TELECOMMUNICATIONS



Chris C.J. Lee
Fontys University of Applied Sciences



Telecommunications

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 05/15/2026

TABLE OF CONTENTS

Licensing

1: Passive Components

- 1.1: Introduction
- 1.2: Passive components at high frequency
- 1.3: Surface-Mount Components
- 1.4: Summary

2: Transmission Lines

- 2.1: Introduction
- 2.2: Transmission Line Characteristics
- 2.3: The Lossless Terminated Line
- 2.4: Summary
- 2.5: References
- 2.6: Exercises

3: Impedance Matching

- 3.1: Introduction
- 3.2: Matching Networks
- 3.3: Impedance Transforming Networks
- 3.4: The L Matching Network
- 3.5: Dealing with Complex Loads
- 3.6: Dealing with Complex Loads
- 3.7: Summary
- 3.8: References
- 3.9: Exercises
- 3.10: Impedance Matching Using Smith Charts
- 3.11: Distributed Matching
- 3.12: Matching Using the Smith Chart

4: Filters

- 4.1: Introduction
- 4.2: Singly and Doubly Terminated Networks
- 4.3: The Lowpass Filter Prototype
- 4.4: The Maximally Flat (Butterworth) Lowpass Approximation
- 4.5: The Chebyshev Lowpass Approximation
- 4.6: Element Extraction
- 4.7: Butterworth and Chebyshev Filters
- 4.8: Impedance and Admittance Inverters
- 4.9: Filter Transformations
- 4.10: Cascaded Line Realization of Filters
- 4.11: Butterworth and Chebyshev Bandpass Filters
- 4.12: Richard's Transformation
- 4.13: Kuroda's and Norton's Network Identities

- 4.14: Inter-resonator Coupled Bandpass Filters
- 4.15: Bandpass Filter Topologies
- 4.16: Case Study- Design of a Bandstop Filter
- 4.17: Active Filters
- 4.18: Transient Response of a Bandpass Filter
- 4.19: Summary
- 4.20: References
- 4.21: Exercises

5: Oscillators

- 5.1: Introduction
- 5.2: Oscillator Theory
- 5.3: Reflection Oscillator
- 5.4: Case Study- Reflection Oscillator
- 5.5: Advanced Discussion of Oscillator Noise
- 5.6: Summary
- 5.7: References
- 5.8: Exercises

6: Antennas and the RF Link

- 6.1: Introduction
- 6.2: RF Antennas
- 6.3: Resonant Antennas
- 6.4: Traveling-Wave Antennas
- 6.5: Antenna Parameters
- 6.6: The RF Link
- 6.7: Summary
- 6.8: References
- 6.9: Exercises
- 6.10: Multipaths and Delay Spread
- 6.11: 4.8 Radio Link Interference
- 6.12: Antenna Array

Index

Glossary

Detailed Licensing

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Passive Components

1.1: Introduction

1.2: Passive components at high frequency

1.3: Surface-Mount Components

1.4: Summary

This page titled 1: Passive Components is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

1.1: Introduction

This chapter examines the high frequency behaviour of some passive components that are standard in low frequency electronics.

This page titled 1.1: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

1.2: Passive components at high frequency

RF inductors and capacitors also have loss and parasitic elements. With inductors there is both series resistance and shunt capacitance mainly from interwinding capacitance, while with capacitors there will be shunt resistance and series inductance. Even the pads and leads of components change the behavior of components. In this section, some examples of the effects of non-ideal behavior will be shown. It will also be possible to play with those examples to see how changing the parasitic elements effects performance. Then the concept of Q will be introduced as a way to quantify the performance of components.

1.2.1: AC resistance

At high frequencies, the current is limited to a small layer near the surface by the skin effect, given by:

$$\delta = \sqrt{\frac{2\rho}{\omega\mu_r\mu_0}} \tag{1.2.1}$$

,

where ρ is the resistivity, ω is the angular frequency, μ_r is the relative magnetic permeability (unity for most materials), and μ_0 is the magnetic permeability of free space. For frequencies in the GHz range (standard WiFi or bluetooth), the skin depth is already around ~10 µm. The resistance of a conductor of length L and cross section A is given by

$$R = \rho\frac{L}{A} \tag{1.2.2}$$

But, even if the conductor is very thick (a few mm in the case of a rectangular waveguide), the majority of the current flows within the skin depth. Thus, A must be calculated using the skin depth rather than the thickness.

AC resistance calculation 1.2.1

Consider a 1 mm diameter copper conductor of length 1 m. What is the DC resistance and the resistance at 10 GHz?

Solution

Copper has a conductivity of 1.7×10^{-8} Ωm. The area of the wire is $A = \pi r^2 = 7.9 \times 10^{-7}$ m^2, so the DC resistance is 0.02 Ω.

At 10 GHz, the skin depth is $\delta = \sqrt{\frac{2\rho}{\omega\mu_r\mu_0}} = 930$ nm. So the area in which the current flows is $A = \pi(\delta^2 + 2r\delta) = 2.9 \times 10^{-9}$ m^2. Thus, the resistance at 10 GHz is 5.8 Ω, nearly 270 times more than the DC resistance.

1.2.2: Lead inductance

Modern high frequency electronics make use of surface mount components to minimize the effect of wire inductance. The inductance due to a wire is given by

$$L = 2l\left[2.3\log\left(\frac{4l}{d}\right) - 0.75\right] \tag{1.2.3}$$

.

where l and d are the length and diameter of the wire. For those interested, this result is derived on pages 302-305 Vol 4, Num. 2 of the Bulletin of the Bureau of Standards, "The self and mutual inductances of linear conductors," Edward B. Rosa. Equation 1.2.3 provides an excellent illustration of why through hole components are so unfavorable for high frequency designs. This is shown in the figure 1.2.1, where the reactance of an ideal 1 pF capacitor with 1 cm leads is plotted.

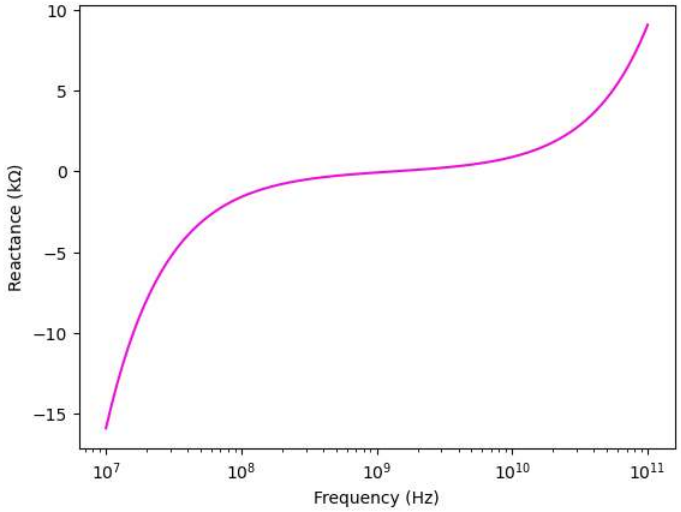


Figure 1.2.1: Reactance of an ideal capacitor with 1 cm leads. Note that the behavior is only capacitive for values below about 200 MHz.

Notice that the reactance is negative, as expected up to a few hundred MHz, but for higher frequencies, the capacitor behaves like an inductor. Hence, when modeling high frequency circuits the parasitic reactances and resistances cannot automatically be neglected.

1.2.3: Introduction to Q

As illustrated in Figure 1.2.1, the reactance of component is only close to the ideal value *below* the (self) resonance frequency of the LC circuit created between the ideal capacitor and the parasitic inductance. This is generally true, so practical inductor or capacitor is limited to operation below the self-resonant frequency determined by the inductance and capacitance itself resonating with its reactive parasitics. The impact of loss is quantified by the Q factor (the quality factor). Q is loosely related to bandwidth in general and the strict relationship is based on the response of a series or parallel connection of a resistor (R), an inductor (L), and a capacitor (C). The response of an RLC network is described by a second-order differential equation with the conclusion that the 3 dB fractional bandwidth of the response (i.e., when the power response is at its half-power level below its peak response) is $1/Q$.

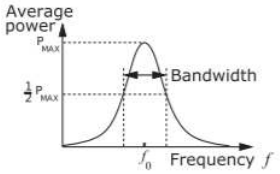


Figure 1.2.2: Transfer characteristic of a resonant circuit. (The transfer function is V/I for the parallel resonant circuit of Figure 1.2.3(a) and I/V for the series resonant circuit of Figure 1.2.3(b).)

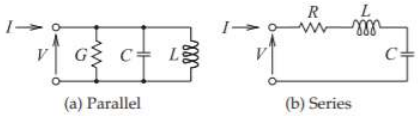


Figure 1.2.3: Second-order resonant circuits.

The fractional bandwidth is$\Delta f/f_0$ where $f_0 = f_r$ is the resonant frequency at the center of the band and Δf is the 3 dB bandwidth. The described relationship between Q and bandwidth is not true for any network other than a second-order circuit, but as a guiding principle, networks with higher Qs will have narrower bandwidths.

1.2.4: Definition of Q

The Q factor of a component at frequency f is defined as the ratio of $2\pi f$ times the maximum energy stored to the energy lost per cycle. In a lumped-element resonant circuit, stored energy is transferred between an inductor, which stores magnetic energy, and a

capacitor, which stores electric energy, and back again every period. Distributed resonators function the same way, exchanging energy stored in electric and magnetic forms, but with the energy stored spatially. The quality factor is

$$Q = 2\pi f_0 \left(\frac{\text{average energy stored in the resonator at } f_0}{\text{power lost in the resonator}} \right) \tag{1.2.4}$$

where f_0 is the resonant frequency.

A simple response is shown in Figure 1.2.2 for a parallel resonant circuit with elements L, C, and $G = 1/R$ (see Figure 1.2.3(a)),

$$Q = \omega_r C/G = 1/(\omega_r LG) \tag{1.2.5}$$

where $f_r = \omega_r/(2\pi)$ is the resonant frequency and is the frequency at which the maximum amount of energy is stored in a resonator. The conductance, G, describes the energy lost in a cycle. For a series resonant circuit (Figure 1.2.3(b)) with L, C, and R elements,

$$Q = \omega_r L/R = 1/(\omega_r CR) \tag{1.2.6}$$

These second-order resonant circuits have a bandpass transfer characteristic (see Figure 1.2.2) with Q being the inverse of the fractional bandwidth of the resonator. The fractional bandwidth, $\Delta f/f$, is measured at the half-power points as shown in Figure 1.2.2. (Δf is also referred to as the two-sided -3 dB and sometimes as the fullwidth half maximum (FWHM) bandwidth

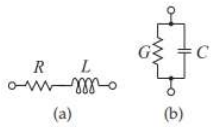


Figure 1.2.4: Loss elements of practical inductors and capacitors: (a) an inductor has a series resistance R; and (b) for a capacitor, the dominant loss mechanism is a shunt conductance $G = 1/R$.

bandwith.) Then

$$Q = f_r/\Delta f \tag{1.2.7}$$

Thus the Q is a measure of the sharpness of the bandpass frequency response. The determination of Q using the measurement of bandwidth together with Equation (1.2.7) is often not very precise, so another definition that uses the much more sensitive phase change at resonance is preferred when measurements are being used. With ϕ being the phase (in radians) of the transfer characteristic, the definition of Q is now

$$Q = \frac{\omega_r}{2} \left| \frac{d\phi}{d\omega} \right| \tag{1.2.8}$$

Equation (1.2.8) is another equivalent definition of Q for parallel RLC or series RLC resonant circuits. It is meaningful to talk about the Q of circuits other than three-element RLC circuits, and then its meaning is always a ratio of the energy stored to the energy dissipated per cycle. The Q of these structures can no longer be determined by bandwidth or by the rate of phase change.

1.2.5: Q of Lumped Elements

Q is also used to characterize the loss of lumped inductors and capacitors. Inductors have a series resistance R, and the main loss mechanism of a capacitor is a shunt conductance G (see Figure 1.2.3).

The Q of an inductor at frequency $f = \omega/(2\pi)$ with a series resistance R and inductance L is

$$Q_{\mathrm{I}} = \frac{\omega L}{R} \tag{1.2.9}$$

Since R is approximately constant with respect to frequency for an inductor, the Q will vary with frequency.

The Q of a capacitor with a shunt conductance G and capacitance C is

$$Q_{\mathrm{C}} = \frac{\omega C}{G} \tag{1.2.10}$$

G is due mainly to relaxation loss mechanisms of the dielectric of a capacitor and so varies linearly with frequency but also it is usually very small. Thus the Q of a capacitor is almost constant with respect to frequency. For microwave components $Q_C \gg Q_I$, and both are smaller than the Q of most transmission line networks. Thus, if the length of a transmission line is not too long for an application, transmission line networks are preferred. If lumped elements must be used, the use of inductors should be minimized.

1.2.6: Loaded Q Factor

The Q of a component as defined in the previous section is called the unloaded Q, Q_U. However, if a component is to be measured or used in any way, it is necessary to couple energy in and out of it. The Q is reduced and thus the resonator bandwidth is increased by the power lost to the external circuit so that the loaded Q, the Q that is measured, is

$$\begin{aligned} Q_L &= 2\pi f_0 \left(\frac{\text{average energy stored in the resonator at } f_0}{\text{power lost in the resonator and to the external circuit}} \right) \\ &= \frac{1}{1/Q + 1/Q_X} \end{aligned} \tag{1.2.11}$$

and

$$Q_X = \left(\frac{1}{Q_L} - \frac{1}{Q_U} \right)^{-1} \tag{1.2.12}$$

where Q_X is called the external Q. Q_L accounts for the power extracted from the resonant circuit. If the loading is kept very small, $Q_L \approx Q_U$.

1.2.7: Summary of the Properties of Q

In summary:

a. Q is properly defined and related to the energy stored in a resonator for a second-order network, one with two reactive elements of opposite types.
b. Q is not well defined for networks with three or more reactive elements.
c. It is only used (as defined or some approximation of it) for guiding the design.

1.2.8: Some example component equivalent circuits

In the following, example equivalent circuits are given for resistors, inductors and capacitors. Not every example of an equivalent circuit focuses on the Q, but, from the circuit diagrams it is clear that all of these circuits are damped oscillators.

An equivalent circuit for a resistor is shown in Figure 1.2.5.

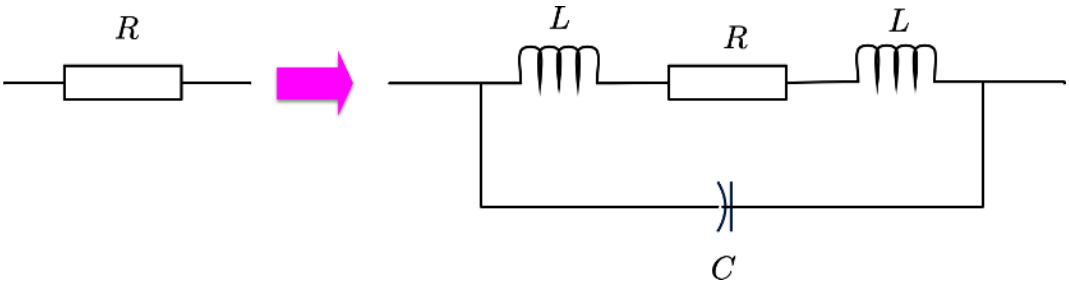


Figure 1.2.5: A non-ideal resistor is replaced by a series of ideal inductances and capacitances.

? Exercise 1.2.1

The equivalent circuit shown in Figure 1.2.5 is modeled in the code below, with a plot to show the real and imaginary components of the effective impedance. The values for the parasitic inductance and capacitance are set quite high so that the effect is easy to see. These values are free to be modified. What range of parasitic inductance and capacitance are allowed for a resistor to function at 2.4 GHz and 5 GHz (WiFi carrier frequencies)?

Answer

First adjust the model range to cover 1 - 10 GHz. You should find that the inductance can be quite high (~1 nF), but the parasitic capacitance needs to be around 1 fF.

The equivalent circuit for a capacitor is shown in Figure 1.2.6.

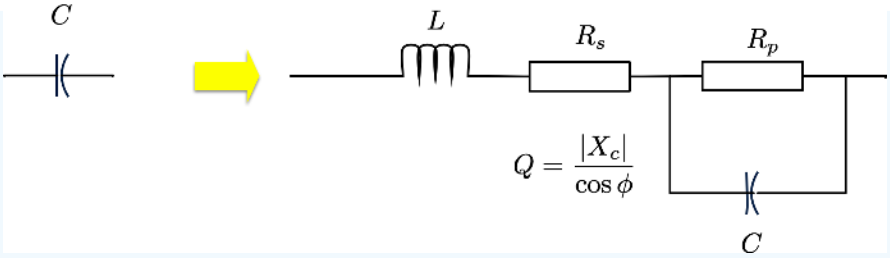


Figure 1.2.6: A non-ideal capacitor has both an inductive and a resistive component. Note that for an ideal capacitor, the power factor $cos\phi = 0$, leading to an infinite Q.

Again, the example in Figure 1.2.6 is modeled in the code below. In the current generation of automotive radar, the transmitters and receivers operate between 77-81 GHz. Use the code below to try and estimate the maximum allowed values for the parasitic inductors and resistors.

Answer

Adjust the simulation range to the range of interest (you can use non-integer values in the logspace function. For instance, a starting value of 10.5 means a starting frequency of ~30 GHz). You should find that the impedance is relatively insensitive to the two resistance values (but the Q is!), while the parasitic inductance should be lower than 0.1 pH and preferably about 0.01 pH.

An equivalent circuit for an inductor is shown in Figure 1.2.7. An ideal inductor, L, has a parallel parasitic capacitance, C_d, and the wire windings have some resistance R_s. At low frequency, the inductor behaviour dominates as it has the lowest reactance, while at high frequency, the capacitor dominates. Thus, an inductor will show capacitive behavior at high frequencies.

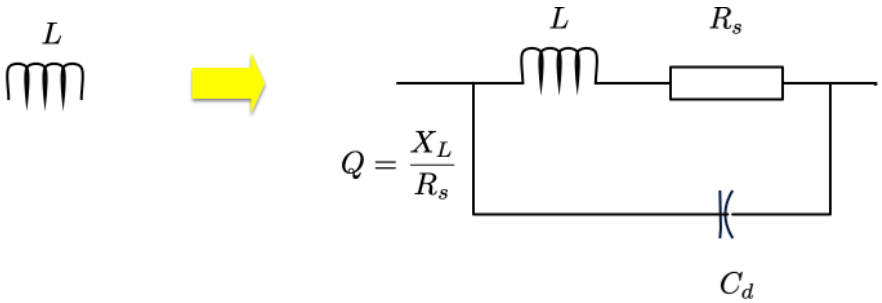


Figure 1.2.7: A non-ideal inductor has both a capacitive and a resistive component. Note that for an ideal inductor $R_s = 0$, leading to an infinite Q.

? Exercise 1.2.3

Below is a model that calculates the impedance of a non-ideal inductor using the equivalent circuit in Figure 1.2.7. A high frequency inductor (Vishay IFCB-402) has the following properties:

L @ 500 MHz (nH)	Self resonance frequency (MHz)	Max DC resistance (Ω)	Q @ 1700 MHz	Q @ 800 MHz	Q @ 100 MHz
1.0	12000	0.15	33	21	7

Use the script below to estimate the parasitic capacitance. Is the tabulated DC resistance a good estimate? If not, what value best fits the rest of the data? The inductance has a tolerance of 0.3 nH. Does adjusting the inductance within tolerance give you a better fit to the tabulated values? What is the maximum frequency at which you would use this inductor? Note, you may want to change the frequency range to get a better estimate.

Answer

You should adjust the DC resistance and inductance so that tabulated and calculated \(Q)\ values are as close to each other as possible. You may have to change the frequency scale so that this is clear to see. Adjust the frequency range to cover only the self resonance frequency and adjust the parasitic capacitance until the self resonance frequency occurs near 12000 MHz (you don't have to be exact). This indicates that the DC resistance is not a good estimate at high frequency. The AC resistance is over twice the DC resistance at 1200 MHz, while at 100 MHz, the AC resistance is insignificant. You should find that the parasitic inductance must be about 0.18 pF. The ideal and calculated inductance start to deviate from each other significantly at 3 GHz.

This page titled 1.2: Passive components at high frequency is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **9.2: Q Factor** by Michael Steer is licensed CC BY-NC 4.0.

1.3: Surface-Mount Components

The majority of the RF and microwave design effort goes into developing modules and interconnecting modules on circuit boards. With these the most common type of component to use is a surface-mount component. Figure 1.3.1(a) shows a two-terminal element, such as a resistor or capacitor, in the form of a surface-mount component. Figure 1.3.1(b and c) show the use of surface-mount components on a microwave circuit board.

A two-terminal surface-mount resistor or capacitor is commonly called a chip resistor or chip capacitor. These can be very small, and the smaller the component often the higher the operating frequency due to reduced parasitic capacitance or inductance. Common sizes of two-terminal chip components are listed in Table 1.3.1. The resonance when a chip capacitor (inductor) resonates with the parasitic inductance (capacitance) is called self-resonance and the resonant frequency is called the self-resonance frequency. Figure 1.3.2(a) shows an inductor in a surface-mount package and details are shown in Figure 1.3.2(b) and the inductor is useable as an inductor at a frequency backed-off from the SRF, see Table 9.4.1.

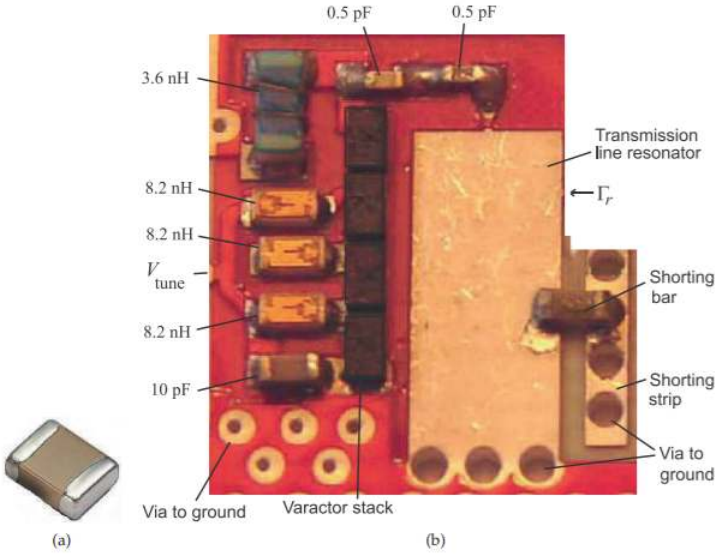


Figure 1.3.1: Circuit board showing the use of surface-mount components: (a) chip resistor or capacitor with metal terminals at the two ends; and (b) a populated RF microstrip circuit board of a 5 GHz tunable resonator [1]. The larger components have dimensions $1.6\ \text{mm} \times 0.8\ \text{mm}$. The shorting bar is a $0\ \Omega$ chip resistor.

Metric Designation	Size (mm × mm)
0402	0.4 × 0.2
0603	0.6 × 0.3
1005	1.0 × 0.5
1608	1.6 × 0.8
2012	2.0 × 1.25
2520	2.5 × 2.0
3216	3.2 × 1.6
3225	3.2 × 2.5
4516	4.5 × 1.6
4532	4.5 × 3.2
5025	5.0 × 2.5
6432	6.4 × 3.2
–	7.4 × 5.1

Table 1.3.1: Sizes and designation of two-terminal surface mount components. Note the designation of a surface-mount component refers to its dimensions in millimeters.

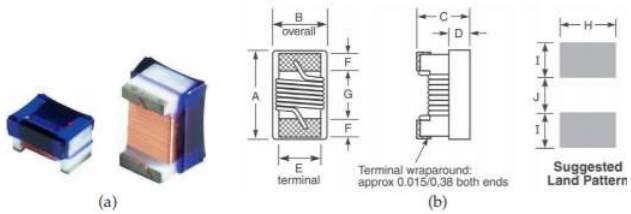


Figure 1.3.2: : Chip inductor: (a) in an 0603 surface-mount package; and (b) schematic showing sizes and pads to be provided on a circuit board ($A = 1.63$ mm,$B = 0.84$ mm, $C = 0.61$ mm, $D = 0.33$ mm, $E = 0.76$ mm, $F = 0.64$ mm, $G = 0.64$ mm, and $H = 1.02$ mm). Copyright Coilcraft, Inc., used with permission [2].

This page titled 1.3: Surface-Mount Components is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **9.3: Surface-Mount Components** by Michael Steer is licensed CC BY-NC 4.0.

1.4: Summary

This chapter introduced the richness of microwave components available to the RF designer. Microwave lumped-element R, L, and C components are carefully constructed so that they function as intended up to 10 or so. They are usually in surface-mount form so that they can be integrated in design while minimizing the parasitic effects introduced by leads.

This page titled 1.4: Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **9.11: Summary** by Michael Steer is licensed CC BY-NC 4.0.

CHAPTER OVERVIEW

2: Transmission Lines

2.1: Introduction
2.2: Transmission Line Characteristics
2.3: The Lossless Terminated Line
2.4: Summary
2.5: References
2.6: Exercises

This page titled 2: Transmission Lines is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

2.1: Introduction

A transmission line stores electric and magnetic energy. As such a line has a circuit form that combines inductors, L_s (for the magnetic energy), capacitors, C_s (for the electric energy), and resistors, R_s (modeling losses), whose values depend on the line geometry and material properties.

The transmission lines considered in this chapter are restricted to just two parallel conductors, as shown in Figure 2.1.1, with the distance between the two wires (i.e., in the transverse direction) being substantially smaller than the wavelengths of the signals on the line. The correct physical interpretation is that the conductors of a transmission line confine and guide an EM field. The EM field contains the energy of the signal and not the current on the line. However, with electrically small transverse dimensions, a two conductor line

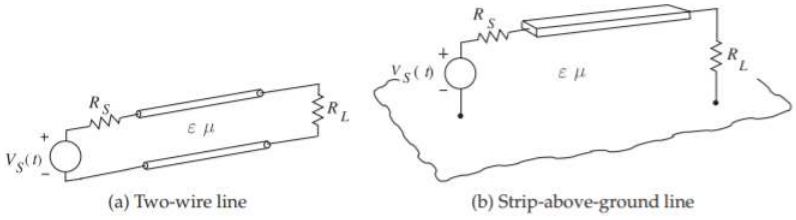


Figure 2.1.1: Two conductor transmission lines.

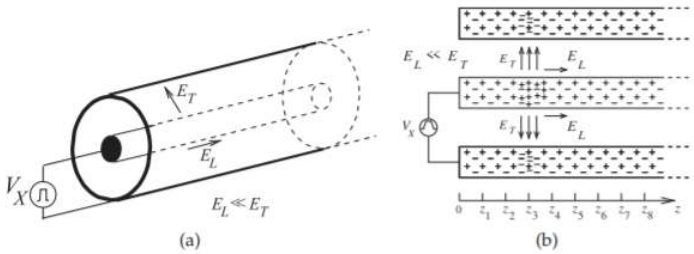


Figure 2.1.2: A coaxial transmission line: (a) three-dimensional view; (b) the line with pulsed voltage source showing the electric fields at an instant in time as a voltage pulse travels down the line.

may be satisfactorily analyzed on the basis of voltages and currents.

Frequency-domain analysis is the best way to understand transmission lines. This transmission line theory with modern developments is presented in Section 3.2 and useful formulas and concepts are developed in Section 3.3 for lossless transmission lines. Section 3.4 presents several configurations of lossless lines that are particularly useful in microwave circuit design and used in many places in this book series.

2.1.1: When Must a Line be Considered a Transmission Line

The key determinant of whether a transmission line can be considered as an invisible connection between two points is whether the signal anywhere along the interconnect has the same value at a particular instant. If the value of the signal (say, voltage) varies along the line (at an instant), then it may be necessary to consider transmission line effects. A typical criterion used is that if the length of the interconnect is less than $1/20\text{th}$ of the wavelength of the highest-frequency component of a signal, then transmission line effects can be safely ignored and the circuit can be modeled as a single RLC circuit [1].

2.1.2: Movement of a Signal on a Transmission Line

When a positive voltage pulse is applied to the center conductor of a coaxial line, as shown in Figure 2.1.2(a), an electric field results that is directed from the center conductor to the outer conductor. Referring to Figure 2.1.2(b), the component of the field that is directed along the shortest path from the center conductor to the outer conductor is denoted E_T, and the subscript T denotes the transverse component of the field as shown. Figure 2.1.2(b) shows the fields in the structure after the pulse has started moving along the line. This is shown in another view in Figure 2.1.3 at four different times. The transverse voltage, V_T, is given by E_T integrated along a path between the inner and outer conductors: $V_T \approx E_T(a - b)$. This is a good measure, provided that the

transverse dimensions are small compared to a wavelength. The voltage pulse exciting the line has a trapezoidal shape and the voltage between the inner and outer conductor has the same form as E_T. As indicated by Maxwell's equations, a change in time of the electric field results in a spatial

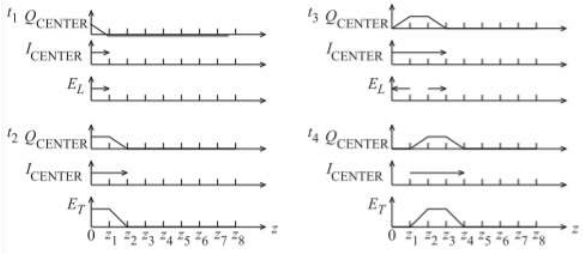


Figure 2.1.3: Fields, currents, and charges on the coaxial transmission line of Figure 2.1.2 at times $t_4 > t_3 > t_2 > t_1$. Q_{CENTER} is the net free charge on the center conductor. I_{CENTER} is the current on the center conductor.

change in the magnetic field and hence current. As a result there is a variation of the current in time and this results in a spatial change of the electric field. The chasing from a time variation to a spatial variation and then back to a time variation causes the pulse to move down the line.

The pulse moves down the line at the **group velocity**, which for a lossless coaxial line is the same as the **phase velocity**, v_p[1]. This is determined by the physical properties of the region between the conductors. The permittivity, ε, describes energy storage associated with the electric field, E, and the energy storage associated with the magnetic field, H, is described by the permeability, μ. It has been determined that

$$v_p = 1/\sqrt{\mu\varepsilon} \tag{2.1.1}$$

In free space $v_p = c = 1/\sqrt{\mu_0\varepsilon_0} = 3 \times 10^8$ m/s. The free-space wavelengths, $\lambda_0 = c/f$, at several frequencies, f, are

f	100 MHz	1 GHz	10 GHz
λ_0	3 m	30 cm	3.0 cm

Table 2.1.1

Example 2.1.1: Transmission Line Wavelength

A length of coaxial line is filled with a dielectric having a relative permittivity, ε_r, of 20 and is designed to be $1/4$ wavelength long at a frequency, f, of 1.850 GHz. ($\mu_r = 1$)

a. What is the free-space wavelength?
b. What is the wavelength of the signal in the dielectric-filled coaxial line?
c. How long is the line?

Solution

a. $\lambda_0 = c/f = 3 \times 10^8/1.85 \times 10^9 = 0.162$ m $= 16.2$ cm
b. Note that for a dielectric-filled line with $\mu_r = 1$, $\lambda = v_p/f = c/(\sqrt{\varepsilon_r} f) = \lambda_0/\sqrt{\varepsilon_r}$, so $\lambda = \lambda_0/\sqrt{\varepsilon_r} = 16.2$ cm$/\sqrt{20} = 3.62$ cm
c. $\lambda_g/4 = 3.62$ cm$/4 = 9.05$ mm

2.1.3: Footnotes

[1] The phase velocity is the apparent velocity of a point of constant phase on a sinewave and is almost frequency independent for a low-loss coaxial line of small transverse dimensions (less than $\lambda/10$).

This page titled 2.1: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **3.1: Introduction** by Michael Steer is licensed CC BY-NC 4.0.

2.2: Transmission Line Characteristics

This section gives, without derivation the standard characteristics of transmission lines. The first section, Section 2.2.1, makes the argument that a circuit with resistors, inductors, and capacitors is a good model for a transmission line. The most important parameters that are derived from transmission line theory are given in Section 2.2.2. The dimensions of some of the quantities that appear in transmission line theory are discussed in Section 2.2.3. Section 2.2.4 summarizes the important parameters of a lossless line and then a particularly important line, the microstrip line, is considered in Section 2.2.5.

2.2.1: Transmission Line RLGC Model

Regardless of the actual structure, a segment of uniform transmission line (i.e., a line with constant cross section along its length) as shown in Figure 2.2.1(a) can be modeled by the circuit shown in Figure 2.2.1(b) with

Resistance along the line	$= R$	
Inductance along the line	$= L$	all specified
Conductance shunting the line	$= G$	per unit length
Capacitance shunting the line	$= C$	

$R,\ L,\ G$, and C are referred to as resistance, inductance, conductance, and capacitance per unit length. (Sometimes p.u.l. is used as shorthand for **per unit length**.) In the metric system, ohms per meter (Ω/m), henries per meter ($\mathrm{H/m}$), siemens per meter ($\mathrm{S/m}$) and farads per meter (F/m), respectively, are used. The values of $R,\ L,\ G$, and C are affected by the geometry of the transmission line and by the electrical properties of the dielectrics and conductors. C describes the ability to store electrical energy and is mostly due to the properties of the dielectric. G describes loss in the dielectric which derives from conduction in the dielectric and from dielectric relaxation. Most microwave substrates have negligible conductivity so dielectric loss dominates. Dielectric relaxation loss results from the movement of charge centers which result in distortion of the dielectric lattice (if a crystal) or molecular structure. The periodic variation of the E field transfers energy from the EM field to mechanical vibrations. R is due to ohmic loss in the metal more than anything else. L describes the ability to store magnetic energy and is mostly a function of geometry, as most materials used with transmission lines have $\mu_r = 1$ (so no more magnetic energy is stored than in a vacuum).

For most lines the effects due to L and C dominate because of the relatively low series resistance and shunt conductance. The propagation characteristics of the line are described by its loss-free, or lossless, equivalent line, although in practice some information about R or G is necessary to determine power losses. The lossless concept is just a useful and good approximation.

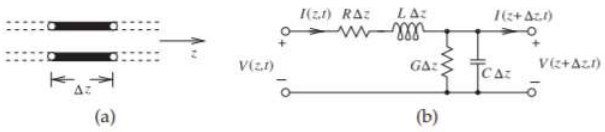


Figure 2.2.1: Transmission line segment: (a) of length Δz; and (b) lumped element model.

2.2.2: Transmission Line Properties

The derivation of transmission line properties can be found here. In this section, the properties are given without derivation and some of their consequences are explored.

The following quantities are defined:

$$\text{Characteristic impedance:} \quad Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{2.2.1}$$

$$\text{Propagation constant:} \quad \gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{2.2.2}$$

$$\text{Attenuation constant:} \quad \alpha = \Re\{\gamma\} \tag{2.2.3}$$

$$\text{Phase constant:} \quad \beta = \Im\{\gamma\} \tag{2.2.4}$$

$$\text{Wavenumber:} \quad k = -j\gamma \tag{2.2.5}$$

$$\text{Phase velocity:} \quad v_p = \omega/\beta \tag{2.2.6}$$

$$\text{Wavelength:} \quad \lambda = \frac{2\pi}{|\gamma|} = \frac{2\pi}{|k|} \tag{2.2.7}$$

where $\omega = 2\pi f$ is the radian frequency and f is the frequency with the SI units of hertz (Hz). The wavenumber k as defined here is used in electromagnetics and where wave propagation is concerned. Considering one of the traveling waves, the phase velocity refers to the apparent velocity of which a point of constant phase on the sinewave appears to move.

The important result here is that a voltage wave (and a current wave) can be defined on a transmission line. One more parameter needs to be introduced: the group velocity,

Note

The symbol ∂ means partial derivative (the derivative of a multi-variable function with respect to just one of the variables).

$$v_g = \frac{\partial \omega}{\partial \beta} \tag{2.2.8}$$

The group velocity is the velocity of a modulated waveform's envelope and describes how fast information propagates. It is the velocity at which the energy (i.e. information) in the waveform moves. Thus group velocity can never be more than the speed of light in a vacuum, c. Phase velocity, however, can be more than c. If the speed at which information moves varies with frequency, then a signal such as a pulse will spread out. Such a line is said to have dispersion. For a lossless, dispersionless line, the group and phase velocity are the same. If the phase velocity is frequency independent, then β is linearly proportional to ω.

Electrical length is used in designs prior to establishing the physical length of a line. The electrical length is expressed either as a fraction of a wavelength or in degrees (or radians), where a wavelength corresponds to 360° (or 2π radians). If ℓ is its physical length, the electrical length of the line in radians is $\beta\ell$.

Example 2.2.1: Physical and Electrical Length

A transmission line is 10 cm long and at the operating frequency the phase constant β is 30 rad/m. What is the electrical length of the line?

Solution

The physical length of the line is $\ell = 10 \text{ cm} = 0.1 \text{ m}$. Then the electrical length of the line is $\ell_e = \beta\ell = (30 \text{ rad/m}) \times 0.1 \text{ m} = 3 \text{ radians}$. The electrical length can also be expressed in terms of wavelength noting that 360° corresponds to 2π radians, which also corresponds to λ. Thus $\ell_e = (3 \text{ radians}) = 3 \times 360/(2\pi) = 171.9^\circ$ or $\ell_e = 3/(2\pi)\lambda = 0.477\lambda$.

Example 2.2.2: $RLGC$ Parameters

A transmission line has the $RLGC$ parameters $R = 100\ \Omega/\text{m}$, $L = 80$ nH/m, $G = 1.6$ S/m, and $C = 200$ pF/m. Consider a traveling wave at 2 GHz on the line.

a. What is the attenuation constant?
b. What is the phase constant?
c. What is the phase velocity?
d. What is the characteristic impedance of the line?
e. What is the group velocity?

Solution

a. $\alpha : \gamma = \alpha + \jmath\beta = \sqrt{(R + \jmath\omega L)(G + \jmath\omega C)}; \quad \omega = 12.57 \cdot 10^9 \text{ rad/s}$

$$\gamma = \sqrt{(100 + \jmath\omega \cdot 80 \cdot 10^{-9})(1.6 + \jmath\omega 200 \times 10^{-12})} = (17.94 + \jmath 51.85)\text{m}^{-1}$$

$$\alpha = \Re\{\gamma\} = 17.94 \text{ Np/m}$$

b. Phase constant: $\beta = \Im\{\gamma\} = 51.85 \text{ rad/m}$
c. Phase velocity:

$$v_p = \frac{\omega}{\beta} = \frac{2\pi f}{\beta} = \frac{12.57 \times 10^9 \text{rad} \cdot \text{s}^{-1}}{51.85 \text{rad} \cdot \text{m}^{-1}} = 2.42 \times 10^8 \text{ m/s}$$

d. $Z_0 = (R + \jmath\omega L)/\gamma = (100 + \jmath\omega \cdot 80 \cdot 10^{-9})/(17.94 + \jmath 51.85) = (17.9 + \jmath 4.3)\Omega$
Note also that $Z_0 = \sqrt{(R + \jmath\omega L)/(G + \jmath\omega C)}$, which yields the same answer.

e. Group velocity:

$$v_g = \frac{\partial\omega}{\partial\beta}|_{f=2 \text{ GHz}}$$

Numerical derivatives will be used, thus $v_g = \Delta\omega/\Delta\beta$. Now β is already known at 2 GHz. At 1.9 GHz, $\gamma = 17.884 + \jmath 49.397 \text{ m}^{-1}$, and so $\beta = 49.397$ rad/m.

$$v_g = \frac{2\pi(2 \text{ GHz} - 1.9 \text{ GHz})}{\beta(2 \text{ GHz}) - \beta(1.9 \text{ GHz})} = \frac{2\pi(2 - 1.9)10^9 \text{ Hz}}{(51.85 - 49.397) \text{ m}^{-1}} = 2.563 \times 10^8 \text{ m/s}$$

(Note that $\text{Hz} = \text{s}^{-1}$. Note that $v_g \neq= v_p$, and so the transmission line has dispersion.)

2.2.3: Dimensions of γ, α, and β

The SI unit of γ are inverse meters (m^{-1}) and the attenuation constant, α, and the phase constant, β, have, strictly speaking, the same units. However, the convention is to introduce the dimensionless quantities Neper and radian to convey additional information. Thus the attenuation constant α has the units of Nepers per meter (Np/m) and the phase constant β has the units radians per meter (rad/m). The unit Neper comes from the name of the $e(= 2.7182818284590452354\ldots)$ symbol (written in upright font and not italics since it is a constant), which is called the Neper.

Note

The name for e derives from John **Napier**, who developed the theory of logarithms [2]. e is sometimes called Euler's constant

The Neper is used in calculating transmission line signal levels, $V(z) = V_0^+ \exp(-\gamma z) + V_0^+ \exp(\gamma z)$, where V^+ and V^- indicate contributions from waves traveling from source to load and load to source. The attenuation and phase constants are often separated and then the attenuation constant describes the decrease in signal amplitude as the signal travels down a transmission line. So when $\alpha\ell = 1$ Np, where ℓ is the length of the line, the signal has decreased to $1/e$ of its original value, and the power drops to $1/e^2$ of its original value. The decrease in signal level represents loss and the units of decibels per meter (dB/m) are used with $1 \text{ Np} = 20 \log e = 8.6858896381 \text{ dB}$. So expressing α as 1 Np/m is the same as saying that the attenuation loss is 8.6859 dB/m. To convert from dB to Np multiply by 0.1151. Thus $\alpha = x \text{ dB/m} = x \times 0.1151 \text{ Np/m}$.

Note

In engineering $\log x \equiv \log_{10} x$ and $\ln x \equiv \log_e x$.

Example 2.2.3: Transmission Line Characteristics

A line has an attenuation of 10 dB/m and a phase constant of 50 radians/m at 2 GHz.

a. What is the complex propagation constant of the transmission line?
b. If the capacitance of the line is 100 pF/m and the conductive loss is zero (i.e., $G = 0$), what is the characteristic impedance of the line?

Solution

a. $\alpha|_{\text{Np}} = 0.1151 \times \alpha|_{\text{dB}} = 0.1151 \times (10 \text{ dB/m}) = 1.151 \text{ Np/m}, \ \beta = 50 \text{ rad/m}$
Propagation constant, $\gamma = \alpha + \jmath\beta = (1.151 + \jmath 50) \text{ m}^{-1}$

b. $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$, and $Z_0 = \sqrt{(R + j\omega L)/(G + j\omega C)}$, therefore $Z_0 = \gamma/(G + j\omega C)$; $w = 2\pi \cdot 2 \times 10^9$ s^{-1}; $G = 0$; $C = 100 \times 10^{-12}$ F, so $Z_0 = 39.8 - j0.916\ \Omega$.

2.2.4: Lossless Transmission Line

If the conductor and dielectric are ideal (i.e., lossless), then R =0= G and the equations for the transmission line characteristics simplify. The transmission line parameters from Equations (2.2.1) and (2.2.2)-(2.2.7) are then

$$Z_0 = \sqrt{\frac{L}{C}} \tag{2.2.9}$$

$$\alpha = 0 \tag{2.2.10}$$

$$\beta = \omega\sqrt{LC} \tag{2.2.11}$$

$$v_p = 1/\sqrt{LC} \tag{2.2.12}$$

$$\lambda_g = \frac{2\pi}{\omega\sqrt{LC}} = \frac{v_p}{f} \tag{2.2.13}$$

2.2.5: Microstrip Line

A microstrip line is shown in Figure 2.2.1(a). This is a commonly used transmission line, as it can be cheaply fabricated using printed circuit board techniques. This line consists of a metal-backed substrate of relative permittivity ε_r on top of which is a metal strip. Above that is air.

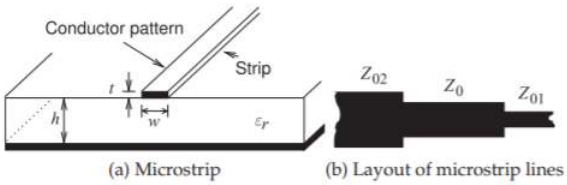


Figure 2.2.1: Microstrip transmission line. The layout (or top) view is commonly used with circuit designs using microstrip. This is the pattern of the strip where (b) shows three lines of different width.

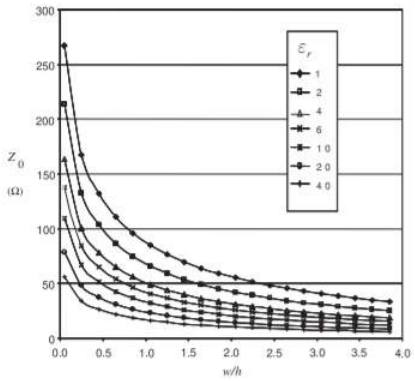


Figure 2.2.2: Dependence of Z_0 of a microstrip line at 1 GHz for various ε_r and aspect (w/h) ratios. Calculated using EM simulation.

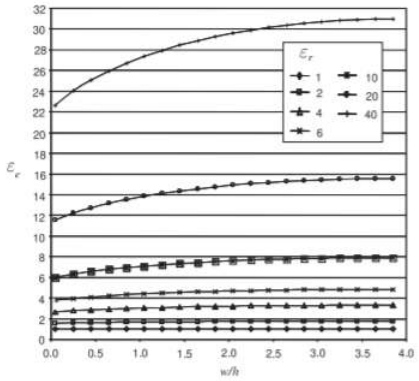


Figure 2.2.3: Dependence of effective relative permittivity ε_e of a microstrip line at 1 GHz for various permittivities and aspect ratios (w/h).

The width of the strip determines the characteristic impedance of the line, and can be approximately calculated by:

$$Z_0 \approx \frac{377}{\sqrt{\epsilon_r}\left(\frac{w}{h}+2\right)} \tag{2.2.14}$$

Note that Equation 2.2.14 is not accurate enough for use in high performance systems (see below), and more accurate analytical expressions have been found. However, for best performance, it is better to use a software tool to calculate the strip dimensions. The characteristic impedance of microstrip lines having various strip widths is shown in Figure 2.2.2 for several substrate permittivities. So the wider the strip and the higher the substrate permittivity, the lower the characteristic impedance of the line. The EM fields are partly in air and partly in the dielectric and an effective permittivity must be used when calculating the electrical length of the line. The results of field simulations of the effective permittivity of lines of various widths and with various substrate permittivities are shown in Figure 2.2.3, where it can be seen that the effective relative permittivity, ε_e, increases for wide strips. This is because more of the EM field is in the substrate. Microstrip transmission line structures are often drawn showing just the layout of the strip, as shown in Figure 2.2.1(b), where the three lines have different characteristic impedances.

2.2.6: Summary

The important takeaway from this section is that a signal moves on a transmission line as forward- and backward-traveling waves. The energy transferred is in the traveling waves. The total voltage and current at a point on the line is the sum of the traveling voltage and current waves, respectively, but the total voltage/current view is not sufficient to describe how a transmission line works. Transmission line theory is developed in terms of traveling voltages and current waves and these are akin to a one-dimensional form of Maxwell's equations.

This page titled 2.2: Transmission Line Characteristics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **3.2: Transmission Line Theory** by Michael Steer is licensed CC BY-NC 4.0.

2.3: The Lossless Terminated Line

Microwave engineers want to work with total voltage and current when possible and the art of design synthesis usually requires relating the total voltage and current world of a lumped element circuit to the traveling voltage world of transmission lines. This section introduces some important abstractions that enable the total voltage and current view of the world to be used with transmission lines. The first step is in Section 2.3.1 where total voltages and currents are related to forward- and backward-traveling voltages and currents. Insight into traveling waves and reflections is presented in Section 2.3.2. Important abstractions are presented first for the input reflection coefficient of a terminated lossless line in Section 2.3.3 and then for the input impedance of the line in Section 2.3.4. The last section, Section 2.3.5, presents a view of the total voltage on the transmission line and describes the voltage standing wave concept.

Nature is **self consistent**. The voltages and currents at any location on a transmission line at any instant in time can only have single values. If a transmission line is terminated with a load $Z_L \neq Z_0$, then to be self consistent and single valued, some portion of the incident wave must be reflected. In short, at a discontinuity in line impedance:

- Voltage is single valued $V = V_i + V_r$
- Current is single valued $I = I_i + I_r$
- Power is conserved $P_i = P_t + P_r + P_a$

where the subscripts i, r, a indicate incident, reflected and absorbed.

Before introducing the theory, it is helpful to examine the cases of a transmission line that is terminated by an open-circuit and a short-circuit. These provide an intuitive understanding for why reflections *must* exist.

Example 2.3.1

An open-circuited transmission line is shown in Figure 2.3.1. What are the amplitudes of the voltage and current at the load? What are the absorbed and reflected powers?

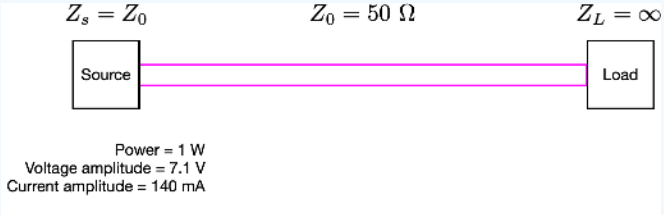


Figure 2.3.1: Transmission line with an open-circuit termination.

Solution

Since the terminating impedance is $Z_L = \infty$, the total current at the load must be zero. Hence, $I_r = -I_i =$ -140 mA. The absorbed power is 0, so the reflected power, $P_r = P_i =$ 1 W. For self consistency, the reflected voltage must be the same as the incident voltage ($V_r = V_i$), so the amplitude of the voltage at the anti-nodes along the line are $2V_i$

Example 2.3.2

An short-circuited transmission line is shown in Figure 2.3.2. What are the amplitudes of the voltage and current at the load? What are the absorbed and reflected powers?

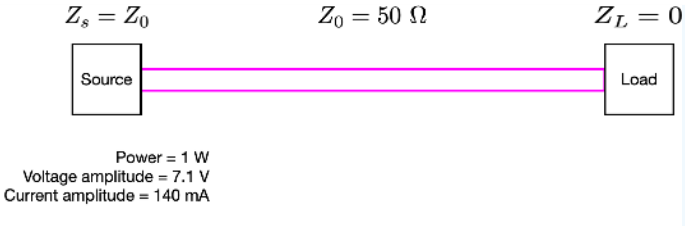


Figure 2.3.2: Transmission line with an short-circuit termination.

Solution

Since the terminating impedance is $Z_L = 0\ \Omega$, the total voltage at the load must be zero. Hence, $V_r = -V_i =$ -7.1 V. The absorbed power is 0, so the reflected power, $P_r = P_i =$ 1 W. The reflected current is equal to the incident current ($I_r = I_i = 140$ mA. Hence the current at the anti-nodes is $2I_i$.

2.3.1: Total Voltage and Current on the Line

Consider the terminated line shown in Figure 2.3.3(a). Assume an incident or forward-traveling wave, with traveling voltage $V_0^+ e^{-\jmath\beta z}$ and current $I_0^+ e^{-\jmath\beta z}$ propagating toward the load Z_L at $z = 0$. The characteristic impedance of the transmission line is the ratio of the voltage and current traveling waves so that

$$\frac{V_0^+(z)}{I_0^+(z)} = \frac{V_0^+ e^{-\jmath\beta z}}{I_0^+ e^{-\jmath\beta z}} = \frac{V_0^+(0)}{I_0^+(0)} = \frac{V_0^+}{I_0^+} = Z_0 \tag{2.3.1}$$

The load Z_L imposes an additional constraint on the relationship of the total voltage and current at $z = 0$:

$$\frac{V_L}{I_L} = \frac{V(z=0)}{I(z=0)} = Z_L \tag{2.3.2}$$

When $Z_L \neq Z_0$ equations 2.3.3 and 2.3.2 are inconsistent unless there is a reflected wave with appropriate amplitude.

The reflected wave, similar to the forward propagating wave, is given by (note the sign change):

$$\frac{V_0^- e^{\jmath\beta z}}{-I_0^- e^{\jmath\beta z}} = \frac{V_0^-}{-I_0^-} = Z_0 \tag{2.3.3}$$

Now the total voltage

$$V(z) = V_0^+ e^{-\jmath\beta z} + V_0^- e^{\jmath\beta z} \tag{2.3.4}$$

and the total current, $I(z)$, is related to the traveling current waves by

$$I(z) = \frac{V_0^+}{Z_0} e^{-\jmath\beta z} - \frac{V_0^-}{Z_0} e^{\jmath\beta z} = I_0^+ e^{-\jmath\beta z} + I_0^- e^{\jmath\beta z} \tag{2.3.5}$$

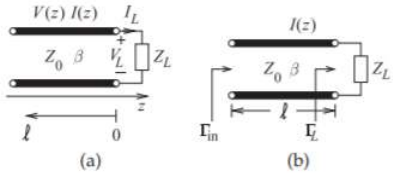


Figure 2.3.3: A terminated transmission line.

Thus at the termination of the line ($z = 0$),

$$\frac{V(0)}{I(0)} = Z_L = Z_0 \frac{V_0^+ + V_0^-}{V_0^+ - V_0^-}$$

This can be rearranged as the ratio of the reflected voltage to the incident voltage:

$$\frac{V_0^-}{V_0^+} = \frac{Z_L - Z_0}{Z_L + Z_0}$$

This ratio is defined as the voltage **reflection coefficient** at the load,

$$\Gamma_L = \Gamma_L^V = \frac{V_0^-(0)}{V_0^+(0)} = \frac{V_0^-}{V_0^+} = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.3.6}$$

That is, at the load

$$V_0^- = \Gamma_L V_0^+ \tag{2.3.7}$$

The relationship in Equation (2.3.6) can be rewritten so that the input load impedance can be obtained from the reflection coefficient:

$$Z_L = Z_0 \frac{1 + \Gamma^V}{1 - \Gamma^V} \tag{2.3.8}$$

Similarly, the current reflection coefficient can be written as

$$\Gamma^I = \frac{I_0^-}{I_0^+} = \frac{-Z_L + Z_0}{Z_L + Z_0} = -\Gamma^V \tag{2.3.9}$$

The voltage reflection coefficient is used most of the time, so the reflection coefficient, Γ, on its own refers to the voltage reflection coefficient, $\Gamma^V = \Gamma$.

Example 2.3.3: Forward- and Backward-Traveling Waves at an Open Circuit

A lossless transmission line is terminated in an open circuit. What is the relationship between the forward- and backward-traveling voltage waves at the end of the line?

Solution

At the end of the line the total current is zero, so that $I^+ + I^- = 0$ and so

$$I^- = -I^+ \tag{2.3.10}$$

The forward- and backward traveling voltages and currents are related to the characteristic impedance by

$$Z_0 = V^+/I^+ = -V^-/I^- \tag{2.3.11}$$

Note the change in sign, as a result of the direction of propagation changing but the positive reference for current is in the same direction. Substituting for I^- at the termination,

$$V^+ = -V^- I^+/I^- = -V^- I^+/(-I^+) = V^- \tag{2.3.12}$$

Thus the total voltage at the end of the line, V_{TOTAL}, is $V^+ + V^- = 2V^+$. Note that the total voltage at the end of the line is twice the incident (forward-traveling) voltage.

Example 2.3.4: Current Reflection Coefficient

A load consists of a shunt connection of a capacitor of 10 pF and a resistor of $60\ \Omega$. The load terminates a lossless $50\ \Omega$ transmission line. The operating frequency is 5 GHz.

a. What is the impedance of the load?
b. What is the normalized impedance of the load (normalized to Z_0 of the line)?
c. What is the reflection coefficient of the load?
d. What is the current reflection coefficient of the load?

Solution

a. $C = 10 \cdot 10^{-12}$ F, $R = 60\,\Omega$; $f = 5 \cdot 10^9$ Hz; $\omega = 2\pi f$; $Z_0 = 50\,\Omega$

$$Z_L = R||C = (1/R + j\omega C)^{-1} = 0.168 - j3.174\,\Omega$$

b. $z_L = Z_L/Z_0 = 3.368 \cdot 10^{-3} - j0.063$

c. This is the voltage reflection coefficient. $\Gamma_L = (z_L - 1)/(z_L + 1) = -0.985 - j0.126 = 0.993\angle 187.3^\circ$.
d. $\Gamma_L^I = -\Gamma_L = 0.985 + j0.126 = 0.993\angle(187.3 - 180)^\circ = 0.993\angle 7.3^\circ$.

2.3.2: Forward- and Backward-Traveling Pulses

Reflections at the end of a line produce a backward-traveling signal. Forward- and backward-traveling pulses are shown in Figure 2.3.4(a) for the situation where the resistance at the end of the line is lower than the characteristic impedance of the line ($Z_L < Z_0$).

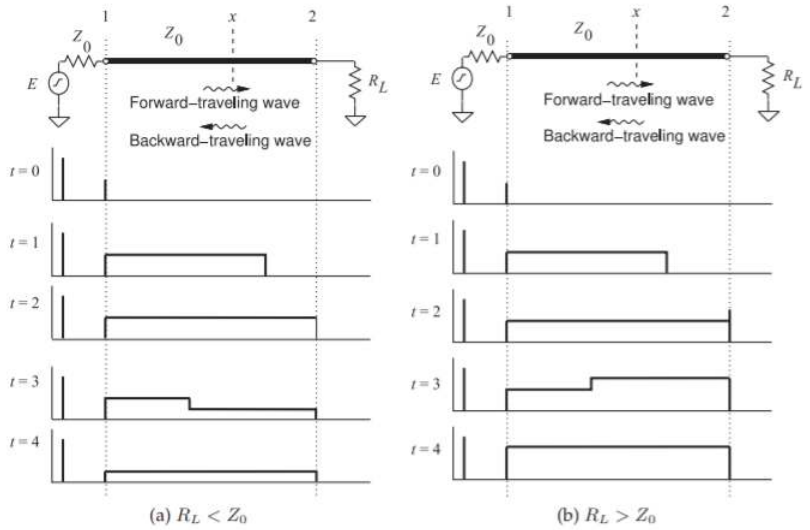


Figure 2.3.4: Reflection of a voltage pulse from a load: (a) when the resistance of the load, R_L is lower than the characteristic impedance of the line, Z_0; and (b) when R_L is greater than Z_0.

The voltage source is a step voltage that is zero for time $t < 0$. At time $t = 0$, the step is applied to the line and it begins traveling down the line, as shown at time $t = 1$. This voltage step moving from left to right is called the forward-traveling voltage wave.

At time $t = 2$, the leading edge of the step reaches the load, and as the load has lower resistance than the characteristic impedance of the line, the total voltage across the load drops below the level of the forward-traveling voltage step. The reflected wave is called the backward-traveling wave and it must be negative, as it adds to the forward-traveling wave to yield the total voltage. Thus the voltage reflection coefficient, Γ, is negative and the total voltage on the line, which is all that can be directly observed, drops. A reflected, smaller, and opposite step signal travels in the backward direction and adds to the forward-traveling step to produce the waveform shown at $t = 3$. The impedance of the source matches the transmission line impedance so that the reflection at the source is zero. The signal on the line at time $t = 4$, the time for round-trip propagation on the line, therefore remains at the lower value. The easiest way to remember the polarity of the reflected pulse is to consider the situation with a short-circuit at the load. Then the total voltage on the line at the load must be zero. The only way this can occur when a signal is incident is if the reflected signal is equal in magnitude but opposite in sign, in this case $\Gamma = -1$. So whenever $|Z_L| < |Z_0|$, the reflected pulse will tend to subtract from the incident pulse.

The opposite situation occurs when the resistance at the load is higher than the characteristic impedance of the line (Figure 3-9(b)). In this case the reflected pulse has the same polarity as the incident signal. Again, to remember this, think of the open-circuited case. The voltage across the load doubles, as the reflected pulse has the same sign as well as magnitude as that of the incident signal, in this case $\Gamma = +1$. This is required so that the total current is zero.

A more illustrative situation is shown in Figure 2.3.5, where a more complicated signal is incident on a load that has a resistance higher than that of the characteristic impedance of the line. The peaking of the voltage that results at the load is typically the design objective in many long digital interconnects, as less overall signal energy needs to be transmitted down the line, or equivalently a lower current drive capability of the source is required to achieve first incidence switching. This is at the price of having reflected signals on the interconnects, but these are dissipated through a combination of line loss and absorption of the reflected signal at the driver.

2.3.3: Input Reflection Coefficient of a Lossless Line

The reflection coefficient looking into a line varies with position along the line as the forward- and backward-traveling waves change in relative phase. Referring to Figure 2.3.4, at a distance ℓ from the load (i.e., $z = -\ell$), the input reflection looking into a terminated lossless line is

$$\Gamma_{\text{in}}|_{z=-\ell} = \frac{V^-(z=-\ell)}{V^+(z=-\ell)} = \frac{V^-(z=0)e^{-\jmath\beta\ell}}{V^+(z=0)e^{+\jmath\beta\ell}} = \frac{V^-(z=0)e^{-\jmath\beta\ell}}{V^+(z=0)e^{+\jmath\beta\ell}} = \Gamma_L e^{-\jmath\beta\ell} \tag{2.3.13}$$

Note that Γ_{in} has the same magnitude as Γ_L but rotates in the clockwise direction (becomes increasingly negative) at twice the rate of increase of the electrical length $\beta\ell$.

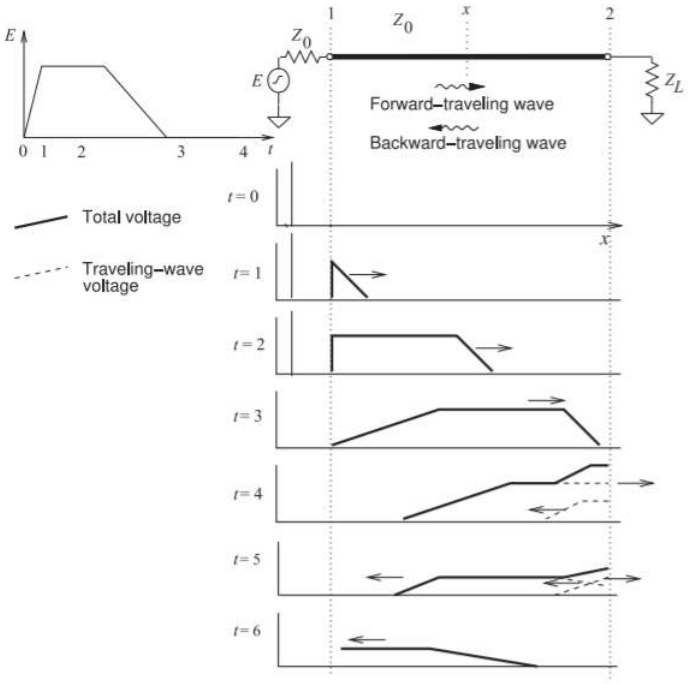


Figure 2.3.5: Reflection of a pulse from a termination with $Z_L > Z_0$. Z_L and Z_0 are real.

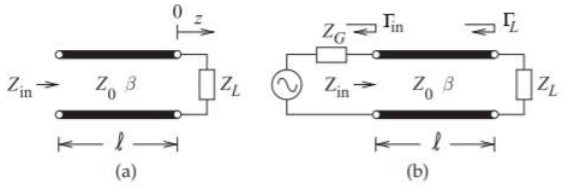


Figure 2.3.6: Terminated transmission line: (a) a transmission line terminated in a load impedance, Z_L, with an input impedance of Z_{in}; and (b) a transmission line with source impedance Z_G and load Z_L.

It is important to graphical concepts introduced later that there be a full appreciation for the angle of Γ_{in} becoming increasingly negative at twice the rate at which the electrical length of the line increases. Figure 2.3.7 is a way of visualizing this. The transmission line here is $\lambda/4$ long with an electrical length of 90° and is terminated in a load with reflection coefficient $\Gamma_L = +1$. At position $z = 0$ the forward-traveling voltage wave is $v^+(t,0) = |V^+|\cos(\omega t)$, and this then propagates down the line in the $+z$ direction. The forward-traveling voltage at point $z = \lambda/8$ at $t = 0$ will be the same as the voltage at $z = 0$ at a time one-eighth of a period in the past. The voltage at $z = \lambda/8$

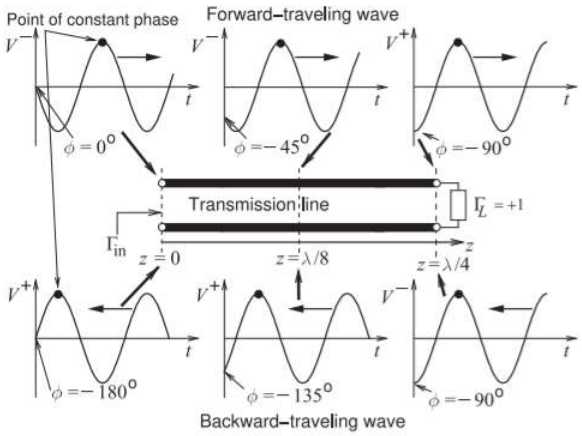


Figure 2.3.7: The forward-traveling wave $v^+(t,z) = |V^+|\cos(\omega t - \beta z) = |V^+|\cos(\omega t + \phi(z))$ and the backward-traveling wave $v^-(t,z) = |V^+|\cos(\omega t + \beta z) = |V^+|\cos[\omega t + \phi(z)]$. The phase, ϕ, of the forward-traveling wave becomes increasingly negative along the line as z increases, and when reflected the phase ϕ of the backward-traveling wave becomes increasingly negative as the wave moves away from the load (i.e. as z decreases).

is $v^+(t,\lambda/8) = |V^+|\cos(\omega t - 2\pi/8)$, i.e. there is a phase rotation of $-45°$. Then at $z = \lambda/4, v^+(t,\lambda/4) = |V^+|\cos(\omega t - 2\pi/4)$, i.e. at time $t = 0$ there is a phase rotation of $-90°$ relative to $v^+(0,0)$, and this is the negative of the electrical length of the line. The voltage wave reflects at the load and becomes a backward-traveling wave. Here $\Gamma_L = +1$ and so, at the load, the phase of the backward- and forward-traveling waves are the same. The backward-traveling wave continues to travel in the $-z$ direction and its phase at $t = 0$ becomes increasingly negative as z gets closer to the input of the line. The phase of the backward-traveling wave at $z = 0$ is rotated $-90°$ with respect to the backward-traveling wave at the load, and has rotated $-180°$ relative to the forward-traveling wave at $z = 0$. For a lossless line, in general, the angle of Γ_{in} = [phase of $V^-(z = 0)$ relative to the phase of $V^+(z = 0)$] + (the phase of Γ_L) = -2(electrical length of the line) + (the phase of Γ_L).

2.3.4: Input Impedance of a Lossless Line

The impedance looking into a lossless line varies with position, as the forward- and backward-traveling waves combine to yield position-dependent total voltage and current. At a distance ℓ from the load (i.e., $z = -\ell$), the input impedance seen looking toward the load is

$$Z_{in}|z = -\ell = \frac{V(z=-\ell)}{I(z=-\ell)} = Z_0\frac{I + |\Gamma|e^{j(\Theta - 2\beta\ell)}}{1 - |\Gamma|e^{j(\Theta-2\beta\ell)}} = Z_0\frac{1+\Gamma_L e^{j(-2\beta\ell)}}{1-\Gamma_L e^{j(-2\beta\ell)}} \tag{2.3.14}$$

Another form is obtained by substituting Equation (2.3.6) in Equation (2.3.14):

$$\begin{aligned} Z_{in} &= Z_0\frac{(Z_L+Z_0)e^{j\beta\ell} + (Z_L - Z_0)e^{-j\beta\ell}}{(Z_L+Z_0)e^{j\beta\ell} - (Z_L - Z_0)e^{-j\beta\ell}} = Z_0\frac{Z_L\cos(\beta\ell) + jZ_0\cos(\beta\ell)}{Z_0\cos(\beta\ell) + jZ_L\cos(\beta\ell)} \\ &= Z_0\frac{Z_L + jZ_0\tan\beta\ell}{Z_0 + jZ_L\tan\beta\ell} \end{aligned} \tag{2.3.15}$$

This is the **lossless telegrapher's equation**. The electrical length, $\beta\ell$, is in radians when used in calculations.

? Exercise 2.3.1

The script below calculates the input impedance for a transmission line, a load and the distance from the load to the source. Answer the following questions

- For a matched load, how does the input impedance change with distance?
- Slowly increase the load from the matched condition, what do you observe?
- Keep the magnitude of the load matched (50 Ω under the default conditions), what happens to the input impedance as the load becomes more reactive?

```
1 import numpy as np
```

```
from matplotlib import pylab as plt
from scipy.constants import c

#choose a frequency here
f = 5e9
omega = 2*np.pi*f
ld = c/f #free space wavelength

#choose a terminating impedance
#Z_real = 30
#Z_imag = np.sqrt(50**2 - Z_real**2)
#Z_l = Z_real + 1.0j*Z_imag
Z_l = 50

#transmission line characteristics
Z_0 = 50
l = 50e-2 #50 cm line
eps = 3.48 #effective dielectric constant for transmission line
ld_g = ld/eps
beta = 2*np.pi/ld_g


```

run | run all | restart | restart & run all

```
<matplotlib.legend.Legend at 0x7f614bdaad00>
```

Answer

- For a matched load the input impedance is constant
- As you increase the load impedance, the input impedance (real and imaginary) starts to change with distance. As the impedance increases, the changes become larger and the changes more rapid, meaning that at some locations, the source may experience something close to a short or open circuit
- Both the imaginary and real part of the impedance oscillate with distance. When the real part is zero, it stops oscillating and stays zero, while the imaginary part follows a tangent function with Z_in varying between $\pm\infty$

2.3.5: Standing Waves and Voltage Standing Wave Ratio

The total voltage on a terminated line is the sum of forward- and backwardtraveling waves. This sum produces what is called a standing wave. Figure 2.3.8 shows the total and traveling waveforms on a line terminated in a reactance and evaluated at times equal to multiples of an eighth of a period. Here the traveling waves have the same amplitude indicating that the termination of the line is reactive, $|\Gamma| = 1$. The interesting property here is that the total voltage appears as a standing wave with fixed points called nodes where the total voltage is always zero. This is more easily seen in Figure 2.3.9(a), where the total voltage is overlaid for many times. If the termination has resistance, then the magnitude of the backward-traveling wave will be less than that of the forward-traveling wave and the overlaid total voltage is as shown in Figure 2.3.9(b).

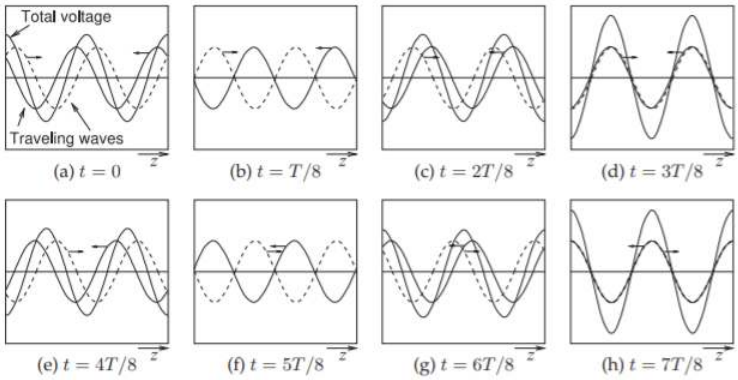
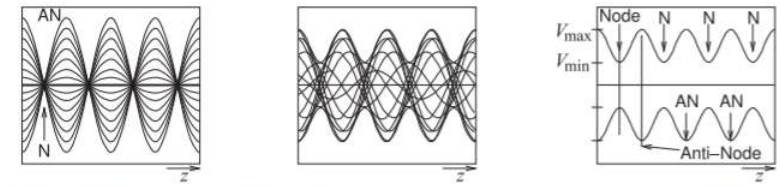


Figure 2.3.8: Evolution of a standing wave as the sum of forward- and backward-traveling waves (to the right and left, respectively) of equal amplitude evaluated at times t equal to eighths of the period T. At $t = T/8$ and $t = 5T/8$ the total voltage everywhere on the

line is zero.



(a) load is a reactance only (b) load is at least partially resistive (c) envelope of (b)

Figure 2.3.9: Standing waves as an overlay of waveforms at many times: (a) when the forward and backward-traveling waves have the same amplitude; (b) when the waves have different amplitudes; and (c) the envelope of the standing wave. N is a node (a minimum) and AN is an antinode (a maximum). Nodes, N, are separated by $\lambda/2$. Antinodes, AN, are separated by $\lambda/2$.

This is still a standing wave, but the minima are now not zero. The envelope of the standing wave is shown in Figure 2.3.9(c), where there is a maximum amplitude $V_{\max}$ and a minimum amplitude $V_{\min}$.

Now this situation will be examined mathematically to relate the standing wave to the reflection coefficient. If $\Gamma = 0$, then the magnitude of the total voltage on the line, $|V(z)|$, is equal to $|V_0^+|$ anywhere on the line. For this reason, such a line is said to be "flat." If there is reflection the magnitude of the total voltage on the line is not constant (see Figure 2.3.9(b)). Thus from Equation (2.3.13):

$$|V(z)| = |V_0^+||1 + \Gamma e^{2\jmath\beta z}| = |V_0^+||1 + \Gamma e^{-2\jmath\beta\ell}| \tag{2.3.16}$$

where $z = -\ell$ is the positive distance measured from the load at $z = 0$ toward the generator. Or, setting $\Gamma = |\Gamma|e^{\jmath\Theta}$,

$$|V(z)| = |V_0^+||1+|\Gamma|e^{\jmath(\Theta-2\beta\ell)}| \tag{2.3.17}$$

where Θ is the phase of the reflection coefficient ($\Gamma = |\Gamma|e^{\jmath\Theta}$) at the load. This result shows that the voltage magnitude oscillates with position z along the line. The maximum value occurs when $e^{\jmath(\Theta-2\beta\ell)} = 1$ and is given by

$$V_{\max} = |V_0^+|(1 + |\Gamma|) \tag{2.3.18}$$

Similarly the minimum value of the total voltage magnitude occurs when the phase term is $e^{\jmath(\Theta-2\beta\ell)} = -1$, and is given by

$$V_{\min} = |V_0^+|(1 - |\Gamma|) \tag{2.3.19}$$

A mismatch can be defined by the **voltage standing wave ratio (VSWR)**:

$$\text{VSWR} = \frac{V_{\max}}{V_{\min}} = \frac{(1 + |\Gamma|)}{(1 - |\Gamma|)} \tag{2.3.20}$$

Also

$$|\Gamma| = \frac{\text{VSWR} - 1}{\text{VSWR} + 1} \tag{2.3.21}$$

Notice that in general Γ is complex, but VSWR is necessarily always real and $1 \le \text{VSWR} \le \infty$. For the matched condition, $\Gamma = 0$ and $\text{VSWR} = 1$, and the closer VSWR is to 1, the closer the load is to being matched to the line and the more power is delivered to the load. The magnitude of the reflection coefficient on a line with a short-circuit or open-circuit load is 1, and in both cases the VSWR is infinite.

? Exercise 2.3.1

Below is a script that calculates the standing wave pattern and standing wave ratio for you. Use the script to verify the results from above. In the text below Z_l is the load impedance and Z_0 is the characteristic impedance of the transmission line. Answer for yourself the following questions:

- What is the difference between $Z_l = \infty$ (use a very large value in the script) and $Z_l = 0\ \Omega$?
- What happens if $|Z_l| = 50\ \Omega$ but transitions from real to complex?
- As you go from $Z_l < Z_0$ to $Z_l > Z_0$, what happens to the nodes and anti-nodes in the standing wave?

```
import numpy as np
from matplotlib import pylab as plt
from scipy.constants import c

#choose a frequency here
f = 5e9
omega = 2*np.pi*f
ld = c/f #free space wavelength

#choose a terminating impedance
Z_l = 50

#transmission line characteristics
Z_0 = 0
l = 50e-2 #50 cm line
eps = 3.48 #effective dielectric constant for transmission line
ld_g = ld/eps
beta = 2*np.pi/ld_g


#calculate reflection coefficient
gamma = (Z_l - Z_0)/(Z_l + Z_0)

step_size = ld_g/100
```

run | run all | restart | restart & run all

```
<matplotlib.legend.Legend at 0x7f09690d9190>
```

Answer

Add texts here. Do not delete this text first.

- For $Z_l = \infty$ and $Z_l = 0$, VSWR is ∞, but the for and infinite load, there is an antinode at the load, and a short, there is a node at the load.
- A purely reactive load has $\text{VSWR} = \infty$ but neither the node nor the antinode are located at the load. As the load becomes more real, the VSWR drops until it reaches zero for $Z_l = Z_0$
- The nodes and antinodes swap

To determine the position of the standing wave maximum, ℓmax, consider Equation (2.3.17) and note that at the maximum

$$\Theta - 2\beta\ell_{\max} = 2n\pi, \quad n = 0, 1, 2, ... \tag{2.3.22}$$

Here Θ is the angle of the reflection coefficient at the load:

$$\Theta - 2n\pi = 2\frac{2\pi}{\lambda_g}\ell_{\max} \tag{2.3.23}$$

Thus the position of the voltage maxima, $\ell_{\max}$, normalized to wavelength is

$$\frac{\ell_{\max}}{\lambda_g} = \frac{1}{2}\left(\frac{\Theta}{2\pi} - n\right), \quad n = 0, -1, -2, ... \tag{2.3.24}$$

Similarly the position of the voltage minima is (using Equation (2.3.17)):

$$\Theta - 2\beta\ell_{\min} = (2n+1)\pi \tag{2.3.25}$$

After rearranging the terms,

$$\frac{\ell_{\min}}{\lambda_g} = \frac{1}{2}\left(\frac{\Theta}{2\pi} - n + \frac{1}{2}\right), \quad n = 0, -1, -2, ... \tag{2.3.26}$$

Summarizing from Equations (2.3.24) and (2.3.26):

1. The distance between two successive maxima is $\lambda_g/2$.
2. The distance between two successive minima is $\lambda_g/2$.

3. The distance between a maximum and an adjacent minimum is $\lambda_g/4$.
4. From the measured VSWR the magnitude of the reflection coefficient $|\Gamma|$ can be found. From the measured $\ell_{\max}$ the angle Θ of Γ can be found. Then from Γ the load impedance can be found.

In a similar manner to that above, the magnitude of the total current on the line is

$$|I(\ell)| = \frac{|V_0^+|}{Z_0}\left|1-|\Gamma|e^{j(\Theta-2\beta\ell)}\right| \tag{2.3.27}$$

Hence the standing wave current is maximum where the standing-wave voltage amplitude is minimum, and minimum where the standing-wave voltage amplitude is maximum.

Now Z_{in} in Equation (2.3.15) is a periodic function of ℓ with period $\lambda/2$ and it varies between $Z_{\max}$ and $Z_{\min}$, where

$$Z_{\max} = \frac{V_{\max}}{I_{\min}} = Z_0 \times \text{VSWR} \quad \text{and} \quad Z_{\min} = \frac{V_{\min}}{I_{\max}} = \frac{Z_0}{\text{VSWR}} \tag{2.3.28}$$

Example 2.3.5: Standing Wave Ratio

In Example 3.3.2 a load consisted of a capacitor of 10 pF in shunt with a resistor of $60\ \Omega$. The load terminated a lossless $50\ \Omega$ transmission line. The operating frequency is 5 GHz.

a. What is the SWR?
b. What is the current standing wave ratio (ISWR)? (When SWR is used on its own it is assumed to refer to VSWR.)

Solution

a. From Example 3.3.2, $\Gamma_L = 0.993\angle 187.3^\circ$ and so

$$\text{VSWR} = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} = \frac{1+0.993}{1-0.993} = 285$$

b. $\text{ISWR} = \text{VSWR} = 285$.

Example 2.3.6: Standing Waves

A load has an impedance $Z_L = 45 + j75\ \Omega$ and the system reference impedance, Z_0, is $100\ \Omega$.

a. What is the reflection coefficient?
b. What is the current reflection coefficient?
c. What is the SWR?
d. What is the ISWR?
e. The power available from a source with a $100\ \Omega$ Thevenin equivalent impedance is 1 mW. The source is connected directly to the load, Z_L. Use the reflection coefficient to calculate the power delivered to Z_L.
f. What is the total power absorbed by the Thevenin equivalent source impedance?
g. Discuss the effect on power flow of inserting a lossless $100\ \Omega$ transmission line between the source and the load.

Solution

a. The voltage reflection coefficient is

$$\begin{aligned}\Gamma_L &= (Z_L - Z_0)/(Z_L + Z_0) = (45 + j75 - 100)/(45 + j75 + 100)\\ &= (93.0\angle(2.204 \text{ rads}))/(163.2\angle(0.4773 \text{ rads}))\\ &= 0.570\angle(1.726 \text{ rads} 0.570\angle 98.9^\circ = -0.0881 + j0.653 = \Gamma^V\end{aligned} \tag{2.3.29}$$

b. The current reflection coefficient is

$$\Gamma^I = -\Gamma^V = 0.0881 - j0.563 = 0.570\angle(98.9^\circ - 180^\circ) = 0.570\angle 81.1^\circ \tag{2.3.30}$$

c. The SWR is the VSWR, so

$$\text{SWR} = \text{VSWR} = \frac{V_{\max}}{V_{\min}} = \frac{1+|\Gamma^V|}{1-|\Gamma^V|} = \frac{1+0.570}{1-0.570} = 3.65 \tag{2.3.31}$$

d. The current SWR is $\text{ISWR} = \text{VSWR}$.

e. To determine the reflection coefficient of the load, begin by developing the Thevenin equivalent circuit of the load. The power available from the source is $P_A = 1$ mW, so the Thevenin equivalent circuit is

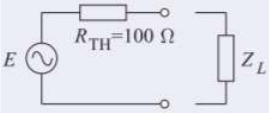


Figure 2.3.8

The power reflected by the load is

$$P_R = P_A|\Gamma_L^2| = 1 \text{ mW} \cdot (0.570)^2 = 0.325 \text{ mW}$$

and the power delivered to the load is

$$P_D = P_A(1-|\Gamma_L^2|) = 0.675 \text{ mW}$$

f. It is tempting to think that the power dissipated in R_{TH} is just P_R. However, this is not correct. Instead, the current in R_{TH} must be determined and then the power dissipated in R_{TH} found. Let the current through $R_{\text{TH}cdot}$ be I, and this is composed of forward and backward-traveling components:

$$I = I^+ + I^- = (1+\Gamma_I)I^+$$

where I^+ is the forward-traveling current wave. Thus

$$P_A = \frac{1}{2}|I^+|^2 R_{\text{TH}} = \frac{1}{2}|I^+|^2 \times 100 = 1 \text{ mW} = 10^{-3} \text{ W}$$

so $I^+ = 4.47$ mA, and

$$I = (1+\Gamma_I)I^+ = (1+0.0881-\jmath 0.563) \times 4.47 \times 10^{-3} \text{ A}, \quad |I| = 5.48 \text{ mA}$$

The power dissipated in R_{TH} is

$$P_{\text{TH}} = \frac{1}{2}|I|^2 R_{\text{TH}} = \frac{1}{2}(5.48 \times 10^{-3})^2 R_{\text{TH}} - 1.50 \text{ mW} \tag{2.3.32}$$

The circuit is that shown in part (e) and so the current in R_{TH} is the same as the current in Z_L. Thus the power delivered to the load Z_L is due to the real part of Z_L:

$$P_D = \frac{1}{2}|I^2|\Re(Z_L) = \frac{1}{2}(5.48 \times 10^{-3})^2 \times 45 = 0.676 \text{ mW} \tag{2.3.33}$$

g. Inserting a transmission line with the same characteristic impedance as the Thevenin equivalent impedance will have no effect on power flow.

2.3.6: Summary

This section related the physics of traveling voltage and current waves on lossless transmission lines to the total voltage and current view. First the input reflection coefficient of a terminated lossless line was developed and from this the input impedance, which is the ratio of total voltage and total current, derived. At any point along a line the amplitude of total voltage varies sinusoidally, tracing out a standing wave pattern along the line and yielding the VSWR metric which is the ratio of the maximum amplitude of the total voltage to the minimum amplitude of that voltage. This is an important metric that is often used to provide an indication of

how good a match, i.e. how small the reflection is, with a $\text{VSWR} = 1$ indicating no reflection and a $\text{VSWR} = \infty$ indicating total reflection, i.e. a reflection coefficient magnitude of 1.

This page titled 2.3: The Lossless Terminated Line is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **3.3: The Lossless Terminated Line** by Michael Steer is licensed CC BY-NC 4.0.

2.4: Summary

In this chapter a classical tr\FACeatment of transmission lines was presented. Transmission lines are distributed elements and form the basis of microwave circuits. A distinguishing feature is that they support forward- and backward-traveling waves and they can be used to implement circuit functions.

The most important of the formulas presented in this chapter are listed here. Reflection coefficients are referenced to an impedance Z_0, the load impedance is Z_L, and a line has a characteristic impedance Z_0, physical length ℓ, and propagation constant γ (or electrical length in radians of $\beta\ell$ where ℓ is the physical length of the line.

Reflection coefficient of a load impedance Z_L

$$\Gamma = \Gamma^V = \frac{Z_L - Z_{\text{REF}}}{Z_L + Z_{\text{REF}}} \quad (3.3.6)$$

Load impedance in terms of reflection coefficient Γ

$$Z_L = Z_{\text{REF}} \frac{1+\Gamma}{1-\Gamma} \quad (3.3.8)$$

Input reflection coefficient of a lossless line of length ℓ

$$\Gamma_{\text{in}} = \Gamma_L e^{-j2\beta\ell} \quad (3.3.13)$$

Reflection coefficient in terms of VSWR

$$|\Gamma| = \frac{\text{VSWR} - 1}{\text{VSWR} + 1} \quad (3.3.21)$$

Input impedance of a lossless line

$$Z_{\text{in}} = Z_0 \frac{Z_L + jZ_0 \tan\beta\ell}{Z_0 + jZ_L \tan\beta\ell} \quad (3.3.15)$$

VSWR in terms of reflection coefficient

$$\text{VSWR} = \frac{(1+|\Gamma|)}{(1-|\Gamma|)} \quad (3.3.20)$$

This page titled 2.4: Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **3.6: Summary** by Michael Steer is licensed CC BY-NC 4.0.

2.5: References

[1] T. Edwards and M. Steer, Foundations for Microstrip Circuit Design. John Wiley & Sons, 2016.

[2] C. Boyer and U. Merzbach, “Invention of logarithms,” in A History of Mathematics, 2nd ed. John Wiley & Sons, 1991.

This page titled 2.5: References is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **3.7: References** by Michael Steer is licensed CC BY-NC 4.0.

2.6: Exercises

1. A coaxial line is short-circuited at one end and is filled with a dielectric with a relative permittivity of 64. [Parallels Example 3.1.1]
   a. What is the free-space wavelength at 18 GHz?
   b. What is the wavelength in the dielectricfilled coaxial line at 18 GHz?
   c. The first resonance of the coaxial resonator is at 18 GHz. What is the physical length of the resonator?
2. A transmission line has the following $RLGC$ parameters: $R = 100\,\Omega/\text{m},\ L = 85$ nH/m, $G = 1$ S/m, and $C = 150$ pF/m. Consider a traveling wave on the transmission line with a frequency of 1 GHz. [Parallels Example 3.2.2]
   a. What is the attenuation constant?
   b. What is the phase constant?
   c. What is the phase velocity?
   d. What is the characteristic impedance of the line?
   e. What is the group velocity?
3. A transmission line has the per-unit length parameters $L = 85$ nH/m, $G = 1$ S/m, and $C = 150$ pF/m. Use a frequency of 1 GHz. [Parallels Example 3.2.2]
   a. What is the phase velocity if $R = 0\,\Omega/\text{m}$?
   b. What is the group velocity if $R = 0\,\Omega/\text{m}$?
   c. If $R = 10$ kΩ/m what is the phase velocity?
   d. If $R = 10$ kΩ/m what is the group velocity?
4. A line is 10 cm long and at the operating frequency the phase constant β is 40 rad/m. What is the electrical length of the line? [Parallels Example 3.2.1]
5. A coaxial transmission line is filled with lossy dielectric material with a relative permittivity of $5 - j0.2$. If the line is air-filled it would have a characteristic impedance of $100\,\Omega$. What is the input impedance of the line if it is 1 km long? Use reasonable approximations. [Hint: Does the termination matter?]
6. A transmission line has the per unit length parameters $R = 2\,\Omega/\text{cm},\ L = 100$ nH/m, $G = 1$ mS/M, $C = 200$ pF/m.
   a. What is the propagation constant of the line at 5 GHz?
   b. What is the characteristic impedance of the line at 5 GHz?
   c. Plot the magnitude of the characteristic impedance versus frequency from 100 MHz to 10 GHz.
7. A line is 20 cm long and at 1 GHz the phase constant β is 20 rad/m. What is the electrical length of the line in degrees?
8. What is the electrical length of a line that is a quarter of a wavelength long,
   a. in degrees?
   b. in radians?
9. A lossless transmission line has an inductance of 8 nH/cm and a capacitance of 40 pF/cm.
   a. What is the characteristic impedance of the line?
   b. What is the phase velocity on the line at 1 GHz?
10. A transmission line has an attenuation of 2 dB/m and a phase constant of 25 radians/m at 2 GHz. [Parallels Example 3.2.3]
    a. What is the complex propagation constant of the transmission line?
    b. If the capacitance of the line is $50\ \text{pF} \cdot \text{m}^{-1}$ and $G = 0$, what is the characteristic impedance of the line?
11. A very low-loss microstrip transmission line has the following per unit length parameters: $R = 2\,\Omega/\text{m},\ L = 80$ nH/m, $C = 200$ pF/m, and $G = 1$ μS/m.
    a. What is the characteristic impedance of the line if loss is ignored?
    b. What is the attenuation constant due to conductor loss?
    c. What is the attenuation constant due to dielectric loss?
12. A lossless transmission line carrying a 1 GHz signal has the following per unit length parameters: $L = 80$ nH/m, $C = 200$ pF/m.
    a. What is the attenuation constant?
    b. What is the phase constant?

c. What is the phase velocity?
d. What is the characteristic impedance of the line?

13. A transmission line has a characteristic impedance Z_0 and is terminated in a load with a reflection coefficient of $0.8\angle 45^\circ$. A forward-traveling voltage wave on the line has a power of $1\ \mathrm{dBm}$.
    1. How much power is reflected by the load?
    2. What is the power delivered to the load?
14. A transmission line has an attenuation of $0.2\ \mathrm{dB/cm}$ and a phase constant of $50\ \mathrm{radians/m}$ at $1\ \mathrm{GHz}$.
    a. What is the complex propagation constant of the transmission line?
    b. If the capacitance of the line is $100\ \mathrm{pF/m}$ and $G = 0$, what is the complex characteristic impedance of the line?
    c. If the line is driven by a source modeled as an ideal voltage and a series impedance, what is the impedance of the source for maximum transfer of power to the transmission line?
    d. If $1\ \mathrm{W}$ is delivered (i.e. in the forwardtraveling wave) to the transmission line by the generator, what is the power in the forward-traveling wave on the line at $2\ \mathrm{m}$ from the generator?
15. A lossless transmission line is driven by a $1\ \mathrm{GHz}$ generator having a Thevenin equivalent impedance of $50\ \Omega$. The transmission line is lossless, has a characteristic impedance of $75\ \Omega$, and is infinitely long. The maximum power that can be delivered to a load attached to the generator is $2\ \mathrm{W}$.
    a. What is the total (phasor) voltage at the input to the transmission line?
    b. What is the magnitude of the forwardtraveling voltage wave at the generator side of the line?
    c. What is the magnitude of the forwardtraveling current wave at the generator side of the line?
16. A transmission line is terminated in a short circuit. What is the ratio of the forward- and backward-traveling voltage waves at the termination? [Parallels Example 3.3.1]
17. A $50\ \Omega$ transmission line is terminated in a $40\ \Omega$ load. What is the ratio of the forward- to the backward-traveling voltage waves at the termination? [Parallels Example 3.3.1]
18. A $50\ \Omega$ transmission line is terminated in an open circuit. What is the ratio of the forwardto the backward-traveling voltage waves at the termination? [Parallels Example 3.3.1]
19. A line has a characteristic impedance Z_0 and is terminated in a load with a reflection coefficient of 0.8. A forward-traveling voltage wave on the line has a power of $1\ \mathrm{W}$.
    a. How much power is reflected by the load?
    b. What is the power delivered to the load?
20. A load consists of a shunt connection of a capacitor of $10\ \mathrm{pF}$ and a resistor of $25\ \Omega$. The load terminates a lossless $50\ \Omega$ transmission line. The operating frequency is $1\ \mathrm{GHz}$. [Parallels Example 3.3.2]
    a. What is the impedance of the load?
    b. What is the normalized impedance of the load (normalized to the characteristic impedance of the line)?
    c. What is the reflection coefficient of the load?
    d. What is the current reflection coefficient of the load?
    e. What is the standing wave ratio (SWR)?
    f. What is the current standing wave ratio (ISWR)?
21. An amplifier is connected to a load by a transmission line matched to the amplifier. If the SWR on the line is 1.5, what percentage of the available amplifier power is absorbed by the load?
22. A load has a reflection coefficient of 0.5 when referred to $50\ \Omega$. The load is at the end of a line with a $50\ \Omega$ characteristic impedance.
    a. If the line has an electrical length of 45°, what is the reflection coefficient calculated at the input of the line?
    b. What is the VSWR on the $50\ \Omega$ line?
23. A $100\ \Omega$ resistor in parallel with a $5\ \mathrm{pF}$ capacitor terminates a $100\ \Omega$ transmission line. Calculate the SWR on the line at $2\ \mathrm{GHz}$.
24. A lossless $50\ \Omega$ transmission line has a $50\ \Omega$ generator at one end and is terminated in $100\ \Omega$. What is the VSWR on the line?
25. A lossless $75\ \Omega$ line is driven by a $75\ \Omega$ generator. The line is terminated in a load that with a reflection coefficient (referred to $50\ \Omega$) of $0.5 + j0.5$. What is the VSWR on the line?

26. A load with a 20 pF capacitor in parallel with a $50\ \Omega$ resistor terminates a $25\ \Omega$ line. The operating frequency is 5 GHz. [Parallels Example 3.3.3]
    a. What is the VSWR?
    b. What is ISWR?
27. A load $Z_L = 55 - j55\ \Omega$ and the system reference impedance, Z_0, is $50\ \Omega$. [Parallels Example 3.3.4]
    a. What is the load reflection coefficient Γ_L?
    b. What is the current reflection coefficient?
    c. What is the VSWR on the line?
    d. What is the ISWR on the line?
    e. Now consider a source connected directly to the load. The source has a Thevenin equivalent impedance $Z_G = 60\ \Omega$ and an available power of 1 W. Use Γ_L to find the power delivered to Z_L.
    f. What is the total power absorbed by Z_G?
28. A load of $100\ \Omega$ is to be matched to a transmission line with a characteristic impedance of $50\ \Omega$. Use a quarter-wave transformer. What is the characteristic impedance of the quarterwave transformer?
29. Determine the characteristic impedance of a quarter-wave transformer used to match a load of $50\ \Omega$ to a generator with a Thevenin equivalent impedance of $75\ \Omega$.
30. A transmission line is to be inserted between a $5\ \Omega$ line and a $50\ \Omega$ load so that there is maximum power transfer to the $50\ \Omega$ load at 20 GHz.
    a. How long is the inserted line in terms of wavelengths at 20 GHz?
    b. What is the characteristic impedance of the line at 20 GHz?

2.6.1: Exercises By Section

†challenging, ‡very challenging

§3.11

§3.22†, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12†, 13

§3.314†, 15, 16, 17, 18, 19†, 20†, 21†, 22, 23, 24, 25, 26, 27

§3.428, 29, 30

2.6.2: Answers to Selected Exercises

10. $0.23 + j25\ \text{m}^{-1}$

29. $61.2\ \Omega$

This page titled 2.6: Exercises is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **3.8: Exercises** by Michael Steer is licensed CC BY-NC 4.0.

CHAPTER OVERVIEW

3: Impedance Matching

3.1: Introduction
3.2: Matching Networks
3.3: Impedance Transforming Networks
3.4: The L Matching Network
3.5: Dealing with Complex Loads
3.6: Dealing with Complex Loads
3.7: Summary
3.8: References
3.9: Exercises
3.10: Impedance Matching Using Smith Charts
3.11: Distributed Matching
3.12: Matching Using the Smith Chart

This page titled 3: Impedance Matching is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

3.1: Introduction

The maximum transfer of signal power is one of the prime objectives in RF and microwave circuit design. Power traverses a network from a source to a load generally through a sequence of two-port networks. Maximum power transfer requires that the Thevenin equivalent impedance of a source be matched to the impedance seen from the source. That is, the source should be presented with the complex conjugate of the source impedance. This is achieved by designing what is called a matching network inserted between the source and load. Design of the matching network is called impedance matching.

Section 3.1.1 describes two common matching objectives. Then design approaches for impedance matching are presented first with an algorithmic approach in Sections 3.1.2–3.1.7 and then a graphical approach based on using a Smith chart in Sections 3.1.8 and 3.1.9.

This page titled 3.1: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.1: Introduction** by Michael Steer is licensed CC BY-NC 4.0.

3.2: Matching Networks

Matching networks are constructed using lossless elements such as lumped capacitors, lumped inductors and transmission lines and so have, ideally, no loss and introduce no additional noise. This section discusses matching objectives and the types of matching networks.

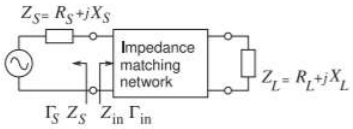


Figure 3.2.1: A source with Thevenin equivalent impedance Z_S and load with impedance Z_L interfaced by a matching network presenting an impedance Z_{in} to the source.

3.2.1: Matching for Zero Reflection or for Maximum Power Transfer

With RF circuits the aim of matching is to achieve maximum power transfer. With reference to Figure 3.2.1 the condition for maximum power transfer is $Z_{\text{in}} = Z_S^*$ which is equivalent to $\Gamma_{in} = \Gamma_s^*$. The proof is as follows:

$$\Gamma_{\text{in}} = \left(\frac{Z_{\text{in}} - Z_{\text{REF}}}{Z_{\text{in}} + Z_{\text{REF}}} \right) \tag{3.2.1}$$

and for maximum power transfer $Z_{\text{in}} = Z_S^*$, so

$$\begin{aligned} \Gamma_{\text{in}}^* &= \frac{Z_{\text{in}} - Z_{\text{REF}}}{Z_{\text{in}} + Z_{\text{REF}}} = \left(\frac{Z_S^* - Z_{\text{REF}}}{Z_S^* + Z_{\text{REF}}} \right)^* = \frac{(Z_S^* - Z_0)^*}{(Z_S^* + Z_0)^*} \\ &= \frac{(Z_S^*)^* - Z_{\text{REF}}^*}{(Z_S^*)^* + Z_{\text{REF}}^*} = \frac{Z_S - Z_{\text{REF}}^*}{Z_S + Z_{\text{REF}}^*} = \Gamma_S \end{aligned} \tag{3.2.2}$$

If Z_{REF} is real, $Z_{\text{REF}}^* = Z_{\text{REF}}$ and then the condition for maximum power transfer is

$$\Gamma_{\text{in}}^* = \frac{Z_S - Z_{\text{REF}}}{Z_S + Z_{\text{REF}}} = \Gamma_S \tag{3.2.3}$$

Thus, provided that Z_{REF} is real, the condition for maximum power transfer in terms of reflection coefficients is $\Gamma_{\text{in}}^* = \Gamma_S$ or $\Gamma_{\text{in}} = \Gamma_S^*$.

3.2.2: Types of Matching Networks

Up to a few gigahertz, lumped inductors and capacitors can be used in matching networks. Above a few gigahertz parasitics result in self-resonance. Lumped elements are also lossy. Segments of transmission lines are also used in matching networks as the loss of a transmission line component is always much less than the loss of a lumped inductor. However the length of a transmission line segment is up to $\lambda/4$ which is far too large to fit in consumer wireless products operating below a few gigahertz. An impedance matching network may consist of

a. Lumped elements only. These are the smallest networks, but have the most stringent limit on the maximum frequency of operation. The relatively high resistive loss of an inductor is the main limiting factor limiting performance. The self resonant frequency of an inductor limits operation to low microwave frequencies.
b. Distributed elements (microstrip or other transmission line circuits) only. These have excellent performance, but their size restricts their use in systems to above a few gigahertz.
c. A hybrid design combining lumped and distributed elements, primarily small sections of lines with capacitors. These lines are shorter than in a design with distributed elements only, but the hybrid design has higher performance than a lumped-element-only design.

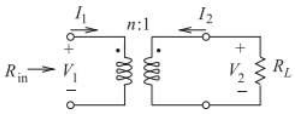


Figure 3.2.2: A transformer as a matching network. Port 1 is on the left or primary side and Port 2 is on the right or secondary side.

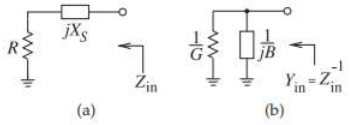


Figure 3.2.2: Matching using a series reactance: (a) the series reactive element; and (b) its equivalent shunt circuit.

This page titled 3.2: Matching Networks is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.2: Matching Networks** by Michael Steer is licensed CC BY-NC 4.0.

3.3: Impedance Transforming Networks

Transformers and reactive elements considered in this section can be used to losslessly transform impedance levels. This is a basic aspect of network design.

3.3.1: The Ideal Transformer

The ideal transformer shown in Figure 10.2.2 can be used to match a load to a source if the source and load impedances are resistances. This will be shown by starting with the constitutive relations of the transformer:

$$V_1 = nV_2 \quad \text{and} \quad I_1 = -I_2/n \tag{3.3.1}$$

Here n is the transformer ratio. For a wire-wound transformer, n is the ratio of the number of windings on the primary side, Port 1, to the number of windings on the secondary side, Port 2. Thus the input resistance, R_{in}, is related to the load resistance, R_L, by

$$R_{\text{in}} = \frac{V_1}{I_1} = -n^2\frac{V_2}{I_2} = n^2 R_L \tag{3.3.2}$$

The matching problem with purely resistive load and source impedances is solved by choosing the appropriate winding ratio, n. However, resistive-only problems are rare at RF, and so other matching circuits must be used.

3.3.2: A Series Reactive Element

Matching using lumped elements is based on the impedance and admittance transforming properties of series and shunt reactive elements. Even a single reactive element can achieve limited impedance matching. Consider the series reactive element shown in Figure 3.2.3(a). Here the reactive element, X_S, is in series with a resistance R. The shunt equivalent of this network is shown in Figure 3.2.3(b) with a shunt susceptance of B. In this transformation the resistance R has been converted to a resistance $R_P = 1/G$. The mathematics describing this transformation is as follows. The input admittance of the series connection (Figure 3.2.3(a)) is

$$Y_{\text{in}}(\omega) = \frac{1}{Z_{\text{in}}(\omega)} = \frac{1}{R + \jmath X_S} = \frac{R}{R^2 + X_S^2} - \jmath\frac{X_S}{R^2 + X_S^2} \tag{3.3.3}$$

Thus the elements of the equivalent shunt network, Figure 3.2.3(b), are

$$G = \frac{R}{R^2 + X_S^2} \quad \text{and} \quad B = -\frac{X_S}{R^2 + X_S^2} \tag{3.3.4}$$

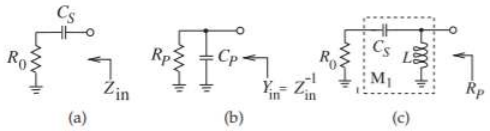


Figure 3.3.1: Impedance transformation by a series reactive element: (a) a resistor with a series capacitor; (b) its equivalent shunt circuit; and (c) an LC network.

The "resistance" of the network, R, has been transformed to a new value,

$$R_P = G^{-1} = \frac{R^2 + X_S^2}{R} > R \tag{3.3.5}$$

This is an important start to matching, as X_S can be chosen to convert R (a load, for example) to any desired resistance value (such as the resistance of a source). However there is still a residual reactance that must be removed to complete the matching network design. Before moving on to the solution of this problem consider the following example.

✔ Example 3.3.1: Capacitive Impedance Transformation

Consider the impedance transforming properties of the capacitive series element in Figure 3.3.1(a). Show that the capacitor can be adjusted to obtain any positive shunt resistance.

Solution

The concept here is that the series resistor and capacitor network has an equivalent shunt circuit that includes a capacitor and a resistor. By adjusting C_S any value can be obtained for R_P. From Equation (3.3.5),

$$R_P = \frac{R_0^2 + (1/\omega^2 C_S^2)}{R_0} = \frac{1 + \omega^2 C_S^2 R_0^2}{\omega^2 C_S^2 R_0} \tag{3.3.6}$$

and the susceptance is

$$B = \frac{(1/\omega C_S)}{R_0^2 + 1/\omega^2 C_S^2} = \omega \frac{C_S}{1 + \omega^2 C_S^2 R_0^2} \tag{3.3.7}$$

Thus

$$C_P = \frac{B}{\omega} = \frac{C_S}{1 + \omega^2 C_S^2 R_0^2} \tag{3.3.8}$$

To match R_0 to a resistive load R_P ($> R_0$) at a radian frequency ω_d, then, from Equation (3.3.6), the series capacitance required, i.e. the design equation for C_S, comes from

$$\omega_d C_S = 1/\sqrt{R_0 R_P - R_0^2} \tag{3.3.9}$$

To complete the matching design, use a shunt inductor L, as shown in Figure 3.3.1(c), where $\omega_d C_P = 1/(\omega_d L)$. The equivalent impedance in Figure 3.3.1(c) is a resistor of value R_P, with a value that can be adjusted by choosing C_S which then requires L to be adjusted.

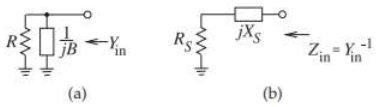


Figure 3.3.2: A resistor with (a) a shunt parallel reactive element where B is a susceptance, and (b) its equivalent series circuit.

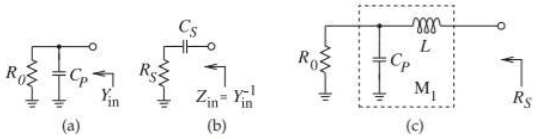


Figure 3.3.3: Parallel-to-series transformation: (a) resistor with shunt capacitor; (b) its equivalent series circuit; and (c) the transforming circuit with added series inductor.

3.3.3: A Parallel Reactive Element

The dual of the series matching procedure is the use of a parallel reactive element, as shown in Figure 3.3.2(a). The input admittance of the shunt circuit

$$Y_{\text{in}} = \frac{1}{R} + \jmath B \tag{3.3.10}$$

This can be converted to a series circuit by calculating $Z_{\text{in}} = 1/Y_{\text{in}}$:

$$Z_{\text{in}} = \frac{R}{1 + \jmath BR} = \frac{R}{1 + B^2 R^2} - \jmath \frac{BR^2}{1 + B^2 R^2} \tag{3.3.11}$$

So

$$R_S = \frac{R}{1 + B^2 R^2} \quad \text{and} \quad X_S = \frac{-BR^2}{1 + B^2 R^2} \tag{3.3.12}$$

Notice that $R_S < R$.

Example 3.3.2: Parallel Tuning

As an example of the use of a parallel reactive element to tune a resistance value, consider the circuit in Figure 3.3.3(a) where a capacitor tunes the effective resistance value so that the series equivalent circuit (Figure 3.3.3(b)) has elements

$$R_S = \frac{R_0}{1+\omega^2 C_P^2 R_0^2} \quad \text{and} \quad X_S = -\frac{\omega C_P R_0^2}{1+\omega^2 C_P^2 R_0^2} = -\frac{1}{\omega C_S} \tag{3.3.13}$$

So

$$C_S = \frac{1+\omega^2 C_P^2 R_0^2}{\omega^2 C_P R_0^2} \tag{3.3.14}$$

Now consider matching R_0 to a resistive load R_S, which is less than R_0 at a given frequency ω_d. This requires that

$$\omega_d C_P = \sqrt{1/(R_S R_0) - 1/R_0^2}$$

To complete the design, use a series inductor to remove the reactive effect of the capacitor, as shown in Figure 3.3.3(C). The value of the inductor required is found from

$$\omega_d L = \frac{1}{\omega_d C_S} \quad \text{that is} \quad L = \frac{1}{\omega_d^2 C_s} \tag{3.3.15}$$

This page titled 3.3: Impedance Transforming Networks is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.3: Impedance Transforming Networks** by Michael Steer is licensed CC BY-NC 4.0.

3.4: The L Matching Network

The examples in the previous two sections suggest the basic concept behind lossless matching of two different resistance levels using an L network:

Step 1: Use a series (shunt) reactive element to transform a smaller (larger) resistance up (down) to a larger (smaller) value with a real part equal to the desired resistance value.

Step 2: Use a shunt (series) reactive element to resonate with (or cancel) the imaginary part of the impedance that results from Step 1.

So a resistance can be transformed to any resistive value by using an LC transforming circuit. A summary of the L matching networks is given in Figure 3.4.1. The two possible cases, $R_S < R_L$ and $R_L < R_S$, will be considered in the following subsections.

3.4.1: Design Equations for $R_S < R_L$

Consider the matching network topology of Figure 3.4.2. Here

$$Z_{\text{in}} = \frac{R_L(\jmath X_P)}{R_L + \jmath X_P} = \frac{R_L X_P^2}{R_L^2 + X_P^2} + \jmath \frac{X_P R_L^2}{R_L^2 + X_P^2} \tag{3.4.1}$$

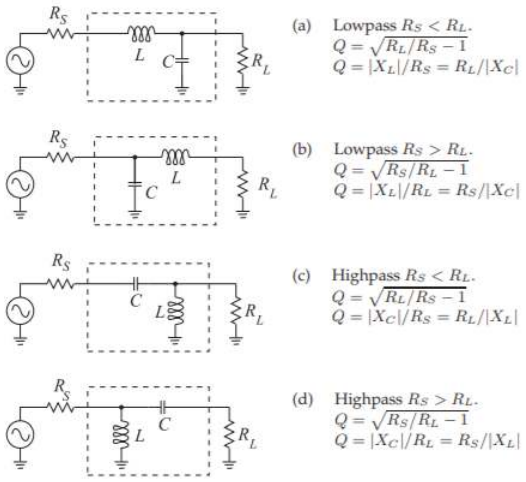


Figure 3.4.1: L matching networks consisting of one shunt reactive element and one series reactive element. (R_S is matched to R_L.) X_C is the reactance of the capacitor C, and X_L is the reactance of the inductor L. Note that with a two-element matching network the Q and thus bandwidth of the match is fixed.

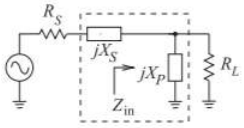


Figure 3.4.2: Two-element matching network topology for $R_S < R_L$. X_S is the series reactance and X_P is the parallel reactance.

and the matching objective is $Z_{\text{in}} = R_S - \jmath X_S$ so that

$$R_s = \frac{R_L X_P^2}{R_L^2 + X_P^2} \quad \text{and} \quad X_S = \frac{-X_P R_L^2}{X_P^2 + R_L^2} \tag{3.4.2}$$

From these

$$\frac{R_S}{R_L} = \frac{1}{(R_L/X_P)^2 + 1} \quad \text{and} \quad -\frac{X_S}{R_S} = \frac{R_L}{X_P} \tag{3.4.3}$$

Introducing the quantities

$$Q_S = \text{ the } Q \text{ of the series leg } = |X_S/R_S| \tag{3.4.4}$$

$$Q_P = \text{ the } Q \text{ of the shunt leg } = |R_L/X_P| \tag{3.4.5}$$

leads to the final design equations for $R_S < R_L$:

$$|Q_S| = |Q_P| = \sqrt{\frac{R_L}{R_S} - 1} \tag{3.4.6}$$

The L matching network principle is that X_P and X_S will be either capacitive or inductive and they will have the opposite sign (i.e., the L matching network comprises one inductor and one capacitor). Also, once R_S and R_L are given, the Q of the network and thus bandwidth is defined; with the L network, the designer does not have a choice of circuit Q.

Example 3.4.1: Matching Network Design

Design a circuit to match a $100\,\Omega$ source to a $1700\,\Omega$ load at 900 MHz. Assume that a DC voltage must also be transferred from the source to the load.

Solution

Here $R_S < R_L$, and so the topology of Figure 3.4.3(a) can be used and there are two versions, one with a series inductor and one with a series capacitor. The series inductor version (see Figure 3.4.3(b)) is chosen as this enables DC bias to be applied. From Equations (3.4.4)–(3.4.6) the design equations are

$$|Q_S| = |Q_P| = \sqrt{\frac{1700}{100} - 1} = \sqrt{16} = 4, \quad \frac{X_S}{R_S} = 4, \quad \text{and} \quad X_S = 4 \cdot 100 = 400 \tag{3.4.7}$$

This indicates that $\omega L = 400\,\Omega$, and so the series element is

$$L = \frac{400}{2\pi \cdot 9 \cdot 10^8} = 70.7 \text{ nH} \tag{3.4.8}$$

For the shunt element next to the load, $|R_L/X_C| = 4$, and so

$$|X_C| = \frac{R_L}{4} = \frac{1700}{4} = 425 \tag{3.4.9}$$

Thus $1/\omega C = 425$ and

$$C = \frac{1}{2\pi \cdot 9 \cdot 10^8 \cdot 425} = 0.416 \text{ pF} \tag{3.4.10}$$

The final matching network design is shown in Figure 3.4.3(c).

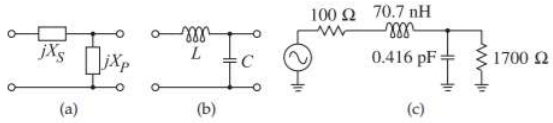


Figure 3.4.3: Matching network development for Example 3.4.1.

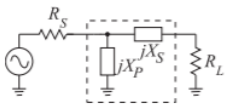


Figure 3.4.4: Two-element matching network topology for $R_S > R_L$.

3.4.2: L Network Design for $R_S > R_L$

For $R_S > R_L$, the topology shown in Figure 3.4.2 is used. The design equations for the L network for $R_S > R_L$ are similarly derived and are

$$|Q_S| = |Q_P| = \sqrt{\frac{R_S}{R_L} - 1} \tag{3.4.11}$$

$$-Q_S = Q_P, \quad Q_S = \frac{X_S}{R_L}, \quad \text{and} \quad Q_P = \frac{R_S}{X_P} \tag{3.4.12}$$

Example 3.4.2: Two-Element Matching Network

Design a passive two-element matching network that will achieve maximum power transfer from a source with an impedance of $50\,\Omega$ to a load with an impedance of $75\,\Omega$. Choose a matching network that will not allow DC to pass.

Solution

$R_L > R_S$, so, from Figure 3.4.1, the appropriate matching network topology is

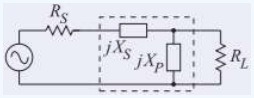


Figure 3.4.5

This topology can be either high pass or low pass depending on the choice of X_S and X_P. Design proceeds by finding the magnitudes of X_S and X_P. In two-element matching the circuit Q is fixed. With $R_L = 75\,\Omega$ and $R_S = 50\,\Omega$.

The Q of the matching network is the same for the series and parallel elements:

$$|Q_S| = \frac{|X_S|}{R_S} = \sqrt{\frac{R_L}{R_S} - 1} = 0.7071 \quad \text{and} \quad |Q_P| = \frac{R_L}{|X_P|} = |Q_S| = 0.7071$$

therefore $|X_S| = R_S \cdot |Q_S| = 50 \cdot 0.7071 = 35.35\,\Omega$. Also

$$|X_P| = R_L/|Q_P| = 75/0.7071 = 106.1\,\Omega$$

Specific element types can now be assigned to X_S and X_P, and note that they must be of opposite type.

The lowpass matching network is

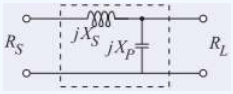


Figure 3.4.6

$$X_S = +35.5\,\Omega, \quad X_P = -106.1\,\Omega$$

The highpass matching network is

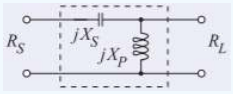


Figure 3.4.7

$$X_S = -35.35\,\Omega, \quad X_P = +106.1\,\Omega$$

This highpass design satisfies the design criterion that DC is not passed, as DC is blocked by the series capacitor.

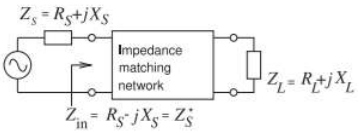


Figure 3.4.8: A matching network matching a complex load to a source with a complex Thevenin impedance.

https://eng.libretexts.org/@go/page/108249

This page titled 3.4: The L Matching Network is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.4: The L Matching Network** by Michael Steer is licensed CC BY-NC 4.0.

3.5: Dealing with Complex Loads

This section presents strategies for dealing with complex loads. In the algorithmic matching approach design proceeds first by ignoring the complex load and source and then accounting for them either through topology choice or canceling their effect through resonance.

3.5.1: Matching

Input and output impedances of transistors, mixers, antennas, etc., contain both resistive and reactive components. Thus a realistic impedance matching problem looks like that shown in Figure 10.4.8. The matching approaches that were presented in the previous sections can be directly applied if X_S and X_L are treated as stray reactances that need to be either canceled or, ideally, used as part of the matching network. There are two basic approaches to handling complex impedances:

a. Absorption: Absorb source and load reactances into the impedance matching network itself. This is done through careful placement of each matching element such that capacitors are placed in parallel with source and load capacitances, and inductors in series with source and load inductances. The stray values are then subtracted from the L and C values for the matching network calculated on the basis of the resistive parts of Z_S and Z_L only. The new (smaller) values, L' and C', constitute the elements of the matching network. Sometimes it is necessary to perform a series-to-parallel, or parallel-to-series, conversion of the source or load impedances so that the reactive elements are in the correct series or shunt arrangement for absorption.
b. Resonance: Resonate source and load reactances with an equal and opposite reactance at the frequency of interest.

The presence of reactance in a load indicates energy storage, and therefore bandwidth limiting. In the above approaches to handling a reactive load, the resonance approach could easily result in a narrowband matching solution. The major problem in matching is often to obtain sufficient bandwidth. What is sufficient will vary depending on the application. To maximize bandwidth the general goal is to minimize the total energy storage. Roughly the total energy stored will be proportional to the sum of the magnitudes of the reactances in the circuit. Of course, the actual energy storage depends on the voltage and current levels, which will themselves vary in the circuit. A good approach leading to large bandwidths is to incorporate the load reactance into the matching network. Thus the choice of appropriate matching network topology is critical. However, if the source and load reactance value is larger than the calculated matching network element value, then absorption on its own cannot be used. In this situation resonance must be combined with absorption. The majority of impedance matching designs are based on a combination of resonance and absorption.

Example 3.5.1: Matching Network Design Using Resonance

For the configuration shown in Figure 3.5.1, design an impedance matching network that will block the flow of DC current from the source to the load. The frequency of operation is 1 GHz. Design the matching network, neglecting the presence of the 10 pF capacitance at the load. Since $R_S = 50\ \Omega < R_L = 500\ \Omega$, and from Figure 10.4.1, consider the topologies of Figures 3.5.2 (a) and 3.5.2(b). The design criterion of blocking flow of DC from the source to the load narrows the choice to the topology of Figure 3.5.2(b).

Solution

Step 1:

$$|Q_S| = \left|\frac{X_S}{R_S}\right| = |Q_P| = \left|\frac{R_L}{X_P}\right| = \sqrt{\frac{R_L}{R_S} - 1} = 3 \quad \text{and} \quad Q_P = \frac{R_L}{X_P} \tag{3.5.1}$$

So $X_P = \omega L = R_L/Q_P = 500/3$. Reducing this gives

$$\omega L = \frac{500}{3}, \quad \text{and so} \quad L = \frac{500}{3 \times 2\pi \times 10^9} = 26.5\ \text{nH} \tag{3.5.2}$$

Similarly $-X_S/R_S = 3$ and so

$$\frac{1/(\omega C)}{R_S} = 3 \quad \text{or} \quad C = \frac{1}{3\omega R_S} = \frac{1}{3 \times 2\pi \times 10^9 \times 50} = 1.06\ \text{pF} \tag{3.5.3}$$

Step 2:

Resonate the 10 pF capacitor using an inductor in parallel:

$$(\omega L')^{-1} = \omega \times 10 \times 10^{-12} \tag{3.5.4}$$

$$L' = \frac{1}{\omega^2 10^{-11}} = \frac{1}{(2\pi)^2 10^{18} \times 10^{-11}} = 2.533 \text{ nH} \tag{3.5.5}$$

Thus Figure 3.5.3(c) is the required matching network. Two inductors are in parallel and the circuit can be simplified to that shown in Figure 3.5.4, where

$$L_X = (26.5 \text{ nH} \parallel 2.533 \text{ nH}) = \frac{2.533 \times 26.5}{2.533 + 26.5} \text{ nH} = 2.312 \text{ nH} \tag{3.5.6}$$

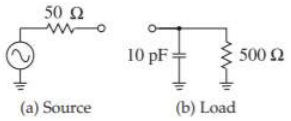


Figure 3.5.1: Matching problem considered in Example 3.5.1.

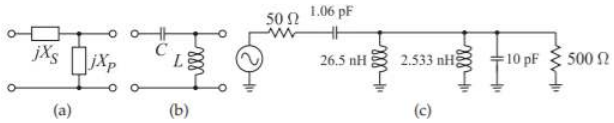


Figure 3.5.2: Matching network topology used in Example 3.5.1: a) and (b) topology; and (c) intermediate matching network.

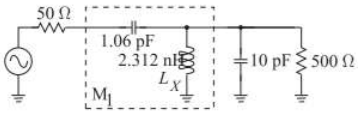


Figure 3.5.3: Final matching network in Example 3.5.1.

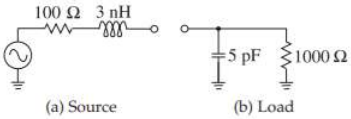


Figure 3.5.4: Matching problem in Example 3.5.2.

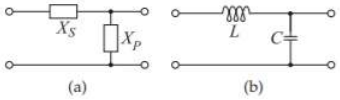


Figure 3.5.5: Topologies referred to in Example 3.5.2.

Example 3.5.2: Matching Network Design Using Resonance and Absorption

For the source and load configurations shown in Figure 3.5.4, design a lowpass impedance matching network at $f = 1$ GHz.

Solution

Since $R_S < R_L$, use the topology shown in Figure 3.5.5(a). For a lowpass response, the topology is that of Figure 3.5.5(b). Notice that absorption is the natural way of handling the 3 nH at the source and the 5 pF at the load. The design process is as follows:

Step 1:

Design the matching network, neglecting the reactive elements at the source and load:

$$|Q_S| = |Q_P| = \sqrt{\frac{R_L}{R_S} - 1} = \sqrt{10 - 1} = 3 \tag{3.5.7}$$

$$\frac{X_S}{R_S} = 3, \quad X_S = 3 \times 100, \quad \omega L = 300 \quad \text{and} \quad L = \frac{300}{2\pi \times 10^9} = 47.75 \text{ nH} \tag{3.5.8}$$

$$\frac{R_P}{X_P} = -3 \quad \text{and} \quad \frac{1000}{-(1/\omega C)} = -3 \quad \text{and} \quad C = \frac{3}{1000 \times 2\pi \times 10^9} = 0.477 \text{ pF} \tag{3.5.9}$$

This design is shown in Figure 3.5.6(a). This is the matching network that matches the $100\ \Omega$ source resistance to the $1000\ \Omega$ load with the source and load reactances ignored.

Step 2:

Figure 3.5.6(b) is the interim matching solution. The source inductance is absorbed into the matching network, reducing the required series inductance of the matching network. The capacitance of the load cannot be fully absorbed. The design for the resistance-only case requires a shunt capacitance of 0.477 pF, but 5 pF is available from the load. Thus there is an excess capacitance of 4.523 pF that must be resonated out by the inductance L'':

$$\frac{1}{\omega L''} = \omega 4.523 \times 10^{-12}. \quad \text{So} \quad L'' = \frac{1}{(2\pi)^2 \times 10^{18} \times 4.523 \times 10^{-12}} = 5.600 \text{ nH} \tag{3.5.10}$$

The final matching network design (Figure 3.5.6(c)) fully absorbs the source inductance into the matching network, but only partly absorbs the load capacitance.

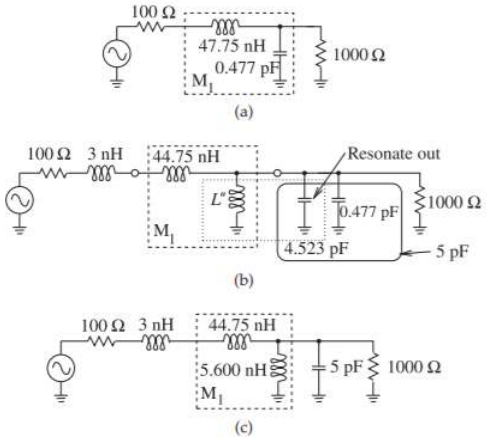


Figure 3.5.6: Evolution of the matching network in Example 3.5.2: (a) matching network design considering only the source and load resistors; (b) matching network with the reactive parts of the source and load impedances included; and (c) final design.

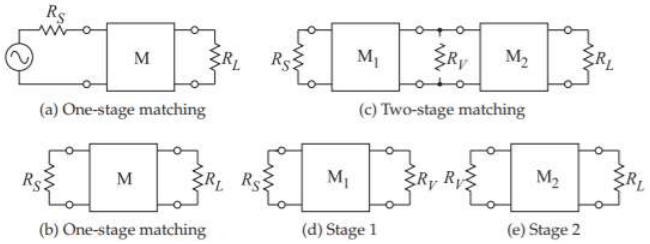


Figure 3.5.7: Matching in stages: (a) matching network M matching R_S to R_L; (b) without an explicit source; (c) two-stage matching with a virtual resistor R_V; (d) matching R_S to R_V; and (e) matching R_V to R_L.

This page titled 3.5: Dealing with Complex Loads is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.5: Dealing with Complex Loads** by Michael Steer is licensed CC BY-NC 4.0.

3.6: Dealing with Complex Loads

The bandwidth of a matching network can be controlled by using multiple matching stages either making the matching bandwidth wider or narrower. This concept is elaborated on in this and several design approaches presented.

3.6.1: Design Concept for Manipulating Bandwidth

The concept for manipulating matching network bandwidth is to do the matching in stages as shown in Figure 3.5.7. Figure 3.5.7(a) shows the one-stage matching problem using the common identification of the matching network as 'M'. The one-stage matching problem is repeated in Figure 3.5.7 without explicitly showing the source generator.

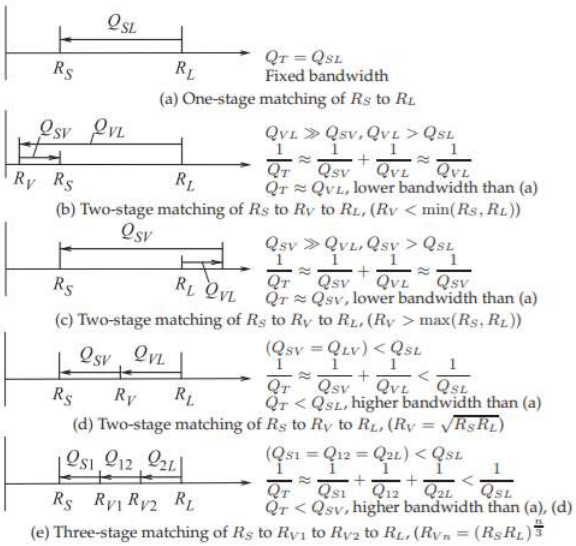


Figure 3.6.1: Effect of multi-stage matching on total circuit Q, Q_T, and matching bandwidth (which is approximately inversely proportional to Q_T.)

A two-stage matching problem is shown in Figure 3.5.7(c) with the introduction of a virtual resistor R_V between the first, M_1, and second, M_2, stage matching networks. R_V is shown as a virtual connection as it is not actually inserted in the circuit. Instead this is a short-hand way of indicating the matching problem to be done in two stages as shown in Figure 3.5.7(d and e) with the first stage matching the source resistor R_S to R_V and the second stage matching R_V to the load resistor R_L. After M_1 has been designed the resistance looking into the right-hand port of M_1, see Figure 3.5.7(d), will be R_V so R_V is the Thevenin equivalent source resistance to M_2. Similarly the input impedance looking into the left-hand port of M_2 is R_V so R_V is the effective load resistor of M_1. Of course these are the impedances at the center frequency and away from the center frequency of the match the input impedances will be complex.

The concept behind multi-stage matching network design is shown in Figure 3.6.1 where the standard one-stage match is shown in Figure 3.6.1(a). While this is shown for $R_L > R_S$ the concept holds for $R_L < R_S$. The arrows follow the that design begins with the load and ends at the source. With the one stage match the circuit Q is fixed and designated here as the total circuit Q, Q_T being the same as the Q of the one-stage R_L to R_S matching network, Q_{SL}. The two-stage match that reduces bandwidth (compared to the one-stage match) is shown in Figure 3.6.1(b). The total Q, Q_T, of the second stage is higher than for the one-stage design because the ratio of R_L to R_V is greater than the ratio of R_V to R_S. Bandwidth can also be reduced relative to the one-stage match by assigning R_V to be greater than both R_L

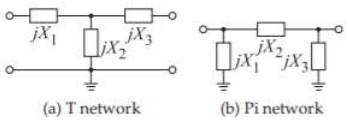


Figure 3.6.2: Two three-element matching networks.

and R_S, see Figure 3.6.1.

Choosing R_V to be between R_S and R_L will result in a circuit with lower Q_T and the bandwidth of the match will increase, see Figure 3.6.1(d). The maximum bandwidth for a two-stage match is to choose R_V as the geometric mean of R_S and R_L. This concept can be extended to multiple stages as shown for a three-stage match in Figure 3.6.1(e).

This section presents various matching network designs for manipulating bandwidth and all are based on the concept of choosing a virtual resistor.

3.6.2: Three-Element Matching Networks

With the L network (i.e., two-element matching), the circuit Q is fixed once the source and load resistances, R_S and R_L, are fixed:

$$Q = \sqrt{\frac{R_L}{R_S} - 1}, \quad (R_L > R_S) \tag{3.6.1}$$

Thus the designer does not have a choice of circuit Q. Breaking the matching problem into parts enables the circuit Q to be controlled. Introducing a third element in the matching network provides the extra degree of freedom in the design for adjusting Q, and hence bandwidth.

Two three-element matching networks, the T network and the Pi network, are shown in Figure 3.6.2. Which network is used depends on

a. the realization constraints associated with the specific design, and
b. the nature of the reactive parts of the source and load impedances and whether they can be used as part of the matching network.

The three-element matching network comprises 2 two-element (or L) matching networks and is used to increase the overall Q and thus narrow bandwidth. Given R_S and R_L, the circuit Q established by an L matching network is the minimum circuit Q available in the three-element matching arrangement. With three-element matching, the Q can only increase, so three-element matching is used for narrowband (high-Q) applications. However, lower Q can be obtained with more than three elements. The next subsections consider matching with more than three elements.

3.6.3: The Pi Network

The Pi network may be thought of as two back-to-back L networks that are used to match the load and the source to a virtual resistance, R_V, placed at the junction between the two networks, as shown in Figure 3.6.3(b). The design of each section of the Pi network is as for the L network matching. R_S is matched to R_V and R_V is matched to R_L.

R_V must be selected smaller than R_S and R_L since it is connected to the series arm of each L section. Furthermore, R_V can be any value that is smaller than the smaller of $R_S,\ R_L$. However, it is customarily used as the design parameter for specifying the desired Q.

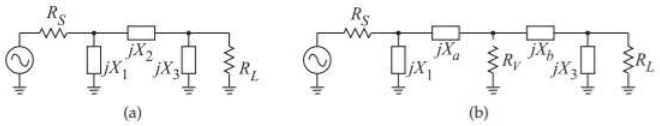


Figure 3.6.3: Pi matching networks: (a) view of a Pi network; and (b) as two back-to-back L networks with a virtual resistance, R_V, between the networks.

As a useful design approximation, the loaded Q of the Pi network can be taken as the Q of the L section with the highest Q:

$$Q = \sqrt{\frac{\max(R_S,\ R_L)}{R_V} - 1} \tag{3.6.2}$$

Given $R_S,\ R_L$, and Q, the above equation yields the value of R_V.

Example 3.6.1: Three-Element Matching Network Design

Design a Pi network to match a $50\ \Omega$ source to a $500\ \Omega$ load. The desired Q is 10. A suitable matching network topology is shown in Figure 3.6.4 together with the virtual resistance, R_V, to be used in design.

Solution

$R_S = 50\ \Omega$ and $R_L = 500\ \Omega$ so $\max(R_S, R_L) = 500\ \Omega$ and so the virtual resistor is

$$R_V = \frac{\max(R_S,\ R_L)}{Q^2+1} = \frac{500}{101} = 4.95\ \Omega \tag{3.6.3}$$

Design proceeds by separately designing the L networks to the left and right of R_V. For the L network on the left,

$$Q_{\text{left}} = \sqrt{\frac{50}{4.95} - 1} = 3.017 \quad \text{so} \quad Q_{\text{left}} = \frac{|X_a|}{R_V} = \frac{R_S}{|X_1|} = 3.017 \tag{3.6.4}$$

Note that X_1 and X_a must be of opposite types (one is capacitive and the other is inductive). The left L network has elements

$$|X_a| = 14.933\ \Omega \quad \text{and} \quad |X_1| = 16.6\ \Omega \tag{3.6.5}$$

For the L network on the right of R_V,

$$Q_{\text{right}} = Q = 10, \quad \text{thus} \quad \frac{|X_b|}{R_V} = \frac{R_L}{X_3} = 10 \tag{3.6.6}$$

$X_b,\ X_3$ are of opposite types, and

$$|X_b| = 49.5\ \Omega \quad \text{and} \quad |X_3| = 50\ \Omega \tag{3.6.7}$$

The resulting Pi network is shown in Figure 3.6.5 with the values

$$|X_1| = 16.6\ \Omega, \quad |X_3| = 50\ \Omega, \quad |X_a| = 14.933\ \Omega, \quad \text{and} \quad |X_b| = 49.5\ \Omega \tag{3.6.8}$$

Note that the pair $X_a,\ X_1$ are of opposite types and similarly $X_b,\ X_3$ are of opposite types. So there are four possible realizations, as shown in Figure 3.6.6.

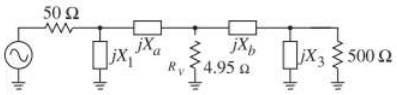


Figure 3.6.4: Matching network problem of Example 3.6.1.

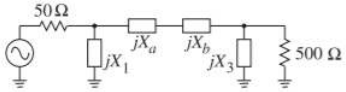


Figure 3.6.5: Final matching network in Example 3.6.1.

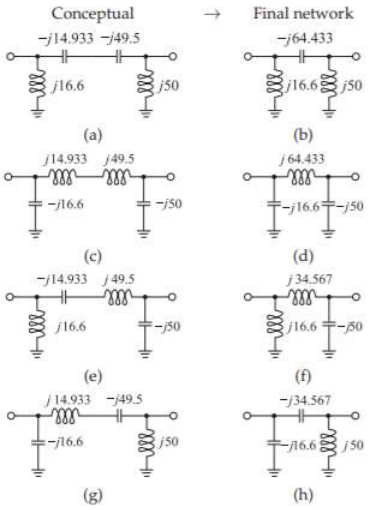


Figure 3.6.6: Four possible Pi matching networks: (a), (c), (e), and (g) conceptual circuits; and (b), (d), (f), and (h), respectively, their final reduced Pi networks.

In the previous example there were four possible realizations of the three-element matching network, and this is true in general. The specific choice of one of the four possible realizations will depend on specific application-related factors such as

a. elimination of stray reactances,
b. the need to pass or block DC current, and
c. the need for harmonic filtering.

It is fortunate that it may be possible to achieve multiple functions with the same network.

Example 3.6.2: Three-Element Matching with Reactive Source and Load

Design a Pi network to match the source to the load shown. The design frequency is 900 MHz and the desired Q is 10.

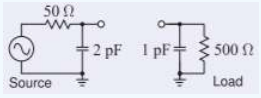


Figure 3.6.7

Solution

The design objective is to arrive at an overall network which has a Q of 10. To achieve this it is necessary to absorb the source and load reactances into the matching network. If they were resonated instead, the overall Q of the network can be expected to higher than the Q of the L matching network on its own.

Design begins by considering the matching of $R_S = 50\ \Omega$ to $R_L = 500\ \Omega$. Since the Q is specified, three (or more) matching elements must be used. The design starting point is shown on the right:

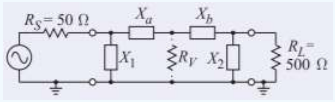


Figure 3.6.8

The virtual resistor $R_V = \max(R_s,\ R_L)/(1 + Q_2) = (500\ \Omega)/(1 + 100) = 4.95\ \Omega$. The left subnetwork with X_1 and X_a has $Q_{\text{LEFT}} = \sqrt{R_S/R_V - 1} = \sqrt{50/4.95 - 1} = 3.017$. The right subnetwork with X_2 and X_b has $Q_{\text{RIGHT}} = \sqrt{R_L/R_V - 1} = \sqrt{500/4.95 - 1} = 10.001$.

Note

Note that Q_{RIGHT} is almost exactly the desired Q of the network and Q_{LEFT} will have little effect on the Q of the overall circuit. Now $Q_{\text{LEFT}} = |X_a|/R_V = R_S/|X_1|$, so $|X_a| = 14.9\ \Omega$ and $|X_1| = 16.57\ \Omega$. $Q_{\text{RIGHT}} = |X_b|/R_V = R_S/|X_2|$, so $|X_b| = 49.5\ \Omega$ and $|X_2| = 50.0\ \Omega$.

X_1 must be chosen to be a capacitor $C_1 = 10.67$ pF so that the 2 pF source capacitance can be absorbed. Similarly X_2 is a capacitor $C_2 = 3.53$ pF. X_a and X_b are both inductors that combine in series for a total inductance $L_3 = 11.38$ nH. This leads to the final design shown below where the matching network is in the dashed box.

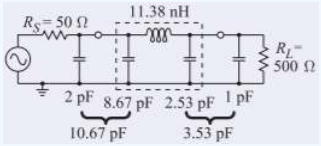


Figure 3.6.9

3.6.4: Matching Network Q Revisited

To demonstrate that the circuit Q established by an L matching network is the minimum circuit Q for a network having at most three elements, consider the design equations for $R_S > R_L$. Referring to Figure 3.6.10,

$$X_1 = \frac{R_S}{Q}, \quad X_3 = R_L\left(\frac{R_S/R_L}{Q^2+1-R_S/R_L}\right)^{\frac{1}{2}}, \quad X_2 = \frac{QR_S + R_S R_L/X_3}{Q^2+1} \tag{3.6.9}$$

Notice that the denominator of X_3 can be written as

$$Q^2 + 1 - \frac{R_S}{R_L} = \left(Q + \sqrt{\frac{R_S}{R_L} - 1}\right)\left(Q - \sqrt{\frac{R_S}{R_L} - 1}\right) \tag{3.6.10}$$

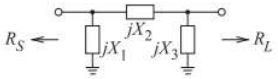


Figure 3.6.10: A three-element matching network.

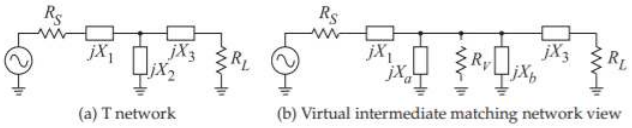


Figure 3.6.11: T network design approach.

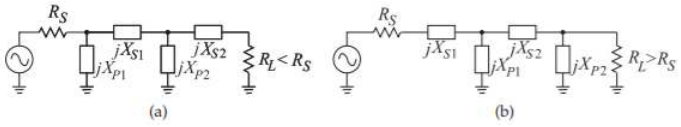


Figure 3.6.12: Broadband matching networks.

Then for a real solution we must have

$$Q \geq \sqrt{\frac{R_S}{R_L} - 1} \tag{3.6.11}$$

and so $Q \geq Q_L$ network. For $Q = Q_{\text{L network}}$, $X_3 \to \infty$ and the Pi network reduces to an L network that has two elements. Thus it is not possible to have a lower Q with a three-element matching network than the Q of a two-element matching network. Thus a three-element matching network must have lower bandwidth than that of a two-element matching network.

3.6.5: The T Network

The T network may be thought of as two back-to-back L networks that are used to match the load and the source to a virtual resistance, R_V, placed at the junction between the two L networks (see Figure 3.6.11). RV must be selected to be larger than both R_S and R_L since it is connected to the shunt leg of each L section. R_V is chosen according to the equation

$$Q = \sqrt{\frac{R_V}{\min(R_S,\ R_L)} - 1} \tag{3.6.12}$$

where Q is the desired loaded Q of the network. Each L network is calculated in exactly the same manner as was done for the Pi network matching. That is, R_S is matched to R_V and R_V is matched to R_L. Once again there will be four possible designs for the T network, given $R_S,\ R_L$, and Q.

3.6.6: Broadband (Low Q) Matching

L network matching does not allow the circuit Q, and hence bandwidth, to be selected. However, Pi network and T network matching allows the circuit Q to be selected independent of the source and load impedances, provided that the chosen Q is larger than that which can be obtained with an L network. Thus the Pi and T networks result in narrower bandwidth designs.

One design solution for broadband matching is to use two (or more) series-connected L sections (see Figure 3.6.12). Design is still based on the concept of a virtual resistor, R_V, placed at the junction of the two L networks (as in Figure 3.6.13), but now R_V is chosen to be between R_S and R_L:

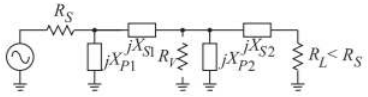


Figure 3.6.13: Matching network with two L networks.

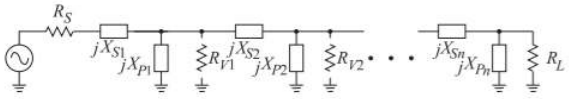


Figure 3.6.14: Cascaded L networks for broadband matching.

$$R_{\min} \le R_V \le R_{\max} \tag{3.6.13}$$

where $R_{\min} = \min(R_L,\ R_S)$ and $R_{\max} = \max(R_L,\ R_S)$. Then one of the two networks will have

$$Q_1 = \sqrt{\frac{R_V}{R_{\min}} - 1} \quad \text{and the other} \quad Q_2 = \sqrt{\frac{R_{\max}}{R_V} - 1} \tag{3.6.14}$$

The maximum bandwidth (minimum Q) available is obtained when

$$Q_1 = Q_2 = \sqrt{\frac{R_V}{R_{\min}} - 1} = \sqrt{\frac{R_{\max}}{R_V} - 1} \tag{3.6.15}$$

That is, the maximum matching bandwidth is obtained when R_V is the geometric mean of R_S and R_L:

$$R_V = \sqrt{R_L R_S} \tag{3.6.16}$$

Even wider bandwidths can be obtained by cascading more than two L networks, as shown in Figure 3.6.14. In this circuit

$$R_S < R_{V1} < R_{V2} \ldots < R_{Vn-1} < R_L \tag{3.6.17}$$

For optimum bandwidth the ratios should be equal,

$$\frac{R_{V_1}}{R_S} = \frac{R_{V_2}}{R_{V_1}} = \frac{R_{V_3}}{R_{V_2}} = \cdots = \frac{R_L}{R_{V_{n-1}}} \tag{3.6.18}$$

and the Q is given by

$$Q = \sqrt{\frac{R_{V_1}}{R_S} - 1} = \sqrt{\frac{R_{V_2}}{R_1} - 1} = \cdots = \sqrt{\frac{R_L}{R_{V_{n-1}}} - 1} \tag{3.6.19}$$

If there are N L networks used in the match, the maximum bandwidth will be obtained if the ith virtual resistor is

$$R_{V_i} = (R_S R_L)^{i/N}, \quad i = 1, \ldots, (N-1) \tag{3.6.20}$$

Example 3.6.3: Two-Section Matching Network Design

Consider matching a $10\ \Omega$ source to a $1000\ \Omega$ load using two L matching networks and designing for a Q of 3. How many matching sections are required?

Solution

Here the approximate Qs achieved with a single L matching network and with an optimum two-section design are compared. For a single L network design

$$Q = \sqrt{\frac{R_L}{R_S} - 1} = 9.95 \tag{3.6.21}$$

Now consider an optimum two-section design:

$$R_V = \sqrt{R_S R_L}; \quad Q_2 = \sqrt{\frac{R_L}{R_V} - 1} = \sqrt{\sqrt{\frac{R_L}{R_S}} - 1} = 3 \tag{3.6.22}$$

Thus the Q is 3 compared to the Q of an L section of 9.95. If the fractional bandwidth is inversely proportional to Q, then the bandwidth of the two-section design is $9.95/3 = 3.32$ times more than that of the L section.

Now consider how many sections are required to obtain a Q of 2:

$$(1 + Q^2) = \frac{R_{V_1}}{R_S} = \frac{R_{V_2}}{R_{V_1}} = \ldots = \frac{R_L}{R_{V_{n-1}}} \Rightarrow \tag{3.6.23}$$

$$(1 + Q^2)^n = \frac{R_L}{R_S} \Rightarrow n \ln(1 + Q^2) = \ln \frac{R_L}{R_S} \Rightarrow n = \frac{\ln(R_L/R_S)}{\ln(1 + Q^2)} \tag{3.6.24}$$

For $Q = 2$ and $R_L/R_S = 100$, $n = 2.86$, which rounds to $n = 3$, and three sections are required.

This page titled 3.6: Dealing with Complex Loads is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.6: Dealing with Complex Loads** by Michael Steer is licensed CC BY-NC 4.0.

3.7: Summary

This chapter presented techniques for impedance matching that achieve maximum power transfer from a source to a load. The simplest matching network uses a series and a shunt element, a two-element matching network, to realize a single-frequency match. This type of impedance matching network uses lumped elements and can be used up to a few gigahertz. Performance is limited by the self-resonant frequency of lumped elements and by their loss, particularly that of inductors. The shunt element can be replaced by a shunt stub, but in most transmission line technologies, including microstrip, the series element cannot be implemented as a stub. Matching networks can also be realized using transmission line segments only, principally shunt stubs and cascaded transmission lines. A tunable double-stub matching network, which uses two stubs separated by a transmission line, is standard equipment in microwave laboratories and facilitates matching of a circuit under development.

The bandwidth of a matching network is set by the maximum allowable reflection coefficient of the terminated network. Two-element matching nearly always results in a narrow match and for typical communications applications often achieves acceptable matching over bandwidths of only 1%–3%. The most significant determinant of the quality of the match that can be achieved is the ratio of the source and load resistances, as well as the reactive energy storage of the source and load.

An important concept in matching network design is a technique for controlling bandwidth. The concept is based on matching to an intermediate resistance, typically designated as R_v. Compared to a two-element network, increased bandwidth is obtained if R_v is the geometric mean of the source and load resistances. This new network consists of two two-element matching networks. If R_v is greater or less than both the source and load resistances, then the bandwidth of the matching network is reduced. The matching network synthesis problem can also be addressed using filter design techniques, and this enables simultaneous control over the quality and bandwidth of the match. It is always a good idea to have no more bandwidth in the system than is needed, as this minimizes the propagation of noise.

This page titled 3.7: Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.7: Summary** by Michael Steer is licensed CC BY-NC 4.0.

3.8: References

[1] R. Collins, Foundations for Microwave Engineering. McGraw Hill, 1966.

[2] M. Steer, Microwave and RF Design, Transmission Lines, 3rd ed. North Carolina State University, 2019.

This page titled 3.8: References is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.8: References** by Michael Steer is licensed CC BY-NC 4.0.

3.9: Exercises

1. Consider the design of a magnetic transformer that will match the $3\,\Omega$ output resistance of a power amplifier (this is the source) to a $50\,\Omega$ load. The secondary of the transformer is on the load side.
   a. What is the ratio of the number of primary turns to the number of secondary turns for ideal matching?
   b. If the transformer ratio could be implemented exactly (the ideal situation), what is the reflection coefficient normalized to $3\,\Omega$ looking into the primary of the transformer with the $50\,\Omega$ load?
   c. What is the ideal return loss of the loaded transformer (looking into the primary)? Express your answer in dB.
   d. If there are 100 secondary windings, how many primary windings are there in your design? Note that the number of windings must be an integer. (This practical situation will be considered in the rest of the problem.)
   e. What is the input resistance of the transformer looking into the primary?
   f. What is the reflection coefficient normalized to $3\,\Omega$ looking into the primary of the transformer with the $50\,\Omega$ load?
   g. What is the actual return loss (in dB) of the loaded transformer (looking into the primary)?
   h. If the maximum available power from the amplifier is $20\ \mathrm{dBm}$, how much power (in dBm) is reflected at the input of the transformer?
   i. Thus, how much power (in dBm) is delivered to the load ignoring loss in the transformer?
2. Consider the design of a magnetic transformer that will match a $50\,\Omega$ output resistance to the $100\,\Omega$ load presented by an amplifier. The secondary of the transformer is on the load (amplifier) side.
   a. What is the ratio of the number of primary turns to the number of secondary turns for ideal matching?
   b. If the transformer ratio could be implemented exactly (the ideal situation), what is the reflection coefficient normalized to $50\,\Omega$ looking into the primary of the transformer with the load?
   c. What is the ideal return loss of the loaded transformer (looking into the primary)? Express your answer in dB.
   d. If there are 20 secondary windings, how many primary windings are there in your design? Note that the number of windings must be an integer? (This situation will be considered in the rest of the problem.)
   e. What is the input resistance of the transformer looking into the primary?
   f. What is the reflection coefficient normalized to $50\,\Omega$ looking into the primary of the loaded transformer?
   g. What is the actual return loss (in dB) of the loaded transformer (looking into the primary)?
   h. If the maximum available power from the source is $-10\ \mathrm{dBm}$, how much power (in dBm) is reflected from the input of the transformer?
   i. Thus, how much power (in dBm) is delivered to the amplifier ignoring loss in the transformer?
3. Consider the design of an L-matching network centered at $1\ \mathrm{GHz}$ that will match the $2\,\Omega$ output resistance of a power amplifier (this is the source) to a $50\,\Omega$ load. [Parallels Example 10.4.1 but note the DC blocking requirement below.]
   a. What is the Q of the matching network?
   b. The matching network must block DC current. Draw the topology of the matching network.
   c. What is the reactance of the series element in the matching network?
   d. What is the reactance of the shunt element in the matching network?
   e. What is the value of the series element in the matching network?
   f. What is the value of the shunt element in the matching network?
   g. Draw and label the final design of your matching network including the source and load resistances.
   h. Approximately, what is the $3\ \mathrm{dB}$ bandwidth of the matching network?
4. Consider the design of an L-matching network centered at $100\ \mathrm{GHz}$ that will match a source with a Thevenin resistance of $50\,\Omega$ to the input of an amplifier presenting a load resistance of $100\,\Omega$ to the matching network. [Parallels Example 10.4.2 but note the DC blocking requirement below.]
   a. What is the Q of the matching network?
   b. The matching network must block DC current. Draw the topology of the matching network.
   c. What is the reactance of the series element in the matching network?
   d. What is the reactance of the shunt element in the matching network?
   e. What is the value of the series element in the matching network?

f. What is the value of the shunt element in the matching network?
g. Draw and label the final design of your matching network including the source and load resistance.
h. Approximately, what is the 3 dB bandwidth of the matching network?

5. Design a Pi network to match the source configuration to the load configuration below. The design frequency is 900 MHz and the desired Q is 10. [Parallels Example 10.6.2]

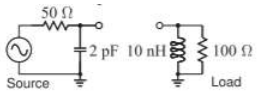


Figure 3.9.1

6. Design a Pi network to match the source configuration to the load configuration below. The design frequency is 900 MHz and the desired Q is 10. [Parallels Example 10.6.2]

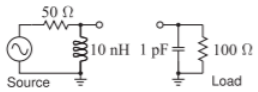


Figure 3.9.2

7. Develop the electrical design of an L-matching network to match the source to the load below.

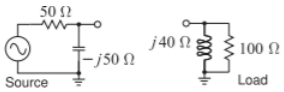


Figure 3.9.3

8. Develop the electrical design of an L-matching network to match the source to the load below.

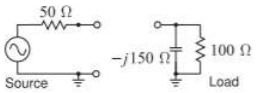


Figure 3.9.4

9. Design a lowpass lumped-element matching network to match the source and load shown below. The design frequency is 1 GHz. You must use a Smith Chart and clearly show your working and derivations. You must develop the final values of the elements.

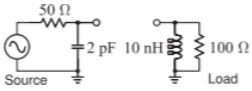


Figure 3.9.5

10. Consider the design of an L-matching network centered at 100 GHz that will match a source with a Thevenin resistance of $50\ \Omega$ to the input of an amplifier presenting a load resistance of $200\ \Omega$ to the matching network. [Parallels Example 10.4.2 but note the DC blocking requirement below.]
    a. What is the Q of the matching network?
    b. The matching network must block DC current. Draw the topology of the matching network.
    c. What is the reactance of the series element in the matching network?
    d. What is the reactance of the shunt element in the matching network?
    e. What is the value of the series element in the matching network?
    f. What is the value of the shunt element in the matching network?
    g. Draw and label the final design of your matching network including the source and load resistance.
    h. Approximately, what is the 3 dB bandwidth of the matching network?
11. Design a two-element matching network to interface a source with a $25\ \Omega$ Thevenin equivalent impedance to a load consisting of a capacitor in parallel with a resistor so that the load admittance is $Y_L = 0.02 + j0.02$ S. Use the absorption method to handle

the reactive load.

12. Design a matching network to interface a source with a $25\ \Omega$ Thevenin equivalent impedance to a load consisting of a capacitor in parallel with a resistor so that the load admittance is $Y_L = 0.01 + j0.01$ S.
    a. If the complexity of the matching network is not limited, what is the minimum Q that could possibly be achieved in the complete network consisting of the matching network and the source and load impedances?
    b. Outline the procedure for designing the matching network for maximum bandwidth if only four elements can be used in the network. You do not need to design the network,
13. Design a Pi network to match the source configuration to the load configuration below. The design frequency is 900 MHz and the desired Q is 10.

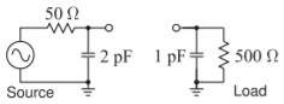


Figure 3.9.6

14. Design a passive matching network that will achieve maximum bandwidth matching from a source with an impedance of $2\ \Omega$ (typical of the output impedance of a power amplifier) to a load with an impedance of $50\ \Omega$. The matching network can have a maximum of three reactive elements. You need only calculate reactances and not the capacitor and inductor values.
15. Design a passive matching network that will achieve maximum bandwidth matching from a source with an impedance of $20\ \Omega$ to a load with an impedance of $125\ \Omega$. The matching network can have a maximum of four reactive elements. You need only calculate reactances and not the capacitor and inductor values.
    a. Will you use two, three, or four elements in your matching network?
    b. With a diagram, and perhaps equations, indicate the design procedure.
    c. Design the matching network. It is sufficient to use reactance values.
16. Design a passive matching network that will achieve maximum bandwidth matching from a source with an impedance of $60\ \Omega$ to a load with an impedance of $5\ \Omega$. The matching network can have a maximum of four reactive elements. You need only calculate reactances and not the capacitor and inductor values.
    a. Will you use two, three, or four elements in your matching network?
    b. With a diagram and perhaps equations, indicate the design procedure.
    c. Design the matching network. It is sufficient to use reactance values.
17. Design a T network to match a $50\ \Omega$ source to a $1000\ \Omega$ load. The desired loaded Q is 15.
18. Repeat Example 10.3.2 with an inductor in series with the load. Show that the inductance can be adjusted to obtain any positive shunt resistance value.
19. Design a three-lumped-element matching network that interfaces a source with an impedance of $5\ \Omega$ to a load with an impedance consisting of a resistor with an impedance of $10\ \Omega$. The network must have a Q of 6.
20. A two-port matching network is shown below with a generator and a load. The generator impedance is $40\ \Omega$ and the load impedance is $Z_L = 50 - j20\ \Omega$. Use a Smith chart to design the matching network.

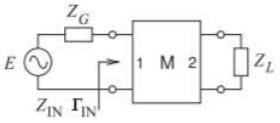


Figure 3.9.7

    a. What is the condition for maximum power transfer from the generator? Express your answer using impedances.
    b. What is the condition for maximum power transfer from the generator? Express your answer using reflection coefficients.
    c. What system reference impedance are you going to use to solve the problem?
    d. Plot Z_L on the Smith chart and label the point. (Remember to use impedance normalization if required.)
    e. Plot ZG on the Smith chart and label the point.
    f. Design a matching network using only transmission lines. Show your work on the Smith chart. You must express the lengths of the lines in terms of electrical length (either degrees or wavelengths). Characteristic impedances of the lines are

required. (You will therefore have a design that consists of one stub and one other length of transmission line.)

21. Use a lossless transmission line and a series reactive element to match a source with a Thevenin equivalent impedance of $25 + j50\ \Omega$ to a load of $100\ \Omega$. (That is, use one transmission line and one series reactance only.)
    a. Draw the matching network with the source and load.
    b. What is the value of the series reactance in the matching network (you can leave this in ohms)?
    c. What is the length and characteristic impedance of the transmission line?
22. Consider a load $Z_L = 80 + j40\ \Omega$. Use the Smith chart to design a matching network consisting of only two transmission lines that will match the load to a generator of $40\ \Omega$.
    a. Draw the matching network with transmission lines. If you use a stub, it should be a short-circuited stub.
    b. Indicate your choice of characteristic impedance for your transmission lines. What is the normalized load impedance? What is the normalized source impedance?
    c. Briefly outline the design procedure you will use. You will need to use Smith chart sketches.
    d. Plot the load and source on a Smith chart.
    e. Complete the design of the matching network, providing the lengths of the transmission lines.
23. A two-port matching network is shown below with a generator and a load. The generator impedance is $40\ \Omega$ and the load impedance is $Z_L = 20 - j50\ \Omega$. Use a Smith chart to design the matching network.

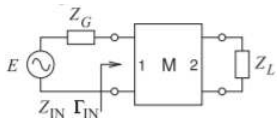


Figure 3.9.8

    a. What is the condition for maximum power transfer from the generator? Express your answer using impedances.
    b. What is the condition for maximum power transfer from the generator? Express your answer using reflection coefficients.
    c. What system reference impedance are you going to use to solve the problem?
    d. Plot Z_L on a Smith chart and label the point. (Remember to use impedance normalization if required.)
    e. Plot Z_G on a Smith chart and label the point.
    f. Design a matching network using only transmission lines and show your work on a Smith chart. You must express the lengths of the lines in terms of electrical length (either degrees or wavelengths long). Characteristic impedances of the lines are required. (You will therefore have a design that consists of one stub and one other length of transmission line.)
24. Use a Smith chart to design a microstrip network to match a load $Z_L = 100 - j100\ \Omega$ to a source $Z_S = 34 - j40\ \Omega$). Use transmission lines only and do not use short-circuited stubs. Use a reference impedance of $40\ \Omega$.
    a. Draw the matching network problem labeling impedances and the impedance looking into the matching network from the source as Z_1.
    b. What is the condition for maximum power transfer in terms of impedances?
    c. What is the condition for maximum power transfer in terms of reflection coefficients?
    d. Identify, i.e. draw, at least two suitable microstrip matching networks.
    e. Develop the electrical design of the matching network using the Smith chart using $40\ \Omega$ lines only. You only need do one design.
    f. Draw the microstrip layout of the matching network identify critical parameters such characteristic impedances and electrical length. Ensure that you identify which is the source side and which is the load side. You do not need to determine the widths of the lies or their physical lengths.
25. Repeat exercise 24 but now with $Z_L = 10 - j40\ \Omega$ and $Z_S = 28 - j28\ \Omega$).
26. Use a Smith chart to design a two-element lumped-element lossless matching network to interface a source with an admittance $Y_S = 6 - j12\ \text{mS}$ to a load with admittance $Y_L = 70 - j50\ \text{mS}$.
27. Use a Smith chart to design a two-element lumped-element lossless matching network to interface a load $Z_L = 50 + j50\ \Omega$ to a source $Z_S = 10\ \Omega$.

3.9.1: Exercises by Section

†challenging

§10.3 1, 2

§10.4 3, 4

§10.5 5, 6, 7, 8, 9, 10, 11†, 12†, 13†

§10.6 14†, 15†, 16†, 17†, 18†, 19†

§10.7 20†, 21†, 22†, 23†

§10.9 24, 25, 26, 27

3.9.2: Answers to Selected Exercises

15. (c) $Q = 1.22467$

18. $C = 1/(\omega_d^2 L_P)$

20. (d) $1.25 - j0.5$

21. (b) $-j50\ \Omega$

23. (f) $40\ \Omega$, $0.085\ \lambda$-long line before load, $40\ \Omega$, $0.076\ \lambda$-long shorted stub

---

This page titled 3.9: Exercises is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

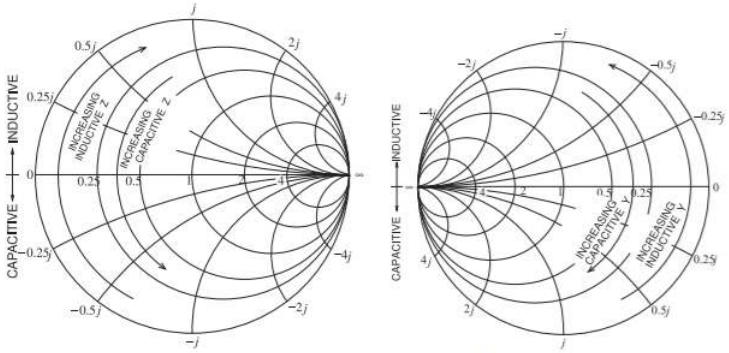
- **10.9: Exercises** by Michael Steer is licensed CC BY-NC 4.0.

3.10: Impedance Matching Using Smith Charts

The lumped-element matching networks presented up to now can also be developed using Smith charts which provide a fairly intuitive approach to network design. With experience it will be found that this is the preferred approach to developing designs, as trade-offs can be captured graphically. Smith chart-based design will be presented using examples.

3.10.1: Two-Element Matching

The examples here build on the preceding lumped-element matching network design and now use the Smith. Capacitive and inductive regions on the Smith chart are shown in Figure 3.10.1. In the design examples presented here, circles of constant resistance or constant conductance are followed and these correspond to varying reactance or susceptance, respectively.



(a) Impedance Smith chart (b) Admittance Smith chart

Figure 3.10.1: Inductive and capacitive regions on Smith charts. Increasing capacitive impedance (Z) indicates smaller capacitance; increasing inductive admittance (Y) indicates smaller inductance.

Example 3.10.1: Two-Element Matching Network Design Using a Smith Chart

Develop a two-element matching network to match a source with an impedance of $R_S = 25\ \Omega$ to a load $R_L = 200\ \Omega$ (see Figure 3.10.2).

Solution

The design objective is to present conjugate matched impedances to the source and load. However, since here the source and load impedances are real, the design objective is $Z_1 = R_S$ and $Z_2 = R_L$. The load and source resistances are plotted on the Smith chart in Figure 3.10.4(a) after choosing a normalization impedance of $Z_0 = 50\ \Omega$ (and so $r_S = R_S/Z_0 = 0.5$ and $r_L = R_L/Z_0 = 4$). The normalized source impedance, r_S, is Point A, and the normalized load impedance, r_L, is Point C. The matching network must be lossless, which means that the design must follow lines of constant resistance (on the impedance part of the Smith chart) or constant conductance (on the admittance part of the Smith chart). So Points A and C must be on the above circles and the circles must intersect if a design is possible. The design can be viewed as moving back from the source toward the load or moving back from the load toward the source. (The views result in identical designs.) Here the view taken is moving back from the source toward the load.

One possible design is shown in Figure 3.10.4(a). From Point A, the line of constant resistance is followed to Point B (there is increasing series reactance along this path). From Point B, the locus follows a line of constant conductance to the final point, Point C. There is also an alternative design that follows the path shown in Figure 3.10.4(b). There are only two designs that have a path from A to B following just two arcs. At this point two designs have been outlined. The next step is assigning element values.

The design shown in Figure 3.10.4(a) begins with r_S followed by a series reactance, x_S, taking the locus from A to B. Then a shunt capacitive susceptance, b_P, takes the locus from B to C and r_L. At Point A the reactance $x_A = 0$, at Point B the reactance $x_B = 1.323$. This value is read off the Smith chart, requiring that an arc as shown be interpolated between the arcs provided. It should be noted that not all versions of Smith charts include negative signs, as the chart becomes too complicated. Thus the user needs to be aware and add signs where appropriate. The normalized series reactance is

$$x_S = x_B - x_A = 1.323 - 0 = 1.323 \tag{3.10.1}$$

that is,

$$X_S = x_s Z_0 = 1.323 \times 50 = 66.1\ \Omega \tag{3.10.2}$$

A shunt capacitive element takes the locus from Point B to Point C and

$$b_P = b_C - b_B = 0 - (-0.661) = 0.661 \tag{3.10.3}$$

so

$$B_P = b_P / Z_0 = 0.661/50 = 13.22\ \text{mS} \quad \text{or} \quad X_P = -1/B_P = -75.6\ \Omega \tag{3.10.4}$$

The final design is shown in Figure 3.10.3.

One of the advantages of using the Smith chart is that the design progresses in stages, with the structure of the design developed before actual numerical values are calculated. Of course, it is difficult to extract accurate values from a chart, so designs are regularly roughed out on a Smith chart and refined using CAD tools. Example 3.10.1 matched a resistive source to a resistive load. The next example considers the matching of complex load and source impedances.

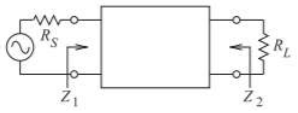


Figure 3.10.2: Design objectives for Example 3.10.1. $R_S = 15\ \Omega,\ R_L = 200\ \Omega$.

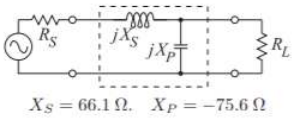


Figure 3.10.3: Final design for Example 3.10.1 using the path shown in Figure 3.10.4(a).

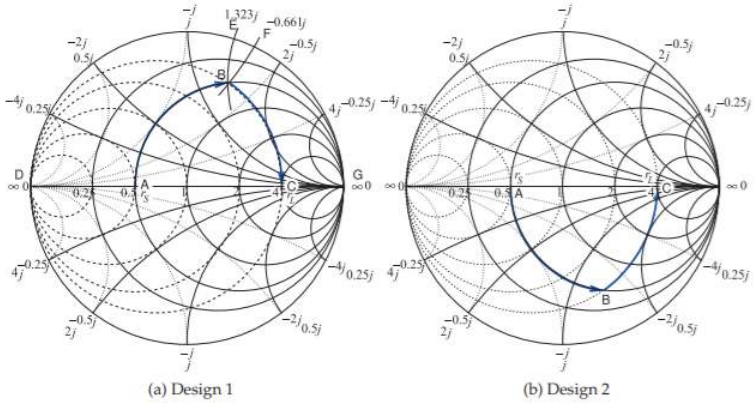


Figure 3.10.4: Alternative designs for Example 3.10.1. The normalization impedance is 50 Ω.

In the earlier algorithmic approach to matching network design absorption and resonance were introduced as strategies for dealing with complex terminations. Design was not always straightforward. This complication disappears with a Smith chart-based design, as it is conceptually not much different from the resistive problem of Example 3.10.1.

Example 3.10.2: Matching Network Design With Complex Impedances

Develop a two-element matching network to match a source with an impedance of $Z_S = 12.5 + 12.5j\ \Omega$ to a load $Z_L = 50 - 50j\ \Omega$, as shown in Figure 3.10.5.

Solution

The design objective is to present conjugate matched impedances to the source and load; that is, $Z_1 = Z_S^*$S and $Z_2 = Z_L^*$. The choice here is to design for Z_1; that is, elements will be inserted in front of Z_L to produce the impedance Z_1. The normalized source and load impedances are plotted in Figure 3.10.6(a) using a normalization impedance of $Z_0 = 50\,\Omega$, so $z_S = Z_S/Z_0 = 0.25 + 0.25j$ (Point S) and $z_L = Z_L/Z_0 = 1 - j$ (Point C).

The impedance to be synthesized is $z_1 = Z_1/Z_0 = z_S^* = 0.25 - 0.25j$ (Point A). The matching network must be lossless, which means that the lumped-element design must follow lines of constant resistance (on the impedance part of the Smith chart) or constant conductance (on the admittance part of the Smith chart). Points A and C must be on the above circles and the circles must intersect if a design is possible.

The design can be viewed as moving back from the load impedance toward the conjugate of the source impedance. The direction of the impedance locus is important. One possible design is shown in Figure 3.10.6(a). From Point C the line of constant conductance is followed to Point B (there is increasing positive [i.e., capacitive] shunt susceptance along this path). From Point B the locus follows a line of constant resistance to the final point, Point A.

The design shown in Figure 3.10.6(a) begins with a shunt susceptance, b_P, taking the locus from Point C to Point B and then a series inductive reactance, x_S, taking the locus to Point A. At Point C the susceptance $b_C = 0.5$, at Point B the susceptance $b_B = 1.323$. This value is read off the Smith chart, requiring that an arc of constant susceptance, as shown, be interpolated between the constant susceptance arcs provided. The normalized shunt susceptance is

$$b_P = b_B - b_C = 1.323 - 05 = 0.823 \tag{3.10.5}$$

that is,

$$B_P = b_P/Z_0 = 0.823/(50\,\Omega) = 16.5 \text{ mS or } X_P = -1/B_P = -60.8\,\Omega \tag{3.10.6}$$

A series reactive element takes the locus from Point B to Point A, so

$$x_S = x_A - x_B = -0.25 - (-0.661) = 0.411 \tag{3.10.7}$$

so

$$X_S = x_S Z_0 = 0.411 \times 50\,\Omega = 20.6\,\Omega \tag{3.10.8}$$

The final design is shown in Figure 3.10.7.

There are only two designs that have a path from Point C to Point A following just two arcs. In Design 1, shown in Figure 3.10.6(a), Path CBA is much shorter than Path CHA for Design 2 shown in Figure 3.10.6(b). The path length is an approximate indication of the total reactance required, and the higher the reactance, the greater the energy storage and hence the narrower the bandwidth of the design. (The actual relative bandwidth depends on the voltage and current levels in the network; the path length criteria, however, is an important rule of thumb.) Thus Design 1 can be expected to have a much

higher bandwidth than Design 2. Since designing broader bandwidth is usually an objective, a design requiring a shorter path on a Smith chart is usually preferable.

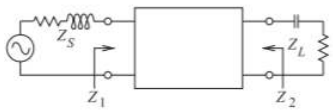


Figure 3.10.5: Design objectives for Example 3.10.2.

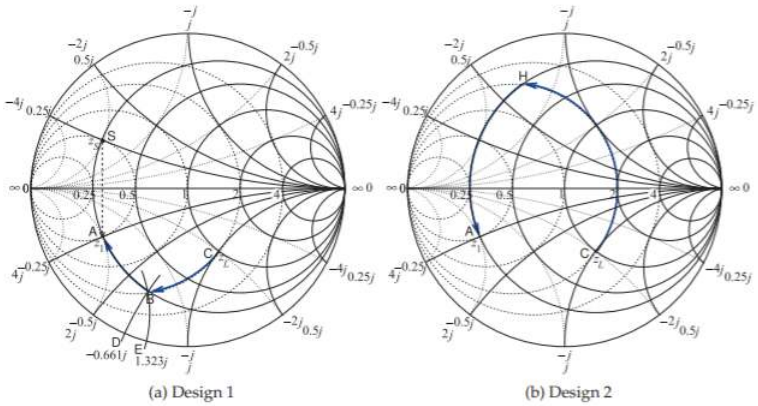


Figure 3.10.6: Smith chart-based designs used in Example 3.10.2. (50 Ω normalization used.)

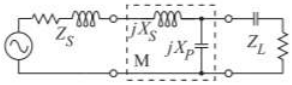


Figure 3.10.7: Final circuit for Design 1 of Example 3.10.2. $X_S = 20.6\,\Omega,\ X_P = -60.8\,\Omega$.

This page titled 3.10: Impedance Matching Using Smith Charts is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.10: Impedance Matching Using Smith Charts** by Michael Steer is licensed CC BY-NC 4.0.

3.11: Distributed Matching

Matching using lumped elements leads to series and shunt lumped elements. The shunt elements can be implemented using shunt transmission lines, as a short length (less than one-quarter wavelength long) of short-circuited transmission line looks like an inductor and a short section of open-circuited transmission line looks like a capacitor. However, in microstrip it is not possible to realize the series elements as lengths of transmission lines. The solution is to use lengths of transmission line together with shunt elements. If space is not at a premium, this is an optimum solution, as transmission lines have much lower loss than a lumped inductor. The series transmission lines rotate the reflection coefficient on the Smith chart.

As with all matching design, using transmission lines begins with a topology in mind. Several topologies are shown in Figure 3.11.1. Figure 3.11.1(a) is the top view of a microstrip matching network with a series transmission line and stub realized as an open-circuited transmission line. Figure 3.11.1(b) is a shorthand schematic for this circuit. Matching network design then becomes a problem of choosing the lengths and characteristic impedances of the lines.

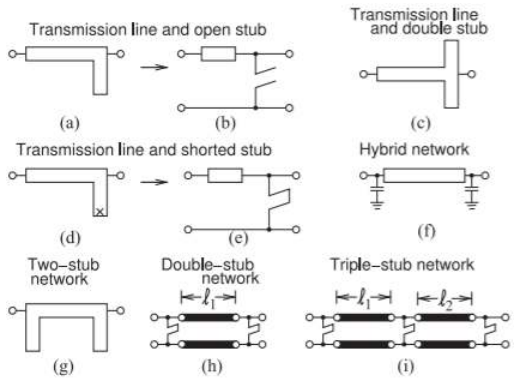


Figure 3.11.1: Matching networks with transmission line elements.

The stub here is used to realize a capacitive shunt element. This network corresponds to two-element matching with a shunt capacitor. The value of the shunt capacitance can be increased using a dual stub, as shown in Figure 3.11.1(c), where the capacitive input impedances of each stub are in parallel. The dual circuit to that in Figure 3.11.1(a) is shown in Figure 3.11.1(d) together with its schematic representation in Figure 3.11.1(e). This circuit has a short-circuited stub that realizes a shunt inductance.

Mixing lumped capacitors with a transmission line element, as shown in Figure 3.11.1(f), realizes a much more space-efficient network design. There are many variations to stub-based matching network design, including the two-stub design in Figure 3.11.1 (g).

A common situation encountered in the laboratory is the matching of circuits that are in development. Laboratory items available for matching include the **stub tuner**, shown in Figure 3.11.2(a), and the **double-stub tuner**, shown in Figure 3.11.2(b). With the double-stub tuner the length of the series transmission line is fixed, but stubs can have variable length using lengths of transmission lines with sliding short circuits. Not all impedances can be matched using a double stub tuner, however. A triple-stub tuner can match all impedances presented to it [1]. The **double-slug tuner** shown in Figure 3.11.1(c) has dielectric slugs each of which introduces a short section of lower impedance line. The slugs are moved up and down the line and avoid the rapid changes in impedances that occur with the stub tuners and as a result the double-slug tuner provides a broader bandwidth match than does the double stub tuner. The **slide-screw slug tuner** shown in Figure 3.11.1(d) can achieve a broadband match. Here a metal slug can be lowered into the slabline changing the impedance of a section of transmission line and mostly affects the magnitude of the reflection coefficient while moving the metal slug along the line mostly affects the phase. This is the type of tuner incorporated in computer-controlled automated tuners.

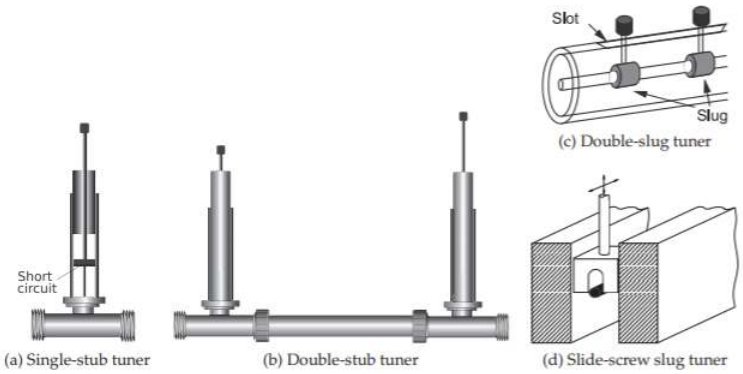


Figure 3.11.2: Laboratory tuners.

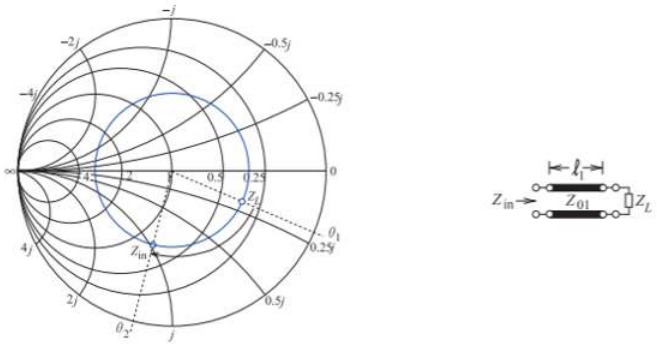


Figure 3.11.3: Rotation of the input impedance of a transmission line on a Smith chart normalized to Z_{01} as the line length increases.

3.11.1: Stub Matching

In this section matching using one series transmission line and one stub will be considered. This corresponds to the microstrip circuit topologies shown in Figures 3.11.1(a and d). First, consider the terminated transmission line shown in Figure 3.11.3. When the length, ℓ_1, of the line is zero, the input impedance of the line, Z_{in}, equals Z_L. How it changes is best described by considering the input reflection coefficient, Γ_{in}, of the line. If the reflection coefficient is normalized to Z_{01}, then the magnitude of Γ_{in} and its phase varies as twice the electrical length of the line. This situation is shown on the Smith chart in Figure 3.11.3, where Z_L is chosen arbitrarily.

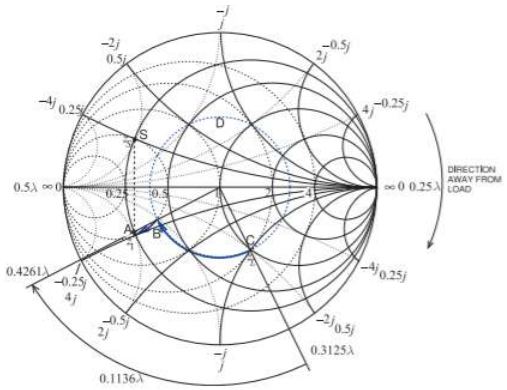


Figure 3.11.4: Design for Example 3.11.1.

The input reflection coefficient of the line rotates in a clockwise direction as the length of the line increases. One way of remembering this is to consider an open-circuited line. When the line length is zero, $Y_{\text{in}} = 0$ and $\Gamma_{\text{in}} = +1$. A short length of this line is capacitive so that its reflection coefficient will be in the bottom half of the Smith chart. A length of line can be used to rotate the impedance to an appropriate point to follow a line of constant conductance to the desired input impedance.

Example 3.11.1: Matching Network Design With a Transmission Line and a Single Stub

Design a two-element matching network to match a source with an impedance $Z_S = 12.5 + 12.5j\,\Omega$ to a load $Z_L = 50 - 50j\,\Omega$, as shown in Figure 10.7.5. This example repeats the design in Example 10.7.1, but now using a transmission line.

Solution

As in Example 10.7.1, choose $Z_0 = 50\ \Omega$ and the design path is from $z_L = Z_L/Z_0 = 1 - j$ to z_s^*, where $z_s = 0.25 + 0.25j$. One possible design solution is indicated in Figure 3.11.4. The line length, ℓ (taking the locus from Point C to Point B), is

$$\ell = 0.4261\lambda - 0.3125\lambda = 0.1136\lambda$$

and the normalized shunt susceptance, b_P (taking the locus from Point B to Point C), is

$$b_P = b_A - b_B = 2 - 1 = 1$$

Thus $X_P = (-1/b_P) \times 50\ \Omega = -50\ \Omega$. The final design is shown in Figure 3.11.2(a). The stub design of Figure 3.11.2(b) follows the procedure described in Example 6.5.1.

3.11.2: 3.11.2 Hybrid Lumped-Distributed Matching

A lossless matching network can have transmission lines as well as inductors and capacitors. If the system reference or normalization impedance is the characteristic impedance of a transmission line, then the locus of the input impedance (or reflection coefficient) of the line with respect to the length of

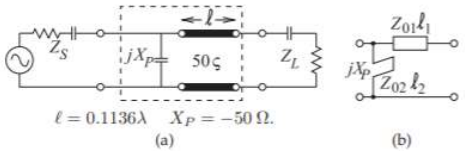


Figure 3.11.2: Single-stub matching network design of Example 3.11.1: (a) electrical design; and (b) electrical design with a shunt stub.

the line is an arc on a circle centered at the origin of the Smith chart. The direction of the arc is clockwise as the electrical length of the line moves away from the load. So a hybrid matching network is possible that combines a length of transmission line with a lumped element (preferably a capacitor rather than an a inductor as the inductor would have a relatively lower Q).

This page titled 3.11: Distributed Matching is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.11: Distributed Matching** by Michael Steer is licensed CC BY-NC 4.0.

3.12: Matching Using the Smith Chart

The purpose of this section is to use the Smith chart to present several design options for matching a source to a load, see Figure 3.12.1. The designs here provide another view of design using the Smith chart.

3.12.1: Locating the Design Points

The first design choice to be made is the reference impedance to use. Here $Z_{\text{REF}} = 50\ \Omega$ will be chosen largely because this is in the center of the design space for microstrip lines. Generally the characteristic impedance, Z_0, of a microstrip line needs to be between $20\ \Omega$ and $100\ \Omega$. A microstrip line with $Z_0 < 20\ \Omega$ will be wide and there is a possibility of multimoding due to transverse resonance. Also a $20\ \Omega$ line is about six times wider than a $50\ \Omega$ line and so takes up a lot of room and there is a good chance that it could be close to other microstrip lines or perhaps the wall of an enclosure. This is based on the rule of thumb (developed in Example 3.4 of [2].) that $Z_0 \propto \sqrt{h/w}$ where h is the substrate thickness and w is the strip width. The thickness is usually fixed. If $Z_0 > 100\ \Omega$ the characteristic impedance is getting close to the wave impedance of free space or of the dielectric of the substrate. As such it is likely that field lines are not tightly constrained by the metal of the strip and the fields can more likely radiate. Then radiation loss can be high or coupling to a neighboring microstrip can be high.

The normalized source and load impedances are $z_S = Z_S/Z_{\text{REF}} = [(29.36 - j12.05)\ \Omega]/(50\ \Omega) = 0.587 - j0.241$ and $z_L = Z_L/Z_{\text{REF}} = [(132.7 - j148.8)\ \Omega/(50\ \Omega) = 2.655 - j2.976 = r_L + jx_L$, respectively. Maximum power transfer requires that the input impedance of the matching network terminated in Z_L be $Z_1 = Z_S^*$, i.e. $z_1 = z_S^* = 0.587 + j0.241 = r_1 + jx_1$. These impedances are plotted on the normalized Smith chart in Figure 3.12.2. The normalized load impedance is Point L. To locate this point the arcs corresponding to the real and imaginary parts of z_L are considered separately.

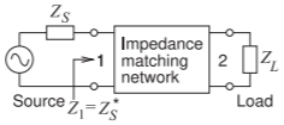


Figure 3.12.1: Matching problem with the matching network between the source and load designed for maxium power transfer. $Z_S = R_S + jX_S = 29.36 - j12.05$, $Z_1 = R_1 + jX_1 = Z_S^* = R_S - jX_S = 29.36 + j12.05$, and $Z_L = R_L + jX_L = 32.7 - j148.8$.

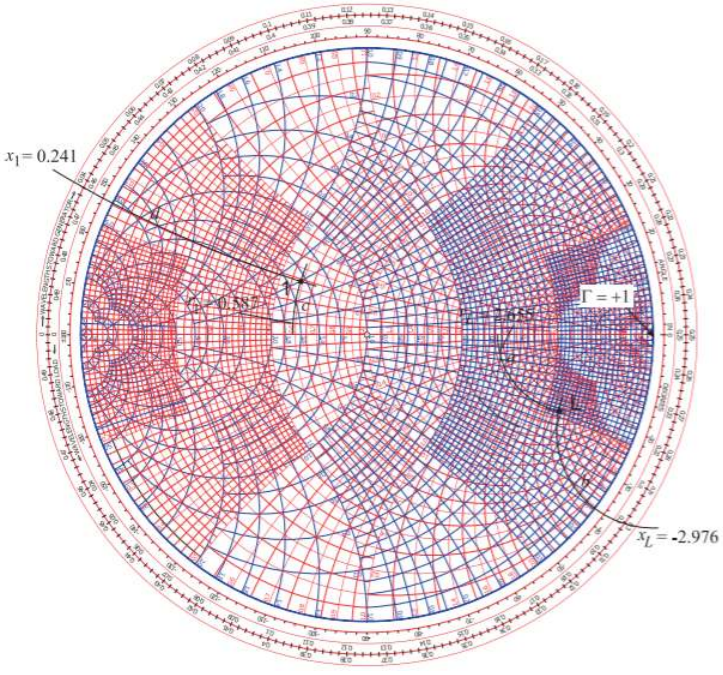


Figure 3.12.2: Locating Z_L at Point L and $Z_1 = Z_S^*$ at Point 1.

The resistive part of z_L is $r_L = 2.655$ and the resistance labels are located on the horizontal axis (or equator) of the Smith chart. There are two sets of labels, one for normalized resistance, r (which is above the horizontal axis), and one for normalized conductance, g (which is below the horizontal axis). The way to remember which is which is to realize that the infinite impedance point is at the $\Gamma = +1$ (open circuit) location on the right of the graph. At the origin (center) of the Smith chart $r = 1 = g$ and to the right of the center the values of r should be greater than one. The closest r labels to $r_L = 2.655$ are $r = 2.0$ and $r = 3.0$. There are five divisions so the unlabeled curves correspond to $2.2, 2.4, ...$ The arc corresponding to $r = 2.655$ must be interpolated and this interpolation is shown as the Path 'a'.

The imaginary part of z_L is $x_L = -2.976$. The labels for the arcs of constant reactance are given adjacent to the unit circle. There are two sets of labels, one for reactance and one for susceptance. To recall which is which, the point of infinite impedance can be used and the required reactance labels should increase towards the $\Gamma = +1$ point. Recall that the Smith chart does not include signs of reactances (there is not enough room) so note must be made that positive reactances are in the top half of the Smith chart and negative reactances are in the bottom half. Since x_L is negative it will be in the bottom half of the Smith chart. The closest labels are $x = 2.0$ (this is actually -2.0) and $x = 3.0$ (this is actually -3.0) so the arc for $x = -2.976$ is interpolated as the Path 'b'. Point L, i.e. z_L, is located at the intersection of Paths 'a' and'b'. The normalized impedance z_1 is located similarly at Point 1 by finding the point of intersection of the $r_1 = 0.587$ arc, Path 'c', and the $x_1 = +0.241$ arc, Path 'd'.

3.12.2: Design Options

By convention design follows a process of beginning with z_L and adding series and shunt elements in front of it evolving the impedance (or reflection coefficient) until the input impedance is $z_1 = z_s^*$. Two electrical designs are shown in Figure 3.12.3 and the corresponding lumped-element and microstrip topologies are shown in Figure 3.12.4. The subscript on the circuit elements correspond to the paths on the Smith chart in Figure 3.12.3. The designs will be elaborated in the following subsections.

3.12.3: Design 1, Hybrid Design

Design 1 on its own is shown in Figure 3.12.5. The concept here is to use a transmission line and a shunt to go from the load Point L to the Point 1. The reason why a shunt element is chosen and not a series element is that the shunt element can be implemented as a stub line and a series element, i.e. a series stub, cannot be implemented in microstrip. A lumped element limits a design to the low microwave range as losses become prohibitively large especially for inductors. Also if a microstrip line is going to be used anyway then a decision has already been made that there is enough room to implement a transmission-line based design and so the shunt lumped element can reasonably be replaced by a stub.

Design follows trial and error. The first attempt, and the one that works here, is to draw a circle through L centered on the origin at Point O. This circle describes a transmission line whose characteristic impedance is the same as the reference impedance of the Smith chart, here $50\ \Omega$. The next step is to draw a circle of constant conductance through Point 1. The combination path from L to 1 needs to lie on these circles and the intermediate point will be where these circles intersect. One other constraint is that with a transmission line the locus (as the line length increases) of the input reflection coefficient of the line must rotate in the clockwise direction. It is seen that there are two points of intersection and the first of these, at A, is chosen in design. So the electrical design is defined by the directed Paths 'g' and 'h'. Path 'g' defines the properties of the transmission line and Path 'h' defines the properties of the shunt element. As 'h' is directed towards the infinite inductive susceptance point, Path 'h' defines an inductor. The topology of this design is shown in Figure 3.12.4(a).

The characteristic impedance of the transmission line (defined by Path 'g') is $Z_{0g} = 50\ \Omega$ and the electrical length of the line is defined by the angle subtended by the arc 'g'. The electrical length of the line is determined from the outermost circular scale which is labeled 'WAVELENGTHS TOWARDS GENERATOR.'

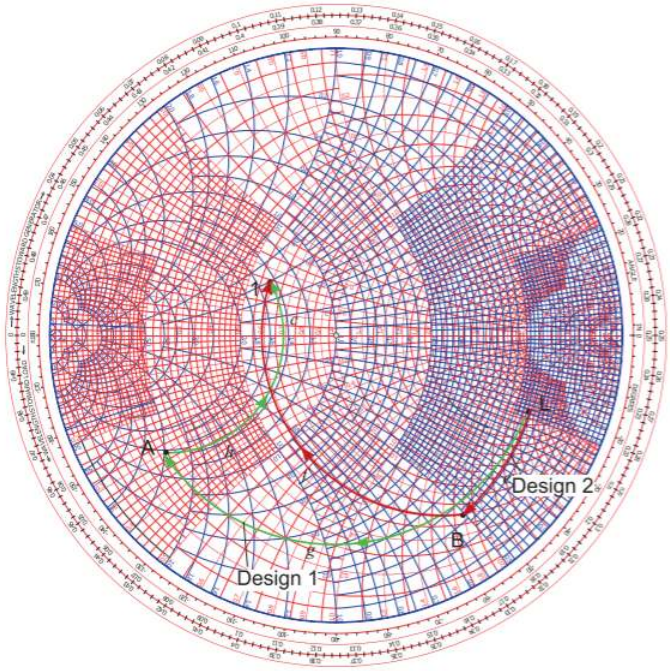


Figure 3.12.3: Two matching network electrical designs matching a load impedance Z_L at Point L to a source Z_S showing $Z_1 = Z_S^*$ at Point 1.

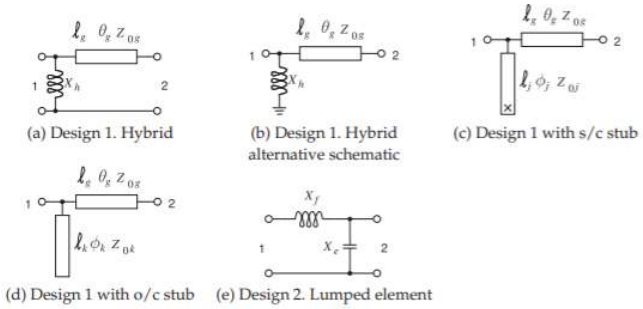


Figure 3.12.4: Matching network topologies using lumped elements and microstrip lines. In the stub layouts × is a via to the ground plane implementing a short circuit (s/c) and an open circuit o/c simply does not show a connection to the microstrip ground plane.

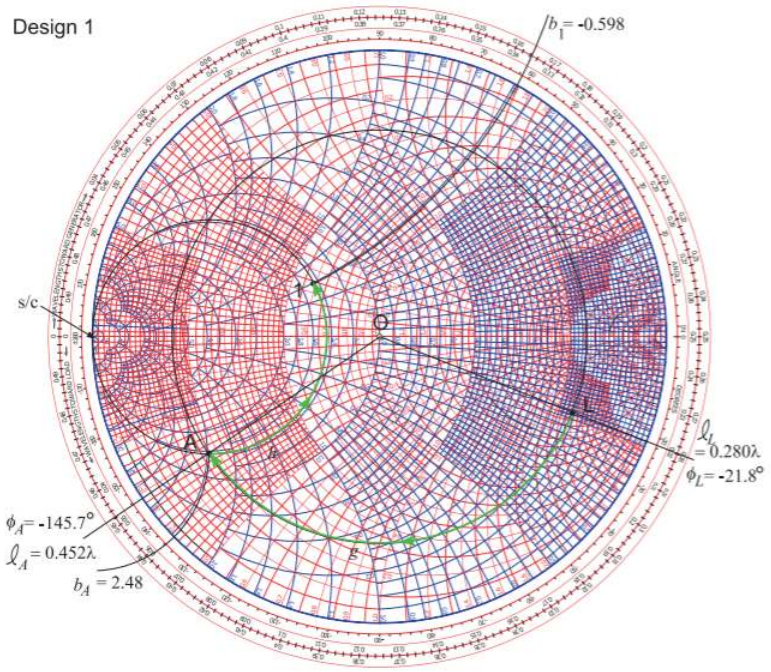


Figure 3.12.5: Design 1. Hybrid design combining a transmission line with a lumped element in shunt. The design is identified by paths ‘g’ and ‘h’.

A line drawn from O through L intersecting the scale has a scale reading of $\ell_L = 0.280\lambda$. Then the scale reading at A is similarly found as $\ell_A = 0.452\lambda$ and so the line length is $\ell_g = \ell_A - \ell_L = 0.452\lambda - 0.280\lambda = 0.172\lambda$. Another way of determining the electrical length of the line is from the change in reflection coefficient angle. For L the reflection coefficient angle is $\phi_L = -21.8^\circ$ read from the innermost circular scale. This angle is just the angle from the polar plot. Then the angle at A is read as $\phi_A = -145.7^\circ$. The difference is $|\phi_A - \phi_L| = |-145.7 - (-21.8^\circ)| = 124.9^\circ$. The electrical length of the line is half the change in reflection coefficient angle and so the electrical length of the line is $\theta_g = \frac{1}{2}124.9^\circ = 62.5^\circ$. Now λ corresponds to an electrical length of 360° so θ_g corresponds to $62.5/360\lambda = 0.174\lambda$ corresponding to the previously determined length of 0.172λ which is very good agreement given that these were derived from graphical readings.

Path ‘h’ defines a shunt inductor and a circle of constant conductance is followed with only the susceptance changing. The susceptance indicated by Path ‘h’ is $b_h = b_1 - b_A$. To obtain b_A extend the circle of constant susceptance through A out to the unit circle. The extended circle crosses the unit circle between the susceptance labels 2.0 and 3.0. A check is that susceptance is positive in the bottom half of the Smith chart so the signs of the labels do not need to be adjusted. There are two scales adjacent to the unit circle, one for normalized susceptance and one for normalized reactance. The intersection is close to the infinite susceptance point at the s/c (short-circuit) so the values that are becoming very large towards s/c are used. Interpolation results in the reading $b_A = 2.48$. A similar process applied to Point 1 results in $b_1 = -0.598$ where the negative sign has been applied to the scale reading since Point 1 is in the top half of the Smith chart. Thus $b_h = b_1 - b_A = -0.598 - 2.48 = -3.08$ and so the normalized reactance of the shunt element is $x_h = -1/b_h = 0.325$. The un-normalized reactance of the shunt element is $X_h = x_h Z_{\text{REF}} = 16.2\ \Omega$.

The final Design 1 hybrid layout is shown in Figure 3.12.4(a) with $X_h = 16.2\ \Omega,\ Z_{0g} = 50\ \Omega$, and $\ell_g = 0.172\lambda$. That is all that is needed to define the electrical design, providing the electrical length in degrees, $\theta_g = 62.5^\circ$ is redundant but provided anyway. The transmission line in Figure 3.12.4(a) is shown as as the top view of the strip of a microstrip line as is commonly done. A more common way of representing this schematic is shown in Figure 3.12.4(b) where the ground connections at Ports 1 and 2 have been removed and the ground connection of the inductor shown separately.

3.12.4: Design 1 with an Open-Circuited Stub

In the previous section Design 1 was left as a hybrid design with a transmission line and a lumped-element inductor. In this section the lumped-element inductor is implemented as an open-circuited stub, see Figure 3.12.4(d). Recall that the $50\ \Omega$-normalized susceptance of the inductor is $b_h = -3.08$. If the stub is also implemented as a $50\ \Omega$ line then b_h can be used unchanged. Point C in

Figure 3.12.6 corresponds to the normalized admittance $0 - j3.08$. The unit circle is the zero conductance circle (and is also the zero resistance circle) and the susceptance is read from the scale adjacent to the unit circle again noting that susceptances in the top half of the Smith chart need to incorporate a negative sign, and the susceptance scale is identified by the susceptance values becoming larger approaching the s/c point. Point C also corresponds to $x_h = -1/b_h = 3.25$ and indeed this is the value read from the normalized reactance scale.

A transmission line needs to be designed to have a normalized input susceptance of $b_h = -3.08$. Choosing an open circuit, o/c, termination the point corresponding to o/c is as identified in the figure. At the o/c point the length scale reads $\ell_{o/c} = 0.250\lambda$. The locus rotates in the clockwise direction up to Point C where the direct electrical reading reading is $\ell_C = 0.050\lambda$. Using this directly to determine the line length $\ell_k = \ell_C - \ell_{o/c} = 0.050\lambda - 0.250\lambda = -0.20\lambda$ which indicates that the stub has a negative length. Clearly an erroneous result. This apparent discrepancy comes about because the length scale resets at the short circuit point where the length scale abruptly goes from 0.5λ to 0λ. Thus the corrected ℓ_C reading needs to have an additional 0.5λ. Thus the corrected value of $\ell_C = (0.5 + 0.050)\lambda = 0.550\lambda$ and $\ell_k = \ell_C - \ell_{o/c} = 0.550\lambda - 0.250\lambda = 0.300\lambda$.

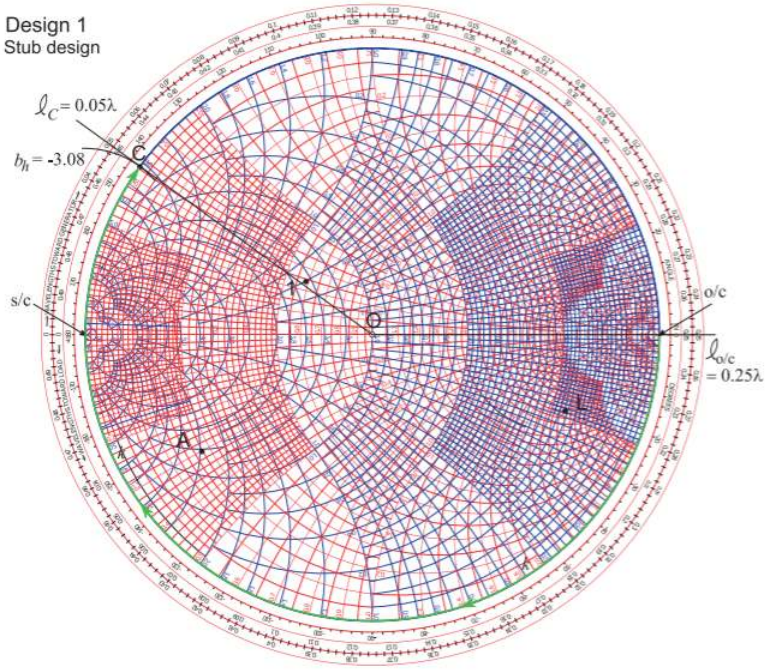


Figure 3.12.6: Design 1. Design of an open-circuit stub having normalized input susceptance b_h.

Thus the final design is as shown in Figure 3.12.4(d) with $Z_{0k} = 50\ \Omega$, and $\ell_g = 0.300\lambda,\ Z_{0g} = 50\ \Omega$, and $\ell_g = 0.172\lambda$.

The stub could also have been implemented as a short-circuit stub as shown in Figure 3.12.4(c). Now the beginning of the line would be at the s/c point and the line length would be 0.050λ.

3.12.5: Design 2, Lumped-Element Design

Design 2 is a lumped-element design and the Smith-chart-based electrical design is shown in Figure 3.12.7 resulting in the schematic shown in Figure 3.12.4(e). Design proceeds by identifying where circles of constant conductance and constant resistance passing through the Points L and 1 intersect. One solution is shown in Figure 3.12.7. A circle of constant conductance passes through L and part of a circle of constant resistance passes through 1. If the circle had continued there would have been a second intersection with the circle through L.

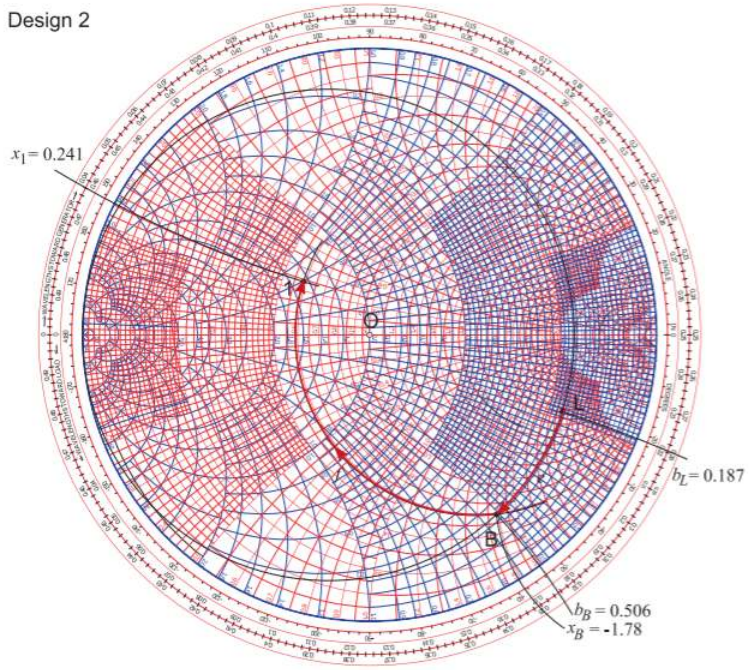


Figure 3.12.7: Design 2.

Both of these intersections mean that there is a shunt element adjacent to the load and series element adjacent to the source. Recall that in a lumped-element design that under no circumstances can the locus a lumped element pass through the short circuit or open-circuit points (the susceptance and reactance infinity points respectively).

Returning to the actual design shown in Figure 3.12.7. The first intersection of the two circles is Point B so that the design is specified by the Paths 'e' and 'f'. Design has largely been completed by identifying these paths and the next stage is determining the circuit elements that correspond to these paths. Path 'e' follows a circle of constant conductance and so indicates a shunt susceptance and the direction of the locus indicates a capacitance. The value of this normalized susceptance is $b_e = b_B - b_L = 0.506 - 0.187 = 0.319$. (Remember to check the signs of the readings.) Path 'f' identifies a series inductor with a reactance $x_f = x_1 - x_B = 0.241 - (-1.78) = 2.02$. The final design is shown in Figure 3.12.4(e) with $X_e = Z_0/b_e = 50/0.319\ \Omega = 158\ \Omega$ and $X_f = Z_0 x_f = 50 \cdot 2.02\ \Omega = 101\ \Omega$.

3.12.6: Summary

This section presented two designs for a matching network. One of the particular benefits of using the Smith chart is identifying topologies and initial design values. Design can then transfer to a microwave circuit simulator. The Smith chart enables back-of-the-envelope design studies. While with experience it is possible to complete many of these steps with a computer-based Smith chart tool, even experienced designers doodle with a printed Smith chart when exploring design options.

This page titled 3.12: Matching Using the Smith Chart is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **10.12: Matching Using the Smith Chart** by Michael Steer is licensed CC BY-NC 4.0.

CHAPTER OVERVIEW

4: Filters

4.1: Introduction
4.2: Singly and Doubly Terminated Networks
4.3: The Lowpass Filter Prototype
4.4: The Maximally Flat (Butterworth) Lowpass Approximation
4.5: The Chebyshev Lowpass Approximation
4.6: Element Extraction
4.7: Butterworth and Chebyshev Filters
4.8: Impedance and Admittance Inverters
4.9: Filter Transformations
4.10: Cascaded Line Realization of Filters
4.11: Butterworth and Chebyshev Bandpass Filters
4.12: Richard's Transformation
4.13: Kuroda's and Norton's Network Identities
4.14: Inter-resonator Coupled Bandpass Filters
4.15: Bandpass Filter Topologies
4.16: Case Study- Design of a Bandstop Filter
4.17: Active Filters
4.18: Transient Response of a Bandpass Filter
4.19: Summary
4.20: References
4.21: Exercises

This page titled 4: Filters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

4.1: Introduction

Filters are the most fundamental of signal processing circuits using energy storage elements to obtain frequency-dependent characteristics. Some of the important filter attributes are (1) controlling noise by not allowing out-of-band noise to propagate in a circuit; (2) keeping signals outside the transmit band, especially harmonics, from being transmitted; and (3) presenting only signals in a specified band to active receive circuitry. At microwave frequencies a filter can consist solely of lumped elements, solely of distributed elements, or a mix of lumped and distributed elements. The distributed realizations can be transmission line-based implementations of the components of lumped-element filter prototypes or, preferably, make use of the particular frequency characteristics found with certain distributed structures. Loss in lumped elements, particularly above a few gigahertz, means that the performance of distributed filters nearly always exceeds that of lumped-element filters. However, since the basic component of a distributed filter is a one-quarter wavelength long transmission line, distributed filters can be prohibitively large below a few gigahertz.

Only a few basic types of responses are required of most RF filters as follows:

a. Lowpass—providing maximum power transfer at frequencies below the corner frequency, f_0. Above f_0, transmission is blocked. See Figure 4.1.1(a).
b. Highpass—passing signals at frequencies above f_0. Below f_0, transmission is blocked. See Figure 4.1.1(b).
c. Bandpass—passing signals at frequencies between lower and upper corner frequencies (defining the passband) and blocking transmission outside the band. This is the most common type of RF filter. See Figure 4.1.1(c).
d. Bandstop (or notch)—which blocks signals between lower and upper corner frequencies (defining the stopband). See Figure 4.1.1(d).
e. Allpass—which equalizes a signal by adjusting the phase generally to correct for phase distortion elsewhere. See Figure 4.1.1 (e).

The above list is not comprehensive, as actual operating conditions may mandate specific frequency profiles. For example, WiFi systems operating at 2.45 GHz are susceptible to potentially large signals being transmitted from nearby cellular phones operating in the 1700–2300 MHz range. Thus a frontend filter for a 2.45 GHz system must ensure very high levels of attenuation over the 1700–2300 MHz range. Thus the optimum solution here, probably, will have an asymmetrical frequency response with high rejection on one side of the passband obtained by accepting lower rejection on the other side.

4.1.1: Filter Prototypes

In the 1960s an approach to RF filter design and synthesis was developed and this is still followed. The approach is to translate the mathematical response of a lowpass filter. A filter with the desired lowpass response is then synthesized using lumped elements and the resulting filter is called a lowpass prototype. The lowpass filter is then transformed so that the new lumped-element filter has the desired RF response, such as highpass or bandpass. In the case of a bandpass filter, each inductor and capacitor of the lowpass prototype becomes a resonator that is coupled to another resonator. In distributed form, this basic resonator is a one-quarter wavelength long transmission line.

Filter synthesis is a systematic approach to realizing circuits with desired frequency characteristics. A filter can also be thought of as a hardware implementation of a differential equation producing a specific impulse,

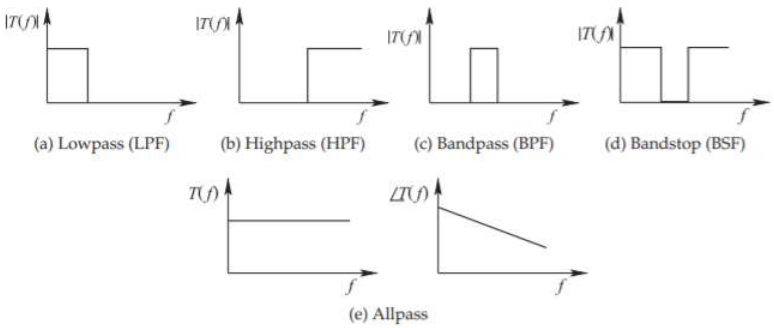


Figure 4.1.1: Ideal filter transfer function, $T(f)$, responses.

step, or frequency response. It is not surprising that a filter can be described this way as the lumped reactive elements, L and C, describe first-order differential equations and their interconnection describes higher-order differential equations.

The synthesis of a filter response begins by identifying the desired differential equation response. The differential equations are specified using the Laplace variable, s, so that in the frequency domain (with $s = j\omega$) an nth-order differential equation becomes an nth-order polynomial expression in s. If possible, and this is usually the case, a lowpass version of a filter can be developed. The conversion to the final filter shape, say a bandpass response, proceeds mathematically through a number of stages. Each stage has a circuit form and each stage is called a filter prototype.

4.1.2: Image Parameter Versus Insertion loss Methods

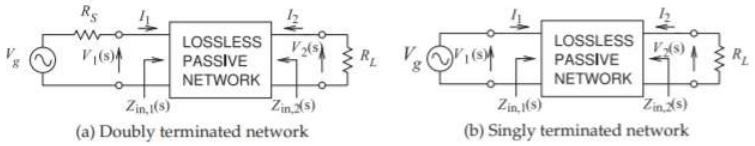
Filter networks can be synthesized using the image parameter method or the reflection coefficient (or insertion loss) method [1]. The image parameter measured is based on cascading two-port networks with each two-port having the essential desired filter response [2]. Several of these two-ports are cascaded to achieve a periodic-like structure that has the final desired filter response. This method is restrictive and arbitrary filter responses can not be obtained. It is rarely used in RF and microwave filter design but is useful in analyzing simple structures with cascades of identical elements [3]. A better approach is to use the insertion loss method as used in this chapter.

This page titled 4.1: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.1: Introduction** has no license indicated.

4.2: Singly and Doubly Terminated Networks

Filters are generally two-port networks and they provide maximum power transfer from a source to a load over a specified frequency range while rejecting the transmission of signals at other frequencies. Two possible filter networks are shown in Figure 4.2.1. The network in Figure 4.2.1(a) is known as a doubly terminated network, as both ports are resistively terminated.



(a) Doubly terminated network (b) Singly terminated network

Figure 4.2.1: Terminated networks.

The network in Figure 4.2.1(b) is called a singly terminated network, as only one port is terminated in a resistor. The doubly terminated network is much closer to the type of network required at RF where loads and source impedances are finite. A singly terminated network is applicable in some RF integrated circuit applications where very little RF power is involved and often when feedback is used. In such cases the output of an RFIC amplifier can approximate an ideal voltage source as the Thevenin equivalent source impedance can be negligible. Most synthesis of singly terminated filter networks uses an analogous procedure to that presented here for doubly terminated filters.

4.2.1: Double Terminated Networks

The established filter synthesis procedure for doubly terminated filter networks focuses on realizing the input reflection coefficient. For a lossless filter the square of the reflection coefficient magnitude is one minus the square of the insertion loss and this is the origin of this method being called the insertion loss method.

The input reflection coefficient of the doubly terminated network in Figure 4.2.1(a) is

$$\Gamma_1(s) = \frac{Z_{\text{in, 1}}(s) - R_S}{Z_{\text{in, 1}}(s) + R_S} \tag{4.2.1}$$

where the reference impedance is the source resistance, RS, and s is the Laplace variable. In the passband of a filter the reflection coefficient is approximately zero.

There are several other parameters used in filter design and these will be introduced now. The **transducer power ratio** (**TPR**) is defined as

$$\begin{aligned} \text{TPR} &= \frac{\text{Maximum power available from the source}}{\text{Power absorbed by the load}} \\ &= \frac{\frac{1}{2}(V_g(s)/2)^2/R_S}{\frac{1}{2}V_2^2(s)/R_L} = \left|\frac{1}{2}\sqrt{\frac{R_L}{R_S}}\frac{V_g(s)}{V_2(s)}\right|^2 \end{aligned} \tag{4.2.2}$$

The transmission coefficient, $T(s)$, of the network is

$$T(s) = \frac{1}{\sqrt{\text{TPR}(s)}} = 2\sqrt{\frac{R_S}{R_L}}\frac{V_2(s)}{V_g(s)} \tag{4.2.3}$$

IF the filter is lossless, the insertion loss (IL) (or **transducer** function) is

$$\text{IL}(s) = \text{TPR}(s) = [1/T(s)]^2 \tag{4.2.4}$$

The next step in development of the filter synthesis procedure is introduction of the **characteristic function**, defined as the ratio of the reflection and transmission coefficients:

$$K(s) = \frac{\Gamma_1(s)}{T(s)} = \frac{N(s)}{D(s)} \tag{4.2.5}$$

where N is the numerator function and D is the denominator function. Ideally the filter network is lossless and so, from the principle of energy conservation, the sum of the magnitudes squared of the transmission and reflection coefficients must be unity:

$$|T(s)|^2 + |\Gamma_1(s)|^2 = 1 \tag{4.2.6}$$

That is,

$$|T(s)|^2 = 1 - |\Gamma_1(s)|^2 \tag{4.2.7}$$

Dividing both sides of Equation (4.2.7) by $|T(s)|^2$ results in

$$1 = \frac{1}{|T(s)|^2} - \frac{|\Gamma_1(s)|^2}{|T(s)|^2} \tag{4.2.8}$$

or

$$1 = |\mathrm{IL}(s)| - |K(s)|^2 \tag{4.2.9}$$

Rearranging Equation (4.2.9) leads to

$$|\mathrm{IL}(s)| = 1 + |K(s)|^2 \tag{4.2.10}$$

and so

$$|T(s)|^2 = \frac{1}{1 + |K(s)|^2} \tag{4.2.11}$$

and

$$|\Gamma_1(s)|^2 = \frac{|K(s)|^2}{1 + |K(s)|^2} \tag{4.2.12}$$

In the above it is seen that both the reflection and transmission coefficients are functions of the characteristic function of the two-port network. A milestone has been reached. In an RF filter, the frequency-dependent insertion loss or transmission coefficient is of most importance, as these are directly related to power flow. Equation (4.2.11) shows that this can be expressed in terms of another function, $K(s)$, which, from Equation (4.2.5), can be expressed as the ratio of $N(s)$ and $D(s)$. For lumped-element circuits, $N(s)$ and $D(s)$ are polynomials.

4.2.2: Lowpass Filter Response

As an example of the relationship of the various filter responses, consider the lossless lowpass filter responses shown in Figure 4.2.1. This filter has a lowpass response with a corner frequency expressed in radians as $\omega_c = 1$ rad/s. Ideally $T(s)$ for $s \leq \jmath$ (note $s = \jmath\omega$) would be one, and for $s > \jmath$, $T(s)$ would be zero. It is not possible to realize such an ideal response, and the response shown in Figure 4.2.2(a) is typical of what can be achieved. The reflection coefficient is shown in Figure 4.2.2(b), with the characteristic function $K(s)$ shown in Figure 4.2.2(c). If $|K(s)|^2$ is expressed as the ratio of two polynomials (i.e, as $N(s)/D(s)$), then it can be seen from Figure 4.2.2(e and f) that the zeros of the reflection coefficient, $|\Gamma_1(s)|^2$, are also the zeros of $N(s)$. Also, it is observed that the zeros of the transmission coefficient are also the zeros of $D(s)$, as shown in Figure 4.2.2(a and c).

The first objective in lumped-element filter design is development of the transfer function of the network in the frequency domain or, equivalently, the s domain.[1] The input-output transfer function of the generic filter in Figure 4.2.3 is

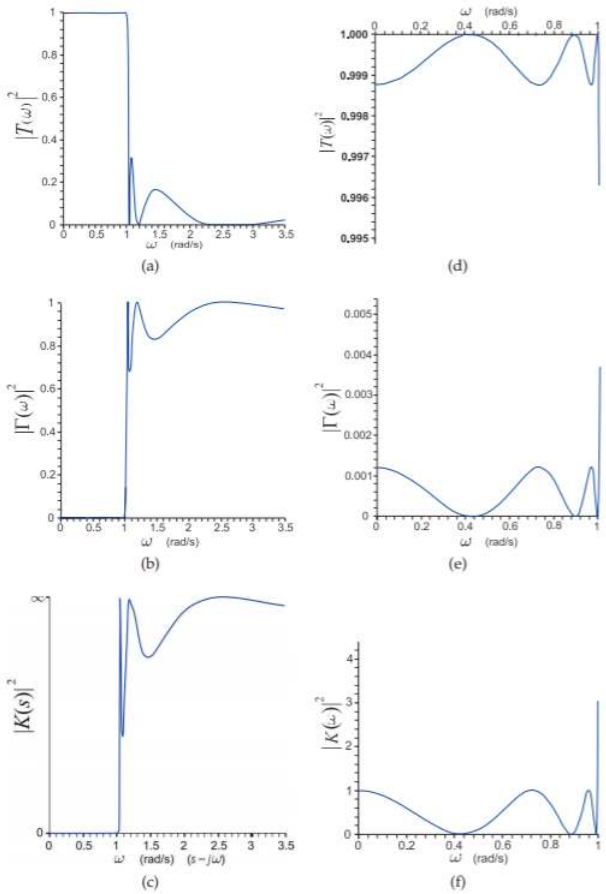


Figure 4.2.2: An example of a lowpass filter in terms of various responses: (a) transmission coefficient; (b) reflection coefficient response; and (c) characteristic function response. Detailed responses are shown in (d), (e), and (f), respectively.

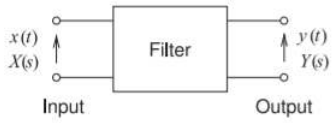


Figure 4.2.3: Generic filter.

$$T(s) = Y(s)/X(s) \tag{4.2.13}$$

and the design procedure is to match this response to a response defined by the ratio of two polynomials:

$$T(s) = \frac{N(s)}{D(s)} = \frac{a_m s^m + a_{m-1}s^{m-1} + \cdots + a_1 s + a_0}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0} \tag{4.2.14}$$

Here N stands for numerator and D for denominator, and these are not the same as those in Equation (4.2.5) (where they were just labels for numerator and denominator). The filter response using a **pole-zero description** can be synthesized so the design process begins by rewriting Equation (4.2.14) explicitly in terms of zeros, z_m, and poles, p_n:

$$T(s) = \frac{N(s)}{D(s)} = \frac{a_m(s+z_1)(s+z_2)\cdots(s+z_{m-1})(s+z_m)}{(s+p_1)(s+p_2)\cdots(s+p_{n-1})(s+p_n)} \tag{4.2.15}$$

Since only the frequency response is of interest, $s = \jmath\omega$, thus simplifying the analysis for sinusoidal signals.

The poles and zeros can be complex numbers and can be plotted on the complex s plane. Conditions imposed by realizable circuits require that $D(s)$ be a **Hurwitz** polynomial,[2] which ensures that its poles are located in the left-half plane. $N(s)$ determines the

location of the transmission zeros of the filter, and the order of $N(s)$ cannot be more than the order of $D(s)$. That is, $n \geq m$ so that the filter has finite or zero response at infinite frequency

Two strategies can be employed in deriving the filter response. The first is to derive the polynomials $N(s)$ and $D(s)$ in Equation (4.2.14). This seems like an open-ended problem, but it was discovered in the 1950s and 1960s that in normal situations there are only a few types of useful responses that are described by a few polynomials, including Butterworth, Bessel, Chebyshev, and Cauer polynomials.

4.2.3: Footnotes

[1] Care is being taken in the use of terminology as s can be complex with some filter types, but these filters are realized using digital signal processing.

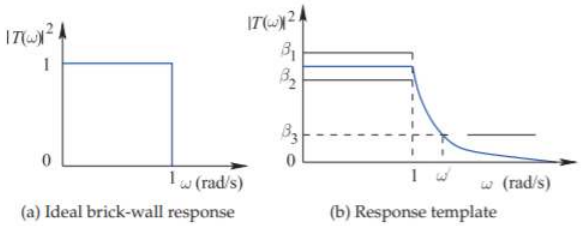
[2] A Hurwitz polynomial is a polynomial whose coefficients are positive real numbers and whose zeros are located in the left-half of the complex s plane.

This page titled 4.2: Singly and Doubly Terminated Networks is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.2: Singly and Doubly Terminated Networks** has no license indicated.

4.3: The Lowpass Filter Prototype

Most filter design is based on the synthesis of a lowpass filter equivalent called a lowpass prototype. Transformations are then used to correct for the actual source and load impedances, frequency range, and the desired filter type, such as highpass or bandpass. An ideal lowpass response is shown in Figure 4.3.1(a), which shows a transmission coefficient of 1 up to a normalized frequency of 1 rad/sec. This type of response defines what is called a **brick-wall filter**, which unfortunately cannot be physically realized. Instead, the response is approximated and specified in terms of a response template (see Figure 4.3.1(b)).



(a) Ideal brick-wall response (b) Response template

Figure 4.3.1: Lowpass filter response.

The template specifies a passband response that is between β_1 and β_2 at frequencies below the corner radian frequency (at $\omega = 1$), and below β_3 above a radian frequency $\omega' > 1$. The lowpass filter is harder to realize the closer the specified response is to the ideal response shown in Figure 4.3.1(a), that is, when

$$(\beta_2 - \beta_1) \to 0, \quad \beta_3 \to 0, \quad \text{and} \quad (\omega' - 1) \to 0 \tag{4.3.1}$$

Several polynomials have particularly interesting characteristics that match the requirements of the response template. At first it may seem surprising that polynomial functions could yield close to ideal filter responses. However, as will be seen, there is something special about these polynomials: they are natural solutions to extreme conditions. For example, the nth-order Butterworth polynomial has the special property that the first n derivatives at $s = 0$ are zero. The other polynomials also have extreme properties, and filters that are synthesized using them also have extreme properties. As well as the maximally flat property resulting from Butterworth polynomials, filter responses obtained using Chebyshev polynomials have the steepest skirts (i.e., fastest transitions from the frequency region where signals are passed to the region where they are blocked). Usually it is one of these extreme cases that is most desirable.

A streamlining of the filter synthesis procedure is obtained by first focusing on the development of normalized lowpass filters having a 1 rad/s corner frequency and 1 Ω reference impedance. Transformations transform a lowpass filter prototype into another filter having the response desired. These transformations give rise to symmetrical responses.

This page titled 4.3: The Lowpass Filter Prototype is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.3: The Lowpass Filter Prototype** has no license indicated.

4.4: The Maximally Flat (Butterworth) Lowpass Approximation

Butterworth filters have a maximally flat response (Figure 4.4.1) that in the time domain corresponds to a critically damped system.

4.4.1: Butterworth Filter Design

Butterworth filters have the transfer function

$$T(s) = \frac{N(s)}{D(s)} = \frac{k}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0} \tag{4.4.1}$$

This is an all-pole response. The characteristic polynomial of the Butterworth filter is

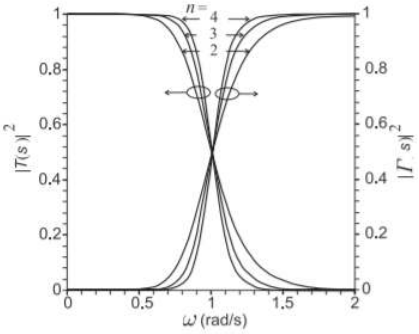


Figure 4.4.1: Maximally flat, or Butterworth, lowpass filter approximation for various orders, n, of the filter.

n	**Factors of** $B_n(s)$
1	$(s+1)$
2	$(s^2+1.4142s+1)$
3	$(s+1)(s^2+s+1)$
4	$(s^2+0.7654s+1)(s^2+1.8478s+1)$
5	$(s+1)(s^2+0.6180s+1)(s^2+1.6180s+1)$
6	$(s^2+0.5176s+1)(s^2+1.4142s+1)(s^2+1.9319s+1)$
7	$(s+1)(s^2+0.4450s+1)(s^2+1.2470s+1)(s^2+1.8019s+1)$
8	$(s^2+0.3902+1)(s^2+1.1111s+1)(s^2+1.6629s+1)(s^2+1.9616s+1)$

Table 4.4.1: Factors of the $B_n(s)$ polynomial.

$$|K(s)|^2 = |s^{2n}| = \omega^{2n} \tag{4.4.2}$$

since $s = \jmath\omega$ and where n is the order of the function. Thus the transmission coefficient is

$$|T(s)|^2 = \frac{1}{1+|K(s)|^2} = \frac{1}{1+|s^{2n}|} = \frac{1}{1+\omega^{2n}} \tag{4.4.3}$$

which is often written as

$$|T(s)|^2 = \frac{1}{1+|K(s)|^2} = \frac{1}{B_n(s)B_n(-s)} \tag{4.4.4}$$

In the filter community, $B_n(s)$ is called the Butterworth polynomial and has the general form

$$B_n(s) = \begin{cases} \prod_{k=1}^{n/2} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\pi\right) + 1\right] & \text{for } n \text{ even} \\ (s+1)\prod_{k=1}^{(n-1)/2} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\pi\right) + 1\right] & \text{for } n \text{ odd} \end{cases} \tag{4.4.5}$$

$B_n(s)$ is given in factorized form in Table 4.4.1. Further factorization of the second-order factors results in complex conjugate roots, so they are generally left in the form shown.

Insight is gained by examining the characteristic polynomial for a real radian frequency, ω (where $s = \jmath\omega$), at the following frequency points:

$$\text{at } \omega = 0 \quad \rightarrow \quad |T(0)|^2 = \frac{1}{1+0} = 1 \tag{4.4.6}$$

$$\text{at } \omega = 1 \quad \rightarrow \quad |T(1)|^2 = \frac{1}{1+1} = \frac{1}{2} \tag{4.4.7}$$

Thus the transmission response of the filter is at half power at the corner frequency $\omega = 1$. Another observation is that the transmission coefficient at $\omega = 0$ and $\omega = 1$ are independent of the degree of the characteristic polynomial. This is shown in Figure 4.4.1, where the responses of Butterworth filters are plotted for three different orders.

4.4.2: Construction of the Transfer Function

Often the characteristic function, K, is specified in terms of the variable ω, and for synthesis purposes the response must be transformed back to the s domain. As an example, consider the maximally flat function of third order with a characteristic polynomial equal to ω^3 (from Equation (4.4.3)):

$$K(\omega) = \omega^3 \quad \text{and} \quad |T(\omega)|^2 = \frac{1}{1+\omega^6} \tag{4.4.8}$$

(Note that here the first six derivatives of $|K(s)|^2$ with respect to s are zero at $s = 0$.) So, from Equation (2.2.7),

$$|\Gamma_1(\omega)|^2 = 1 - \frac{1}{1+\omega^6} = \frac{\omega^6}{1+\omega^6} \tag{4.4.9}$$

At real frequencies (i.e., frequencies lying on the imaginary axis in the s plane)

$$s = \jmath\omega, \quad \text{so} \quad \omega = s/\jmath = -\jmath s \tag{4.4.10}$$

Thus, in the s domain, the reflection coefficient becomes

$$|\Gamma_1(\jmath\omega)|^2 = |\Gamma_1(\jmath(-\jmath s))|^2 = |\Gamma_1(s)|^2 = \frac{(-\jmath s)^6}{1+(-\jmath s)^6} \tag{4.4.11}$$

That is,

$$|\Gamma_1(s)|^2 = \frac{-s^6}{1-s^6} = \Gamma_1(s)\Gamma_1(-s) \tag{4.4.12}$$

Factoring the denominator polynomial yields

$$1 - s^6 = (1-s)(1+s)(s^2+s+1)(s^2-s+1) \tag{4.4.13}$$

By choosing those factors with roots only in the left-half plane (required for a realizable network), the zeros of the denominator are obtained. Now the numerator of $|\Gamma_1(s)|^2$ is easily factored, and since this is third order, it only has three reflection zeros at DC. The resulting function is

$$\Gamma_1(s) = \frac{s^3}{(s+1)(s^2+s+1)} \tag{4.4.14}$$

and so

$$T(s) = \frac{1}{(s+1)(s^2+s+1)} \tag{4.4.15}$$

This example illustrates how $\Gamma_1(s)$ and $T(s)$ can be obtained from $K(s)$. This procedure is generalized in the next section.

4.4.3: nth-Order Reflection Approximation

Following the procedure outlined in the previous section, generalization leads to the nth-order Butterworth response:

$$\begin{aligned} |\Gamma_1(s)|^2 &= \Gamma_1(s)\cdot\Gamma_1(-s) \\ &= \frac{(-s^2)^n}{1+(-s^2)^n} = \frac{(-s^2)^n}{\prod_{i=1}^{n}(s-s_i)\cdot\prod_{j=n+1}^{2n}(s-s_j)} \end{aligned} \tag{4.4.16}$$

So $|\Gamma_1(s)|^2$ has n roots (these are the s_is) lying in the left-half s plane, and n roots (the s_js) lying in the right-half s plane. Note that j is used as an index and $\jmath$ ($\jmath$, without the dot) represents $\sqrt{-1}$. It is reasonable to group all of the left-half plane roots together (a circuit could not be synthesized if there were a right-half plane root) so that

$$\Gamma_1(s) = \frac{(-s)^n}{\prod_{i=1}^{n}(s-s_i)} \tag{4.4.17}$$

Solving for the roots of the denominator of Equation (4.4.16) (i.e., finding the roots of $1+(-s^2)^n=0$) yields the following roots in the left-half s plane:

$$s_i = \exp\left\{\jmath(2i-1+n)\frac{\pi}{2n}\right\} \quad i=1,2,\ldots,n \tag{4.4.18}$$

where $\exp(\jmath\theta)=\cos(\theta)+\jmath\sin(\theta)$.

Example 4.4.1: Reflection Coefficient Derivation for Butterworth Filter

Develop the reflection coefficient from the roots of the third-order $(n=3)$ Butterworth lowpass filter prototype.

Solution

From Equation (4.4.18) the three roots are

$$s_1 = \exp\left\{\jmath(2\times1-1+3)\frac{\pi}{2\times3}\right\} = \exp\left\{\jmath\frac{2}{3}\pi\right\} \tag{4.4.19}$$

$$s_2 = \exp\left\{\jmath(2\times2-1+3)\frac{\pi}{2\times3}\right\} = \exp\{\jmath\,\pi\} \tag{4.4.20}$$

$$s_3 = \exp\left\{\jmath(2\times3-1+3)\frac{\pi}{2\times3}\right\} = \exp\left\{\jmath\frac{4}{3}\pi\right\} \tag{4.4.21}$$

and the input reflection coefficient of the filter normalized to $1\ \Omega$ is

$$\Gamma_1(s) = \frac{s^3}{\left(s+\frac{1}{2}-\jmath\frac{\sqrt{3}}{2}\right)(s+1)\left(s+\frac{1}{2}+\jmath\frac{\sqrt{3}}{2}\right)} = \frac{s^3}{(s+1)(s^2+s+1)} \tag{4.4.22}$$

This is the same as Equation (4.4.14).

4.4.4: Bandwidth Consideration

At the corner frequency of the Butterworth response, that is, at 1 rad/s, the transmission response is 3 dB down from its maximum transfer response.

Order, n							
$n=3$	4	5	6	7	8	9	10
0.798 rad/s	0.844	0.873	0.901	0.908	0.919	0.928	0.935

Table 4.4.2: Radian frequencies at which the response of a Butterworth filter is 1 dB down for a corner frequency $f_0 = 0.1592$ Hz ($\omega_0 = 1$ rad/s). At f_0 the transmission response is down 3 dB.

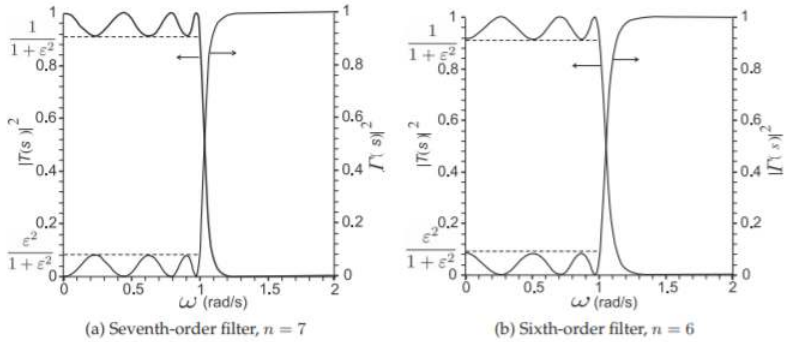


Figure 4.4.2: Chebyshev lowpass filter responses. In decibels, the ripple is $R_{\text{dB}} = -10\log[1/(1+\varepsilon^2)] = 10\log(1+\varepsilon^2)$, ε is called the ripple factor.

Sometimes microwave filters are designed to have 1 dB bandwidths. The radian frequencies at which the responses of various orders of Butterworth filters are 1 dB down are given in Table 4.4.2. By frequency scaling the Butterworth response the filter can be design for a specified 1 dB bandwidth.

This page titled 4.4: The Maximally Flat (Butterworth) Lowpass Approximation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.4: The Maximally Flat (Butterworth) Lowpass Approximation** has no license indicated.

4.5: The Chebyshev Lowpass Approximation

The maximally flat approximation to the ideal lowpass filter response is best near the origin but not so good near the band edge. Chebyshev filters have better responses near the band edge, with lower insertion loss near the edges, but at the cost of ripples in the passband. Example reflection and transmission responses are shown in Figure 2.4.2 for a seventh-order and a sixth-order Chebyshev lowpass filter.

4.5.1: Chebyshev Filter Design

The general form of the Chebyshev transmission coefficient is

$$|T(s)|^2 = \frac{1}{1+\varepsilon^2|K(s)|^2} \tag{4.5.1}$$

where ε is the ripple factor and defines the **passband ripple** (**PBR**):

$$\text{PBR} = 1+\varepsilon^2, \quad \text{or in decibels} \quad R_{\text{dB}} = \text{PBR}|_{\text{dB}} = 10\log(1+\varepsilon^2) \tag{4.5.2}$$

The PBR can be seen in the transmission response, $|T(s)|^2$, in Figure 2.4.2. In the passband the peaks of the lossless filter response have $|T(s)|^2 = 1$ and the minimums of the ripple response all have $|T(s)|^2 = 1/(1+\varepsilon^2) = 1/\text{PBR}$. Consequently Chebyshev filters are also known as **equiripple** all-pole lowpass filters. Also note that the corner radian frequency, $\omega = 1$ for the lowpass filter prototype, has a transmission response (i.e., insertion loss IL) of $|T(s)|^2 = 1/(1+\varepsilon^2)$, whereas the Butterworth transmission response was at half power at the corner frequency. For the Chebyshev filter, the insertion loss at the corner frequency is the ripple:

$$\text{IL} = 1R_{\text{dB}} = 10\log(1+\varepsilon^2) \tag{4.5.3}$$

For the nth-order Chebyshev (lowpass filter) approximation, the square of the characteristic function is

$$|K_n(\omega)|^2 = \begin{cases} \cos^2[n\ \cos^{-1}(\omega)] & -1 \le \omega \le 1 \\ \cosh^2[n\cosh^{-1}(|\omega|)] & \omega \le -1,\ \omega \ge 1 \end{cases} \tag{4.5.4}$$

which can be expressed as a polynomial. For example, with $n = 3$,

$$K_3(\omega) = 4\omega^3 - 3\omega, \quad \text{for all } \omega \tag{4.5.5}$$

(This equivalence was derived by Pafnuty Chebyshev.) It is surprising that the trigonometric expression has such a simple polynomial equivalence. From Equation (2.2.11) the transmission coefficient is (for $-1 \le \omega \le 1$)

$$|T(\omega)|^2 = \frac{1}{1+\varepsilon^2\cos^2[n\cos^{-1}(\omega)]} \tag{4.5.6}$$

and the reflection coefficient is

$$|\Gamma_1(\omega)|^2 = \frac{\varepsilon^2\cos^2[n\cos^{-1}(\omega)]}{1+\varepsilon^2\cos^2[n\cos^{-1}(\omega)]} \tag{4.5.7}$$

Factorizing the denominator of either Equation (4.5.6) or Equation (4.5.7) yields the following roots (of the denominators of $\Gamma_1(s)$ and $T(s)$):

$$\begin{aligned} s_i = &\sin\left[\frac{(2i-1)\pi}{2n}\right]\sinh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{\varepsilon}\right)\right] \\ &+\jmath\cos\left[\frac{(2i-1)\pi}{2n}\right]\cosh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{\varepsilon}\right)\right] \quad i = 1,2,\ldots,n \end{aligned} \tag{4.5.8}$$

The roots of the numerator of $\Gamma_1(s)$ in the s plane are

$$s_k = \jmath\cos\frac{(2k-1)\pi}{2n} \quad k = 1,2,\ldots,n \tag{4.5.9}$$

Equations (4.5.8) and (4.5.9) can be used to obtain the reflection and transmission coefficients directly in the s domain.

4.5.2: Chebyshev Approximation and Recursion

The characteristic function of the Chebyshev approximation can be obtained from the recursion formula,

$$K_n(\omega) = 2\omega K_{n-1}(\omega) - K_{n-2}(\omega) \tag{4.5.10}$$

Response 1 dB **down**				
Ripple	$n=3$	$n=5$	$n=7$	$n=9$
0.01 dB	1.564	1.192	1.097	1.058
0.1 dB	1.202	1.071	1.036	1.022
0.2 dB	1.127	1.045	1.023	1.014
1 dB	1.000	1.000	1.000	1.000
3 dB	–	–	–	–
Response 3 dB **down**				
Ripple	$n=3$	$n=5$	$n=7$	$n=9$
0.01 dB	1.877	1.291	1.145	1.087
0.1 dB	1.389	1.134	1.068	1.041
0.2 dB	1.284	1.099	1.050	1.030
1 dB	1.095	1.0338	1.017	1.010
3 dB	1.000	1.000	1.000	1.000

Table 4.5.1: Radian frequencies at which the transmission response of an nth Chebyshev filter is down 1 dB and 3 dB for a corner frequency $\omega_0 = 1$ rad/s. (Note that ω_0 is the radian frequency at which the transmission response of a Chebyshev filter is down by the ripple, see Figure 2.4.2.)

with

$$K_1(\omega) = \omega; \qquad K_2(\omega) = 2\omega^2 - 1 \tag{4.5.11}$$

For example, with $n = 3$,

$$K_3(\omega) = 2\omega K_{3-1}(\omega) - K_{3-2}(\omega) \tag{4.5.12}$$

$$= 2\omega(2\omega^2 - 1) - \omega = 4\omega^3 - 2\omega - \omega = 4\omega^3 - 3\omega \tag{4.5.13}$$

4.5.3: Bandwidth Consideration

At the corner frequency of the Chebyshev filter the transmission response is down by the amount of the ripple. This can be seen in Figure 2.4.2. However, the bandwidth of a filter is usually specified in terms of its 1 dB or 3 dB bandwidth at which the transmission response is down 1 dB or 3 dB, respectively, from its maximum response. The radian frequencies at which the responses of various orders of Chebyshev filters are 1 dB down and 3 dB down are given in Table 4.5.1. By frequency scaling the Chebyshev response, the filter can be designed for a specified 1 dB or 3 dB bandwidth.

This page titled 4.5: The Chebyshev Lowpass Approximation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.5: The Chebyshev Lowpass Approximation** has no license indicated.

4.6: Element Extraction

In the previous two sections the mathematical responses of Butterworth and Chebyshev filters were derived for various orders. In this section it will be shown how these filters can be implemented with inductors and capacitors using what is called **ladder synthesis** [4].

4.6.1: Ladder Synthesis

To obtain the element values yielding the desired transfer function, an impedance or admittance function must first be obtained. The impedance or admittance function can be readily obtained from the input reflection coefficient of a network, but for now the focus is on synthesizing a given impedance function. A general impedance function can be expressed as

$$Z(s) = \frac{a_n(s^2+\omega_1^2)(s^2+\omega_3^2)(s^2+\omega_5^2)\cdots}{b_m s(s^2+\omega_2^2)(s^2+\omega_4^2)(s^2+\omega_6^2)\cdots} \tag{4.6.1}$$

where a_n and b_m are constants. This can be realized using L and C elements in a network terminated by a resistor, provided that the degree of the numerator and denominator differ by no more than unity (i.e., $|m-n| \leq 1$). In

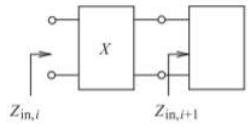


Figure 4.6.1: Extraction of a network X to reduce an impedance $Z_{\text{in},\,i}$ to a lower-order impedance $Z_{\text{in},\,i+1}$.

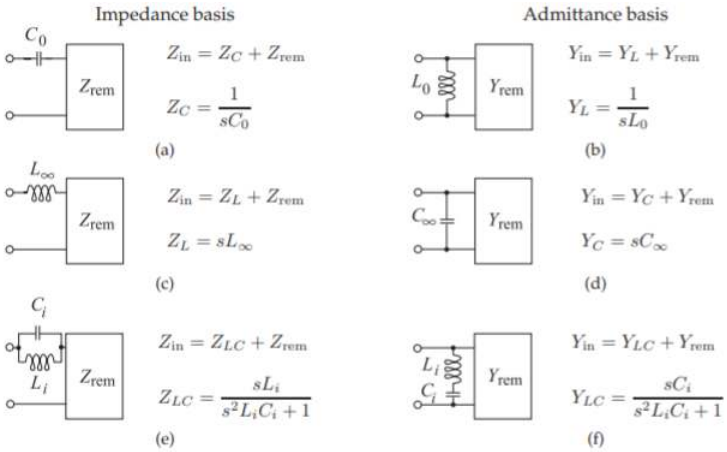


Figure 4.6.2: Synthesis of impedance and admittance functions. Starting with an impedance function $Z(s)$: (a) extraction of a series capacitor; (c) extraction of a series inductor; and (e) extraction of a series parallel LC block i. Starting with an admittance function $Y(s)$: (b) extraction of a shunt inductor; (d) extraction of a shunt capacitor; and (f) extraction of a shunt series LC block.

the case of a doubly terminated network, this resistor is the load. The element extraction procedure, shown in Figure 4.6.1, involves extracting a network X from $Z_{\text{in},\,i}$, leaving a reduced-order impedance $Z_{\text{in},\,i+1}$.

The extraction of inductors and capacitors is illustrated in Figure 4.6.2. Thus, following the extraction of an element or a pair of elements an impedance, Z_{rem}, or admittance, Y_{rem}, remains that can be similarly simplified. For example, and referring to Figure 4.6.2(a), $Z(s) = 1/(sC) + Z_{\text{rem}}$. So a pole of $Z(s)$ at DC requires the extraction of a series capacitor of value (see Figure 4.6.2(a))

$$C_0 = \left.\frac{1}{sZ(s)}\right|_{s=0} \tag{4.6.2}$$

while a pole at infinity requires the extraction of a series inductor of value (see Figure 4.6.2(c))

$$L_\infty = \left.\frac{Z(s)}{s}\right|_{s=\infty} \tag{4.6.3}$$

Another possibility is a pole at a finite frequency (call this ω_0), which requires the extraction of a series parallel LC block, as shown in Figure 4.6.2(e), with elements of value

$$C_i = \left.\frac{s}{(s^2+\omega_2^2)Z(s)}\right|_{s=j\omega_0} \quad \text{and} \quad L_i = \frac{1}{\omega_0^2 C_i} \tag{4.6.4}$$

The extraction process can also be carried out on an admittance basis. First

$$Y(s) = \frac{b_m s(s^2+\omega_2^2)(s^2+\omega_4^2)(s^2+\omega_s^2)\cdots}{a_n(s^2+\omega_1^2)(s^2+\omega_3^2)(s^2+\omega_5^2)\cdots} \tag{4.6.5}$$

Now a pole at zero requires the extraction of a shunt inductor of value (see Figure 4.6.2(b))

$$L_0 = \left.\frac{1}{sY(s)}\right|_{s=0} \tag{4.6.6}$$

and a pole at infinity requires the extraction of a shunt capacitor of value (see Figure 4.6.2(d))

$$C_\infty = \left.\frac{Y(s)}{s}\right|_{s=\infty} \tag{4.6.7}$$

A pole at a finite frequency requires the extraction of a shunt-series LC block (as shown in Figure 4.6.2(f)) with values

$$L_i = \left.\frac{s}{(s^2+\omega_0^2)Y(s)}\right|_{s=j\omega_0} \quad \text{and} \quad C_i = \frac{1}{\omega_0^2 L_i} \tag{4.6.8}$$

Many aspects of filter synthesis can seem abstract when presented in full generality. Consequently it is common to illustrate filter synthesis concepts using examples. Following this time-honored tradition, an example is now presented.

Example 4.6.1: Element Extraction for a Third-Order Lowpass Filter

A third-order maximally flat filter has the reflection coefficient

$$\Gamma_1(s) = \frac{s^3}{(s+1)(s^2+s+1)} \tag{4.6.9}$$

Synthesize this filter as a doubly terminated network.

Solution

The reflection coefficient function (Equation (4.6.9)) has all its poles located at infinity, so the corresponding network realization must be made of simple L or C elements and terminated in a resistor. Hence, referring to Figure 2.2.1 and considering a $1\ \Omega$ system,

$$Z_{\text{in},\,1}(s) = \frac{1+\Gamma_1(s)}{1-\Gamma_1(s)} = \frac{2s^3+2s^2+2s+1}{2s^2+2s+1} \tag{4.6.10}$$

Note that the input impedance approaches infinity as the frequency goes to infinity, hence a series inductor must be extracted. The value of this inductor is

$$L_{\infty 1} = \left.\frac{Z_{\text{in},\,1}(s)}{s}\right|_{s=\infty} = 1\ \text{H} \tag{4.6.11}$$

The filter is developed by extracting one element at a time. Following the extraction of the first element, the second-stage impedance is left. Now the impedance function is

$$\begin{aligned} Z_{\text{in},\,2}(s) = Z_{\text{in},\,1}(s) - sL_{\infty 1} &= \frac{2s^3+2s^2+2s+1}{2s^2+2s+1} - sL_{\infty 1} \\ &= \frac{2s^3+2s^2+2s+1-s(2s^2+2s+1)}{2s^2+2s+1} = \frac{s+1}{2s^2+2s+1} \end{aligned}$$

Note that the stage impedance above, $Z_{\text{in},\,2}$, approaches zero as the frequency goes to infinity. There is not a single series element that would cause this. However, the stage admittance function,

$$Y_{\text{in, 2}}(s) = \frac{1}{Z_{\text{in, 2}}(s)} = \frac{2s^2 + 2s + 1}{s + 1} \tag{4.6.12}$$

goes to infinity as the frequency approaches infinity and so a shunt capacitor is extracted:

$$Y_{\text{in, 3}}(s) = Y_{\text{in, 2}}(s) - sC_{\infty 2} = \frac{1}{s + 1} \tag{4.6.13}$$

where

$$C_{\infty 2} = 2F \tag{4.6.14}$$

So sometimes it is more convenient to consider extraction of an admittance and sometimes it is better to consider extraction of an impedance.

By examining the remaining stage impedance, it is seen that a pole exists at infinity, and so a series inductor, $L_{\infty 3}$, is extracted. The value of this inductor comes from

$$Z_{\text{in, 3}} = \frac{1}{Y_{\text{in, 3}}} = s + 1\,\Omega \tag{4.6.15}$$

and so the inductor value is

$$L_{\infty 3} = \left.\frac{s + 1}{s}\right|_{S=\infty} = 1\text{ H} \tag{4.6.16}$$

The final step is to extract a load of value 1 as follows:

$$Z_{\text{in, 4}} = Z_{\text{in, s}} - sL_{\infty 3} = 1\,\Omega \tag{4.6.17}$$

This example synthesized a doubly terminated network. The resulting network, called a ladder circuit, is shown in Figure 4.6.3. The left-most $1\,\Omega$ resistor is part of the source.

This circuit has a dual form consisting of two shunt capacitors separated by a series inductor. The dual circuit derives from realizing the admittance function obtained from the reflection coefficient. Other network extraction techniques are presented in Scanlan and Levy [4, 5] and Matthaei et al. [1].

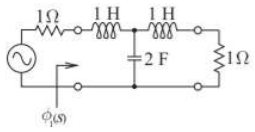


Figure 4.6.3: Synthesized maximally flat network with a thirdorder lowpass reflection response.

4.6.2: Summary

The input impedance function of a lumped-element circuit can always be expressed as the ratio of two polynomials in s and the order of the numerator and the denominator polynomials can differ by at most one [5]. If the orders differ by one, then a single inductor or capacitor can always be extracted, however, the remaining impedance function may not be realizable. This indicates that a more complex LC (and possibly R) combination is required. To be able to systematically extract arbitrarily complex circuits, a long list of possible functions, such as those shown in Figure 4.6.2, is required. For most of the circuits of interest the LC combinations shown in Figure 4.6.2 are sufficient. The next example describes impedance function extraction that requires an LC combination.

Example 4.6.2: Element Extraction of an Impedance Function

Realize the impedance function $Z_w = \frac{4s^3+4s^2+2s+2}{4s^2+2s+1}$.

Solution

The order of the numerator is 1 greater than the order of the denominator and this indicates that a series inductor is perhaps present. The series inductance is

$$L_1 = \left.\frac{Z_w(s)}{(s)}\right|_{z=\infty} = 1 \text{ H} \tag{4.6.18}$$

The remaining impedance is

$$Z_{\text{in, 2}} = Z_w - sL_1 = \frac{4s^3 + 4s^2 + 2s + 2}{4s^2 + 2s + 1} - s = \frac{2s^2 + s + 2}{4s^2 + 2s + 1} \tag{4.6.19}$$

The numerator and denominator of $Z_{\text{in, 2}}$ have the same order. Therefore a simple L or C element cannot be used to reduce the complexity of the impedance function. Thus an initial series inductor was not the right choice and the extraction must backtrack.

Figure 4.6.2 shows several element combinations that can be used to reduce the complexity of an impedance function. Insight into which alternative to choose comes from factoring z_w, and note that real roots are required, thus

$$Z_w = \frac{4s^3 + 4s^2 + 2s + 2}{4s^2 + 2s + 1} = \frac{(2s^2 + 1)(2s + 2)}{4s^2 + 2s + 1} \tag{4.6.20}$$

Examination of Figure 4.6.2 reveals that a ready fit to Z_w is not found. Instead consider the admittance function

$$Y_w = \frac{1}{Z_w} = \frac{4s^2 + 2s + 1}{(2s^2 + 1)(2s + 2)} \tag{4.6.21}$$

So the reduction shown in Figure 4.6.2(f) looks like the right candidate. The general choice for the element is

$$y_x = \frac{as}{bs^2 + 1} \tag{4.6.22}$$

Choosing $b = 2$ now reduces complexity (since part of the factored denominator of Y_w now occurs), so

$$Y_w = \frac{as}{2s^2 + 1} + \left(\frac{4s^2 + 2s + 1}{(2s^2 + 1)(2s + 2)} - \frac{as}{2s^2 + 1}\right) \tag{4.6.23}$$

$$= \frac{as}{2s^2 + 1} + \left(\frac{(4 - 2a)s^2 + (2 - 2a)s + 1}{(2s^2 + 1)(2s + 2)}\right) \tag{4.6.24}$$

Choose $a = 1$,

$$Y_w = \frac{s}{(2s^2 + 1)} + \frac{2s^2 + 1}{(2s^2 + 1)(2s + 2)} = \frac{s}{2s^2 + 1} + \frac{1}{2s + 2} \tag{4.6.25}$$

$$= \frac{s}{2s^2 + 1} + Y_{\text{in, 2}} \tag{4.6.26}$$

So $C_1 L_1 = b = 1$, $C_1 = a = 1$ F, $L_1 = 2$ H, and

$$Y_{\text{in, 2}} = \frac{1}{(2s + 2)} \quad \text{or} \quad Z_{\text{in, 2}} = \frac{1}{Y_{\text{in, 2}}} = 2s + 2 \tag{4.6.27}$$

The final network is

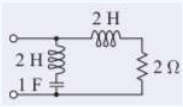


Figure 4.6.4

This page titled 4.6: Element Extraction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.6: Element Extraction** has no license indicated.

4.7: Butterworth and Chebyshev Filters

The nth-order lowpass filters constructed from the Butterworth and Chebyshev polynomials have the ladder circuit forms of Figure 4.7.1(a or b). Figure 4.7.1 uses several shorthand notations commonly used with filters. First, note that there are two prototype forms designated Type 1 and Type 2, and these are referred to as duals of each other. The two prototype forms have identical responses with the same numerical element values $g_1, \ldots, g_n$. Consider the Type 1 prototype of Figure 4.7.1(a). The right-most element is the resistive load, which is also known as the $(n+1)$th element. The next element to the left of this is either a shunt capacitor (of value g_n) if n is even, or a series inductor (of value g_n) if n is odd. So for the Type 1 prototype, the shunt capacitor next to the load does not exist if n is odd. The same interpretation applies to the circuit in Figure 4.7.1(b).

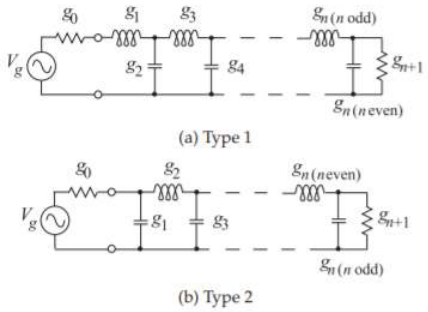


Figure 4.7.1: Filter prototypes in the Cauer topology. Here n is the order of the filter.

Order, n	2	3	4	5	6	7	8	9
g_1	1.4142	1	0.7654	0.6180	0.5176	0.4450	0.3902	0.3473
g_2	1.4142	2	1.8478	1.6180	1.4142	1.2470	1.1111	1
g_3	1	1	1.8478	2	1.9318	1.8019	1.6629	1.5321
g_4		1	0.7654	1.6180	1.9318	2	1.9615	1.8794
g_5			1	0.6180	1.4142	1.8019	1.9615	2
g_6				1	0.5176	1.2470	1.6629	1.8794
g_7					1	0.4450	1.1111	1.5321
g_8						1	0.3902	1
g_9							1	0.3473
g_{10}								1

Table 4.7.1: Coefficients of the Butterworth lowpass prototype filter normalized to a radian corner frequency of 1 rad/s and a 1 Ω system impedance (i.e., $g_0 = 1 = g_{n+1}$).

4.7.1: Butterworth Filter

A generalization of the example of the previous section leads to a formula for the element values of a ladder circuit implementing a Butterworth lowpass filter. For a maximally flat or Butterworth response the element values of the circuit in Figure 4.7.1(a and b) are

$$g_r = 2\sin\left\{(2r-1)\frac{\pi}{2n}\right\} \quad r = 1, 2, 3, \ldots, n \tag{4.7.1}$$

and $g_0 = 1 = g_{n+1}$. Table 4.7.1 lists the coefficients of Butterworth lowpass prototype filters up to ninth order.

Example 4.7.1: Fourth-Order Butterworth Lowpass Filter

Derive the fourth-order Butterworth lowpass prototype of Type 1.

Solution

From Equation (4.7.1),

$$g_1 = 2\sin[\pi/(2\cdot 4)] = 0.765369 \text{ H} \tag{4.7.2}$$
$$g_2 = 2\sin[3\pi/(2\cdot 4)] = 1.847759 \text{ F} \tag{4.7.3}$$
$$g_3 = 2\sin[5\pi/(2\cdot 4)] = 1.847759 \text{ H} \tag{4.7.4}$$
$$g_4 = 2\sin[7\pi/(2\cdot 4)] = 0.765369 \text{ F} \tag{4.7.5}$$

Thus the fourth-order Butterworth lowpass prototype circuit with a corner frequency of 1 rad/s is as shown in Figure 4.7.2.

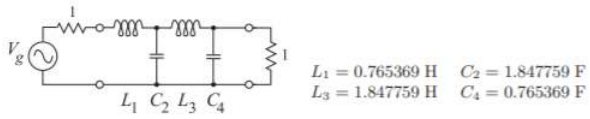


Figure 4.7.2: Fourthorder Butterworth lowpass filter prototype.

4.7.2: Chebyshev Filter

For a Chebyshev response, the element values of the lowpass prototype shown in Figure 4.7.1 are found from the recursive formula [1, 6, 7]:

$$g_0 = 1 \quad g_1 = \frac{2a_1}{\gamma} \tag{4.7.6}$$

$$g_{n+1} = \begin{Bmatrix} 1 & n \text{ odd} \\ \tanh^2(\beta/4) & n \text{ even } G_n \text{ series inductor} \\ \coth^2(\beta/4) & n \text{ even } G_n \text{ shunt capacitor} \end{Bmatrix} \tag{4.7.7}$$

$$g_k = \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}}, \quad k = 1, 2, \ldots, n \tag{4.7.8}$$

$$a_k = \sin\left[\frac{(2k-1)\pi}{2n}\right] \quad k = 1, 2, \ldots, n \tag{4.7.9}$$

where

$$\gamma = \sinh\left(\frac{\beta}{2n}\right) \tag{4.7.10}$$

$$b_k = \gamma^2 + \sin^2\left(\frac{k\pi}{n}\right) \quad k = 1, 2, \ldots, n \tag{4.7.11}$$

$$\beta = \ln\left[\coth\left(\frac{R_{\text{dB}}}{2\cdot 20\log(2)}\right)\right] = \ln\left[\coth\left(\frac{R_{\text{dB}}}{17.3717793}\right)\right] \tag{4.7.12}$$

$$R_{\text{dB}} = 10\log(1+\varepsilon^2) \tag{4.7.13}$$

n is the order of the filter, and ε is the ripple factor and defines the level of the ripple in absolute terms. R_{dB} is the ripple expressed in decibels (the ripple is generally specified in decibels).

An interesting point to note here is that the source resistor, the value of which is given by g_0, and terminating resistor, the value of which is given by g_{n+1}, are only equal for odd-order filters. For an even-order Chebyshev filter the terminating resistor, g_{n+1}, will be different and a function of the filter ripple. Because it is generally desirable to have identical source and load impedances, Chebyshev filters are nearly always restricted to odd order. Thus the odd-order Chebyshev prototypes are as shown in Figure 4.7.3.

Also, for an odd-degree function (n is odd) there is a perfect match at DC,

$$|T(0)|^2 = 1 \tag{4.7.14}$$

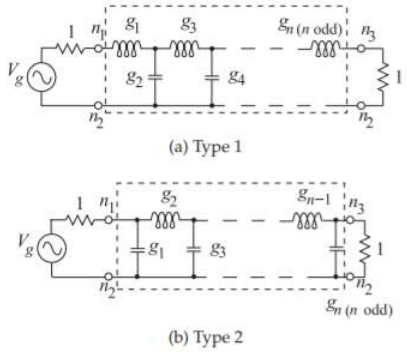


Figure 4.7.3: Odd-order Chebyshev lowpass filter prototypes in the Cauer topology. Here n is the order of the filter.

Order Ripple	$n = 3$					
	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.62918	1.03156	1.22754	2.02359	3.34874	0.85158
g_2	0.97028	1.14740	1.15254	0.99410	0.71170	1.10316
g_3	0.62918	1.03156	1.22754	2.02359	3.34874	0.85158
Order Ripple	$n = 5$					
	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.75633	1.14681	1.33944	2.13488	3.48129	0.97140
g_2	1.30492	1.37121	1.33702	1.09111	0.76192	1.37208
g_3	1.57731	1.97500	2.16605	3.00092	4.53755	1.80136
g_4	1.30492	1.37121	1.33702	1.09111	0.76192	1.37208
g_5	0.75633	1.14681	1.33944	2.13488	3.48129	0.97140
Order Ripple	$n = 7$					
	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.79694	1.18118	1.37226	2.16656	3.51852	1.00794
g_2	1.39242	1.42281	1.37820	1.11151	0.77220	1.43678
g_3	1.74813	2.09667	2.27566	3.09364	4.63898	1.93981
g_4	1.63313	1.57340	1.50016	1.17352	0.80381	1.62196
g_5	1.74813	2.09667	2.27566	3.09364	4.63898	1.93981
g_6	1.39242	1.42281	1.37820	1.11151	0.77220	1.43678
g_7	0.79694	1.18118	1.37226	2.16656	3.51852	1.00794
Order Ripple	$n = 9$					

Order Ripple	$n = 3$					
	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.81446	1.19567	1.38603	2.17972	3.53394	1.02347
g_2	1.42706	1.44260	1.39389	1.11918	0.76604	1.46186
g_3	1.80436	2.13455	2.30932	3.12143	4.66906	1.98372
g_4	1.71254	1.61672	1.53405	1.18967	0.81181	1.67776
g_5	1.90579	2.20537	2.37280	3.17463	4.72701	2.06485
g_6	1.71254	1.61672	1.53405	1.18967	0.81181	1.67776
g_7	1.80436	2.13455	2.30932	3.12143	4.66906	1.98372
g_8	1.42706	1.44260	1.39389	1.11918	0.76604	1.46186
g_9	0.81446	1.19567	1.38603	2.17972	3.53394	1.02347

Table 4.7.2: Coefficients of a Chebyshev lowpass prototype filter normalized to a radian corner frequency of $\omega_0 = 1$ rad/s and a $1\ \Omega$ system impedance (i.e., $g_0 = 1 = g_{n+1}$). The ripple factor, ε, is related to the ripple in decibels by Equation (4.7.13) (e.g., $\varepsilon = 0.1$ is a ripple of 0.0432 dB). (Note that ω_0 is the radian frequency at which the transmission response of a Chebyshev filter is down by the ripple, see Figure 2.4.2.)

while for an even-degree function (i.e., n is even) a mismatch exists of value

$$|T(0)|^2 = \frac{4R_L}{(R_L+1)^2} = \frac{1}{1+\varepsilon^2} \tag{4.7.15}$$

so that

$$R_L = g_{n+1} = \left[\varepsilon + \sqrt{(1+\varepsilon^2)}\right]^2 \tag{4.7.16}$$

Coefficients of several Chebyshev lowpass prototype filters with different levels of ripple and odd orders up to ninth order are given in Table 4.7.2.

4.7.3: Summary

A Butterworth filter has a monotonic response without ripple, but a relatively slow transition from the passband to the stopband. A Chebyshev filter has a rapid transition but has ripple in either the stopband or passband. Butterworth and Chebyshev filters are special cases of **elliptical filters**, which are also called **Cauer filters**. In general, an elliptical filter has ripple in both the stopband and the passband. The level of the ripple can be selected

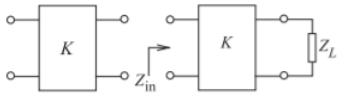


Figure 4.7.4: Impedance inverter (of impedance K in ohms): (a) represented as a two-port; and (b) the two-port terminated in a load.

independently in each band. With zero ripple in the stopband, but ripple in the passband, an elliptical filter becomes a **Type I Chebyshev filter**. With zero ripple in the passband, but ripple in the stopband, an elliptical filter becomes a **Type II Chebyshev filter**. With no ripple in either band the elliptical filter becomes a Butterworth filter. With ripple in both the passband and stopband, the transition between the passband and stopband can be made more abrupt or alternatively the tolerance to component variations increased.

Another type of filter is the **Bessel filter** which has maximally flat group delay in the passband, which means that the phase response has maximum linearity across the passband. The **Legendre filter** (also known as the **optimum "L" filter**) has a high transition rate from passband to stopband for a given filter order, and also has a monotonic frequency response (i.e., without ripple). It is a compromise between the Butterworth filter, with monotonic frequency response but slower transition and the Chebyshev filter, which has a faster transition but ripples in the frequency response.

More in-depth discussions of a large class of filters along with coefficient tables and coefficient formulas are available in Matthaei et al. [1], Hunter [3], Daniels [8], Lutovac et al. [9], and in most other books dedicated solely to microwave filters.

This page titled 4.7: Butterworth and Chebyshev Filters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.7: Butterworth and Chebyshev Filters** has no license indicated.

4.8: Impedance and Admittance Inverters

Inverters are two-port networks used in many RF and microwave filters. The input impedance of an inverter terminated in an impedance Z_L is $1/Z_L$. Impedance and admittance inverters are the same network, with the distinction being whether siemens or ohms are used to define them. An inverter is sometimes called a **unit element** (**UE**). At frequencies of a few hundred megahertz and below an inverter can be realized using operational and transconductance amplifiers. At microwave frequencies the simplest inverter is a one-quarter wavelength long line. In RF and microwave filter design they are used to convert a series element into a shunt element. It is much easier to realize shunt elements in distributed circuits than series elements. Similar circuit transformations enable an inductor to be replaced by a capacitor.

The schematic representation of an impedance inverter is shown in Figure, 2.7.4(a). The constitutive property of the inverter is that the input impedance of the terminated impedance inverter in Figure 2.7.4(b) is

$$Z_{\text{in}} = \frac{K^2}{Z_L} \tag{4.8.1}$$

So the inverter both inverts the load impedance and scales it. Similarly, if Port 1 is terminated in Z_L the input impedance at Port 2 is Z_{in} as defined above.

An impedance inverter has the value K (in ohms), and sometimes K is called the characteristic impedance of the inverter. Sometimes K is just

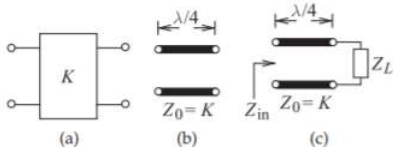


Figure 4.8.1: Inverter equivalence: (a) two-port impedance inverter (of impedance K): (b) a quarter-wave transmission line of characteristic impedance $Z_0 = K$; and (c) a terminated one-quarter wavelength long line.

called the impedance of the inverter. For an admittance inverter J is used and is called the characteristic admittance of the inverter, and sometimes just the admittance of the inverter. They are related as $J = 1/K$. In Section 2.4.6 of [10] it is shown that a $\lambda/4$ long line with a load has an input impedance that is the inverse of the load, normalized by the square of the characteristic impedance of the line. So an inverter can be realized at microwave frequencies using a one-quarter wavelength long transmission line (see Figure 4.8.1(b)). For the configuration shown in Figure 4.8.1(c),

$$Z_{\text{in}} = \frac{K^2}{Z_L} \tag{4.8.2}$$

4.8.1: Properties of an Impedance Inverter

An impedance inverter has the $ABCD$ matrix

$$\mathbf{T} = \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \tag{4.8.3}$$

where K is called the characteristic impedance of the inverter. With a load impedance, Z_L (at Port 2), the input impedance (at Port 1) is (as expected)

$$Z_{\text{in}}(s) = \frac{AZ_L + B}{CZ_L + D} = \frac{\jmath K}{(\jmath/K)Z_L} = \frac{K^2}{Z_L} \tag{4.8.4}$$

Now the $ABCD$ matrix of the transmission line of Figure 4.8.1(b) is

$$\begin{bmatrix} \cos\theta & \jmath Z_0 \sin\theta \\ (\jmath/Z_0)\sin\theta & \cos\theta \end{bmatrix} \tag{4.8.5}$$

which is identical to Equation (4.8.3) when the electrical length is $\theta = \pi/2$ (i.e., when the line is $\lambda/4$ long). The inverter is shown in Figure 2.7.4(a) as a two-port and its implementation as a $\lambda/4$ long line is shown in Figure 2.7.4(c). The bandwidth over which the

line realizes an impedance inverter is limited, however, as it is an ideal inverter only at the frequency at which it is $\lambda/4$ long.

4.8.2: Replacement of a Series Inductor by a Shunt Capacitor

A series inductor can be replaced by a shunt capacitor surrounded by a pair of inverters followed by a negative unity transformer (i.e., an inverter with $K = 1$). This equivalence is shown in Figure 4.8.2 and this will now be shown mathematically.

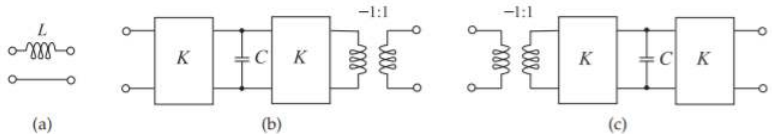


Figure 4.8.2: Equivalent realizations of a series inductor: (a) as a two-port; (b) its realization using a capacitor, inverters of characteristic impedance K, and a negative unity transformer; and (c) an alternative realization. $C = L/K^2$.

From Table 2.4.1 of [11], the $ABCD$ matrix of the series inductor shown in Figure 4.8.2(a) (which has an impedance of sL) is

$$\mathbf{T}_L = \begin{bmatrix} 1 & sL \\ 0 & 1 \end{bmatrix} \tag{4.8.6}$$

and the $ABCD$ matrix of the shunt capacitor (which has an admittance of sC) is, from Table 2.4.1 of of [11],

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \tag{4.8.7}$$

The $ABCD$ matrix of an inverter with K in ohms (generally the unit is dropped and ohms is assumed) is

$$\mathbf{T}_2 = \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \tag{4.8.8}$$

and finally, the $ABCD$ matrix of a negative unity transformer, $n = -1$, is, from Table 2.4.1 of [11],

$$\mathbf{T}_3 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \tag{4.8.9}$$

Then the $ABCD$ matrix of the cascade shown in Figure 4.8.2(b) is

$$\begin{aligned} \mathbf{T}_C &= \mathbf{T}_2\mathbf{T}_1\mathbf{T}_2\mathbf{T}_3 \\ &= \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \begin{bmatrix} 0 & -\jmath K \\ -\jmath/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} \jmath sCK & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} 0 & -\jmath K \\ -\jmath/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & sCK^2 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{4.8.10}$$

Thus $T_C = T_L$ if $L = CK^2$ (compare Equations (4.8.6) and (4.8.10)). Thus a series inductor can be replaced by a shunt capacitor with an inverter before and after it and with a negative unity transformer. The unity transformer

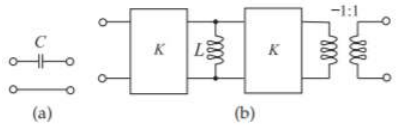


Figure 4.8.3: A series capacitor: (a) as a twoport; (b) its realization using a shunt inductor, inverters and negative unity transformer' $L = CK^2$.

may also be placed at the first port, as in Figure 4.8.2(c). Thus the two-ports shown in Figure 4.8.2 are all electrically identical, with the limitation being the frequency range over which the inverter can be realized. An interesting and important observation is that as

a result of the characteristic impedance of the inverter (e.g., $50\ \Omega$), a small shunt capacitor can be used to realize a large series inductance value.

Example 4.8.1: Inductor Synthesis Using an Inverter

Consider the network of Figure 4.8.2(c) with inverters having a characteristic impedance of $50\ \Omega$. What value of inductance is realized using a 10 pF capacitor?

Solution

$K = 50$, so $L = CK^2 = 10^{-11} \cdot 2500 = 25$ nH.

4.8.3: Replacement of a Series Capacitor by a Shunt Inductor

A series capacitor can be replaced by a shunt inductor plus inverters and a negative transformer (see Figure 4.8.3). The $ABCD$ parameters of the series capacitor in Figure 4.8.3(a) are

$$\mathbf{T} = \begin{bmatrix} 1 & 1/sC \\ 0 & 1 \end{bmatrix} \tag{4.8.11}$$

and here it is shown that the cascade in Figure 4.8.3(b) has the same $ABCD$ parameters. The cascade in Figure 4.8.3(b) has the $ABCD$ parameters

$$\begin{aligned} \mathbf{T} &= \begin{bmatrix} 0 & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/sL & 1 \end{bmatrix} \begin{bmatrix} 0 & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/sL & 1 \end{bmatrix} \begin{bmatrix} 0 & -jK \\ -j/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} jK/sL & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} 0 & -jK \\ -j/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & K^2/sL \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{4.8.12}$$

So the series capacitor, C, can be realized using a shunt inductor, L, inverters, and a negative unity transformer, and $C = L/K^2$. It is unlikely that this transformation would be exploited, as much better lower-loss capacitors can be realized at RF than inductors.

4.8.4: Ladder Prototype with Impedance Inverters

The transformations discussed in the previous two sections can be used to advantage in simplifying filters. In this section the transformations are

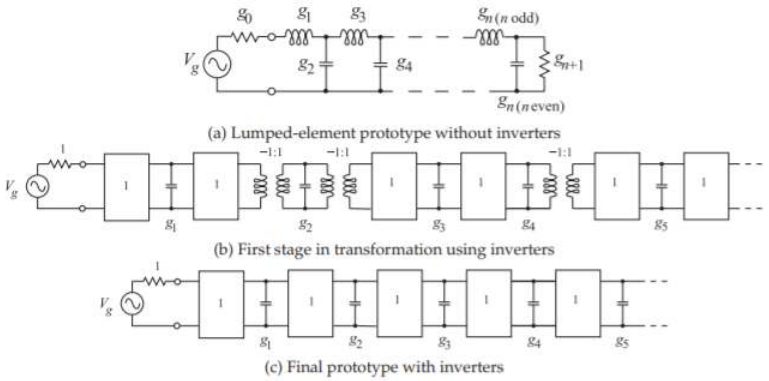


Figure 4.8.4: Ladder prototype filters using impedance inverters: (a) lumped-element prototype; (b) first stage in transformation using inverters; and (c) final stage.

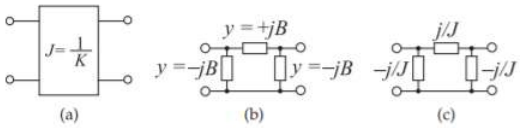


Figure 4.8.5: Admittance inverter: (a) as a two-port; (b) realized using lumped elements with $B = -J$; and (c) lumped equivalent circuit (the element values in (c) are impedances).

applied to the prototype lowpass ladder filter shown in Figure 4.8.4(a). The inductors in the ladder circuit are a particular problem as they have considerable resistance at microwave frequencies. The series inductors can be replaced by a circuit with capacitors, inverters, and transformers, as shown in Figure 4.8.4(b). This simplifies further to the realization shown in Figure 4.8.4(c), as the negative unity transformers only affect the phase of the transmission coefficient. So a lowpass ladder filter can be realized using just capacitors and inverters.

4.8.5: Lumped-Element Realization of an Inverter

The admittance inverter is functionally the same as the impedance inverter (see Figure 2.7.4(a)) and the schematic is the same (see Figure 4.8.5(a)). As will be shown, an inverter can be realized using frequency-invariant lossless elements (i.e., elements whose reactance or susceptance do not vary with frequency) using the network of Figure 4.8.5(b). Recall that J is used to identify an admittance inverter and K identifies an impedance inverter. If not specified by the context, the inverter (with value specified by a number) defaults to being an impedance inverter. Alternatively units can be used to indicate which type of inverter is being used. The function of

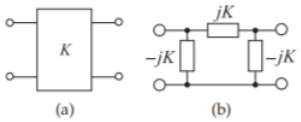


Figure 4.8.6: Impedance inverter: (a) as a two-port; and (b) its lumped equivalent circuit (the element values in (b) are impedances).

the inverter is the same in any case; both can be realized by one-quarter wavelength long lines, for example. For the remainder of this chapter it will be more convenient, most of the time, to use the admittance inverter, as many calculations will be in terms of admittances since most lumped elements in filters synthesis will be in shunt.

Now it will be shown that the lumped-element network of Figure 4.8.5(b) realizes an inverter. To do this the inverter and the lumped-element network must have the same two-port parameters. First, the $ABCD$ matrix of an inverter of characteristic admittance J is

$$\mathbf{T}_j = \begin{bmatrix} 0 & \jmath/J \\ \jmath J & 0 \end{bmatrix} \tag{4.8.13}$$

Referring to Table 2-1 of [11], the circuit of Figure 4.8.5(b) has the $ABCD$ matrix

$$\begin{aligned} T &= \begin{bmatrix} 1 & 0 \\ -\jmath B & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/(\jmath B) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath B & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1/(\jmath B) \\ -\jmath B & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath B & 1 \end{bmatrix} = \begin{bmatrix} 0 & -\jmath/B \\ -\jmath B & 0 \end{bmatrix} \end{aligned} \tag{4.8.14}$$

where B is the susceptance of the frequency-invariant elements. Equation (4.8.14) is identical to Equation (4.8.13) if $B = -J$. More practical equivalents of the circuit of Figure 4.8.5(b) can be derived, as shown later.

For completeness, the lumped-element equivalent of the impedance inverter is shown in Figure 4.8.6 (derived from Figure 4.8.5 with $J = 1/K$).

Example 4.8.2: Lumped Inverter Analysis

Demonstrate that Figure 4.8.5(b) is a lumped-element admittance inverter.

Solution

Terminating the network in Figure 4.8.5(b) results in the network shown in Figure 4.8.7(a). This is relabeled in Figure 4.8.7(b), where the elements are admittances. Then

$$y_{\text{in}} = y_3 // (y_1 \$ (y_2 + y_L)) \tag{4.8.15}$$

where // indicates "in parallel with" and \$ indicates "in series with." These are common shorthand notations in circuit calculations. Continuing on from Equation (4.8.15),

$$y_{\text{in}} = y_3 + \left[\frac{1}{y_1} + \left(\frac{1}{y_2 + y_L}\right)\right]^{-1} = y_3 + \left[\frac{y_2 + y_L + y_1}{y_1(y_2 + y_L)}\right]^{-1} \tag{4.8.16}$$

Substituting $y_1 = \jmath B$ and $y_2 = y_3 = -\jmath B$, this becomes

$$y_{\text{in}} = -\jmath B + \frac{\jmath B(y_L - \jmath B)}{y_L} = \frac{-\jmath B y_L + \jmath B y_L + B^2}{y_L} = \frac{B^2}{y_L} \tag{4.8.17}$$

Thus the lumped-element circuit of Figure 4.8.5(b) is a lumped-element admittance inverter of value B (in siemens).

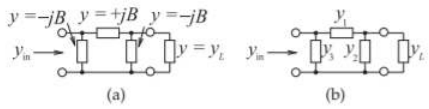


Figure 4.8.7: Terminated lumped-element admittance inverter.

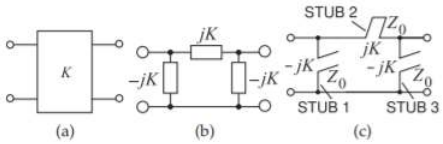


Figure 4.8.8: Narrowband inverter equivalents at frequency f_0: (a) impedance inverter with characteristic impedance K; (b) lumped-element equivalent network; and (c) inverter realized by short- and open-circuited stubs.

4.8.6: Narrowband Realization of an Inverter Using Transmission Line Stubs

In this section it will be shown that an impedance inverter can be implemented using short- and open-circuited stubs. The match is good over a narrow band centered at frequency f_0. An impedance inverter is shown in Figure 4.8.8(a) and its equivalent lumped-element network is shown in Figure 4.8.8(b). A stub-based implementation is shown in Figure 4.8.8(c), where there are short- and open-circuited stubs of characteristic impedance Z_0. The input impedance of the stubs is shown at the inputs of the stubs. The stubs have an electrical length θ at f_0 and the stubs are one-quarter wavelength long (i.e., resonant) at what is called the commensurate frequency, f_r.

Now it will be shown that the network of Figure 4.8.8(c) is a good representation of the inverter at f_0. This is done by matching $ABCD$ parameters. The $ABCD$ parameter matrix of an inverter is

$$T = \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \tag{4.8.18}$$

and, at frequency f_0, the $ABCD$ parameter matrix of the stub circuit of Figure 4.8.8(c) is

$$
\begin{aligned}
T &= \begin{bmatrix} 1 & 0 \\ -1/[\jmath Z_0 \tan(\theta)] & 1 \end{bmatrix} \begin{bmatrix} 1 & \jmath Z_0 \tan(\theta) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/[\jmath Z_0 \tan(\theta)] & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \jmath Z_0 \tan(\theta) \\ -1/[\jmath Z_0 \tan(\theta)] & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/[\jmath Z_0 \tan(\theta)] & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & \jmath Z_0 \tan(\theta) \\ \jmath/[Z_0 \tan(\theta)] & 0 \end{bmatrix}
\end{aligned} \tag{4.8.19}
$$

Thus, equating Equations (4.8.18) and (4.8.19), the stub network is a good representation of the inverter if

$$K = Z_0 \tan(\theta) \tag{4.8.20}$$

and so the required characteristic impedance of each stub at frequency f_0 is

$$Z_0 = \frac{K}{\tan(\theta)} = \frac{K}{\tan\left(\frac{\pi}{2}\frac{f_0}{f_r}\right)} \tag{4.8.21}$$

4.8.6.1: Special Case, $f_r = 2f_0$

In most designs the stub resonant frequency, f_r (also called the commensurate frequency), is chosen to be twice that of the center frequency of the design, f_0. So with $f_r = 2f_0$, then at f_0

$$Z_0 = \frac{K}{\tan\left(\frac{\pi}{2}\frac{f_0}{2f_0}\right)} = \frac{K}{\tan \pi/4} = K \tag{4.8.22}$$

and the input impedance of the stub is $\jmath K$. So the characteristic impedance of the transmission line stub is $Z_0 = K$.

This page titled 4.8: Impedance and Admittance Inverters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.8: Impedance and Admittance Inverters** has no license indicated.

4.9: Filter Transformations

So far the discussion has centered around lowpass filters. Filter design technology has developed so that the design of a corresponding lowpass filter is the essential first step, and this is used as a prototype to derive a filter with other characteristics. The three most important transformations are listed here:

1. Impedance scaling: A lowpass filter prototype is referenced to a standard impedance. Usually $1\ \Omega$ is used, so the reference source and load resistances are also $1\ \Omega$. To reference to a higher or lower impedance, scaling of the impedance of all elements of the filter is required.
2. Corner frequency scaling: The corner frequency of a lowpass filter prototype is normalized to $1\ \mathrm{rad/s}$. To reference to another frequency, the values of the elements must be altered so that they have the same impedance at the scaled frequency.
3. Filter type transformation: These transformations enable the circuit of a lowpass filter to be converted to a circuit with a bandpass, bandstop, or highpass response. The concept is that the response at DC is to be replicated at infinite frequency for the highpass (so capacitors become inductors, etc.); to be replicated at the center of the passband for a bandpass filter (so capacitors become a shunt LC circuit); and to be inverted at the center of a bandstop filter (so capacitors become a series LC circuit).

The transformations can be performed in any order. The key point is that the complexity of the lowpass circuit design is minimal since the number of elements does not grow until the filter type transformation is performed, so this step is often done last.

4.9.1: Impedance Transformation

The lowpass prototypes discussed up to now have been referred to a $1\ \Omega$ system. The system impedance can be changed to any level by simply scaling the impedances of all the circuit elements in a filter by the same amount, as shown in Figure 4.9.1. The impedance transformation follows the same procedure for all filter types (the procedure is the same for lowpass, highpass, bandpass, and bandstop filters, for example). It should also be noted that impedance scaling of a transmission line is simply multiplication of the line's characteristic impedance by the scale factor.

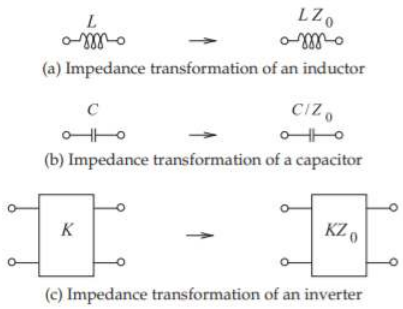


Figure 4.9.1: Impedance transformations. The impedances of the elements are increased by a factor Z_0.

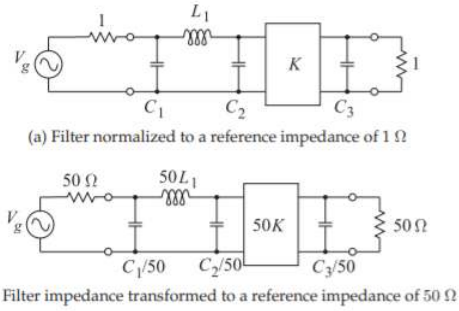


Figure 4.9.2: Impedance transformation of an example filter.

Example 4.9.1: Lowpass Filter Design

Consider the lowpass filter, with an inverter, shown in Figure 4.9.2(a). This filter is referenced to 1 Ω, as the source and load impedances are both 1 Ω. Redesign the filter so that the same frequency response is obtained with 50 Ω source and load impedances.

Solution

It is necessary to impedance transform from 1 Ω to 50 Ω. The resulting filter is shown in Figure 4.9.2(b). Each element has an impedance (resistance or reactance) that is 50 times larger than it had in the 1 Ω prototype.

4.9.2: Frequency Transformation: Lowpass

Frequency transformations differ depending on the type of the filter. So it is normal to frequency transform the lowpass prototype before converting the prototype to another form (such as bandpass). The lowpass prototypes normally have a band-edge or cutoff frequency at an angular frequency of unity, that is, 1 rad/s. The band-edge frequency can be transformed from unity to an arbitrary angular frequency ω_c, as shown in Figure 4.9.3, by scaling the reactive elements, as shown in Figures 4.9.4 and 4.9.5(a), so that

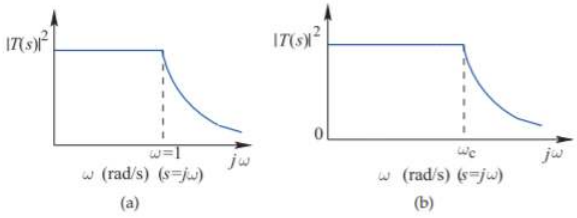


Figure 4.9.3: Frequency transformation of a lowpass filter response from (a) one normalized to a corner frequency of 1 rad/s, to (b) one with a radian corner frequency of ω_c.

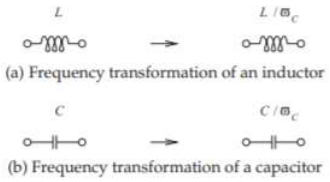


Figure 4.9.4: Frequency transformations. The impedance of the new (scaled) element at the new (scaled) frequency is the same as it was at the original frequency.

they have the same impedance at the transformed frequency as they did at the original frequency. The inverter is unchanged, as are the source and load resistances, since these are frequency independent.

4.9.3: Lowpass to Highpass Transformation

The transformation of a lowpass filter prototype to a highpass filter is shown diagrammatically in Figure 4.9.6. Mathematically ω in the transfer function of the lowpass prototype is replaced by $-1/\omega$, that is

$$T_{\text{highpass}}(\omega) = T_{\text{lowpass}}(-1/\omega) \tag{4.9.1}$$

In terms of a lumped element in the lowpass prototype circuit, if an element has an impedance $j\omega L$ and L is frequency independent, then the corresponding element in the highpass prototype filter has an impedance $1/(j\omega^2 L)$. Thus reactive elements are transformed as shown in Figure 4.9.5(b), where ω_0 is the corner frequency of both the lowpass and highpass prototype circuits. So inductors are transformed into capacitors and capacitors into inductors. For example, the odd-order lowpass filter prototypes shown in Figure 2.7.3 are transformed into the highpass filters shown in Figure 4.9.7.

4.9.4: Lowpass to Bandpass Transformation

Understanding the transformation of the lowpass filter into its corresponding bandpass form requires that the lowpass filter be considered with both its positive and negative frequency responses, as shown in Figure 4.9.8(a). This response is shifted in

frequency to obtain the bandpass response shown in Figure 4.9.8(b). Mathematically the radian frequency, ω, in the response

(a) LOWPASS PROTOTYPE TO LOWPASS TRANSFORMATION

Lowpass prototype	Lowpass element	Reactance transformation ω_0 = corner frequency
C_0	C_1	$C_1 = C_0/\omega_0$
L_0	L_1	$L_1 = L_0/\omega_0$

(b) LOWPASS PROTOTYPE TO HIGHPASS TRANSFORMATION

Lowpass prototype	Highpass element	Reactance transformation ω_0 = corner frequency
C_0	L_1	$L_1 = 1/(\omega_0 C_0)$
L_0	C_1	$C_1 = 1/(\omega_0 L_0)$

(c) LOWPASS PROTOTYPE TO BANDPASS TRANSFORMATION

Lowpass prototype	Bandpass element	Reactance transformation ω_0 = center of passband
C_0	L_1 in series with C_1	$C_1 = \alpha C_0/\omega_0 \quad L_1 = 1/(\alpha C_0 \omega_0)$
L_0	L_1 in parallel with C_1	$L_1 = \alpha L_0/\omega_0 \quad C_1 = 1/(\alpha L_0 \omega_0)$

(d) LOWPASS PROTOTYPE TO BANDSTOP TRANSFORMATION

Lowpass prototype	Bandstop element	Reactance transformation ω_0 = center of stopband
C_0	L_1 in parallel with C_1	$L_1 = \alpha/(C_0\omega_0) \quad C_1 = C_0/(\omega_0\alpha)$
L_0	L_1 in series with C_1	$L_1 = L_0/(\alpha\omega_0) \quad C_1 = \alpha/(L_0\omega_0)$

Figure 4.9.5: Transformations of the elements of a prototype lowpass filter to obtain specific filter types. The corner frequency of the lowpass prototype is 1 rad/s. In the transformations to bandpass and bandstop filters, $\omega_0 = 1/\sqrt{L_1 C_1} = \sqrt{\omega_1 \omega_2}$; ω_1 and ω_2 are the band-edge frequencies, and α is the transformation constant, $\alpha = \omega_0/(\omega_2 - \omega_1)$.

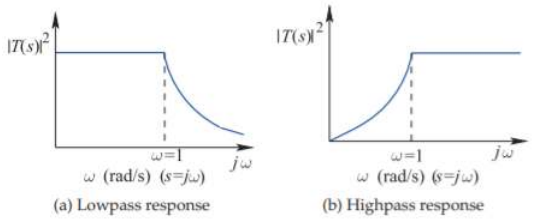


Figure 4.9.6: Lowpass to highpass transformation.

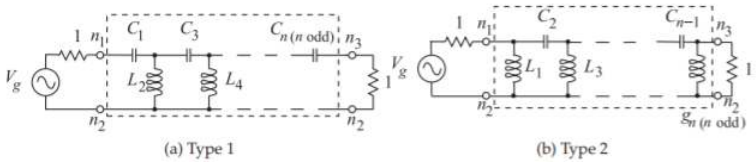
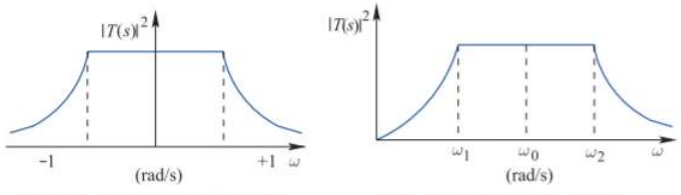


Figure 4.9.7: Odd-order Chebyshev highpass filter prototypes in the Cauer topology. Here n is the order of the filter.



https://eng.libretexts.org/@go/page/113728

Figure 4.9.8: Frequency responses in lowpass to bandpass transformation ($s = j\omega$).

function is replaced by its bandpass form,

$$\omega \rightarrow \left[\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right] \tag{4.9.2}$$

That is,

$$T_{\text{bandpass}}(\omega) = T_{\text{lowpass}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \tag{4.9.3}$$

and so the transfer function of the bandpass filter is derived from the transfer function of the bandpass filter with ω replaced by $(\omega/\omega_0 \ -\omega_0/\omega)$. This separately maps the -1 and $+1$ band-edge radian frequencies of the lowpass response to the bandpass frequencies ω_1 and ω_2:

$$-1 \rightarrow \left[\frac{\omega_1}{\omega_0} - \frac{\omega_0}{\omega_1}\right] \quad \text{and} \quad +1 \rightarrow \left[\frac{\omega_2}{\omega_0} - \frac{\omega_0}{\omega_2}\right] \tag{4.9.4}$$

Solving the above equations simultaneously yields the center frequency ω0 and the band-edge frequencies ω_1 and ω_2 with

$$\omega_0 = \sqrt{\omega_1 \omega_2} \tag{4.9.5}$$

and the so-called transformation constant

$$\alpha = \frac{\omega_0}{\omega_2 - \omega_1} \tag{4.9.6}$$

The resulting element conversions are given in Figure 4.9.5(c).

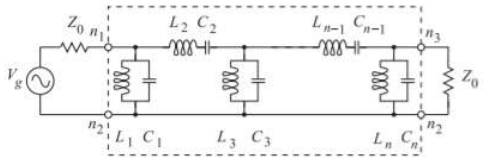


Figure 4.9.9: Lumped-element odd-order (nth-order) Chebyshev bandpass filter prototypes in the Type II Cauer topology.

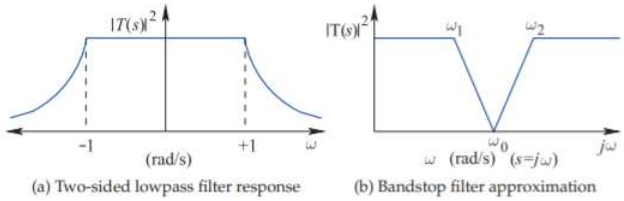


Figure 4.9.10: Frequency responses in lowpass to bandstop transformation.

As an example, a lumped-element Type 2 Cauer bandpass filter is shown in Figure 4.9.9. The shunt LC combination and the series LC combinations are resonators resonant at the center frequency of the filter. Here the filter is normalized to Z_0 source and load impedances. As these are the same, this filter topology only applies for an odd-order filter. Combining transformations, the element values of a lumped bandpass filter with center radian frequency $\omega_0 = 2\pi f_0$ and radian bandwidth $\omega_{\text{BW}} = 2\pi(f_2 - f_1)$ are as follows (g_r is from the lowpass prototype):

$$C_r = \begin{cases} \frac{g_r}{\omega_{\text{BW}} Z_0} & r = \text{odd} \\ \frac{\omega_{\text{BW}}}{\omega_0^2 g_r Z_0} & r = \text{even} \end{cases} \quad \text{and} \quad L_r = \begin{cases} \frac{\omega_{\text{BW}} Z_0}{\omega_0^2 g_r} & r = \text{odd} \\ \frac{g_r Z_0}{\omega_{\text{BW}}} & r = \text{even} \end{cases} \tag{4.9.7}$$

4.9.5: Lowpass to Bandstop Transformation

Again, consider both the positive and negative frequency responses of the lowpass filter prototype, as shown in Figure 4.9.10(a). This response is shifted in frequency to obtain the bandstop response shown in Figure 4.9.10(b). Mathematically the frequency, ω, in the response function is replaced by its bandstop form:

$$\omega \to \left[\alpha\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)\right]^{-1} \tag{4.9.8}$$

That is,

$$T_{\text{bandstop}}(\omega) = T_{\text{lowpass}}\left(\frac{1}{\alpha}\left[\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right]^{-1}\right) \tag{4.9.9}$$

The center frequency (corresponding to DC in the lowpass prototype response) is

$$\omega_0 = \sqrt{\omega_1 \omega_2} \tag{4.9.10}$$

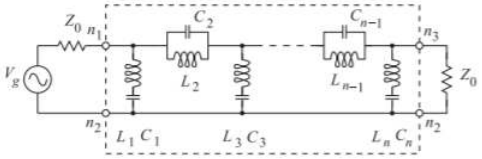


Figure 4.9.11: Lumped-element odd-order (nth-order) Chebyshev bandstop filter prototypes in the type II Cauer topology.

and the transformation constant is

$$\alpha = \frac{\omega_0}{\omega_2 - \omega_1} \tag{4.9.11}$$

where ω_1 and ω_2 are the band-edge radian frequencies. The resulting element conversions are given in Figure 4.9.5(d).

Combining transformations, the element values of a lumped bandstop filter with center radian frequency $\omega_0 = 2\pi f_0$ and radian bandwidth $\omega_{\text{BW}} = 2\pi(f_2 - f_1)$ are as follows:

$$C_r = \begin{cases} \frac{g_r \omega_{\text{BW}}}{\omega_0^2 Z_0} & r = \text{odd} \\ \frac{1}{\omega_{\text{BW}} g_r Z_0} & r = \text{even} \end{cases} \quad \text{and} \quad L_r = \begin{cases} \frac{Z_0}{\omega_{\text{BW}} g_r} & r = \text{odd} \\ \frac{g_r \omega_{\text{BW}} Z_0}{\omega_0^2} & r = \text{even} \end{cases} \tag{4.9.12}$$

A lumped-element type II Cauer bandstop filter is shown in Figure 4.9.11. The parallel LC combination and the series LC combinations are resonators resonant at the center frequency of the filter. The parallel LC resonator is an open circuit at the center frequency of the stopband and the series LC resonators are short circuits. The LC resonators are implemented using resonators, usually transmission line segments and not lumped components.

4.9.6: Transformed Ladder Prototypes

Combining the filter type transformations, and with appropriate use of inverters, the original lowpass prototype ladder filter and its various filter type transforms are shown in Figure 2.10.1.

This page titled 4.9: Filter Transformations is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.9: Filter Transformations** has no license indicated.

4.10: Cascaded Line Realization of Filters

In this section, filters are presented that use cascaded sections of transmission line, realizing series inductors and shunt capacitors. For example, a short (less than one-quarter wavelength long) length of relatively high-impedance line behaves predominantly as a series inductance. Also, a very short (much less than one-quarter wavelength long) length of relatively low-impedance line acts predominantly as a shunt capacitance. So a Pi network of lumped elements can be realized with alternate sections of low- and high-impedance microstrip lines. Such an inductive line is shown in Figure 4.10.2(a), which has the two equivalent circuits shown in Figure 4.10.2(b and c). Basic transmission line theory gives the input reactance of the line of length, ℓ (with a low-impedance load),

$$X_L = Z_0 \sin(2\pi\ell/\lambda_g) \tag{4.10.1}$$

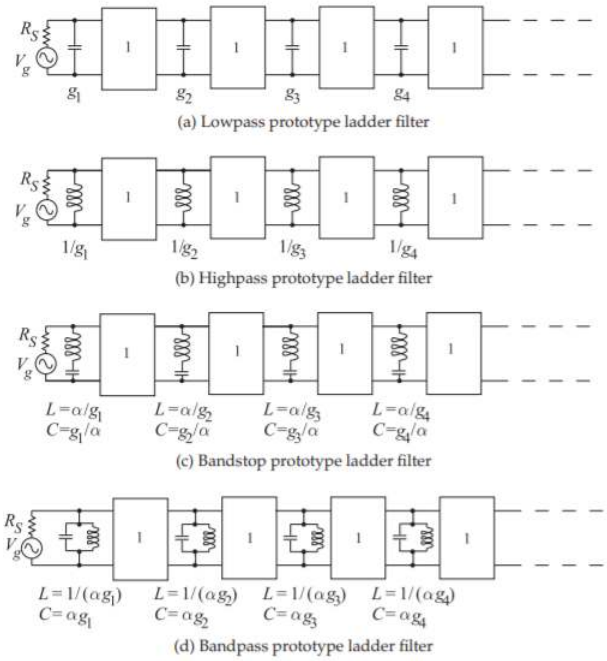


Figure 4.10.1: Ladder prototype filters.

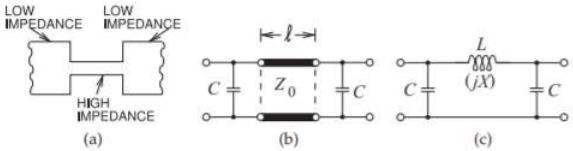


Figure 4.10.2: Inductive length of line with adjacent capacitive lines: (a) microstrip form; (b) lumped equivalent circuit; and (c) lumped-distributed equivalent circuit.

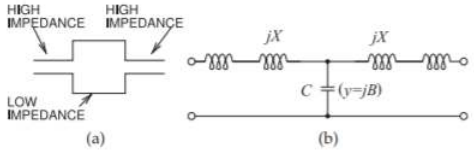
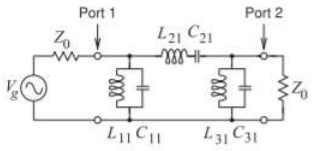


Figure 4.10.3: Capacitive length of line with adjacent inductive lines: (a) microstrip; and (b) lumped equivalent (the left-most and rightmost series inductors come from the high-impedance lines).



Butterworth	Chebyshev
$C_{11} = C_{31} = 31.8310$ pF	$C_{11} = C_{31} = 41.5369$ pF
$L_{11} = L_{31} = 795.775$ pH	$L_{11} = L_{31} = 609.824$ pH
$C_{21} = 159.155$ fF	$C_{21} = 188.299$ fF
$L_{21} = 159.155$ nH	$L_{21} = 134.522$ nH
	ripple, $\varepsilon = 0.1 = 0.043$ dB

Figure 4.10.4: Lumped-element 3rd-order bandpass filters in a $Z_0 = 50\,\Omega$ system with center frequency $f_0 = 1$ GHz and a 3 dB bandwidth of 10%.

so that the length of this predominantly inductive line is

$$\ell = \frac{\lambda_g}{2\pi}\sin^{-1}\left(\frac{\omega L}{Z_0}\right) \tag{4.10.2}$$

Previously it was shown that a short length of line having a relatively low characteristic impedance yields a capacitive element and this is shown, together with its equivalent circuit, in Figure 4.10.3. The predominating shunt capacitance is determined by first considering the susceptance

$$B = \frac{1}{Z_0}\sin\left(\frac{2\pi\ell}{\lambda_g}\right) \tag{4.10.3}$$

so that

$$\ell = \frac{\lambda_g}{2\pi}\sin^{-1}(\omega C Z_0) \tag{4.10.4}$$

Thus a cascade of low-impedance and high-impedance lines can realize (approximately) an LC ladder network.

It is tempting to take filters synthesized in lumped-element form and realize them in the above manner with transmission lines. The series element realization concept presented in this section could be extended using shorted and open stubs to better realize shunt elements. The problem with this approach is that the resulting filters are narrowband and the response outside the desired operating range is unpredictable. It is far better to employ the Richards's transformation, considered in Section 2.12, which is a much broader bandwidth technique for realizing filters in distributed form.

This page titled 4.10: Cascaded Line Realization of Filters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.10: Cascaded Line Realization of Filters** has no license indicated.

4.11: Butterworth and Chebyshev Bandpass Filters

Designs of Butterworth and Chebyshev filters with center frequencies of 1 GHz and 3 dB bandwidths of 10% are shown in Figure 2.10.4. Their transmission, S_{21}, and reflection, S_{11}, responses are shown in Figure $4.11.1$. The filter skirts, i.e. the transitions from the passband to the stopbands, is steeper for the Chebyshev filter than for the Butterworth filter as expected. The ripple of the Chebyshev filter is very small and not seen in this plot. In a real filter there will be loss and low-level ripples in the magnitude response

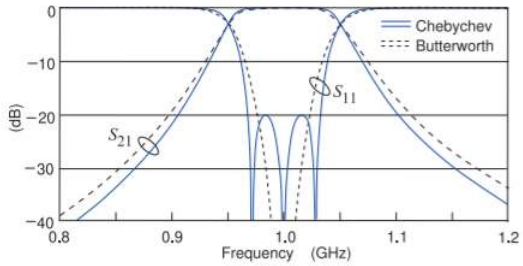


Figure $4.11.1$: Insertion loss (S_{21}) and return loss (S_{11}) of the Butterworth and Chebychev lumped-element bandpass filters in a $50\ \Omega$ system.

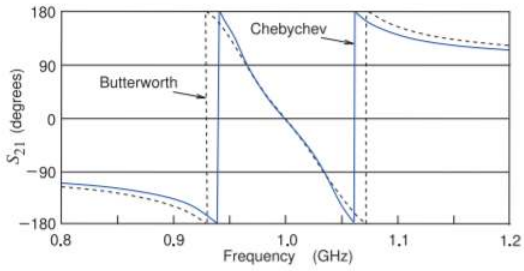


Figure $4.11.2$: Phase of the transmission response (S_{21}) of the Butterworth and Chebyshev lumped-element filters. The discontinuities seen in the phase from -180° to $+180^\circ$ are artifacts and the phases of the filters reduce monotonically with increasing frequency. (For example, the phase at 0.8 GHz is $-110^\circ + 360^\circ = 250^\circ$.)

disappear. However even with loss the steep skirts of the Chebyshev transmission response remain. Also, even with loss, the impact of the Chebyshev ripples is clearly seen in the reflection (S_{11}) response. In Figure $4.11.1$ three distinct S_{11} zeros are seen and these correspond to the three poles of the Chebyshev filter's S_{21} response, but of course we cannot see these. (In the Laplace transfer function there are 3 complex poles each pair being transformed from one of the three poles of the lowpass prototype.) The Butterworth filter also has three (complex) S_{11} zeros and these are all at the center frequency of the bandpass filter, 1 GHz.

Another characteristic that differs between the Chebyshev and Butterworth responses is seen in their phase responses plotted in Figure $4.11.2$. Each pole in the S_{21} characteristic causes a 90° phase change. The three complex poles (i.e. six actual poles) of S_{21} then result in six 90° phase changes in S_{21} for a total phase change of 450°. Small ripples are seen in the Chebyshev phase responses in the pass band while the phase changes for the Butterworth filter are smooth. (The Chebyshev phase ripples remain even with low-level loss.)

The magnitude and phase responses, Figures $4.11.1$ and $4.11.2$, do not provide complete visualization of the filter characteristics. Additional insight is provided in the S_{11} loci on the Smith chart, see Figure $4.11.3$. Figure $4.11.3$ shows the S_{11} characteristics plotted on Smith charts. (In the passband the S_{21} locus would be very close to the unit circle and little of value is observed in the passband.) There is a wealth of information here. First consider the response for the Chebyshev filter, Figure $4.11.3$(a). As frequency increases from 0.8 GHz, the locus of S_{11} is first close to the unit circle and then approaches the origin

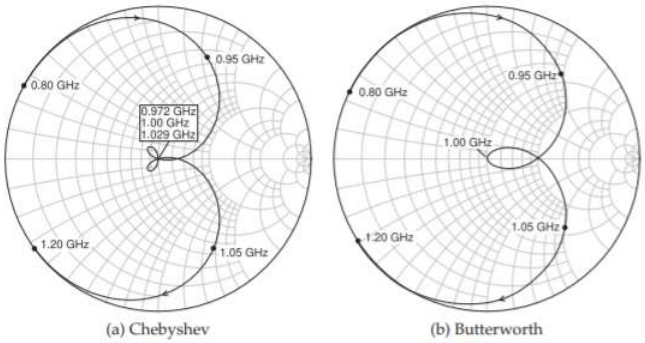


Figure 4.11.3: Smith chart plot of S_{11} of the Butterworth and Chebychev lumped-element filters.

of the polar plot as the frequency approaches the passband of the filter. A special characteristic of the Chebyshev response is the looping which here results in three passes of the locus through the origin. These are the three zeros. Eventually the locus of S_{11} increases as frequency increases above the passband. A loop is also seen in the Butterworth response, Figure 4.11.3(b). This loop goes through the origin and while it seems that there is just one zero there are actually three. What is happening is that as frequency increases and as the locus of the Butterworth S_{11} response approaches the origin, the movement of the locus with respect to frequency slows down. The best way to be convinced that there are three zeros of S_{11} is to look at the transmission phase $(\angle S_{21})$ response in Figure 4.11.1.

Comprehensive visualization of the filter response requires the rectangular plots of the magnitudes of S_{21} and S_{11}, Figure 4.11.1, of S_{21} phase, Figure 4.11.2, and the Smith chart plot of S_{11}, Figure 4.11.3. The phase plot convinces you of the number of zeros in your design which is important in interpreting the Butterworth results. A physical implementation of the design will not be exact and so tuning is required. Then the most important characterizations are the rectangular magnitude and Smith chart responses. Here the loops on the Smith chart, even if they do not go through the origin exactly, distinguish the Butterworth and Chebyshev responses. Matching may be required to shift the loops to the origin of the polar plot.

This page titled 4.11: Butterworth and Chebyshev Bandpass Filters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.11: Butterworth and Chebyshev Bandpass Filters** has no license indicated.

4.12: Richard's Transformation

Richards's transformation is a remarkable scheme that takes into account the actual properties of transmission lines, yielding broadband transmission line-based implementations of lumped-element filter prototypes [12, 13, 14, 15].

4.12.1: Richard's Transformation and Transmission Lines

Consider a section of transmission line of electrical length θ with $ABCD$ parameters

$$T = \begin{bmatrix} \cos(\theta) & \jmath Z_0 \sin(\theta) \\ \jmath/Z_0 \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{4.12.1}$$

If this line is terminated in a load, Z_L, then its input impedance is

$$Z_{\text{in}}(\theta) = \frac{\cos(\theta) Z_L + \jmath Z_0 \sin(\theta)}{\jmath/Z_0 \sin(\theta) Z_L + \cos(\theta)} \tag{4.12.2}$$

Now examine two extreme conditions. As the load impedance increases, eventually becoming an open circuit, the input impedance of a line with electrical length θ is defined in terms of a cotangent of the electrical length:

$$Z_L \to \infty \quad \Rightarrow Z_{\text{in}}(\theta) = \frac{Z_0}{\jmath} \cot(\theta) \tag{4.12.3}$$

$$Y_{\text{in}}(\theta) = \jmath Y_0 \tan(\theta) \tag{4.12.4}$$

As the load impedance reduces to become a short circuit, the input impedance of a line with electrical length θ is defined in terms of a tangent of the electrical length:

$$Z_L \to 0 \quad \Rightarrow \quad Z_{\text{in}}(\theta) = \jmath Z_0 \tan(\theta) \tag{4.12.5}$$

These results lead to the Richards's transformation, which replaces the Laplace variable, s, by Richards's variable, S, where $S = \jmath\alpha \tan(\theta)$. This transformation is written

$$s \to S = \jmath\alpha \tan(\theta) \tag{4.12.6}$$

For now α and θ are constants that can be chosen as design variables. θ, of course, is the electrical length of the line. Also, α must have the units of impedance and it is the characteristic impedance of the transmission line.

Applying the Richards's transformation to a capacitor, the admittance of the element is transformed as follows:

$$y = sC \to Y = SC = \jmath\alpha C \tan(\theta) \tag{4.12.7}$$

so that the capacitor is transformed into an open-circuited stub with characteristic admittance

$$Y_0 = \alpha C \tag{4.12.8}$$

If a lumped-element capacitor with admittance $y = sC$ is to be realized using a transmission line, the admittance $Y = SC = \jmath\alpha C \tan(\theta)$ is instead realized. There are two parameters to select to realize this admittance. The first, α, is the characteristic admittance of the transmission line (and for any given transmission line topology there is a minimum and maximum characteristic admittance or impedance that can be realized), and θ is the electrical length of the line.

Applying the transformation to an inductor, the impedance of the element is transformed as follows:

$$Z = sL \to Z = SL = \jmath\alpha L \tan(\theta) \tag{4.12.9}$$

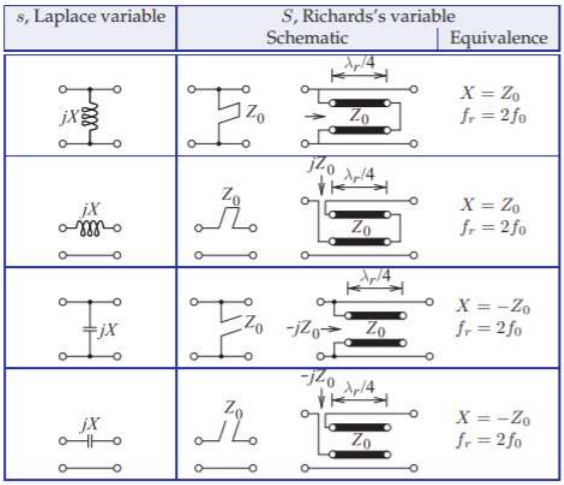


Figure 4.12.1: Equivalences resulting from Richards's transformation. With $f_r = 2f_0$ the transmission line stubs are one-eighth wavelength long at f_0.

so that the inductor is transformed into a short-circuited stub with characteristic impedance

$$Z_0 = \alpha L \tag{4.12.10}$$

Thus the Richards's transform converts an inductor into a short-circuited stub and a capacitor into an open-circuited stub.

4.12.2: Richard's Transformation and Stubs

There is a duality between stubs and inductors and capacitors; they are coupled by Richards's transformation. One of the important quantities used in the transformation is the commensurate frequency, f_r, which most often is chosen as twice the operating frequency, f_0. Considering stubs that are one-quarter wavelength long at f_r, the duality is as shown in Figure 4.12.1.

4.12.3: Richard's Transformation Applied to a Lowpass Filter

In this section, Richards's transformation is used to realize a lumped-element filter in distributed form. The design example begins by considering a Chebyshev lowpass filter with the transmission characteristic

$$|T(s)|^2 = \frac{1}{1+\varepsilon^2|K(s)|^2} \tag{4.12.11}$$

With the Richards's transformation $(s \to \jmath\omega \to \jmath\alpha\tan(\theta))$ this becomes

$$|T(\jmath\alpha\tan\theta)|^2 = \frac{1}{1+\varepsilon^2|K(\jmath\alpha\tan\theta))|^2} \tag{4.12.12}$$

Thus the passband edge at $\omega = 1$ is mapped to $\omega = \theta_1$ as

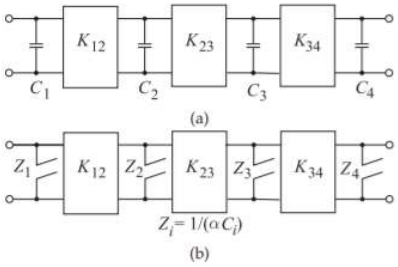
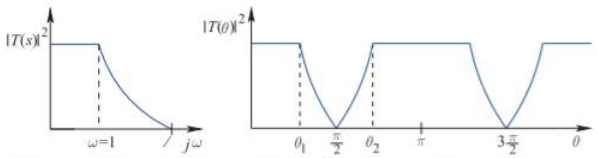
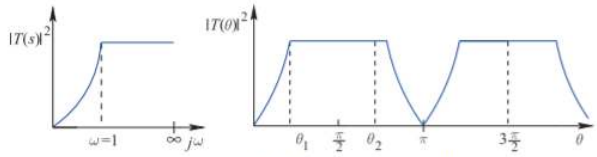


Figure 4.12.2: Lowpass prototypes: (a) lowpass prototype as a ladder filter with inverters; and (b) lowpass distributed prototype with open-circuited stubs (the impedance looking into the stub is indicated).



(a) Lowpass filter response (b) Distributed lowpass filter response

Figure 4.12.3: Lowpass to distributed lowpass transformation. $s = j\omega$.



(a) Highpass filter response (b) Distributed highpass filter response

Figure 4.12.4: Highpass to distributed highpass transformation. $s = j\omega$.

$$\omega = 1 \rightarrow \alpha \tan(\theta_1) \tag{4.12.13}$$

so that

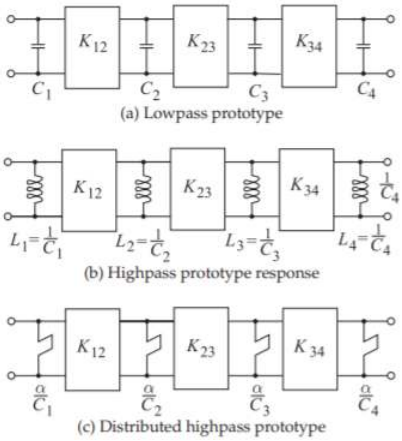
$$\alpha = \frac{1}{\tan(\theta_1)} \tag{4.12.14}$$

Recalling that a capacitor is transformed into an open-circuited stub (see Equation (4.12.8)), Richards's transformation applied to the lowpass filter prototype results in a filter with transmission line elements only, as shown Figure 4.12.2, provided that the inverters are realized using transmission lines.

The implementation of a lowpass filter in distributed form results in passbands and stopbands that repeat in frequency, as shown in Figure 4.12.3. This occurs because transmission lines are used to realize lumped elements and the two-port parameters of a transmission line repeat every wavelength (or one-half wavelength in some cases). For example, the input impedance of a stub is the same whether it is one-half wavelength long or one wavelength long.

4.12.4: Richard's Transformation Applied to a Highpass Filter

With reference to Figure 4.12.4(a), the passband edge at $\omega = 1$ is mapped to θ_1, as shown in Figure 4.12.4(b). This means that the passband at $\omega = 1$ is mapped



(a) Lowpass prototype

(b) Highpass prototype response

(c) Distributed highpass prototype

Figure 4.12.5: Ladder prototype transformation. (The input impedances of the stubs are indicated in (c).)

to θ_1 as (note that θ_1 is the electrical length at the band edge)

$$\omega = 1 \rightarrow \alpha \tan(\theta_1) \tag{4.12.15}$$

so that

$$\alpha = \frac{1}{\tan(\theta_1)} \tag{4.12.16}$$

The sequence of steps transforming a lumped lowpass prototype to its distributed highpass prototype are shown in Figure 4.12.5. As previously discussed, all the inverters can be approximated by transmission lines of length $\pi/2$ at the corner frequency of the filter.

This page titled 4.12: Richard's Transformation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.12: Richards's Transformation** has no license indicated.

4.13: Kuroda's and Norton's Network Identities

Kuroda's and Norton's network identities are a number of pairs of equivalent networks that facilitate the transformation from one prototype to another, particularly transformations that enable realizable transmission line implementations.

4.13.1: Kuroda's Identities

Kuroda's identities embody a number of specific manipulations using impedance or admittance inverters. They are particularly useful in implementing Richards's transformations as they physically separate transmission line stubs, transform series stubs into shunt stubs, and can change characteristic impedances that are either too small or too high to practically realizable values. Kuroda's identities are a number of equivalent two-port networks, as shown in Figure 4.13.1. The proof is derived by obtaining the $ABCD$ parameters of the two ports similar to the technique used throughout this chapter. The identities shown in Figure 4.13.1 are narrowband. With stubs replacing the lumped elements, Kuroda's identities then have broader bandwidths. Kuroda's identities with stubs are shown in Figure 4.13.2. The major use of these identities is to transform designs with series stubs (in addition to possible shunt stubs) into designs with shunt

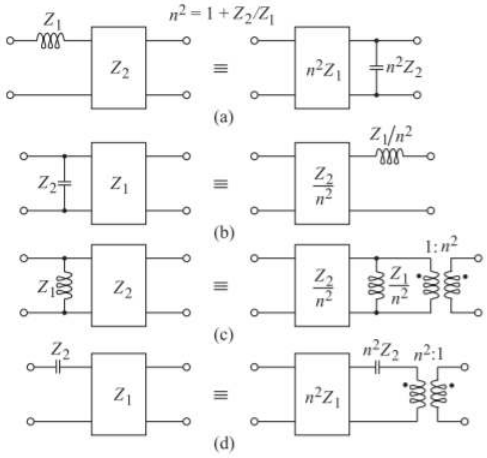


Figure 4.13.1: Kuroda's identities. Here the inverters are impedance inverters and the designation refers to the impedance of the inverter. Recall that an inverter of impedance Z_1 can be realized by a one-quarter wavelength long transmission line of characteristic impedance Z_1. (As usual, element impedances are indicated.)

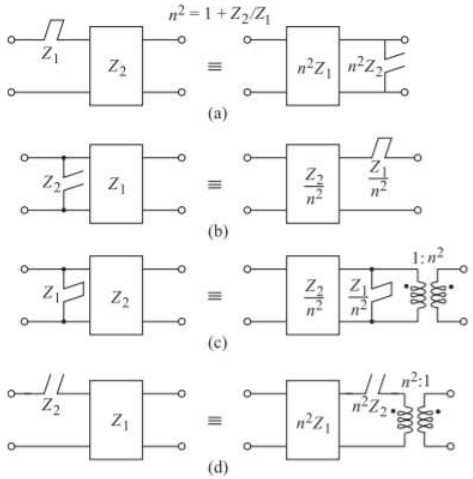


Figure 4.13.2: The stub form of Kuroda's identities with impedance inverters. The stub impedances shown are the input impedances of the stubs.

stubs only.

To see how these identities are used, consider the identity shown in Figure 4.13.1(a). The network on the left has a series inductor that, using transmission lines, is realized by a series stub. A series stub cannot be realized in most transmission line technologies, including microstrip. Using the identity shown on the right in Figure 4.13.1(a), the series stub is replaced by the shunt stub used to realize the shunt capacitor. At the same time, impedance scaling

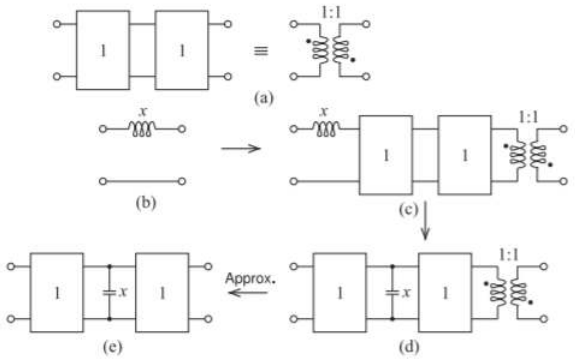


Figure 4.13.3: Transformation of a series inductor into a shunt capacitor between inverters.

can be used. If the impedance of the inverter in the network on the left is too low, then it can be scaled by a factor n^2, where $n^2 = 1 + Z_2/Z_1$.

One use of Kuroda's transforms is to convert a series inductor into a shunt capacitor. Consider the inductor transformation shown in Figure 4.13.3. Figure 4.13.3(a) shows that two inverters in cascade is electrically equivalent to an inverting transformer. Since each inverter is a unity inverter (either a $1\ \Omega$ impedance inverter or a $1\ \mathrm{S}$ admittance inverter), the equivalence is a unitary inverting transformer that corresponds to a 180° phase change. So the transformation of a series inductor begins with the transformation of the inductor (with value x) in Figure 4.13.3(b) into the network of Figure 4.13.3(c) in which the inverter cascade and the inverting transformer cancel each other out. Using the Kuroda identity in Figure 4.13.2(a), the network in Figure 4.13.3(c) converts to the electrically identical network in Figure 4.13.3(d). Since the 180° rotation usually does not matter in circuits, the final transformation shown in Figure 4.13.3(e) is usually acceptable. (The rotation would only matter if there was another path between the input and output as then phasing would affect the way signals on multiple paths combined.) Now the capacitor in Figure 4.13.3(e) has the numerical value x, the same as the numerical value of the original inductor. This is a result of using unitary inverters. The key result here is that a series inductor is equivalent to a shunt capacitor flanked by two inverters.

4.13.2: Norton's Identities

Norton's transformations enable the magnitude of lumped element values to be scaled [16, 17]. Norton's transformations are shown in Figure 4.13.4 where K is the scaling factor. Additional elements are introduced in these transformations including possibly negatively valued elements. Generally these elements can be combined with other elements so that the elements to be realized are positive (e.g. positive capacitor) and the transformer can also be replaced through subsequent transformations using, for example, Kuroda's identities.

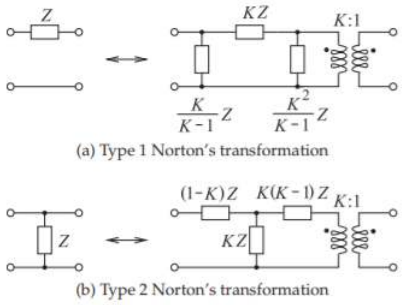


Figure 4.13.4: Norton's transformations.

This page titled 4.13: Kuroda's and Norton's Network Identities is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.13: Kuroda's and Norton's Network Identities** has no license indicated.

4.14: Inter-resonator Coupled Bandpass Filters

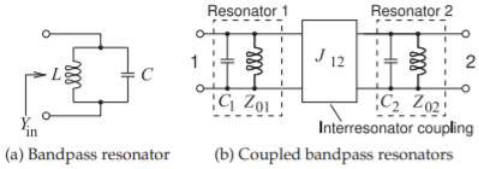


Figure 4.14.1: Bandpass resonators.

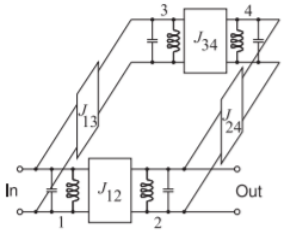


Figure 4.14.2: The electrical design of a cross-coupled filter. There are four resonators labeled 1, 2, 3, and 4. The bandpass resonators are coupled by admittance inverters.

The synthesis method that has been described so far leads to bandpass filters that are a cascade of coupled bandpass resonators. A bandpass filter then comprises pairs of resonators of the type shown in Figure 4.14.1(a) that are coupled using what is called inter-resonator coupling (see Figure 4.14.1(b)). A generalization of this architecture is that a bandpass filter comprises coupled resonators that do not need to be in a linear cascade. Coupling the output of one resonator latter in a cascade to the input of an earlier resonator is called cross coupling, and a filter that uses this arrangement is said to be a crosscoupled resonator filter or a cross-coupled filter. By reusing resonators in this way, fewer resonators are required to achieve a specified bandpass response and the filter can have higher performance or be smaller. An example of the electrical design of a cross-coupled filter is shown in Figure 4.14.2.

Design of a cross-coupled bandpass filter is not as systematic as the design of a ladder filter. The design concept is based on using a number of resonators where each resonator on its own has the same resonant

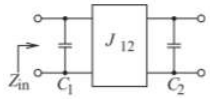


Figure 4.14.3: Lowpass filter approximation with capacitors separated by an admittance inverter.

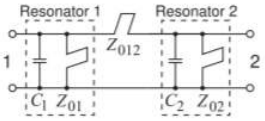


Figure 4.14.4: Segment of a filter with two resonators. $C_1 = C_2 = 2.7171$ pF, $Z_{01} = Z_{02} = 67.7535\ \Omega$, and $Z_{012} = 442.3836\ \Omega$. The stubs are resonant at $f_r = 2f_0$ where the filter center frequency $f_0 = 1$ GHz. Thus the stubs are $\lambda/8$ long at f_0.

frequency f_0, where f_0 is the center frequency of the filter. The desired filter characteristics are obtained by adjusting the inter-resonator coupling, that is, the coupling between pairs of resonators. Design is not a synthesis process based on a series of mathematical transformations, but instead is based on the understanding of the impact of changing the coupling. The design process will now be described for a pair of resonators.

4.14.1: Coupling of a Pair of Resonators

A segment of a bandpass filter fundamentally comprises a pair of coupled bandpass resonators. This is based on the lowpass prototype with a pair of capacitors separated by an admittance inverter (see Figure 4.14.3). In transforming this to a bandpass filter with a center frequency of f_0, each capacitor is replaced by a resonator such as that shown in Figure 4.14.1(a). This leads to a pair of

resonators, each resonant at f_0, that are coupled by an inverter (see Figure 4.14.1(b)). The value of the inverter determines the level of coupling of the resonators.

Consider the circuit in Figure 4.14.4, which consists of a pair of coupled resonators, with each resonator resonant at $f_0 = 1$ GHz, and the resonators are coupled by the series stub with characteristic impedance Z_{012} (i.e., the Z_{012} stub). The circuit of Figure 4.14.4 is a segment of a Chebyshev 1 GHz bandpass filter with a fractional bandwidth of 10%. This is an interim stage of a bandpass filter design[1] and the system impedance of this circuit segment is 139.4 Ω. The magnitude and phase of the $139.4 - \Omega$ S_{21} and S_{11} parameters of this circuit are shown in Figure 4.14.5(a), and the phase of S_{21} is shown in Figure 4.14.5(b). Specific features to identify here are the two peaks in the phase response at f_1 and f_2. These are also the frequencies of peak transmission response where $|S_{21}| = 1$. The bandwidth defined by f_1 and f_2 is called the coupling bandwidth $(= f_2 - f_1)$ or the bandwidth of the filter (note that this is not the 3-dB bandwidth of the filter). The level of coupling determines the coupling bandwidth. That is, changing the level of coupling provided by the Z_{012} stub changes the position of f_1 and f_2. An interpretation of what is happening is that by coupling the two resonators, the resonate frequency of the resonators on their own, f_0, has been split by the coupling, resulting in two peaks at f_1 and f_2. The higher the coupling, the larger the coupling bandwidth, and the lower the coupling, the smaller the coupling bandwidth. For example, a narrowband bandpass filter comprises

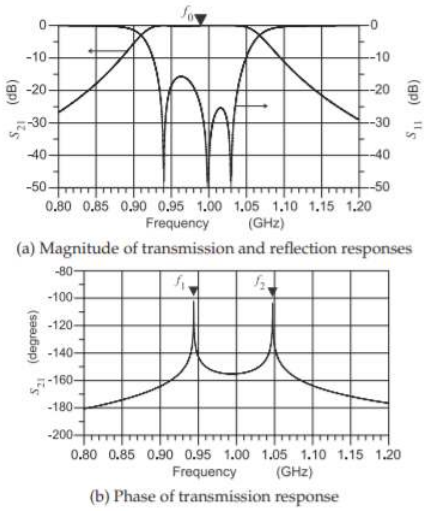


Figure 4.14.5: Response of the coupled resonator network in Figure 4.14.4.

resonators that have low-level coupling.

So following the choice of an appropriate architecture, that is, the number of resonators and the coupling arrangement, design and manual optimization becomes one of changing the level of coupling between pairs of resonators. Design is not as clean as for ladder filters, but the advantages of a design with fewer resonators is a compelling justification. Invariably, extensive (perhaps a day's worth) manual adjustment of the coupling of the final fabricated filter is required. Overall this is an approach worth adopting for high-value filters such as those used in a basestation, where the high unit costs resulting from manual tuning of individual filters can be justified. This type of filter design is considered further in [3, 18, 19, 20, 21].

4.14.2: Relationship of Ladder Synthesis and Inter-resonator Coupling

The relationship of design focused on interresonator coupling and design based on the ladder synthesis approach can be demonstrated for a pair of resonators as follows. Consider the lowpass prototype circuit of Figure 4.14.3. The input impedance of this circuit is

$$Z_{\text{in}}(s) = [sC_1 + J_{12}^2/(sC_2)]^{-1} \tag{4.14.1}$$

With $s = \jmath\omega$, the poles of Z_{in} occur when

$$J_{12}^2 - \omega^2 C_1 C_2 = 0 \tag{4.14.2}$$

thus the radian frequencies of the poles are

$$\omega = \pm \frac{J_{12}}{\sqrt{C_1 C_2}} \tag{4.14.3}$$

Applying the bandpass transformation (so that the capacitors are replaced by the resonators of Figure 4.14.1(a)) results in the bandpass circuit of Figure 4.14.1(b). With this transformation the frequency is transformed as

$$\omega \to \alpha \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) \tag{4.14.4}$$

That is, the transmission coefficient of the bandpass filter is

$$T_{\text{BPF}}(\omega) = T_{\text{LPF}} \left(\alpha \left[\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right] \right) \tag{4.14.5}$$

where $T_{\text{LPF}}(\omega)$ is the transmission coefficient of the lowpass filter at the radian frequency ω. The poles of the bandpass filter are at ω_1, where

$$\alpha \left(\frac{\omega_1}{\omega_0} - \frac{\omega_0}{\omega_1} \right) = + \frac{J_{12}}{\sqrt{C_1 C_2}} \tag{4.14.6}$$

and ω_2, where

$$\alpha \left(\frac{\omega_2}{\omega_0} - \frac{\omega_0}{\omega_2} \right) = - \frac{J_{12}}{\sqrt{C_1 C_2}} \tag{4.14.7}$$

Then the coupling bandwidth of the filter is

$$f_1 - f_2 = \frac{\omega_1 - \omega_2}{2\pi} = \frac{J_{12}\omega_0}{2\pi\alpha\sqrt{C_1 C_2}} \tag{4.14.8}$$

Thus the coupling bandwidth can be directly related to the steps in the ladder-based synthesis of a bandpass ladder filter.

4.14.3: Group Delay

Group delay of a network, and in particular of a filter, is the delay to send information through a network. It is the time required for a modulated carrier signal to appear at the output of a filter after being applied to the input of the filter. Phase delay is a similar measure of delay but does not describe the time it takes to send information. It describes the delay to send a particular phase of a single sinewave. Group delay is a steady-sate concept and so only approximately captures the transient response of a filter.

If a two-port has the transmission coefficient

$$T(s) = S_{21} = a + jb = t\angle\varphi \tag{4.14.9}$$

where a and b are its real and imaginary parts, t is its magnitude, and φ is its phase. With $s = j\omega$, the phase of T is

$$\varphi(\omega) = \tan^{-1}\left(\frac{b}{a}\right) \tag{4.14.10}$$

The group delay, τ_D, is the negative of the derivative of this phase as follows:

$$\text{Group delay } = \tau_g(\omega) = -\frac{d\varphi}{d\omega} = -\frac{d}{d\omega}\left[\tan^{-1}\left(\frac{b}{a}\right)\right] \tag{4.14.11}$$

This compares to the phase delay:

$$\text{Phase delay } = \tau_\varphi(\omega) = -\frac{\varphi}{\omega} = -\tan^{-1}\left(\frac{b}{a}\right) \tag{4.14.12}$$

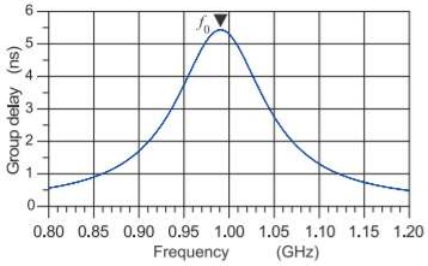


Figure 4.14.6: Group delay of a resonator.

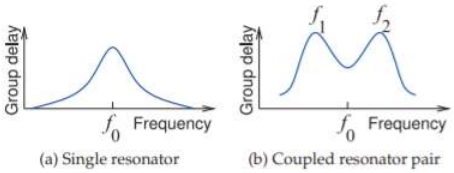


Figure 4.14.7: Group delay of resonators in a filter. Additional delay is introduced by the coupling inverter.

Figure 4.14.6(c) shows the group delay response of one of the resonators in Figure 4.14.4, where the group delay peaks at the resonant frequency of the resonator f_0. The peak of the group delay occurs at the resonant frequency of the resonator without loading. From examination of the phase of S_{21} (see Figure 4.14.5(b)), it is clear that the group delay peaks very close to f_1 and f_2. So the coupled resonator pair has two peaks in the group delay. This can be seen in Figure 4.14.7, which plots the group delay of a single resonator and a coupled pair of resonators.

4.14.4: Footnotes

[1] Specifically these resonators are the left-most resonators in a design developed in Chapter 3 (see Figure 3.4.16).

This page titled 4.14: Inter-resonator Coupled Bandpass Filters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.14: Inter-resonator Coupled Bandpass Filters** has no license indicated.

4.15: Bandpass Filter Topologies

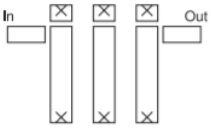


Figure 4.15.1: Microstrip layout of a third-order combline filter based on coupled lines. Each of the vertical microstrip line forms a resonator with the top gap capacitors. The input and output gap capacitors provide impedance matching.

The essential structure of a bandpass filter comprises resonators that are coupled to each other. A large number of filter architectures based on this concept have been deployed. A parallel coupled line (PCL) filter called a combline filter is shown in Figure 4.15.1. Here the microstrip lines form the resonator and the coupling of parallel lines provides the interresonator coupling. A variation on a filter with PCL resonators is shown in Figure 4.15.2 [22]. This filter uses edge-coupled microstrip resonators that are $\lambda/2$ long (at the center frequency). Figure 4.15.3 shows another distributed bandpass filter topology utilizing end-coupled microstrip resonators. That is, the gaps provide the interresonator coupling. All transmission line bandpass filters have spurious passbands [23, 24, 25]. The root cause of the spurious responses derives from the transformation of the parallel LC resonators into their transmission line form.



Figure 4.15.2: Microstrip layout of a parallel coupled bandpass filter.



Figure 4.15.3: General microstrip layout for an end-coupled bandpass filter (series coupling gaps between cascaded straight resonator elements).

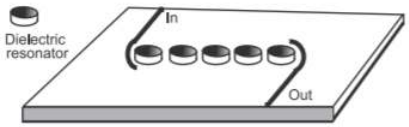


Figure 4.15.4: A microstrip coupled dielectric resonator bandpass filter configuration.

A **dielectric-resonator**-based bandpass filter is shown in Figure 4.15.4 [26]. Here the pucks are the bandpass resonators and they are coupled and the evanescent fields outside the pucks provide the interresonator coupling. The pucks shown are cylinders of high-permittivity material (typically having a relative permittivity of 500–85000) with an approximate magnetic wall at the cylindrical surface of the puck. Thus the puck resonates when its diameter is approximately $\lambda/2$.[1]

With all these filters the interresonator coupling functions as an inverter. Thus the basic functional unit of the bandpass filters is the coupled resonator structure shown in Figure 2.14.1(b).

4.15.1: Footnotes

[1] A better estimate is developed from the zeros of Bessel functions, as the fields inside the pucks have a Bessel function dependence (this is the form of the solution of the wave equation in cylindrical coordinates).

This page titled 4.15: Bandpass Filter Topologies is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.15: Bandpass Filter Topologies** has no license indicated.

4.16: Case Study- Design of a Bandstop Filter

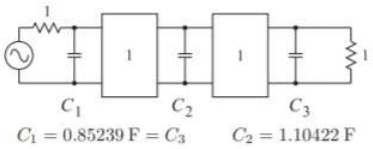


Figure 4.16.1: Bandpass filter approximation.

The design of a bandstop filter begins with the lowpass filter prototype shown in Figure 4.16.1. To the lowpass prototype, the highpass transformation is applied to obtain the highpass prototype of Figure 4.16.2. Picking a center frequency of approximately 1 GHz and corner frequencies of $f_1 = 950$ MHz and $f_2 = 1050$ MHz, corresponding to a bandwidth of approximately 10%, the bandstop transformation is now applied. This results in the prototype of Figure 4.16.3. Finally, scaling the system impedance to 50 Ω leads to the prototype of Figure 4.16.4. The impedance inverters must remain set at 50 Ω in order to obtain a broad match. In the passband, energy will pass at all frequencies.

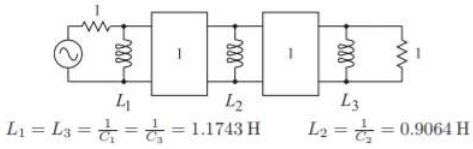


Figure 4.16.2: Bandstop filter prototype.

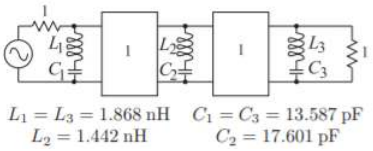


Figure 4.16.3: Bandstop filter following transformation from the lowpass prototype.

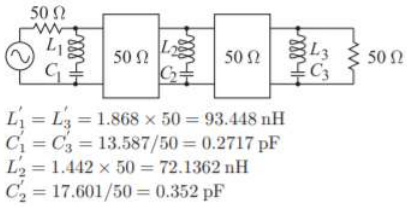


Figure 4.16.4: Bandstop filter after impedance transformation.

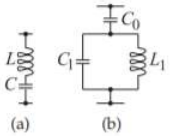


Figure 4.16.5: Transformations of the resonators in the bandstop filter to obtain realizable values. The series LC resonator in (a) is transformed to the form in (b).

In Figure 4.16.4, the inductor values are relatively large and the capacitor values are relatively small so that it will be difficult to realize the filter in either lumped or distributed forms. These values must be scaled to obtain realizable values. One possible transformation is shown in Figure 4.16.5. To establish that the left-hand and right-hand networks are equivalent, at least near one frequency, the impedances and derivatives must be matched. For the circuit in Figure 4.16.5(a),

$$Z_1 = \jmath \frac{\omega^2 LC - 1}{\omega C} \tag{4.16.1}$$

and

$$\frac{dZ_1}{d\omega} = \jmath\frac{\omega^2 LC + 1}{\omega^2 C} \tag{4.16.2}$$

and for the circuit in Figure 4.16.5(b),

$$Z_2 = \jmath\frac{\omega^2 L_1 C_1 - 1 + \omega^2 L_1 C_0}{\omega C_0(1 - \omega^2 L_1 C_1)} \tag{4.16.3}$$

and

$$\frac{dZ_2}{d\omega} = \jmath\frac{\omega^4 L_1^2 C_1^2 - 2\omega^2 L_1 C_1 + \omega^4 L_1^2 C_0 C_1 + \omega^2 L_1 C_0 + 1}{\omega^2 C_0(\omega^2 L_1 C_1 - 1)^2} \tag{4.16.4}$$

Equating the above enables C_0 and C_1 to be found for a chosen value of L_1. Thus the series LC resonators in Figure 4.16.6(a and c) are replaced by the

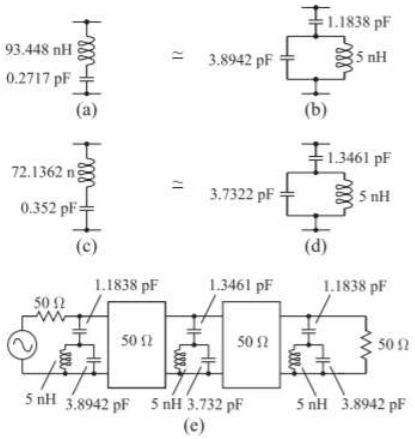


Figure 4.16.6: Intermediate bandstop filter prototype.

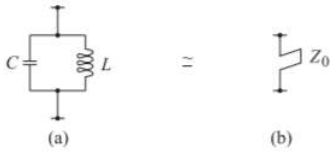


Figure 4.16.7: Equivalence of shunt bandpass resonator to shunt short-circuited stub.

networks in Figure 4.16.6(b and d). This results in the filter of Figure 4.16.6(e).

At this stage the bandpass resonators are then equated to short-circuited stubs by equating the admittance, Y_1, of the lumped circuit in Figure 4.16.7(a) with the admittance, Y_2, of the stub in Figure 4.16.7(b). That is, by equating

$$Y_1 = \jmath\frac{(\omega^2 CL - 1)}{\omega L} \quad \text{and} \quad Y_2 = \frac{1}{\jmath Z_0 \tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right)} \tag{4.16.5}$$

The characteristic impedance of the stub, Z_0, is selected so that the frequency, ω_r, is not too far above the upper band-edge frequency of the filter, in this case 1.05 GHz. Choosing $Z_0 = 20\ \Omega$ results in the stub transformations shown in Figure 4.16.8(a–d). The bandstop filter prototype with stubs is shown in Figure 4.16.8(e). The final physical layout of the bandstop filter is shown in Figure 4.16.9. The response of the final bandstop filter design is shown in Figure 4.16.10.

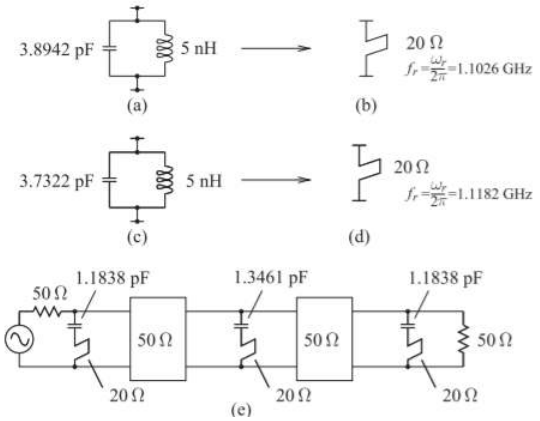


Figure 4.16.8: Bandstop filter prototype using stub approximations. The stub in (b) is the transmission line approximation of the parallel resonant circuit in (a). The stub in (d) approximates the circuit in (c).

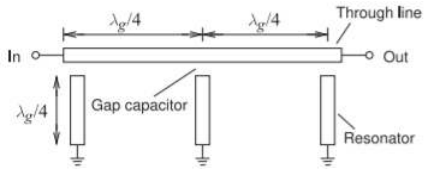


Figure 4.16.9: Physical layout of a bandstop filter in microstrip.

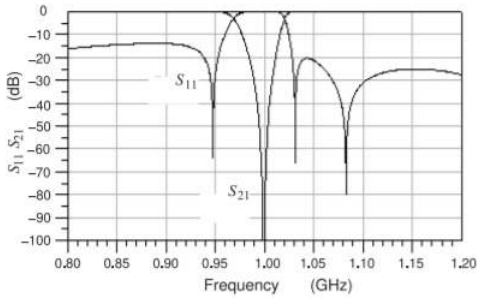


Figure 4.16.10: Response of the bandstop filter shown in Figure 4.16.9.

This page titled 4.16: Case Study- Design of a Bandstop Filter is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.16: Case Study- Design of a Bandstop Filter** has no license indicated.

4.17: Active Filters

With active circuits, losses in lumped-element components can be compensated for by the gain of active devices resulting in filters with high Q. Inverters can also be realized efficiently. The net result is that active filters provide a performance per unit area advantage over passive element realizations. Of course, active filters are limited to low-level signals (such as in the receive circuitry, and prior to the power amplifier stages) as the full power of the signal must be handled by the active devices in the active filter. With active devices, additional noise is introduced and so the noise figure of an active filter must be considered at RF where bandwidths are significant.

At microwave frequencies active filters based on traditional low-frequency concepts have relatively low Q. Higher Q and narrower-band applications require active inductors or distributed techniques. The smallest active filters are the ones that use active inductors and the largest use distributed transmission line elements.

Sometimes the main role of a filter is to limit the dynamic range of signals presented to active circuits, thereby limiting distortion. Clearly active filters are not suitable when large out-of-band signals are a concern.

4.17.1: Radio Frequency Active Filters

There are several problems with lumped-element filters. High-order filters require many poles and zeros that must be precisely placed, requiring small tolerances. With active filters it is possible, to some degree, to achieve higher-order filters with fewer reactive elements and so the tolerancing problem is considerably reduced. Also, there are a few designs that can be tuned electronically. Lumped-element bandpass filters and highpass filters require inductors. Inductors are a major problem, as RF inductors are lossy and also only relatively small values can be obtained. On-chip inductors take considerable area and can thereby dominate the cost of RFICs. With active filters, feedback and capacitors can be used to realize inductor-like characteristics, defined as a positive imaginary impedance increasing with frequency at least over a small frequency range. Many times, and especially at lower radio frequencies, lumped inductors are no longer required. The size of capacitors in filters is another issue, and using feedback and gain, the effective value of capacitors can be magnified. All this comes with limits. The principle limit is that the cutoff frequency of the transistors must be much higher than the operating frequency of the filters. This is a cost incurred whenever feedback is used. A rough rule of thumb is that the transistor cutoff frequency should be about 10 times the operating frequency. So for a 5 GHz bandpass filter, a 50 GHz transistor process is required. The other significant cost is that active filters are only suitable for small signals. The efficiency, a measure of the amount of DC power used in producing the RF output signal, is then not a consideration. Active filters cannot be used where signal levels are likely to be large (such as at the output of a power amplifier or at the receiver input, due to likely out-of-band signals).

The essential aspect of low-RF active filters is the use of feedback, with reactive feedback elements, to realize a frequency-domain transfer function with poles and zeros. The aim is not to synthesize effective capacitors and inductors, but to focus on the overall response. Active filters are generally developed as stages, with each stage realizing a second- or higher-order transfer function. Design is simulation driven, as this is the only way to account for the active device parasitics.

In operational amplifier design, feedback is used to achieve gain stability, but at the cost of reduced bandwidth. If reactive feedback elements are used, high-order transfer functions can be obtained with just a few reactive elements. In Figure $4.17.1$, a twin-T notch network is connected in the operational amplifier feedback path to obtain a bandpass filter [27]. Away from the notch frequency the feedback path has a high impedance and the overall amplifier gain falls off. As the signal frequency approaches the notch frequency, the feedback path becomes effective and the amplifier gain increases. The notch frequency, f_n, is proportional to the RC product of the feedback components and the Q is proportional to the amplifier gain. The governing equations for this filter are

$$f_n = 1/(2\pi RC) \quad \text{and} \quad Q = (1+G)/4 \tag{4.17.1}$$

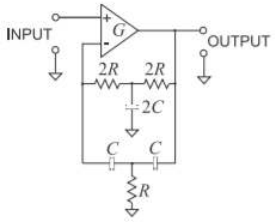


Figure 4.17.1: Active twin-T bandpass filter: center frequency $f_0 = \pi RC/4$ and $Q = (G+1)/4$.

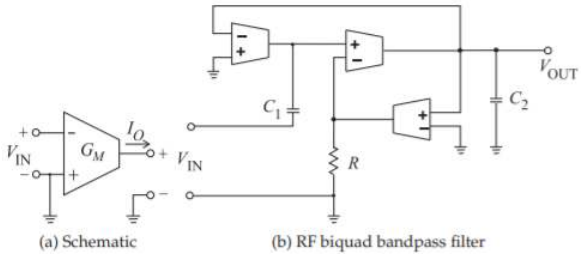


Figure 4.17.2: Operational transconductance amplifier with transconductance G_M. In (a) $I_O = G_M V_{IN}$.

4.17.2: Biquadratic Filters

Biquadratic filters are an important class of filters and are the basic building block used in analog and RF active filters. Biquad filters commonly use operational **transconductance amplifiers** (**OTAs**), the schematic of which is shown in Figure 4.17.2(a). Here $I_O = G_M V_{IN}$, which is also the basic characteristic of an individual transistor. These circuits have characteristics typical of operational amplifiers, including gain stability and resistance to component tolerance variations. OTA-based filters have the advantage of easy tunability, with pole and zero frequencies electronically adjustable. An RF bandpass **biquad** filter is shown in Figure 4.17.2(b). Stages such as this are cascaded to realize high-order filters.

Biquadratic filters implement the transfer function

$$H(s) = \frac{V_{OUT}}{V_{IN}} = \frac{N(s)}{D(s)} = \frac{a_2 s^2 + a_1 s + a_0}{s^2 + b_1 s + b_0} = \frac{a_2 (s+z_1)(s+z_2)}{(s+p_1)(s+p_2)} \tag{4.17.2}$$

This is referred to as a biquadratic function and filters that implement this function are called biquadratic filters or simply biquad filters. There are several special types of biquad filters.

The form of the biquadratic function for a second-order lowpass filter is

$$H_{LP}(s) = \frac{a_0}{s^2 + b_1 s + b_0} = \frac{K\omega_p^2}{s^2 + (\omega_p/Q_p)s + \omega_p^2} \tag{4.17.3}$$

$H_{LP}(s)$ has a double zero at $s = \infty$ and thus negligible response at very high frequencies.

The highpass form of the biquadratic filter is described by

$$H_{HP}(s) = \frac{a_2 s^2}{s^2 + b_1 s + b_0} = \frac{K s^2}{s^2 + (\omega_p/Q_p)s + \omega_p^2} \tag{4.17.4}$$

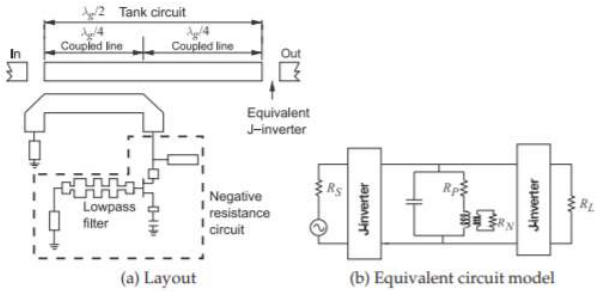


Figure 4.17.3: Active coupled distributed filter. After [30].

The response of $H_{\mathrm{HP}}(s)$ at high frequencies is K and the response at very low frequencies goes to zero.

The bandpass form of the biquadratic filter is described by

$$H_{\mathrm{BP}}(s) = \frac{a_1 s}{s^2 + b_1 s + b_0} = \frac{K(\omega_p/Q_p)s}{s^2 + (\omega_p/Q_p)s + \omega_p^2} \tag{4.17.5}$$

The response of $H_{\mathrm{BP}}(s)$ at high and low frequencies goes to zero, and only at and near the center frequency, $\omega = \omega_p$, is there a reasonable response.

The band-reject or notch form of the biquadratic filter is described by

$$H_{\mathrm{BR}}(s) = \frac{a_2 s^2 + a_0}{s^2 + b_1 s + b_0} = \frac{K(s^2 + \omega_z^2)}{s^2 + (\omega_p/Q_p)s + \omega_p^2} \tag{4.17.6}$$

The response of $H_{\mathrm{BR}}(s)$ at high and low frequencies is high, but there is a double zero at the notch frequency, $\omega = \omega_z$, where the response is very low.

With all of the biquadratic filters, the sharpness of the response is determined by Q_p. The edge frequency or center frequency is ω_p for the bandpass, lowpass, and highpass filters, and the notch frequency is ω_z for the bandstop filter.

4.17.3: Distributed Active Filters

At high microwave frequencies, good results can be obtained by combining distributed elements with a gain stage. Commonly the active devices are used as coupling devices. A negative resistance can be produced using the active devices that can compensate for the losses associated with the transmission line elements or lumped passive elements in the filter. The concept uses a resonant tank circuit formed by a transmission line or lumped-element resonator with a negative resistance coupled into the tank circuit by the active devices [28, 29, 30]. This concept is illustrated in Figure 4.17.3. Here a one-half wavelength long resonator forms a resonant circuit (commonly called a tank circuit in this context). A negative resistance is coupled into the tank circuit to compensate for losses in the resonator and the result is effectively a lossless tank circuit.

An active resonator circuit is shown in Figure 4.17.4. The design of the lossless resonator is based on the negative resistance obtained when the capacitor is connected at the source of the MESFET. This element in the

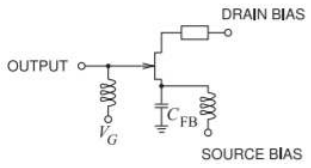


Figure 4.17.4: An active resonator circuit. After [31].

source provides a feedback path between the output of the circuit, the drain-to-ground voltage, and the input gate-to-source voltage. An increase in the drain-source current leads to a voltage at the source that changes the gate-source voltage. This induces a negative resistance that is adjusted through the feedback capacitance, C_{FB}, to compensate for inductor losses. An additional inductor, L_P, is added at the gate. The inductor resonates with the series combination of C_{FB} and C_{GS}. C_{FB} can be implemented using a varactor diode to enable electronic tuning.

This page titled 4.17: Active Filters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.17: Active Filters** has no license indicated.

4.18: Transient Response of a Bandpass Filter

Figure 4.18.1: Lumped-element 5th-order Chebyshev filter.

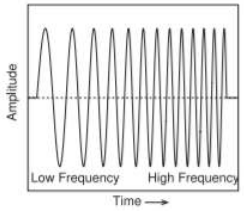


Figure 4.18.2: A linear chirp.

A bandpass filter is normally characterized by its center frequency and bandwidth, and sinusoidal signals that are within the bandwidth are transmitted by the filter and those outside the passband are reflected by the filter. However, if the signal changes frequency quickly, the transient response of the filter can be unexpected [32, 33].

Figure 4.18.1 is a fifth-order filter with each pair of shunt resonators coupled by a series resonator. In response to an RF pulse (a pulse of sinusoidally varying RF), the filter passes energy in-band when all of the resonators have reached steady state. When the RF pulse is removed, the resonators will loose energy and eventually the RF signals on the resonators disappear. There will be a finite time to “charge” and “discharge” the filter’s resonators.

A classic test of the RF response over frequency is to use a linear RF chirp, as shown in Figure 4.18.2. With a linear **chirp**, the frequency of the signal in the pulse changes linearly and smoothly over time from one frequency to another. Figure 4.18.3(a) presents the transient response at the output of

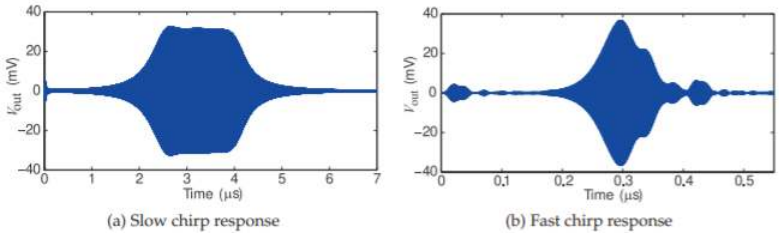


Figure 4.18.3: Output transient response of a 3rd-order Chebyshev filter with a center frequency of 1 GHz and a 30 MHz bandwidth excited by a -20 dBm linear chirp from 950 MHz to 1050 MHz: (a) chirp rate $= 20$ MHz/μs, (b) chirp rate $= 400$ MHz/μs. At the slower chirp rate, the filter response is approximately a superposition of steady-state responses as the frequency changes (i.e., the response is quasi-stationary). At the higher chirp rate, the filter response can no longer be approximated as a sum of steady-state responses (i.e., it is no longer quasi-stationary). After [34].

a 30 MHz-bandwidth 1 GHz filter excited by a 1 μs-long linear chirp from 950 MHz to 1050 MHz with a chirp rate of 20 MHz/μs. This is a relatively slow chirp. The chirp starts at 0 μs and stops at 5 μs. Thus the frequency of the chirp is in-band at 1.75 μs and is out-of-band at 3.25 μs. The chirp is sufficiently slow that the transient transmission response corresponds to the frequency response of the filter. However, if the chirp is fast, the resonators may not reach steady state, and then the transient response of the filter cannot be simply extrapolated from its frequency-domain response. This is seen in Figure 4.18.3(b), which is the response of the filter to a fast chirp with a chirp rate of 400 MHz/μs and a chirp duration of 250 ns, and beginning at 950 MHz and ending at 1050 MHz. The RF chirp begins out of band at 0 ns, becomes in-band at 0.0875 μs, and goes out-of-band again at 0.1625 μs before turning off at 0.25 μs. Interestingly, there is an almost immediate output response, around 0.03 μs when the RF chirp is out of band. Overall the chirped-filter response does not follow the frequency response of the filter.

Another view of the same phenomenon is examined for a 7th-order Chebyshev filter with a center frequency of 900 GHz and 34 MHz bandwidth from 883 MHz to 917 MHz. The frequency-domain transmission response of the filter is shown in Figure 4.18.4. The transmission responses to 0.25 μs-long RF pulses of different frequencies are shown in Figure 4.18.5. The frequency of the RF pulse is shown in each subplot. The top-left plot is the response when the RF pulse is in-band. Here the RF pulse is applied at 0 μs

and removed at $0.25\ \mu s$. An initial delay in the transmission response and what is referred to as a **long-tail response** is seen when the RF pulse is removed. The other plots in Figure $4.18.5$ show the filter response as the RF is backed off from the passband. So, even when the RF is out of band, there can be an appreciable output from the filter at the beginning and end of the RF pulse.

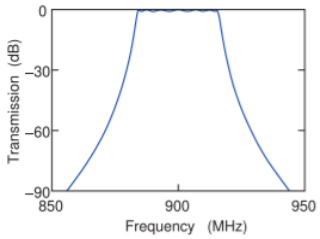


Figure $4.18.4$: Frequency-domain transmission response of a 7th-order Chebyshev filter with a center frequency of 900 GHz and 34 MHz bandwidth from 883 MHz to 917 MHz.

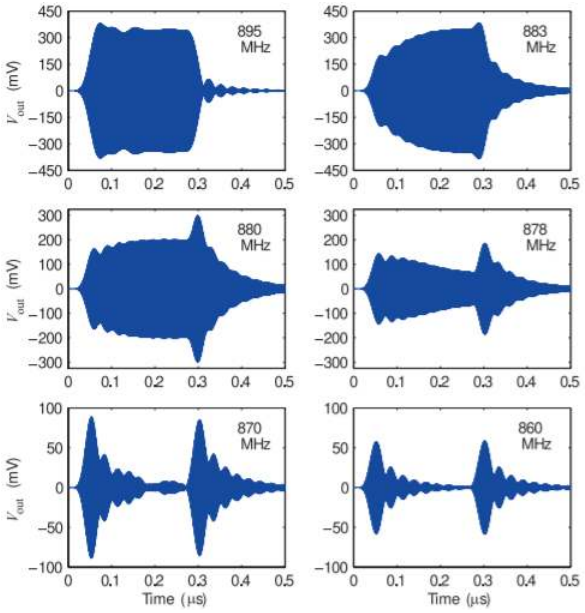


Figure $4.18.5$: Transient response at the output of the 900 MHz 7th-order Chebyshev bandpass filter with a passband from 883 MHz to 917 MHz. The frequency response is shown in Figure $4.18.4$. After [34].

This page titled 4.18: Transient Response of a Bandpass Filter is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.18: Transient Response of a Bandpass Filter** has no license indicated.

4.19: Summary

Radio frequency and microwave filter design combines the mathematical synthesis of a circuit with the required performance and the intuitive realization that particular structures inherently have a desired frequency selectivity. The art of filter design is to direct the mathematical synthesis and companion circuit so that the circuit structures in the synthesized filter prototype match the functionality of physical microwave structures. A very important step in this process is the development of circuit equivalent models of physical structures. Once a microwave filter with appropriate topology has been designed and the close-to-final physical filter is designed, the physical layout is optimized in a circuit simulator to account for parasitics and higher-order EM effects.

A similar design procedure applies to active filter designs where particular circuit topologies, for example, the biquad network blocks, are identified and exploited. The synthesis approach provides design insight and exploitation all parameters of a transfer function, leading to optimum network topologies. Synthesis can be time consuming and specialized, but it is the only way to develop filters with optimum performance. Insight gained during design identifies yield issues and provides the basis for inventing new filter topologies.

4.19.1: Further Reading

The design of microwave filters has a rich tradition. Many excellent books and articles have been written about microwave filter design techniques. Books and articles with extensive treatments are references [1, 3, 4, 14, 16, 17, 18, 19, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46]. However the most important systematic approaches to microwave filter design were presented in this chapter. Many journal and conference papers present topologies in various technologies that can be used to implement filters. The best way to locate these papers is to search using the specifics of the technology of interest. For example “wideband microstrip bandstop filters on glass substrates” will yield a list of papers on the topic. There are many small companies that specialize in different types of filter design. For very high performance filters, for example in basestation applications, the number of designers and vendors is quite small and filter designers follow the literature and patents closely. For designers building microwave systems with small to medium volumes the best approach is to use microstrip design. For bandpass microstrip design the most common choice is to use filter design based on coupled microstrip design which is considered in the next chapter.

This page titled 4.19: Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.19: Summary** has no license indicated.

4.20: References

[1] G. Matthaei, L. Young, and E. Jones, *Microwave Filters, Impedance-Matching Networks and Coupling Structures.* McGraw-Hill, 1965, reprinted in 1980, Artech House.

[2] D. M. Pozar, *Microwave engineering*, 4th ed. John Wiley & Sons, 2012.

[3] I. Hunter, *Theory and Design of Microwave Filters.* IEE Press, 2001.

[4] J. Scanlan and R. Levy, *Circuit Theory.* Oliver & Boyd, 1973, vol. 2.

[5] ——, *Circuit Theory*. Oliver & Boyd, 1973, vol. 1.

[6] V. Belevitch, "Techbyshev filters and amplifier networks," *Wireless Engineer*, pp. 106– 110, Apr. 1952.

[7] H. Orchard, "Formulae for ladder filters," *Wireless Engineer*, pp. 3–5, Jan. 1953.

[8] R. Daniels, *Approximation Methods for Electronic Filter Design*. McGraw-Hill, 1974.

[9] M. Lutovac, D. Miroslav, D. Tosic, V. Dejan, and B. Evans, *Filter Design for Signal Processing using MATLAB and Mathematica.* Prentice Hall, 2001.

[10] M. Steer, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[11] ——, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

[12] P. Richards, "Resistor-transmission-line circuits," *Proc. of the IRE*, vol. 36, no. 2, pp. 217– 220, Feb. 1948.

[13] ——, "General impedance-function theory," *Quarterly Applied Mathematics*, vol. 6, no. 6, pp. 21–29, 1948.

[14] J. Rhodes, *Theory of Electrical Filters.* John Wiley & Sons, 1976.

[15] H. Baher, *Synthesis of Electrical Networks.* John Wiley & Sons, 1984.

[16] M. E. Van Valkenburg, *Introduction to modern network synthesis.* John Wiley & Sons, 1960.

[17] A. I. Zverev et al., *Handbook of filter synthesis.* John Wiley & Sons, 1967.

[18] R. Cameron, R. Mansour, and C. Kudsia, *Microwave Filters for Communication Systems: Fundamentals, Design and Applications.* John Wiley and Sons, 2007.

[19] R. Cameron, "Advanced coupling matrix synthesis techniques for microwave filters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 51, no. 1, pp. 1–10, Jan. 2003.

[20] I. Hunter, J. Rhodes, and V. Dassonville, "Dual-mode filters with conductor-loaded dielectric resonators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 47, no. 12, pp. 2304–2311, Dec. 1999.

[21] W. Fathelbab and M. Steer, "A reconfigurable bandpass filter for RF/microwave multifunctional systems," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 3, pp. 1111–1116, Mar. 2005.

[22] A. Nakamura, K. Hibino, F. Yamamoto, and S. Kamihashi, "Expert system for microwave filter design," in *1990 IEEE MTT-S Int. Microwave Symp. Dig.*, May 1990, pp. 1183–1186.

[23] W. Fathelbab and M. Steer, "Broadband network design," in *Multifunctional Adaptive Microwave Circuits and Systems*, M. Steer and W. Palmer, Eds., 2008, ch. 8.

[24] ——, "Parallel-coupled line filters with enhanced stopband performance," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 12, pp. 3774–3781, Dec. 2005.

[25] P. Cheong, S.-W. Fok, and K.-W. Tam, "Miniaturized parallel coupled-line bandpass filter with spurious-response suppression," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 5, pp. 1810–1816, May 2005.

[26] I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design.* John Wiley & Sons, 1988.

[27] F. Rosenbaum, R. Gregory, W. Richard, W. Ou, F. Kuhns, and T. Trimble, "An MMIC twin-tee active bandpass filter," in *1993 IEEE MTT-S Int. Microwave Symp. Dig.*, 1993, pp. 361–364.

[28] B. Kapilevich, “Active microwave filters,” *Telecommunication Radio Engineering*, vol. 4, no. 2, pp. 51–58, Feb. 1985.

[29] C. Chang and T. Itoh, “Narrowband planar microwave active filter,” *Electronics Letters*, vol. 25, no. 18, pp. 1228–1229, Aug. 1989.

[30] C.-Y. Chang and T. Itoh, “A varactor-tuned, active microwave band-pass filter,” in *1990 IEEE MTT-S Int. Microwave Symp. Dig.*, May 1990, pp. 499–502.

[31] U. Karacaoglu and I. Robertson, “High selectivity varactor-tuned MMIC bandpass filter using lossless active resonators,” in *IEEE 1994 Microwave and Millimeter-Wave Monolithic Circuits Symp. Dig. of Papers*, May 1994, pp. 237–240.

[32] G. Mazzaro, M. Steer, and K. Gard, “Intermodulation distortion in narrowband amplifier circuits,” *IET Microwaves, Antennas Propagation*, vol. 4, no. 9, pp. 1149–1156, Sep. 2010.

[33] G. Mazzaro, M. Steer, K. Gard, and A. Walker, “Response of RF networks to transient waveforms: Interference in frequency-hopped communications,” *IEEE Trans. on Microwave Theory and Techniques*, vol. 56, no. 12, pp. 2808–2814, Dec. 2008.

[34] G. Mazzaro, “Time-frequency effects in wireless communication systems,” Ph.D. dissertation, North Carolina State University, 2009.

[35] J. Malherbe, *Microwave Transmission Line Filters.* Artech House, 1979.

[36] J.-S. Hong and M. Lancaster, *Microstrip Filters for RF/Microwave Applications.* John Wiley & Sons, 2001.

[37] H. Howe, *Stripline Circuit Design.* Artech House, 1974.

[38] R. E. Collin, *Foundations for Microwave Engineering.* John Wiley & Sons, 2007.

[39] D. G. Swanson and W. J. Hoefer, *Microwave circuit modeling using electromagnetic field simulation.* Artech House, 2003.

[40] W.-K. Chen, *Theory and design of broadband matching networks.* Pergamon Press, 1976.

[41] ——, *Passive and Active Filters: Theory and Implementations.* John Wiley & Sons, 1986.

[42] H. J. Blinchikoff and A. I. Zverev, *Filtering in the time and frequency domains.* Krieger Publishing Co., 1986.

[43] L. Zhu, S. Sun, and R. Li, *Microwave Bandpass Filters for Wideband Communications.* John Wiley & Sons, 2011.

[44] J.-S. Hong, *Microstrip Filters for RF/Microwave Applications*, 2nd ed. John Wiley & Sons, 2011.

[45] N. Kinayman and M. Aksun, *Modern microwave circuits.* Artech House, 2005.

[46] R. M. Foster, “Academic and theoretical aspects of circuit theory,” *Proc. IRE*, vol. 50, no. 5, pp. 866–871, May 1962.

This page titled 4.20: References is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.20: References** has no license indicated.

4.21: Exercises

1. The characteristic function of a doubly terminated network is $K(s) = s^2$.
   a. What is the magnitude-squared transmission coefficient ($|T(s)|^2$)?
   b. What is the magnitude-squared reflection coefficient ($|\Gamma(s)|^2$)?
   c. What is transmission coefficient $T(s)$ for a circuit that can be realized using positive R, L, and C elements)?
2. The characteristic function of a doubly terminated network is $K(s) = s^4$. What is the magnitude-squared transmission coefficient ($|\Gamma(s)|^2$)?
3. Consider the design of a fourth-order lowpass Butterworth filter. [This problem follows the development in Section 2.4.]
   a. What is the magnitude-squared characteristic polynomial, $|K(s)|^2$, of the Butterworth filter?
   b. What is the magnitude-squared transmission coefficient (or transfer function)?
   c. What is the magnitude-squared reflection coefficient function?
   d. Derive the reflection coefficient function (i.e., $\Gamma(s)$). Write down the reflection coefficient in factorized form using up to second-order factors.
   e. What are the roots of the numerator polynomial of the reflection coefficient function?
   f. What are the roots of the denominator polynomial of the reflection coefficient function?
   g. Identify the conjugate pole pairs in the factorized reflection coefficient.
   h. Plot the poles and zeros of the reflection coefficient on the complex s plane.
4. Derive the reflection coefficient poles of a second-order Butterworth filter and write out the reflection coefficient with nominator and denominator polynomials, that is not in factorized form. [Parallels Example 2.4.1]
5. Derive the reflection coefficient poles and zeros of a fourth-order Chebyshev filter with a ripple factor, ε, of 0.1.
6. Synthesize the impedance function

$$Z_x = \frac{s^3 + s^2 + 2s + 1}{s^2 + s + 1}$$

   That is, develop the RLC circuit that realizes Z_x. [Parallels Example 2.6.1]
7. Synthesize the impedance function

$$Z_w = \frac{4s^2 + 2s + 1}{8s^3 + 8s^2 + 2s + 1}$$

   That is, develop the RLC circuit that realizes Z_w. [Parallels Example 2.6.2]
8. Synthesize the impedance function

$$Z_x = \frac{4s^2 + 2s + 1}{4s^2 + 1}$$

   That is, develop the RLC circuit that realizes Z_x. [Parallels Example 2.6.1. You may also want to consult Figure 2.6.2.]
9. Synthesize the impedance function

$$Z_x = \frac{4s^4 + 2s^3 + 5s^2 + 2s + 1}{4s^4 + 4s^3 + 7s^2 + s + 1}$$

   That is, develop the RLC circuit that realizes Z_x. [Parallels Example 2.6.1. You may also want to consult Figure 2.6.2.]
10. Develop the lowpass prototype of a fifth-order Butterworth lowpass filter. There may be more than one solution. That is, draw the circuit of the lowpass filter prototype with element values.
11. Develop the lowpass prototype of a fifth-order Chebyshev lowpass filter with 1 dB ripple and 1 rad/s corner frequency.
12. Develop the lowpass prototype of a ninth-order Chebyshev lowpass filter with 0.01 dB ripple and 1 rad/s corner frequency.

13. A 0.04 S admittance inverter is to be implemented in microstrip using a single length of transmission line. The effective permittivity of the line is 9 and the design center frequency is 10 GHz.
    a. What is the characteristic impedance of the transmission line?
    b. What is the wavelength in millimeters at the design center frequency in free space?
    c. What is the wavelength in millimeters at the design center frequency in microstrip?
    d. What is electrical length of the microstrip transmission line in degrees at the design center frequency?
    e. What is the length of the microstrip transmission line in millimeters?
14. In Section 2.8.2 it was seen that a series inductor can be replaced by a shunt capacitor with inverters and a negative unity transformer. If the inverter is realized with a one-quarter wavelength long transmission line of characteristic impedance $50\ \Omega$:
    a. Derive the $ABCD$ parameters of the cascade of Figure 2.8.2(c) with $50\ \Omega$ inverters.
    b. What is the value of the shunt capacitance in the cascade required to realize a 1 nH inductor?
15. A series inductor of 10 pH must be realized by an equivalent circuit using shunt capacitors and sections of one-quarter wavelength long $1\ \Omega$ transmission line. Design the equivalent circuit. [Hint: The one-quarter wavelength long lines are impedance inverters.]
16. A series inductor of 10 nH must be realized by an equivalent circuit using shunt capacitors and sections of one-quarter wavelength long $50\ \Omega$ transmission line. [Hint: The one-quarter wavelength long lines are impedance inverters.] Design the equivalent circuit.
17. At 5 GHz, a series 5 nH inductor is to be realized using one or more $75\ \Omega$ impedance inverters, a unity transformer, and a capacitor. What is the value of the capacitor?
18. In Section 2.8.3 it was seen that a series capacitor can be replaced by a shunt inductor with inverters and a negative unity transformer. Consider that the inverters are realized with a one-quarter wavelength long transmission line of characteristic impedance $100\ \Omega$.
    a. Derive the $ABCD$ parameters of the cascade of Figure 2.8.3 with the $100\ \Omega$ inverters.
    b. What is the value of the shunt inductance in the cascade required to realize a 1 pH capacitor?
19. A $50\ \Omega$ impedance inverter is to be realized using three resonant stubs. The center frequency of the design is f_0. The first resonant frequency of the stubs is $f_r = 2f_0$.
    a. Draw the circuit using stubs. On your diagram indicate the input impedance and characteristic impedance of each of the stubs if $f_r = 2f_0$.
    b. What is the input impedance of a shorted one-eighth wavelength long transmission line if the characteristic impedance of the line is Z_{01}?
    c. What is the input impedance of an open-ended one-eighth wavelength long transmission?
    d. What is the input impedance of a shorted one-eighth wavelength long transmission if the characteristic impedance of the line is Z_{02}?
    e. What is the length of each of the stubs in the inverter in terms of the wavelength at the frequency f_0?
20. Design a third-order Type 1 Chebyshev highpass filter with a corner frequency of 1 GHz, a system impedance of $50\ \Omega$, and 0.2 dB ripple. There are a number of steps in the design, and to demonstrate that you understand them you are asked to complete the partial designs indicated below. A Cauer 1 lowpass filter prototype is shown below with ω_c being the corner radian frequency, $f_c = \omega_c/(2\pi)$ being the corner frequency, and Z_0 being the system impedance.

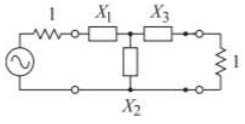


Figure 4.21.1

    a. Design an LPF with $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$.
    b. Design a HPF with $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$.
    c. Design a HPF with $f_c = 1$ GHz, $Z_0 = 1\ \Omega$.
    d. Design a HPF with $f_c = 1$ GHz, $Z_0 = 50\ \Omega$.

21. This problem considers the design of a Butterworth bandpass filter at 900 MHz.
    a. Design an LC second-order Butterworth lowpass filter with a corner frequency of 1 rad/s in a $1\ \Omega$ system.
    b. Using the above filter prototype, design a lowpass filter with a corner frequency of 900 MHz.
    c. Design a second-order Butterworth bandpass filter at 900 MHz using the lowpass filter prototype in (a). Use a fractional bandwidth of 0.1 and a system impedance of $50\ \Omega$.
    d. What is the 3 dB bandwidth of the filter in (c)?
22. Design a third-order maximally flat bandpass filter prototype in a $50\ \Omega$ system centered at 1 GHz with a 10% bandwidth. The lowpass prototype of a third-order maximally flat filter is shown in Figure 2.6.3.
    a. Convert the prototype lowpass filter to a lowpass filter with inverters and capacitors only; that is, remove the series inductors.
    b. Scale the filter to take the corner frequency from 1 rad/s to 1 GHz.
    c. Transform the lowpass filter into a bandpass filter. That is, replace each shunt capacitor by a parallel LC network. This step will establish the bandwidth of the filter.
    d. Transform the system impedance of the filter from 1 to $50\ \Omega$.
23. The lowpass prototype of a fourth-order lowpass Butterworth filter is shown below. The corner frequency is 1 rad/s.

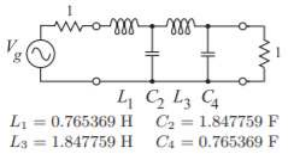


Figure 4.21.2

Based on this, develop a fourth-order Butterworth bandpass filter prototype centered at 10^9 rad/s with a fractional bandwidth of 5%.

a. Scale the lowpass prototype to have a corner frequency of 10^9 rad/s. Draw the prototype with element values.
b. Draw the schematic of the lumped-element fourth-order Butterworth bandpass prototype based on the original lowpass filter prototype.
c. Derive the element values of the lumped-element bandpass filter prototype in a $75\ \Omega$ system.

24. Design a third-order Type 2 Chebyshev highpass filter with a corner frequency of 1 GHz, a system impedance of $50\ \Omega$, and 0.2 dB ripple. There are a number of steps in the design, and to demonstrate that you understand them you are asked to complete the table below. For each stage of the filter synthesis you must indicate whether the element is an inductance or a capacitance by writing L or C in the appropriate cell. Other cells require a numeric value and you must include units. The X element is identified in the prototype below. A Cauer 2 lowpass filter prototype is shown with ωc being the corner radian frequency, $f_c = \omega_c/(2\pi)$ being the corner frequency, and Z_0 being the system impedance.

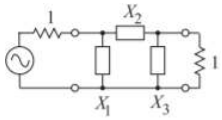


Figure 4.21.3

    a. Complete the LPF (lowpass filter) column of the table with $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$.
    b. Complete the HPF (highpass filter) column of the table with $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$.
    c. Complete the second HPF column of the table with $f_c = 1$ GHz, $Z_0 = 1\ \Omega$.
    d. Complete the third HPF column of the table with $f_c = 1$ GHz, $Z_0 = 50\ \Omega$.

ELEMENT	LPF	
	$\omega_c = 1$ rad/s, $Z_0 = 1\,\Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

Table 4.21.1

ELEMENT	HPF	
	$\omega_c = 1$ rad/s, $Z_0 = 1\,\Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

Table 4.21.2

ELEMENT	HPF	
	$f_c = 1$ GHz, $Z_0 = 1\,\Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

Table 4.21.3

ELEMENT	HPF	
	$f_c = 1$ GHz, $Z_0 = 50\,\Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

Table 4.21.4

25. What is the input impedance of a shorted $\lambda/8$ long transmission line with a characteristic impedance of Z_0?
26. What is the input impedance of an open-circuited $\lambda/8$ long transmission line with a characteristic impedance of Z_0?
27. Apply Richards's transformation to a shunt inductor with a reactance of $50\,\Omega$. What is the electrical length of the shorted stub if the stub has a characteristic impedance of $50\,\Omega$?
28. Apply Richards's transformation to a shunt capacitor with a reactance of $-50\,\Omega$. What is the characteristic impedance of the open-circuited stub if the electrical length of the stub is one-quarter wavelength long?

4.21.1: Exercises by Section

†challenging, ‡very challenging

§2.2 1†, 2

§2.4 3†, 4†

§2.5 5†

§2.6 6†, 7†, 8†, 9†

§2.7 10, 11†, 12†

§2.8 13, 14†, 15†, 16†, 17†, 18†, 19†

§2.9 20†, 21†, 2 † 2, 23†, 24†

§2.12 25, 26, 27†, 28†

4.21.2: Answers to Selected Exercises

1. (c) $\Gamma(s) = \frac{s^2}{s^2+\sqrt{2}s+1}$
2. $|T(w)|^2 = 1/(1+\omega^8)$
3. (f) $-0.38 + \jmath 0.92$, $-0.38 - \jmath 0.92$, $-0.92 + \jmath 0.38$, $-0.38 - \jmath 0.92$
7.

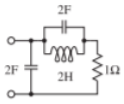


Figure 4.21.4

13. (c) 9
14. (a) $\begin{bmatrix} 1 & sC(50)^2 \\ 0 & 1 \end{bmatrix}$
17. 889 fF
18. 10 nH
19. (c) $-Z_2 = -\jmath 50\,\Omega$
24. 1 GHz, 50 Ω, HPF : $X_3 = 6.5$ nH

This page titled 4.21: Exercises is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **2.21: Exercises** has no license indicated.

CHAPTER OVERVIEW

5: Oscillators

5.1: Introduction
5.2: Oscillator Theory
5.3: Reflection Oscillator
5.4: Case Study- Reflection Oscillator
5.5: Advanced Discussion of Oscillator Noise
5.6: Summary
5.7: References
5.8: Exercises

This page titled 5: Oscillators is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

5.1: Introduction

Most RF oscillators generate sinusoidal signals that are either used to drive mixers or, if modulated, to produce frequency modulated signals directly. In some designs the microwave oscillators drive flip-flop circuits that produce periodic square-wave signals with multiple phases as required for the LO drives of I/Q modulators. RF and microwave oscillators can be designed using either a two-port or a one-port approach. The classic treatment of oscillators is based on a two-port gain device with a feedback loop, but the oscillator can nearly always be viewed, and thus more conveniently designed at microwave frequencies, as a one-port in which a resonant circuit, called the **tank** circuit, is connected to an active circuit that presents a negative resistance. However, the stability, noise, and start-up analyses of an oscillator are based on a two-port with feedback.

All microwave oscillator designs are based on one of three basic oscillator configurations. Suitable configurations must have as few reactive elements as possible while enabling stable single-frequency oscillation. These configurations have three or four reactive element. Mapping a microwave oscillator design on to one of the standard oscillator designs is not simple mainly because active devices at microwave frequencies have substantial parasitic capacitances and also adjustments must be made to accommodate biasing. Stability of a microwave oscillator is of great concern and the essence is that there should be as few significant energy storage elements as possible. If there are parasitic energy storage elements these should be quite small or absorbed into the capacitors of one of the basic oscillator configurations. With more energy storage elements than necessary the chances of unwanted resonances is much higher and hence instability of an oscillator is more likely.

At microwave frequencies the Q of lumped elements is limited. A lumped inductor in the tank circuit has a particularly low Q and if there is room it is replaced by a transmission line. Oscillators that are fixed in frequency can use a resonant circuit with high Q circuit elements. This contrasts with voltage-controlled oscillators (VCOs) that have a lossy variable element, nearly always a varactor, in the tank circuit in what is now a low Q resonant circuit. The voltage-controlled variation of (invariably) the capacitance of this element changes the frequency of the oscillator.

A microwave oscillator could be realized on-chip or realized as a hybrid circuit with a packaged active device and packaged lumped-elements and possibly transmission lines for the resonant circuit. A hybrid design has much more flexibility than an on-chip design, and if designed correctly has better performance than a monolithically integrated design. Hybrid design techniques are much more mature than chip-based designs but over time some of the techniques used with hybrid designs will migrate to on-chip designs. Also, on-chip designs are preferably differential with transistors in a push-pull, i.e. differential, configuration. Better performance of an on-chip oscillator can be obtained by using an off-chip resonator.

Oscillator theory derives from the analysis of a two-port with gain and feedback. This theory is described in Section 5.2. The following sections explore practical oscillator configurations. Section 5.3 presents the design technique for designing a fixed-frequency microwave oscillator and this is followed up with a design case study in Section 5.4. The distinguishing feature here is that the resonant circuit consists of high Q elements. Section 5.5 presents a design approach for a voltage-controlled microwave oscillator. A state-of-the-art case study of a $5\ \mathrm{GHz}$ VCO design is presented in Section 5.6. Section 5.7 describes the design of an on-chip differential oscillator. The last two sections, Sections 5.8 and 5.9 in this chapter describe oscillator phase noise, a characteristic that is the fundamental performance limiting parameter of an oscillator as, for example, it affects the sensitivity and performance of communication systems and the range of radar.

This page titled 5.1: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.1: Introduction** has no license indicated.

5.2: Oscillator Theory

Microwave oscillators are usually implemented as reflection oscillators with two connected one-port circuits with one being an active device configured as a one-port and presenting a negative conductance, and a second oneport network being the tank or resonator network which must be designed to have specific admittance versus frequency characteristics that ensure stability. However the great body of network theory available derives from control theory and general circuit theory, and is based on a two-port oscillator with linear frequency-selective feedback.

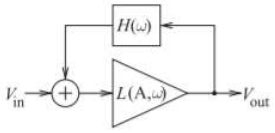


Figure 5.2.1: Representation of a feedback oscillator based on a two-port active device.

5.2.1: Theory of Oscillation

The two-port view of oscillators is based on the amplifier plus feedback loop shown in Figure 5.2.1. Here $L(A,\omega)$ describes the characteristics of an amplifier whose transfer function is dependent on both amplitude, A, and radian frequency, ω, and has a magnitude greater than 1. $H(\omega)$ describes the transfer characteristics of a linear feedback network that is dependent only on frequency. The output of the feedback system is described by

$$V_{\text{out}} = L(A,\omega)V_{\text{in}} + L(A,\omega)H(\omega)V_{\text{out}} \tag{5.2.1}$$

and so

$$V_{\text{out}} = \frac{L(A,\omega)V_{\text{in}}}{1 - L(A,\omega)H(\omega)} \tag{5.2.2}$$

The aim of most oscillator design is to use an active device whose characteristics are independent of frequency but whose transfer response depends on the output amplitude. Then Equation (5.2.2) becomes

$$V_{\text{out}} = \frac{L(A)V_{\text{in}}}{1 - L(A)H(\omega)} \tag{5.2.3}$$

Oscillation begins with input noise when the oscillator is powered on. If the denominator of Equation (5.2.3) is close to zero, oscillations build up at a frequency determined by the feedback network. As the amplitude of the oscillations builds, $L(A)$ compresses until the denominator is finite but close to zero[1], and there are stable oscillations. Stable oscillation is no accident and is the result of careful design, and much depends on the nature of the feedback network. When an amplifier becomes unstable, for example, the signal produced is often chaotic, with rapid variations in amplitude and frequency, and is not the single frequency sinusoid required of an oscillator.

5.2.2: Basic Oscillator Configurations

There are two dominant types of feedback networks, the Pi- and T-type networks, shown in Figure 5.2.2. Three Pi-type networks have proven to be particularly suited to the amplitude saturation characteristics of FET and BJT active devices and, ignoring parasitic capacitances, they result in stable oscillation. These are the Hartley, Colpitts, and Clapp circuits shown in Figure 5.2.3. The grounds shown can be considered to be common terminals instead.

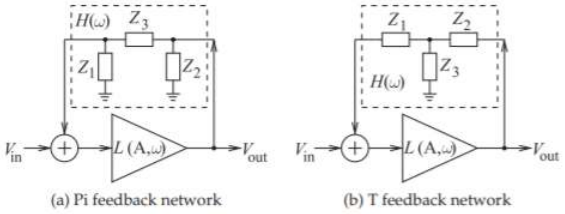


Figure 5.2.2: Feedback oscillator with Pi- and T-type feedback networks.

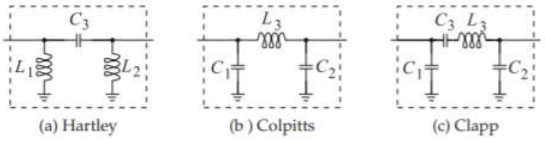


Figure 5.2.3: Basic oscillator feedback networks.

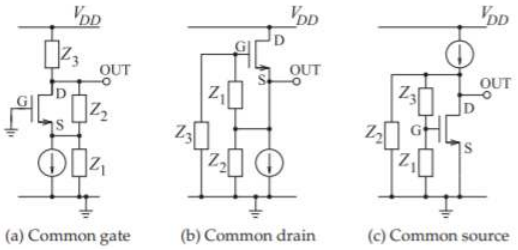


Figure 5.2.4: Circuit schematics for FET feedback oscillators using Pi-type feedback networks. The current sources provide bias.

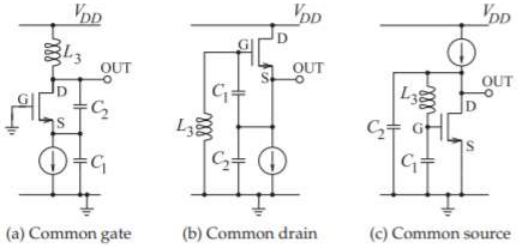


Figure 5.2.5: Circuit schematics for Colpitts FET oscillators. The current sources provide bias. (In (a) C_1 is across the input of the gate-source transistor with connection through the ground.)

The Hartley and Colpitt's basic oscillator configurations have three large energy storage elements. The Clapp oscillator has four significant energy storage elements but it is important because sometimes it is desirable to use a very high Q resonator such as a piezoelectric crystal which is electrically modeled as a series LC circuit.

Single transistor oscillators using the feedback networks arranged in common-gate, common-source or common-drain configurations are shown in Figure 5.2.4. FET Colpitts oscillators are shown in Figure 5.2.5. The Colpitt's configuration is the most common configuration for a microwave oscillator

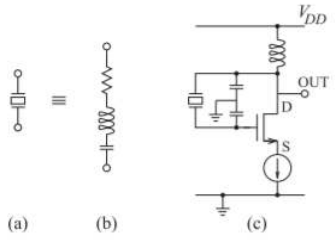


Figure 5.2.6: Crystal oscillator: (a) schematic symbol for a crystal; (b) its equivalent circuit; and (c) a FET crystal oscillator. The most commomly used form of a crystal oscillator is a Clapp oscillator because of the series LC electrical model of a piezolectric crystal.

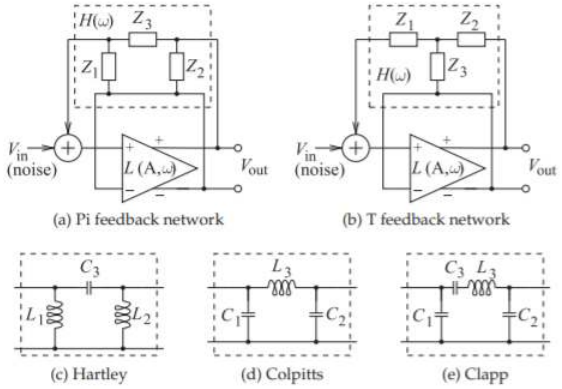


Figure 5.2.7: Differential feedback oscillator with Pi- and T-type feedback networks.

as it can absorb the parasitic capacitances of a transistor into the C_1 and C_2 capacitors, as in Figures 5.2.5(a and c). The circuits shown in Figures 5.2.2-5.2.5 are the bases of nearly every microwave oscillator. Many modifications lead to better stability, compensate for transistor parasitics, and accommodate differential signaling.

Crystal references, for example, are commonly used for fixed-frequency oscillators, as the piezoelectric (usually quartz) crystals used have a very high Q and the oscillation frequency is very stable, typically to a few parts in a billion, and better if the crystal is temperature stabilized. A circuit with a crystal-based feedback loop creating a Clapp FET oscillator is shown in Figure 5.2.6. Common quartz reference crystals available as off-the-shelf components are 10 MHz, 20 MHz, and 40 MHz but are available up to 300 MHz. So they are not microwave oscillators but they are used as reference signals in a phase-locked loop to precisely set the frequency of microwave VCOs. Fixed-frequency microwave oscillators do not have this method available for stabilizing the oscillation frequency.

The Hartley, Colpitts, and Clapp networks are also widely used as the bases of oscillators in CMOS RFICs, where the feedback network is closely approximated using differential circuits [5, 6, 7, 8, 9, 10]. The basic differential configurations are shown in Figure 5.2.7.

Example 5.2.1: Common Emitter BJT Clapp Oscillator

Draw the schematic of a common-emitter BJT Clapp oscillator without biasing and then draw it with biasing.

Solution

The first step is defining the input and the output connections to be used in the oscillator configuration. There are several ways to do this and the solution here considers two configurations, Circuit A and Circuit B.



Figure 5.2.8

Thus the possible common-emitter Clapp oscillators are (as in Figure 5.2.4 but with a BJT)

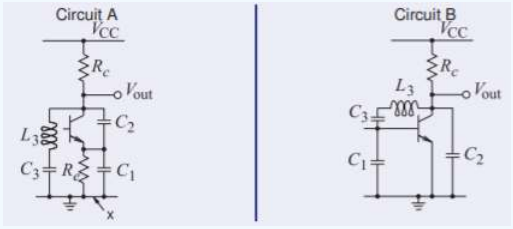


Figure 5.2.9

In Circuit A the point x is the input terminal of the oscillator and the emitter terminal is the common point of the basic Clapp oscillator configuration. Appropriate bias must be applied to the base of the transistor. Also it would be good not to deliver power to R_c and instead deliver it to the load. In the circuits below a choke inductor L_c, an open circuit at RF, is used to apply bias. L_c also enables a large output voltage swing. C_c is a large coupling capacitor and has very low impedance at RF and blocks DC.

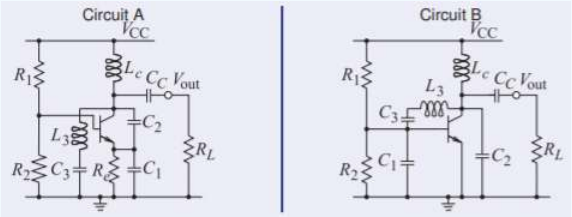


Figure 5.2.10

Of these Circuit B is the more attractive as the parasitic base-emitter capacitance will be absorbed into C_1, and the parasitic collector-emitter capacitance will be absorbed into C_2. Neither configuration can absorb the collector-base capacitance, C_{cb}, so extra care must be made that C_{cb} does not result in instability.

5.2.3: Footnotes

[1] This description of oscillation, based on Equation (5.2.3), is behind the erroneous **Barkhausen stability criterion**, which is also known as the **Barkhausen oscillation criterion**. Barkhausen himself used the criterion, known as the **Barkhausen criterion**, to establish the frequency of oscillation as $L(A, \omega)H(\omega) = 1$. This was misinterpreted as a stability or oscillation criterion. It is a necessary criterion for two-port feedback oscillation, but not sufficient. It does not indicate whether a system is unstable. Instead, the Nyquist criterion is the necessary and sufficient criterion for oscillation in feedback oscillators [1, 2, 3, 4]. The Barkhausen criterion should not be used in determining whether oscillation occurs.

This page titled 5.2: Oscillator Theory is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.2: Oscillator Theory** has no license indicated.

5.3: Reflection Oscillator

Design of stable microwave oscillators traditionally uses the one-port oscillator stability criterion outlined by Kurokawa [11, 12]. Most two-port oscillators can be designed by casting them in the form of a one-port oscillator. In applying the condition, each of the networks—the active device, the resonator load, and the device termination—are characterized as one-ports.

5.3.1: Kurokawa Oscillation Condition

The oscillation condition for a stable reflection oscillator can be expressed in terms of the admittances of the resonator network, $Y_r = G_r + \jmath B_r$, and of the active network, $Y_d = G_d + \jmath B_d$, see Figure 5.3.1(a). The Kurokawa oscillator condition establishes that for stable (single-frequency) oscillation [11]

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega} - \frac{\partial B_d}{\partial V}\frac{\partial G_r}{\partial \omega}\right)\Big|_{V=V_0,\omega=\omega_0} > 0 \tag{5.3.1}$$

where the subscript 0 refers to the operating point, r refers to the resonator, and d refers to the active device. The active device and resonator characteristics shown in Figures 5.3.1(b and c) satisfy the Kurokawa condition. In Equation (5.3.1) V_0 is the amplitude of the oscillation at the interface of the active and resonator networks. If the condition in Equation (5.3.1) is not met, then the oscillator may simultaneously oscillate at multiple frequencies. The Kurokawa oscillation condition must be met at all times which includes during start-up, including when bias is applied, of the oscillator.

With a fixed-frequency oscillator the resonator network is linear so G_r and B_r are independent of amplitude and G_r (by design) is independent of frequency. B_r varies with frequency. Ideally the active device has a frequency independent G_d and amplitude independent B_d. Achieving this is a major design task. With these conditions $\partial B_d/\partial V \approx 0$ and $\partial G_r/\partial \omega_0 \approx 0$ and the

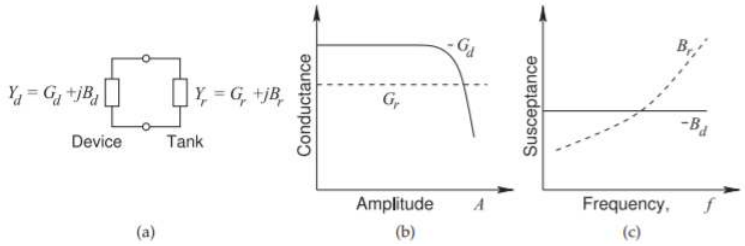


Figure 5.3.1: Reflection oscillator operation: (a) one-port oscillator; (b) as the amplitude of the oscillation increases, the magnitude of the device conductance, $|G_d|$, decreases while the conductance of the tank circuit, G_r, is constant; and (c) as the frequency of the oscillation increases, the susceptance of the tank circuit, B_r, changes while, B_d (ideally) does not change.

Kurokawa condition in Equation (5.3.1) simplifies to

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega_0}\right)\Big|_{V=V_0,\omega=\omega_0} > 0 \tag{5.3.2}$$

Thus design is greatly simplified for a fixed-frequency oscillator with a high-Q resonator.

5.3.2: Reflection Oscillator Design Approach

When a device with admittance $Y_d = G_d + \jmath B_d$ is connected in shunt to a resonator of admittance $Y_r = G_r + \jmath B_r$ (see Figure 5.3.1 (a)), the voltage amplitude A and radian frequency ω of the resulting equilibrium oscillation are determined when $-G_d(A) = G_r(\omega)$ and $-B_d(A) = B_r(\omega)$. Here the assumption is that the device conductance only is a strong function of voltage amplitude, while the resonator admittance is a function only of angular frequency. This condition can be represented graphically by first denoting the locus of the negative of the device's complex admittance as $-Y_d(A) = -[G_d(A) + \jmath B_d(A)]$ (also referred to as the inverse device reflection coefficient, or $1/\Gamma$ locus (sometimes referred to as the $1/S$ locus)) and the locus of the resonator admittance as $Y_r(\omega) = G_r(\omega) + \jmath B_r(\omega)$. Then, for stable single-frequency oscillation, the intersection of these loci occurs at a single point (i.e., at a single amplitude and frequency combination).

In most oscillator design, the aim is to make the device admittance independent of frequency, and, of course, the admittance of the linear tank circuit is independent of the amplitude of the oscillating signal. These oscillation conditions are depicted in Figure 5.3.1

(b and c). In Figure 5.3.1(b), as the amplitude of the oscillation increases, the magnitude of the device conductance, $|G_d|$, decreases while the conductance of the tank circuit, G_r, is constant. In Figure 5.3.1(c), as the frequency of the oscillation increases, the susceptance of the tank circuit, B_r, changes while B_d (ideally) is slow to change. The intersections define the amplitude and frequency of the oscillation.

If the device admittance is dependent on frequency, then it is difficult to avoid multiple intersections of the $-Y_d(1/\Gamma_d)$ and $Y_r(\Gamma_r)$ loci as viewed in the complex plane. The angle of intersection of the Y_r and $-Y_d$ loci is an important indicator of stability relating to multiple oscillations, oscillator start-up problems, and excess noise [13]. Thus the appropriate angular intersection of these loci is critical. It is difficult to achieve all of the objectives in design unless Y_d is frequency independent.

Ideally, resonator design requires that Q be maximized so that $G_r \approx 0$. This can be achieved with a fixed-frequency microwave oscillator as the resonator can be implemented with a capacitor and a transmission line segment both of which have very low loss. However the tunable capacitors in a VCO design are lossy and so the Q is not high. The two types of oscillators need a different design approach. Furthermore, for a VCO, voltage tuning of the resonator must satisfy the stability criteria, including a single point of intersection and appropriate angle of intersection, over the tuning range. With emphasis on these characteristics and since there are device capacitive parasitics, achieving a proper stable resonator-device interface can be troublesome. An alternative and equally viable approach to stability analysis of a broad class of oscillators, particularly for those using three-terminal devices, is application of the two-port criteria developed for amplifier stability assessment. However, the one-port design approach is preferred by microwave designers because the one-port connection is closer to the intended operation. The one-port assessment of oscillator stability is not unlike the Bode criteria applied to two-port feedback systems [14, 15]. However, unlike the two-port open-loop assessment of stability, the one-port characterization technique is conveniently aligned with the measurements that can be made by a VNA [16, 17]. As well, the nonlinear limiting effect of the active device is readily measured.

5.3.3: Summary

Microwave oscillator design invariably uses the reflection oscillator approach in which a one-port active device network is connected to a one-port resonator network. Design is complicated enough and it is necessary to simplify design and limit the design space. The procedure almost universally followed is to design for the characteristics shown in Figure 5.3.1. Design of a fixed-frequency oscillator is further simplified because the conductance of the resonator is almost zero.

This page titled 5.3: Reflection Oscillator is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.3: Reflection Oscillator** has no license indicated.

5.4: Case Study- Reflection Oscillator

In this case study[1] the design of the reflection oscillator shown in Figure 5.4.1(a) is examined. This is an 18 GHz common-gate oscillator with series inductive feedback provided by the transmission line TL_1. This circuit is derived from the Colpitts oscillator configuration. The resonator is resonant considerably below the oscillation frequency, and so presents a capacitance to the source of the transistor at the oscillation frequency.

5.4.1: Design Procedure

The circuit in Figure 5.4.1(a) is a modified common-base Colpitts oscillator and the RF equivalent circuit required to understand oscillator operation, and how the circuit corresponds to the standard Colpitts oscillator, are shown in Figure 5.4.2. This is the most common fixed-frequency microwave oscillator topology.

The Colpitts feedback network is shown in Figure 5.4.2(a) where an inductor provides feedback from the output to the input of the active device. Returning to the modified Colpitts oscillator of Figure 5.4.1, L_1 is a large choke inductor presenting an RF open circuit and C_1 is a large DC blocking capacitor presenting an RF short circuit. The gate transmission line, TL_1, presents a small inductance and most importantly provides inductive feedback from the output to the input of the active device and closely corresponds to L_3 in the Colpitts feedback network of Figure 5.4.2(a). The gate-source and drain-source parasitic capacitances of the transistor, C_{GS} and C_{DS}, correspond approximately to capacitors C_1 and C_2 in the Colpitts feedback network. This results in the RF equivalent circuit shown in Figure 5.4.2(b). In Figure 5.4.2(b) X_R is the reactance of the resonator and compensates for the nonideal nature of the modified Colpitts oscillator. Stable operation of this oscillator requires that X_R be capacitive but have a variation with frequency substantially less than that of a capacitor. This is accomplished by having the oscillation frequency above the resonant frequency of the resonator.

In Figure 5.4.1 the reflection oscillator is composed of two networks, the active device network and the resonator network. The third network shown, the OscProbe network, is a probe used to control oscillator analysis using the harmonic balance method. The design approach here is to develop a topology incorporating the active device that presents a negative resistance at the interface, X, between the active device network and the resonator network. Since the active device and the resonator are each best modeled as parallel circuits, it is best to refer to admittance and so the active device network presents a negative conductance to the resonator. In addition, the small-signal admittances of the active network, $Y_d = G_d + jB_d$, and of the resonator network, $Y_r = G_r + jB_r$, are shown in Figure 5.4.3. Here the active network has a small-signal conductance that is negative and a small-signal

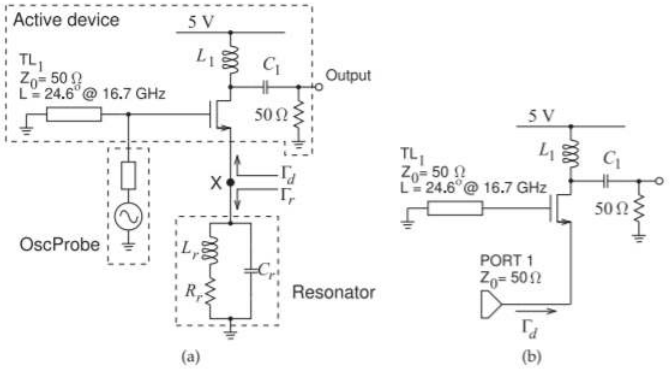


Figure 5.4.1: Reflection oscillator: (a) complete oscillator circuit used in simulation; and (b) configuration for measuring the large-signal reflection coefficient of the active device. $L_r = 5.6$ nH, $R_r = 10\,\Omega$, $C_r = 445$ fF, $L_1 = 15$ nH, and $C_1 = 10$ pF. The resonant frequency of the resonator is 3.19 GHz, but this is not the oscillation frequency. It presents the required slope of susceptance with respect to frequency at the oscillator frequency of 17.76 GHz.

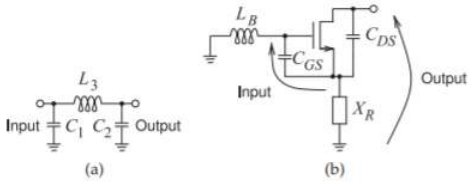


Figure 5.4.2: Operation of the modified Colpitts oscillator: (a) Colpitts feedback network; (b) the essential RF equivalent circuit of the Modified Colpitts oscillator of Figure 5.4.1.

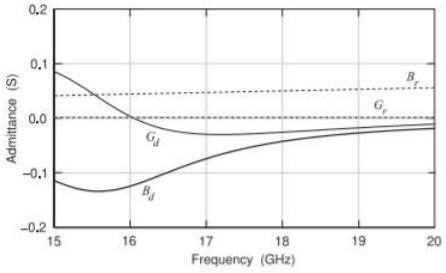


Figure 5.4.3: Small-signal admittance of the active device, $Y_d = G_d + jB_d$, and admittance of the resonator, $Y_r = G_r + jB_r$. B_d varies with frequency largely because of the frequency-dependent feedback provided by TL_1.

susceptance that is inductive. The resonator has negligible conductance and a capacitive susceptance.

The admittance looking into the source of the active network depends on the level of the signal at X in Figure 5.4.1. However, the admittance of the resonator is independent of the signal level since the resonator is a linear network. The design strategy is then to develop a feedback network, here the transmission line, TL_1, in the gate of the FET, that then presents a negative conductance to a frequency-selective structure, the resonator network.

The effect of the signal level is examined using the active network test circuit shown in Figure 5.4.1(b). Here a $50\ \Omega$ generator, at Port 1, drives the source of the active device. The power of the signal at the port is varied and it is found that the negative conductance varies from -0.0274 S at a small applied signal, to -0.0224 S at an applied signal level of -10 dBm, to -0.004 S at -7 dBm, and to 0.001 S at -6 dBm. Oscillation will occur when the conductance looking into the source terminal of the active device is approximately zero (since the conductance of the resonator is negligible) and the susceptances of the active network and the resonator network cancel.

Unfortunately the susceptance of the active network changes as the signal level varies since the active device capacitances are nonlinear. The effect of this will now be examined. It is instructive to view the effect of signal level on the oscillation condition by considering the reflection coefficients Γ_d for the active device network and Γ_r for the resonator network. This is shown in Figure 5.4.4, where a polar plot is used since $|\Gamma_d| > 1$ (except for a very large signal at X). Γ_r is approximately on the unit circle and is capacitive being in the lower half plane. Γ_d is shown for several signal levels, with a signal level of -20 dBm corresponding to the small-signal condition. As the signal level increases eventually to -6 dBm, both the conductance and susceptance of the active device change. The device conductance becomes positive as Γ_d crosses inside the unit circle. Resonance occurs when $\Gamma_d\Gamma_r = 1$. Since $G_r \approx 0$, this is when $\Gamma_d \approx 1/\Gamma_r$. That is, resonance will occur at the frequency where the locus of Γ_d intersects with $1/\Gamma_r$.

The results presented in Figure 5.4.4 for the reflection coefficient of the active network in Figure 5.4.1(b) are derived from a large-signal solution calculated using harmonic balance analysis. In most harmonic balance analyses it is only necessary to consider a few harmonics to obtain good results. In the simulation here, five harmonics are considered, with the fundamental frequency set by the frequency of the source at the driven port. The fundamental frequency was stepped from 15 GHz to 20 GHz. Oscillator

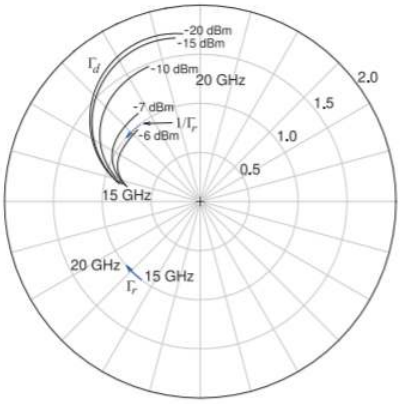


Figure 5.4.4: Reflection coefficient of the resonator, Γ_r, and of the active device, Γ_d, at various signal levels. Γ is plotted on a polar plot with the outer circle corresponding to $|\Gamma| = 2$. As desired, the locus of $1/\Gamma_r$ is parallel to Γ_d at a fixed signal level (here about -6 dB), and the direction of $1/\Gamma_r$ with increasing frequency is opposite that of Γ_d. The close parallel match is ensured by adjusting the susceptance-versus-frequency slope of the resonator. Considering the intersection of $1/\Gamma_r$ and Γ_d is equivalent to considering the intersection of $1/\Gamma_d$ and Γ_r.

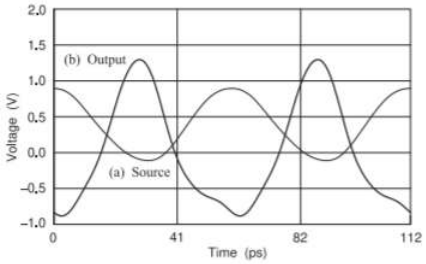
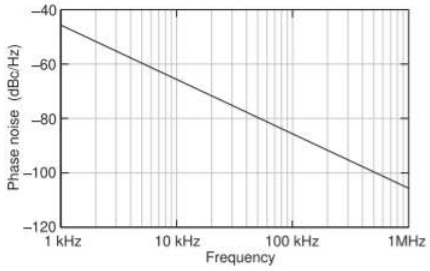


Figure 5.4.5: Voltage waveforms at the output of the oscillator (Curve (b)), and at the source terminal of the active device, Terminal x, which is the interface between the active device and the resonator (Curve (a)).

simulation also uses harmonic balance analysis, but now the frequency of oscillation is not known ahead of time. It is necessary to introduce another condition to enable the simulator to find the oscillation frequency. One of the techniques used in harmonic balance simulators is the introduction of an oscillator probe element such as the OscProbe element shown in Figure 5.4.1(a). The frequency of the source, F_{OSC}, in the OscProbe element is initially guessed by the simulator and the series impedance of the OscProbe element is a short circuit at the oscillation frequency and an open circuit at the harmonics. This extra condition is included in the harmonic balance equations and F_{OSC} is allowed to vary as well as the amplitude of the source. The oscillation solution is obtained when the current through the impedance in the OscProbe element is zero. The oscillation frequency is found to be 17.76 GHz. Note that this is not the resonant frequency of the resonator, as here the active device presents an inductance to the resonator. The resonator must present an effective capacitance that has the required susceptance-versus-frequency slope so that the frequency locus of $1/\Gamma_r(f)$ is parallel but oppositely directed to that of $\Gamma_d(f)$ (see Figure 5.4.4).

Figure 5.4.5 shows the waveforms at the output of the oscillator (Curve (b)) and at the interface (X) between the resonator and active device networks



https://eng.libretexts.org/@go/page/113745

Figure 5.4.6: Phase noise of the reflection oscillator.

(Curve (a)). The output of the oscillator would be followed by a bandpass filter so that a single harmonic-free sinewave is presented to the external output terminal.

The phase noise of the oscillator is calculated from the oscillator steadystate conditions and is shown in Figure 5.4.6. The phase noise level is relatively high and this is principally due to the resistance R_r in the resonator, which models the loss in the lumped inductor. Performance would be improved if, instead, a distributed circuit (with transmission lines) was used to realize the required admittance and its derivative at the oscillation frequency.

5.4.2: Summary

The zero combined suspectance point in Figure 5.4.3 was at 17.76 GHz while the oscillation frequency shown in Figure 5.4.5 is at 17.29 GHz. The reason for this discrepancy is that the susceptance plot in Figure 5.4.3 is based on small-signal conditions whereas the simulation is for a large signal and includes harmonics. It can be expected that the oscillation frequency calculated using harmonic balance analysis will be different from that obtained from a small-signal analysis. Consider the Γ_d curves in Figure 5.4.4. (Generating this data took many simulation runs and did not consider harmonics.) The magnitude of Γ_d reduces as the applied signal level increases. This rotation is not desired and means that the susceptance of the active device is amplitude-dependent. This corresponds to the magnitude of the negative conductance of the active device getting smaller. Note that there is also a small rotation of Γ_d indicating that the susceptance of the active device also changes as the signal level increases. This means that the oscillation frequency will depend on the amplitude of the oscillation. In addition, the applied signal used to calculate the Γ_d at different power levels is not the actual oscillating signal.

5.4.3: Footnotes

[1] AWR Design Environment Project File: Reflection_Oscillator.emp.

This page titled 5.4: Case Study- Reflection Oscillator is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.4: Case Study- Reflection Oscillator** has no license indicated.

5.5: Advanced Discussion of Oscillator Noise

This section presents a discussion of oscillator noise, and particularly the rapid increase of phase noise close to the carrier. Noise can be partitioned into amplitude and phase noise components. The nonlinear saturation of an oscillator suppresses amplitude noise so only phase noise is of concern. While usually associated with oscillators, phase noise is also added to a signal by an amplifier.

There is not a consensus as to the origins of close-to-carrier phase noise. This section begins with observations of oscillator noise in the frequency domain and in the time domain. Then three theories of excess oscillator noise, Leeson's theory, the linear time-invariant model, and the chaotic map model, are presented.

Not having a complete model of the physical origin of phase noise means that a simulator cannot reliably predict the phase noise of an oscillator. Also designing an oscillator with good phase noise performance currently relies heavily on experience and projections based on what has been achieved by a designer previously.

5.5.1: Observations of Oscillator Noise in the Frequency Domain

The most puzzling noise observed with oscillators is the noise observed at a small frequency offset from the carrier (i.e., the average oscillation signal). To develop an appreciation for the breadth of observations, the signals produced by several different oscillators will be considered. First, Figure 5.5.2

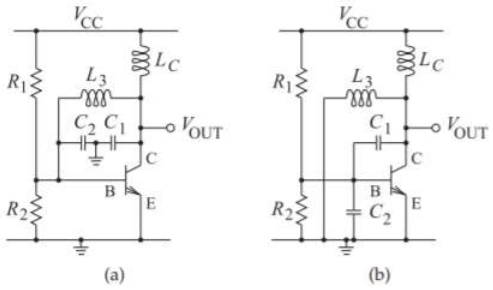


Figure 5.5.1: Common emitter BJT Colpitts oscillators: (a) configuration with a feedback network between the collector and base of the transistor; and (b) alternative configuration.

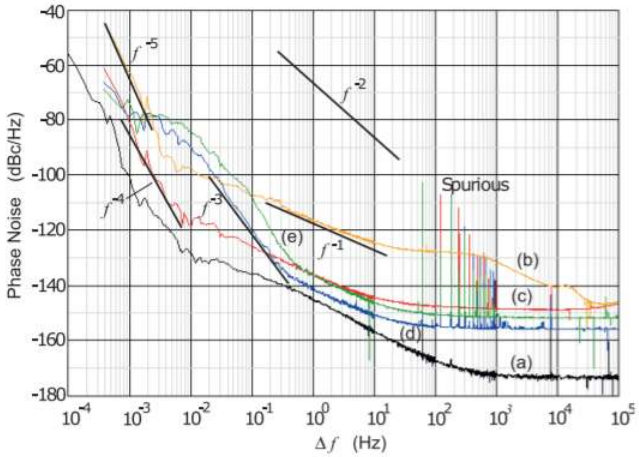


Figure 5.5.2: Measured phase noise of low-frequency oscillators: (a) instrument noise floor; (b) HP 5087A frequency distribution amplifier at 5 MHz (used to drive the external reference input of several test instruments using a single high-quality oscillator); (c) TADD-1 frequency distribution amplifier at 10 MHz; (d) TADD-1 frequency distribution amplifier at 5 MHz; (e) Spectracom 8140T frequency distribution amplifier at 10 MHz. Five phase noise regions are identified as f^{-5}, f^{-4}, f^{-3}, f^{-1}, and white noise. The spurious signals are related to injected harmonics of the 60 Hz power mains. Used with permission of John Ackermann [23].

is a plot of the phase noise observed at the output of several oscillators and amplifiers operating at 5 MHz and 10 MHz. Curve (a) is the noise floor of the noise measurement instrument and spurious tones are observed at multiples of 60 Hz, the power mains

frequency. Curves (b), (c), (d), and (e) show phase noise varying in straight-line segments. Being a log-log plot, these curves show phase noise varying as f^{-5}, f^{-4}, f^{-3}, f^{-1}, and f^{0}. None of the phase noise plots here show a region with an f^{-2} dependence, although this is observed with other oscillators.

Another oscillator to consider is the VCO circuit shown in Figure 5.5.3 [24]. This is a 50 MHz VCO with a semiconductor varactor being the variable element with a zero-bias capacitance of 100 pF. The capacitance of the varactor is controlled by the voltage, V_b. With $V_b = 0$ V, the phase noise shown in Figure 5.5.4 was observed. The distinct phase noise regions have frequency dependencies of f^{0}, f^{-1}, f^{-2}, and f^{-3}. The phase noise of this oscillator is plotted again in Figure 5.5.5 for three different varactor bias voltages. The phase noise characteristics of the oscillator change even though the underlying physical sources of noise do not change (of course). Curve (a), with $V_b = 6$ V, and Curve (b), with $V_b = 0$ V, have an f^{-1} region around 20 kHz (see Figure 5.5.4 for more details), but the f^{-1} region is not observed in Curve (c) where $V_b = 18$ V. One interpretation is that the crossover frequencies have shifted. So what is particularly interesting here is

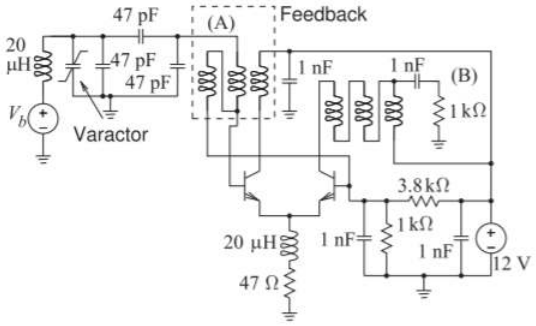


Figure 5.5.3: Varactor-tuned VCO schematic, from [24].

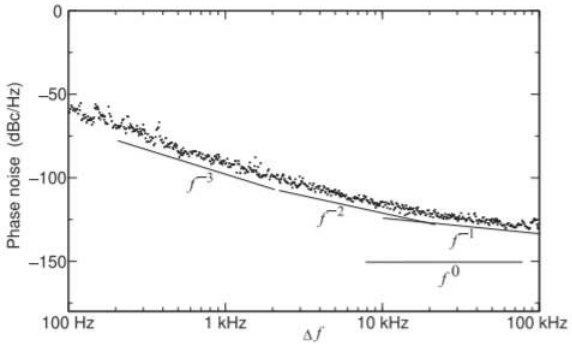


Figure 5.5.4: Measured phase noise of a 50 MHz BJT varactor-based VCO with the varactor biased at 0 V [25, 26]. Three phase noise regions are identified as f^{-3} (having a slope of −9 dB/octave), f^{-2} (having a slope of −6 dB/octave), and f^{-1} (having a slope of −3 dB/octave).

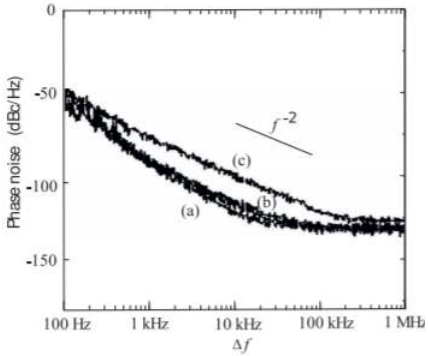


Figure 5.5.5: Measured phase noise of a 50 MHz BJT varactor-based VCO at three varactor bias voltages: (a) 6 V; (b) 0 V; and (c) 18 V [25]. The varactor breakdown voltage is 30 V Curve (b) was also plotted in Figure 5.5.4.

https://eng.libretexts.org/@go/page/113749

that the same physical source of noise can be manifested quite differently at the output of an oscillator when the circuit bias is changed.

The third phase noise example is for a 2.4 GHz power oscillator that has the output spectrum shown in Figure $5.5.6$ with regions having dependencies of f^{-3} and f^{-0}, but nothing in between. (The slight increase in noise power spectral density at 40 kHz offset is due to dynamics of the oscillator's

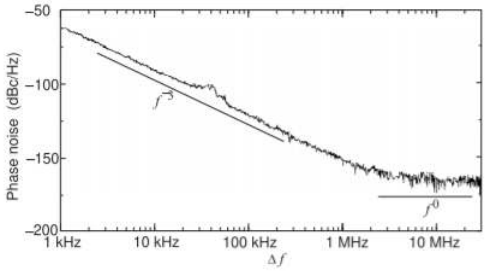


Figure $5.5.6$: Phase noise of a 2.4 GHz power oscillator with an output power of 34.5 dBm [27, p. 323]. Two phase noise regions are identified as f^{-3} (having a slope of -9 dB/octave) and white noise (with an f^0 dependency).

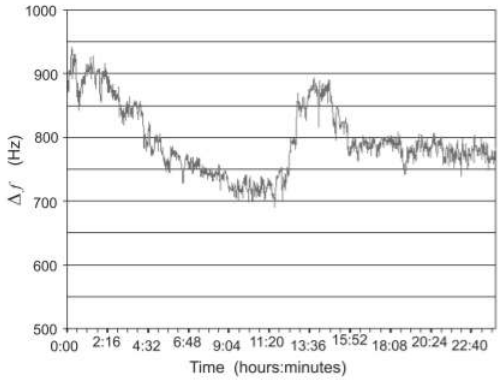


Figure $5.5.7$: Long-term stability of a 10 GHz oscillator measured over a 24 hour interval after being on for 3 weeks. Used with permission of John Ackermann [23].

feedback loop.) Finally, the 5 GHz oscillator considered in Section 5.6 has f^{-3} and f^{-2} phase noise regions (see Figure 5.6.15).

So the whole range of phase noise dependencies on frequency offset are observed, but the universal observation is that the dependency of the noise power spectral density is to a non-positive integer power of frequency (i.e., $f^{-n},\ n = 0, 1, ...$).

5.5.2: Observations of Oscillator Noise in the Time Domain

An important time-domain characterization of noise is referred to as random walk noise. An example of this is the variation of the oscillation frequency over a long period of time. The long-term stability of a 10 GHz oscillator is shown in Figure $5.5.7$. This noise cannot be characterized in the frequency domain and instead is described by its **Allan variance**, $\sigma_y^2(\tau)$, or **Allan deviation**, $\sigma_y(\tau) = \sqrt{\sigma_y^2(\tau)}$, defined as follows.

If the frequency measured at time t is $f(t)$ and the nominal oscillation frequency is f_n, then the fractional frequency at time t is defined as

$$y(t) = \frac{f(t) - f_n}{f_n} \tag{5.5.1}$$

Then the average fractional frequency over an observation time interval τ is defined as

$$\overline{y}(t, \tau) = \frac{1}{\tau} \int_0^{\tau} y(t + t_v) dt_v \tag{5.5.2}$$

This leads to the definition of the Allan variance as

$$\sigma_y^2(\tau) = \frac{1}{2}\langle \overline{y}_{n+1} - \overline{y}_n \rangle \tag{5.5.3}$$

where τ is the observation period and $\overline{y}_n$ is the nth fractional frequency average over the time interval τ. Note that there is no dead-time between the nth and $(n+1)$th measurement time intervals.

The random walk shown in Figure 5.5.7 is an important clue to unraveling flicker noise. Figure 5.5.7 shows long-term memory and here it is shown that there is memory over several hours. Even on smaller time scales, random walk is apparent and there is a self-similar property—the hallmark of chaotic behavior. Could this random walk effect and $1/f$ noise arise from the same physical process? Most likely, but there is no accepted theory.

5.5.3: Excess Oscillator Noise: The Leeson Effect and Flicker Noise

As seen in Figures 5.5.2 to 5.5.6, oscillators have noise that increases as the offset Δf, from the mean oscillation frequency decreases. This noise has separable regions where noise varies as Δf^n, where n is an integer ranging from 0 to -5. There are transition regions between these discrete states, but there is not a region where n is a fractional number. Not all of the discrete states are observed because, presumably, either the crossover frequencies have changed order, or the frequency offset, Δf, was not low enough.

In 1966 Leeson [28] examined the effect of feedback on noise in oscillators (see Figure 5.2.1). The phase noise mechanism treated by this analysis is now called the Leeson effect. Leeson showed that white phase noise and white flicker noise (white here meaning independent of frequency) of the amplifier in the feedback loop translate to noise on the oscillation signal with power law dependencies of f^{-2}, called white frequency noise, and f^{-3}, called flicker frequency noise, respectively. These were the dominant "nonwhite" forms of noise observed in his time. However, his analysis did not predict the level of the noise accurately and sometimes was off by an order of magnitude.

The Leeson effect is briefly summarized here. First, it was observed that nearly every physical system has fluctuations that vary as $1/f$ at low frequencies. This includes electrical devices such as the amplifier in an oscillator feedback loop. This leads to equal amplitude phase and amplitude noise superimposed on the oscillation. Since noise is small, the amplitude fluctuations are suppressed by the saturation of the active device so that the only noise observed in good designs is phase noise. Leeson determined that the oscillator phase noise has a region with Δf^{-3} dependence that is due to low-frequency f^{-1} noise (i.e. around DC), a Δf^{-2} region due to white noise in the bandwidth of the oscillator's tank circuit, and also a white noise region outside the bandwidth of the tank circuit. The basis for the development of Leeson's oscillator phase noise model is shown in Figure 5.5.8. Mathematically [28],

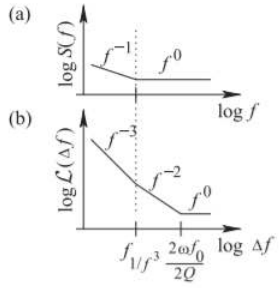


Figure 5.5.8: Derivation of oscillator noise spectra: (a) the noise spectra of an electronic material with noise increasing as the frequency reduces; and (b) the noise spectra close to the oscillation frequency of an oscillator.

$$\mathcal{L}(\Delta f) = \mathcal{L}(\Delta\omega) = \frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right] \tag{5.5.4}$$

where Q is the loaded Q factor of the oscillator's tank circuit and F is an empirical factor. $\mathcal{L}$ has the units of $\text{radians}^2/\text{Hz}$, or in decibels,

$$\mathcal{L}|_{\text{dB}}(\Delta f) = 10\log\left\{\frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right]\right\} \tag{5.5.5}$$

which has the units of $\mathrm{dB/Hz}$ or more usually expressed as "decibels below the carrier" of power P_0, or $\mathrm{dBc/Hz}$. This is the power at a specified offset such as the phase noise of a 5.05 GHz oscillator at an offset of 1 MHz and with an output power of 0 dBm being -130 dBc/Hz [19].

The derivation of the oscillator noise characteristics from first principles, which led to Equations (5.5.4) and (5.5.5), predicts noise levels that are much lower than those observed in practice [29, 30]. As well, the prediction inherent in Equation (5.5.4) is that by increasing the Q of the tank circuit the noise level will be reduced. However, this is not always obtained in practice. A further complication is that Equation (5.5.4) provides no mechanism for the generation of $1/(\Delta f)$ and $1/(\Delta f)^3$ noise in the oscillator phase noise spectrum. An adhoc modification of Equation (5.5.4) accounts for this:

$$\mathcal{L}(\Delta f) = \frac{2FkT}{P_0}\left[1+\left(\frac{f_0}{2Q\Delta f}\right)^2\right]\left(1+\frac{f_{c-3}}{|\Delta f|}\right) \tag{5.5.6}$$

Given the inadequacy of this model it is just as well to use

$$\mathcal{L}(\Delta f) = \sum_{i=0}^{-5} b_i f_i^n \tag{5.5.7}$$

where the b_i coefficients are extracted from measurements.

The Leeson effect can be stated as oscillator phase noise being upconverted white noise around DC.

5.5.4: Excess Oscillator Noise: Linear Time-Variant Model

The Leeson effect model described in the previous subsection uses a linear time-invariant model of the oscillator and does not consider down-conversion of noise from frequencies near harmonics. The linear time-variant model, also called the Hajimiri and Lee model, incorporates these higher-order conversion mechanisms [31].

Noise injected into an oscillator has a different impact depending on whether it is injected at the peak of the oscillating signal or at the zero-crossings. The noise injected at the peaks of the oscillating signal are quenched by the saturating effect of the active device in the oscillator. However, noise at or near the zero-crossings of the waveform introduces jitter and phase noise. This effect on phase noise can be described by an impulse sensitivity function [31]. Consider an impulse injected at phase $x = \omega_0 t$, then the time-domain impulse response is

$$h_\phi(t,\tau) = \frac{\Gamma(\omega_0 t)}{q_{\max}} u(t-\tau) \tag{5.5.8}$$

where $\Gamma(\,)$ is the impulse sensitivity function, $q_{\max}$ is the maximum charge displacement on the capacitor forming the tank circuit, t is the observation time, and τ is the time of the excitation. The excess phase of the oscillator (the additional phase induced onto the phase of the carrier) is

$$\phi(t) = \frac{1}{q_{\max}} \int_{-\infty}^{t} \Gamma(\omega_0 t) i(\tau) d\tau \tag{5.5.9}$$

where $i(\tau)$ is the noise current injected in the oscillator.

$\Gamma(\,)$ can be derived approximately for some oscillators such as the CMOS LC oscillator in [32], where it was shown that the impulse sensitivity function can be expressed as a Fourier series with a fundamental component corresponding to the frequency of oscillation. The excess phase at the zero-crossings of the oscillator is

$$\phi(t) = \frac{1}{q_{\max}}\left[c_0 \int_{-\infty}^{t} i(\tau) d\tau + \sum_{m=1}^{\infty} c_m \int_{-\infty}^{t} i(\tau)\cos(n\omega_0\tau) d\tau\right] \tag{5.5.10}$$

where c_m are coefficients of the Fourier series. The first term with the c_0 coefficient indicates noise that is up-converted from baseband, while the term in the summation is the contribution to the oscillator phase noise due to down-conversion of noise near the harmonic frequencies. With the assumption that the noise at the harmonics is white noise with mean-square current $\overline{i_n^2}$, then the noise spectral density is [32, 33]

$$\mathcal{L}(\Delta\omega) = 10\log\left(\frac{\bar{i}_n^2}{\Delta f}\frac{\sum_{m=0}^{\infty} c_m^2}{4q_{\max}^2\Delta\omega^2}\right) \tag{5.5.11}$$

Here $\Delta f = 1$ Hz for noise in a 1 Hz bandwidth and $\Delta\omega$ is the radian offset frequency. Thus white noise at the harmonics is down-converted to f^{-2} noise at the oscillation frequency, and f^{-1} noise at baseband is up-converted to f^{-3} noise at the oscillation frequency.

The time-variant model provides a richer description of phase noise on an oscillating signal than does the Leeson model, but it does not describe f^{-1}, or f^{-n}, $n > 3$, noise that is observed with oscillators.

Thus the Hajimiri and Lee model relates to down-conversion of white noise at harmonics of the oscillation frequency to the near carrier noise which is in addition to Leeson's model of up-converted near-DC white noise. The Hajimiri and Lee, and Leeson models of phase noise led to designers developing microwave oscillators with significantly lower phase design yet there remains appreciable near-carrier phase noise.

5.5.5: Excess Oscillator Noise: Chaotic Maps and Flicker Noise

While not firmly established, it is possible that flicker noise originates from nonlinear dynamics and chaos [25, 26, 34, 35, 36, 37, 38, 39, 40, 41]. In this model flicker noise derives from a nonlinear process with delayed feedback. The mathematical foundation is well established [42], describing a phenomenon called intermittency [43] that occurs when a physical process transitions between stable periodic states and chaotic states. Inherent to some forms of intermittency is long-term memory with a $1/f$ spectrum [44]. This has been established for many physical and biological systems.

5.5.5.1: Logistics Map

A classic example of **intermittency**, and the first widely accepted, is the following model of population dynamics. If t_n denotes discrete time and (the real number) x_n denotes the ratio of the existing population to the maximum possible population at t_n (so x_n is between 0 and 1), then what is called the logistics map provides the population ratio, $x_{(n+1)}$, at time t_{n+1}. The logistics map is [45]

$$F_\lambda(x) = \lambda x_n(1 - x_n) \tag{5.5.12}$$

and so

$$x_{n+1} = F_\lambda(x) \tag{5.5.13}$$

Here λ is a positive number representing the combined rate of reproduction and starvation. So environmental conditions determine λ, which is constrained so that $0 < \lambda \leq 4$. Depending on λ, the logistics map (i.e., Equation (5.5.12)) will produce a stable population or a random population depending on the value of λ, with $\lambda = 4$ producing white noise.

Thermal noise produces random fluctuations in the amplitude and phase of a sinusoidal signal that is being processed in a nonlinear electronic system such as an amplifier or an oscillator. Denoting the thermal amplitude fluctuations by $a_{t,I}(t)$ and the thermal phase fluctuations by $\phi_{t,I}(t)$, a sinusoidal signal with mean amplitude A and an initial phase of zero is

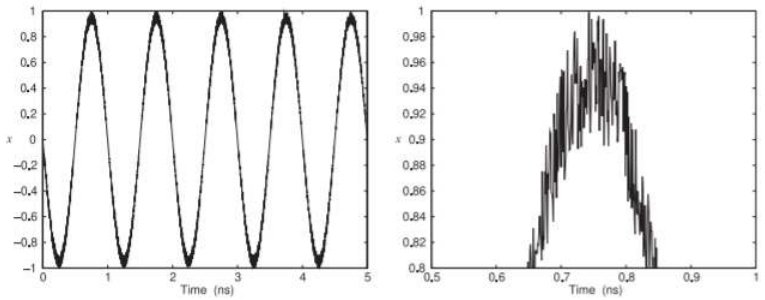
$$x(t) = A[1 + a_{t,I}(t)]\cos[\omega t + \phi_{t,I}(t)] \tag{5.5.14}$$

Using the logistics map with $\lambda = 4$ (which produces white noise) to determine $a_{t,I}(t)$ and $\phi_{t,I}(t)$, a sinusoidal signal with thermal (white) noise is as shown in Figure 5.5.9. Using $a_{t,I}(t)$ and $\phi_{t,I}(t)$ determined from a Gaussian distribution would yield the same qualitative result. Of course most of this noise would be easy to remove by bandpass filtering but thermal noise will still appear within the finite bandwidth of the signal. It is just easier to visualize the effect of noise by plotting it on this scale.

Equation (5.5.12) is a simple nonlinear equation with delayed feedback that mixes x over time. What is called the rate of mixing describes the extent of correlation to past events and can be thought of as an exponential decay rate. However, with the logistics map, $F_\lambda(x)$ in Equation (5.5.12), the rate of this mixing is not controllable.

5.5.5.2: Logarithmic Map

There are many maps that will lead to $1/(\Delta f)$ effects and one of the most convenient to use in modeling flicker noise in electronics is called the loga-



(a) Waveform (b) Detail

Figure 5.5.9: Sinusoidal signal with superimposed white noise calculated using the logistics map rather than calculating noise as a Gaussian process.

rithmic map [46, 47]:

$$F_\beta(x) = \begin{cases} x(1 + Y(\beta)x|\log(x)|^{1+\beta}) & \text{if} \quad 0 \le x \le 1/2 \\ 2x - 1 & \text{if} \quad 1/2 < x \le 1 \end{cases} \tag{5.5.15}$$

and so

$$x_{n+1} = F_\beta(x) \tag{5.5.16}$$

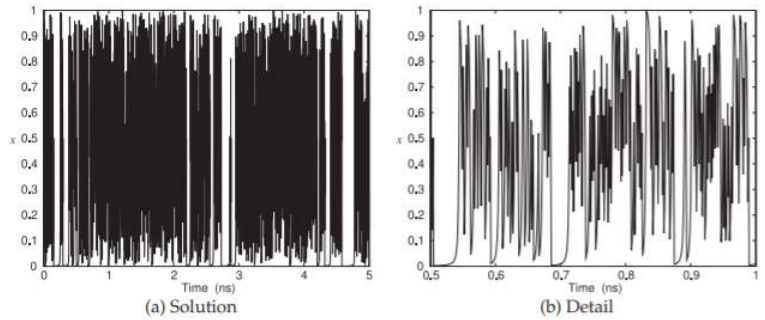
$F_\beta(x)$ is defined on the interval $0 < x \le 1$ and $Y(\beta) = 2(\log 2)^{-(1+\beta)}$ is chosen to ensure that $\lim_{x\to 1/2} -f_\beta(x) = 1$. (Note that the map is discontinuous at $x = 1/2$.) If Δt is a fixed time interval, then $x(t + \Delta t) = F_\beta(x(t))$.

The logarithmic map uses only a single parameter, β, that controls the rate of mixing and thus the long-term memory property. The solution of the logarithmic map, Equation (5.5.15), with $\beta = 0.000005$ is shown in Figure 5.5.10.

The Fourier transform of the sequence plotted in Figure 5.5.10 is shown in Figure 5.5.11 and its autocorrelation is shown in Figure 5.5.12. The spectrum in Figure 5.5.11 (above 1 Hz) has an $f^{-0.5}$ dependence, and since the sequence corresponds to voltage, squaring this yields a $1/f$ power characteristic. The autocorrelation plot in Figure 5.5.12 shows the slow long-term decay in the correlation with respect to the discrete interval, i, between the sequence points. That is, the logarithmic map describes a process with slowly decaying correlations. The extended correlation is a measure of the rate of mixing. This interpretation corresponds very well to the understanding of physical systems, and in particular to electronic systems. It can be shown [47] that the rate of decay of correlation of this map is bounded as

$$R(n) \le B(\log n)^{-\beta} \tag{5.5.17}$$

where n is the nth time interval. Thus the map is said to describe a logarithmic mixing rate that can be made as slow as desired by varying the value of β. It is this long-range dependence that produces f^{-1} (and f^{-2}, f^{-3}, etc.) noise. In an oscillator these result in phase noise with $1/(\Delta f)$, $1/(\Delta f)^2$, $1/(\Delta f)^3$, etc. characteristics. In semiconductors, for example, the



(a) Solution (b) Detail

Figure 5.5.10: Solution of the logarithmic map of Equation (5.5.15) with (a random initial seed) $x_0 = 0.477347,\ \beta = 0.000005$, and $t_{n+1} - t_n = 1$ ps.

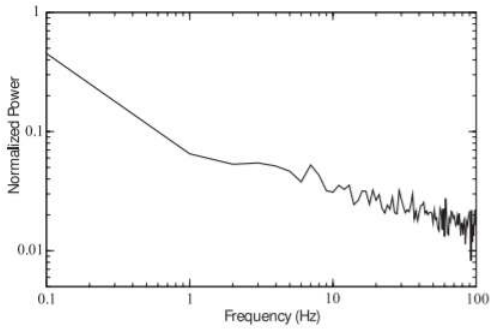


Figure 5.5.11: Spectrum of the logarithmic map with $\beta = 0.000005$.

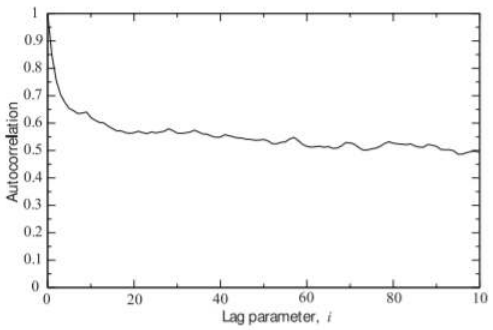


Figure 5.5.12: Correlation plot of the logarithmic map with $\beta = 0.000005$ for a million-point sequence.

rate of mixing is influenced by the density of traps [48] and lattice scattering [49] with the lower the density of traps (i.e. better quality semiconductor material) and the lower the amount of scattering, the lower the rate of mixing and hence the lower the level of flicker noise. The delayed feedback in the chaotic-map model is consistent with trapping and the observation that reducing traps improves phase noise performance.

The long-range mixing indicated by the slow decay of the correlation function (see Figure 5.5.12) is key to the f^{-1} response. Another function that yields a long-term correlation is the Ornstein–Uhlenbeck process [50, 51, 52] developed to describe Brownian motion. The autocorrelation function of this process decays exponentially and predicts f^{-2} noise but not f^{-1} noise [52]. The Ornstein–Uhlenbeck process decays too rapidly to predict the f^{-1} response. This is discussed further in [52].

So the solution of Equation (5.5.16) (the logarithmic map), shown in Figure 5.5.10, has complicated dynamics with long periods of stability with rapid transitions between stable and rapidly varying levels. The sequence of x_ns depends on the starting condition (i.e., x_0), but no matter how it starts, the power spectrum of the solution has an inverse frequency dependence (i.e., it is exactly f^{-1}). The logarithmic map, as with all chaotic maps, describes a nonlinear process with delayed feedback. This matches the situation in physical, biological, chemical, and financial systems. Since nearly every physical process can be described as a (perhaps weak) nonlinear process with delayed feedback, the widespread observation of $1/f$ fluctuations is not surprising. So the basis of $1/f$ noise is the most basic of physical processes.

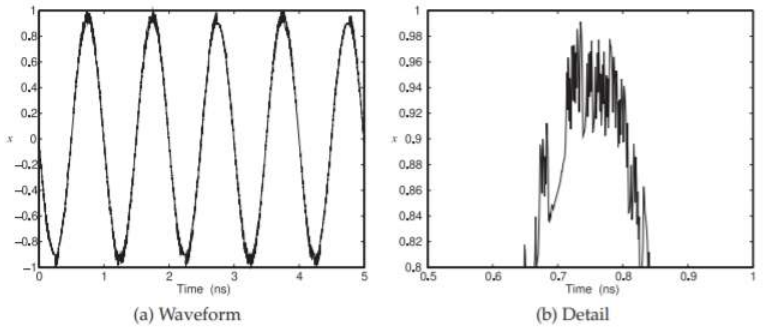
Intermittency (described by chaotic maps) results in random fluctuations in the amplitude and phase of a sinusoidal signal processed by a nonlinear electronic system such as an amplifier or oscillator. Denoting the amplitude intermittency by $a_I(t)$ and the phase intermittency by $\phi_I(t)$, a sinusoidal signal with mean amplitude A and an initial phase of zero is

$$x(t) = A[1 + a_I(t)]\cos[\omega t + \phi_I(t)] \tag{5.5.18}$$

This signal is shown in Figure 5.5.13 with the logarithmic intermittency fluctuations $a_I(t)$ and $\phi_I(t)$ as shown in Figure 5.5.10, calculated using different seeds. The effect of intermittency fluctuations is greatly exaggerated here for visualization purposes. In practice, the fluctuations at the scale shown in Figure 5.5.13 could be eliminated using a bandpass filter. However, the fluctuations are self-similar (another property of chaotic processes) and is repeated at all scales. In a bandpass electronic system the in-band amplitude fluctuations are suppressed by device nonlinearity, but the phase fluctuations appear as phase noise on an oscillator.

5.5.6: Summary

Three oscillator phase noise models were presented. The Leeson model is based on up-conversion of white noise from baseband producing noise around the oscillator carrier with an f^{-2} dependency. The Hajimiri and Lee model is based on a linear time-varying model of the oscillator with f^{-1} noise at baseband resulting in oscillator phase noise with a $1/(\Delta f)^3$ dependency, and white noise at the baseband and harmonics resulting in noise around the oscillating frequency with an $1/(\Delta f)^2$ dependency. Upconversion of noise has been demonstrated as a mechanism that describes



(a) Waveform (b) Detail

Figure 5.5.13: Sinusoidal signal with superimposed intermittency noise. Using Equation (5.5.18) with $A = 0.9, a_I(t)$ scaled to the interval $[0, 0.09]$, $a_I(0) = 0.477347$ (before scaling), $\phi_I(t)$ scaled to the interval $[0, 5]$ radians, and $\phi_I(0) = 0.00915926$ (before scaling).

some of the observed oscillator noise. Neither the Leeson nor the linear timevarying models describe the full set of observations of phase noise with $1/(\Delta f)^5, \ldots 1/(\Delta f)$ dependencies.

The chaotic map model is physically appealing and describes the origin of flicker noise as the time-delayed feedback of the output of a nonlinear process. This can produce a chaotic response called intermittency that embodies long-term memory. It has been shown through simulation that this model predicts the $1/(\Delta f)^3$, $1/(\Delta f)^2$, $1/(\Delta f)^1$, and $1/(\Delta f)^0$ dependencies of phase noise. It is also consistent with the random walk seen in time-domain observations of oscillator noise. However, the chaotic map-based model has not yet led to a compact formula for phase noise similar to Leeson's formula. The development of a compact phase noise model (e.g., like Leeson's model) will not be simple as integer calculus and transfer function-based analyses cannot be directly used with a chaotic map. However, it is clear that the description of phase noise is getting close to a satisfying physical explanation.

Phase noise can also be induced by vibrations [53, 54], and spurious signals from the environment (such as those coupled from the power mains) can also appear as phase noise.

This page titled 5.5: Advanced Discussion of Oscillator Noise is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.8: Advanced Discussion of Oscillator Noise** has no license indicated.

5.6: Summary

This chapter focused on the design of two categories of microwave oscillators: fixed-frequency oscillators and VCOs. All microwave oscillators can be considered to be an amplifier with a feedback network with the amplifier establishing the amplitude of oscillation and the feedback network setting the frequency of oscillation. As well, nearly all microwave oscillators can be considered as reflection oscillators in which the active device, with appropriate feedback, presents a negative resistance, or equivalently, a negative conductance to a linear frequency-selective circuit. Whether design based on negative conductance or negative resistance is used depends on which, resistance or conductance, reduces in magnitude as the signal level increases. Since transistors are essentially voltage-controlled current sources, and the current tends to saturate at high output levels, the natural view is to consider transistor-based oscillators as having a negative conductance that reduces in magnitude as the signal level increases. With negativeconductance reflection oscillators part or all of the feedback network appears as a linear two-terminal, that is one-port, circuit in parallel with the negative conductance from the active device. This two-terminal circuit is often called the resonator or tank circuit.

The chapter presented two case studies of oscillator design. This enabled design decisions to be illustrated that could not be presented in an algorithmic way. While the design case studies considered common-base Colpitts designs, the principles apply to other types of oscillators. Still the Colpitts common-base/common-drain oscillator core has proved to yield microwave VCOs with stellar performance. RFICs necessarily use differential circuits, but even then they are conveniently designed as negative conductance oscillators. Treating the design problem as that of interconnected one-port circuits considerably simplifies making design trade-offs, and so makes it easier to achieve an optimal VCO.

VCO design is the most complicated of microwave designs with significant trade-offs of low phase noise, stability, broad tuning range, rapid turn-on transient, high output power, and high efficiency. Several decades ago fixed-frequency oscillators were most common. With these the feedback network incorporates a high-Q resonant element that results in the phase noise on the oscillating signal being insignificant in nearly all microwave applications. The high-Q resonator of the fixed-frequency oscillators largely assures single frequency of operation, and the desired constant amplitude output is assured by ensuring the active device enters saturation. With a VCO, stability is not as easy to achieve. Stability here refers to the oscillator producing a sinewave of a single frequency and constant amplitude.

With VCOs the resonator is nearly always of relatively low Q, as it must be variable. Then a major concern in oscillator design is phase noise appearing on the output signal. Managing phase noise is complicated because the origins of phase noise are not well known and so physically based device models, that would enable the accurate determination of phase noise in computer-based RF circuit simulation, are not available. For noncompetitive VCOs there are many noise sources other than the intrinsic phase noise of an active device that dominate the phase noise of the output signal. For example, it is known that up-conversion of low-frequency noise, and down-conversion of harmonic noise, produce oscillator phase noise. However, once these sources of phase noise are minimized in design, there is a remaining intrinsic phase noise component. It is this phase noise, intrinsic to active devices, that is the concern of competitive VCO design. While the origins of intrinsic phase noise in well-designed oscillators is not completely understood, there are best practices to follow that minimize the coupling of external noise to the oscillating signal. External noise at low frequencies will be up-converted unless care is taken. Good design practice is to eliminate low-frequency noise from the power supply and from the tuning voltage source for VCOs. Some more specific guidelines:

1. Good grounding is required with decoupling capacitors between the supply and ground.
2. Attention must be given to the signal return path for the supply and tuning voltage source to avoid common impedance coupling. Any noise on the tuning voltage in a VCO will result in phase noise on the VCO output. In a microstrip circuit, only minimum metal on the microstrip layer should be removed with floating metal connected to the ground through multiple vias. This suppresses substrate modes, minimizes parasitic coupling, and minimizes electro-thermal passive intermodulation distortion.
3. The oscillator output should drive a resistive load and it is common to use a resistive pad (i.e., attenuator) followed by a bandpass filter. This reduces the effect of the load on the oscillation frequency. Also it is common to isolate the tank circuit from the load as was seen in the case studies presented in this chapter.
4. Internal oscillator paths should be as small as possible to minimize the coupling of noise from the environment.

A VCO can be incorporated in a phase-locked loop which further reduces phase-noise and locks the oscillator to a low-frequency highly stable frequency reference. The oscillation frequency is then controlled by the fractional frequency division in the phase-locked loop.

This page titled 5.6: Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.10: Summary** has no license indicated.

5.7: References

[1] M. Odyniec, Ed., *RF and Microwave Oscillator Design.* Artech House, 2002.

[2] A. Rohde, U.L. Poddar and G. Bock, *The Design of Modern Microwave Oscillators for Wireless Applications.* Wiley, 2005.

[3] E. Faulkner, *Introduction to the Theory of Linear Systems.* Chapman & Hall, 1969.

[4] A. Pippard, *Response and Stability.* Cambridge University Press, 1985.

[5] L. Dai and R. Harjani, *Design of High Performance CMOS Voltage-Controlled Oscillators.* Kluwer Academic Publishers, 2003.

[6] M. Tiebout, *Low Power VCO Design in CMOS.* Springer, 2006.

[7] A. Aktas and M. Ismail, *CMOS PLLs and VCOs for 4G wireless.* Kluwer, 2004.

[8] D. Pederson and K. Mayaram, *Analog Integrated Circuits for Communication: Principles, Simulation and Design.* Springer, 2008.

[9] B. Razavi, *Design of Analog CMOS Integrated Circuits.* McGraw-Hill, 2001.

[10] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits.* Cambridge University Press, 2004.

[11] K. Kurokawa, "Some basic characteristics of broadband negative resistance oscillator circuits," *The Bell System Technical J.*, vol. 48, no. 69, pp. 1937–1955, Jul.-Aug. 1969.

[12] G. Gonzalez, *Foundations of Oscillator Circuit Design.* Artech House, 2007.

[13] C. Rauscher, "Large-signal technique for designing single-frequency and voltage-controlled GaAs FET oscillators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 29, no. 4, pp. 293–304, Apr. 1981.

[14] R. Jackson, "Criteria for the onset of oscillation in microwave circuits," *IEEE Trans. On Microwave Theory and Techniques*, vol. 40, no. 3, pp. 566–569, Mar. 1992.

[15] D. Maclean, *Evaluating Feedback in Amplifiers and Oscillators: Theory, Design and Analogue Applications.* Research Studies Press LTD., 2004.

[16] J. Boyles, "The oscillator as a reflection amplifier: an intuitive approach to oscillator design," *Microwave Journal*, vol. 29, no. 6, pp. 83–98, Jun. 1986.

[17] D. Esdale and M. Howes, "A reflection coefficient approach to the design of one-port negative impedance oscillators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 29, no. 8, pp. 770–776, Aug. 1981.

[18] G. Gonzalez and O. Sosa, "On the design of a series-feedback network in a transistor negative-resistance oscillator," *IEEE Trans. on Microwave Theory and Techniques*, vol. 47, no. 1, pp. 42–47, Jan. 1999.

[19] A. Victor and M. Steer, "Reflection coefficient shaping of a 5-GHz voltage-tuned oscillator for improved tuning," *IEEE Trans. on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2488–2494, Dec. 2007.

[20] A. Knights and M. Kelly, "Laterally stacked varactor formed by ion implantation," *Electronics Letters*, vol. 35, no. 10, pp. 846–847, May 1999.

[21] M. Steer, *Microwave and RF Design, Modules*, 3rd ed. North Carolina State University, 2019.

[22] P. J. Topham, A. Dearn, and G. Parkinson, "GaAs bipolar broadband oscillators," in *21st European Microwave Conf.*, Sep. 1991, pp. 178–183.

[23] http //www.febo.com.

[24] A. Victor, J. Nath, D. Ghosh, B. Boyette, J. Maria, M. Steer, A. Kingon, and G. Stauf, "Noise characteristics of an oscillator with a barium strontium titanate (BST) varactor," *IEE Proc., Part H, Microwaves, Antennas and Propagation*, vol. 153, no. 1, pp. 96–102, 2006.

[25] N. Kriplani, A. Victor, and M. Steer, "Timedomain modelling of phase noise in an oscillator," in *36th European Microwave Conf.*, 2006, pp. 514–517.

[26] N. Kriplani, S. Luniya, and M. Steer, “Integrated deterministic and stochastic simulation of electronic circuits: Application to large signal–noise analysis,” *Int. Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 21, no. 6, pp. 381–394, 2008.

[27] A. Victor, “Microwave power oscillator utilizing thin-film ferroelectic varactors,” Ph.D. dissertation, North Carolina State University, 2010.

[28] D. Leeson, “A simple model of feedback oscillator noise spectrum,” *Proc. of the IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.

[29] T. Lee and A. Hajimiri, “Oscillator phase noise: a tutorial,” *IEEE J. of Solid-State Circuits*, vol. 35, no. 3, pp. 326–336, Mar. 2000.

[30] S. Kogan, *Electronic Noise and Fluctuations in Solids.* Cambridge University Press, 1996.

[31] A. Hajimiri and T. Lee, “A general theory of phase noise in electrical oscillators,” *IEEE J. of Solid-State Circuits*, vol. 33, no. 2, pp. 179– 194, Feb. 1998.

[32] ——, “Design issues in cmos differential lc oscillators,” *IEEE J. of Solid-State Circuits*, vol. 34, no. 5, pp. 717–724, May 1999.

[33] ——, “Corrections to ”a general theory of phase noise in electrical oscillators”,” *IEEE J. of Solid-State Circuits*, vol. 33, no. 6, p. 928, Jun. 1998.

[34] G. Wornell and A. Oppenheim, *Signal Processing with Fractals: A Wavelet-Based Approach.* Prentice Hall, 1996.

[35] G. Wornell, “Wavelet-based representations for the $1/f$ family of fractal processes,” *Proc. of the IEEE*, vol. 81, no. 10, pp. 1428–1450, Oct. 1993.

[36] N. Kriplani, “Modelling colored noise under large-signal conditions,” Ph.D. dissertation, North Carolina State University, 2005.

[37] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, and M. Brammer, “Colored noise and computational inference in neurophysiological (fmri) time series analysis: resampling methods in time and wavelet domains,” *Human Brain Mapping*, vol. 12, no. 2, pp. 61–78, 2000.

[38] B. Mandelbrot, *The Fractal Geometry of Nature.* Times Books, 1982.

[39] B. Mandelbrot and J. Van Ness, “Fractional brownian motions, fractional noises and applications,” *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[40] M. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* Dover Publications, 2009.

[41] R. Voss, *Fractals in Nature: From Characterization to Simulation.* Springer-Verlag, 1988.

[42] R. Devaney, *A First Course in Chaotic Dynamical Systems.* Perseus Book, 1992.

[43] Y. Pomeau and P. Manneville, “Intermittent transition to turbulence in dissipative dynamical systems,” *Commun. Math. Phys*, vol. 74, no. 2, pp. 189–197, 1980.

[44] R. Bhansali, M. Holland, and P. Kokoszka, “Chaotic maps with slowly decaying correlations and intermittency,” in *Fields Institute Communications: Asymptotic Methods in Stochastics*, L. Horv´ath and B. Szyszkowicz, Eds., 2004, p. •.

[45] R. May, “Simple mathematical models with very complicated dynamics,” *Nature*, vol. 261, no. 5560, pp. 459–467, June 1976.

[46] T. Kawabe and Y. Kondo, “Intermittent chaos generated by logarithmic map,” *Progress of Theoretical Physics*, vol. 86, no. 3, pp. 581–586, 1991.

[47] M. Holland, “Slowly mixing systems and intermittency maps,” *Ergodic Theory and Dynamical System*, vol. 25, no. 1, pp. 133–159, Feb. 2005.

[48] E. Simoen, A. Mercha, L. Pantisano, C. Claeys, and E. Young, “Low-frequency noise behavior of SiO2–HfO2 dual-layer gate dielectric nMOSFETs with different interfacial oxide thickness,” *IEEE Trans. on Electron Devices*, vol. 51, no. 5, pp. 780–784, May 2004.

[49] F. Hooge, T. Kleinpenning, and L. Vandamme, "•experimental studies on $1/f$ noise," *Reports on Progress in Physics*, 1981.

[50] G. Uhlenbeck and L. Ornstein, "On the theory of brownian motion," *Physical Review*, vol. 36, pp. 823–841, Sep. 1930.

[51] E. Bibbona, G. Panfilo, and P. Tavella, "The ornstein-uhlenbeck process as a model of a low pass filtered white noise," *Metrologia*, vol. 45, no. 6, pp. S117–S126, Dec. 2008.

[52] A. Suarez, *Analysis and Design of Autonomous Microwave Circuits.* John Wiley & Sons, Inc., 2009.

[53] A. Hati, C. Nelson, and D. Howe, "Effect of vibration on P and AM noise of oscillatory and non-oscillatory components at 10 GHz," in *2009 IEEE Int. Frequency Control Symp., Joint with the 22nd European Frequency and Time Forum*, Apr. 2009, pp. 524–529.

[54] W. Robins, *Phase Noise in Signal Sources.* Peter Peregrinus Ltd., 1982.

[55] M. Buckingham, *Noise in Electronic Devices and Systems.* Ellis Horwood, 1983.

[56] P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE.* McGrawHill, 1988.

[57] A. van der Ziel, X. Zhang, and A. Pawlikiewicz, "Location of $1/f$ noise sources in BJT's and HBJT's—i. theory," *IEEE Trans. on Electron Devices*, vol. 33, no. 9, pp. 1371–1376, Sep. 1986.

[58] C. Green and B. Jones, "$1/f$ noise in bipolar transistors," *J. Phys. D: Appl. Phys.*, vol. 18, pp. 77–91, 1985.

This page titled 5.7: References is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.11: References** has no license indicated.

5.8: Exercises

1. Draw the schematic of an op-amp-based Hartley oscillator circuit.
2. Consider the circuit below. This is the equivalent circuit of an active device with a negative conductance $= 1/R_D$ connected to a tank circuit comprising capacitor C_T and inductor L_T, and then a load R_L. The oscillation amplitude will adjust so that $R_D = R_L$.

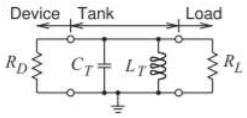


Figure 5.8.1

   a. Write down a formula for the oscillation frequency f_0.
   b. What is f_0 if $C_T = 1$ pF and $L_T = 1$ nH?

3. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit $C_T = 0.1$ pF and $L_T = 0.5$ nH. What is the oscillation frequency assuming that there is sufficient negative conductance for oscillation to occur?

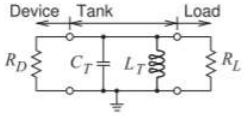


Figure 5.8.2

4. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor C_L in parallel with a resistor R_L. The oscillation amplitude will adjust so that $R_D = R_L$.

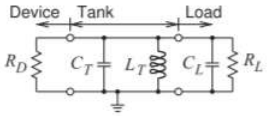


Figure 5.8.3

   a. Write down a symbolic formula for the oscillation frequency f_0.
   b. What is f_0 if $C_L = 0.2$ pF and $R_L = 50\ \Omega$?

5. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit $C_T = 0.1$ pF and $L_T = 0.5$ nH. The capacitance of the load is $C_L = 0.05$ pF and the load resistance is $50\ \Omega$. What is the oscillation frequency assuming that there is sufficient negative conductance for oscillation to occur?

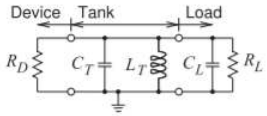


Figure 5.8.4

6. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor $C_L = 0.2$ pF in parallel with a resistor $R_L = 50\ \Omega$. Between the tank circuit is a 3 dB $50\ \Omega$ attenuator (i.e., the system is designed for system impedance of $50\ \Omega$). The oscillation amplitude will adjust so that $R_D = R_L$.

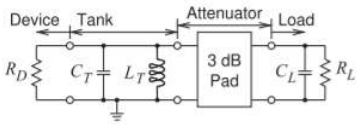


Figure 5.8.5

a. Ignore the load capacitor C_L, what is the frequency of oscillation, f_0, of the oscillator?
b. From now on consider the load capacitance. What is the oscillation frequency without the attenuator?
c. At the frequency calculated in (a), what is the admittance looking into the attenuator from the tank circuit with the load comprising C_L and R_L? [Hint use the resistive PI network equivalent of the attenuator.]
d. What is the equivalent shunt resistor and capacitor circuit looking into the attenuator from the oscillator? This is the effective load seen by the active device and tank circuit.
e. What is the oscillation frequency with the attenuator and the RC load?
f. Using your results above, discuss the effect of the attenuator on reducing the sensitivity of oscillation on the loading conditions.

7. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor $C_L = 0.2$ pF in parallel with a resistor $R_L = 50\ \Omega$. Between the tank circuit is a 3 dB 50 Ω attenuator (ti.e., the system is designed for system impedance of 50 Ω) and a bandpass filter with an insertion loss at the oscillation frequency of 2 dB. The oscillation amplitude will adjust so that $R_D = R_L$.

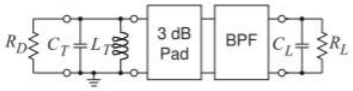


Figure 5.8.6

a. Ignore the load capacitor C_L, what is the frequency of oscillation, f_0, of the oscillator?
b. From now on consider the load capacitance. What is the oscillation frequency without the attenuator and bandpass filter?
c. At the frequency calculated in (a), what is the admittance looking into the attenuator from the tank circuit with the bandpass filter and the load comprising C_L and R_L? [Hint: Consider that the effect of the insertion loss of the filter is an attenuation. Consider using a resistive PI network equivalent of the attenuator plus bandpass filter.]
d. What is the equivalent shunt resistor and capacitor circuit looking into the attenuator for the oscillator? This is the effective load seen by the active device and tank circuit.
e. What is the oscillation frequency with the attenuator and the RC load?

8. A reflection oscillator is shown below along with the Γ_r and Γ_d loci plotted on a Smith chart. Γ_r, the reflection coefficient looking into the resonator, rotates clockwise as frequency increases. Γ_d, the reflection coefficient of the active device, is amplitude dependent but is frequency independent. The arrowed Γ_d curve plots the loci of Γ_d as amplitude increases. [Hint: Consider Figure 5.5.4.]

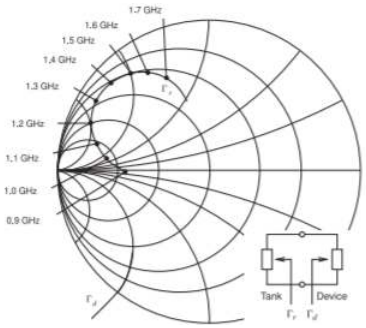


Figure 5.8.7

a. On the Γ_r plane show the region of the Smith Chart identifying oscillation.

b. On the Γ_d plane show the region of the Smith Chart identifying oscillation.
c. What is the frequency of oscillation?

9. Consider an oscillator that can be modeled as shunt connections of a linear conductance, G_r, a linear susceptance, B_r, a nonlinear or device conductance, G_d, and a nonlinear or device susceptance, B_d.
   a. Draw the circuit.
   b. Using $G_r,\ B_r,\ G_d$, and B_d, write down the Kurokawa oscillator condition that establishes stable single frequency oscillation.
   c. What are common design choices $G_r,\ B_r,\ G_d$, and B_d made that simplify the Kurokawa oscillator condition? Write down the resulting simplified Kurokawa oscillation condition.
10. An oscillator has a linear conductance, G_r, a linear susceptance, B_r, a nonlinear or device conductance, G_d, and a nonlinear or device susceptance, B_d, all in shunt. The reflection coefficient looking into the linear network is Γ_r and the reflection coefficient looking into the device is Γ_d.
    a. Use a Smith chart sketch to show the loci of the linear and nonlinear admittances (or equivalently their reflection coefficients Γ_r and Γ_d) for the case when loss in the linear network is low (i.e. $G_r \approx 0$). Do this when the oscillating signal is small and when it has reached steady state. Indicate the steady-state oscillation point. (You can use the negative or the inverse of either Γ_r or Γ_d as appropriate.)
    b. Use a Smith chart sketch to show the loci of the linear and nonlinear admittances (or equivalently their reflection coefficients for the case when loss in the linear network is high (i.e. G_r cannot be ignored). Do this when the oscillating signal is small and when it has reached steady state. (You can use the negative of the inverse of either Γ_r or Γ_d as appropriate.)
11. The equivalent circuit of an oscillator is shown below with $R_L = 50\,\Omega$, $C_T = 1$ pF and inductance $L_T = 0.1$ nH.

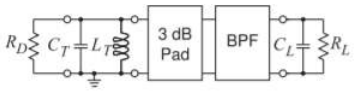


Figure 5.8.8

    a. What is the oscillation frequency assuming that there is sufficient negative conductance for oscillation to occur?
    b. What is R_D when there is oscillation?
12. A reflection oscillator is shown below along with the Γ_r and Γ_d loci plotted on a Smith chart. Γ_r, the reflection coefficient looking into the resonator, rotates clockwise as frequency increases. Γ_d, the reflection coefficient of the active device, is amplitude dependent but is frequency independent. The arrowed curve plots the loci of Γd as amplitude increases. [Hint: Consider Figure 5.5.4.]

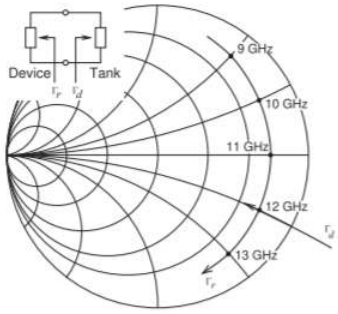


Figure 5.8.9

    a. On the Γ_r plane show the region of the Smith Chart identifying oscillation.
    b. On the Γ_d plane show the region of the Smith Chart identifying oscillation.
    c. What is the frequency of oscillation?
13. The case study presented in Section 5.6 described the design of a 5 GHz VCO. The output of the oscillator was followed by a resistive Pi attenuator and then a bandpass filter. What is the attenuation (in decibels) and the system impedance of the attenuator. [Hint: You will need to carefully read Section 5.6.]

14. Consider a common emitter BJT Clapp oscillator.
    a. Draw the schematic of the oscillator without the bias circuit.
    b. Redraw the oscillator circuit including the bias circuit.
15. Consider a BJT Hartley oscillator.
    a. Draw the schematic of the oscillator circuit in the common-base configuration. Do not show the bias circuit.
    b. Redraw the schematic of the oscillator circuit in the common-base configuration, this time showing the bias circuit.
16. Consider a FET Clapp oscillator.
    a. Draw the schematic of the oscillator circuit in the common drain configuration. Do not show the bias circuit.
    b. Redraw the schematic of the oscillator circuit, this time showing the bias circuit.
17. Consider a FET Hartley oscillator.
    a. Draw the schematic of the oscillator circuit in the common source configuration. Do not show the bias circuit.
    b. Redraw the schematic of the oscillator circuit, this time showing the bias circuit.
18. Consider a common source FET Clapp oscillator.
    a. Draw the schematic of the oscillator without biasing.
    b. Redraw the oscillator circuit including bias current sources.
19. Derive an expression for the oscillation frequency of the Colpitts BJT oscillator in the common emitter configuration shown in Figure 5.8.1(b).
20. A two-port feedback oscillator is shown in Figure 5.2.1.
    a. Draw the schematic of a feedback Colpitts oscillator.
    b. Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are $C_1 = C_2 = 2$ pF and $L_3 = 5$ nH. Ignore any phase shift introduced by the amplifier.
21. A two-port feedback oscillator is shown in Figure 5.2.1.
    a. Draw the schematic of a feedback Colpitts oscillator.
    b. Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are $C_1 = 1$ pF, $C_2 = 3$ pF, and $L_3 = 1$ nH. Ignore any phase shift introduced by the amplifier.
22. A two-port feedback oscillator is shown in Figure 5.2.1.
    a. Draw the schematic of a feedback Colpitts oscillator.
    b. Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are $C_1 = 5$ pF, $C_2 = 1$ pF, and $L_3 = 10$ nH? Ignore any phase shift introduced by the amplifier.
23. A negative-gm differential FET VCO, as shown in Figure 5.7.1, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a 1 kΩ load. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate. What is the oscillation frequency of the oscillator?
24. A negative-gm differential FET VCO, as shown in Figure 5.7.1, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a $50\,\Omega$ differential load that is in parallel with a 0.5 pF capacitor. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate.
    a. Draw the schematic of the oscillator with the load.
    b. Draw the equivalent tank circuit of the oscillator. This will need to include the effect of the load.
    c. What is the oscillation frequency of oscillator?
25. A negative-gm differential FET VCO, as shown in Figure 5.7.1, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a $50\,\Omega$ differential load that is in

parallel with a 1 pF capacitor. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate. What is the oscillation frequency of the oscillator?

26. A common emitter Colpitts oscillator, as shown in Figure 5.8.1(a), has $C_1 = 1$ pF, $C_2 = 2$ pF, and $L_3 = 2$ nH. L_C is a choke inductor. $V_{CC} = 5$ V and the circuit is biased so that $g_m = 1$ mS. Ignore the internal parasitics of the transistor. What is the oscillation frequency of the oscillator? [Parallels Example 5.7.1]
27. A common emitter Colpitts oscillator, as shown in Figure 5.8.1(a), has $C_1 = 0.1$ pF, $C_2 = 0.2$ pF, and $L_3 = 0.5$ nH. L_C is a choke inductor. $V_{CC} = 5$ V and the circuit is biased so that $g_m = 1$ mS. Ignore the internal parasitics of the transistor. What is the oscillation frequency of the oscillator? [Parallels Example 5.7.1]
28. A digital communication system has a symbol rate of 1 MS/s.
    a. What is the highest frequency phase noise that will affect the bit error rate of the communication system? (It could be 1 MHz, 2 MHZ, no limit, 0.5 MHz, etc.)
    b. What changes would you make to the system (e.g., added components such as an attenuator, filter, amplifier, etc.) that will reduce the impact of high-offset phase noise?
29. A QPSK communication system has a transmitted bit rate of 100 kbit/s. Consider that the QPSK modulation scheme is ideal.
    a. What is the symbol rate?
    b. What would you do to the system (e.g., added components) to reduce the impact of high-offset phase noise?
30. A QPSK communication system has a transmitted bit rate of 1 Mbit/s. The QPSK modulation scheme is ideal (2 bit/s/Hz. DSP processing is such that phase noise slower than the duration of 5 symbols has no affect on the communication system throughput.
    a. What is the symbol rate?
    b. What is the lowest frequency phase noise that will affect the communication system?
31. An oscillator has phase noise that reduces away from the oscillation center frequency, f_{osc} at the rate of $1/\Delta f^2$ where Δf is the frequency offset from f_{osc}. If the phase noise at 100 MHz frequency offset from the oscillation frequency is -100 dBc/Hz, what is the phase noise at 10 MHz?
32. The phase noise of an oscillator measured at 1 MHz is -136 dBc/Hz. If the phase noise varies as $1/(\Delta f)$, where Δf is the offset from the center frequency of oscillation, what is the phase noise at 100 kHz offset?
33. A phase-locked microwave oscillator typically utilizes a low-Q oscillator. For such an oscillator the phase noise at the frequency that affects microwave systems has an inverse square relationship to frequency. The phase noise measured at 100 kHz is -106 dBc/Hz, what is the phase noise referred to 1 MHz?
34. A phase-locked microwave oscillator typically utilizes a low-Q oscillator. For such an oscillator the phase noise at the frequency that affects microwave systems often has an inverse square relationship to frequency. The phase noise measured at 1 MHz is -125 dBc/Hz, what is the phase noise at 100 kHz?

5.8.1: Exercises by Section

†challenging, ‡very challenging

§5.2 1

§5.3 2†, 3†, 4†, 5†, 6‡, 7‡, 8, 9, 10, 11

§5.5 12

§5.6 13†

§5.7 14†, 15†, 16†, 17†, 18†, 19†, 20†, 21†, 22†, 23†, 24†, 25†, 26†, 27†

§5.8 28†, 29†, 30†, 31, 32†, 33†, 34†

5.8.2: Answers to Selected Exercises

4. 4.594 GHz
6. (d) $49.8\,\Omega,\ 99.6$ fF
7. (e) 4.883 MHz

13. $49.8\ \Omega,\ 2.97\ \text{dB}$

19. $\sqrt{\frac{C_1+C_2}{L_3C_1C_2}}$
20. $1.592\ \text{GHz}$

23. $25.16\ \text{GHz}$
24. (c) $10.27\ \text{GHz}$

30. $100\ \text{GHz}$

34. $-105\ \text{dBc/Hz}$

This page titled 5.8: Exercises is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **5.12: Exercises** has no license indicated.

CHAPTER OVERVIEW

6: Antennas and the RF Link

6.1: Introduction
6.2: RF Antennas
6.3: Resonant Antennas
6.4: Traveling-Wave Antennas
6.5: Antenna Parameters
6.6: The RF Link
6.7: Summary
6.8: References
6.9: Exercises
6.10: Multipaths and Delay Spread
6.11: 4.8 Radio Link Interference
6.12: Antenna Array

This page titled 6: Antennas and the RF Link is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

6.1: Introduction

An antenna interfaces circuits with free-space with a transmit antenna converting a guided wave signal on a transmission line to an electromagnetic (EM) wave propagating in free space, while a receive antenna is a transducer that converts a free-space EM wave to a guided wave on a transmission line and eventually to a receiver circuit. Together the transmit and receive antennas are part of the RF link. The RF link is the path between the output of the transmitter circuit and the input of the receiver circuit (see Figure 6.1.1). Usually this path includes the cable from the transmitter to the transmit antenna, the transmit antenna itself, the propagation path, the receive antenna, and the transmission line connecting the receive antenna to the receiver circuit. The received signal is much smaller than the transmitted signal. The overwhelming majority of the loss is from the propagation path as the EM signal spreads out, and usually diffracts, reflects, and is partially blocked by objects such as hills and buildings.

The first half of this chapter is concerned with the properties of antennas. One of the characteristics of antennas is that the energy can be focused in a particular direction, a phenomenon captured by the concept of antenna gain, which can partially compensate for path loss. The second half of this chapter considers modeling the RF link and the geographical arrangement of antennas that manage interference from other radios while providing support for as many users as possible.

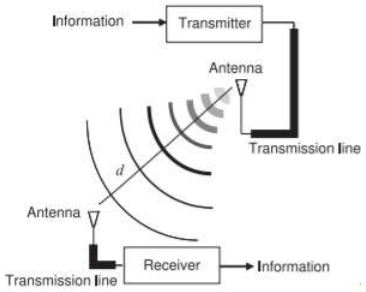


Figure 6.1.1: RF link.

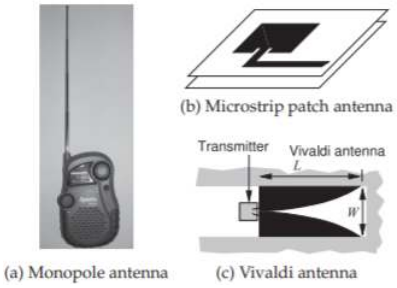


Figure 6.1.2: Representative resonant, (a) and (b), and traveling-wave, (c), antennas.

Example 6.1.1: Interference

In the figure there are two transmitters, Tx_1 and Tx_2, operating at the same power level, and one receiver, Rx. Tx_1 is an intentional transmitter and its signal is intended to be received at Rx. Tx_1 is separated from Rx by $D_1 = 2\ \mathrm{km}$. Tx_2 uses the same frequency channel as Tx_1, and as far as RX is concerned it transmits an interfering signal. Assume that the antennas are omnidirectional (i.e., they transmit and receive signals equally in all directions) and that the transmitted power density drops off as $1/d^2$, where d is the distance from the transmitter. Calculate the signal-to-interference ratio (SIR) at Rx.

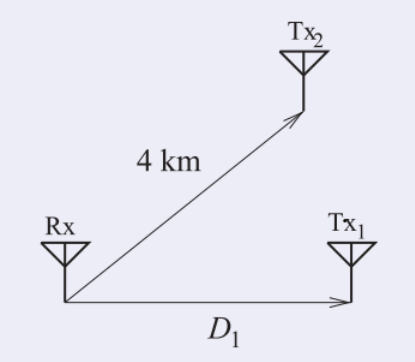


Figure 6.1.3

Solution

$D_1 = 2$ km and $D_2 = 4$ km.

P_1 is the signal power transmitted by Tx_1 and received at Rx.

P_2 is the interference power transmitted by Tx_2 and received at Rx.

So SIR $= \frac{P_1}{P_2} = \left(\frac{D_2}{D_1}\right)^2 = 4 = 6.02$ dB.

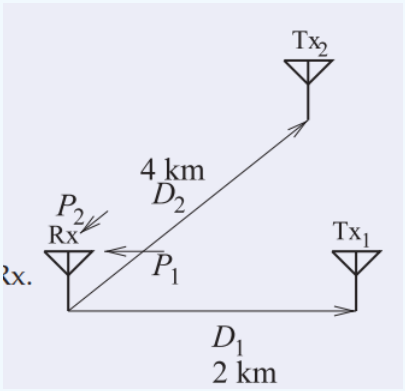


Figure 6.1.4

This page titled 6.1: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.1: Introduction** by Michael Steer is licensed CC BY-NC 4.0.

6.2: RF Antennas

Antennas are of two fundamental types: **resonant** and **traveling-wave antennas**, see Figure 4.1.2. Resonant antennas establish a standing wave of current with required resonance usually established when the antenna section is either a quarter- or half-wavelength long. These antennas are also known as standing-wave antennas. Resonant antennas are inherently narrowband because of the resonance required to establish a large standing wave of current. Figures 4.1.2(a and b) show two representative resonant antennas. The physics of the operation of a resonant antenna is understood by considering the time domain. First consider the physical operation of a transmit antenna. When a sinusoidal voltage is applied to the conductor

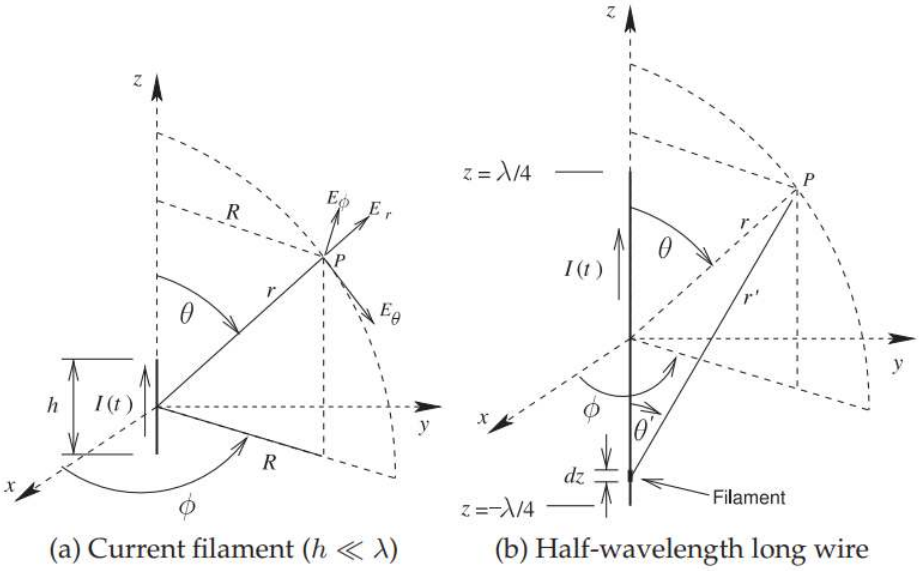


Figure 6.2.1: Wire antennas. The distance from the center of an antenna to the field point P is r. $R = r \sin\theta$.

of an antenna, charges, i.e. free electrons, accelerate or decelerate under the influence of an applied voltage source which typically arrives at the antenna from a traveling wave voltage on a transmission line. When the charges accelerate (or decelerate) they produce an EM field which radiates away from the antenna. At a point on the antenna there is a current sinusoidally varying in time, and the net acceleration of the charges (in charge per second per second) is also sinusoidal with an amplitude that is directly proportional to the amplitude of the current sinewave. Hence having a large standing wave of current, when the antenna resonates sinusoidally, results in a large charge acceleration and hence large radiated fields.

When an EM field impinges upon a conductor the field causes charges to accelerate and hence induce a voltage that propagates on a transmission line connected to the receive antenna. Traveling-wave antennas, an example of one is shown in Figures 4.1.2(c), operate as extended lines that gradually flare out so that a traveling wave on the original transmission line transitions into free space. Traveling-wave antennas tend to be two or more wavelengths long at the lowest frequency of operation. While relatively long, they are broadband, many 3 or more octaves wide (e.g., extending from 500 MHz to 4 GHz or more). These antennas are sometimes referred to as **aperture antennas**.

This page titled 6.2: RF Antennas is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.2: RF Antennas** by Michael Steer is licensed CC BY-NC 4.0.

6.3: Resonant Antennas

With a resonant antenna the current on the antenna is directly related to the amplitude of the radiated EM field. Resonance ensures that the standing wave current on the antenna is high.

6.3.1: Radiation from a Current Filament

The fields radiated by a resonant antenna are most conveniently calculated by considering the distribution of current on the antenna. The analysis begins by considering a short filament of current, see Figure 4.2.1(a). Considering the sinusoidal steady state at radian frequency ω, the current on the filament with phase χ is $I(t) = |I_0|\cos(\omega t + \chi)$, so that $I_0 = |I_0|e^{-j\chi}$ is the phasor of the current on the filament. The length of the filament is h, but it has no other dimensions, that is, it is considered to be infinitely thin.

Resonant antennas are conveniently modeled as being made up of an array of current filaments with spacings and lengths being a tiny fraction of a wavelength. Wire antennas are even simpler and can be considered to be a line of current filaments. Ramo, Whinnery, and Van Duzer [1] calculated the spherical EM fields at the point P with the spherical coordinates (ϕ, θ, r) generated by the z-directed current filament centered at the origin in Figure 4.2.1. The total EM field components in phasor form are

$$H_\phi = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\beta}{r} + \frac{1}{r^2}\right)\sin\theta, \quad \overline{H_\phi} = H_\phi\hat{\phi} \tag{6.3.1}$$

$$E_r = \frac{I_0 h}{r\pi} e^{-j\beta r}\left(\frac{2\eta}{r^2} + \frac{2}{j\omega\varepsilon_0 r^3}\right)\cos\theta, \quad \overline{E_r} = E_r\hat{\mathbf{r}} \tag{6.3.2}$$

$$E_\theta = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\omega\mu_0}{r} + \frac{1}{j\omega\varepsilon_0 r^3} + \frac{\eta}{r^2}\right)\sin\theta, \quad \overline{E_\theta} = E_\theta\hat{\theta} \tag{6.3.3}$$

where η is the free-space characteristic impedance, ε_0 is the permittivity of free space, and μ_0 is the permeability of free space. The variable k is called the **wavenumber** and $k = 2\pi/\lambda = \omega\sqrt{\mu_0\varepsilon_0}$. The $e^{-j\beta r}$ terms describe the variation of the phase of the field as the field propagates away from the filament. Equations (6.3.1)-(6.3.3) are the complete fields with the $1/r^2$ and $1/r^3$ dependence describing the near-field components. In the far field, i.e. $r \gg \lambda$, the components with $1/r^2$ and $1/r^3$ dependence become negligible and the field components left are the propagating components H_ϕ and E_θ:

$$H_\phi = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\beta}{r}\right)\sin\theta, \quad E_r = 0, \quad \text{and} \quad E_\theta = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\omega\mu_0}{r}\right)\sin\theta \tag{6.3.4}$$

Now consider the fields in the plane normal to the filament, that is, with $\theta = \pi/2$ radians so that $\sin\theta = 1$. The fields are now

$$H_\phi = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\beta}{r}\right) \quad \text{and} \quad E_\theta = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\omega\mu_0}{r}\right) \tag{6.3.5}$$

and the wave impedance is

$$\eta = \frac{E_\theta}{H_\phi} = \frac{I_0 h}{4\pi} e^{-j\beta r}\frac{j\omega\mu_0}{r}\left(\frac{I_0 h}{4\pi} e^{-j\beta r}\frac{j\beta}{r}\right)^{-1} = \frac{\omega\mu_0}{\beta} \tag{6.3.6}$$

Note that the strength of the fields is directly proportional to the magnitude of the current. This proves to be very useful in understanding spurious radiation from microwave structures. Now $\beta = \omega\sqrt{\mu_0\epsilon_0}$, so

$$\eta = \frac{\omega\mu_0}{\omega\sqrt{\mu_0\epsilon_0}} = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377\ \Omega \tag{6.3.7}$$

as expected. Thus an antenna can be viewed as having the inherent function of an impedance transformer converting from the lower characteristic impedance of a transmission line (often $50\ \Omega$) to the $377\ \Omega$ characteristic impedance of free space. Further comments can be made about the propagating fields (Equation (6.3.4)). The EM field propagates in all directions except not directly in line with the filament. For fixed r, the amplitude of the propagating field increases sinusoidally with respect to θ until it is maximum in the direction normal to the filament.

The power radiated is obtained using the **Poynting vector**, which is the cross-product of the propagating electric and magnetic fields. From this the time-average propagating power density is (with the SI units of $\mathrm{W/m^2}$)

$$P_R = \frac{1}{2}\Re(E_\theta H_\phi^* = \frac{\eta\beta^2|I_0|^2h^2}{32\pi^2r^2}\sin^2\theta \tag{6.3.8}$$

and the power density is proportional to $1/r^2$. In Equation (6.3.8) $\Re(\cdots)$ indicates that the real part is taken.

6.3.2: Finite-Length Wire Antennas

The EM wave launched by a wire antenna of finite length is obtained by considering the wire as being made up of many filaments and the field is then the superposition of the fields from each filament. As an example, consider the antenna in Figure 4.2.1(b) where the wire is half a wavelength long. As a good approximation the current on the wire is a standing wave and the current on the wire is in phase so that the current phasor is

$$I(z) = I_0 \cos(kz) \tag{6.3.9}$$

From Equation (6.3.4) and referring to Figure 4.2.1 the fields in the far field are

$$H_\phi = \int_{-\lambda/4}^{\lambda/4} \frac{I_0 \cos(kz)}{4\pi} e^{-\jmath kr'} \left(\frac{\jmath k}{r'}\right) \sin\theta' dz \tag{6.3.10}$$

$$E_\theta = \int_{-\lambda/4}^{\lambda/4} \frac{I_0 \cos(kz)}{4\pi} e^{-\jmath kr'} \left(\frac{\jmath\omega\mu_0}{r'}\right) \sin\theta' dz \tag{6.3.11}$$

where θ' is the angle from the filament to the point P. Now $k = 2\pi/\lambda$ and at the ends of the wire $z = \pm\lambda/4$ where $\cos(kz) = \cos(\pm\pi/2) = 0$. Evaluating the equations is analytically involved and will not be done here. The net result is that the fields are further concentrated in the plane normal to the wire. At large r, of at least several wavelengths distant from the antenna, only the field components decreasing as $1/r$ are significant. At large r the phase differences of the contributions from the filaments is significant and results in shaping of the fields. The geometry to be used in calculating the far field is shown in Figure 6.3.1(a). The phase contribution of each filament, relative to that at $z = 0$, is $(kz\sin\theta)/\lambda$ and Equations (6.3.10) and (6.3.11) become

$$H_\phi = I_0 \left(\frac{\jmath j}{4\pi r}\right) \sin(\theta) e^{-\jmath kr} \int_{-\lambda/4}^{\lambda/4} \frac{\sin(kz)}{4\pi} \sin(z\sin(\theta)) dz \tag{6.3.12}$$

$$E_\theta = I_0 \left(\frac{\jmath\omega\mu_0}{4\pi r}\right) \sin(\theta)) e^{-\jmath kr} \int_{-\lambda/4}^{\lambda/4} \sin(kz) \sin(z\sin(\theta)) dz \tag{6.3.13}$$

Figure 6.3.1(b) is a plot of the near-field electric field in the $y-z$ plane calculating E_r and E_θ (recall that $E_\phi = 0$) every 90°. Further from the antenna the E_r component rapidly reduces in size, and E_θ dominates.

A summary of the implications of the above equations are, first, that the strength of the radiated electric and magnetic fields are proportional to the

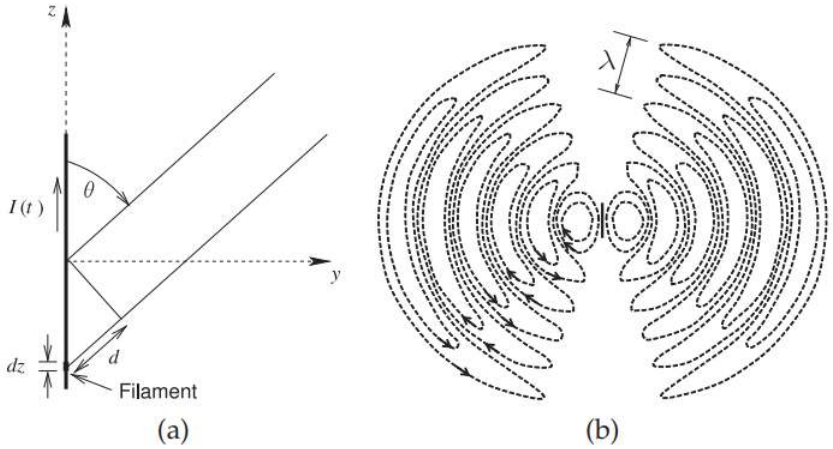


Figure 6.3.1: Wire antenna: (a) geometry for calculating contributions from current filaments of length dz with coordinate z ($d = -z\sin\theta$); and (b) instantaneous electric field in the $y-z$ plane due to a $\lambda/2$ long current element. There is also a magnetic field.

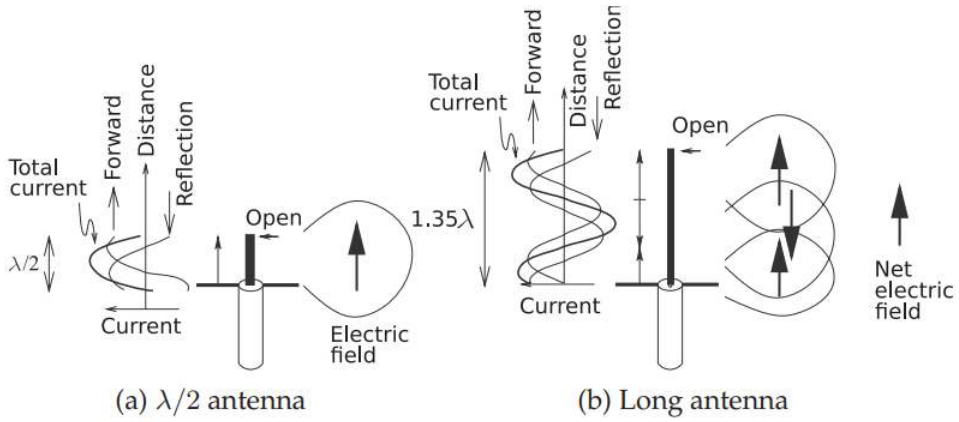


Figure 6.3.2: Monopole antenna showing total current and forwardand backward-traveling currents: (a) a $\frac{1}{2}\lambda$-long antenna; and (b) a relatively long antenna.

current on the wire antenna. So establishing a standing current wave and hence magnifying the current is important to the efficiency of a wire antenna. A second result is that the power density of freely propagating EM fields in the far field is proportional to $1/r^2$, where r is the distance from the antenna. A third interpretation is that the longer the antenna, the flatter the radiated transmission profile; that is, the radiated energy is more tightly confined to the $x-y$ (i.e. $\Theta = 0$) plane. For the wire antenna the peak radiated field is in the plane normal to the antenna, and thus the wire antenna is generally oriented vertically so that transmission is in the plane of the earth and power is not radiated unnecessarily into the ground or into the sky.

To obtain an efficient resonant antenna, all of the current should be pointed in the same direction at a particular time. One way of achieving this is to establish a standing wave, as shown in Figure 6.3.2(a). At the open-circuited end, the current reflects so that the total current at the end of the wire is zero. The initial and reflected current waves combine to create a standing wave. Provided that the antenna is sufficiently short, all of the total current— the standing wave—is pointed in the same direction. The optimum length is about a half wavelength. If the wire is longer, the contributions to the field from the oppositely directed current segments cancel (see Figure 6.3.2(b)).

In Figure 6.3.2(a) a coaxial cable is attached to the monopole antenna below the ground plane and often a series capacitor between the cable and the antenna provides a low level of coupling leading to a larger standing wave. The capacitor also approximately matches the characteristic impedance of the cable to the input impedance Z_{in} of the antenna. If the length of the monopole is reduced to one-quarter wavelength long it is again resonant, and the input impedance, Z_{in}, is found to be $36\ \Omega$. Then a $50\ \Omega$ cable can

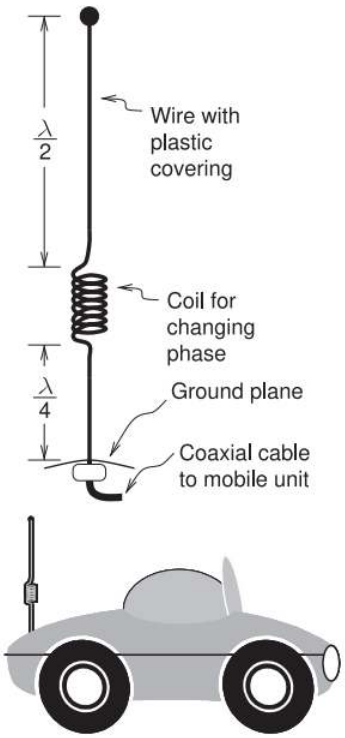


Figure 6.3.3: Mobile antenna with phasing coil extending the effective length of the antenna.

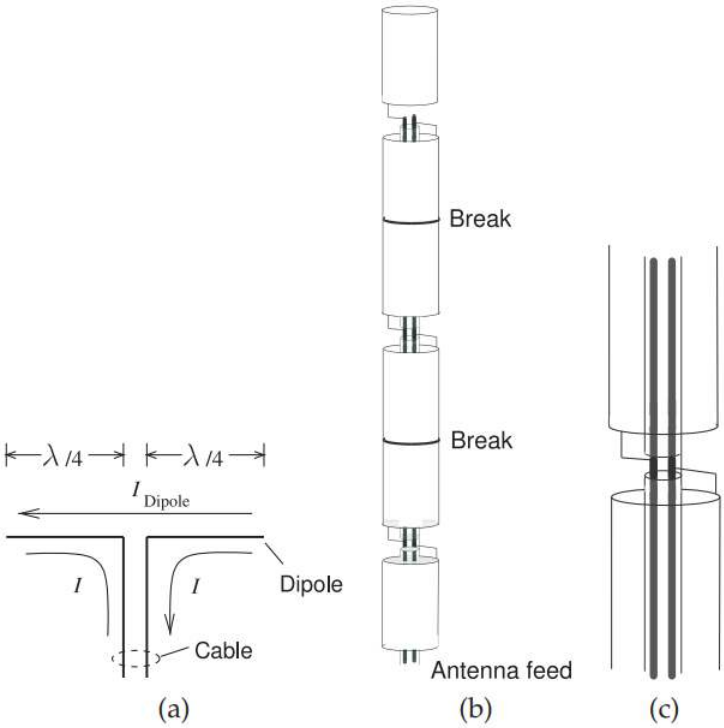


Figure 6.3.4: Dipole antenna: (a) current distribution; (b) stacked dipole antenna; and (c) detail of the connection in a stacked dipole antenna.

be directly connected to the antenna and there is only a small mismatch and nearly all the power is transferred to the antenna and then radiated.

Another variation on the monopole is shown in Figure 6.3.3, where the key component is the phasing coil. The phasing coil (with a wire length of $\lambda/2$) rotates the electrical angle of the current phasor on the line so that the current on the $\lambda/4$ segment is in the same direction as on the $\lambda/2$ segment. The result is that the two straight segments of the loaded monopole radiate a more tightly confined EM field. The phasing coil itself does not radiate (much).

Another ingenious solution to obtaining a longer effective wire antenna with same-directed current (and hence a more tightly confined RF beam) is the stacked dipole antenna 6.3.4. The basis of the antenna is a dipole as shown in Figure 6.3.4(a). The cable has two conductors that have equal amplitude currents, I, but flowing as shown. The wire section is coupled to the cable so that the currents on the two conductors realize a single effective current I_{dipole} on the dipole antenna. The stacked dipole shown in Figure 6.3.4(b) takes this geometric arrangement further. Now the radiating element is hollow and a coaxial cable is passed through the antenna elements and the half-wavelength sections are fed separately to effectively create a wire antenna that is several wavelengths long with the current pointing in one direction. Most cellular antennas using wire antennas are stacked dipole antennas.

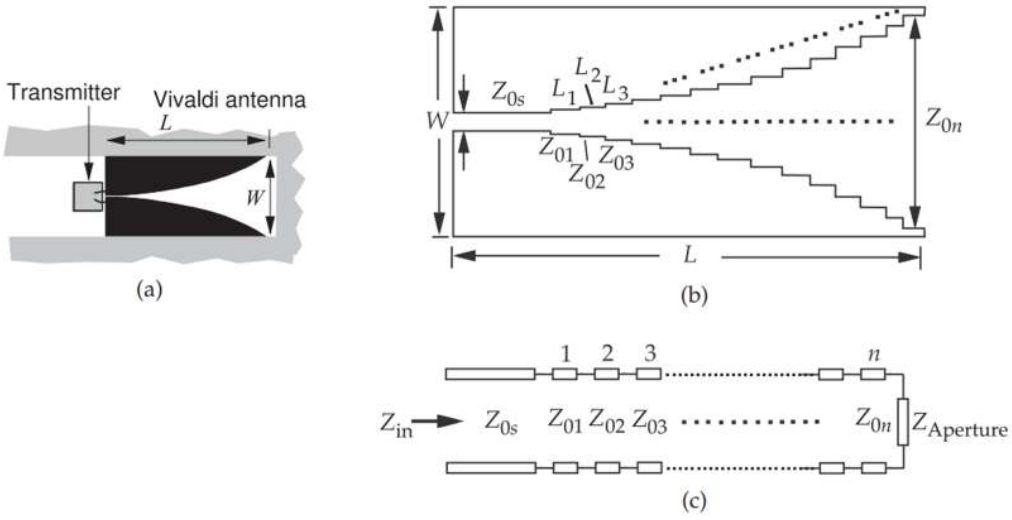


Figure 6.3.5: Vivaldi antenna showing design procedure: (a) the antenna; (b) a stepped approximation; and (c) transmission line approximation. In (a) the black region is a metal sheet.

6.3.2.1: Summary

Standing waves of current can be realized by resonant structures other than wires. A microstrip patch antenna, see Figure 4.1.2(b), is an example, but the underlying principle is that an array of current filaments generates EM components that combine to create a propagating field. Resonant antennas are inherently narrowband because of the reliance on the establishment of a standing wave. A relative bandwidth of 5%–10% is typical.

This page titled 6.3: Resonant Antennas is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.3: Resonant Antennas** by Michael Steer is licensed CC BY-NC 4.0.

6.4: Traveling-Wave Antennas

Traveling-wave antennas have the characteristics of broad bandwidth and large size. These antennas begin as a transmission line structure that flares out slowly, providing a low reflection transition from a transmission line to free space. The bandwidth can be very large and is primarily dependent on how gradual the transition is

One of the more interesting traveling-wave antennas is the **Vivaldi antenna** of Figure 4.3.5(a). The Vivaldi antenna is an extension of a slotline in which the fields are confined in the space between two metal sheets in the same plane. The slotline spacing increases gradually in an exponential manner, much like that of a Vivaldi violin (from which it gets its name), over a distance of a wavelength or more. A circuit model is shown in Figures 4.3.5(b and c) where the antenna is modeled as a cascade of many transmission lines of slowly increasing characteristic impedance, Z_0. Since the Z_0 progression is gradual there are low-level reflections at the transmission line interfaces. The forward-traveling wave on the antenna continues to propagate with a negligible reflected field. Eventually the slot opens sufficiently that the effective impedance of the slot is that of free space and the traveling wave continues to propagate in air

The other traveling-wave antennas work similarly and all are at least a wavelength long, with the central concept being a gradual taper from the

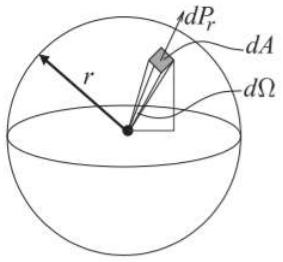


Figure 6.4.1: Free-space spreading loss. The incremental power, dP_r, intercepted by the shaded region of incremental area dA is proportional to $1/r^2$. The solid angle subtended by the shaded area is the incremental solid angle $d\Omega$. The integral of dA over the surface of the sphere, i.e. the area of the sphere is $4\pi r^2$. The total solid angle subtended by the sphere is the integral of $d\Omega$ over the sphere and is 4π steradians (or 4π sr).

characteristic impedance of the originating transmission line to free space. The final aperture is at least one-half wavelength across so that the fields can curl on themselves (i.e., loop back on themselves) and are self-supporting as they leave the antenna.

This page titled 6.4: Traveling-Wave Antennas is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.4: Traveling-Wave Antennas** by Michael Steer is licensed CC BY-NC 4.0.

6.5: Antenna Parameters

This section introduces a number of antenna metrics that are used to characterize antenna performance.

6.5.1: Radiation Density and Radiation Intensity

Antennas do not radiate equally in all directions concentrating radiated power in (usually) one direction called the **main** (or **major**) **lobe** of the antenna. This focusing effect is called **directivity**. The power in a particular direction is characterized by the radiation density and the radiation intensity metrics. The radiation density, S_r, is the power per unit area with the SI units of $\mathrm{W/m^2}$, and will be maximum in the main lobe. Referring to Figure 4.4.1 with an antenna located at the center of the sphere of radius r and radiating a total power P_r, S_r is the incremental radiated power dP_r passing through the incremental shaded region of area, dA:

$$S_r = \frac{dP_r}{dA} \tag{6.5.1}$$

S_r reduces with distance falling off as $1/r^2$ in free space. For a practical antenna S_r will vary across the surface of the sphere. The total power radiated is the closed integral over the surface S of the sphere:

$$P_r = \oint_S dP_r = \oint_S S_r dA \tag{6.5.2}$$

An alternative measure of power concentration is the radiation intensity U which is in terms of the incremental solid angle $d\Omega$ subtended by dA so that $d\Omega = dA/r^2$ and (with the SI units of $\mathrm{W/steradian}$ or $\mathrm{W/sr}$)

$$U = \frac{dP_r}{d\Omega} = \frac{dP_r}{dA} r^2 = r^2 S_r \tag{6.5.3}$$

6.5.1.1: Isotropic Antenna

It is useful to reference the directivity of an antenna with respect to a fictitious isotropic antenna that has no loss and radiates equally in all directions so that S_r is only a function of r. Then integrating over the surface of the sphere yields the total radiated power

$$P_r|_{\mathrm{Isotropic}} = \oint_S dP_r = \oint_S S_r dA = S_r \oint_S dA = S_r 4\pi r^2 = 4\pi U \tag{6.5.4}$$

Since the isotropic antenna has no loss the power input to the antenna P_{IN} is equal to the power radiated $P_r = P_{\mathrm{IN}}$. Thus for an isotropic antenna

$$S_r = \frac{P_r}{4\pi r^2} = \frac{P_{\mathrm{IN}}}{4\pi r^2} \tag{6.5.5}$$

$$U|_{\mathrm{Isotropic}} = r^2 S_r = \frac{P_{\mathrm{IN}}}{4\pi r^2} \tag{6.5.6}$$

6.5.1.2: Antenna Efficiency

Antenna efficiency, sometimes called the **radiation efficiency**, describes losses in an antenna principally due to resistive (I^2R) losses. Resonant antennas work by creating a large current that is maximized through the generation of a standing wave at resonance. There is a lot of current, and even just a little resistance results in substantial resistive loss. The power that is reflected from the input of the antenna is usually small. The total radiated power (in all directions), P_r, is the power input to the antenna less losses. The antenna efficiency, η_A is therefore defined as

$$\eta_A = P_r / P_{\mathrm{IN}} \tag{6.5.7}$$

where P_{IN} is the power input to the antenna and $\eta_A < 1$ and usually expressed as a percentage. Antenna efficiency is very close to one for many antennas, but can be 50% for microstrip patch antennas.

Antenna loss refers to the same mechanism that gives rise to antenna efficiency. Thus an antenna with an antenna efficiency of 50% has an antenna loss of $3\ \mathrm{dB}$. Generally losses are resistive due to I^2R loss and mismatch loss of the antenna that occurs when the input impedance is not matched to the impedance of the cable connected to the antenna. Because of confusion with antenna gain (they are not the opposite of each other) the use of the term ‘antenna loss’ is discouraged and instead ‘antenna efficiency’ preferred.

6.5.2: Directivity and Antenna Gain

The directivity of an antenna, D, is the ratio of the radiated power density to that of an isotropic antenna with the same total radiated power, P_r:

$$D = \frac{S_r}{S_r|_{\text{Isotropic}}} = \frac{U}{U|_{\text{Isotropic}}} \tag{6.5.8}$$

where S_r and U refer to the actual antenna and the power densities and intensities are measured at the same distance from the antennas. For an actual antenna D is dependent on the direction from the antenna, see Figure 6.5.1. The maximum value of D will be in the direction of the main lobe of the antenna and this is called **directivity gain**.

The focusing property of an antenna is characterized by comparing the radiated power density to that of an isotropic antenna with the same input power. The antenna gain, G_A, is the maximum value of D when the power input $P_{\text{IN}} = P_r/\eta_A$ to the antenna and the isotropic antenna are the same:

$$G_A = \eta_A \max(D) \tag{6.5.9}$$

Antenna	Type	Figure	Gain (dBi)	Notes
Lossless isotropic antenna			0	
$\lambda/2$ dipole	Resonant	4.3.4(a)	2	$R_{\text{in}} = 73\,\Omega$
3λ diameter parabolic dish	Traveling	—	38	$R_{\text{in}} = \text{match}$
Patch	Resonant	4.1.2(b)	9	$R_{\text{in}} = \text{match}$
Vivaldi	Traveling	4.1.2(c)	10	$R_{\text{in}} = \text{match}$
$\lambda/4$ monopole on ground	Resonant	4.3.2(a)	2	$R_{\text{in}} = 36\,\Omega$
$5/8\lambda$ monopole on ground	Resonant	4.3.2(a)	3	Matching required

Table 6.5.1: Several antenna systems. $R_{\text{in}} = \text{match}$ for resonant antennas indicates that the antenna can be designed to have an input impedance matching that of a feed cable. Traveling wave antennas are intrinsically matched.

Losses in the antenna are accounted for by the efficiency term η_A.

In Equation (6.5.9) G_A is a gain factor and is often expressed in terms of decibels (taking 10 times the log of G_A) but dBi (with 'i' standing for withrespect-to isotropic) is used to indicate that it is not a power gain in the same sense as amplifier gain. G_A instead is the ratio of power densities for two different antennas. For example, an antenna that focuses power in one direction increasing the peak radiated power density by a factor of 20 relative to that of an isotropic antenna has an antenna gain of 13 dBi. With care G_A can be often used in calculations of power as as with amplifier gain.

Since it is almost impossible to calculate internal antenna losses, antenna gain is invariably only measured. The input power to an antenna can be measured and the peak radiated power density, $P_D|_{\text{Maximum}}$, measured in the far field at several wavelengths distant (at $r \gg \lambda$). This is compared to the power density from an ideal isotropic antenna at the same distance with the same input power. Antenna gain is determined from

$$G_A = \frac{\text{Maximum radiated power per unit area}}{\text{Maximum radiated power per unit area for an isotropic antenna}} \tag{6.5.10}$$

$$= \frac{S_r|_{\text{Maximum}}}{S_r|_{\text{Isotropic}}} = 4\pi r^2 \frac{P_D|_{\text{Maximum}}}{P_{\text{IN}}} \tag{6.5.11}$$

$$= 4\pi \frac{\text{Maximum radiated power per unit solid angle}}{\text{Total input power to the antenna}}$$

$$= 4\pi \frac{(dP_r/d\Omega)|_{\text{Maximum}}}{P_{\text{IN}}} = 4\pi r^2 \frac{(dP_r/dA)|_{\text{Maximum}}}{P_{\text{IN}}} \tag{6.5.12}$$

The antenna gains of common resonant and traveling-wave antennas are given in Table 6.5.1. In free space the antenna gain determined using Equation (6.5.9) is independent of distance. Antenna gain is measured on an antenna range using a calibrated receive antenna and care taken to avoid reflections from objects, especially from the ground.

The losses of an antenna are incorporated in the antenna gain which is defined in terms of the power input to the antenna, see Equation (6.5.12). Thus in calculations of radiated power using antenna gain, there is no need to separately account for resistive losses in the antenna.

In summary, antennas concentrate the radiated power in one direction so that the density of the power radiated in the direction of the peak field is higher than the power density from an isotropic antenna. Power radiated from a base station antenna, such as that shown in Figure 6.5.2, is concentrated in a region that looks like a toroid or, more closely, a balloon squashed at its north and south poles. Then the antenna does not radiate much power into space and will concentrate power in a region skimming the surface of

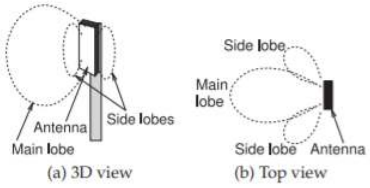


Figure 6.5.1: Field pattern produced by a microstrip antenna.

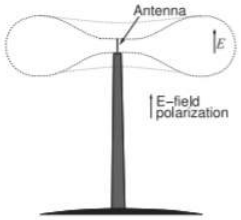


Figure 6.5.2: A base station transmitter pattern.

the earth. Antenna gain is a measure of the effectiveness of an antenna to concentrate power in one direction. Thus, in free space due to power spreading out the maximum power density (in SI units of $\mathrm{W/m^2}$) at a distance r is

$$P_D = \frac{G_A P_{\mathrm{IN}}}{4\pi r^2} \tag{6.5.13}$$

where $4\pi r^2$ is the area of a sphere of radius r and P_{IN} is the input power.

Measurements of antenna gain are used to derive antenna efficiency. It is impossible to measure or simulate the resistive and dielectric losses of an antenna directly. Antenna efficiency is obtained using theoretical calculations of antenna gain assuming no losses in the antenna itself. This is compared to the measured antenna gain yielding the antenna efficiency.

Example 6.5.1: Antenna Gain

A base station antenna has an antenna gain, G_A, of 11 dBi and a 40 W input. The transmitted power density falls off with distance d as $1/d^2$. What is the peak power density at 5 km?

Solution

A sphere of radius 5 km has an area $A = 4\pi r^2 = 3.142 \cdot 10^8 m^2$; $G_A = 11$ dBi $= 12.6$. In the direction of peak radiated power, the power density at 5 km is

$$P_D = \frac{P_{\mathrm{IN}} G_A}{A} = \frac{40 \cdot 12.6 \text{ W}}{3.142 \cdot 10^8 \text{ m}^2} = 1.603\ \mu\mathrm{W/m^2}$$

Example 6.5.2: Antenna Efficiency

A antenna has an antenna gain of 13 dBi and an antenna efficiency of 50% and all of the loss is due to resistive losses and resistance of metals is proportional to temperature. The RF signal input to the antenna has a power of 40 W.

a. What is the input power in dBm?
$P_{\text{in}} = 40 \text{ W} = 46.02 \text{ dBm}$.

b. What is the total power transmitted in dBm?

$$P_{\text{Radiated}} = 50\% \text{ of } P_{\text{IN}} = 20 \text{ W or } 43.01 \text{ dBm}.$$
$$\text{Alternatively, } P_{\text{Radiated}} = 46.02 \text{ dBm} - 3 \text{ dB} = 43.02 \text{ dBm}.$$

c. If the antenna is cooled to near absolute zero so that it is lossless, what would the antenna gain be?
The antenna gain would increase by 3 dB and antenna gain incorporates both directivity and antenna losses. So the gain of the cooled antenna is 16 dBi.

6.5.3: Effective Isotropic Radiated Power

A transmit antenna does not radiate power equally in all directions and for a receiver in the main lobe of the transmit antenna it is as though there is an isotropic transmit antenna with a much higher input power. This concept is incorporated in the effective isotropic radiated power (**EIRP**):

$$\text{EIRP} = P_{\text{IN}} G_A \tag{6.5.14}$$

This uses antenna gain to derive the total power that would be radiated by an isotropic antenna producing the same (peak) power density as the actual antenna. Regulatory limits on the power levels of transmitters are in terms of EIRP rather than the total radiated power. Sometimes **equivalent radiated power** (**ERP**) is used instead of EIRP with the same meaning.

6.5.4: Effective Aperture Size

Effective aperture size is defined so that the power density at a receive antenna when multiplied by its effective aperture size, A_R, yields the power output from the antenna at its connector. Since antennas are linear and reciprocal it is is to be expected that there is a relationship between effective aperture size and antenna gain.

An antenna has an effective size that is more than its actual physical size. This is because of its influence on the EM fields around it. The power captured by an antenna is the effective aperture size (or area) multiplied by the transmitted power density. That is, the effective aperture size of an antenna is the area of the surface that captures all of the power passing through it and delivers this power to the output terminals of the antenna.

The effective aperture area of a receive antenna, A_R, is related to the receive antenna gain, G_R, as follows [2, 3] (note that A_e is often used if it is not necessary to distinguish antennas):

$$A_R = \frac{G_R \lambda^2}{4\pi} \tag{6.5.15}$$

where λ is the wavelength of the radio signal[1]. The effective aperture area of an antenna can have little to do with its physical size; e.g., a wire antenna has almost no physical size but has a significant effective aperture size.

If S_r is the transmitted power density at the receive antenna, the power received is

$$P_R = P_D A_R = P_D \frac{G_R \lambda^2}{4\pi} \tag{6.5.16}$$

Here P_R is the power delivered at the output connector of the receive antenna, as loss in the receive antenna is incorporated in G_R and A_R. The total power radiated by the transmit antenna in the direction of maximum power density is given by multiplying the power input to the transmit antenna, P_T,[2] by the antenna gain of the transmit antenna, G_T. This can be converted to the power density at a distance d (ignoring multipath effects),

$$S_r = \frac{P_T G_T}{4\pi d^2} \tag{6.5.17}$$

and the power delivered by the receive antenna is

$$P_R = S_r A_R = \frac{P_T G_T}{4\pi d^2}\frac{G_R \lambda^2}{4\pi} = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2 \tag{6.5.18}$$

That is,

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2 \tag{6.5.19}$$

This equation is known as the **Friis transmission equation** or the **Friis transmission formula**.

Equation (6.5.19) provides a powerful tool for evaluating the power received in a communication system and modifications have been developed to account for impairments that are encountered. One of the assumptions in the development of Equation (6.5.19) is that the polarization of the receive antenna matches that of the transmitted signal. The transmit antenna will radiate a signal with an electric field with a particular polarization, i.e. E field orientation. Propagation through air has little effect on the polarization of a signal, although atmospheric conditions can rotate the polarization slightly. A polarization mismatch factor could be included in Equation (6.5.19). For ground-based communications, such as cellular communications, reflections and diffraction have a major impact on the power of the received signal and techniques for handling these effects are considered in the next section.

6.5.5: Summary

This section introduced several metrics for characterizing antennas:

Metric	Equation	Description
S_r	(6.5.1)	Radiated power density, $\mathrm{W/m^2}$
U	(6.5.3)	Radiation intensity W/sr
η_A	(6.5.15)	Antenna efficiency
D	(6.5.8)	Antenna directivity
G_A	(6.5.10)	Antenna gain, used with a transmit antenna
A_e	(6.5.15)	Effective aperture area, used with a receive antennaa
EIRP	(6.5.14)	Equivalent isotropic radiated power

Example 6.5.3: Point-to-Point Communication

In a point-to-point communication system, a parabolic receive antenna has an antenna gain of 60 dBi. If the signal is 60 GHz and the power density at the receive antenna is $1\ \mathrm{pW/cm^2}$, what is the power at the output of the receive antenna connected to the RF electronics?

Solution

The first step is determining the effective aperture area, A_R, of the parabolic antenna. The frequency is 60 GHz, and so the wavelength (λ) is 5 mm. Also note that $G_R = 60\ \mathrm{dBi} = 10^6$. From Equation (6.5.15)

$$A_R = \frac{G_R \lambda^2}{4\pi} = \frac{10^6 \cdot 0.005^2}{4\pi} = 1.989\ \mathrm{m^2} \tag{6.5.20}$$

Using Equation (6.5.16) $P_D = 1\ \mathrm{pW/cm^2} = 10\ \mathrm{nW/m^2}$, the total power delivered to the RF receiver electronics (at the output of the receive antenna) is

$$P_R = P_D A_R = 10\ \mathrm{nW} \cdot \mathrm{m^{-2}} \cdot 1.989\ \mathrm{m^2} = 19.89\ \mathrm{nW} \tag{6.5.21}$$

6.5.6: Footnotes

[1] The derivation of Equation (6.5.15) invokes a thought experiment with an antenna connected to a resistor in thermal equilibrium and each in separate thermally isolated chambers with black body walls [4, 5]. The available thermal (Johnson-Nyquist) noise power from the resistor is radiated by the antenna (incorporating antenna gain) and absorbed by the walls of the antenna's chamber. The same amount of power, as black body radiation from the walls of the antenna's chamber, must be received by the antenna (incorporating effective aperture size) and delivered to the resistor where it is dissipated as heat. At frequencies up to infrared and at room temperature (and hence at radio frequencies) the power density of black body radiation increases linearly with temperature

and as the square of frequency (i.e. as $1/\lambda^2$) whereas Johnson-Nyquist noise is independent of frequency but increases linearly with temperature. The derivation is exact for an isotropic antenna and assumed to apply to all antennas.

[2] Note that this was P_{IN} previously and here the subscript T is used to indicate the power input to the transmit antenna.

This page titled 6.5: Antenna Parameters is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.5: Antenna Parameters** by Michael Steer is licensed CC BY-NC 4.0.

6.6: The RF Link

The RF link is between a transmit antenna and a receive antenna. Sometimes the RF link includes the antenna and sometimes it does not, this will be clear from the context, but usually it includes the antennas. The principle source of link loss is the spreading out of the EM field as it propagates. In the absence of any other effects (such as atmospheric loss and reflections), the power density reduces as $1/d^2$, where d is distance, and this is called the **line-of-sight** (**LOS**) situation. In this section the propagation path is first described along with its impairments including propagation on multiple paths between a transmit antenna and a receive antenna. The resonant scattering is described as well as multiple fading effects due to reflections, diffraction, rain, and other atmospheric effects.

6.6.1: Propagation Path

When the radiated signal reflects and diffracts there are multiple propagation paths that result in what is called fading, as the paths constructively and destructively combine at the receiver. The direct path and other paths, called multipaths, which are reflected or diffracted by ground, buildings, and other objects do not arrive at the receiver in phase leading to time-varying constructive and destructive combining, called multipath interference. Of these destructive combining is much worse as it can reduce a signal level below what it would be if propagation was in free space. In urban areas, 10 or 20 paths can have significant power in them and these combine at the receive antenna [6].

When the signal on one of the paths dominates, this is usually the LOS path, fading is called **Rician fading**. With LOS and a single ground reflection, the situation is the classic Rician fading as shown in Figure $6.6.1$(a). This is a special situation, as the ground changes the phase of the signal upon reflection, generally by 180°. When the receiver is a long way from the base station, the lengths of the two paths are almost identical and the level of the signals in the two paths are almost the same. The net result is that these two signals almost cancel, and so instead of the power falling off by $1/d^2$, it falls off by $1/d^3$. When there are many paths and all have similar amplitude signals, fading is called **Rayleigh fading**. In an urban area such as that shown in Figure $6.6.1$(b), there are many significant multipaths and the power falls off by $1/d^4$ and sometimes faster.

Common paths encountered in cellular radio are shown in Figure $6.6.2$. As a rough guide, in the single-digit gigahertz range each diffraction and scattering event reduces the signal received by 20 dB. The knife-edge diffraction scenario is shown in more detail in Figure $6.6.3$. This case is fairly easy to analyze and can be used to estimate the effects of individual obstructions. The diffraction model is derived from the theory of half-infinite screen diffraction [7]. First, calculate the parameter ν from the geometry of

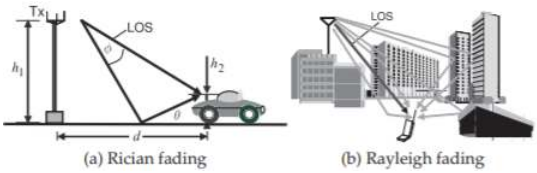


Figure $6.6.1$: Multipath propagation: (a) line of sight (LOS) and ground reflection paths only; and (b) in an urban environment.

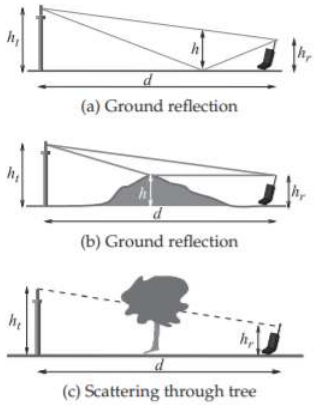


Figure $6.6.2$: Common paths contributing to multipath propagation.

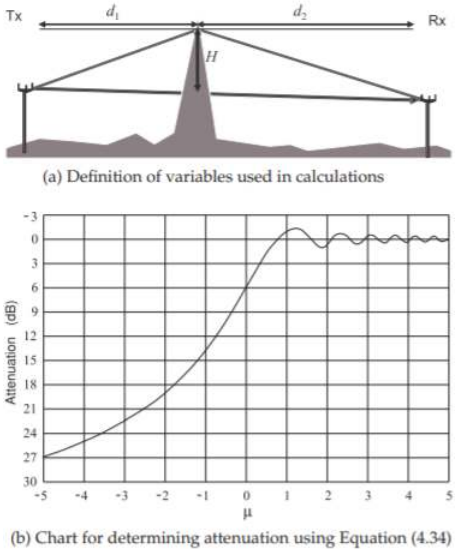


Figure 6.6.3: Knife-edge diffraction.

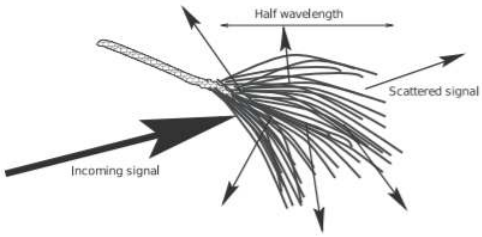


Figure 6.6.4: Pine needles scattering an incoming EM signal.

the path using

$$\nu = -H\sqrt{\frac{2}{\lambda}\left(\frac{1}{d_1}+\frac{1}{d_2}\right)} \tag{6.6.1}$$

Next, consult the plot in Figure 6.6.3(b) to obtain the diffraction loss (or attenuation). This loss should be added (using decibels) to the otherwise determined path loss to obtain the total path loss. Other losses such as reflection cancelation still apply, but are computed independently for the path sections before and after the obstruction.

6.6.2: Resonant Scattering

Propagation is rarely from point to point as the path is often obstructed and so non-LOS (**NLOS**). One type of event that reduces transmission is scattering. The level of the effect depends on the size of the objects causing scattering. Here the effect of the pine needles of Figure 6.6.4 will be considered. The pine needles (as most objects in the environment) conduct electricity, especially when wet. When an EM field is incident an individual needle acts as a wire antenna, with the current maximum when the pine needle is one-half wavelength long. At this length, the “needle” antenna supports a standing wave and will re-radiate the signal in all directions. This is scattering, and there is a considerable loss in the direction of propagation of the original fields. As well, there is loss due to a needle not being a very good conductor, so EM energy is lost as heat. The effect of scattering is frequency and size dependent. A typical pine needle is 15 cm long, which is exactly $\lambda/2$ at 1 GHz, and so a stand of pine trees have an extraordinary impact on cellular communications at 1 GHz. As a rough guide, 20 dB of a signal is lost when passing through a small stand of pine trees. At 2 GHz, a similar impact comes from other leaves, and they do not need to look like wire antennas. Consider an oak leaf having a dimension of 7.5 cm. This is $\lambda/2$ at 2 GHz, another dominant cellular frequency. So scattering results, but now the loss is season dependent, with the loss due to scattering being considerably smaller in winter (when trees such as oaks do not have leaves) than in summer.

6.6.3: Fading

Fading refers to the variation of the received signal with time or when the position of transmit or receive antennas is changed. The variation in signal strength is much greater than would be expected from changes in path

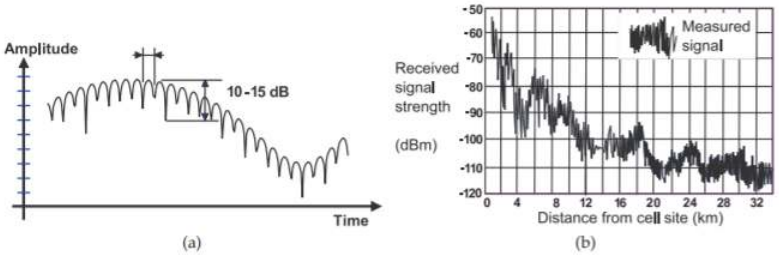


Figure 6.6.5: Fast and slow fading: (a) in time as a radio and obstructions move; and (b) in distance.

length. Several forms of fading are identified and associated with different physical effects. The most important fading types are flat fading, multipath fading, and rain fading.

6.6.3.1: Flat Fading

Temperature variations of the atmosphere between the transmit antenna and receive antenna give rise to what is called flat fading and sometimes called **thermal fading**. This fading is called flat because it is independent of frequency. One form of flat fading is due to refraction, which occurs when different layers of the atmosphere have different densities and thus dielectric permittivities increasing or decreasing away from the surface of the earth. The temperature profile can increase away from the earth surface or reduce depending on whether the temperature of the earth is higher than that of the air and is commonly associated with the beginning and end of the day. **Temperature inversions** can also occur where the temperature profile in air does not uniformly increase or decrease, thus causing a layer with a higher permittivity than that of the air above or below. RF energy gets trapped in this layer, reflecting from the top and bottom of the inversion layer. This is called **ducting**. In point-to-point communication systems the transmit and receive antennas are mounted high on towers and then reflection from ground objects is often small. In such cases flat fading is the most commonly observed phenomena and minor fluctuations of several decibels in receive signal level are common throughout the day. However, when temperature variations are extreme, ducting can severely impact communications reducing signal levels by up to 20 dB.

6.6.3.2: Shadow Fading

Shadow fading occurs when the LOS path is blocked by an obstruction such as a building or hill. The full signal strength returns when the obstruction or receiver has moved to restore the direct path. Shadow fading is also called slow fading and the characteristic of the channel is regarded as being relatively constant over a short time. The amplitude response varies in time and distance and is shown in Figure 6.6.5. This figure shows both fast fades that are 10 to 15 dB deep and slow or shadow fades that are 20 to 30 dB deep.

6.6.3.3: Multipath Fading

Multipath fading is the most common fading when either the transmit antenna or the receive antenna are close to the ground, near obstructing buildings or terrains, or inside a building. In such situations there are many reflections that combine destructively and constructively. Multipath fading is also called **fast fading**, as the characteristics of the channel can change significantly in a few milliseconds. Two types of multipath fading—Rician fading and Rayleigh fading—will be considered.

Multipath is principally a problem when the line of sight between transmit and receive antennas is obscured. However, where there is line of sight the signal reflecting from the ground immediately in front of a receive antenna can sometimes largely cancel the line-of-sight signal. In time, intervening objects can move and the propagation characteristics of the various paths can change due to thermal variations. All this adds to the randomness of fast fades. Multipath fading of 20 dB can occur for a small percentage of the time on time scales of many seconds when there are few propagation paths (e.g. in a rural area) to a large percentage of the time many times per second in a dense urban environment when there are many paths. Constructive combining does increase the signal level momentarily, but there is no advantage to this. Destructive combining can result in deep fades of 20 dB impacting communications and forcing the communication system to accommodate either by using higher average powers or using strategies

such as multiple antennas or spreading the communication signal over a wide bandwidth since fades tend to be 500 kHz to 1 MHz wide at all frequencies.

6.6.3.4: Rician Fading

Rician fading occurs when there is one dominant RF path, usually the LOS path, and one or more other paths. The main situation is when there is an LOS path and a ground reflection as shown in Figure 6.6.1(a). The signal received is the sum of two signals:

$$v_r(t) = C[\cos(\omega_c t) + r\cos(\omega_c t + \phi)] \tag{6.6.2}$$

where $C\cos(\omega_c t)$ is the LOS signal, r is the ratio of the amplitude of the signal reflected from the ground and the LOS signal, and ϕ is the relative phase. When the distance between the transmitted and received signal is large, the ground reflection will be a glancing reflection and the LOS and reflected path will be almost exactly the same length. Ground reflection will usually introduce a 180° phase rotation in the reflected signal and r could be 0.8–0.9. Thus there will be nearly complete cancelation of the signal [8].

6.6.3.5: Rayleigh Fading

Rayleigh fading, or **fast fading**, in a multipath environment results from destructive cancellation as individual paths of similar amplitude drift in and out of phase as the receiver and sources of multiple reflections, diffractions, and refractions move. With multiple paths between a transmitter and a receiver, different components of the received signal arrive at different times. When the line of sight is blocked, the received signal on as many as 20 of these paths can have appreciable signal power. The time between when the first significant received signal is received and the last is received is called the

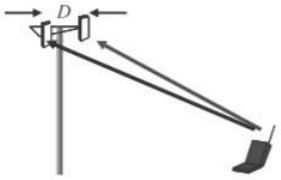


Figure 6.6.6: Two receive antennas used to achieve **space diversity** and overcome the effects of Rayleigh fading.

delay spread. The delay spread is largest inside buildings. A typical office building has a delay spread of 30 ns, and an office building with highly reflective walls and large open spaces has a **delay spread** of up to 250 ns [9, 10, 11], similar delay spreads are obtained at all RF frequencies [12]. Rayleigh fading is the type of fading that occurs when the inverse of the delay spread is small compared to the bandwidth of the received signal. Then when the multipaths combine destructively the entire signal within a bandwidth of 500 kHz to 1 MHz is suppressed. This bandwidth is experimentally observed and is related to the delay spread of individual paths, each with random amplitude and phase. If a communication signal has a bandwidth that is entirely within the suppression bandwidth of a Rayleigh fade, the situation is called frequency flat fading. Some communication signals have bandwidths greater than the Rayleigh fade bandwidth, so it is possible to recover from Raleigh fades if error correction is used.

The result of Rayleigh fading is rapid amplitude variation with respect to frequency. In Figure 6.6.5(a) the amplitude varies in time depending on the speed of moving objects. This graph could just as well be a plot of amplitude versus distance or amplitude versus frequency. A measured amplitude response is shown in Figure 6.6.5(b). Although somewhat random, the fades occur over distances roughly $\frac{1}{2}\lambda$ apart and 500 kHz wide. The 500 kHz width is almost independent of frequency from hundreds of megahertz to 100 GHz. The probability of receiving a signal x dB below the time-averaged received signal power is approximately 10^{-x} [13]. Rayleigh fading is named after the statistical model that describes this.

Fortunately deep Rayleigh fades are very short, last a small percentage of the time, and slight changes to the propagation environment can circumvent their effects. One strategy for overcoming fades is to use two receive antennas as shown in Figure 6.6.6. Here the signals received by two antennas separated by several wavelengths will rarely fade concurrently. In practice, the required separation for good decorrelation is found to be 10 to 20λ.

In an LOS wireless system (e.g. a point-to-point link without multipath), Rayleigh fading is due to rapidly changing atmospheric conditions, with the refractive index of small regions varying. These fades occur over a few seconds.

6.6.3.6: Rain Fading

Rain fading is due to both the amount of rain and the size of individual rain drops and fading occurs over periods of minutes to hours. Propagation through the atmosphere is affected by absorption by molecules in the air, fog, and rain, and by scattering by rain drops. Figure 6.6.7 shows the attenuation in decibels per kilometer from 3 GHz to 300 GHz. Curve (a) shows the

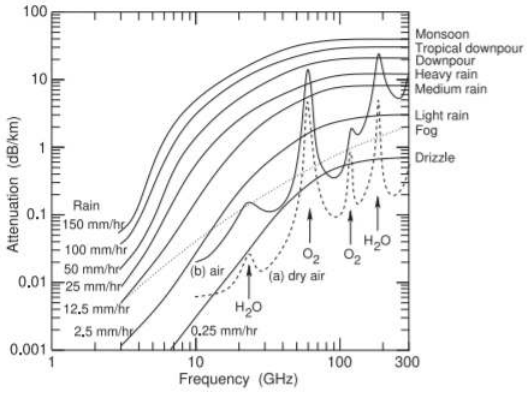


Figure 6.6.7: Excess attenuation due to atmospheric conditions showing the effect of rain on RF transmission at sea level. Curve (a) is atmospheric attenuation, due to excitation of molecular resonances, of very dry air at 0°C, curve (b) is for typical air (i.e. less dry) at 20°C. The attenuation shown for fog and rain is additional (in dB) to the atmospheric absorption shown as curve (b).

attenuation in dry air at 0°C with very low attenuation at a few gigahertz with several attenuation peaks as frequency increases. The first peak in attenuation is due to the resonance of water molecules in water vapor. The resonance peaks at 23 GHz but loss starts increasing before that. The next absorption peak is at 60 GHz due to the resonance of oxygen molecules. Two other absorption peaks are observed going up to 300 GHz. Curve (b) is for a less dry atmosphere at 20°C. The dotted curve shows the additional effect of fog on attenuation and then a family of curves shows the effect of rain on propagation. Below 10 GHz the effect of rain is very little and is safely ignored below 5 GHz. The attenuation due to rain increases with frequency and this derives largely from scattering and is related to the RF signal's wavelength relative to the circumference of rain drops.

6.6.3.7: Fading Due to Ducting

Normally the temperature and density of air, and thus its refractive index, drops with increasing height above earth. As a result, radio waves will be refracted toward the earth as shown in Figure 6.6.18(a). However, during periods of stable weather characterized by a high-pressure system, the temperature can rise with increasing height before eventually falling, creating what is called temperature inversion. This can occur at heights of tens to hundreds of meters. Temperature inversion is most common in summer, but can also occur when the temperature is dropping quickly, such as at sunset. At sunset the inversion layer may occur at a meter or so above the ground as the earth cools rapidly. The denser colder air above the ground has a higher permittivity than the air above and this results in an inversion layer that has a higher refractive index than the air above and below. RF waves can become trapped in the inversion layer, as any RF energy leaving the temperature inversion layer is refracted back into the layer. This effect is called ducting. Ducting can also occur when a cold air mass is overrun by warm air.

Fading due to ducting occurs when the receiver wanders in and out of the ducting layer, as the ducting layer is not stable, increasing and decreasing in

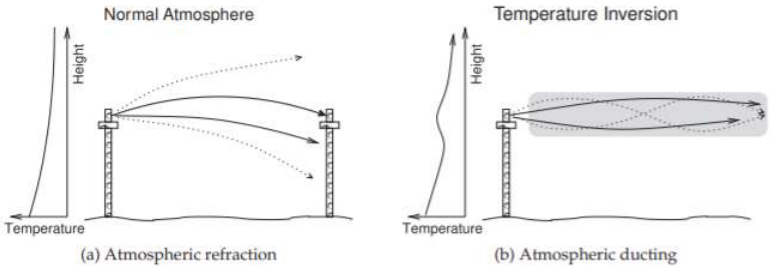


Figure 6.6.8: Fading resulting from ducting: (a) normal atmospheric refraction (normally the temperature of air drops with increasing height and the lower refractive index at high heights results in a concave refraction); and (b) atmospheric ducting (resulting from temperature inversion inducing an air layer with higher dielectric constant than the surrounding air).

strength and in geometry.

6.6.3.8: Summary of Fading

The fades of most concern in a mobile wireless system are the deep fades resulting from destructive interference of multiple reflections. These fades vary rapidly (over a few milliseconds) if a handset is moving at vehicular speeds but occur slowly when the transmitter and receiver are fixed. Fades can be viewed as deep amplitude modulation, and so so modulation is restricted to phase shift keying schemes when a transmitter and receive moving at vehicular speeds relative to each other.

6.6.4: Link Loss and Path Loss

With transmit and receive antennas included in the RF link, the usual case, link loss is defined as the ratio of the power input to the transmit antenna (P_T) to the power delivered by the receive antenna (P_R). Rearranging Equation (4.5.18) and taking logarithms yields the total line-of-sight (LOS) link loss, $L_{\text{LINK, LOS}}$, between the input of the transmit antenna and the output of the receive antenna separated by distance d is (in decibels):

$$L_{\text{LINK, LOS}}|_{\text{dB}} = 10\log\left(\frac{P_T}{P_R}\right) = 10\log\left(\frac{P_T}{P_D A_R}\right) \tag{6.6.3}$$

$$= 10\log\left[P_T\left(\frac{4\pi d^2}{P_T G_T}\right)\left(\frac{4\pi}{\lambda^2 G_R}\right)\right] \tag{6.6.4}$$

$$= 10\log\left[\left(\frac{1}{G_T G_R}\right)\left(\frac{4\pi d}{\lambda}\right)^2\right] \tag{6.6.5}$$

$$= -10\log G_T - 10\log G_R + 20\log\left(\frac{4\pi d}{\lambda}\right) \tag{6.6.6}$$

The last term includes d and is called the LOS path loss (in decibels):

$$L_{\text{PATH, LOS}}|_{\text{dB}} = 20\log\left(\frac{4\pi d}{\lambda}\right) \tag{6.6.7}$$

This is the preferred form of the expression for path loss, as it can be used directly in calculating link loss using the antenna gains of the transmit and receive antennas without the exercise of calculating the **effective aperture size** of the receive antenna.

An alternative definition of path loss comes directly from Equation (4.5.11) and is called the LOS path loss of the first kind:

$${}^1L_{\text{PATH, LOS}} = 4\pi d^2 \tag{6.6.8}$$

but this is not commonly used.

Multipath effects result in losses that are proportional to d^n [14, 15] so that the general path loss, including multipath effects, is (in decibels)

$$\begin{aligned} L_{\text{PATH}}|_{\text{dB}} &= L_{\text{PATH, LOS}}|_{\text{dB}} + \text{excess loss}|_{\text{dB}} \\ &= 20\log\left(\frac{4\pi d}{\lambda}\right) + 10(n-2)\log\left(\frac{d}{1\text{ m}}\right) \\ &= 20\log\left[\frac{4\pi(1\text{ m})}{\lambda}\right] + 10(2)\log\left(\frac{d}{1\text{ m}}\right) + 10(n-2)\log\left(\frac{d}{1\text{ m}}\right) \\ &= 10n\log[d/(1\text{ m})] + C \end{aligned} \tag{6.6.9}$$

where the distance d and wavelength λ are in meters, and C is a constant that captures the effect of wavelength. Here,

$$C = 20\log[4\pi(1\text{ m})/\lambda] \tag{6.6.10}$$

Combining this with Equation (6.6.6) yields the link loss between the input of the transmit antenna and the output of the receive antenna:

$$L_{\text{LINK}}|_{\text{dB}} = -G_T|_{\text{dB}} - G_R|_{\text{dB}} + 10n\log[d/(1\text{ m})] + C \tag{6.6.11}$$

As you can imagine, a few constants, here n and C, cannot capture the full complexity of the propagation environment. Many models have been developed to capture particular environments better and incorporate mast height, experimental correction factors, and statistical parameters.

The path loss between two antennas is exactly the same in both directions when the frequency of the signal in each direction is the same. This is radio link reciprocity. Many communication systems use different frequencies in the two directions, and then the links are not reciprocal.

Example 6.6.1: Power Density

A communication system operating in a dense urban environment has a power density rolloff of $1/d^{3.5}$ between the base station transmit antenna and the mobile receive antenna. At $10\ \text{m}$ from the transmit antenna, the power density is $0.3167\ \text{W/m}^2$. What is the power density at the receive antenna located at $1\ \text{km}$ from the base station?

Solution

$P_D(10\ \text{m}) = 0.3167\ \text{W/m}^2$ and let $d_c = 10\ \text{m}$, so at $d = 1\ \text{km}$, the power density $P_D(1\ \text{km})$ is obtained from

$$\frac{P_D(1\ \text{km})}{P_D(10\ \text{m})} = \frac{d_c^{3.5}}{d^{3.5}} = \frac{10^{3.5}}{1000^{3.5}} = 10^{-7} \tag{6.6.12}$$

so

$$P_D(1\ \text{km}) = P_D(10\ \text{m}) \cdot 10^{-7} = 31.7\ \text{nW/m}^2 \tag{6.6.13}$$

Example 6.6.2: Link Loss

A $5.6\ \text{GHz}$ communication system uses a transmit antenna with an antenna gain G_T of $35\ \text{dB}$ and a receive antenna with an antenna gain G_R of $6\ \text{dB}$. If the distance between the antennas is $200\ \text{m}$, what is the link loss if the power density reduces as $1/d^3$? The link loss here is between the input to the transmit antenna and the output from the receive antenna.

Solution

The link loss is provided by Equation (6.6.11),

$$L_{\text{LINK}}|_{\text{dB}} = -G_T - G_R + 10n\log[d/(1\ \text{m})] + C$$

and C comes from Equation (6.6.10), where $\lambda = 5.36\ \text{cm}$. So

$$C = 20\log\left(\frac{4\pi}{\lambda}\right) = 20\log\left(\frac{4\pi}{0.0536}\right) = 47.4\ \text{dB}$$

With $n = 3$ and $d = 200\ \text{m}$,

$$L_{\text{LINK}}|_{\text{dB}} = -35 - 6 + 10 \cdot 3 \cdot \log(200) + 47.4\ \text{dB} = 75.4\ \text{dB}$$

Example 6.6.3: Radiated Power Density

In free space, radiated power density drops off with distance d as $1/d^2$. However, in a terrestrial environment there are multiple paths between a transmitter and a receiver, with the dominant paths being the direct LOS path and the path involving reflection off the ground. Reflection from the ground partially cancels the signal in the direct path, and in a semi-urban environment results in an attenuation loss of $40\ \text{dB}$ per decade of distance (instead of the $20\ \text{dB}$ per decade of distance roll-off in free space). Consider a transmitter that has a power density of $1\ \text{W/m}^2$ at a distance of $1\ \text{m}$ from the transmitter.

a. The power density falls off as $1/d^n$, where d is distance and n is an index. What is n?
b. At what distance from the transmit antenna will the power density reach $1\ \mu\text{W} \cdot \text{m}^{-2}$?

Solution

a. Power drops off by $40\ \text{dB}$ per decade of distance. $40\ \text{dB}$ corresponds to a factor of $10,000 (= 10^4)$. So, at distance d, the power density $P_D(d) = k/d^n$ (k is a constant). At a decade of distance, $10d$, $P_D(10d) = k/(10d)^n = P_D(d)/10000$, thus

$$\frac{k}{10^n d^n} = \frac{1}{10,000}\frac{k}{d^n}; \quad 10^n = 10,000 \Rightarrow n = 4$$

b. At $d = 1$ m, $P_D(1\text{ m}) = 1\text{ W/m}^2$. At a distance x,

$$P_D(x) = 1\,\mu\text{W/m}^2 = \frac{k}{x^4}\text{m}^2 = \frac{k}{x^4} \rightarrow x^4 = \frac{1}{10^{-6}} \quad \text{and so} \quad x = 31.6\text{ m}$$

6.6.5: Fresnel Zones

In long-distance wireless communications from one fixed site to another, the intent is to use the LOS path and avoid reflections. Such systems are called **point-to-point links**. Thus avoiding reflections is an important consideration in design. The dominant propagation paths in a point-to-point system are shown in Figure 6.6.9. The refracted wave path arises because of density variations in the air producing a permittivity profile that varies with height, as shown in Figure 6.6.10(a). This effect is called **beam bending**. The other important propagation path to consider is the reflected wave from the ground, which can be important if the ground is too close to the propagation path. Both of these effects will be considered in this section.

As radio waves propagate, they spread out in a plane perpendicular to the direction of propagation, the power density of the radio waves then reduces with distance from the centerline. One of the consequences of this is that an obstruction that is not in the LOS path can still interfere with signal propagation. The appropriate clearance is determined from the Fresnel zones, which are shown in Figure 6.6.11. The direct LOS path between the antennas has a length d. If there is a reflecting object near the LOS path, then there can be a second path between the transmit and receive antennas. The path from the first antenna to the circle defined by the first Fresnel zone and then to the second antenna has a path length $d + \lambda/2$, and so at the receive antenna this signal is 180° out of phase with the LOS signal and there will be

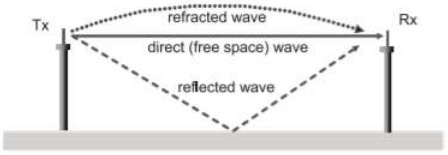


Figure 6.6.9: Three point-to-point characteristic propagation paths: line of sight, reflection, and refraction.

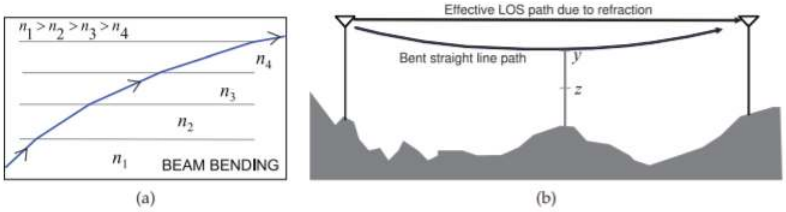


Figure 6.6.10: Beam bending by density variation in air: (a) refraction index profile with the air density reducing with height; and (b) incorporating beam bending in a curved-earth model.

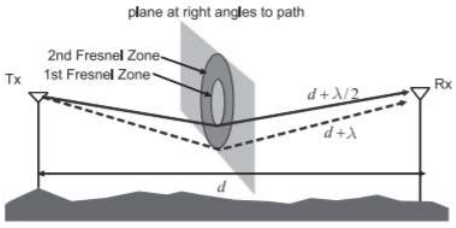


Figure 6.6.11: Fresnel zones in a plane perpendicular to the LOS path, which has length d.

cancelation. The radius of the nth Fresnel zone at a point P is

$$F_n = \sqrt{\frac{n\lambda d_1 d_2}{d_1 + d_2}} \tag{6.6.14}$$

https://eng.libretexts.org/@go/page/113764

where d_1 is the distance from the first antenna to P, d_2 is the distance from the second antenna to P (so $d = d_1 + d_2$), and λ is the wavelength of the propagating signal. Ninety percent of the energy in the wave is in the first Fresnel zone. A guideline is that an obstacle should be separated from the direct path by a distance more than the radius of the first Fresnel zone. Antenna heights are increased so that the beam achieves one or more Fresnel zone clearances.

The above analysis can be used even with beam bending. A convenient way of accommodating beam bending is to use a curved-earth model, as shown in Figure 6.6.10(b), so that subsequent calculations can use LOS considerations [16]. The amount of beam bending to account for comes from experimental surveys. In Figure 6.6.10(b) the original clearance from a hill to the first Fresnel zone is z. With beam bending the clearance increases to y.

Since a receive antenna is unlikely to be large enough to capture all of the energy contained in the first Fresnel zone, the effect of signal blocking by an obstruction that encroaches the first Fresnel zone is of little concern. What is of concern is the destructive combining of a reflected signal with the signal in the main LOS beam.

Example 6.6.4: Fresnel Zone Clearance

A transmit antenna and a receive antenna are separated by 10 km and operate at 2 GHz.

a. What is the radius of the first Fresnel zone?
b. What is the radius of the second Fresnel zone?
c. To ensure LOS propagation, what should the clearance be from the direct line between the antennas and obstructions such as hills and vegetation?

Solution

The radius of the Fresnel zone is calculated at the midpoint so that $d_1 = d_2 = d/2 = 5$ km. Also $f = 2$ GHz, $\lambda = 15$ cm.

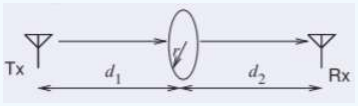


Figure 6.6.12

a. The radius of the first Fresnel zone is, from Equation (6.6.14),

$$r_1 = F_1 = \sqrt{\frac{\lambda d_1 d_2}{d_1 + d_2}} = \sqrt{\frac{\lambda d}{4}} = \sqrt{\frac{(0.15)(10^4)}{4}} = 19.36 \text{ m}$$

b. At the midpoint the radius of the second Fresnel zone is

$$r_2 = F_2 = \sqrt{\frac{2\lambda d_1 d_2}{d_1 + d_2}} = \sqrt{\frac{\lambda d}{2}} = \sqrt{\frac{(0.15)(10^4)}{2}} = 27.39 \text{ m}$$

c. Ninety percent of the energy in the beam is contained within the first Fresnel zone of radius r_1. Obstructions within the first Fresnel zone will result in a significant fraction of the beam being blocked. In addition, reflections from the obstruction could destructively combine with the main beam and reduce the received signal level even beyond the reduction caused by part of the beam being blocked. Two criteria are commonly used to determine the clearance required to avoid signal obstruction.
One criterion used is that the minimum clearance between the direct beam and an obstruction is $0.6r_1$. So the minimum clearance is $0.6r_1 = 11.6$ m.
A more conservative criterion is that the minimum clearance should be $r_1 = 19.36$ m.

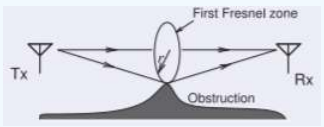


Figure 6.6.13

6.6.6: Propagation Model in the Mobile Environment

RF propagation in the mobile environment cannot be accurately derived. Instead, a fit to measurements is often used. One of the models is the Okumura–Hata model [17], which calculates the path loss as

$$L_{\mathrm{PATH}}|_{(\mathrm{dB})} = 69.55 + 26.16 \log f - 13.82 \log H + (44.9 - 6.55 \log H) \cdot \log d + c \tag{6.6.15}$$

where f is the frequency (in MHz), d is the distance between the base station and terminal (in km), H is the effective height of the base station antenna (in m), and c is an environment correction factor ($c = 0$ dB in a dense urban area, $c = -5$ dB in an urban area, $c = -10$ dB in a suburban area, and $c = -17$ dB in a rural area, for $f = 1$ GHz and $H = 1.5$ m).

More sophisticated characterization of the propagation environment uses ray-tracing models to follow individual propagation paths. The ray-tracing models are based on deterministic methods using terrain data and calculate paths accounting for obstruction and reflection analyses. Each refraction and reflection event is characterized either experimentally or through detailed EM simulations. Appropriate algorithms are applied for best compliance with radio physics. Commonly required inputs to these models include frequency, distance from the transmitter to the receiver, effective base station height, obstacle height and geometry, radius of the first Fresnel zone, forest height/roof height, distance between buildings, arbitrary loss allowances based on land use (forest, water, etc.), and loss allowances for penetration of buildings and vehicles. Such a technique was used to calculate the radio coverage diagrams on the inside front cover of this book.

There are many propagation models for different frequency ranges and different environments. The considerable effort put into developing reliable models is because being able to predict signal coverage is essential to efficient design of basestation layout.

This page titled 6.6: The RF Link is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.6: The RF Link** by Michael Steer is licensed CC BY-NC 4.0.

6.7: Summary

This chapter discussed the impacts on communication integrity of imperfections in the RF link from the output of the transmitter to the input of the receiver. Prior to cellular communications becoming so important, only the LOS communication path was considered and the impact of fading, multiple reflections, and delay spread was regarded as bothersome. Most commonly the solution to the problems introduced by these effects was either to shift to another frequency, to increase the operating power, or install more basestations. With the advent of cellular communications, and digital communications in general, techniques have been developed to overcome the negative impacts of fading without boosting signal levels to the level where they severely impact the operation of other radios. So while many aspects of RF propagation are random, concepts and statistical models have been developed that enable design choices to be made that permit digital communication systems to operate in what would have once been regarded as a hostile environment.

This page titled 6.7: Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.7: Summary** by Michael Steer is licensed CC BY-NC 4.0.

6.8: References

[1] S. Ramo, J. Whinnery, and T. Van Duzer, Fields and Waves in Communication Electronics. John Wiley & Sons, 1965.

[2] C. A. Balanis, “Antenna theory: A review,” Proc. IEEE, vol. 80, no. 1, pp. 7–23, 1992.

[3] C. Balanis, Antenna Theory: Analysis and Design. John Wiley & Sons, 2005.

[4] T. L. Wilson, K. Rohlfs, and S. H ¨uttemeister, Tools of radio astronomy, 6th ed. Springer, 2013.

[5] J. L. Pawsey and R. N. Bracewell, “Radio astronomy,” Oxford, Clarendon Press, 1955.

[6] J. Doble, Introduction to Radio Propagation for Fixed and Mobile Communications. Norwood, MA, USA: Artech House, Inc., 1996.

[7] M. Born and E. Wolf, “Two-dimensional diffraction of a plane wave by a half-plane,,” in Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction of Light, 7th ed. Cambridge University Press, 1999, section 11.5.

[8] A. Abdi, C. Tepedelenlioglu, M. Kaveh, and G. Giannakis, “On the estimation of the k parameter for the rice fading distribution,” IEEE Communications Letters, vol. 5, no. 3, pp. 92–94, Mar. 2001.

[9] A. Saleh and R. Valenzuela, “A statistical model for indoor multipath propagation,” IEEE J. on Selected Areas in Communications, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[10] D. Devasirvatham, “Time delay spread measurements of wideband radio signals within a building,” Electronics Letters, vol. 20, no. 23, pp. 950–951, 8 1984.

[11] A. Acampora and J. Winters, “System applications for wireless indoor communications,” IEEE Communications Magazine, vol. 25, no. 8, pp. 11–20, Aug. 1987.

[12] A. Siamarou and M. Al-Nuaimi, “Multipath delay spread and signal level measurements for indoor wireless radio channels at 62.4 ghz,” in 2001 IEEE VTS 53rd Vehicular Technology Conf., 2001, pp. 454–458.

[13] R. Gibson and R. Murray, “Systems organization for multichannel cordless telephones,” in Int. Zurich Seminar on Digital Communicatio, 1982, pp. 59–61.

[14] J. Andersen, T. Rappaport, and S. Yoshida, “Propagation measurements and models for wireless communications channels,” IEEE Communications Magazine, vol. 33, no. 1, pp. 42–49, Jan. 1995.

[15] T. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, “A survey of various propagation models for mobile communication,” IEEE Antennas and Propagation Magazine, vol. 45, no. 3, pp. 51–82, Jun. 2003.

[16] M. Murotani and K. Kohriyama, “Representation by approximate numerical formulas of radio-relay antenna beam bending due to atmospheric refraction,” IEICE Trans., vol. E72- E, no. 2, pp. 101–103, Feb. 1989.

[17] J. Seybold, Introduction to RF Propagation. John Wiley & Sons, 2005.

[18] P. Papazian, “Basic transmission loss and delay spread measurements for frequencies between 430 and 5750 mhz,” IEEE transactions on antennas and propagation, vol. 53, no. 2, pp. 694–701, 2005.

[19] J. Van Rees, “Measurements of the wideband radio channel characteristics for rural, residential, and suburban areas,” IEEE Transactions on Vehicular Technology, vol. 36, no. 1, pp. 2–6, 1987.

[20] J. A. Wepman, J. R. Hoffman, and L. H. Loew, “Characterization of macrocellular pcs propagation channels in the 1850-1990 MHz band,” in Proceedings of 1994 3rd IEEE International Conference on Universal Personal Communications, 1994, pp. 165–170.

[21] E. S. Sousa, V. M. Jovanovic, and C. Daigneault, “Delay spread measurements for the digital cellular channel in toronto,” IEEE Transactions on Vehicular Technology, vol. 43, no. 4, pp. 837–847, 1994.

[22] D. Devasirvatham, “Radio propagation studies in a small city for universal portable communications,” in 38th IEEE Vehicular Technology Conference, 1988, pp. 100–104.

[23] D. Cox and R. Leck, “Distributions of multipath delay spread and average excess delay for 910-MHz urban mobile radio paths,” IEEE Transactions on Antennas and Propagation, vol. 23, no. 2, pp. 206–213, 1975.

[24] T. S. Rappaport, S. Y. Seidel, and R. Singh, “900-MHz multipath propagation measurements for us digital cellular radiotelephone,” IEEE Transactions on Vehicular Technology, vol. 39, no. 2, pp. 132–139, 1990.

[25] J.-P. De Weck, P. Merki, and R. W. Lorenz, “Power delay profiles measured in mountainous terrain (radiowave propagation),” in 38th IEEE Vehicular Technology Conference, 1988, pp. 105–112.

[26] T. Tanaka, S. Kozono, and A. Akeyama, “Urban multipath propagation delay characteristics in mobile communications,” Electronics and Communications in Japan (Part I: Communications), vol. 74, no. 8, pp. 80–88, 1991.

[27] Y. Azar, G. N. Wong, K. Wang, R. Mayzus, J. K. Schulz, H. Zhao, F. Gutierrez Jr, D. Hwang, and T. S. Rappaport, “28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in new york city.” in ICC, 2013, pp. 5143–5147.

[28] T. S. Rappaport, Y. Qiao, J. I. Tamir, J. N. Murdock, and E. Ben-Dor, “Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems,” in 2012 IEEE Radio and Wireless Symposium. IEEE, 2012, pp. 151–154.

[29] Phased array radiation pattern, Phased array radiation pattern.gif, By Maxter315 [CC BY-SA 4.0 (creativecommons.org/licenses/bysa/4.0)], from Wikimedia Commons.

This page titled 6.8: References is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.8: References** by Michael Steer is licensed CC BY-NC 4.0.

6.9: Exercises

1. An antenna only radiates 45% of the power input to it. The rest is lost as heat. What input power (in dBm) is required to radiate $30\ \mathrm{dBm}$?
2. The output stage of an RF front end consists of an amplifier followed by a filter and then an antenna. The amplifier has a gain of $27\ \mathrm{dB}$, the filter has a loss of $1.9\ \mathrm{dB}$, and of the power input to the antenna, 35% is lost as heat due to resistive losses. If the power input to the amplifier is $30\ \mathrm{dBm}$, calculate the following:
   a. What is the power input to the amplifier in watts?
   b. Express the loss of the antenna in dB.
   c. What is the total gain of the RF front end (amplifier + filter)?
   d. What is the total power radiated by the antenna in dBm?
   e. What is the total power radiated by the antenna in mW?
3. The output stage of an RF front end consists of an amplifier followed by a filter and then an antenna. The amplifier has a gain of $27\ \mathrm{dB}$, the filter has a loss of $1.9\ \mathrm{dB}$, and of the power input to the antenna, 45% is lost as heat due to resistive losses. If the power input to the amplifier is $30\ \mathrm{dBm}$, calculate the following:
   a. What is the power input to the amplifier?
   b. Express the loss of the antenna in decibels.
   c. What is the total gain of the RF front end (amplifier + filter)?
   d. What is the total power radiated by the antenna in dBm?
   e. What is the total power radiated by the antenna in milliwatts?
4. In the figure below there are two transmitters, Tx_1 and Tx_2, operating at the same power level, and one receiver, Rx. Tx_2 is an intentional transmitter and its signal is intended to be received at Rx. Tx_1 uses the same frequency channel as Tx_2, but it transmits an interfering signal. [Parallels Example 4.1.1]

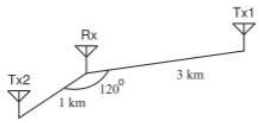


Figure $6.9.1$

   a. Assume that the antennas are omnidirectional and, being in a semiurban area, that the transmitted power density drops off as $1/d^3$, where d is the distance from the transmitter. Calculate the SIR at Rx. Express your answer in decibels.
   b. Now consider a directional antenna at Rx while the transmit antennas remain omnidirectional. The antenna at Rx is directed toward the transmitter Tx_2 and the antenna gain is $6\ \mathrm{dB}$. In the direction of Tx_1 the effective antenna gain is $-3\ \mathrm{dB}$. Now recalculate the SIR. Express your answer in dB.
5. Thirty five percent of the power input to an antenna is lost as heat, what is the loss of the antenna in dB.
6. Only 65% of the power input to an antenna is radiated with the rest lost to dissipation in the antenna, what is the gain of the antenna in dB? (This is not the antenna gain.)
7. The efficiency of an antenna is 66%. If the power input to the antenna is $10\ \mathrm{W}$ what is the power radiated by the antenna in dBm?
8. An antenna with an input of $1\ \mathrm{W}$ operates in free space and has an antenna gain of $12\ \mathrm{dBi}$. What is the maximum power density at $100\ \mathrm{m}$ from the antenna?
9. A transmitter has an antenna with an antenna gain of $10\ \mathrm{dBi}$, the resistive losses of the antenna are 50%, and the power input to the antenna is $1\ \mathrm{W}$. What is the EIRP in watts?
10. A transmitter has an antenna with an antenna gain of $20\ \mathrm{dBi}$, the resistive losses of the antenna are 50%, and the power input to the antenna is $100\ \mathrm{mW}$. What is the EIRP in watts?
11. An antenna with an antenna gain of $8\ \mathrm{dBi}$ radiates $6.67\ \mathrm{W}$. What is the EIRP in watts? Assume that the antenna is 100% efficient.
12. An antenna has an antenna gain of $10\ \mathrm{dBi}$ and a $40\ \mathrm{W}$ input signal. What is the EIRP in watts?

13. An antenna with 5 W of input power has an antenna gain of 20 dBi and an antenna efficiency of 25% and all of the loss is due to resistive losses in the antenna. [Parallels Example 4.5.2]
    a. How much power in dBm is lost as heat in the antenna?
    b. How much power in dBm is radiated by the antenna?
    c. What is the EIRP in dBW?
14. An antenna with an efficiency of 50% has an antenna gain of 12 dBi and radiates 100 W. What is the EIRP in watts?
15. An antenna with an efficiency of 75% and an antenna gain of 10 dBi. If the power input to the antenna is 100 W,
    a. what is the total power in dBm radiated by the antenna?
    b. what is the EIRP in dBm?
16. On a resonant antenna a large current is established by creating a standing wave. The current peaking that thus results establishes a strong electric field (and hence magnetic field) that radiates away from the antenna. A typical dipole loses 15% of the power input to it as resistive (I^2R) losses and has an antenna gain of 10 dBi measured at 50 m. Consider a base station dipole antenna that has 100 W input to it. Also consider that the transmitted power density falls off with distance d as $1/d^3$. Hint, calculate the power density at 50 m. [Parallels Example 4.5.1]
    a. What is the input power in dBm?
    b. What is the power transmitted in dBm?
    c. What is the power density at 1 km? Express your answer as $\mathrm{W/m^2}$.
    d. What is the power captured by a receive antenna (at 1 km) that has an effective antenna aperture of 6 $\mathrm{cm^2}$? Express your answer in first dBm and then watts.
    e. If the background noise level captured by the antenna is 1 pW, what is the SNR in decibels? Ignore interference that comes from other transmitters.
17. A communication system operating at 2.5 GHz includes a transmit antenna with an antenna gain of 12 dBi and a receive antenna with an effective aperture area of 20 $\mathrm{cm^2}$. The distance between the two antennas is 100 m.
    a. What is the antenna gain of the receive antenna?
    b. If the input to the transmit antenna is 1 W, what is the power density at the receive antenna if the power falls off as $1/d^2$, where d is the distance from the transmit antenna?
    c. Thus what is the power delivered at the output of the receive antenna?
18. Consider a point-to-point communication system. Parabolic antennas are mounted high on a mast so that ground effects do not exist, thus power falls off as $1/d^2$. The gain of the transmit antenna is 20 dBi and the gain of the receive antenna is 15 dBi. The distance between the antennas is 10 km. The effective area of the receive antenna is 3 $\mathrm{cm^2}$. If the power input to the transmit antenna is 600 mW, what is the power delivered at the output of the receive antenna?
19. Consider a 28 GHz point-to-point communication system. Parabolic antennas are mounted high on a mast so that ground effects do not exist, thus power falls off as $1/d^2$. The gain of the transmit antenna is 20 dBi and the gain of the receive antenna is 15 dBi. The distance between the antennas is 10 km. If the power output from the receive antenna is 10 pW, what is the power input to the transmit antenna?
20. An antenna has an effective aperture area of 20 $\mathrm{cm^2}$. What is the antenna gain of the antenna at 2.5 GHz?
21. An antenna operating at 28 GHz has an antenna gain of 50 dBi. What is the effective aperture area of the antenna?
22. A 15 GHz receive antenna has an antenna gain of 20 dBi. If the power density at the receive antenna is 1 $\mathrm{nW/cm^2}$, what is the power at the output of the antenna? [Parallels Example 4.6.3]
23. A microstrip patch antenna operating at 2 GHz has an efficiency of 66% and an antenna gain of 8 dBi. The power input to the antenna is 10 W.
    a. What is the power, in dBm, radiated by the antenna?
    b. What is the equivalent isotropic radiated power (EIRP) in watts?
    c. What is the power density, in $\mu\mathrm{W/m^2}$, at 1 km if ground effects are ignored?
    d. Because of multipath effects, the power density drops off as $1/d^4$, where d is distance. What is the power density, in $\mathrm{nW/m^2}$, at 1 km if the power density is 100 $\mathrm{mW/m^2}$ at 10 m from the transmit antenna?
24. A communication system operating at 10 GHz uses a microstrip patch antenna as a transmit antenna and a dipole antenna as a receive antenna. The transmit antenna is directly connected to the transmitter and the output power of the transmitter is 30 W.

The transmit antenna has an antenna gain of 9 dBi and an antenna efficiency of 60%. The receive antenna has an antenna gain of 2 dBi and a radiation efficiency of 80%. The receive antenna is connected to a receiver by a 10 m long cable with a loss of 0.1 dB/m. The link between the transmit and receive antenna is sufficiently elevated that ground effects and multipath effects are insignificant.

a. What is the output power of the transmitter in dBm?
b. What is the EIRP in dBm?
c. The transmitted power will drop off as $1/d^n$ (d is distance). What is n?
d. What is the peak power density in $\mu\text{W/m}^2$ at 1 km?
e. What is the effective aperture size of the receive antenna in m^2?
f. If the radiated power density at the receive antenna is $\mu\text{W/m}^2$, what is the signal power at the output of the receive antenna in dBm?
g. What is the total cable loss in dB?
h. What is the power presented to the receiver in dBm?

25. A communication system operating at 10 GHz uses a microstrip patch antenna as a transmit antenna and a dipole antenna as a receive antenna. The transmit antenna is connected to the transmitter by a 20 m long cable with a loss of 0.2 dB/m and the output power of the transmitter is 30 W. The transmit antenna has an antenna gain of 9 dBi and an antenna efficiency of 60%. The link between the transmit and receive antenna is sufficiently elevated that ground effects and multipath effects are insignificant.

a. What is the output power of the transmitter in dBm?
b. What is the cable loss between the transmitter and the antenna?
c. What is the total power radiated by the transmit antenna in dBm?
d. What is the power lost in the antenna as resistive losses and spurious radiation? Express your answer in dBm.
e. What is the EIRP of the transmitter in dBm?
f. The transmitted power will drop off as $1/d^n$ (d is distance). What is n?
g. What is the peak power density in $\mu\text{W/m}^2$ at 1 km?

26. Stacked dipole antennas are often found at the top of cellphone masts, particularly for large cells and operating frequencies below 1 GHz. These antennas have an efficiency that is close to 90%. Consider an antenna that has 40 W of input power, an antenna gain of 10 dBi, and transmits a signal at 900 MHz.

a. What is the EIRP in watts?
b. If the power density drops as $1/d^3$, where d is the distance from the transmit tower, what is the power density at 1 km if the power density is 100 mW/m^2 at 10 m?

27. Consider an 18 GHz point-to-point communication system. Parabolic antennas are mounted on masts and the LOS between the antennas is just above the tree line. As a result, power falls off as $1/d^3$, where d is the distance between the antennas. The gain of the transmit antenna is 20 dBi and the gain of the receive antenna is 15 dBi. The antennas are aligned so that they are in each other's main beam. The distance between the antennas is 1 km. The transmit antenna is driven by a power amplifier with an output power of 100 W. The amplifier drives a coaxial cable that is connected between the amplifier and the transmit antenna. The cable loses 75% of its power due to resistive losses. On the receive side, the receive antenna is directly connected to a masthead amplifier with a gain of 10 dB and then a short cable with a loss of 3 dB before entering the receive base station.

a. Draw the signal path.
b. What is the loss and gain of the transmitter coaxial cable in decibels?
c. What percentage of the power input to the receive coaxial cable is lost in the receive cable?
d. Express the power of the transmit amplifier in dBW and dBm.
e. What is the propagation loss in decibels?
f. Determine the total power in watts delivered to the receive base station.

28. Consider a point-to-point communication system. Parabolic antennas are mounted high on a mast so that ground effects are minimal. Thus power density falls off as $1/d^{2.3}$, where d is the distance from the transmitter. The gain of the transmit antenna is 15 dBi and the gain of the receive antenna is 12 dBi. These antenna gains are normalized to a distance of 1 m. The distance

between the antennas is 15 km. The output power of the receive antenna must be 1 pW. The RF frequency is 2 GHz; treat the antennas as lossless.

a. What is the received power in dBm?
b. What is the path loss in decibels?
c. What is the link loss in decibels?
d. Using the link loss, calculate the input power, P_T, of the transmitter. Express the answer in dBm.
e. What is the aperture area of the receiver in square meters?
f. Determine the radiated power density at the receiver in terms of the transmitter input power. That is, if P_T is the power input to the transmit antenna, determine the power density, P_D, at the receive antenna where $P_D = xP_T$. What is x in units of m^{-2}?
g. Using the power density calculation and the aperture area, calculate P_T in watts.
h. What is P_T in dBm? This should be the same as the answer you calculated in (d).
i. What is the total power radiated by the transmit antenna in dBm?

29. Two identical antennas are used in a point-to-point communication system, each having a gain of 50 dBi. The system has an operating frequency of 28 GHz and the antennas are at the top of masts 100 m tall. The RF link between the antennas consists only of the direct line-of-sight path.

a. What is the effective aperture area of each antenna?
b. How does the power density of the propagating signal rolloff with distance.
c. If the separation of the transmit and receive antennas is 10 km, what is the path loss in decibels?
d. If the separation of the transmit and receive antennas is 10 km, what is the link loss in decibels?

30. A transmitter and receiver operating at 2 GHz are at the same level, but the direct path between them is blocked by a building and the signal must diffract over the building for a communication link to be established. This is a classic knife-edge diffraction situation. The transmit and receive antennas are each separated from the building by 4 km and the building is 20 m higher than the antennas (which are at the same height). Consider that the building is very thin. It has been found that the path loss can be determined by considering loss due to free-space propagation and loss due to diffraction over the knife edge.

a. What is the additional attenuation (in decibels) due to diffraction?
b. If the operating frequency is 100 MHz, what is the attenuation (in decibels) due to diffraction?
c. If the operating frequency is 10 GHz, what is the attenuation (in decibels) due to diffraction?

31. A hill is 1 km from a transmit antenna and 2 km from a receive antenna. The receive and transmit antennas are at the same height and the hill is 20 m above the height of the antennas. What is the additional loss caused by diffraction over the top of the hill? Treat the hill as a knife-edge and the operating frequency is 1 GHz.

32. A 1 GHz point-to-point link has two major transmission paths. One is a LOS path and the other includes reflection from the ground so that the power density of the transmitted signal rollsoff as $1/d^{2.5}$ where d is the distance from the transmit antenna. At 10 m the power density from the transmit antenna is 100 $\mathrm{mW/m}^2$.

a. What is the power density at 1 km.
b. If the receive antenna has an antenna gain of 30 dBi, what is the effective aperture area of the receive antenna?
c. What is the power of the signal at the output of the receive antenna?

33. Two identical antennas are used in a point-to-point communication system, each having a gain of 30 dBi. The system has an operating frequency of 14 GHz and the antennas are at the top of masts 100 m tall. The RF link between the antennas consists only of the direct LOS path.

a. What is the effective aperture area of each antenna?
b. How does the power density of the propagating signal rolloff with distance?
c. If the separation of the transmit and receive antennas is 10 km, what is the path loss? Ignore atmospheric loss.

34. The three main cellular communication bands are centered around 450 MHz, 900 MHz, and 2 GHz. Compare these three bands in terms of multipath effects, diffraction around buildings, object (such as a wall) penetration, scattering from trees and parts of trees, and ability to follow the curvature of hills. Complete the table below with the relative attributes: high, medium, and low.

	Characteristic	450 MHz	900 MHz	2 GHz
ic	Multipath			
ic	Scattering			
ic	Penetration			
ic	Following curvature			
ic	Range			
ic	Antenna size			
ic	Atmospheric loss			

Table 6.9.1

35. Describe the difference in multipath effects in a central city area compared to multipath effects in a desert. Your description should be approximately 4 lines long and not use a diagram.
36. Wireless LAN systems can operate at 2.4 GHz, 5.6 GHz, 40 GHz and 60 GHz. Contrast with explanation the performance of these schemes inside a building in terms of range.
37. At 60 GHz the atmosphere strongly attenuates a signal. Discuss the origin of this and indicate an advantage and a disadvantage.
38. Short answer questions. Each part requires a short paragraph of about five lines and a figure, where appropriate, to illustrate your understanding.
    a. Cellular communications systems use two frequency bands to communicate between the basestation and the mobile unit. The bands are generally separated by 50 MHz or so. Which band (higher or lower) is used for the downlink from the basestation to the mobile unit and what are the reasons behind this choice?
    b. Describe at least two types of interference in a cellular system from the perspective of a mobile handset.
39. The three main cellular communication bands are centered around 450 MHz, 900 MHz, and 2 GHz. Compare these three bands in terms of multipath effects, diffraction around buildings, object (such as a wall) penetration, scattering from trees and parts of trees, and the ability to follow the curvature of hills. Use a table and indicate the relative attributes: high, medium, and low. WAS 10(C)
40. Describe Rayleigh fading in approximately 4 lines and without using a diagram.
41. In several sentences and using a diagram describe Rayleigh fading and the impact it has on radio communications.
42. A transmit antenna and a receive antenna are separated by 1 km and operate at 1 GHz. What is the radius of the first Fresnel zone at 0.5 km from each antenna? [Parallels Example 4.6.4]
43. A transmit antenna and a receive antenna are separated by 40 km and operate at 10 GHz. [Parallels Example 4.6.4]
    a. What is the radius of the first Fresnel zone at the midpoint between the antennas?
    b. What is the radius of the second Fresnel zone?
    c. To ensure LOS propagation, what should the clearance be from the direct line between the antennas and obstructions such as hills?
44. A transmitter and receiver operate at 100 MHz, are at the same level, and are separated by 4 km. The signal must diffract over a building half way between the antennas that is 20 m higher than the direct path between the antennas. What is the attenuation (in decibels) due to diffraction?
45. A transmitter and receiver operate at 10 GHz, are at the same level, and are 4 km apart. The signal must diffract over a building that is half way between the antennas and is 20 m higher than the line between the antennas. What is the attenuation (in dB) due to diffraction?
46. The path from a transmit antenna to a receive antenna is elevated so that ground and multipath effects are insignificant. The power radiated by the transmit antenna drop off as $1/d^n$ where d is distance, what is n?
47. A communication system has a power density roll-off of $1/d^{2.5}$ between a transmit antenna and a mobile receive antenna which are separated by 10 km. At 10 m from the transmit antenna, the power density is 10 $\mathrm{W/m^2}$. What is the power density at the receive antenna? [Parallels Example 4.6.1]

48. A 900 MHz communication system uses a transmit antenna with an antenna gain G_T of 3 dB and a receive antenna with an antenna gain G_R of 0 dB. If the distance between the antennas is 200 m, what is the link loss from the input to the transmit antenna and the output of the receive antenna if the power density reduces as $1/d^{2.5}$? [Parallels Example 4.6.2]
49. A transmitter has a power density of $100\ \mathrm{mW/m^2}$ at a distance of 1 m from the transmitter. The power density falls off as 33 dB per decade of distance. At what distance from the transmit antenna will the power density reach $1\ \mu\mathrm{W}\cdot\mathrm{m}^{-2}$? [Parallels Example 4.6.3]
50. Describe at least two types of interference in a cellular system from the perspective of a mobile handset.
51. In a cellular system, a signal is intentionally transmitted from a base station nominally located in the center of a cell to a mobile unit in the same cell. However, nearby transmitters using the same channel cause interference. In Figure 4.8.4, a mobile unit is located at the edge of a cell and uses frequency channel A. Many nearby transmitters also operate using channel A and the six nearest transmitters can be considered as causing significant interference. Consider that the mobile unit is a distance r from its cell's transmitter along the line connecting two channel A base stations, that the transmitters all operate at the same power level, and that the distance between base stations operating using channel A is $3r$. This three-cell cluster operates in a suburban area and the power density drops off with distance d as $1/d^{3.5}$ due to multipath effects. What is the SIR at the mobile unit? Express your answer in decibels. [Parallels Example 4.8.1]
52. A cellular system uses a three-cell cluster.
    a. Treating cells as equal sized hexagons with towers in the center of each cell, draw the cell map including all cells within 3 cell diameters of the main channel. Label the main cell and other cells using the same frequencies as A.
    b. Consider a mobile unit at the edge of a cell transmitting the intended signal from the tower at the center of that cell. Identify the interfering towers.
    c. If ground effects and multipath effects are negligible, what is the power roll-off factor if the distance between a tower and the mobile unit is d? That is, what is n if power falls off as $1/d^n$?
    d. If trisector antennas are used, identify the interfering cells and approximately determine the improvement in SIR compared to using a nonsectorized antenna? You do not need to do detailed calculations.
53. Describe trisector antennas in 4 lines and without using a diagram.

6.9.1: Exercises By Section

†challenging, ‡very challenging

§14.31†

§14.52†, 3†, 4‡, 5, 6, 7†, 8†, 9†, 10, 11, 12, 13, 14, 15, 16‡, 17, 18, 19, 22

§14.623‡, 24‡, 25‡, 26‡, 27‡, 28‡, 29†, 30†, 31†, 32‡, 33†, 34‡, 35†, 36, 37, 38†, 39, 40, 41, 42†, 43, 44, 45, 46, 47, 48, 49

§14.850†, 51‡, 52‡, 53†

6.9.2: Answers to Selected Exercises

2. (e) 53.23 dBm

17. $0.251\ \mu$W
18. 14.3 pW

23. (b) 63.1 W

25. (e) 49.77 dBm

27. (f) 708 pW
28. 107.65 dB

32. $7.16\ \mathrm{m^2}$

42. 8.66 m

51. 7.33 dB

This page titled 6.9: Exercises is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.9: Exercises** by Michael Steer is licensed CC BY-NC 4.0.

6.10: Multipaths and Delay Spread

Multipath was discussed previously in Sections 4.6.1 and 4.6.3 in the context of reduction in signal amplitude due to the destructive interference of a signal traveling on several paths. Another effect of multipath is that instances of a signal traveling from a transmitter to a receiver on different paths will arrive at a receiver with various delays. These delays can differ by tens or thousands of nanoseconds which is much longer than the period of a microwave signal, e.g. a 1 GHz signal has a period of 1 ns.

6.10.1: Delay Spread

Each of the paths supports what is called a signal instance. For a single-tone signal the effect of various delays is for the signal instances to arrive at the receiver with very different phases. As well, the signal instances invariably have different amplitudes resulting in a composite received signal that is an unpredictable combination of constructive and destructive interference. Sometimes the constructive interference will result in a stronger composite received signal but this is not that important. What is much more significant is the destructive interference which will result in signal fades with the signal received being very small at times. Consider two signal instances $y_1(t)$ and $y_2(t)$, traveling from a transmitter to a receiver on paths 1 and 2 respectively. If $y_1(t)$ and $y_2(t)$ have the same amplitude and phase then the combined signal will only be increased by 6 dB over a signal that travels on just one path. However if $y_1(t)$ and $y_2(t)$ have the same amplitude but phases that differ by 180° then there will be total cancelation and no composite signal will be received. So destructive interference is much more significant than constructive interference.

For cellular radio, in a rural area there can be two or three significant paths, usually taken as paths having signal instances that are within 20 dB of the largest signal instance. In an urban environment there can be tens of significant paths because there are many reflections from buildings and often there is not an LOS path which would otherwise have the largest signal instance. With multiple paths it is unlikely that there will be total cancelation of the received signal but it is very likely that the received signal will be smaller than if there was one path.

Early cellular radio had narrow bandwidths, e.g. 200 kHz for 2G's GSM cellular system, and low-order modulation, e.g. 2 bits per symbol and it was sufficient to use the phase-based concepts of constructive and destructive interference to understand the effects of multipath. As channel bandwidths increased and the order of modulation increased it was necessary to consider the actual physical effect of the paths having different delays.

If there is an LOS path, say with no building in the way, then the LOS signal instance will be stronger than a signal instance that travels on any other path. It will also have the smallest delay. This is partly because each non-LOS (NLOS) path will travel further and hence spread out more, and also because there will be signal loss at reflections. If there is scattering or diffraction on a path the amplitude of this signal instance on that path will be even smaller. Even if there is an LOS path, if there are multiple NLOS paths then the combined NLOS signal instances could indeed be larger than the LOS signal instance. If there is not an LOS path then one of the NLOS paths will become

Frequency (GHz)	**Tx-Rx distance** (km)	**Environment**	**max. excess delay** $\tau_{d,\,\max}$ (ns)	**median excess delay** $\tau_{d,\,\text{med}}$ (ns)	
0.43	3.8†	urban	1300	900	[18]
0.90	2.2	rural	800	–	[19]
0.90	5	rural	1900	–	[20]
0.90	2	suburban	3700	300	[21]
0.90	2.2	suburban	900	300	[19]
0.90	0.6	urban	700	200	[22]
0.90	3.5	urban	5000	1100	[23]
0.90	7.0	urban	15000	1900	[24]
0.90	1.0	urban	2000	700	[21]
0.90	13	mountain$	3800	–	[25]
0.90	6.0	mountain$	1800	–	[26]
1.35	3.8†	urban	1400	850	[18]
2.26	3.8†	urban	1400	800	[18]
5.75	3.8†	urban	1000	300	[18]
28	0.052	urban LOS	754	< 200	[27]
28	0.097	urban NLOS	1388	200	[27]
38	0.2†	urban LOS	12	1.5	[28]
38	0.2†	urban NLOS	133	14	[28]

Table 6.10.1: Measured excess delays for various carrier frequencies, Tx-Rx distances and multipath environments. Data has been normalized where needed so that $\tau_{d,\,\max}$, is the maximum excess delay that is exceeded only 1% of the time. †d ranges from 0.04 km to 3.8 km and here $\tau_{d,\,\max}$ is the maximum delay exceeded only 10% of the time. ‡ high gain steerable antennas. $ urban area surrounded by mountains

the largest received signal instance and it may not have the shortest delay. Generally only signal instances that are within 20 dB of the largest signal instance must be considered and then the delay, also referred to as 'delay (20 dB)', that matters is the excess delay, τ_d, which is the difference between the earliest and latest arriving signal instances within that 20 dB window. It is essential that a communications link be maintained a certain percentage of the time, e.g. 99%, for a specified level of service. This minimizes the number of times packets must be re-transmitted and assures acceptable operation of a system. Thus the maximum excess delay, $\tau_{d,\,\max}$ becomes an important metric and determines the interval for which the received signal is not valid, the required guard band to prevent signals from an earlier symbol interval overlapping with the current symbol interval, and also the maximum useful bandwidth of a signal.

It is found that the delay spread increases with distance between the transmitter and receiver, the Tx-RX distance, increases with frequency, and increases with a richer multipath environment. For example, a rural area has few paths but large cell sizes are used and the Tx-Rx distance is relatively large resulting in a large excess delay but a dense urban area has small cells but a rich multipath environment with many reflections from buildings. Measured excess delays for various situations are given in Table 6.10.1. There is a lot of variation in excess delays but it is clearly reduced with smaller cells and a less rich multipath environment.

It is also found that as a mobile unit moves the excess delay and the number of significant paths varies tens or hundreds of times per second depending on the environment but also the mobility, e.g. pedestrian versus highway speeds. However over a millisecond

the propagation is found to be effectively fixed and consequently systems are designed with data sent in packets that are usually no longer than a millisecond.

Examining Table 6.10.1 further it is observed that the 2G and 3G cellular bands are below 2 GHz and the delay spreads can be several milliseconds and even larger with larger Tx-Rx distances in an urban area. At the higher microwave frequencies up to 6 GHz covering the operating frequency range of 4G, the delay spread is smaller.

The 28 GHz and 38 GHz bands are where 5G operates with short Tx-Rx distances. Here the number of paths above 20 dB is around 4 to 8 for Tx-Rx separations of 35 to 193 m [27, 28].

The spreads are relatively low and this is largely a result of using high gain steerable antennas which are essential to 5G operation. The high gain antennas reduce the spread of multipath path lengths, and hence reduce the excess delay spread, concentrating the reflected paths close to the transmit and receive antennas.

The prior discussions on multipath effects and the RF link consider a single frequency tone propagating from a transmit antenna and a receive antenna. When a single-tone signal travels on multiple paths there will be destructive and/or constructive interference.

6.10.2: Intersymbol Interference

Intersymbol interference (**ISI**) occurs when a symbol traveling on one path interferes with the symbol traveling on another path. This is the result of the paths having different delays. The effect of excess delay on the received composite signal is indicated in Figure 6.10.1. Consider the situation where there are four paths with the excess delay of path 1 being 0 by definition, i.e. $\tau_{d,1} = 0$, the excess delay of path 2, $\tau_{d,2} = 100$ ns, and so on, $\tau_{d,3} = 200$ ns, and $\tau_{d,4} = 300$ ns. The 2G–5G cellular systems transmit a series of symbols, here A, B, C, D, and E, which have a symbol duration which differs by standard. In Figure 6.10.1 four paths are considered and the four signal instances each travel on a different path and are combined in the receiver's antenna to yield a composite received signal. The symbols of each of the signal instances overlap causing ISI.

With the 2G GSM system symbol lengths are 3690 ns and the first 300 ns of a symbol instance traveling on the first path to arrive has interference from the previous symbol traveling on the other three paths, see Figure 6.10.1(a). In this case the first 300 ns of the 3690 ns of the symbol is garbled. In this example the delay spread is just 300 ns but this is conservative. Table 6.10.1 indicates that in some environments that the excess delay spread 1% of the time can be as long as the symbol interval for 2G/GSM (operating below 2 GHz). So even if the signal strength is high reception may not be possible because of ISI.

With 3G the packet length is 260 ns so with an excess delay spread of 300 ns for the four path example the symbol interference is severe, see Figure 6.10.1(b), and as noted 300 ns is a very conservative estimate of the excess delay spread. One of the features of 3G is a method for separating out the individual signal instances. Only when there are a very large number of significant paths does 3G have problems. In an urban area sometimes a communication link cannot be established with 2G and 3G even though the signal strength is high.

The 4G system minimizes the problem of ISI by having a very long symbol interval of 71.35 mus and a 300 ns excess delay spread is a small fraction of the symbol length. The symbol interval is considerably longer than any of the maximum delay spreads given in Table 6.10.1. Also the 4G standard

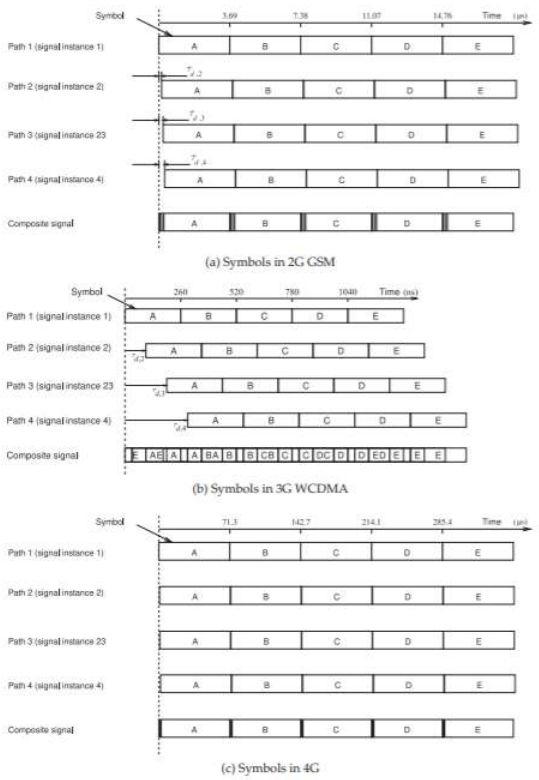


Figure 6.10.1: The effect of excess delay on symbol interference when there are four paths. The excess delay of each path, $\tau_{d,x}$, increases by 100 ns. Five symbols are shown: A, B, C, D, and E. The 2G GSM standard has a symbol duration of 3.6928 μs, the 3G WCDMA standard has a symbol duration of 260 ns, and the 4G OFDM standard has a symbol duration of 71.35 μs. The situation with 5G is similar to that for 4G.

has a guard time band known as a cyclic prefix. (The cyclic prefix is a bit more than a guard band as it involves repeating the end of symbol but this extra feature will be discussed latter.) The cyclic prefix can be either 4.7 μs, 5.2 μs, 16.7 μs, or 33.3 μs accommodating various excess delay spreads. This is more than enough to avoid the problem of ISI as any of the maximum excess delays listed in Table 6.10.1 are less than the maximum cyclic prefix length. The long symbol interval implies a very narrow bandwidth or subchannel bandwidth. A 4G system communicates with a user using a very large number of sub-channels thus supporting high data rates.

6.10.3: Summary

Various excess delays on different paths causes intersymbol interference. High levels of intersymbol interference results in the failure to establish a communication link. Excess delay is a fundamental limitation with the 2G/GSM system and the only solution is to use very small cells in urban environments which have many significant signal paths. The 3G system employs a coding technique that enables the first few different paths to be resolved thus limiting the problem of ISI but not eliminating it. The 4G system, and 5G operates similarly, employs a long guard time, the cyclic prefix, to avoid ISI completely.

This page titled 6.10: Multipaths and Delay Spread is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.10: Multipaths and Delay Spread** by Michael Steer is licensed CC BY-NC 4.0.

6.11: 4.8 Radio Link Interference

Various excess delays on different paths causes intersymbol interference. High levels of intersymbol interference results in the failure to establish a communication link. Excess delay is a fundamental limitation with the 2G/GSM system and the only solution is to use very small cells in urban environments which have many significant signal paths. The 3G system employs a coding technique that enables the first few different paths to be resolved thus limiting the problem of ISI but not eliminating it. The 4G system, and 5G operates similarly, employs a long guard time, the cyclic prefix, to avoid ISI completely

6.11.1: Frequency Reuse Plan

Radio systems using the same channel are geographically spaced to control interference. In a cellular system, the coverage areas are arranged in cells represented by the hexagons in Figure 6.11.1. The actual shape of the cells is influenced by obstructions such as hills and buildings. In 1G and 2G cellular systems the cells are arranged in clusters such as the $3-$, $7-$, and 12-cell clusters shown, and the total number of channels available is divided among the cells in a cluster with the full set of frequency channels repeated in each cluster. (More about 3G–5G arrangements later.) The size of the clusters is the major component of what is called the **frequency reuse** plan. So a three-cell cluster has a spacing of approximately one cell diameter to the next cell using the same frequency channels. The signal level transmitted from the original cell will interfere with the signal in its corresponding cell in adjacent clusters. The level of interference is reduced with 7-cell and then 12-cell clusters. There is also background noise coming from cosmic sources as well as artificial sources, but in a cellular system, interference from other radios operating in the same system nearly always dominates.

The use of directional antennas at the base station increases the SIR. The interference pattern obtained using a **trisector antenna** (each segment of the antenna providing 120° of coverage) is shown in Figure 6.11.2. The trisector antenna can be arranged so that the transmit and receive antennas are

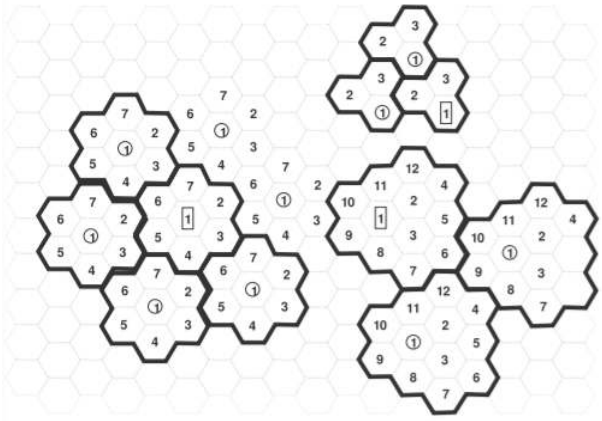


Figure 6.11.1: Cells arranged in clusters. Shown are $3-$ cell clusters (top right), $7-$ cell clusters (bottom left), and $12-$cell clusters (bottom right).

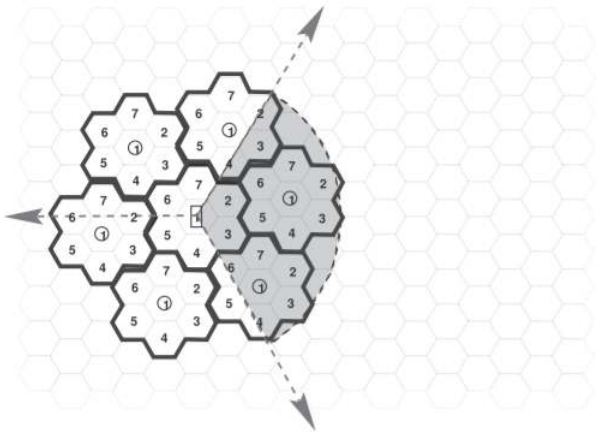


Figure 6.11.2: Cells in a cellular radio system with a trisector antenna, such as in Figure 6.11.3, showing the area of possible interference as a shaded region.

separated (see Figure 6.11.3), and can also be arranged to tilt the coverage toward the ground (e.g., see the antenna in Figure 4.1.2(l)). It is also possible to use higher-order antenna sectoring and to have smaller cells (achieved possibly by relatively low-power transmissions) to provide higher levels of coverage at critical regions such as intersections of roads. These smaller cells are referred to as microcells or picocells.

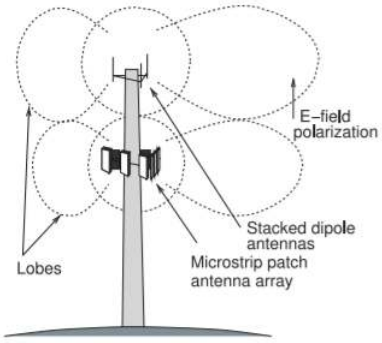


Figure 6.11.3: Base station with trisector stacked-dipole transmit and microstrip patch receive antennas.

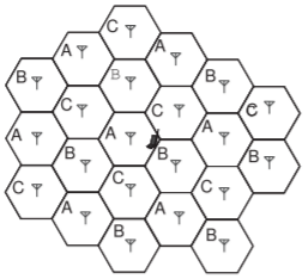


Figure 6.11.4: Three-cell cluster with mobile unit at the edge of cells A, B, and C.

Example 6.11.1: Cellular Interference

In a cellular system, a signal is intentionally transmitted from a base station nominally located in the center of a cell to a mobile unit in the same cell. However, nearby transmitters using the same channel cause interference. In Figure 6.11.4, a mobile unit is located at the edge of a cell and uses frequency channel A. Many nearby transmitters also operate using channel A and the six nearest transmitters can be considered as causing significant interference. Consider that the mobile unit is a distance r from its cell's transmitter along the line connecting two channel A base stations, that the transmitters all operate at the same power level, and that the distance between base stations operating using channel A is $3r$. This three-cell cluster operates in a suburban area and the power density drops off with distance d as $1/d^3$ due to multipath effects. What is the SIR at the mobile unit?

Solution

There are seven close towers transmitting signals on the same channel. Call these A_1–A_7 (M = mobile unit, A = transmitters). A_1 is the desired signal and A_2–A_7 are interferers. The distances from the transmitters to the mobile unit are d_1–d_7, and the powers from the transmitters received at the mobile unit are P_1–P_7.

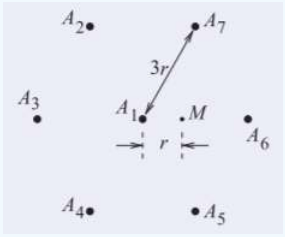


Figure 6.11.5

Now SIR $= \frac{P_1}{P_2+P_3+P_4+P_5+P_6+P_7}$

This problem requires d_2–d_7 to be determined.

Consider the diagram to the right with $d_1 = r$. Also, the distance from A_2 to A_7 is $3r$. Now $d_3 = 3r + r = 4r$, $d_6 = 3r - r = 2r$. So $d_2 = \left[(X+r)^2 + Y^2\right]^{\frac{1}{2}} = (2.5^2 + 2.6^2)r = 3.607r = d_4$. ($X = 3r\sin 30^\circ = 1.5r$ and $Y = 3r\cos 30^\circ = 2.6r$.) Similarly $d_7 = d_5 = \left[(1.5-1)^2 + 2.6^2\right]^{\frac{1}{2}} = 2.648r$.

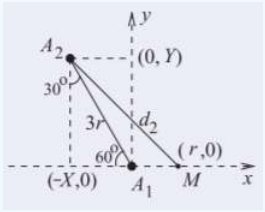


Figure 6.11.6

$$\frac{P_2}{P_1} = \frac{d_1}{d_2} = \frac{1}{3.607^3} = 0.0213 = \frac{P_4}{P_1} \qquad \frac{P_3}{P_1} = \frac{d_1}{d_3} = \frac{1}{4^3} = 0.0156$$
$$\frac{P_7}{P_1} = \frac{d_1}{d_7} = \frac{1}{2.648^3} = 0.0539 = \frac{P_5}{P_1} \qquad \frac{P_6}{P_1} = \frac{d_1}{d_6} = \frac{1}{2^3} = 0.1250$$

$$\text{SIR} = (0.0213 + 0.0156 + 0.0213 + 0.0539 + 0.1250 + 0.0539)^{-1} = 3.44 = 5.36 \text{ dB}$$

6.11.2: Summary

The clustering of cells as described above applies most directly to 1G and 2G cellular systems. The 3G system does not use clusters of cells but instead uses high frequency codes to separate users. There is a system awareness of managing interference but more complicated than the clustering arrangements. The 4G and 5G systems take frequency reuse algorithms to a whole new level performing global minimization of interference of wide areas. Still there is the basic concept that radios in adjacent cells operating on the same frequency channel can interfere with each other and careful attention must be given to frequency reuse.

The region of signal coverage by a basestation is greatly influenced by terrain and is not a neat hexagon. See for example the coverage map for a rural area shown in Figure 4.9.1(a). To ensure adequate coverage basestations must be placed on a terrain-aware grid and not a regular grid. A wasted expense in operating a cellular system would be installing more basestations than are needed for optimum coverage. Thus cellular service providers simulate coverage before installing basestations. A significant characteristic of urban coverage as shown in Figures 4.9.1(b and c) is the "**urban canyon effect**" whereby coverage is concentrated along roads with signals bouncing off buildings directing signal propagation down the roadway. Even areas close to a transmitter but behind buildings are shadowed while distant areas near roads can have good coverage. This effect is also known as the urban **wave-guiding effect** or **wave-channeling effect**. The signal blocking by buildings is shown in Figure 4.9.1(e) where it is seen that there is some diffraction over buildings but shadowing is severe after two or more diffraction events. The indoor environment also affects coverage as shown in Figure 4.9.1(d) for WiFi.

This page titled 6.11: 4.8 Radio Link Interference is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.11: 4.8 Radio Link Interference** by Michael Steer is licensed CC BY-NC 4.0.

6.12: Antenna Array

An antenna array comprises multiple radiating elements, i.e. individual antennas, and focuses a transmit beam in a desired direction. In Figure 6.12.2 the field pattern in the plane of the earth produced by an array of 30 antenna elements arranged horizontally is shown. The fields from each

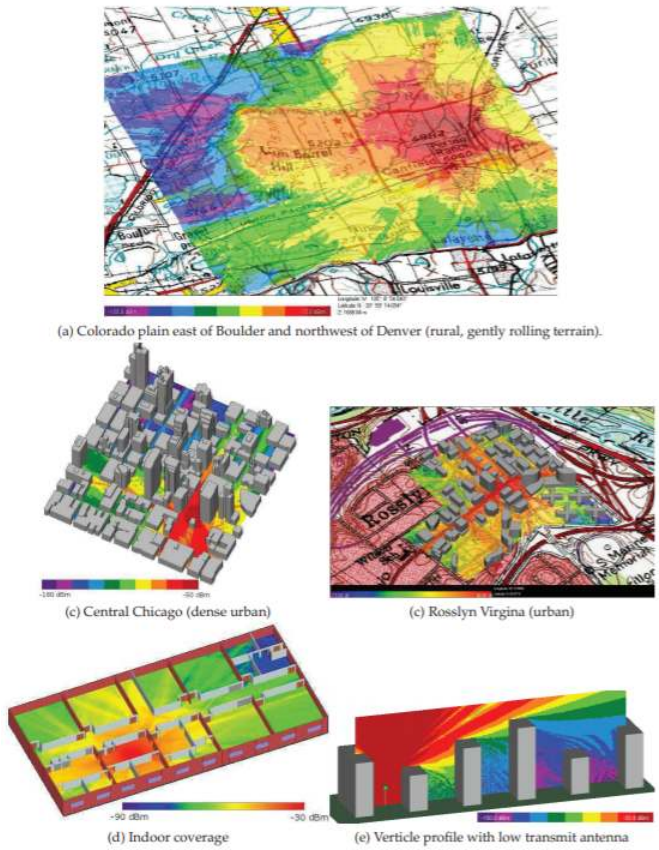


Figure 6.12.1: Radio coverage profiles with 900 MHz transmitter except for (d) which is for 2.45 GHz. Calculated using Wireless Insite®. Copyright Remcom, Inc. Used with permission.

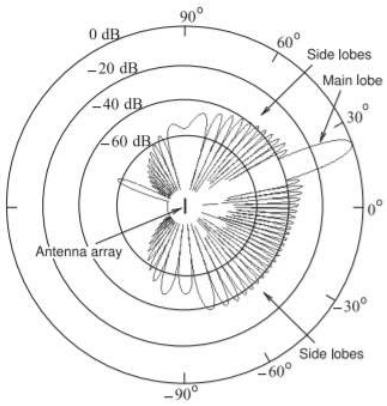


Figure 6.12.2: Electric field pattern from a 30 element array of antennas spaced 0.65λ apart. The sidelobe levels are about 40 dB below the power level of the main lobe. The same signal is presented to the antenna elements except that phases of the signal at each antennas is adjusted to produce a main beam directed at 20 degrees. The signals to each antenna are thus correlated. After [29].

antenna element combined to narrow and strengthen the main beam. Side (or grating) lobes are produced and these will result in some interference but their level here is 40 dB below that of the main beam. This is another way of managing interference in a

cellular system but the direction to the mobile unit must be known. Antenna arrays are used in 5G. The affect of the array is to increase the power density of the main beam and the density relative to that of an isotropic antenna is called the directional gain of the array, D_{Array}, and is the product of the antenna gain, G_A, of an individual antenna element and the array gain, G_{Array}:

$$D_{\text{Array}} = G_{\text{Array}} G_A \tag{6.12.1}$$

The maximum value of G_{Array} is N for an N element array. So the maximum directional gain of the array in dBi is

$$D_{\text{Array}}|_{\text{dBi}} = G_A|_{\text{dBi}} + 10 \log N \tag{6.12.2}$$

This page titled 6.12: Antenna Array is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Chris Lee.

- **4.12: Antenna Array** by Michael Steer is licensed CC BY-NC 4.0.

Index

L

L network
- 3.4: The L Matching Network

R

resonant antennas
- 6.3: Resonant Antennas

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Telecommunications

Webpages: 80

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC 4.0: 97.5% (78 pages)
- Undeclared: 2.5% (2 pages)

By Page

- Telecommunications - *CC BY-NC 4.0*
  - Front Matter - *CC BY-NC 4.0*
    - TitlePage - *CC BY-NC 4.0*
    - InfoPage - *CC BY-NC 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY-NC 4.0*
  - 1: Passive Components - *CC BY-NC 4.0*
    - 1.1: Introduction - *CC BY-NC 4.0*
    - 1.2: Passive components at high frequency - *CC BY-NC 4.0*
    - 1.3: Surface-Mount Components - *CC BY-NC 4.0*
    - 1.4: Summary - *CC BY-NC 4.0*
  - 2: Transmission Lines - *CC BY-NC 4.0*
    - 2.1: Introduction - *CC BY-NC 4.0*
    - 2.2: Transmission Line Characteristics - *CC BY-NC 4.0*
    - 2.3: The Lossless Terminated Line - *CC BY-NC 4.0*
    - 2.4: Summary - *CC BY-NC 4.0*
    - 2.5: References - *CC BY-NC 4.0*
    - 2.6: Exercises - *CC BY-NC 4.0*
  - 3: Impedance Matching - *CC BY-NC 4.0*
    - 3.1: Introduction - *CC BY-NC 4.0*
    - 3.2: Matching Networks - *CC BY-NC 4.0*
    - 3.3: Impedance Transforming Networks - *CC BY-NC 4.0*
    - 3.4: The L Matching Network - *CC BY-NC 4.0*
    - 3.5: Dealing with Complex Loads - *CC BY-NC 4.0*
    - 3.6: Dealing with Complex Loads - *CC BY-NC 4.0*
    - 3.7: Summary - *CC BY-NC 4.0*
    - 3.8: References - *CC BY-NC 4.0*
    - 3.9: Exercises - *CC BY-NC 4.0*
    - 3.10: Impedance Matching Using Smith Charts - *CC BY-NC 4.0*
    - 3.11: Distributed Matching - *CC BY-NC 4.0*
    - 3.12: Matching Using the Smith Chart - *CC BY-NC 4.0*
  - 4: Filters - *CC BY-NC 4.0*
    - 4.1: Introduction - *CC BY-NC 4.0*
    - 4.2: Singly and Doubly Terminated Networks - *CC BY-NC 4.0*
    - 4.3: The Lowpass Filter Prototype - *CC BY-NC 4.0*
    - 4.4: The Maximally Flat (Butterworth) Lowpass Approximation - *CC BY-NC 4.0*
    - 4.5: The Chebyshev Lowpass Approximation - *CC BY-NC 4.0*
    - 4.6: Element Extraction - *CC BY-NC 4.0*
    - 4.7: Butterworth and Chebyshev Filters - *CC BY-NC 4.0*
    - 4.8: Impedance and Admittance Inverters - *CC BY-NC 4.0*
    - 4.9: Filter Transformations - *CC BY-NC 4.0*
    - 4.10: Cascaded Line Realization of Filters - *CC BY-NC 4.0*
    - 4.11: Butterworth and Chebyshev Bandpass Filters - *CC BY-NC 4.0*
    - 4.12: Richard's Transformation - *CC BY-NC 4.0*
    - 4.13: Kuroda's and Norton's Network Identities - *CC BY-NC 4.0*
    - 4.14: Inter-resonator Coupled Bandpass Filters - *CC BY-NC 4.0*
    - 4.15: Bandpass Filter Topologies - *CC BY-NC 4.0*
    - 4.16: Case Study- Design of a Bandstop Filter - *CC BY-NC 4.0*
    - 4.17: Active Filters - *CC BY-NC 4.0*
    - 4.18: Transient Response of a Bandpass Filter - *CC BY-NC 4.0*
    - 4.19: Summary - *CC BY-NC 4.0*
    - 4.20: References - *CC BY-NC 4.0*
    - 4.21: Exercises - *CC BY-NC 4.0*
  - 5: Oscillators - *CC BY-NC 4.0*

- 5.1: Introduction - *CC BY-NC 4.0*
- 5.2: Oscillator Theory - *CC BY-NC 4.0*
- 5.3: Reflection Oscillator - *CC BY-NC 4.0*
- 5.4: Case Study- Reflection Oscillator - *CC BY-NC 4.0*
- 5.5: Advanced Discussion of Oscillator Noise - *CC BY-NC 4.0*
- 5.6: Summary - *CC BY-NC 4.0*
- 5.7: References - *CC BY-NC 4.0*
- 5.8: Exercises - *CC BY-NC 4.0*

6: Antennas and the RF Link - *CC BY-NC 4.0*

- 6.1: Introduction - *CC BY-NC 4.0*
- 6.2: RF Antennas - *CC BY-NC 4.0*
- 6.3: Resonant Antennas - *CC BY-NC 4.0*
- 6.4: Traveling-Wave Antennas - *CC BY-NC 4.0*
- 6.5: Antenna Parameters - *CC BY-NC 4.0*
- 6.6: The RF Link - *CC BY-NC 4.0*
- 6.7: Summary - *CC BY-NC 4.0*
- 6.8: References - *CC BY-NC 4.0*
- 6.9: Exercises - *CC BY-NC 4.0*
- 6.10: Multipaths and Delay Spread - *CC BY-NC 4.0*
- 6.11: 4.8 Radio Link Interference - *CC BY-NC 4.0*
- 6.12: Antenna Array - *CC BY-NC 4.0*

Back Matter - *CC BY-NC 4.0*

- Index - *CC BY-NC 4.0*
- Glossary - *CC BY-NC 4.0*
- Detailed Licensing - *CC BY-NC 4.0*
- Detailed Licensing - *Undeclared*

Detailed Licensing

Overview

Title: Telecommunications

Webpages: 80

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC 4.0: 97.5% (78 pages)
- Undeclared: 2.5% (2 pages)

By Page

- Telecommunications - *CC BY-NC 4.0*
  - Front Matter - *CC BY-NC 4.0*
    - TitlePage - *CC BY-NC 4.0*
    - InfoPage - *CC BY-NC 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY-NC 4.0*
  - 1: Passive Components - *CC BY-NC 4.0*
    - 1.1: Introduction - *CC BY-NC 4.0*
    - 1.2: Passive components at high frequency - *CC BY-NC 4.0*
    - 1.3: Surface-Mount Components - *CC BY-NC 4.0*
    - 1.4: Summary - *CC BY-NC 4.0*
  - 2: Transmission Lines - *CC BY-NC 4.0*
    - 2.1: Introduction - *CC BY-NC 4.0*
    - 2.2: Transmission Line Characteristics - *CC BY-NC 4.0*
    - 2.3: The Lossless Terminated Line - *CC BY-NC 4.0*
    - 2.4: Summary - *CC BY-NC 4.0*
    - 2.5: References - *CC BY-NC 4.0*
    - 2.6: Exercises - *CC BY-NC 4.0*
  - 3: Impedance Matching - *CC BY-NC 4.0*
    - 3.1: Introduction - *CC BY-NC 4.0*
    - 3.2: Matching Networks - *CC BY-NC 4.0*
    - 3.3: Impedance Transforming Networks - *CC BY-NC 4.0*
    - 3.4: The L Matching Network - *CC BY-NC 4.0*
    - 3.5: Dealing with Complex Loads - *CC BY-NC 4.0*
    - 3.6: Dealing with Complex Loads - *CC BY-NC 4.0*
    - 3.7: Summary - *CC BY-NC 4.0*
    - 3.8: References - *CC BY-NC 4.0*
    - 3.9: Exercises - *CC BY-NC 4.0*
    - 3.10: Impedance Matching Using Smith Charts - *CC BY-NC 4.0*
    - 3.11: Distributed Matching - *CC BY-NC 4.0*
    - 3.12: Matching Using the Smith Chart - *CC BY-NC 4.0*
  - 4: Filters - *CC BY-NC 4.0*
    - 4.1: Introduction - *CC BY-NC 4.0*
    - 4.2: Singly and Doubly Terminated Networks - *CC BY-NC 4.0*
    - 4.3: The Lowpass Filter Prototype - *CC BY-NC 4.0*
    - 4.4: The Maximally Flat (Butterworth) Lowpass Approximation - *CC BY-NC 4.0*
    - 4.5: The Chebyshev Lowpass Approximation - *CC BY-NC 4.0*
    - 4.6: Element Extraction - *CC BY-NC 4.0*
    - 4.7: Butterworth and Chebyshev Filters - *CC BY-NC 4.0*
    - 4.8: Impedance and Admittance Inverters - *CC BY-NC 4.0*
    - 4.9: Filter Transformations - *CC BY-NC 4.0*
    - 4.10: Cascaded Line Realization of Filters - *CC BY-NC 4.0*
    - 4.11: Butterworth and Chebyshev Bandpass Filters - *CC BY-NC 4.0*
    - 4.12: Richard's Transformation - *CC BY-NC 4.0*
    - 4.13: Kuroda's and Norton's Network Identities - *CC BY-NC 4.0*
    - 4.14: Inter-resonator Coupled Bandpass Filters - *CC BY-NC 4.0*
    - 4.15: Bandpass Filter Topologies - *CC BY-NC 4.0*
    - 4.16: Case Study- Design of a Bandstop Filter - *CC BY-NC 4.0*
    - 4.17: Active Filters - *CC BY-NC 4.0*
    - 4.18: Transient Response of a Bandpass Filter - *CC BY-NC 4.0*
    - 4.19: Summary - *CC BY-NC 4.0*
    - 4.20: References - *CC BY-NC 4.0*
    - 4.21: Exercises - *CC BY-NC 4.0*
  - 5: Oscillators - *CC BY-NC 4.0*

- 5.1: Introduction - *CC BY-NC 4.0*
- 5.2: Oscillator Theory - *CC BY-NC 4.0*
- 5.3: Reflection Oscillator - *CC BY-NC 4.0*
- 5.4: Case Study- Reflection Oscillator - *CC BY-NC 4.0*
- 5.5: Advanced Discussion of Oscillator Noise - *CC BY-NC 4.0*
- 5.6: Summary - *CC BY-NC 4.0*
- 5.7: References - *CC BY-NC 4.0*
- 5.8: Exercises - *CC BY-NC 4.0*

6: Antennas and the RF Link - *CC BY-NC 4.0*

- 6.1: Introduction - *CC BY-NC 4.0*
- 6.2: RF Antennas - *CC BY-NC 4.0*
- 6.3: Resonant Antennas - *CC BY-NC 4.0*
- 6.4: Traveling-Wave Antennas - *CC BY-NC 4.0*
- 6.5: Antenna Parameters - *CC BY-NC 4.0*
- 6.6: The RF Link - *CC BY-NC 4.0*
- 6.7: Summary - *CC BY-NC 4.0*
- 6.8: References - *CC BY-NC 4.0*
- 6.9: Exercises - *CC BY-NC 4.0*
- 6.10: Multipaths and Delay Spread - *CC BY-NC 4.0*
- 6.11: 4.8 Radio Link Interference - *CC BY-NC 4.0*
- 6.12: Antenna Array - *CC BY-NC 4.0*

Back Matter - *CC BY-NC 4.0*

- Index - *CC BY-NC 4.0*
- Glossary - *CC BY-NC 4.0*
- Detailed Licensing - *CC BY-NC 4.0*
- Detailed Licensing - *Undeclared*