result at the output terminals will be zero volts. It is called *linear* because the core's freedom of motion is straight-line.

The AC voltage output by an LVDT indicates the position of the movable core. Zero volts means that the core is centered. The further away the core is from center position, the greater percentage of input ("excitation") voltage will be seen at the output. The phase of the output voltage relative to the excitation voltage indicates which direction from center the core is offset.

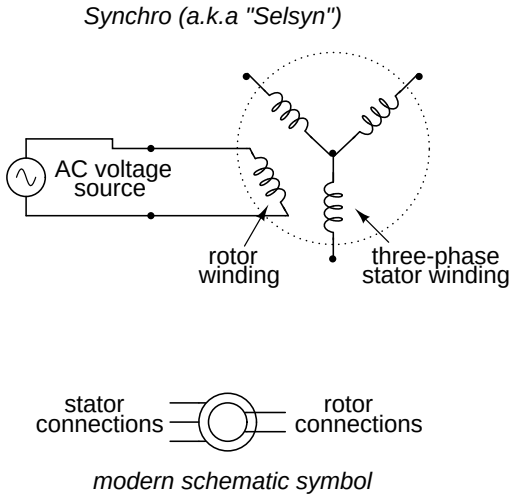
The primary advantage of an LVDT over a potentiometer for position sensing is the absence of physical contact between the moving and stationary parts. The core does not contact the wire windings, but slides in and out within a nonconducting tube. Thus, the LVDT does not "wear" like a potentiometer, nor is there the possibility of creating an arc.

Excitation of the LVDT is typically 10 volts RMS or less, at frequencies ranging from power line to the high audio (20 kHz) range. One potential disadvantage of the LVDT is its response time, which is mostly dependent on the frequency of the AC voltage source. If very quick response times are desired, the frequency must be higher to allow whatever voltage-sensing circuits enough cycles of AC to determine voltage level as the core is moved. To illustrate the potential problem here, imagine this exaggerated scenario: an LVDT powered by a 60 Hz voltage source, with the core being moved in and out hundreds of times per second. The output of this LVDT wouldn't even look like a sine wave because the core would be moved throughout its range of motion before the AC source voltage could complete a single cycle! It would be almost impossible to determine instantaneous core position if it moves faster than the instantaneous source voltage does.

A variation on the LVDT is the RVDT, or **R**otary **V**ariable **D**ifferential **T**ransformer. This device works on almost the same principle, except that the core revolves on a shaft instead of moving in a

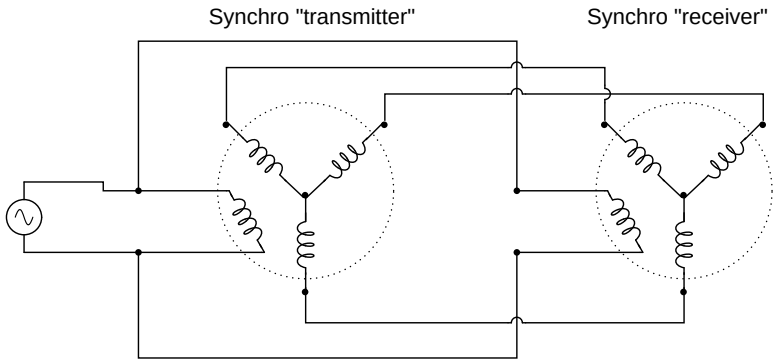
straight line. RVDT's can be constructed for limited motion of 360^o (full-circle) motion.

Continuing with this principle, we have what is known as a *Synchro* or *Selsyn*, which is a device constructed a lot like a wound-rotor polyphase AC motor or generator. The rotor is free to revolve a full 360^o, just like a motor. On the rotor is a single winding connected to a source of AC voltage, much like the primary winding of an LVDT. The stator windings are usually in the form of a three-phase Y, although synchros with more than three phases have been built:



Voltages induced in the stator windings from the rotor's AC excitation are *not* phase-shifted by 120^o as in a real three-phase generator. If the rotor were energized with DC current rather than AC and the shaft spun continuously, then the voltages would be true three-phase. But this is not how a synchro is designed to be operated. Rather, this is a *position-sensing* device much like an RVDT, except that its output signal is much more definite. With the rotor energized by AC, the stator winding voltages will be proportional in magnitude to the angular position of the rotor, phase either 0^o or 180^o shifted, like a regular LVDT or RVDT. You could think of it as a transformer with one primary winding and three secondary windings, each secondary winding oriented at a unique angle. As the rotor is slowly turned, each winding in turn will line up directly with the rotor, producing full voltage, while the other windings will produce something less than full voltage.

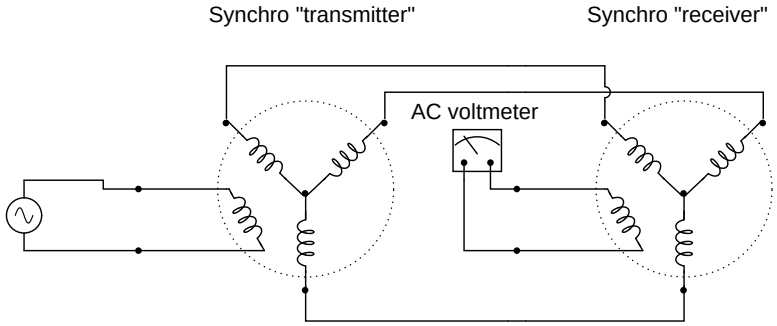
Synchros are often used in pairs. With their rotors connected in parallel and energized by the same AC voltage source, their shafts will match position to a high degree of accuracy:



The receiver rotor will turn to match position with the transmitter rotor so long as the two rotors remain energized.

Such "transmitter/receiver" pairs have been used on ships to relay rudder position, or to relay navigational gyro position over fairly long distances. The only difference between the "transmitter" and the "receiver" is which one gets turned by an outside force. The "receiver" can just as easily be used as the "transmitter" by forcing its shaft to turn and letting the synchro on the left match position.

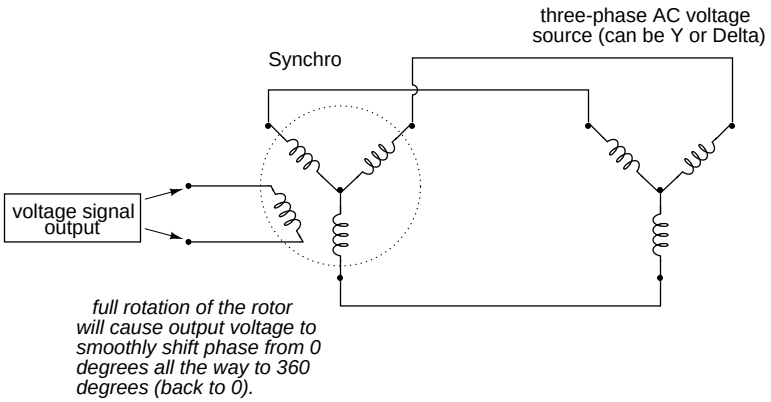
If the receiver's rotor is left unpowered, it will act as a position-error detector, generating an AC voltage at the rotor if the shaft is anything other than 90^o or 270^o shifted from the shaft position of the transmitter. The receiver rotor will no longer generate any torque and consequently will no longer automatically match position with the transmitter's:



AC voltmeter registers a voltage if receiver rotor position isn't exactly 90 or 270 degrees shifted from the position of the transmitter rotor.

This can be thought of almost as a sort of bridge circuit that achieves balance only if the receiver shaft is brought to one of two (matching) positions with the transmitter shaft.

One rather ingenious application of the synchro is in the creation of a phase-shifting device, provided that the stator is energized by three-phase AC:



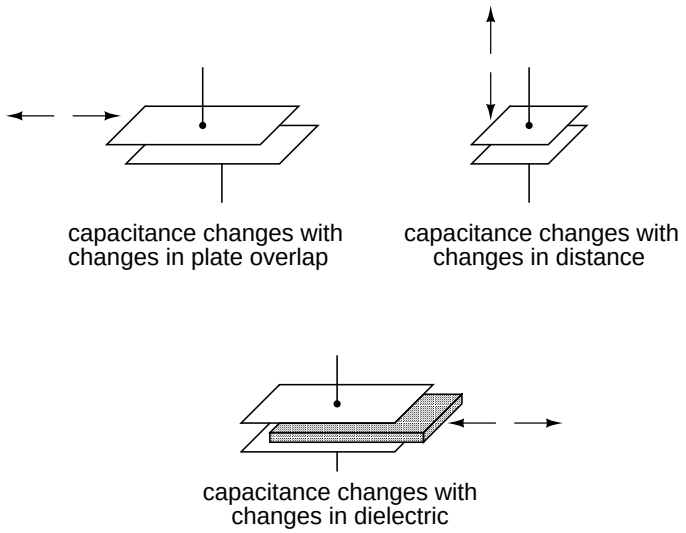
As the synchro's rotor is turned, the rotor coil will progressively align with each stator coil, their respective magnetic fields being 120^o phase-shifted from one another. In between those positions, these phase-shifted fields will mix to produce a rotor voltage somewhere between 0^o, 120^o, or 240^o shift. The practical result is a device capable of providing an infinitely variable-phase AC voltage with the twist of a knob (attached to the rotor shaft).

So far the transducers discussed have all been of the inductive variety. However, it is possible to make transducers which operate on variable capacitance as well, AC being used to sense the change in capacitance and generate a variable output voltage.

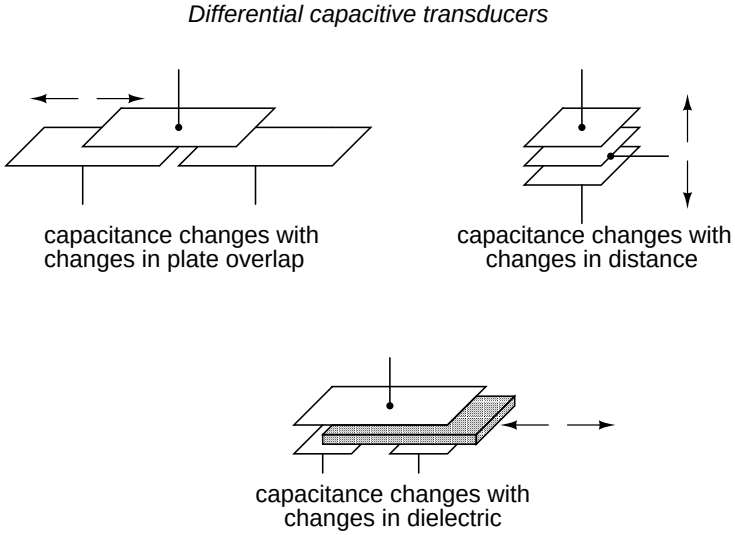
Remember that the capacitance between two conductive surfaces varies with three major factors: the overlapping area of those two surfaces, the distance between them, and the dielectric constant of the material in between the surfaces. If two out of three of these variables can be fixed (stabilized) and the third allowed to vary, then any measurement of capacitance between the surfaces will be solely indicative of changes in that third variable.

Medical researchers have long made use of capacitive sensing to detect physiological changes in living bodies. As early as 1907, a German researcher named H. Cremer placed two metal plates on either side of a beating frog heart and measured the capacitance changes resulting from the heart alternately filling and emptying itself of blood. Similar measurements have been performed on human beings with metal plates placed on the chest and back, recording respiratory and cardiac action by means of capacitance changes. For more precise capacitive measurements of organ activity, metal probes have been inserted into organs (especially the heart) on the tips of catheter tubes, capacitance being measured between the metal probe and the body of the subject. With a sufficiently high AC excitation frequency and sensitive enough voltage detector, not just the pumping action but also the *sounds* of the active heart may be readily interpreted.

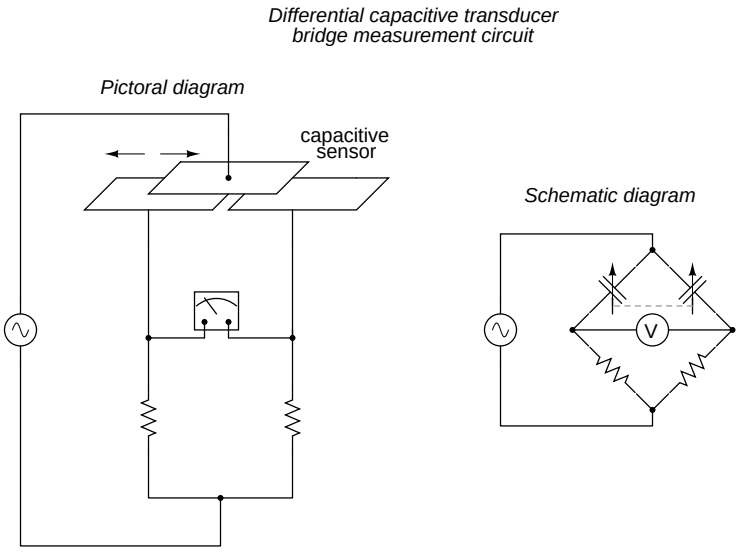
Like inductive transducers, capacitive transducers can also be made to be self-contained units, unlike the direct physiological examples described above. Some transducers work by making one of the capacitor plates movable, either in such a way as to vary the overlapping area or the distance between the plates. Other transducers work by moving a dielectric material in and out between two fixed plates:



Transducers with greater sensitivity and immunity to changes in other variables can be obtained by way of differential design, much like the concept behind the LVDT (Linear Variable *Differential* Transformer). Here are a few examples of differential capacitive transducers:



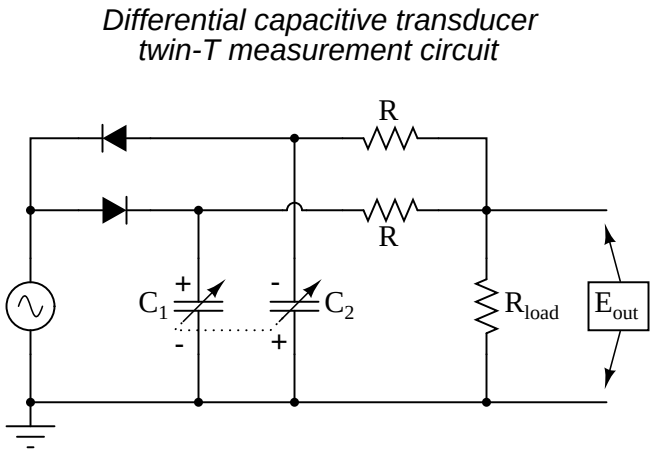
As you can see, all of the differential devices shown in the above illustration have *three* wire connections rather than two: one wire for each of the "end" plates and one for the "common" plate. As the capacitance between one of the "end" plates and the "common" plate changes, the capacitance between the other "end" plate and the "common" plate is such to change in the opposite direction. This kind of transducer lends itself very well to implementation in a bridge circuit:



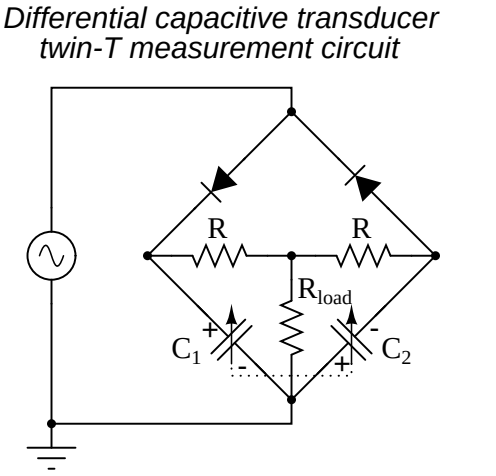
Capacitive transducers provide relatively small capacitances for a measurement circuit to operate with, typically in the *pico*farad range. Because of this, high power supply frequencies (in the megahertz range!) are usually required to reduce these capacitive reactances to reasonable levels. Given the small capacitances provided by typical capacitive transducers, stray capacitances have the potential of being major sources of measurement error. Good conductor shielding is *essential* for reliable and accurate capacitive transducer circuitry!

The bridge circuit is not the only way to effectively interpret the differential capacitance output of such a transducer, but it is one of the simplest to implement and understand. As with the LVDT, the voltage output of the bridge is proportional to the displacement of the transducer action from its center position, and the direction of offset will be indicated by phase shift. This kind of bridge circuit is similar in function to the kind used with strain gauges: it is not intended to be in a "balanced" condition all the time, but rather the degree of imbalance represents the magnitude of the quantity being measured.

An interesting alternative to the bridge circuit for interpreting differential capacitance is the *twin-T*. It requires the use of diodes, those "one-way valves" for electric current mentioned earlier in the chapter:



This circuit might be better understood if re-drawn to resemble more of a bridge configuration:



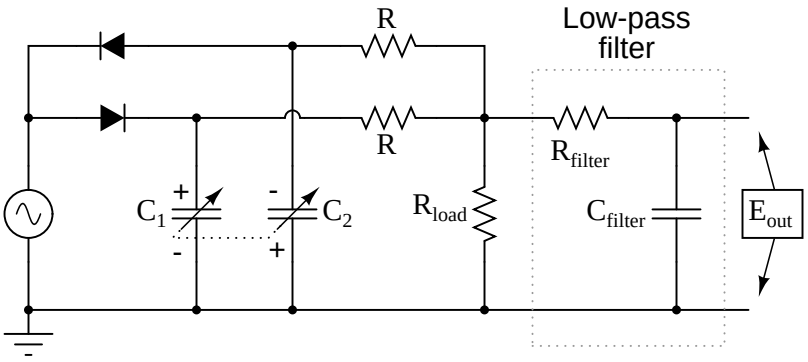
Capacitor C_1 is charged by the AC voltage source during every positive half-cycle (positive as measured in reference to the ground point), while C_2 is charged during every negative half-cycle. While one capacitor is being charged, the other capacitor discharges (at a slower rate than it was charged) through the three-resistor network. As a consequence, C_1 maintains a positive DC voltage with respect to ground, and C_2 a negative DC voltage with respect to ground.

If the capacitive transducer is displaced from center position, one capacitor will increase in capacitance while the other will decrease. This has little effect on the peak voltage charge of each capacitor, as there is negligible resistance in the charging current path from source to capacitor, resulting in a very short time constant (τ). However, when it comes time to discharge through the resistors, the capacitor with the greater capacitance value will hold its charge longer, resulting in a greater average DC voltage over time than the lesser-value capacitor.

The load resistor (R_{load}), connected at one end to the point between the two equal-value resistors (R) and at the other end to ground, will drop no DC voltage if the two capacitors' DC voltage charges

are equal in magnitude. If, on the other hand, one capacitor maintains a greater DC voltage charge than the other due to a difference in capacitance, the load resistor will drop a voltage proportional to the difference between these voltages. Thus, differential capacitance is translated into a DC voltage across the load resistor.

Across the load resistor, there is both AC and DC voltage present, with only the DC voltage being significant to the difference in capacitance. If desired, a low-pass filter may be added to the output of this circuit to block the AC, leaving only a DC signal to be interpreted by measurement circuitry:



As a measurement circuit for differential capacitive sensors, the twin-T configuration enjoys many advantages over the standard bridge configuration. First and foremost, transducer displacement is indicated by a simple DC voltage, not an AC voltage whose magnitude *and* phase must be interpreted to tell which capacitance is greater. Furthermore, given the proper component values and power supply output, this DC output signal may be strong enough to directly drive an electromechanical meter movement, eliminating the need for an amplifier circuit. Another important advantage is that all important circuit elements have one terminal directly connected to ground: the source, the load resistor, and both capacitors are all ground-referenced. This helps minimize the ill effects of stray capacitance commonly plaguing bridge measurement circuits, likewise eliminating the need for compensatory measures such as the Wagner earth.

This circuit is also easy to specify parts for. Normally, a measurement circuit incorporating complementary diodes requires the selection of "matched" diodes for good accuracy. Not so with this circuit! So long as the power supply voltage is significantly greater than the deviation in voltage drop between the two diodes, the effects of mismatch are minimal and contribute little to measurement error. Furthermore, supply frequency variations have a relatively low impact on gain (how much output voltage is developed for a given amount of transducer displacement), and square-wave supply voltage works as well as sine-wave, assuming a 50% duty cycle (equal positive and negative half-cycles), of course.

Personal experience with using this circuit has confirmed its impressive performance. Not only is it easy to prototype and test, but its relative insensitivity to stray capacitance and its high output voltage as compared to traditional bridge circuits makes it a very robust alternative.

12.7 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information.

Jason Starck (June 2000): HTML document formatting, which led to a much better-looking second edition.

Chapter 13

AC MOTORS

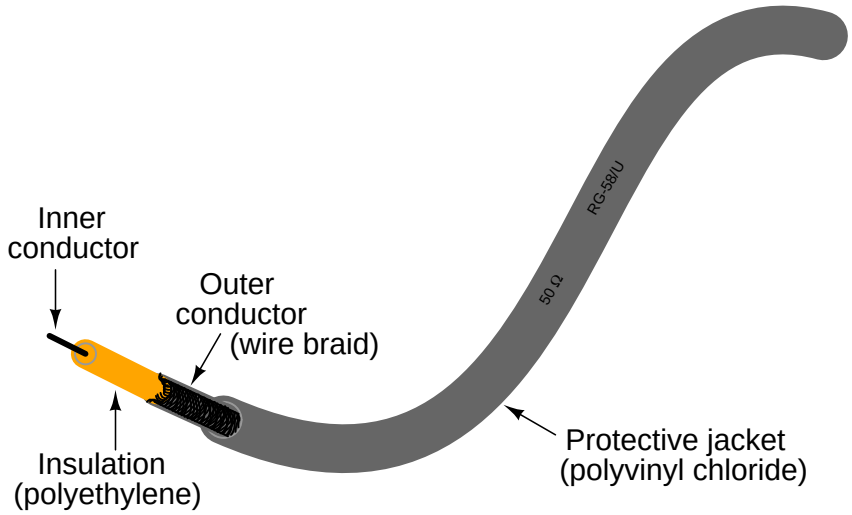
***** PENDING *****

Chapter 14

TRANSMISSION LINES

14.1 A 50-ohm cable?

Early in my explorations of electricity, I came across a length of *coaxial cable* with the label "50 ohms" printed along its outer sheath. Now, coaxial cable is a two-conductor cable made of a single conductor surrounded by a braided wire jacket, with a plastic insulating material separating the two. As such, the outer (braided) conductor completely surrounds the inner (single wire) conductor, the two conductors insulated from each other for the entire length of the cable. This type of cabling is often used to conduct weak (low-amplitude) voltage signals, due to its excellent ability to shield such signals from external interference.



I was mystified by the "50 ohms" label on this coaxial cable. How could two conductors, insulated from each other by a relatively thick layer of plastic, have 50 ohms of resistance between them? Measuring resistance between the outer and inner conductors with my ohmmeter, I found it to be infinite (open-circuit), just as I would have expected from two insulated conductors. Measuring each of the two conductors' resistances from one end of the cable to the other indicated nearly zero

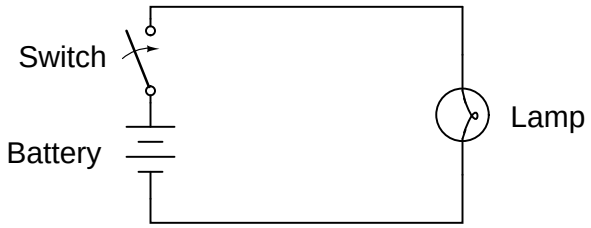
ohms of resistance: again, exactly what I would have expected from continuous, unbroken lengths of wire. Nowhere was I able to measure 50 Ω of resistance on this cable, regardless of which points I connected my ohmmeter between.

What I didn't understand at the time was the cable's response to short-duration voltage "pulses" and high-frequency AC signals. Continuous direct current (DC) – such as that used by my ohmmeter to check the cable's resistance – shows the two conductors to be completely insulated from each other, with nearly infinite resistance between the two. However, due to the effects of capacitance and inductance distributed along the length of the cable, the cable's response to rapidly-changing voltages is such that it acts as a *finite* impedance, drawing current proportional to an applied voltage. What we would normally dismiss as being just a pair of wires becomes an important circuit element in the presence of transient and high-frequency AC signals, with characteristic properties all its own. When expressing such properties, we refer to the wire pair as a *transmission line*.

This chapter explores transmission line behavior. Many transmission line effects do not appear in significant measure in AC circuits of powerline frequency (50 or 60 Hz), or in continuous DC circuits, and so we haven't had to concern ourselves with them in our study of electric circuits thus far. However, in circuits involving high frequencies and/or extremely long cable lengths, the effects are very significant. Practical applications of transmission line effects abound in radio-frequency ("RF") communication circuitry, including computer networks, and in low-frequency circuits subject to voltage transients ("surges") such as lightning strikes on power lines.

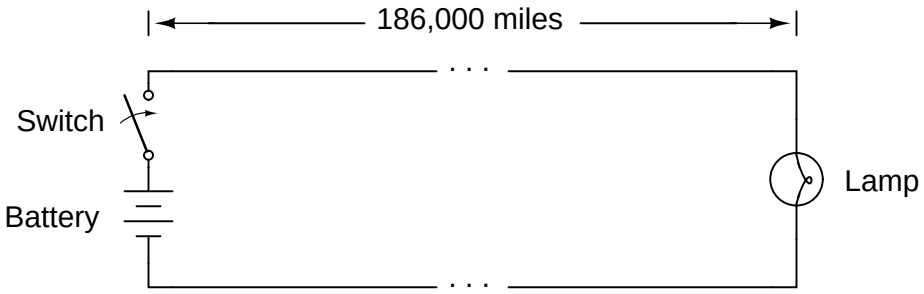
14.2 Circuits and the speed of light

Suppose we had a simple one-battery, one-lamp circuit controlled by a switch. When the switch is closed, the lamp immediately lights. When the switch is opened, the lamp immediately darkens:



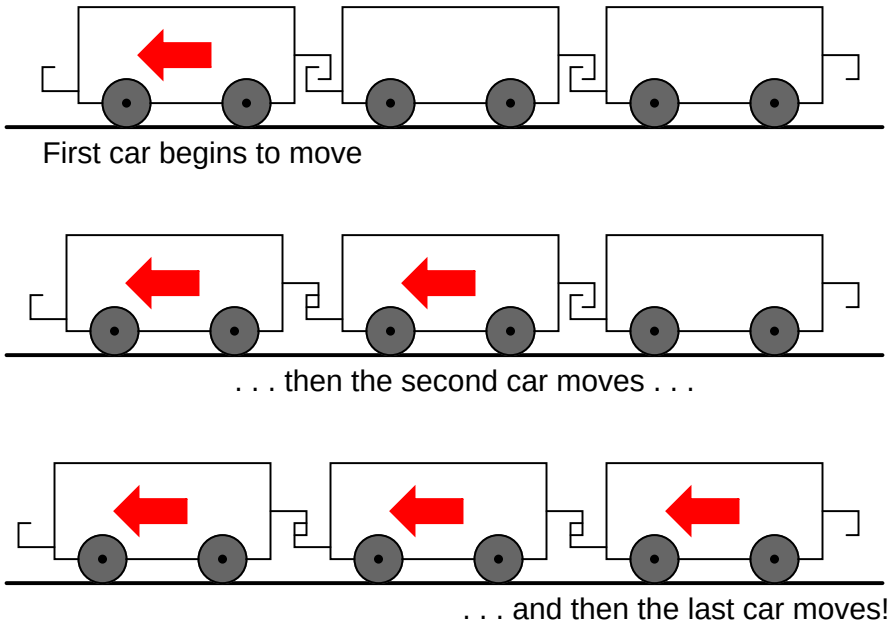
Actually, an incandescent lamp takes a short time for its filament to warm up and emit light after receiving an electric current of sufficient magnitude to power it, so the effect is not instant. However, what I'd like to focus on is the immediacy of the electric current itself, not the response time of the lamp filament. For all practical purposes, the effect of switch action is instant at the lamp's location. Although electrons move through wires very slowly, the overall effect of electrons pushing against each other happens at the speed of light (approximately 186,000 miles per *second*!).

What would happen, though, if the wires carrying power to the lamp were 186,000 miles long? Since we know the effects of electricity do have a finite speed (albeit very fast), a set of very long wires should introduce a time delay into the circuit, delaying the switch's action on the lamp:



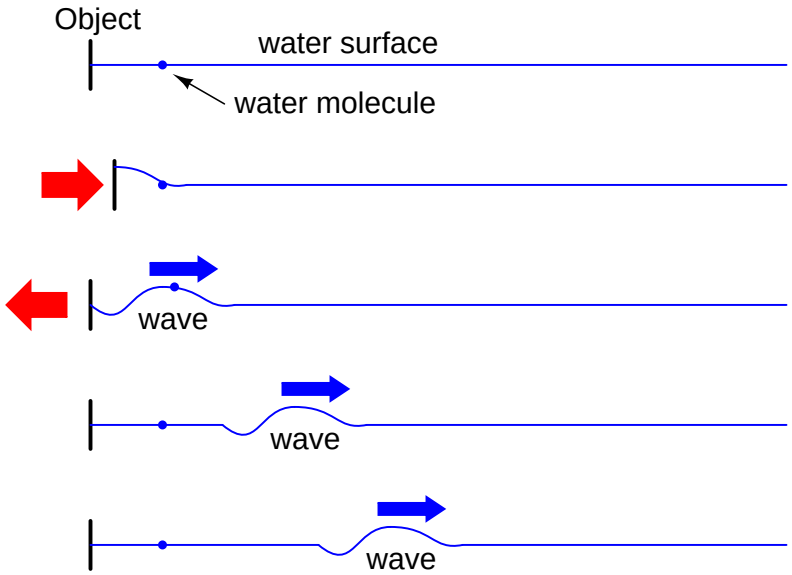
Assuming no warm-up time for the lamp filament, and no resistance along the 372,000 mile length of both wires, the lamp would light up approximately one second after the switch closure. Although the construction and operation of superconducting wires 372,000 miles in length would pose enormous practical problems, it is theoretically possible, and so this "thought experiment" is valid. When the switch is opened again, the lamp will continue to receive power for one second of time after the switch opens, then it will de-energize.

One way of envisioning this is to imagine the electrons within a conductor as rail cars in a train: linked together with a small amount of "slack" or "play" in the couplings. When one rail car (electron) begins to move, it pushes on the one ahead of it and pulls on the one behind it, but not before the slack is relieved from the couplings. Thus, motion is transferred from car to car (from electron to electron) at a maximum velocity limited by the coupling slack, resulting in a much faster transfer of motion from the left end of the train (circuit) to the right end than the actual speed of the cars (electrons):



Another analogy, perhaps more fitting for the subject of transmission lines, is that of waves in water. Suppose a flat, wall-shaped object is suddenly moved horizontally along the surface of water, so as to produce a wave ahead of it. The wave will travel as water molecules bump into each other,

transferring wave motion along the water's surface far faster than the water molecules themselves are actually traveling:

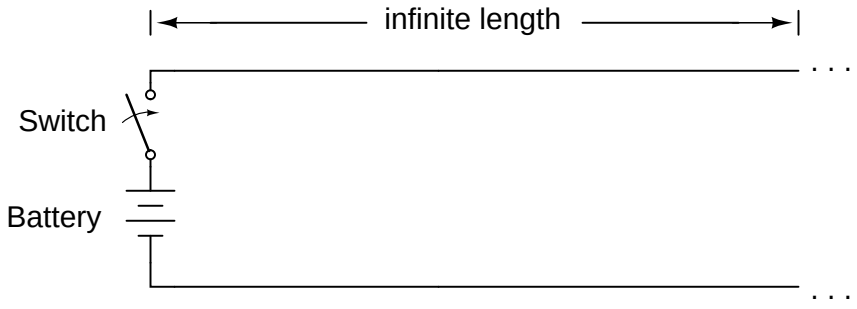


Likewise, electron motion "coupling" travels approximately at the speed of light, although the electrons themselves don't move that quickly. In a very long circuit, this "coupling" speed would become noticeable to a human observer in the form of a short time delay between switch action and lamp action.

- **REVIEW:**
- In an electric circuit, the effects of electron motion travel approximately at the speed of light, although electrons within the conductors do not travel anywhere near that velocity.

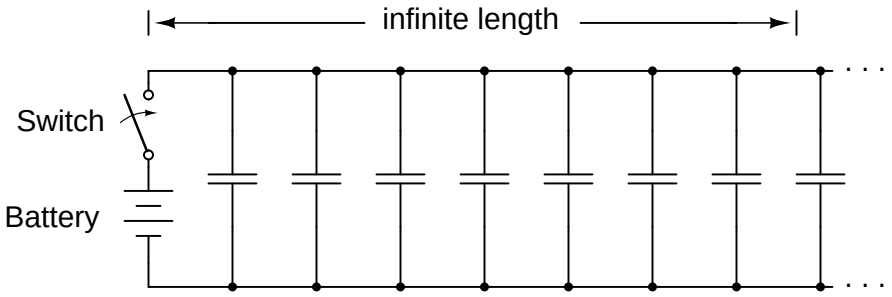
14.3 Characteristic impedance

Suppose, though, that we had a set of parallel wires of *infinite* length, with no lamp at the end. What would happen when we close the switch? Being that there is no longer a load at the end of the wires, this circuit is open. Would there be no current at all?



Despite being able to avoid wire resistance through the use of superconductors in this "thought experiment," we cannot eliminate capacitance along the wires' lengths. *Any* pair of conductors separated by an insulating medium creates capacitance between those conductors:

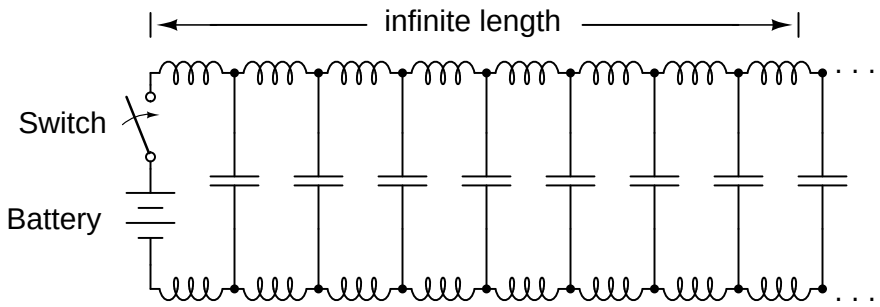
Equivalent circuit showing stray capacitance between conductors

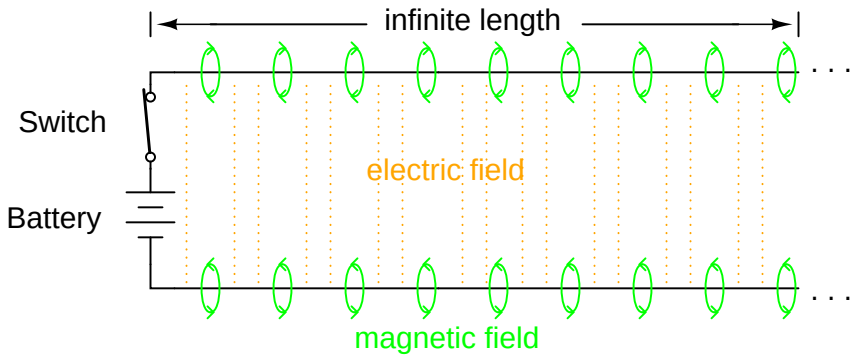


Voltage applied between two conductors creates an electric field between those conductors. Energy is stored in this electric field, and this storage of energy results in an opposition to change in voltage. The reaction of a capacitance against changes in voltage is described by the equation i = C(de/dt), which tells us that current will be drawn proportional to the voltage's rate of change over time. Thus, when the switch is closed, the capacitance between conductors will react against the sudden voltage increase by charging up and drawing current from the source. According to the equation, an instant rise in applied voltage (as produced by perfect switch closure) gives rise to an infinite charging current.

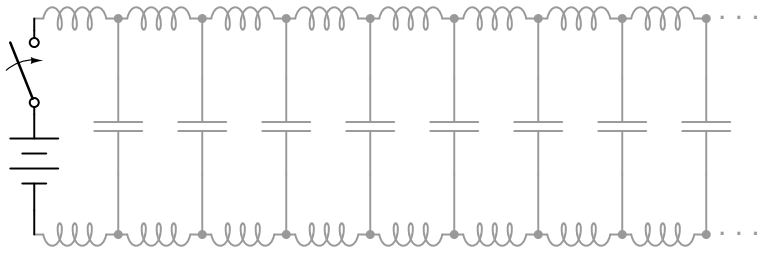
However, the current drawn by a pair of parallel wires will not be infinite, because there exists series impedance along the wires due to inductance. Remember that current through *any* conductor develops a magnetic field of proportional magnitude. Energy is stored in this magnetic field, and this storage of energy results in an opposition to change in current. Each wire develops a magnetic field as it carries charging current for the capacitance between the wires, and in so doing drops voltage according to the inductance equation e = L(di/dt). This voltage drop limits the voltage rate-of-change across the distributed capacitance, preventing the current from ever reaching an infinite magnitude:

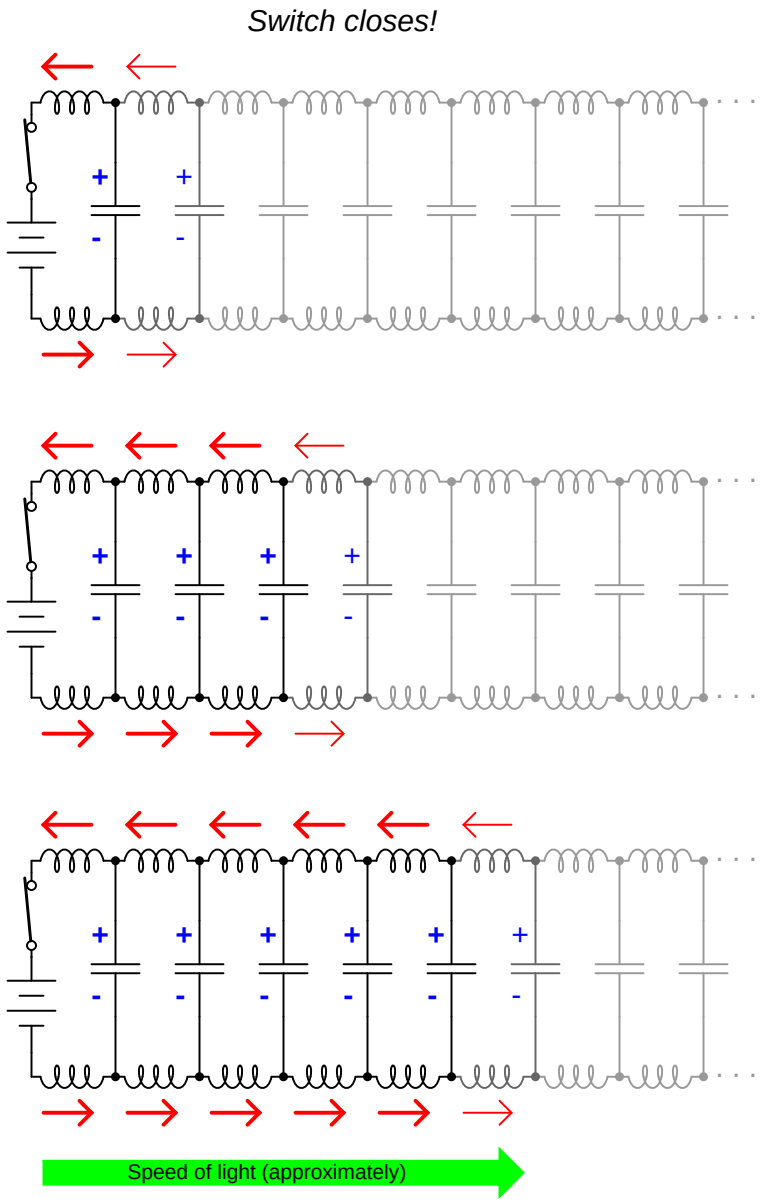
Equivalent circuit showing stray inductance and capacitance





Because the electrons in the two wires transfer motion to and from each other at nearly the speed of light, the "wave front" of voltage and current change will propagate down the length of the wires at that same velocity, resulting in the distributed capacitance and inductance progressively charging to full voltage and current, respectively, like this:





The end result of these interactions is a constant current of limited magnitude through the battery source. Since the wires are infinitely long, their distributed capacitance will never fully charge to the source voltage, and their distributed inductance will never allow unlimited charging current. In other words, this pair of wires will draw current from the source so long as the switch is closed, behaving as a constant load. No longer are the wires merely conductors of electrical current and carriers of voltage, but now constitute a circuit component in themselves, with unique characteristics. No longer are the two wires merely *a pair of conductors*, but rather a *transmission line*.

As a constant load, the transmission line's response to applied voltage is resistive rather than reactive, despite being comprised purely of inductance and capacitance (assuming superconducting wires with zero resistance). We can say this because there is no difference from the battery's perspective between a resistor eternally dissipating energy and an infinite transmission line eternally absorbing energy. The impedance (resistance) of this line in ohms is called the *characteristic impedance*, and it is fixed by the geometry of the two conductors. For a parallel-wire line with air insulation, the characteristic impedance may be calculated as such:

$$Z_0 = \frac{276}{\sqrt{k}} \log \frac{d}{r}$$

Where,

Z_0 = Characteristic impedance of line
d = Distance between conductor centers
r = Conductor radius
k = Relative permittivity of insulation between conductors

If the transmission line is coaxial in construction, the characteristic impedance follows a different equation:

$$Z_0 = \frac{138}{\sqrt{k}} \log \frac{d_1}{d_2}$$

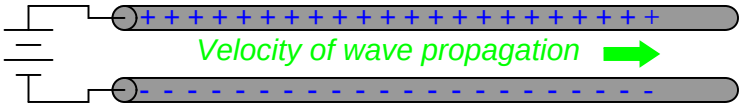
Where,

Z_0 = Characteristic impedance of line
d_1 = Inside diameter of outer conductor
d_2 = Outside diameter of inner conductor
k = Relative permittivity of insulation between conductors

In both equations, identical units of measurement must be used in both terms of the fraction. If the insulating material is other than air (or a vacuum), both the characteristic impedance and the propagation velocity will be affected. The ratio of a transmission line's true propagation velocity

and the speed of light in a vacuum is called the *velocity factor* of that line.

Velocity factor is purely a factor of the insulating material's relative permittivity (otherwise known as its *dielectric constant*), defined as the ratio of a material's electric field permittivity to that of a pure vacuum. The velocity factor of any cable type – coaxial or otherwise – may be calculated quite simply by the following formula:



$$\text{Velocity factor} = \frac{v}{c} = \frac{1}{\sqrt{k}}$$

Where,

v = Velocity of wave propagation

c = Velocity of light in a vacuum

k = Relative permittivity of insulation between conductors

Characteristic impedance is also known as *natural impedance*, and it refers to the equivalent resistance of a transmission line if it were infinitely long, owing to distributed capacitance and inductance as the voltage and current "waves" propagate along its length at a propagation velocity equal to some large fraction of light speed.

It can be seen in either of the first two equations that a transmission line's characteristic impedance (Z_0) increases as the conductor spacing increases. If the conductors are moved away from each other, the distributed capacitance will decrease (greater spacing between capacitor "plates"), and the distributed inductance will increase (less cancellation of the two opposing magnetic fields). Less parallel capacitance and more series inductance results in a smaller current drawn by the line for any given amount of applied voltage, which by definition is a greater impedance. Conversely, bringing the two conductors closer together increases the parallel capacitance and decreases the series inductance. Both changes result in a larger current drawn for a given applied voltage, equating to a lesser impedance.

Barring any dissipative effects such as dielectric "leakage" and conductor resistance, the characteristic impedance of a transmission line is equal to the square root of the ratio of the line's inductance per unit length divided by the line's capacitance per unit length:

$$Z_0 = \sqrt{\frac{L}{C}}$$

Where,

Z_0 = Characteristic impedance of line

L = Inductance per unit length of line

C = Capacitance per unit length of line

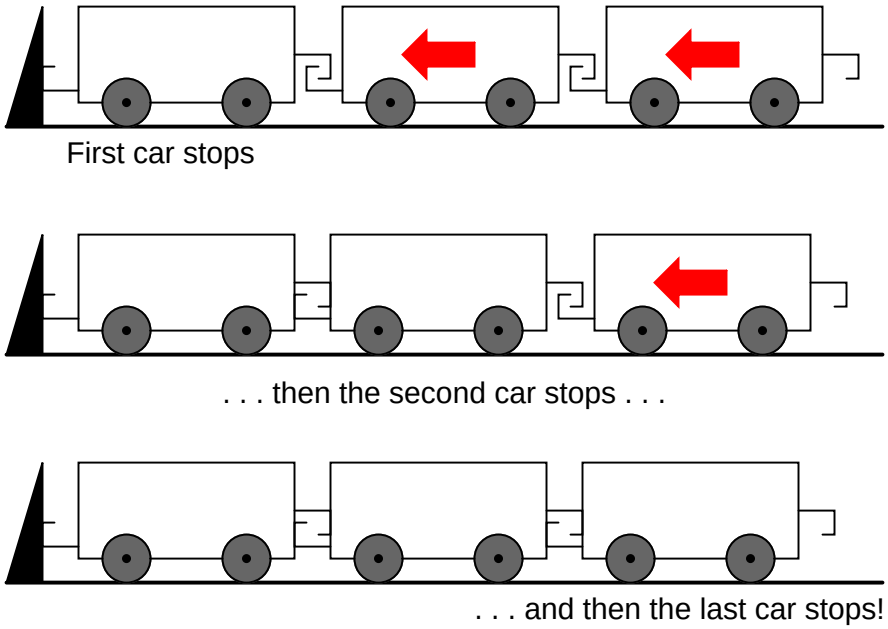
- **REVIEW:**
- A *transmission line* is a pair of parallel conductors exhibiting certain characteristics due to distributed capacitance and inductance along its length.
- When a voltage is suddenly applied to one end of a transmission line, both a voltage "wave" and a current "wave" propagate along the line at nearly light speed.
- If a DC voltage is applied to one end of an infinitely long transmission line, the line will draw current from the DC source as though it were a constant resistance.
- The *characteristic impedance* (Z_0) of a transmission line is the resistance it would exhibit if it were infinite in length. This is entirely different from leakage resistance of the dielectric separating the two conductors, and the metallic resistance of the wires themselves. Characteristic impedance is purely a function of the capacitance and inductance distributed along the line's length, and would exist even if the dielectric were perfect (infinite parallel resistance) and the wires superconducting (zero series resistance).
- *Velocity factor* is a fractional value relating a transmission line's propagation speed to the speed of light in a vacuum. Values range between 0.66 and 0.80 for typical two-wire lines and coaxial cables. For any cable type, it is equal to the reciprocal (1/x) of the square root of the relative permittivity of the cable's insulation.

14.4 Finite-length transmission lines

A transmission line of infinite length is an interesting abstraction, but physically impossible. All transmission lines have some finite length, and as such do not behave precisely the same as an infinite line. If that piece of 50 Ω "RG-58/U" cable I measured with an ohmmeter years ago had been infinitely long, I actually would have been able to measure 50 Ω worth of resistance between the inner and outer conductors. But it was not infinite in length, and so it measured as "open" (infinite resistance).

Nonetheless, the characteristic impedance rating of a transmission line is important even when dealing with limited lengths. An older term for characteristic impedance, which I like for its descriptive value, is *surge impedance*. If a transient voltage (a "surge") is applied to the end of a transmission line, the line will draw a current proportional to the surge voltage magnitude divided by the line's surge impedance (I=E/Z). This simple, Ohm's Law relationship between current and voltage will hold true for a limited period of time, but not indefinitely.

If the end of a transmission line is open-circuited – that is, left unconnected – the current "wave" propagating down the line's length will have to stop at the end, since electrons cannot flow where there is no continuing path. This abrupt cessation of current at the line's end causes a "pile-up" to occur along the length of the transmission line, as the electrons successively find no place to go. Imagine a train traveling down the track with slack between the rail car couplings: if the lead car suddenly crashes into an immovable barricade, it will come to a stop, causing the one behind it to come to a stop as soon as the first coupling slack is taken up, which causes the next rail car to stop as soon as the next coupling's slack is taken up, and so on until the last rail car stops. The train does not come to a halt together, but rather in sequence from first car to last:



A signal propagating from the source-end of a transmission line to the load-end is called an *incident wave*. The propagation of a signal from load-end to source-end (such as what happened in this example with current encountering the end of an open-circuited transmission line) is called a *reflected wave*.

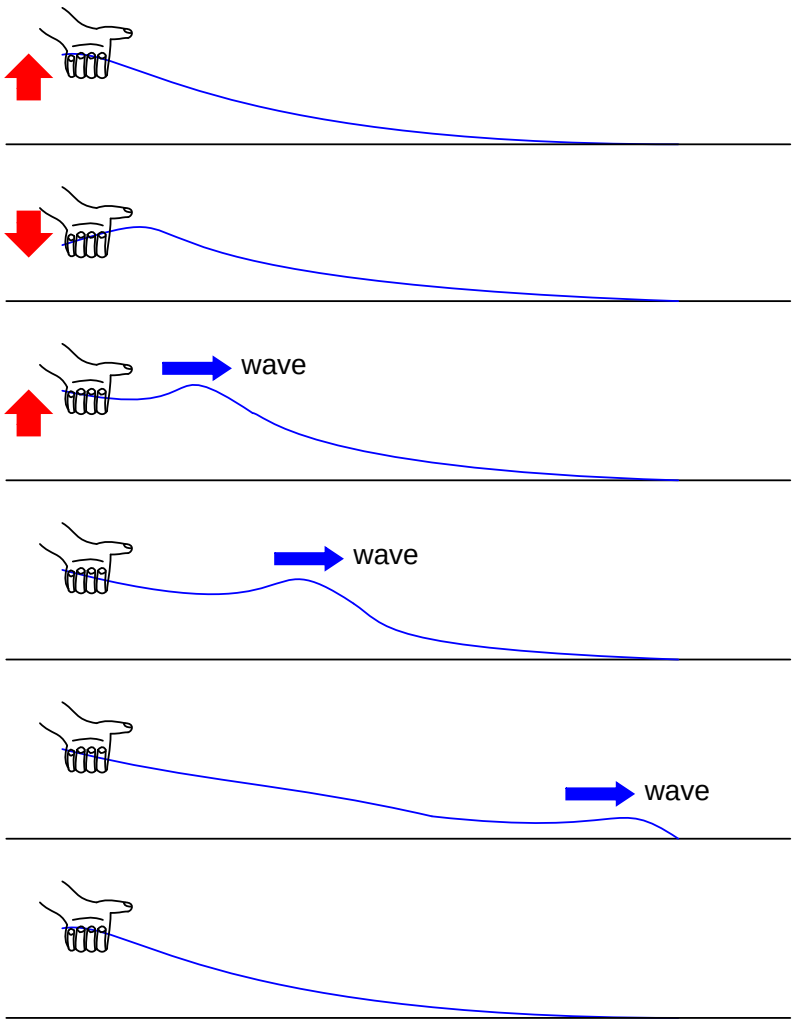
When this electron "pile-up" propagates back to the battery, current at the battery ceases, and the line acts as a simple open circuit. All this happens very quickly for transmission lines of reasonable length, and so an ohmmeter measurement of the line never reveals the brief time period where the line actually behaves as a resistor. For a mile-long cable with a velocity factor of 0.66 (signal propagation velocity is 66% of light speed, or 122,760 miles per second), it takes only 1/122,760 of a second (8.146 microseconds) for a signal to travel from one end to the other. For the current signal to reach the line's end and "reflect" back to the source, the round-trip time is twice this figure, or 16.292 μs.

High-speed measurement instruments are able to detect this transit time from source to line-end and back to source again, and may be used for the purpose of determining a cable's length. This technique may also be used for determining the presence *and* location of a break in one or both of the cable's conductors, since a current will "reflect" off the wire break just as it will off the end of an open-circuited cable. Instruments designed for such purposes are called *time-domain reflectometers* (TDRs). The basic principle is identical to that of sonar range-finding: generating a sound pulse and measuring the time it takes for the echo to return.

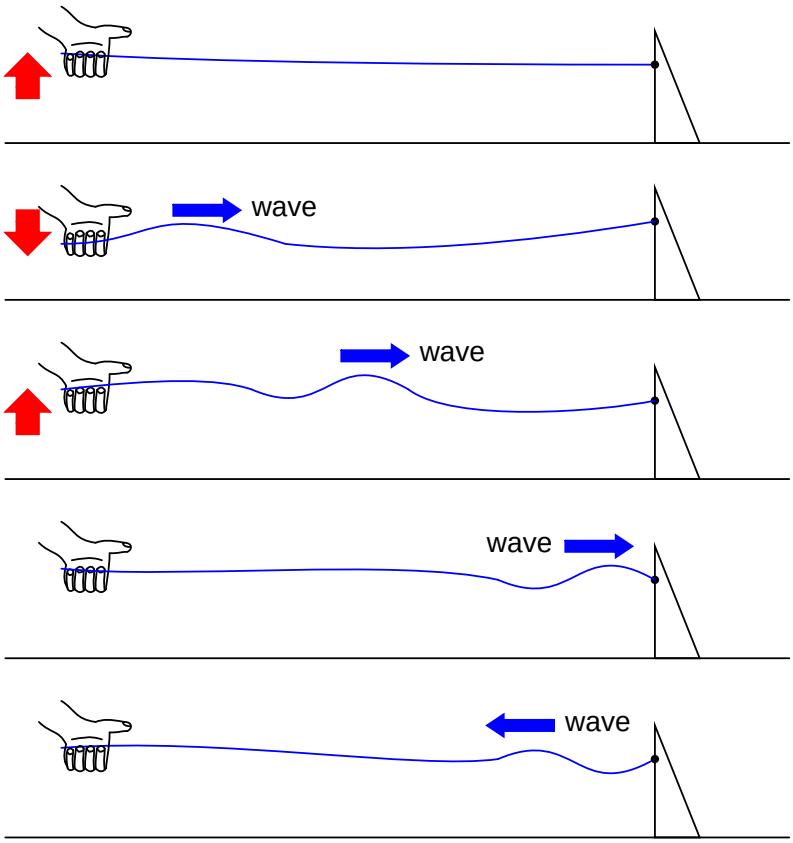
A similar phenomenon takes place if the end of a transmission line is short-circuited: when the voltage wave-front reaches the end of the line, it is reflected back to the source, because voltage cannot exist between two electrically common points. When this reflected wave reaches the source, the source sees the entire transmission line as a short-circuit. Again, this happens as quickly as the signal can propagate round-trip down and up the transmission line at whatever velocity allowed by the dielectric material between the line's conductors.

A simple experiment illustrates the phenomenon of wave reflection in transmission lines. Take

a length of rope by one end and "whip" it with a rapid up-and-down motion of the wrist. A wave may be seen traveling down the rope's length until it dissipates entirely due to friction:



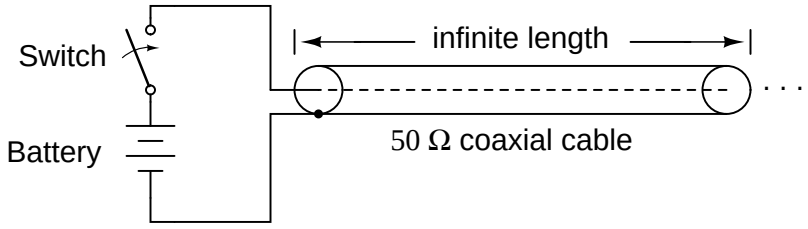
This is analogous to a long transmission line with internal loss: the signal steadily grows weaker as it propagates down the line's length, never reflecting back to the source. However, if the far end of the rope is secured to a solid object at a point prior to the incident wave's total dissipation, a second wave will be reflected back to your hand:



Usually, the purpose of a transmission line is to convey electrical energy from one point to another. Even if the signals are intended for information only, and not to power some significant load device, the ideal situation would be for all of the original signal energy to travel from the source to the load, and then be completely absorbed or dissipated by the load for maximum signal-to-noise ratio. Thus, "loss" along the length of a transmission line is undesirable, as are reflected waves, since reflected energy is energy not delivered to the end device.

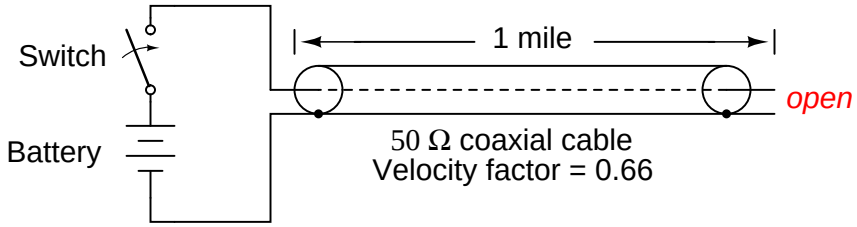
Reflections may be eliminated from the transmission line if the load's impedance exactly equals the characteristic ("surge") impedance of the line. For example, a 50 Ω coaxial cable that is either open-circuited or short-circuited will reflect all of the incident energy back to the source. However, if a 50 Ω resistor is connected at the end of the cable, there will be no reflected energy, all signal energy being dissipated by the resistor.

This makes perfect sense if we return to our hypothetical, infinite-length transmission line example. A transmission line of 50 Ω characteristic impedance and infinite length behaves exactly like a 50 Ω resistance as measured from one end. If we cut this line to some finite length, it will behave as a 50 Ω resistor to a constant source of DC voltage for a brief time, but then behave like an open- or a short-circuit, depending on what condition we leave the cut end of the line: open or shorted. However, if we *terminate* the line with a 50 Ω resistor, the line will once again behave as a 50 Ω resistor, indefinitely: the same as if it were of infinite length again:



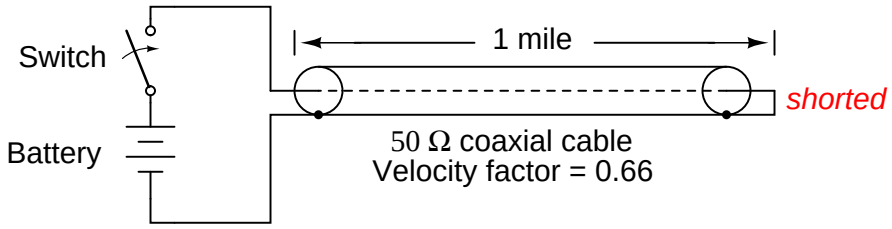
Cable's behavior from perspective of battery:

Exactly like a 50 Ω resistor



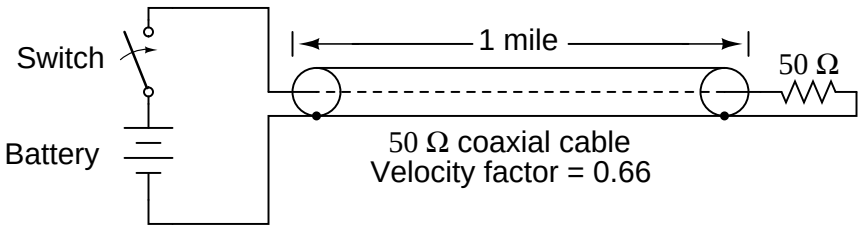
Cable's behavior from perspective of battery:

Like a 50 Ω resistor for 16.292 μs,
then like an open (infinite resistance)



Cable's behavior from perspective of battery:

Like a 50 Ω resistor for 16.292 μs,
then like a short (zero resistance)



Cable's behavior from perspective of battery:

Exactly like a 50 Ω resistor

In essence, a terminating resistor matching the natural impedance of the transmission line makes the line "appear" infinitely long from the perspective of the source, because a resistor has the ability to eternally dissipate energy in the same way a transmission line of infinite length is able to eternally absorb energy.

Reflected waves will also manifest if the terminating resistance isn't precisely equal to the characteristic impedance of the transmission line, not just if the line is left unconnected (open) or jumpered (shorted). Though the energy reflection will not be total with a terminating impedance of slight mismatch, it will be partial. This happens whether or not the terminating resistance is *greater* or *less* than the line's characteristic impedance.

Re-reflections of a reflected wave may also occur at the *source end* of a transmission line, if the source's internal impedance (Thevenin equivalent impedance) is not exactly equal to the line's characteristic impedance. A reflected wave returning back to the source will be dissipated entirely if the source impedance matches the line's, but will be reflected back toward the line end like another incident wave, at least partially, if the source impedance does not match the line. This type of reflection may be particularly troublesome, as it makes it appear that the source has transmitted another pulse.

- **REVIEW:**

- Characteristic impedance is also known as *surge impedance*, due to the temporarily resistive behavior of any length transmission line.

- A finite-length transmission line will appear to a DC voltage source as a constant resistance for some short time, then as whatever impedance the line is terminated with. Therefore, an open-ended cable simply reads "open" when measured with an ohmmeter, and "shorted" when its end is short-circuited.

- A transient ("surge") signal applied to one end of an open-ended or short-circuited transmission line will "reflect" off the far end of the line as a secondary wave. A signal traveling on a transmission line from source to load is called an *incident wave*; a signal "bounced" off the end of a transmission line, traveling from load to source, is called a *reflected wave*.

- Reflected waves will also appear in transmission lines terminated by resistors not precisely matching the characteristic impedance.

- A finite-length transmission line may be made to appear infinite in length if terminated by a resistor of equal value to the line's characteristic impedance. This eliminates all signal reflections.

- A reflected wave may become re-reflected off the source-end of a transmission line if the source's internal impedance does not match the line's characteristic impedance. This re-reflected wave will appear, of course, like another pulse signal transmitted from the source.

14.5 "Long" and "short" transmission lines

In DC and low-frequency AC circuits, the characteristic impedance of parallel wires is usually ignored. This includes the use of coaxial cables in instrument circuits, often employed to protect weak voltage signals from being corrupted by induced "noise" caused by stray electric and magnetic fields. This is due to the relatively short timespans in which reflections take place in the line, as compared to the period of the waveforms or pulses of the significant signals in the circuit. As we saw in the last section, if a transmission line is connected to a DC voltage source, it will behave as a resistor equal in value to the line's characteristic impedance only for as long as it takes the incident pulse to reach the end of the line and return as a reflected pulse, back to the source. After that time (a brief 16.292 μs for the mile-long coaxial cable of the last example), the source "sees" only the terminating impedance, whatever that may be.

If the circuit in question handles low-frequency AC power, such short time delays introduced by a transmission line between when the AC source outputs a voltage peak and when the source "sees" that peak loaded by the terminating impedance (round-trip time for the incident wave to reach the line's end and reflect back to the source) are of little consequence. Even though we know that signal magnitudes along the line's length are not equal at any given time due to signal propagation at (nearly) the speed of light, the actual phase difference between start-of-line and end-of-line signals is negligible, because line-length propagations occur within a very small fraction of the AC waveform's period. For all practical purposes, we can say that voltage along all respective points on a low-frequency, two-conductor line are equal and in-phase with each other at any given point in time.

In these cases, we can say that the transmission lines in question are *electrically short*, because their propagation effects are much quicker than the periods of the conducted signals. By contrast, an *electrically long* line is one where the propagation time is a large fraction or even a multiple of the signal period. A "long" line is generally considered to be one where the source's signal waveform completes at least a quarter-cycle (90^o of "rotation") before the incident signal reaches line's end. Up until this chapter in the *Lessons In Electric Circuits* book series, all connecting lines were assumed to be electrically short.

To put this into perspective, we need to express the distance traveled by a voltage or current signal along a transmission line in relation to its source frequency. An AC waveform with a frequency of 60 Hz completes one cycle in 16.66 ms. At light speed (186,000 m/s), this equates to a distance of 3100 miles that a voltage or current signal will propagate in that time. If the velocity factor of the transmission line is less than 1, the propagation velocity will be less than 186,000 miles per second, and the distance less by the same factor. But even if we used the coaxial cable's velocity factor from the last example (0.66), the distance is still a very long 2046 miles! Whatever distance we calculate for a given frequency is called the *wavelength* of the signal.

A simple formula for calculating wavelength is as follows:

$$\lambda = \frac{v}{f}$$

Where,

λ = Wavelength

v = Velocity of propagation

f = Frequency of signal

The lower-case Greek letter "lambda" (λ) represents wavelength, in whatever unit of length used in the velocity figure (if miles per second, then wavelength in miles; if meters per second, then wavelength in meters). Velocity of propagation is usually the speed of light when calculating signal wavelength in open air or in a vacuum, but will be less if the transmission line has a velocity factor less than 1.

If a "long" line is considered to be one at least 1/4 wavelength in length, you can see why all connecting lines in the circuits discussed thusfar have been assumed "short." For a 60 Hz AC power system, power lines would have to exceed 775 miles in length before the effects of propagation time became significant. Cables connecting an audio amplifier to speakers would have to be over 4.65 miles in length before line reflections would significantly impact a 10 kHz audio signal!

When dealing with radio-frequency systems, though, transmission line length is far from trivial. Consider a 100 MHz radio signal: its wavelength is a mere 9.8202 feet, even at the full propagation velocity of light (186,000 m/s). A transmission line carrying this signal would not have to be more than about 2-1/2 feet in length to be considered "long!" With a cable velocity factor of 0.66, this critical length shrinks to 1.62 feet.

When an electrical source is connected to a load via a "short" transmission line, the load's impedance dominates the circuit. This is to say, when the line is short, its own characteristic impedance is of little consequence to the circuit's behavior. We see this when testing a coaxial cable with an ohmmeter: the cable reads "open" from center conductor to outer conductor if the cable end is left unterminated. Though the line acts as a resistor for a very brief period of time after the meter is connected (about 50 Ω for an RG-58/U cable), it immediately thereafter behaves as a simple "open circuit:" the impedance of the line's open end. Since the combined response time of an ohmmeter and the human being using it *greatly exceeds* the round-trip propagation time up and down the cable, it is "electrically short" for this application, and we only register the terminating (load) impedance. It is the extreme speed of the propagated signal that makes us unable to detect the cable's 50 Ω transient impedance with an ohmmeter.

If we use a coaxial cable to conduct a DC voltage or current to a load, and no component in the circuit is capable of measuring or responding quickly enough to "notice" a reflected wave, the cable is considered "electrically short" and its impedance is irrelevant to circuit function. Note how the electrical "shortness" of a cable is relative to the application: in a DC circuit where voltage and current values change slowly, nearly any physical length of cable would be considered "short" from the standpoint of characteristic impedance and reflected waves. Taking the same length of cable, though, and using it to conduct a high-frequency AC signal could result in a vastly different assessment of that cable's "shortness!"

When a source is connected to a load via a "long" transmission line, the line's own characteristic impedance dominates over load impedance in determining circuit behavior. In other words,

an electrically "long" line acts as the principal component in the circuit, its own characteristics overshadowing the load's. With a source connected to one end of the cable and a load to the other, current drawn from the source is a function primarily of the line and not the load. This is increasingly true the longer the transmission line is. Consider our hypothetical 50 Ω cable of infinite length, surely the ultimate example of a "long" transmission line: no matter what kind of load we connect to one end of this line, the source (connected to the other end) will only see 50 Ω of impedance, because the line's infinite length prevents the signal from *ever reaching* the end where the load is connected. In this scenario, line impedance exclusively defines circuit behavior, rendering the load completely irrelevant.

The most effective way to minimize the impact of transmission line length on circuit behavior is to match the line's characteristic impedance to the load impedance. If the load impedance is equal to the line impedance, then *any* signal source connected to the other end of the line will "see" the exact same impedance, and will have the exact same amount of current drawn from it, regardless of line length. In this condition of perfect impedance matching, line length only affects the amount of time delay from signal departure at the source to signal arrival at the load. However, perfect matching of line and load impedances is not always practical or possible.

The next section discusses the effects of "long" transmission lines, especially when line length happens to match specific fractions or multiples of signal wavelength.

- **REVIEW:**

- Coaxial cabling is sometimes used in DC and low-frequency AC circuits as well as in high-frequency circuits, for the excellent immunity to induced "noise" that it provides for signals.

- When the period of a transmitted voltage or current signal greatly exceeds the propagation time for a transmission line, the line is considered *electrically short*. Conversely, when the propagation time is a large fraction or multiple of the signal's period, the line is considered *electrically long*.

- A signal's *wavelength* is the physical distance it will propagate in the timespan of one period. Wavelength is calculated by the formula λ=v/f, where "λ" is the wavelength, "v" is the propagation velocity, and "f" is the signal frequency.

- A rule-of-thumb for transmission line "shortness" is that the line must be at least 1/4 wavelength before it is considered "long."

- In a circuit with a "short" line, the terminating (load) impedance dominates circuit behavior. The source effectively sees nothing but the load's impedance, barring any resistive losses in the transmission line.

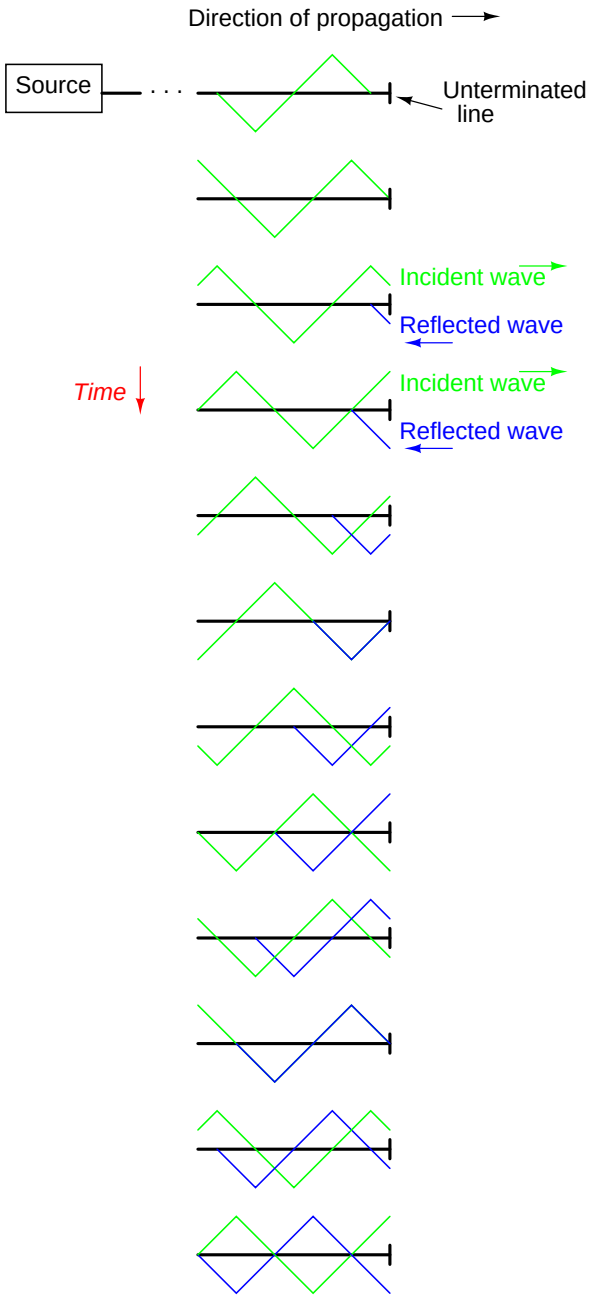
- In a circuit with a "long" line, the line's own characteristic impedance dominates circuit behavior. The ultimate example of this is a transmission line of infinite length: since the signal will *never* reach the load impedance, the source only "sees" the cable's characteristic impedance.

- When a transmission line is terminated by a load precisely matching its impedance, there are no reflected waves and thus no problems with line length.

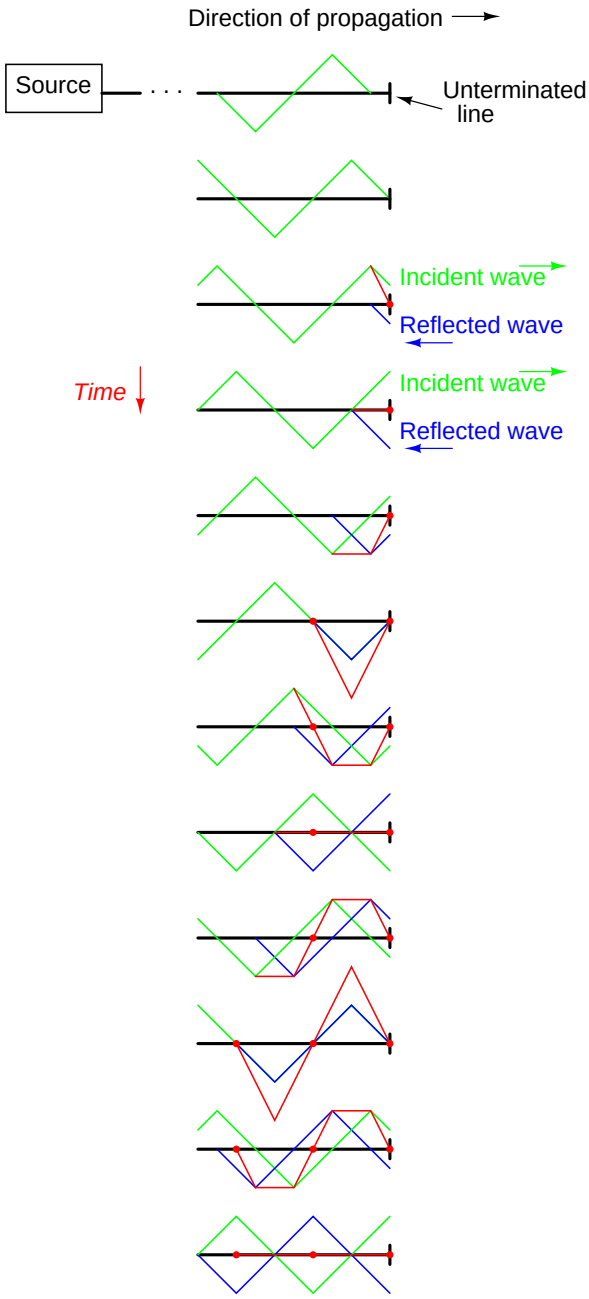
14.6 Standing waves and resonance

Whenever there is a mismatch of impedance between transmission line and load, reflections will occur. If the incident signal is a continuous AC waveform, these reflections will mix with more of the oncoming incident waveform to produce stationary waveforms called *standing waves*.

The following illustration shows how a triangle-shaped incident waveform turns into a mirror-image reflection upon reaching the line's unterminated end. The transmission line in this illustrative sequence is shown as a single, thick line rather than a pair of wires, for simplicity's sake. The incident wave is shown traveling from left to right, while the reflected wave travels from right to left:

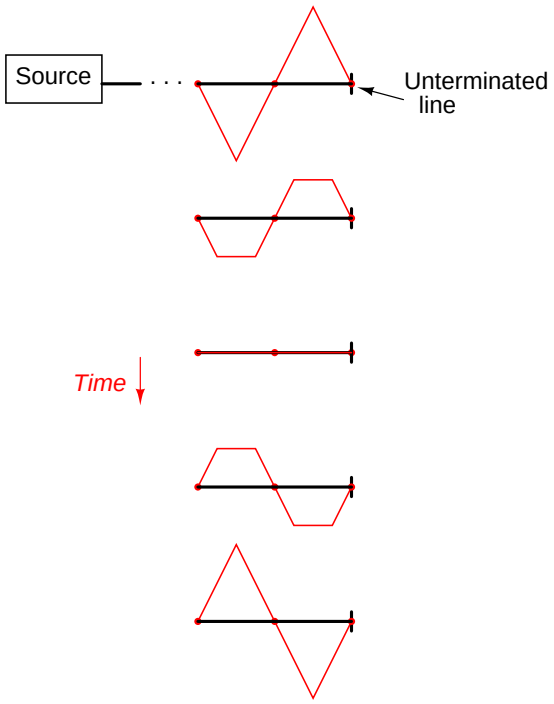


If we add the two waveforms together, we find that a third, stationary waveform is created along the line's length:

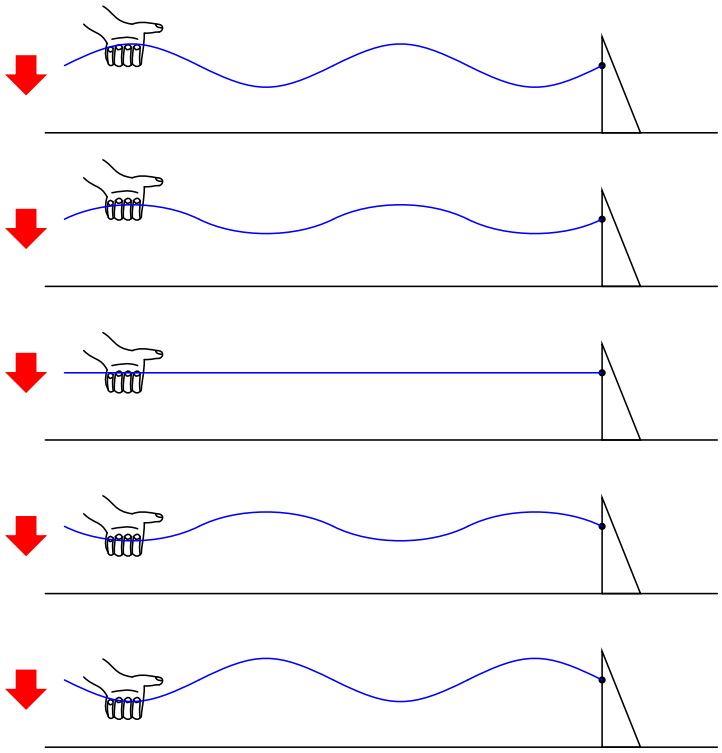


This third, "standing" wave, in fact, represents the only voltage along the line, being the representative sum of incident and reflected voltage waves. It oscillates in instantaneous magnitude, but does not propagate down the cable's length like the incident or reflected waveforms causing it. Note the dots along the line length marking the "zero" points of the standing wave (where the incident

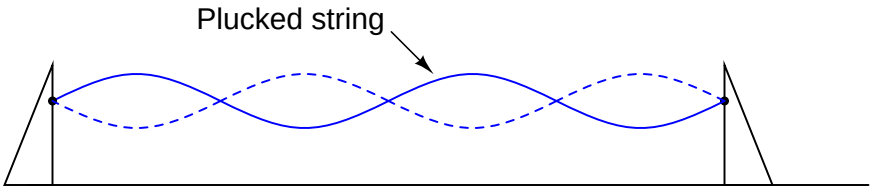
and reflected waves cancel each other), and how those points never change position:



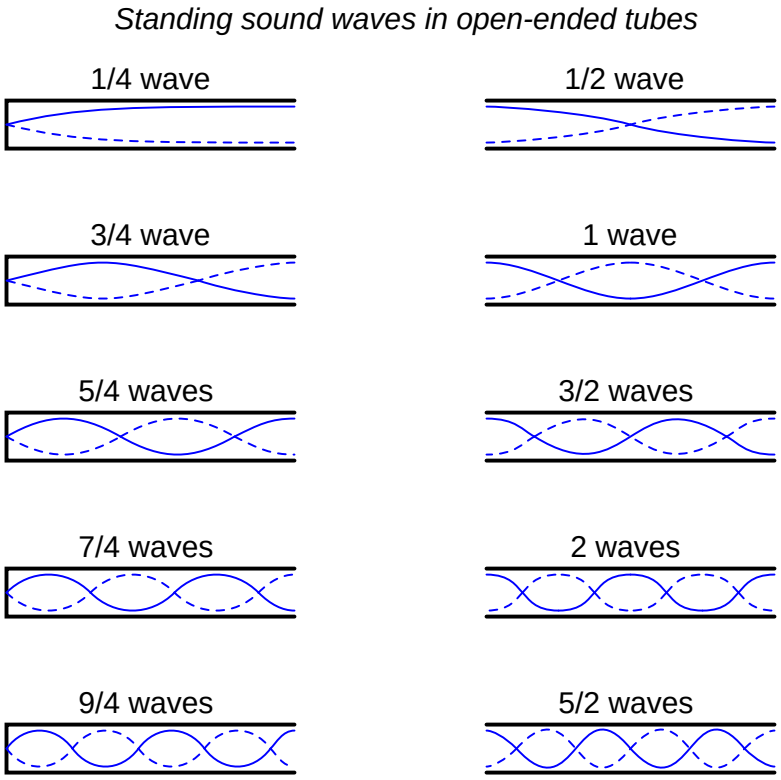
Standing waves are quite abundant in the physical world. Consider a string or rope, shaken at one end, and tied down at the other (only one half-cycle of hand motion shown, moving downward):



Both the nodes (points of little or no vibration) and the antinodes (points of maximum vibration) remain fixed along the length of the string or rope. The effect is most pronounced when the free end is shaken at just the right frequency. Plucked strings exhibit the same "standing wave" behavior, with "nodes" of maximum and minimum vibration along their length. The major difference between a plucked string and a shaken string is that the plucked string supplies its own "correct" frequency of vibration to maximize the standing-wave effect:



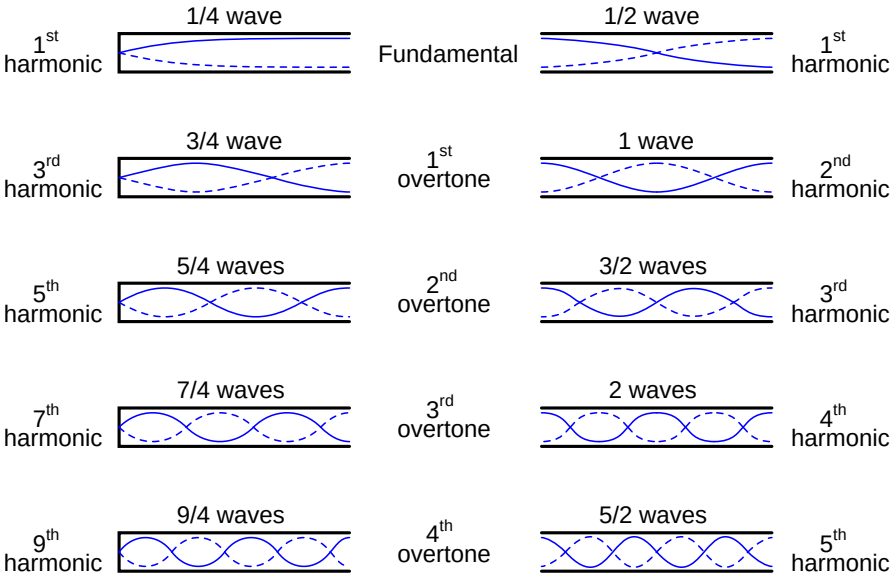
Wind blowing across an open-ended tube also produces standing waves; this time, the waves are vibrations of air molecules (sound) within the tube rather than vibrations of a solid object. Whether the standing wave terminates in a node (minimum amplitude) or an antinode (maximum amplitude) depends on whether the other end of the tube is open or closed:



A closed tube end must be a wave node, while an open tube end must be an antinode. By analogy, the anchored end of a vibrating string must be a node, while the free end (if there is any) must be an antinode.

Note how there is more than one wavelength suitable for producing standing waves of vibrating air within a tube that precisely match the tube's end points. This is true for all standing-wave systems: standing waves will resonate with the system for any frequency (wavelength) correlating to the node/antinode points of the system. Another way of saying this is that there are multiple resonant frequencies for any system supporting standing waves.

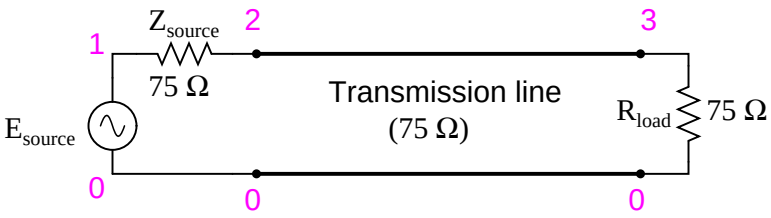
All higher frequencies are integer-multiples of the lowest (fundamental) frequency for the system. The sequential progression of harmonics from one resonant frequency to the next defines the *overtone* frequencies for the system:



The actual frequencies (measured in Hertz) for any of these harmonics or overtones depends on the physical length of the tube and the waves' propagation velocity, which is the speed of sound in air.

Because transmission lines support standing waves, and force these waves to possess nodes and antinodes according to the type of termination impedance at the load end, they also exhibit resonance at frequencies determined by physical length and propagation velocity. Transmission line resonance, though, is a bit more complex than resonance of strings or of air in tubes, because we must consider both voltage waves and current waves.

This complexity is made easier to understand by way of computer simulation. To begin, let's examine a perfectly matched source, transmission line, and load. All components have an impedance of 75 Ω:

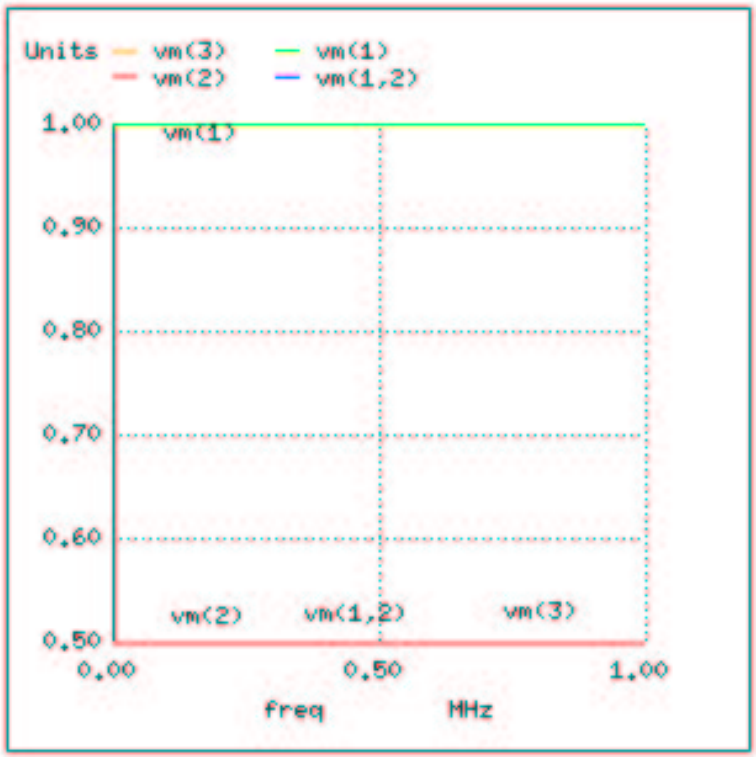


Using SPICE to simulate the circuit, we'll specify the transmission line (`t1`) with a 75 Ω characteristic impedance (`z0=75`) and a propagation delay of 1 microsecond (`td=1u`). This is a convenient method for expressing the physical length of a transmission line: the amount of time it takes a wave to propagate down its entire length. If this were a real 75 Ω cable – perhaps a type "RG-59B/U" coaxial cable, the type commonly used for cable television distribution – with a velocity factor of 0.66, it would be about 648 feet long. Since 1 μs is the period of a 1 MHz signal, I'll choose to sweep the frequency of the AC source from (nearly) zero to that figure, to see how the system reacts when exposed to signals ranging from DC to 1 wavelength.

Here is the SPICE netlist for the circuit shown above:

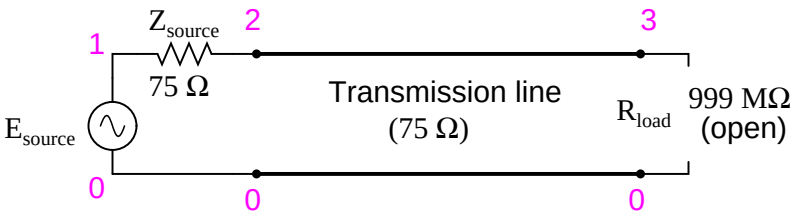
```
Transmission line
v1 1 0 ac 1 sin
rsource 1 2 75
t1 2 0 3 0 z0=75 td=1u
rload 3 0 75
.ac lin 101 1m 1meg
* Using "Nutmeg" program to plot analysis
.end
```

Running this simulation and plotting the source impedance drop (as an indication of current), the source voltage, the line's source-end voltage, and the load voltage, we see that the source voltage – shown as `vm(1)` (voltage magnitude between node 1 and the implied ground point of node 0) on the graphic plot – registers a steady 1 volt, while every other voltage registers a steady 0.5 volts:

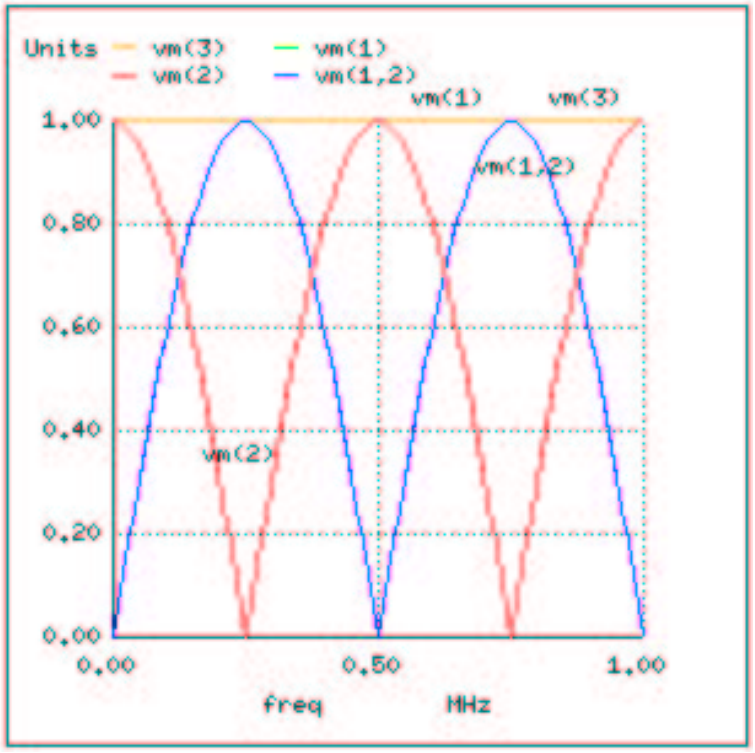


In a system where all impedances are perfectly matched, there can be no standing waves, and therefore no resonant "peaks" or "valleys" in the Bode plot.

Now, let's change the load impedance to 999 MΩ, to simulate an open-ended transmission line. We should definitely see some reflections on the line now as the frequency is swept from 1 mHz to 1 MHz:

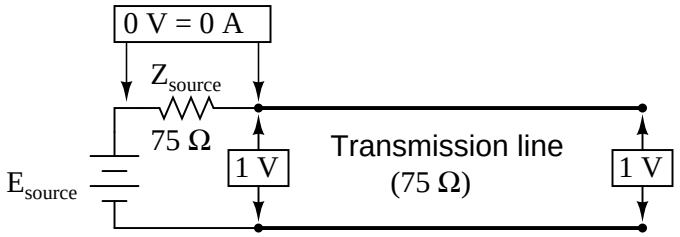



```
Transmission line
v1 1 0 ac 1 sin
rsource 1 2 75
t1 2 0 3 0 z0=75 td=1u
rload 3 0 999meg
.ac lin 101 1m 1meg
* Using "Nutmeg" program to plot analysis
.end
```

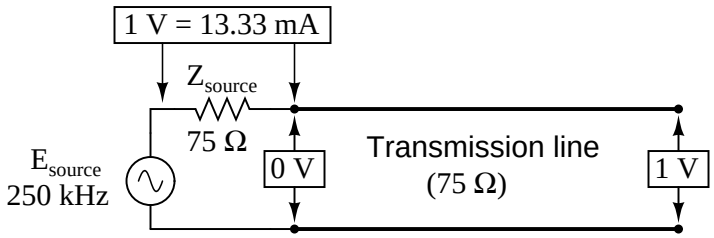


Here, both the supply voltage `vm(1)` and the line's load-end voltage `vm(3)` remain steady at 1 volt. The other voltages dip and peak at different frequencies along the sweep range of 1 mHz to 1 MHz. There are five points of interest along the horizontal axis of the analysis: 0 Hz, 250 kHz, 500 kHz, 750 kHz, and 1 MHz. We will investigate each one with regard to voltage and current at different points of the circuit.

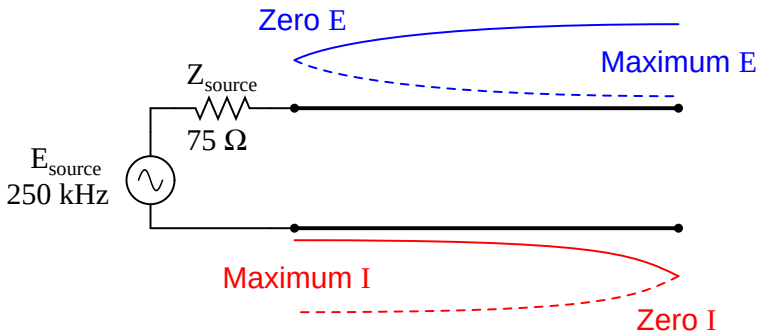
At 0 Hz (actually 1 mHz), the signal is practically DC, and the circuit behaves much as it would given a 1-volt DC battery source. There is no circuit current, as indicated by zero voltage drop across the source impedance (Z_{source}: `vm(1,2)`), and full source voltage present at the source-end of the transmission line (voltage measured between node 2 and node 0: `vm(2)`).



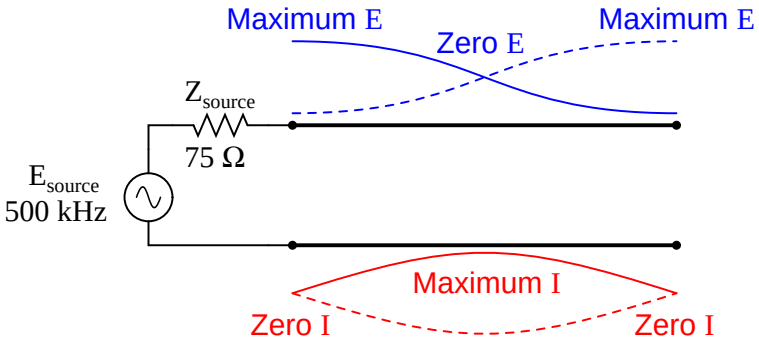
At 250 kHz, we see zero voltage and maximum current at the source-end of the transmission line, yet still full voltage at the load-end:



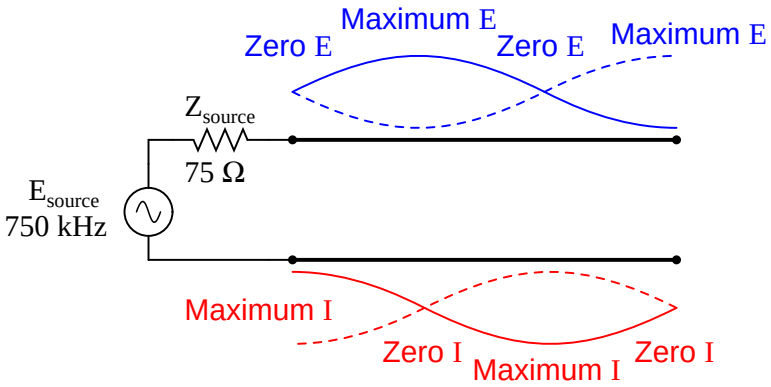
You might be wondering, how can this be? How can we get full source voltage at the line's open end while there is zero voltage at its entrance? The answer is found in the paradox of the standing wave. With a source frequency of 250 kHz, the line's length is precisely right for 1/4 wavelength to fit from end to end. With the line's load end open-circuited, there can be no current, but there will be voltage. Therefore, the load-end of an open-circuited transmission line is a current node (zero point) and a voltage antinode (maximum amplitude):



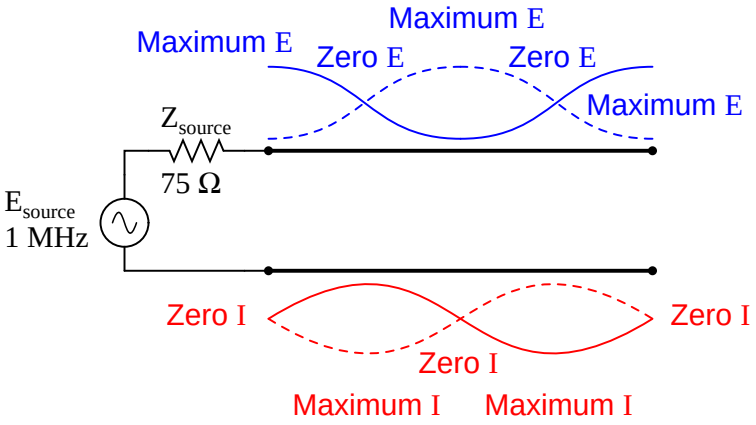
At 500 kHz, exactly one-half of a standing wave rests on the transmission line, and here we see another point in the analysis where the source current drops off to nothing and the source-end voltage of the transmission line rises again to full voltage:



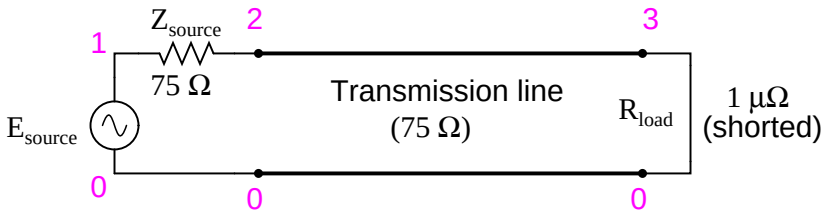
At 750 kHz, the plot looks a lot like it was at 250 kHz: zero source-end voltage (`vm(2)`) and maximum current (`vm(1,2)`). This is due to 3/4 of a wave poised along the transmission line, resulting in the source "seeing" a short-circuit where it connects to the transmission line, even though the other end of the line is open-circuited:



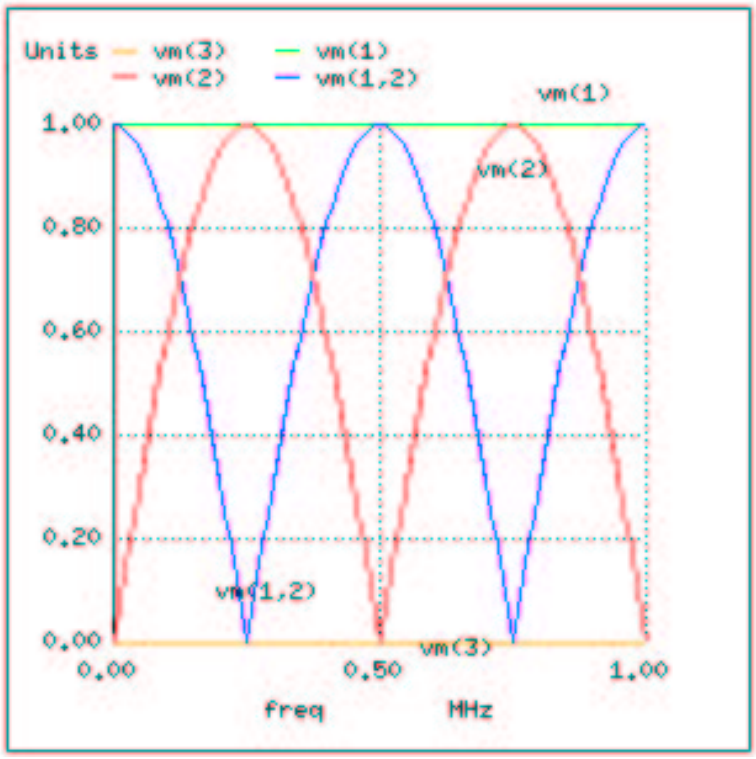
When the supply frequency sweeps up to 1 MHz, a full standing wave exists on the transmission line. At this point, the source-end of the line experiences the same voltage and current amplitudes as the load-end: full voltage and zero current. In essence, the source "sees" an open circuit at the point where it connects to the transmission line.

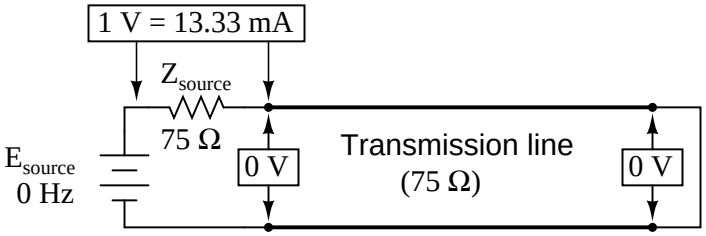


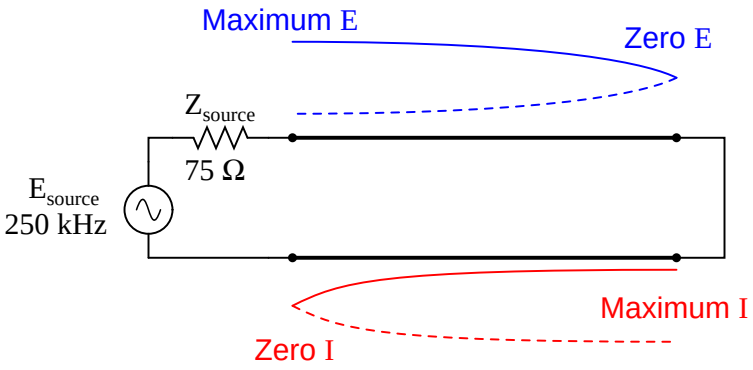
In a similar fashion, a short-circuited transmission line generates standing waves, although the node and antinode assignments for voltage and current are reversed: at the shorted end of the line, there will be zero voltage (node) and maximum current (antinode). What follows is the SPICE simulation and illustrations of what happens at all the interesting frequencies: 0 Hz, 250 kHz, 500 kHz, 750 kHz, and 1 MHz. The short-circuit jumper is simulated by a 1 $\mu\Omega$ load impedance:

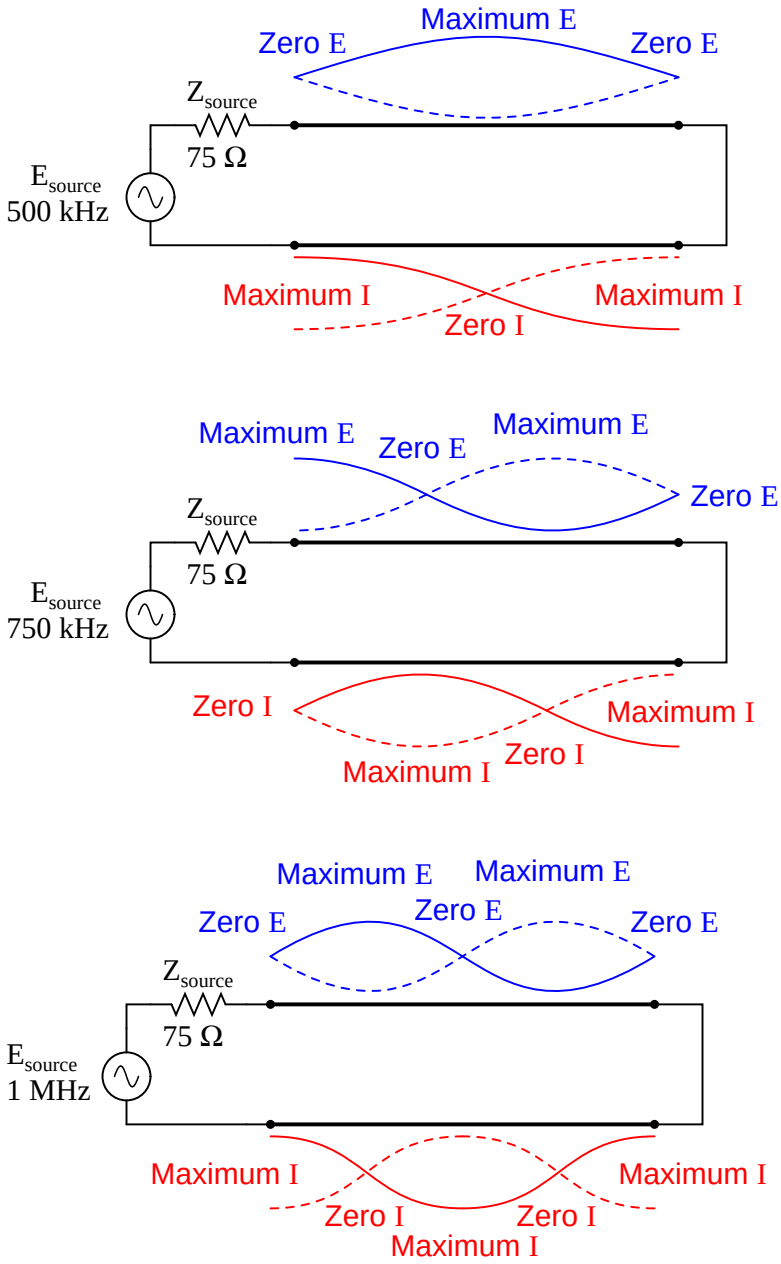



```
Transmission line
v1 1 0 ac 1 sin
rsource 1 2 75
t1 2 0 3 0 z0=75 td=1u
rload 3 0 1u
.ac lin 101 1m 1meg
* Using "Nutmeg" program to plot analysis
.end
```


Units
vm(3)
vm(1)
vm(2)
vm(1,2)
vm(1)
vm(2)
1.00
0.80
0.60
0.40
0.20
0.00
vm(1,2)
vm(3)
0.00
0.50
1.00
freq
MHz


1 V = 13.33 mA
Z source
75 Ω
E source
0 Hz
0 V
Transmission line
(75 Ω)
0 V

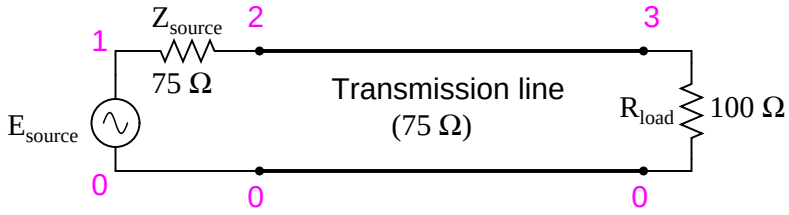

Maximum E
Zero E
Z source
75 Ω
E source
250 kHz
Maximum I
Zero I



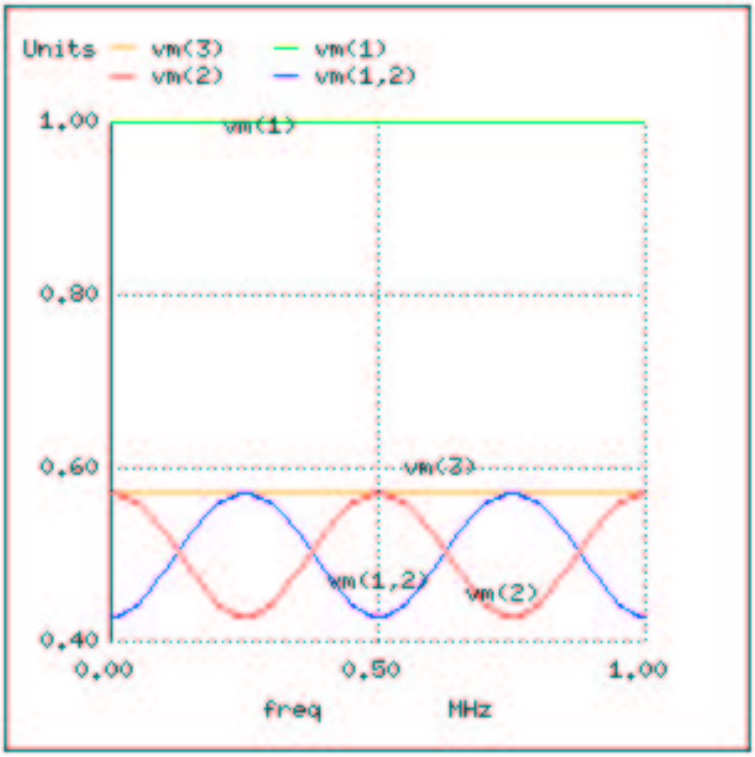
In both these circuit examples, an open-circuited line and a short-circuited line, the energy reflection is total: 100% of the incident wave reaching the line's end gets reflected back toward the source. If, however, the transmission line is terminated in some impedance other than an open or a short, the reflections will be less intense, as will be the difference between minimum and maximum values of voltage and current along the line.

Suppose we were to terminate our example line with a 100 Ω resistor instead of a 75 Ω resistor.

Examine the results of the corresponding SPICE analysis to see the effects of impedance mismatch at different source frequencies:




```
Transmission line
v1 1 0 ac 1 sin
rsource 1 2 75
t1 2 0 3 0 z0=75 td=1u
rload 3 0 100
.ac lin 101 1m 1meg
* Using "Nutmeg" program to plot analysis
.end
```



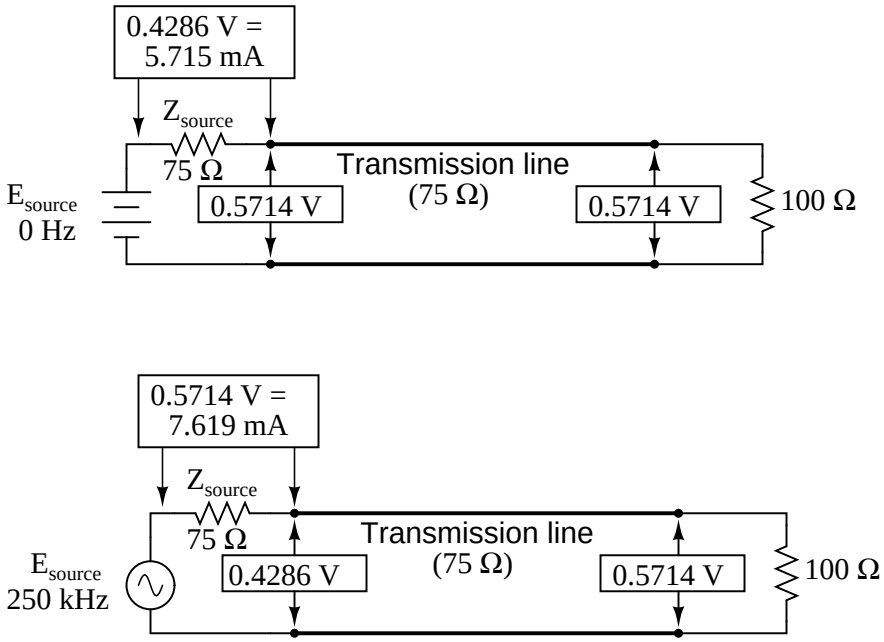
If we run another SPICE analysis, this time printing numerical results rather than plotting them, we can discover exactly what is happening at all the interesting frequencies (DC, 250 kHz, 500 kHz,

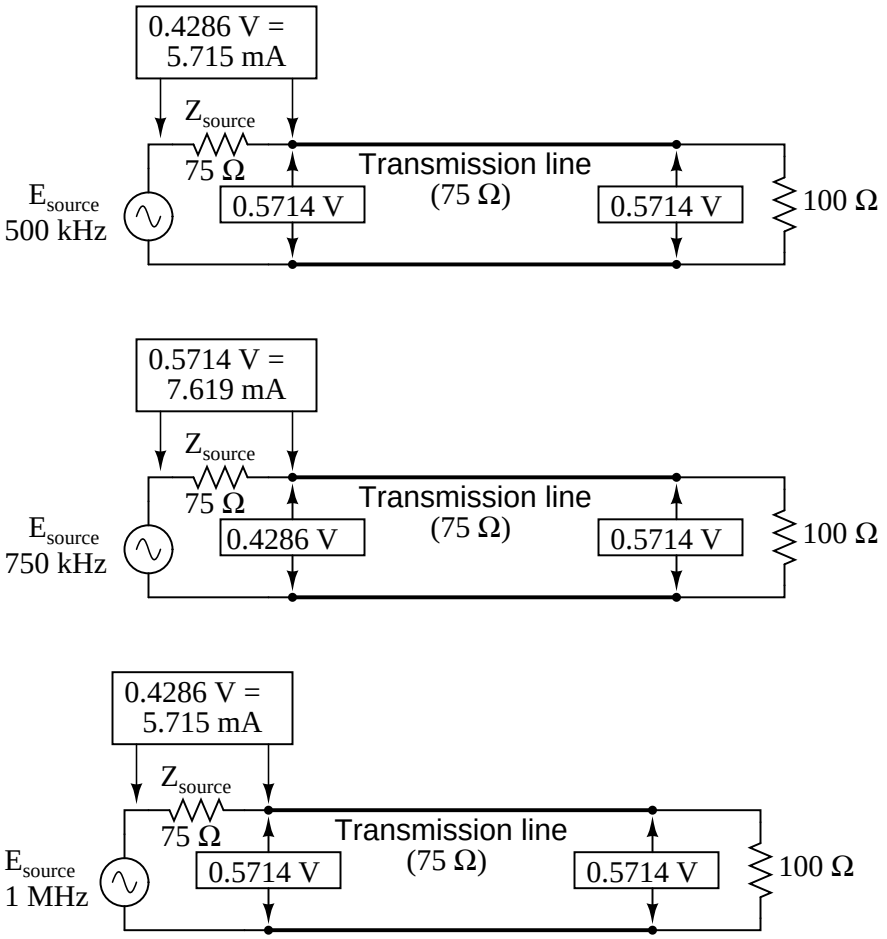
750 kHz, and 1 MHz):

```
Transmission line
v1 1 0 ac 1 sin
rsource 1 2 75
t1 2 0 3 0 z0=75 td=1u
rload 3 0 100
.ac lin 5 1m 1meg
.print ac v(1,2) v(1) v(2) v(3)
.end
```

```
freq          v(1,2)      v(1)        v(2)        v(3)
1.000E-03     4.286E-01   1.000E+00   5.714E-01   5.714E-01
2.500E+05     5.714E-01   1.000E+00   4.286E-01   5.714E-01
5.000E+05     4.286E-01   1.000E+00   5.714E-01   5.714E-01
7.500E+05     5.714E-01   1.000E+00   4.286E-01   5.714E-01
1.000E+06     4.286E-01   1.000E+00   5.714E-01   5.714E-01
```

At all frequencies, the source voltage, `v(1)`, remains steady at 1 volt, as it should. The load voltage, `v(3)`, also remains steady, but at a lesser voltage: 0.5714 volts. However, both the line input voltage (`v(2)`) and the voltage dropped across the source's 75 Ω impedance (`v(1,2)`, indicating current drawn from the source) vary with frequency.





At odd harmonics of the fundamental frequency (250 kHz and 750 kHz), we see differing levels of voltage at each end of the transmission line, because at those frequencies the standing waves terminate at one end in a node and at the other end in an antinode. Unlike the open-circuited and short-circuited transmission line examples, the maximum and minimum voltage levels along this transmission line do not reach the same extreme values of 0% and 100% source voltage, but we still have points of "minimum" and "maximum" voltage. The same holds true for current: if the line's terminating impedance is mismatched to the line's characteristic impedance, we will have points of minimum and maximum current at certain fixed locations on the line, corresponding to the standing current wave's nodes and antinodes, respectively.

One way of expressing the severity of standing waves is as a ratio of maximum amplitude (antinode) to minimum amplitude (node), for voltage or for current. When a line is terminated by an open or a short, this *standing wave ratio*, or *SWR* is valued at infinity, since the minimum amplitude will be zero, and any finite value divided by zero results in an infinite (actually, "undefined") quotient. In this example, with a 75 Ω line terminated by a 100 Ω impedance, the SWR will be finite: 1.333, calculated by taking the maximum line voltage at either 250 kHz or 750 kHz (0.5714 volts) and dividing by the minimum line voltage (0.4286 volts).

Standing wave ratio may also be calculated by taking the line's terminating impedance and the line's characteristic impedance, and dividing the larger of the two values by the smaller. In this example, the terminating impedance of 100 Ω divided by the characteristic impedance of 75 Ω yields a quotient of exactly 1.333, matching the previous calculation very closely.

$$\text{SWR} = \frac{E_{maximum}}{E_{minimum}} = \frac{I_{maximum}}{I_{minimum}}$$

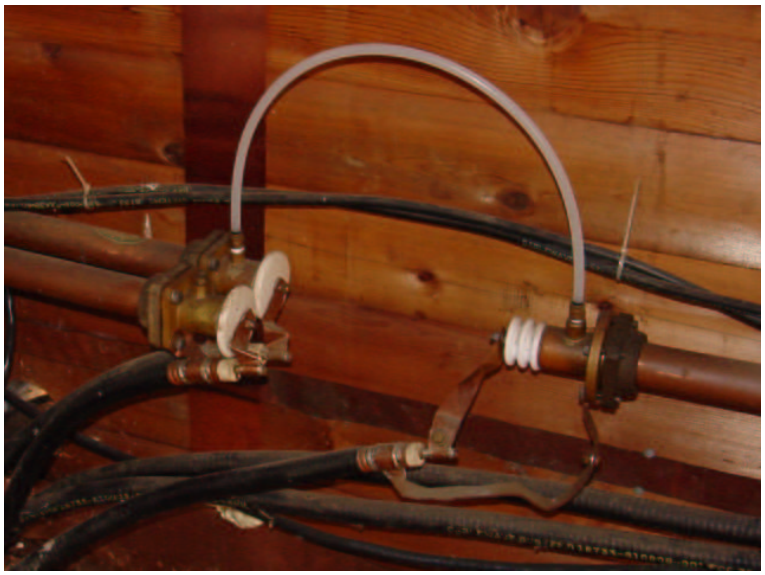
$$\text{SWR} = \frac{Z_{load}}{Z_0} \quad \textit{or} \quad \frac{Z_0}{Z_{load}}$$

which ever is greater

A perfectly terminated transmission line will have an SWR of 1, since voltage at any location along the line's length will be the same, and likewise for current. Again, this is usually considered ideal, not only because reflected waves constitute energy not delivered to the load, but because the high values of voltage and current created by the antinodes of standing waves may over-stress the transmission line's insulation (high voltage) and conductors (high current), respectively.

Also, a transmission line with a high SWR tends to act as an antenna, radiating electromagnetic energy away from the line, rather than channeling all of it to the load. This is usually undesirable, as the radiated energy may "couple" with nearby conductors, producing signal interference. An interesting footnote to this point is that antenna structures – which typically resemble open- or short-circuited transmission lines – are often designed to operate at *high* standing wave ratios, for the very reason of maximizing signal radiation and reception.

The following photograph shows a set of transmission lines at a junction point in a radio transmitter system. The large, copper tubes with ceramic insulator caps at the ends are rigid coaxial transmission lines of 50 Ω characteristic impedance. These lines carry RF power from the radio transmitter circuit to a small, wooden shelter at the base of an antenna structure, and from that shelter on to other shelters with other antenna structures:



Flexible coaxial cable connected to the rigid lines (also of 50 Ω characteristic impedance) conduct the RF power to capacitive and inductive "phasing" networks inside the shelter. The white, plastic tube joining two of the rigid lines together carries "filling" gas from one sealed line to the other. The lines are gas-filled to avoid collecting moisture inside them, which would be a definite problem for a coaxial line. Note the flat, copper "straps" used as jumper wires to connect the conductors of the flexible coaxial cables to the conductors of the rigid lines. Why flat straps of copper and not round wires? Because of the skin effect, which renders most of the cross-sectional area of a round conductor useless at radio frequencies.

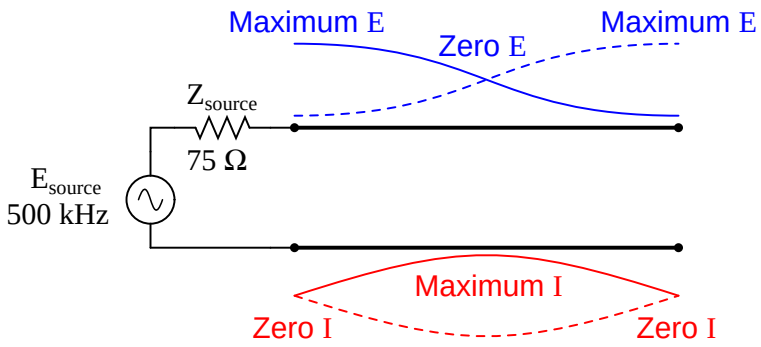
Like many transmission lines, these are operated at low SWR conditions. As we will see in the next section, though, the phenomenon of standing waves in transmission lines is not always undesirable, as it may be exploited to perform a useful function: impedance transformation.

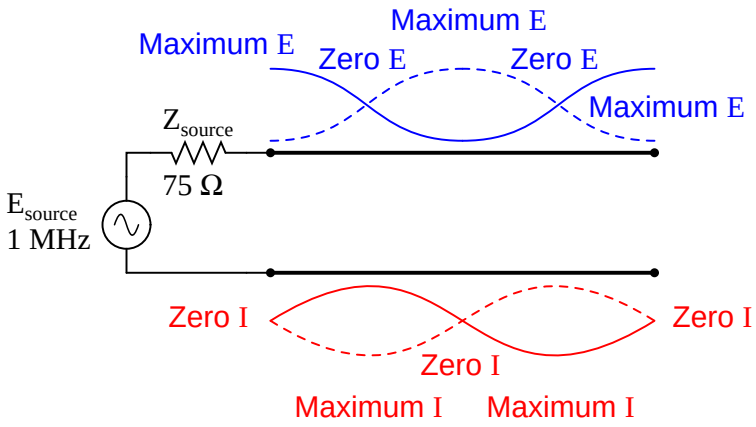
- **REVIEW:**
- *Standing waves* are waves of voltage and current which do not propagate (i.e. they are stationary), but are the result of interference between incident and reflected waves along a transmission line.
- A *node* is a point on a standing wave of *minimum* amplitude.
- An *antinode* is a point on a standing wave of *maximum* amplitude.
- Standing waves can only exist in a transmission line when the terminating impedance does not match the line's characteristic impedance. In a perfectly terminated line, there are no reflected waves, and therefore no standing waves at all.
- At certain frequencies, the nodes and antinodes of standing waves will correlate with the ends of a transmission line, resulting in *resonance*.

- The lowest-frequency resonant point on a transmission line is where the line is one quarter-wavelength long. Resonant points exist at every harmonic (integer-multiple) frequency of the fundamental (quarter-wavelength).
- *Standing wave ratio*, or *SWR*, is the ratio of maximum standing wave amplitude to minimum standing wave amplitude. It may also be calculated by dividing termination impedance by characteristic impedance, or visa-versa, which ever yields the greatest quotient. A line with no standing waves (perfectly matched: Z_{load} to Z_0) has an SWR equal to 1.
- Transmission lines may be damaged by the high maximum amplitudes of standing waves. Voltage antinodes may break down insulation between conductors, and current antinodes may overheat conductors.

14.7 Impedance transformation

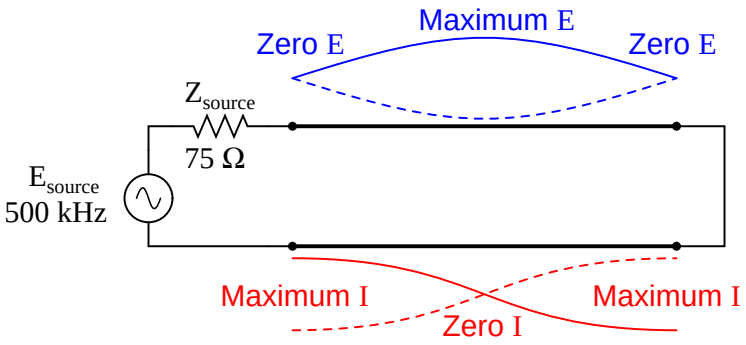
Standing waves at the resonant frequency points of an open- or short-circuited transmission line produce unusual effects. When the signal frequency is such that exactly 1/2 wave or some multiple thereof matches the line's length, the source "sees" the load impedance as it is. The following pair of illustrations shows an open-circuited line operating at 1/2 and 1 wavelength frequencies:

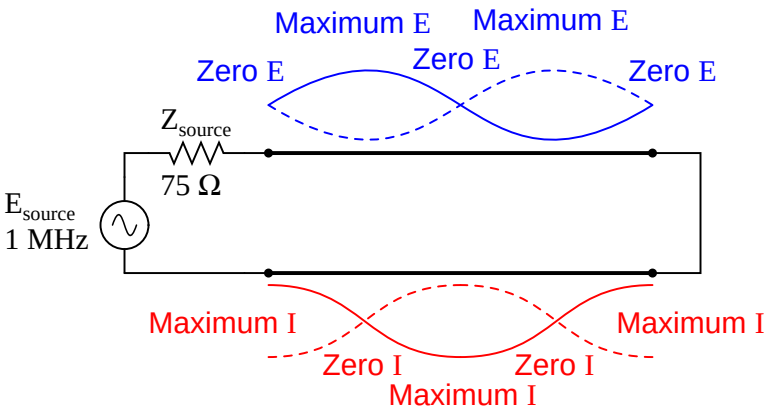




In either case, the line has voltage antinodes at both ends, and current nodes at both ends. That is to say, there is maximum voltage and minimum current at either end of the line, which corresponds to the condition of an open circuit. The fact that this condition exists at *both* ends of the line tells us that the line faithfully reproduces its terminating impedance at the source end, so that the source "sees" an open circuit where it connects to the transmission line, just as if it were directly open-circuited.

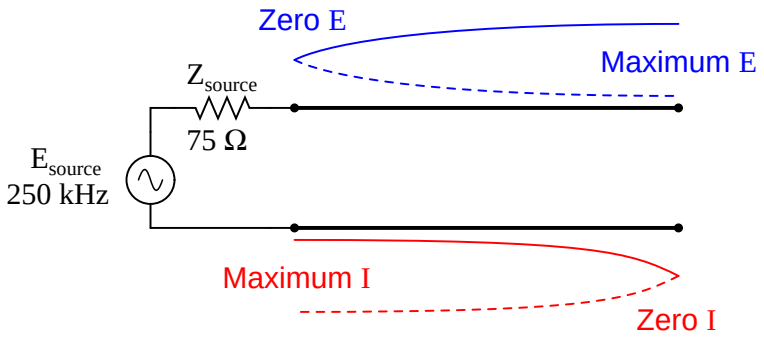
The same is true if the transmission line is terminated by a short: at signal frequencies corresponding to 1/2 wavelength or some multiple thereof, the source "sees" a short circuit, with minimum voltage and maximum current present at the connection points between source and transmission line:

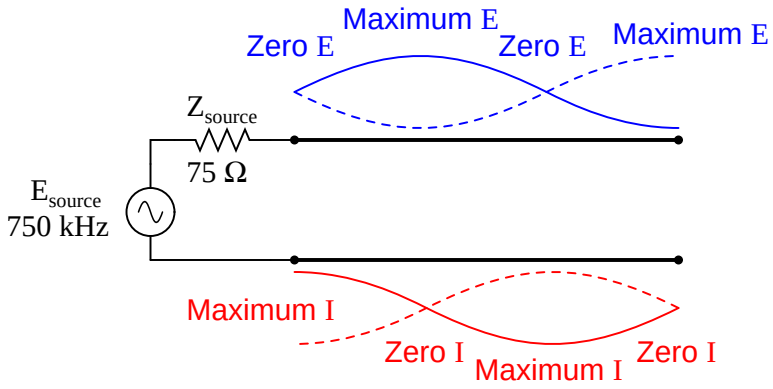




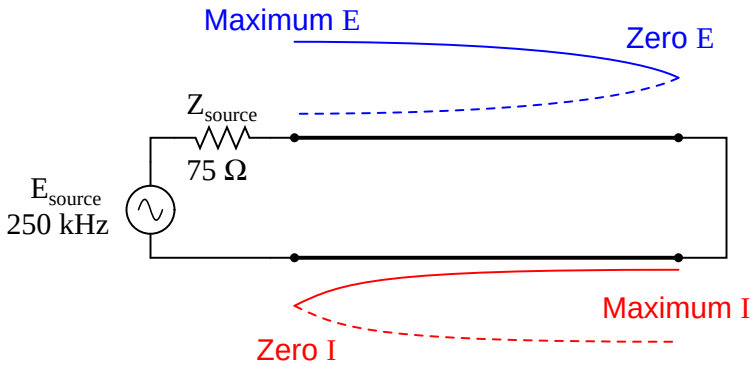
However, if the signal frequency is such that the line resonates at *1/4* wavelength or some multiple thereof, the source will "see" the exact opposite of the termination impedance. That is, if the line is open-circuited, the source will "see" a short-circuit at the point where it connects to the line; and if the line is short-circuited, the source will "see" an open circuit:

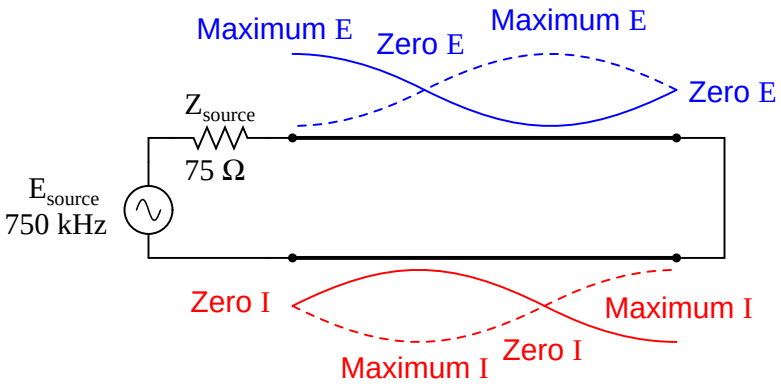
Line open-circuited; source "sees" a short circuit:





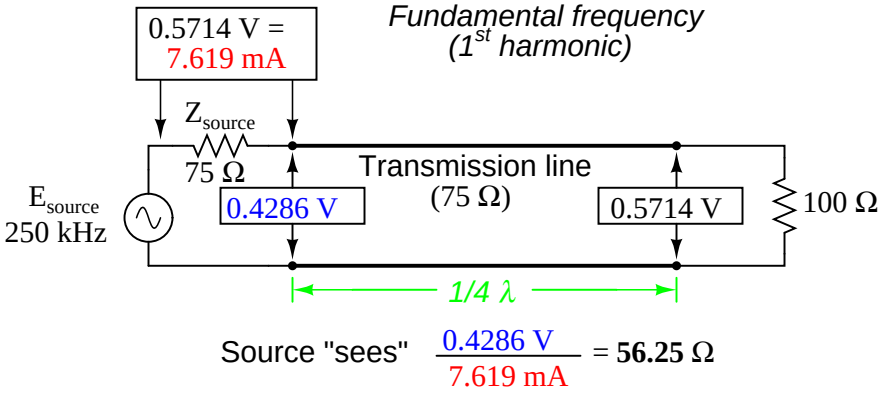
Line short-circuited; source "sees" an open circuit:

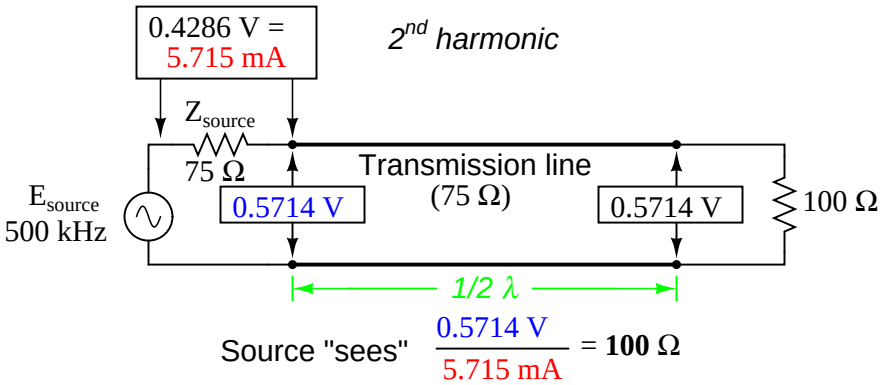


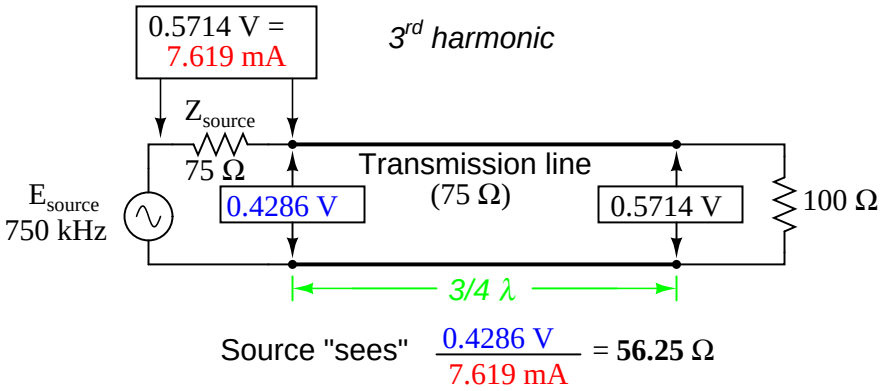


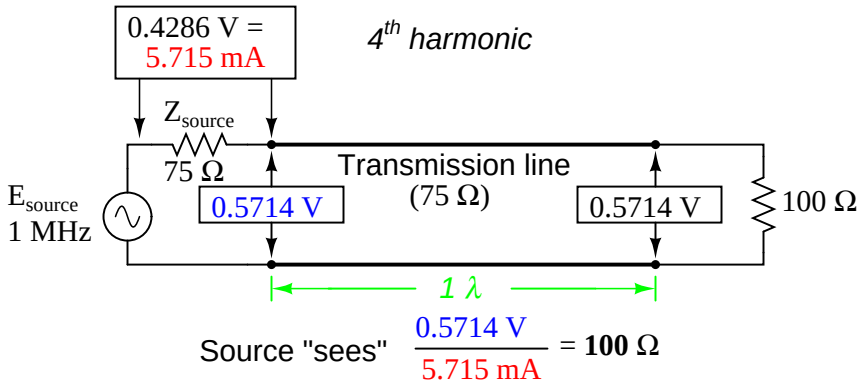
At these frequencies, the transmission line is actually functioning as an *impedance transformer*, transforming an infinite impedance into zero impedance, or visa-versa. Of course, this only occurs at resonant points resulting in a standing wave of 1/4 cycle (the line's fundamental, resonant frequency) or some odd multiple (3/4, 5/4, 7/4, 9/4 . . .), but if the signal frequency is known and unchanging, this phenomenon may be used to match otherwise unmatched impedances to each other.

Take for instance the example circuit from the last section where a 75 Ω source connects to a 75 Ω transmission line, terminating in a 100 Ω load impedance. From the numerical figures obtained via SPICE, let's determine what impedance the source "sees" at its end of the transmission line at the line's resonant frequencies:




0.4286 V =
5.715 mA
2nd harmonic
Zsource
75 Ω
Transmission line
(75 Ω)
Esource
500 kHz
0.5714 V
0.5714 V
100 Ω
1/2 λ
Source "sees" 0.5714 V / 5.715 mA = 100 Ω


0.5714 V =
7.619 mA
3rd harmonic
Zsource
75 Ω
Transmission line
(75 Ω)
Esource
750 kHz
0.4286 V
0.5714 V
100 Ω
3/4 λ
Source "sees" 0.4286 V / 7.619 mA = 56.25 Ω



A simple equation relates line impedance (Z_0), load impedance (Z_{load}), and input impedance (Z_{input}) for an unmatched transmission line operating at an odd harmonic of its fundamental frequency:

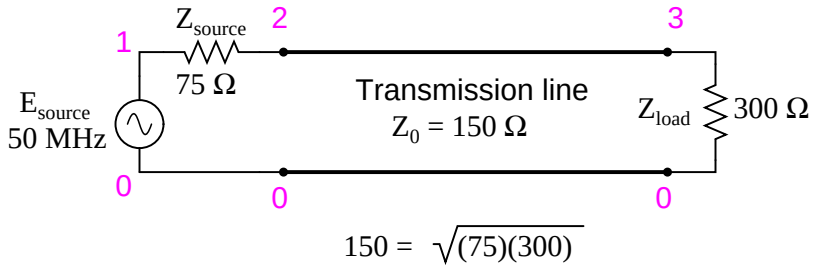
$$Z_0 = \sqrt{Z_{input} Z_{load}}$$

One practical application of this principle would be to match a 300 Ω load to a 75 Ω signal source at a frequency of 50 MHz. All we need to do is calculate the proper transmission line impedance (Z_0), and length so that exactly 1/4 of a wave will "stand" on the line at a frequency of 50 MHz.

First, calculating the line impedance: taking the 75 Ω we desire the source to "see" at the source-end of the transmission line, and multiplying by the 300 Ω load resistance, we obtain a figure of 22,500. Taking the square root of 22,500 yields 150 Ω for a characteristic line impedance.

Now, to calculate the necessary line length: assuming that our cable has a velocity factor of 0.85, and using a speed-of-light figure of 186,000 miles per second, the velocity of propagation will be 158,100 miles per second. Taking this velocity and dividing by the signal frequency gives us a wavelength of 0.003162 miles, or 16.695 feet. Since we only need one-quarter of this length for the cable to support a quarter-wave, the requisite cable length is 4.1738 feet.

Here is a schematic diagram for the circuit, showing node numbers for the SPICE analysis we're about to run:



We can specify the cable length in SPICE in terms of time delay from beginning to end. Since the frequency is 50 MHz, the signal period will be the reciprocal of that, or 20 nano-seconds (20 ns). One-quarter of that time (5 ns) will be the time delay of a transmission line one-quarter wavelength long:

```
Transmission line
```

```
v1 1 0 ac 1 sin
rsource 1 2 75
t1 2 0 3 0 z0=150 td=5n
rload 3 0 300
.ac lin 1 50meg 50meg
.print ac v(1,2) v(1) v(2) v(3)
.end

freq          v(1,2)      v(1)        v(2)        v(3)
5.000E+07     5.000E-01   1.000E+00   5.000E-01   1.000E+00
```

At a frequency of 50 MHz, our 1-volt signal source drops half of its voltage across the series 75 Ω impedance (`v(1,2)`) and the other half of its voltage across the input terminals of the transmission line (`v(2)`). This means the source "thinks" it is powering a 75 Ω load. The actual load impedance, however, receives a full 1 volt, as indicated by the 1.000 figure at `v(3)`. With 0.5 volt dropped across 75 Ω, the source is dissipating 3.333 mW of power: the same as dissipated by 1 volt across the 300 Ω load, indicating a perfect match of impedance, according to the Maximum Power Transfer Theorem. The 1/4-wavelength, 150 Ω, transmission line segment has successfully matched the 300 Ω load to the 75 Ω source.

Bear in mind, of course, that this only works for 50 MHz and its odd-numbered harmonics. For any other signal frequency to receive the same benefit of matched impedances, the 150 Ω line would have to lengthened or shortened accordingly so that it was exactly 1/4 wavelength long.

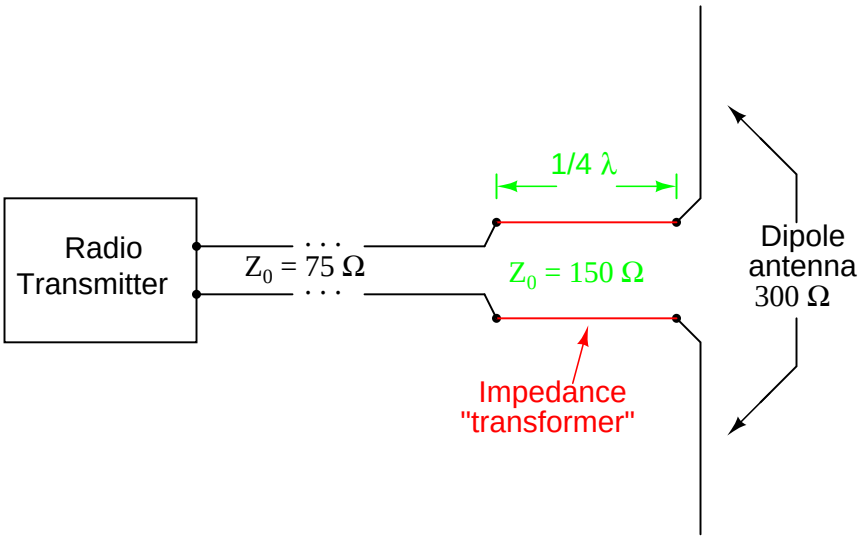
Strangely enough, the exact same line can also match a 75 Ω load to a 300 Ω source, demonstrating how this phenomenon of impedance transformation is fundamentally different in principle from that of a conventional, two-winding transformer:

```
Transmission line
v1 1 0 ac 1 sin
rsource 1 2 300
t1 2 0 3 0 z0=150 td=5n
rload 3 0 75
.ac lin 1 50meg 50meg
.print ac v(1,2) v(1) v(2) v(3)
.end

freq          v(1,2)      v(1)        v(2)        v(3)
5.000E+07     5.000E-01   1.000E+00   5.000E-01   2.500E-01
```

Here, we see the 1-volt source voltage equally split between the 300 Ω source impedance (`v(1,2)`) and the line's input (`v(2)`), indicating that the load "appears" as a 300 Ω impedance from the source's perspective where it connects to the transmission line. This 0.5 volt drop across the source's 300 Ω internal impedance yields a power figure of 833.33 μW, the same as the 0.25 volts across the 75 Ω load, as indicated by voltage figure `v(3)`. Once again, the impedance values of source and load have been matched by the transmission line segment.

This technique of impedance matching is often used to match the differing impedance values of transmission line and antenna in radio transmitter systems, because the transmitter's frequency is

generally well-known and unchanging. The use of an impedance "transformer" 1/4 wavelength in length provides impedance matching using the shortest conductor length possible.



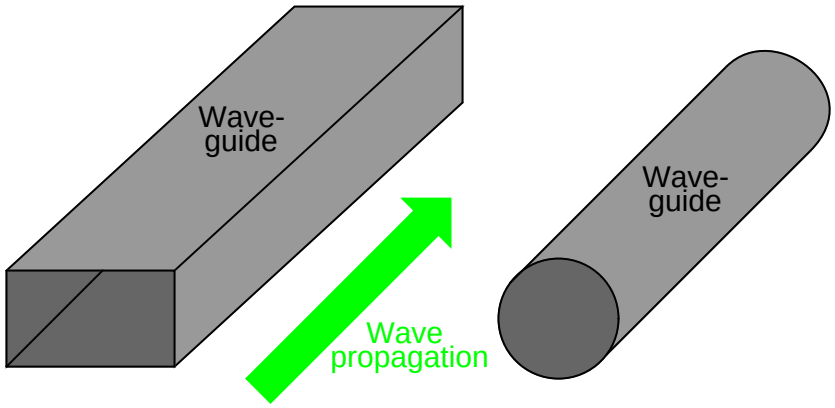
- **REVIEW:**

- A transmission line with standing waves may be used to match different impedance values if operated at the correct frequency(ies).

- When operated at a frequency corresponding to a standing wave of 1/4-wavelength along the transmission line, the line's characteristic impedance necessary for impedance transformation must be equal to the square root of the product of the source's impedance and the load's impedance.

14.8 Waveguides

A *waveguide* is a special form of transmission line consisting of a hollow, metal tube. The tube wall provides distributed inductance, while the empty space between the tube walls provide distributed capacitance:

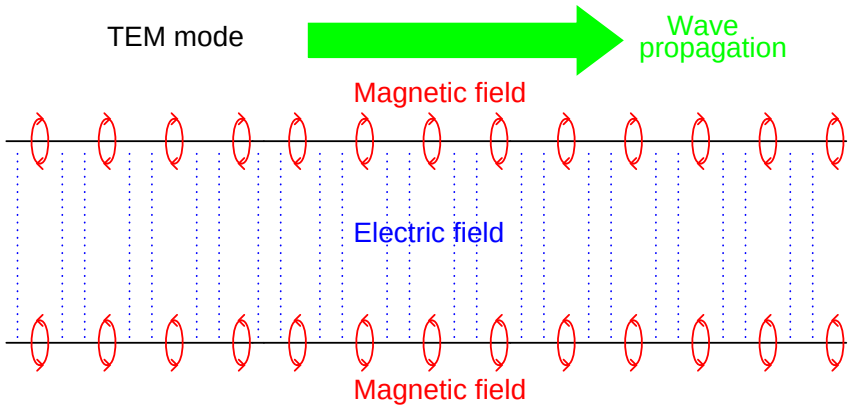


Waveguides are practical only for signals of extremely high frequency, where the wavelength approaches the cross-sectional dimensions of the waveguide. Below such frequencies, waveguides are useless as electrical transmission lines.

When functioning as transmission lines, though, waveguides are considerably simpler than two-conductor cables – especially coaxial cables – in their manufacture and maintenance. With only a single conductor (the waveguide's "shell"), there are no concerns with proper conductor-to-conductor spacing, or of the consistency of the dielectric material, since the only dielectric in a waveguide is air. Moisture is not as severe a problem in waveguides as it is within coaxial cables, either, and so waveguides are often spared the necessity of gas "filling."

Waveguides may be thought of as conduits for electromagnetic energy, the waveguide itself acting as nothing more than a "director" of the energy rather than as a signal conductor in the normal sense of the word. In a sense, all transmission lines function as conduits of electromagnetic energy when transporting pulses or high-frequency waves, directing the waves as the banks of a river direct a tidal wave. However, because waveguides are single-conductor elements, the propagation of electrical energy down a waveguide is of a very different nature than the propagation of electrical energy down a two-conductor transmission line.

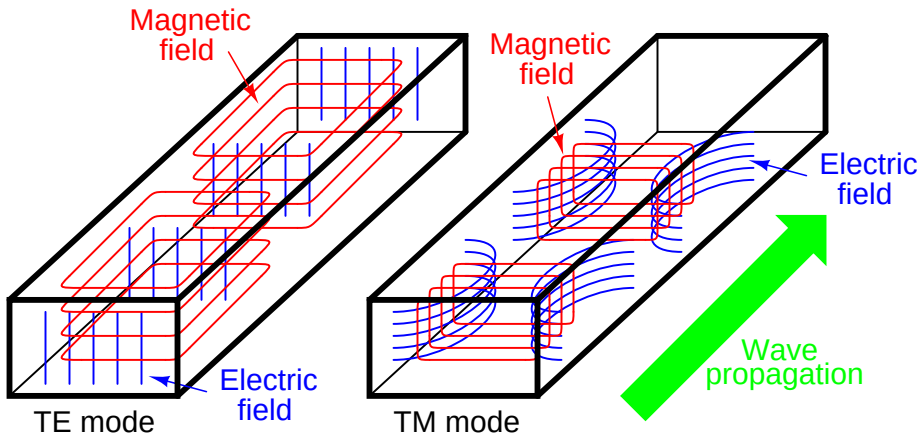
All electromagnetic waves consist of electric and magnetic fields propagating in the same direction of travel, but perpendicular to each other. Along the length of a normal transmission line, both electric and magnetic fields are perpendicular (transverse) to the direction of wave travel. This is known as the *principal mode*, or *TEM* (**T**ransverse **E**lectric and **M**agnetic) mode. This mode of wave propagation can exist only where there are two conductors, and it is the dominant mode of wave propagation where the cross-sectional dimensions of the transmission line are small compared to the wavelength of the signal.



Both field planes perpendicular (transverse) to direction of signal propagation.

At *microwave* signal frequencies (between 100 MHz and 300 GHz), two-conductor transmission lines of any substantial length operating in standard TEM mode become impractical. Lines small enough in cross-sectional dimension to maintain TEM mode signal propagation for microwave signals tend to have low voltage ratings, and suffer from large, parasitic power losses due to conductor "skin" and dielectric effects. Fortunately, though, at these short wavelengths there exist other modes of propagation that are not as "lossy," if a conductive tube is used rather than two parallel conductors. It is at these high frequencies that waveguides become practical.

When an electromagnetic wave propagates down a hollow tube, only one of the fields – either electric or magnetic – will actually be transverse to the wave's direction of travel. The other field will "loop" longitudinally to the direction of travel, but still be perpendicular to the other field. Whichever field remains transverse to the direction of travel determines whether the wave propagates in *TE* mode (**T**ransverse **E**lectric) or *TM* (**T**ransverse **M**agnetic) mode.

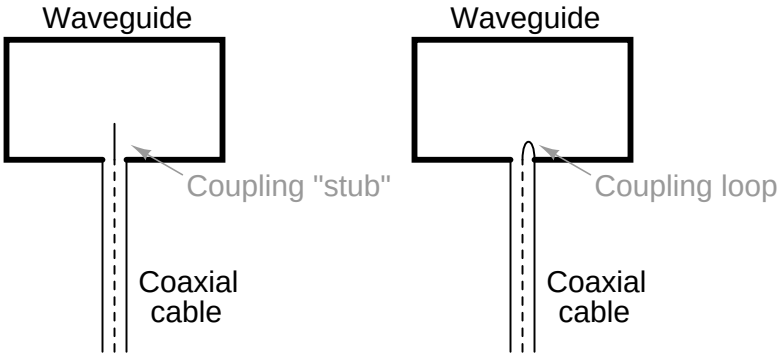


Magnetic flux lines appear as continuous loops
Electric flux lines appear with beginning and end points

Many variations of each mode exist for a given waveguide, and a full discussion of this is subject

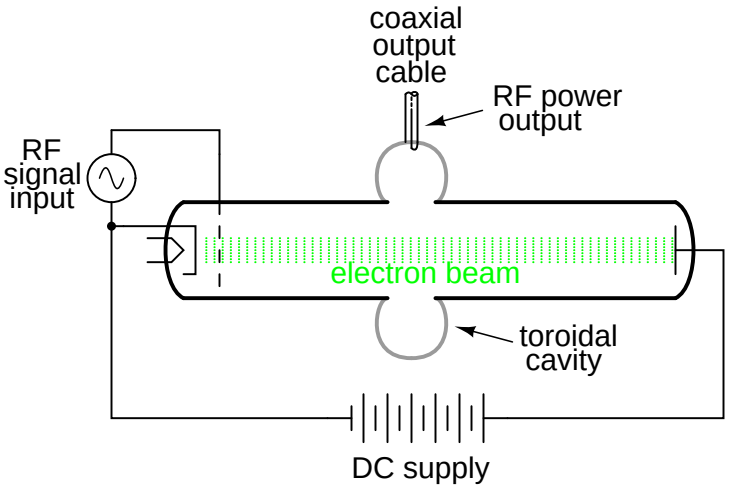
well beyond the scope of this book.

Signals are typically introduced to and extracted from waveguides by means of small antenna-like coupling devices inserted into the waveguide. Sometimes these coupling elements take the form of a dipole, which is nothing more than two open-ended stub wires of appropriate length. Other times, the coupler is a single stub (a half-dipole, similar in principle to a "whip" antenna, $1/4\lambda$ in physical length), or a short loop of wire terminated on the inside surface of the waveguide:



In some cases, such as a class of vacuum tube devices called *inductive output tubes* (the so-called *klystron* tube falls into this category), a "cavity" formed of conductive material may intercept electromagnetic energy from a modulated beam of electrons, having no contact with the beam itself:

The inductive output tube (IOT)

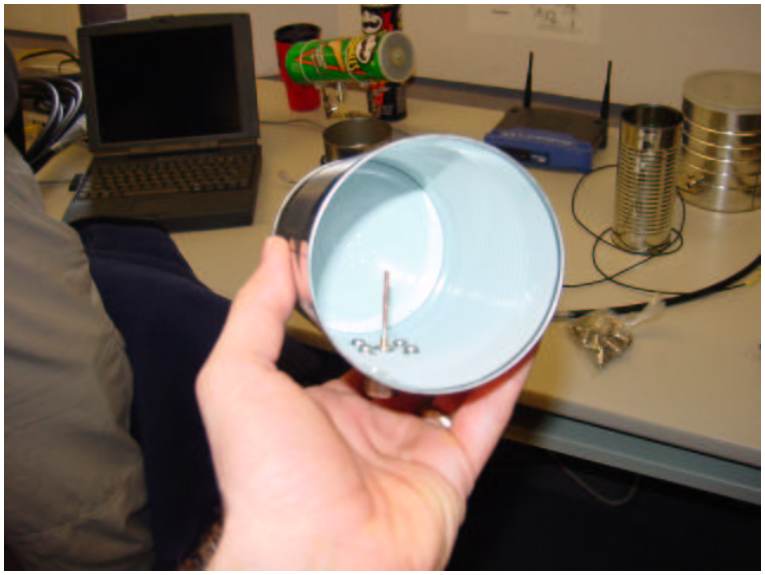


Just as transmission lines are able to function as resonant elements in a circuit, especially when terminated by a short-circuit or an open-circuit, a dead-ended waveguide may also resonate at particular frequencies. When used as such, the device is called a *cavity resonator*. Inductive output tubes use toroid-shaped cavity resonators to maximize the power transfer efficiency between the electron beam and the output cable.

A cavity's resonant frequency may be altered by changing its physical dimensions. To this end,

cavities with movable plates, screws, and other mechanical elements for tuning are manufactured to provide coarse resonant frequency adjustment.

If a resonant cavity is made open on one end, it functions as a unidirectional antenna. The following photograph shows a home-made waveguide formed from a tin can, used as an antenna for a 2.4 GHz signal in an "802.11b" computer communication network. The coupling element is a quarter-wave stub: nothing more than a piece of solid copper wire about 1-1/4 inches in length extending from the center of a coaxial cable connector penetrating the side of the can:



A few more tin-can antennae may be seen in the background, one of them a "Pringles" potato chip can. Although this can is of cardboard (paper) construction, its metallic inner lining provides the necessary conductivity to function as a waveguide. Some of the cans in the background still have their plastic lids in place. The plastic, being nonconductive, does not interfere with the RF signal, but functions as a physical barrier to prevent rain, snow, dust, and other physical contaminants from entering the waveguide. "Real" waveguide antennae use similar barriers to physically enclose the tube, yet allow electromagnetic energy to pass unimpeded.

- **REVIEW:**

- *Waveguides* are metal tubes functioning as "conduits" for carrying electromagnetic waves. They are practical only for signals of extremely high frequency, where the signal wavelength approaches the cross-sectional dimensions of the waveguide.

- Wave propagation through a waveguide may be classified into two broad categories: *TE* (Transverse Electric), or *TM* (Transverse Magnetic), depending on which field (electric or magnetic) is perpendicular (transverse) to the direction of wave travel. Wave travel along a standard, two-conductor transmission line is of the *TEM* (Transverse Electric and Magnetic) mode, where both fields are oriented perpendicular to the direction of travel. TEM mode is only possible with two conductors and cannot exist in a waveguide.

- A dead-ended waveguide serving as a resonant element in a microwave circuit is called a *cavity resonator*.
- A cavity resonator with an open end functions as a unidirectional antenna, sending or receiving RF energy to/from the direction of the open end.

Chapter 15

ABOUT THIS BOOK

15.1 Purpose

They say that necessity is the mother of invention. At least in the case of this book, that adage is true. As an industrial electronics instructor, I was forced to use a sub-standard textbook during my first year of teaching. My students were daily frustrated with the many typographical errors and obscure explanations in this book, having spent much time at home struggling to comprehend the material within. Worse yet were the many incorrect answers in the back of the book to selected problems. Adding insult to injury was the $100+ price.

Contacting the publisher proved to be an exercise in futility. Even though the particular text I was using had been in print and in popular use for a couple of years, they claimed my complaint was the first they'd ever heard. My request to review the draft for the next edition of their book was met with disinterest on their part, and I resolved to find an alternative text.

Finding a suitable alternative was more difficult than I had imagined. Sure, there were plenty of texts in print, but the really good books seemed a bit too heavy on the math and the less intimidating books omitted a lot of information I felt was important. Some of the best books were out of print, and those that were still being printed were quite expensive.

It was out of frustration that I compiled *Lessons in Electric Circuits* from notes and ideas I had been collecting for years. My primary goal was to put readable, high-quality information into the hands of my students, but a secondary goal was to make the book as affordable as possible. Over the years, I had experienced the benefit of receiving free instruction and encouragement in my pursuit of learning electronics from many people, including several teachers of mine in elementary and high school. Their selfless assistance played a key role in my own studies, paving the way for a rewarding career and fascinating hobby. If only I could extend the gift of their help by giving to other people what they gave to me . . .

So, I decided to make the book freely available. More than that, I decided to make it "open," following the same development model used in the making of free software (most notably the various UNIX utilities released by the Free Software Foundation, and the Linux operating system, whose fame is growing even as I write). The goal was to copyright the text – so as to protect my authorship – but expressly allow anyone to distribute and/or modify the text to suit their own needs with a minimum of legal encumbrance. This willful and formal revoking of standard distribution limitations

under copyright is whimsically termed *copyleft*. Anyone can "copyleft" their creative work simply by appending a notice to that effect on their work, but several Licenses already exist, covering the fine legal points in great detail.

The first such License I applied to my work was the GPL – General Public License – of the Free Software Foundation (GNU). The GPL, however, is intended to copyleft works of computer software, and although its introductory language is broad enough to cover works of text, its wording is not as clear as it could be for that application. When other, less specific copyleft Licenses began appearing within the free software community, I chose one of them (the Design Science License, or DSL) as the official notice for my project.

In "copylefting" this text, I guaranteed that no instructor would be limited by a text insufficient for their needs, as I had been with error-ridden textbooks from major publishers. I'm sure this book in its initial form will not satisfy everyone, but anyone has the freedom to change it, leveraging my efforts to suit variant and individual requirements. For the beginning student of electronics, learn what you can from this book, editing it as you feel necessary if you come across a useful piece of information. Then, if you pass it on to someone else, you will be giving them something better than what you received. For the instructor or electronics professional, feel free to use this as a reference manual, adding or editing to your heart's content. The only "catch" is this: if you plan to distribute your modified version of this text, you must give credit where credit is due (to me, the original author, and anyone else whose modifications are contained in your version), and you must ensure that whoever you give the text to is aware of their freedom to similarly share and edit the text. The next chapter covers this process in more detail.

It must be mentioned that although I strive to maintain technical accuracy in all of this book's content, the subject matter is broad and harbors many potential dangers. Electricity maims and kills without provocation, and deserves the utmost respect. I strongly encourage experimentation on the part of the reader, but only with circuits powered by small batteries where there is no risk of electric shock, fire, explosion, etc. High-power electric circuits should be left to the care of trained professionals! The Design Science License clearly states that neither I nor any contributors to this book bear any liability for what is done with its contents.

15.2 The use of SPICE

One of the best ways to learn how things work is to follow the inductive approach: to observe specific instances of things working and derive general conclusions from those observations. In science education, labwork is the traditionally accepted venue for this type of learning, although in many cases labs are designed by educators to reinforce principles previously learned through lecture or textbook reading, rather than to allow the student to learn on their own through a truly exploratory process.

Having taught myself most of the electronics that I know, I appreciate the sense of frustration students may have in teaching themselves from books. Although electronic components are typically inexpensive, not everyone has the means or opportunity to set up a laboratory in their own homes, and when things go wrong there's no one to ask for help. Most textbooks seem to approach the task of education from a deductive perspective: tell the student how things are supposed to work, then apply those principles to specific instances that the student may or may not be able to explore by themselves. The inductive approach, as useful as it is, is hard to find in the pages of a book.

However, textbooks don't have to be this way. I discovered this when I started to learn a

computer program called SPICE. It is a text-based piece of software intended to model circuits and provide analyses of voltage, current, frequency, etc. Although nothing is quite as good as building real circuits to gain knowledge in electronics, computer simulation is an excellent alternative. In learning how to use this powerful tool, I made a discovery: SPICE could be used within a textbook to present circuit simulations to allow students to "observe" the phenomena for themselves. This way, the readers could learn the concepts inductively (by interpreting SPICE's output) as well as deductively (by interpreting my explanations). Furthermore, in seeing SPICE used over and over again, they should be able to understand how to use it themselves, providing a perfectly safe means of experimentation on their own computers with circuit simulations of their own design.

Another advantage to including computer analyses in a textbook is the empirical verification it adds to the concepts presented. Without demonstrations, the reader is left to take the author's statements on faith, trusting that what has been written is indeed accurate. The problem with faith, of course, is that it is only as good as the authority in which it is placed and the accuracy of interpretation through which it is understood. Authors, like all human beings, are liable to err and/or communicate poorly. With demonstrations, however, the reader can immediately see for themselves that what the author describes is indeed true. Demonstrations also serve to clarify the meaning of the text with concrete examples.

SPICE is introduced early in volume I (DC) of this book series, and hopefully in a gentle enough way that it doesn't create confusion. For those wishing to learn more, a chapter in the Reference volume (volume V) contains an overview of SPICE with many example circuits. There may be more flashy (graphic) circuit simulation programs in existence, but SPICE is free, a virtue complementing the charitable philosophy of this book very nicely.

15.3 Acknowledgements

First, I wish to thank my wife, whose patience during those many and long evenings (and weekends!) of typing has been extraordinary.

I also wish to thank those whose open-source software development efforts have made this endeavor all the more affordable and pleasurable. The following is a list of various free computer software used to make this book, and the respective programmers:

- *GNU/Linux* Operating System – Linus Torvalds, Richard Stallman, and a host of others too numerous to mention.
- *Vim* text editor – Bram Moolenaar and others.
- *Xcircuit* drafting program – Tim Edwards.
- *SPICE* circuit simulation program – too many contributors to mention.
- *Nutmeg* post-processor program for SPICE – Wayne Christopher.
- TEX text processing system – Donald Knuth and others.
- *Texinfo* document formatting system – Free Software Foundation.
- LATEX document formatting system – Leslie Lamport and others.

- *Gimp* image manipulation program – too many contributors to mention.
- *Winscope* signal analysis software – Dr. Constantin Zeldovich. (Free for personal and academic use.)

Appreciation is also extended to Robert L. Boylestad, whose first edition of *Introductory Circuit Analysis* taught me more about electric circuits than any other book. Other important texts in my electronics studies include the 1939 edition of *The "Radio" Handbook*, Bernard Grob's second edition of *Introduction to Electronics I*, and Forrest Mims' original *Engineer's Notebook*.

Thanks to the staff of the Bellingham Antique Radio Museum, who were generous enough to let me terrorize their establishment with my camera and flash unit. Similar thanks to Jim Swartos and KARI radio in Blaine, Washington for a very informative tour of their expanded (50 kW) facilities as well as their vintage transmitter equipment.

I wish to specifically thank Jeffrey Elkner and all those at Yorktown High School for being willing to host my book as part of their Open Book Project, and to make the first effort in contributing to its form and content. Thanks also to David Sweet (website: `(http://www.andamooka.org)`) and Ben Crowell (website: `(http://www.lightandmatter.com)`) for providing encouragement, constructive criticism, and a wider audience for the online version of this book.

Thanks to Michael Stutz for drafting his Design Science License, and to Richard Stallman for pioneering the concept of copyleft.

Last but certainly not least, many thanks to my parents and those teachers of mine who saw in me a desire to learn about electricity, and who kindled that flame into a passion for discovery and intellectual adventure. I honor you by helping others as you have helped me.

Tony Kuphaldt, April 2002

"A candle loses nothing of its light when lighting another"
Kahlil Gibran

Chapter 16

CONTRIBUTOR LIST

16.1 How to contribute to this book

As a copylefted work, this book is open to revision and expansion by any interested parties. The only "catch" is that credit must be given where credit is due. This *is* a copyrighted work: it is *not* in the public domain!

If you wish to cite portions of this book in a work of your own, you must follow the same guidelines as for any other copyrighted work. Here is a sample from the Design Science License:

```
The Work is copyright the Author.  All rights to the Work are reserved
by the Author, except as specifically described below. This License
describes the terms and conditions under which the Author permits you
to copy, distribute and modify copies of the Work.

In addition, you may refer to the Work, talk about it, and (as
dictated by "fair use") quote from it, just as you would any
copyrighted material under copyright law.

Your right to operate, perform, read or otherwise interpret and/or
execute the Work is unrestricted; however, you do so at your own risk,
because the Work comes WITHOUT ANY WARRANTY -- see Section 7 ("NO
WARRANTY") below.
```

If you wish to modify this book in any way, you must document the nature of those modifications in the "Credits" section along with your name, and ideally, information concerning how you may be contacted. Again, the Design Science License:

```
Permission is granted to modify or sample from a copy of the Work,
producing a derivative work, and to distribute the derivative work
under the terms described in the section for distribution above,
provided that the following terms are met:
```

```
(a) The new, derivative work is published under the terms of this
License.

(b) The derivative work is given a new name, so that its name or
title can not be confused with the Work, or with a version of
the Work, in any way.

(c) Appropriate authorship credit is given: for the differences
between the Work and the new derivative work, authorship is
attributed to you, while the material sampled or used from
the Work remains attributed to the original Author; appropriate
notice must be included with the new work indicating the nature
and the dates of any modifications of the Work made by you.
```

Given the complexities and security issues surrounding the maintenance of files comprising this book, it is recommended that you submit any revisions or expansions to the original author (Tony R. Kuphaldt). You are, of course, welcome to modify this book directly by editing your own personal copy, but we would all stand to benefit from your contributions if your ideas were incorporated into the online "master copy" where all the world can see it.

16.2 Credits

All entries arranged in alphabetical order of surname. Major contributions are listed by individual name with some detail on the nature of the contribution(s), date, contact info, etc. Minor contributions (typo corrections, etc.) are listed by name only for reasons of brevity. Please understand that when I classify a contribution as "minor," it is in no way inferior to the effort or value of a "major" contribution, just smaller in the sense of less text changed. Any and all contributions are gratefully accepted. I am indebted to all those who have given freely of their own knowledge, time, and resources to make this a better book!

16.2.1 Tony R. Kuphaldt

- **Date(s) of contribution(s):** 1996 to present
- **Nature of contribution:** Original author.
- **Contact at:** `liec0@lycos.com`

16.2.2 Jason Starck

- **Date(s) of contribution(s):** May-June 2000
- **Nature of contribution:** HTML formatting, some error corrections.
- **Contact at:** `jstarck@yhslug.tux.org`

16.2.3 Your name here

- **Date(s) of contribution(s):** Month and year of contribution
- **Nature of contribution:** Insert text here, describing how you contributed to the book.
- **Contact at:** my_email@provider.net

16.2.4 Typo corrections and other "minor" contributions

- *The students of Bellingham Technical College's Instrumentation program.*
- **Bart Anderson** (January 2004) Corrected conceptual and safety errors regarding Tesla coils.
- **Ed Beroset** (May 2002) Suggested better ways to illustrate the meaning of the prefix "poly-" in chapter 10.
- **Duane Damiano** (February 2003) Pointed out magnetic polarity error in DC generator illustration.
- **Harvey Lew** (February 2003) Typo correction in Basic AC chapter: word "circuit" should have been "circle".
- **Jim Palmer** (May 2002) Typo correction on complex number math.
- **Don Stalkowski** (June 2002) Technical help with PostScript-to-PDF file format conversion.
- **John Symonds** (March 2002) Suggested an improved explanation of the unit "Hertz."
- **Joseph Teichman** (June 2002) Suggestion and technical help regarding use of PNG images instead of JPEG.
- **Mark D. Zarella** (April 2002) Suggested an improved explanation for the "average" value of a waveform.

Chapter 17

DESIGN SCIENCE LICENSE

Copyright © 1999-2000 Michael Stutz stutz@dsl.org
Verbatim copying of this document is permitted, in any medium.

17.1 0. Preamble

Copyright law gives certain exclusive rights to the author of a work, including the rights to copy, modify and distribute the work (the "reproductive," "adaptative," and "distribution" rights).

The idea of "copyleft" is to willfully revoke the exclusivity of those rights under certain terms and conditions, so that anyone can copy and distribute the work or properly attributed derivative works, while all copies remain under the same terms and conditions as the original.

The intent of this license is to be a general "copyleft" that can be applied to any kind of work that has protection under copyright. This license states those certain conditions under which a work published under its terms may be copied, distributed, and modified.

Whereas "design science" is a strategy for the development of artifacts as a way to reform the environment (not people) and subsequently improve the universal standard of living, this Design Science License was written and deployed as a strategy for promoting the progress of science and art through reform of the environment.

17.2 1. Definitions

"License" shall mean this Design Science License. The License applies to any work which contains a notice placed by the work's copyright holder stating that it is published under the terms of this Design Science License.

"Work" shall mean such an aforementioned work. The License also applies to the output of the Work, only if said output constitutes a "derivative work" of the licensed Work as defined by copyright law.

"Object Form" shall mean an executable or performable form of the Work, being an embodiment of the Work in some tangible medium.

"Source Data" shall mean the origin of the Object Form, being the entire, machine-readable, preferred form of the Work for copying and for human modification (usually the language, encoding

or format in which composed or recorded by the Author); plus any accompanying files, scripts or other data necessary for installation, configuration or compilation of the Work.

(Examples of "Source Data" include, but are not limited to, the following: if the Work is an image file composed and edited in 'PNG' format, then the original PNG source file is the Source Data; if the Work is an MPEG 1.0 layer 3 digital audio recording made from a 'WAV' format audio file recording of an analog source, then the original WAV file is the Source Data; if the Work was composed as an unformatted plaintext file, then that file is the the Source Data; if the Work was composed in LaTeX, the LaTeX file(s) and any image files and/or custom macros necessary for compilation constitute the Source Data.)

"Author" shall mean the copyright holder(s) of the Work.

The individual licensees are referred to as "you."

17.3 2. Rights and copyright

The Work is copyright the Author. All rights to the Work are reserved by the Author, except as specifically described below. This License describes the terms and conditions under which the Author permits you to copy, distribute and modify copies of the Work.

In addition, you may refer to the Work, talk about it, and (as dictated by "fair use") quote from it, just as you would any copyrighted material under copyright law.

Your right to operate, perform, read or otherwise interpret and/or execute the Work is unrestricted; however, you do so at your own risk, because the Work comes WITHOUT ANY WARRANTY – see Section 7 ("NO WARRANTY") below.

17.4 3. Copying and distribution

Permission is granted to distribute, publish or otherwise present verbatim copies of the entire Source Data of the Work, in any medium, provided that full copyright notice and disclaimer of warranty, where applicable, is conspicuously published on all copies, and a copy of this License is distributed along with the Work.

Permission is granted to distribute, publish or otherwise present copies of the Object Form of the Work, in any medium, under the terms for distribution of Source Data above and also provided that one of the following additional conditions are met:

(a) The Source Data is included in the same distribution, distributed under the terms of this License; or

(b) A written offer is included with the distribution, valid for at least three years or for as long as the distribution is in print (whichever is longer), with a publicly-accessible address (such as a URL on the Internet) where, for a charge not greater than transportation and media costs, anyone may receive a copy of the Source Data of the Work distributed according to the section above; or

(c) A third party's written offer for obtaining the Source Data at no cost, as described in paragraph (b) above, is included with the distribution. This option is valid only if you are a non-commercial party, and only if you received the Object Form of the Work along with such an offer.

You may copy and distribute the Work either gratis or for a fee, and if desired, you may offer warranty protection for the Work.

The aggregation of the Work with other works which are not based on the Work – such as but not limited to inclusion in a publication, broadcast, compilation, or other media – does not bring the other works in the scope of the License; nor does such aggregation void the terms of the License for the Work.

17.5 4. Modification

Permission is granted to modify or sample from a copy of the Work, producing a derivative work, and to distribute the derivative work under the terms described in the section for distribution above, provided that the following terms are met:

(a) The new, derivative work is published under the terms of this License.

(b) The derivative work is given a new name, so that its name or title can not be confused with the Work, or with a version of the Work, in any way.

(c) Appropriate authorship credit is given: for the differences between the Work and the new derivative work, authorship is attributed to you, while the material sampled or used from the Work remains attributed to the original Author; appropriate notice must be included with the new work indicating the nature and the dates of any modifications of the Work made by you.

17.6 5. No restrictions

You may not impose any further restrictions on the Work or any of its derivative works beyond those restrictions described in this License.

17.7 6. Acceptance

Copying, distributing or modifying the Work (including but not limited to sampling from the Work in a new work) indicates acceptance of these terms. If you do not follow the terms of this License, any rights granted to you by the License are null and void. The copying, distribution or modification of the Work outside of the terms described in this License is expressly prohibited by law.

If for any reason, conditions are imposed on you that forbid you to fulfill the conditions of this License, you may not copy, distribute or modify the Work at all.

If any part of this License is found to be in conflict with the law, that part shall be interpreted in its broadest meaning consistent with the law, and no other parts of the License shall be affected.

17.8 7. No warranty

THE WORK IS PROVIDED "AS IS," AND COMES WITH ABSOLUTELY NO WARRANTY, EXPRESS OR IMPLIED, TO THE EXTENT PERMITTED BY APPLICABLE LAW, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

17.9 8. Disclaimer of liability

IN NO EVENT SHALL THE AUTHOR OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS WORK, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

END OF TERMS AND CONDITIONS

```
[$Id:  dsl.txt,v 1.25 2000/03/14 13:14:14 m Exp m $]
```

Index

λ, symbol for wavelength, 391
ω, symbol for angular velocity, 59, 80, 303
400 Hz AC power, aircraft, 248

Absolute value, vector, 35
AC, 1
AC circuit analysis, 112
Admittance, 112
Alternating current, 1
Alternator, 2, 269
Ammeter, 335
Amp-turn, unit, 202
Amplifier, 243
Amplifier, magnetic, 245
Amplitude, 12
Amplitude, average, 13
Amplitude, peak, 12
Amplitude, peak-to-peak, 12
Amplitude, RMS, 15
Amplitude, vector, 30
Analysis, AC circuit, 112
Analysis, Fourier, 149
Analysis, frequency domain, 152
Analysis, parallel, 100
Analysis, series, 95
Analysis, series-parallel, 103, 112
Analysis, time domain, 152
Analyzer, spectrum, 151
Angle, vector, 35
Angular velocity, 59, 80, 303
Antenna, 22, 73
Antinode, 397
Antiresonance, 133
Antiresonance, mechanical, 136
Arithmetic, with complex numbers, 40
Artifact, measurement, 160
Atomic clock, 345
Attenuation, 187
Autotransformer, 228
Average amplitude, AC, 13, 339

B, symbol for magnetic flux density, 112
B, symbol for susceptance, 112
Band-elimination filter, 188
Band-pass filter, 185
Band-reject filter, 188
Band-stop filter, 188
Bode plot, 123, 152, 175
Boosting, AC voltage sources, 263
Boosting, transformer connection, 229
Bridge circuit, 354, 369
Bridge, Maxwell-Wien, 359
Bridge, symmetrical, 357
Bridge, Wheatstone, 344
Bridge, Wien, 358
Brush, generator, 2
Bucking, AC voltage sources, 263
Bucking, transformer connection, 229

C, symbol for capacitance, 78
Cable, coaxial, 375
Cable, shielded, 140
Cable, twisted pair, 140
Calculus, 56, 78
Capacitance, 78
Capacitive coupling, 139, 173
Capacitive reactance, 78, 80
Capacitor, 78
Capacitor, decoupling, 180
Capacitor, multiplier, 337
Capacitor, variable, 367
Cathode Ray Tube, 338
Celsius, unit, 8
Centigrade, unit, 8

Characteristic impedance, 381
Chord, music, 142
Class, transformer heat rating, 256
Clock, atomic, 345
Coaxial cable, 375
Coil, primary, 204
Coil, secondary, 204
Coil, Tesla, 241
Color (musical), 143
Common-mode voltage, 222
Complex number, 21, 27, 323
Complex number arithmetic, 40
Conductance, 111
Conservation of Energy, Law of, 217
Core loss, 133
Counter, 344
Coupling, signal, 139, 173
CPS, unit, 8
Crest, 12
Crest factor, 16
Crossover network, 175
CRT, 338
Crystal, 344
CT, 240, 340
Current transformer, 240, 340
Current, line, 282
Current, phase, 282
Cutoff frequency, 178, 184
Cycle, 7

D'Arsonval meter movement, 16, 335
DC, 1
DC equivalent, AC measurement, 15
Decoupling capacitor, 180
Degree vs. radian, 59
Delta configuration, 281
Derivative, calculus, 251
Detector, null, 354
Dielectric constant, 383
Dielectric heating, 90
Diode, 153, 154, 335
Dipole antenna, 23
Direct current, 1
Distortion, inductor current, 203
Domain, frequency, 152
Domain, time, 152
Dot convention, transformer, 225
Duty cycle, 371

e, symbol for instantaneous voltage, 55, 56, 77, 78
E, symbol for voltage, 61, 83
ECG, 8
Eddy current, 71
Eddy current loss, 133, 247
Effective resistance, 71
EKG, 8
Electric field, 22, 115
Electrocardiograph, 8
Electrolytic capacitor, 91
Electromagnetic induction, 2
Electromagnetic wave, 22
Electrostatic meter movement, 338
Energy, kinetic, 117
Energy, potential, 117
Equalizer, 175
Equalizer, graphic, 151
Equivalent, AC to DC, 15
Exciting current, 203, 207

f, symbol for frequency, 58, 80
Factor, crest, 16
Factor, form, 16
Factor, power, 319
Farad, 78
Fast Fourier Transform, 151, 354
Ferrite, 71
Ferroresonant transformer, 232
FFT, 151, 354
Field, electric, 22, 115
Field, magnetic, 22, 115
Figure, Lissajous, 346
Filter, 127, 175
Filter, band-elimination, 188
Filter, band-pass, 185
Filter, band-reject, 188
Filter, band-stop, 188
Filter, high-pass, 181
Filter, low-pass, 176
Filter, notch, 188
Filter, resonant, 190
Form factor, 16

Fourier analysis, 149
Fourier Transform, 149
Frequency, 8, 342
Frequency meter, 342
Frequency, cutoff, 178, 184
Frequency, fundamental, 142, 398
Full-wave rectification, 156
Function, sine, 6
Fundamental frequency, 142, 398

G, symbol for conductance, 111
Generator, 2, 269
Graphic equalizer, 151
Ground, 31

Half-wave rectification, 154
Hall effect, 350
Harmonic, 142, 293, 398
Harmonic sequence, 316
Harmonic, even vs. odd, 168, 293
Harmonics and waveform symmetry, 168, 293
Harmonics, triplen, 303, 315
Headphones, as sensitive null detector, 355
Heating, dielectric, 90
Heating, inductive, 71
Henry, 56
Hertz, unit, 8
High-pass filter, 181
Hot conductor, 262
Hyperbolic function, trigonometry, 41
Hysteresis, 248

i, imaginary operator, 36
I, symbol for current, 61, 83
i, symbol for instantaneous current, 55, 77, 78
i, symbol for instantaneous voltage, 56
Imaginary number, 37
Impedance, 27, 61, 82, 95, 112
Impedance matching, 235, 236
Impedance, characteristic, 381
Incident wave, 385
Inductance, 56
Inductance, leakage, 211, 213, 249
Induction, electromagnetic, 2
Induction, mutual, 3
Inductive coupling, 139, 173
Inductive heating, 71
Inductive reactance, 56, 58
Inductor, 56
Inrush current, transformer, 252
Instantaneous value, 55, 77
Integral, calculus, 251
Iron-vane meter movement, 16, 338
Isolation transformer, 222
Isolation, transformer, 220

j, imaginary operator, 36
Joule, 352
Joule's Law, 47

KCL, 19, 96
Keyboard, piano, 10
Kirchhoff's Current Law, 19, 96
Kirchhoff's Voltage Law, 19, 47, 96
Klystron tube, 422
KVL, 19, 47, 96

L, symbol for inductance, 56
Lagging phase shift, 21, 57, 79, 329
Laminated iron core, 248
Leading phase shift, 21, 57, 79, 329
Leakage inductance, 211, 213, 249
Lenz's Law, 56
Line, polyphase system, 282
Lissajous figure, 346
Litz wire, 71
Load, nonlinear, 352
Loop antenna, 23
Low-pass filter, 176
LVDT, 363

M, symbol for mutual inductance, 205
Magnetic amplifier, 245
Magnetic field, 22, 115
Magnetic field, rotating, 278
Magnetizing current, 203
Magnetomotive force, 202
Magnetostriction, 255
Magnitude, 12
Maximum Power Transfer Theorem, 236
Maxwell-Wien bridge circuit, 359
Meter movement, 335

Meter, power factor, 330
Mho, unit, 111
Microwaves, 421
MMF, 202
Modulus. vector, 35
Motor, 275
Motor, induction, 277
Motor, synchronous, 277
Motor/generator set, 217
Multiplier, 337
Mutual inductance, 205
Mutual induction, 3

Natural impedance, 383
Negative sequence, 316
Neutral conductor, 262
Node, vs. antinode, 397
Noise, transformer, 255
Non-sinusoidal, 11
Nonlinear components, 154, 293, 352
Nonsinusoidal, 139
Norton's Theorem, 236
Notation, polar, 35
Notation, rectangular, 35
Notch filter, 188
Null detector, 354
Null detector, AC bridge, 355
Null meter, 354
Number, complex, 21, 27, 323
Number, imaginary, 37
Number, real, 37
Number, scalar, 27, 323

Octave, 10
Ohm's Law, 19, 47, 83, 96
Ohm's Law , 61
Ohm, unit, 112
Oscillation, 117
Oscillator, 127
Oscilloscope, 8, 346
Overtone, 142, 354, 398

p, symbol for instantaneous power, 56, 78
P, symbol for true power, 323
Parallel analysis, 100
Parallel circuit rules, 96
Parallel LC resonance, 121
PCB, 181
Peak, 12
Peak-to-peak, 12
Peaking transformer, 250
Pendulum, 115
Period, 8
Permanent magnet moving coil, 335
Permittivity, relative, 383
Phase, 20
Phase rotation, 271
Phase sequence, 272
Phase shift, 20
Phase shift, vector, 30
Phase, transformer, 222
Piano, 10
Piezoelectricity, 344
Pitch (musical), 9
PMMC, 335
Polar notation, 35
Polarity, AC, 42, 52, 262
Pole, alternator, 270
Polyphase, 265, 268
Positive sequence, 316
Potential transformer, 239, 340
Potentiometer, 363
Powdered iron core, 248
Power factor, 319
Power factor meter, 330
Power quality, 352
Power quality meter, 353
Power triangle, 325
Power, apparent, 323
Power, negative, 58
Power, reactive, 323
Power, true, 323
Primary coil, 204
Primary transformer coil, 4
Principal mode, 420
Printed circuit board, 181
PT, 239, 340
Pythagorean Theorem, 39

Q, quality factor, 345
Q, symbol for quality factor, 72
Q, symbol for reactive power, 323

Quality factor, 72
Quartz crystal, 344

R, symbol for resistance, 61, 83, 111
Radian, angular measurement, 59, 80
Radio, 22, 73
Radio wave, 22
Ratio, transformer, 216
Reactance, 112
Reactance, capacitive, 78, 80
Reactance, inductive, 56, 58
Real number, 37
Rectangular notation, 35
Rectification, full-wave, 156
Rectification, half-wave, 154
Rectifier, 335
Rectifier, silicon-controlled, 153
Reflected wave, 385
Reflectometer, time-domain, 385
Relative permittivity, 383
Reluctance, 202
Resistance, 111
Resistance, AC, 73
Resistance, DC, 73
Resistance, effective, 71
Resistor, multiplier, 337, 340
Resistor, shunt, 340
Resonance, 118, 398
Resonance, mechanical, 136, 342
Resonance, parallel, 119
Resonance, parallel LC, 121
Resonance, series LC, 123
Resonance, series-parallel, 128
Resonance, transformer and inductor, 249
Resonant filter, 190
Resonant frequency formula, 119
Resonant frequency meter, 343
Resonate, 118
Resultant vector, 51
RF: Radio Frequency, 73, 376
RMS, 15, 338
Root-Mean-Square, 15, 338
Rotating magnetic field, 278
Rules, parallel circuits, 96
Rules, series circuits, 96
RVDT, 364

S, symbol for apparent power, 323
Saturable reactor, 243
Sawtooth wave, 11
Sawtooth wave , 167
Scalar number, 27, 323
SCR, 153
Secondary coil, 204
Secondary transformer coil, 4
Selectivity, 191
Self-inductance, 203
Selsyn, 365
Sequence, harmonic, 316
Sequence, phase, 272
Series analysis, 95
Series circuit rules, 96
Series LC resonance, 123
Series-parallel analysis, 103, 112
Shield grounding, 140
Shielded cable, 140
SHM, 118
Siemens, unit, 111
Silicon-controlled rectifier, 153
Simple Harmonic Motion, 118
Sine function, 6
Sine wave, 6
Single-phase, 259, 264
Sinusoidal, 11, 139
Skin effect, 70, 72, 133, 241, 411, 421
Sound waves, 9
Spectrum analyzer, 151, 353
Spectrum, frequency, 353
SPICE, 49
Split-phase, 263
Square wave, 11, 144
Standard, measurement, 344
Standing wave ratio, 409
Standing waves, 393
Star configuration, 270, 281
Superposition Theorem, 171, 267
Surge impedance, 384
Susceptance, 112
Switch, tap, 228
SWR, 409
Synchro, 365

Tank circuit, 118, 233

Tap switch, 228
TE mode, 421
TEM mode, 420
Tesla Coil, 241
Theorem, Maximum Power Transfer, 236
Theorem, Norton's, 236
Theorem, Pythagorean, 39
Theorem, Superposition, 171, 267
Theorem, Thevenin's, 236
Thevenin's Theorem, 236
Three-phase, 265, 268
Three-wire DC system, 265
Timbre, 143
Time-domain reflectometer, 385
TM mode, 421
Transducer, 363
Transductor, 245
Transform, Fourier, 149
Transformer, 3, 207
Transformer coils, primary and secondary, 4
Transformer core, laminated, 248
Transformer core, powdered iron, 248
Transformer inrush current, 252
Transformer isolation, 220
Transformer ratio, 216
Transformer, ferroresonant, 232
Transformer, peaking, 250
Transformer, step-down, 216
Transformer, step-up, 216
Transformer, variable, 228, 363
Transistor, 153
Transmission line, 376
Triangle wave, 11
Triangle wave , 166
Triangle, power, 325
Triplen harmonics, 303, 315
True-RMS meter, 16
Tube, vacuum, 226
Tuner circuit, radio, 127, 193
Twin-T circuit, differential capacitance, 369
Twisted pair cable, 140

Unit, amp-turn, 202
Unit, Celsius, 8
Unit, Centigrade, 8
Unit, CPS, 8
Unit, farad, 78
Unit, henry, 56
Unit, Hertz, 8
Unit, joule, 352
Unit, mho, 111
Unit, ohm, 112
Unit, siemens, 111
Unit, volt-amp, 246, 323
Unit, volt-amp-reactive, 323
Unit, watt, 323

v, symbol for instantaneous voltage, 56, 78
VA, unit, 323
Vacuum tube, 226
Value, instantaneous, 55, 77
VAR, unit, 323
Variable capacitor, 367
Variable transformer, 228, 363
Variac, 230
Vector, 27, 323
Vector amplitude, 30
Vector angle, 35
Vector length, 35
Vector magnitude, 35
Vector modulus, 35
Vector phase shift, 30
Vector sum, 51
Vector, absolute value, 35
Velocity factor, transmission line, 382
Vibrating reed frequency meter, 342
Volt-amp, 246
Volt-amp, unit, 323
Volt-amp-reactive, unit, 323
Voltage polarity, AC, 42, 52, 262
Voltage regulation, 231
Voltage, common-mode, 222
Voltage, line, 282
Voltage, phase, 282
Voltmeter, 335

Wagner earth, 360
Watt, unit, 323
Wave, electromagnetic, 22
Wave, sawtooth, 11, 167
Wave, sine, 6
Wave, square, 11, 144

Wave, triangle, 11, 166
Waveform symmetry and harmonics, 168, 293
Waveform, nonsinusoidal, 139
Waveform, sinusoidal, 139
Waveguide, 419
Wavelength, 390
Weston meter movement, 16
Wheatstone bridge, 344
Wien bridge circuit, 358
Winding, primary, 209
Winding, secondary, 209
Wire, Litz , 71

X, symbol for reactance, 58, 61, 80, 83, 112
Xtal, 344

Y configuration, 270, 281
Y, symbol for admittance, 112

Z, symbol for impedance, 61, 82, 83, 112
Zero sequence, 316