CHRIS BOWICK WITH JOHN BLYLER AND CHERYL AJLUNI

RF CIRCUIT design

SECOND EDITION

A best-seller now thoroughly revised

Two new chapters on RF Front-End Design and RF Design Tools

Perfect for the practical, hard-working RF professional

Newnes

PREFACE

A great deal has changed since Chris Bowick's *RF Circuit Design* was first published, some 25 years ago. In fact, we could just say that the RF industry has changed quite a bit since the days of Marconi and Tesla—both technological visionaries woven into the fabric of history as the men who enabled radio communications. Who could have envisioned that their innovations in the late 1800's would lay the groundwork for the eventual creation of the radio—a key component in all mobile and portable communications systems that exist today? Or, that their contributions would one day lead to such a compelling array of RF applications, ranging from radar to the cordless telephone and everything in between. Today, the radio stands as the backbone of the wireless industry. It is in virtually every wireless device, whether a cellular phone, measurement/instrumentation system used in manufacturing, satellite communications system, television or the WLAN.

Of course, back in the early 1980s when this book was first written, RF was generally seen as a defense/military technology. It was utilized in the United States weapons arsenal as well as for things like radar and anti-jamming devices. In 1985, that image of RF changed when the FCC essentially made several bands of wireless spectrum, the Industrial, Scientific, and Medical (ISM) bands, available to the public on a license-free basis. By doing so—and perhaps without even fully comprehending the momentum its actions would eventually create—the FCC planted the seeds of what would one day be a multibillion-dollar industry.

Today that industry is being driven not by aerospace and defense, but rather by the consumer demand for wireless applications that allow "anytime, anywhere" connectivity. And, it is being enabled by a range of new and emerging radio protocols such as Bluetooth®, Wi-Fi (802.11 WLAN), WiMAX, and ZigBee®, in addition to 3G and 4G cellular technologies like CDMA, EGPRS, GSM, and Long Term Evolution (LTE). For evidence of this fact, one needs look no further than the cellular handset. Within one decade, between roughly the years 1990 and 2000, this application emerged from a very small scale semiprofessional niche, to become an almost omnipresent device, with the number of users equal to 18% of the world population. Today, nearly 2 billion people use mobile phones on a daily basis—not just for their voice services, but for a growing number of social and mobile, data-centric Internet applications. Thanks to the mobile phone and service telecommunications industry revolution, average consumers today not only expect pervasive, ubiquitous mobility, they are demanding it.

But what will the future hold for the consumer RF application space? The answer to that question seems fairly well-defined as the RF industry now finds itself rallying behind a single goal: to realize true convergence. In other words, the future of the RF industry lies in its ability to enable next-generation mobile devices to cross all of the boundaries of the RF spectrum. Essentially then, this converged mobile device would bring together traditionally disparate functionality (e.g., mobile phone, television, PC and PDA) on the mobile platform.

Again, nowhere is the progress of the converged mobile device more apparent than with the cellular handset. It offers the ideal platform on which RF standards and technologies can converge to deliver a whole host of new functionality and capabilities that, as a society, we may not even yet be able to imagine. Movement in that direction has already begun. According to analysts with the IDC Worldwide Mobile Phone Tracker service, the converged mobile device market grew an estimated 42 percent in 2006 for a total of over 80 million units. In the fourth quarter alone, vendors shipped a total of 23.5 million devices, 33 percent more than the same quarter a year ago. That's a fairly remarkable accomplishment considering that, prior to the mid-nineties, the possibility of true RF convergence was thought unreachable. The mixing, sampling and direct-conversion technologies were simply deemed too clunky and limited to provide the foundation necessary for implementation of such a vision.

Regardless of how and when the goal of true convergence is finally realized, one thing has become imminently clear in the midst of all the growth and innovation of the past twenty five years—the RF industry is alive and well. More importantly, it is well primed for a future full of continuing innovation and market growth.

Of course, while all of these changes created a wealth of business opportunities in the RF industry, they also created new challenges for RF engineers pushing the limits of design further and further. Today, new opportunities signal new design challenges which engineers—whether experts in RF technology or not—will likely have to face.

One key challenge is how to accommodate the need for multi-band reception in cellular handsets. Another stems from the need for higher bandwidth at higher frequencies which, in turn, means that the critical dimensions of relevant parasitic elements shrink. As a result, layout elements that once could be ignored (e.g., interconnect, contact areas and holes, and bond pads) become non-negligible and influence circuit performance.

In response to these and other challenges, the electronics industry has innovated, and continues to innovate. Consider, for example, that roughly 25 years ago or so, electronic design automation (EDA) was just an infant industry, particularly for high-frequency RF and microwave engineering. While a few tools were commercially available, rather than use these solutions, most companies opted to develop their own high-frequency design tools. As the design process became more complex and the in-house tools too costly to develop and maintain, engineers turned to design automation to address their needs. Thanks to innovation from a variety of EDA companies, engineers now have access to a full gamut of RF/microwave EDA products and methodologies to aid them with everything from design and analysis to verification.

But the innovation doesn't stop there. RF front-end architectures have and will continue to evolve in step with cellular handsets sporting multi-band reception. Multi-band subsystems and shrinking element sizes have coupled with ongoing trends toward lower cost and decreasing time-to-market to create the need for tightly integrated RF front-ends and transceiver circuits. These high levels of system integration have in turn given rise to single-chip modules that incorporate front-end filters, amplifiers and mixes. But implementing single-chip RF front-end designs requires a balance of performance trade-offs between the interfacing subsystems, namely, the antenna and digital baseband systems. Achieving the required system performance when implementing integrated RF front-ends means that analog designers must now work more closely with their digital baseband counterpart, thus leading to greater integration of the traditional analog–digital design teams.

Other areas of innovation in the RF industry will come from improved RF power transistors that promise to give wireless infrastructure power amplifiers new levels of performance with better reliability and ruggedness. RFICs hope to extend the role of CMOS to enable emerging mobile handsets to deliver multimedia functions from a compact package at lower cost. Incumbents like gallium arsenide (GaAs) have moved to higher voltages to keep the pace going. Additionally, power amplifier-duplexer-filter modules will rapidly displace separate components in multi-band W-CDMA radios. Single-chip multimode transceivers will displace separate EDGE and W-CDMA/HSDPA transceivers in W-EDGE handsets. And, to better handle parasitic and high-speed effects on circuits, accurate modeling and back-annotation of ever-smaller layout elements will become critical, as will accurate electromagnetic (EM) modeling of RF on-chip structures like coils and interconnect.

Still further innovation will come from emerging technologies in RF such as gallium nitride and micro-electro-mechanical systems (MEMS). In the latter case, these advanced micromachined devices are being integrated with CMOS signal processing and conditioning circuits for high-volume markets such as mobile phones and portable electronics. According to market research firm ABI Research, by 2008 use of MEMs in mobile phones will take off. This is due to the technology's small size, flexibility and performance advantages, all of which are critical to enabling the adaptive, multifunction handsets of the future.

It is this type of innovation, coupled with the continuously changing landscape of existing application and market opportunities, which has prompted a renewed look at the content in *RF Circuit Design*. It quickly became clear that, in order for this book to continue to serve its purpose as your hands-on guide to RF circuit design, changes were required. As a result, this new 25th anniversary edition comes to you with updated information on existing topics like resonant circuits, impedance matching and RF amplifier design, as well as new content pertaining to RF front-end design and RF design tools. This information is applicable to any engineer working in today's dynamically changing RF industry, as well as for those true visionaries working on the cusp of the information/communication/entertainment market convergence which the RF industry now inspires.

Cheryl Ajluni and John Blyler

ACKNOWLEDGMENTS

No man—or woman—is an island. Many very busy people helped to make this update of Chris's original book possible. Here are just a few of the main contributors—old friends and new—who gave generously of their time and expertise in the review of the RF Front-End chapter of this book: Special thanks to George Zafiropoulos, VP of Marketing, at Synopsys for also rekindling my interest in amateur radio; Colin Warwick, RF Product Manager, The MathWorks, Inc., (Thanks for a very thorough review!); Rick Lazansky, R&D Manager, Agilent EEs of EDA; David Ewing, Director of Software Engineering at Synapse and George Opsahl, President of Clearbrook Technology.

One of the most challenging tasks in preparing any technical piece is the selection of the right case study. This task was made easier for me by the help of both Analog Devices, Inc., and by Jean Rousset, consultant to Agilent Technologies.

This update would not have been possible without the help of Cheryl Ajluni—my co-author, friend, and former editor of Penton's *Wireless Systems Design* magazine. Additional thanks to Jack Browne, editor of *Microwave and RF* magazine, for his insights and content sharing at a critical juncture during my writing. Last but not least, I thank the two most important people to any published book author—namely the acquisition editor, Rachel Roumeliotis and the project manager, Anne B. McGee at Elsevier. Great job, everyone!

John Blyler

This revised version of RF Circuit Design would not have been possible were it not for the tireless efforts of many friends and colleagues, to all of whom I offer my utmost gratitude and respect. Their technical contributions, reviews and honest opinions helped me more than they will ever know. With that said, I want to offer special thanks to Doron Aronson, Michael C'deBaca, Joseph Curcurio, John Dunn, Suzanne Graham, Sonia Harrison, Victoria Juarez de Savin, Jim Lev, Daren McClearnon, Tom Quan, Mark Ravenstahl, Craig Schmidt, Dave Smith, Janet Smith, Heidi Vantulden, and Per Viklund; as well as the following companies: Agilent Technologies, Ansoft, Applied Wave Research, Cadence Design Systems, Mentor Graphics, Microwave Software, and The MathWorks, Inc.

To all of the folks at Elsevier who contributed in some way to this book—Anne B. McGee, Ganesan Murugesan and Rachel Roumeliotis—your work ethic, constant assistance and patience have been very much appreciated.

To Cindy Shamieh, whose excellent research skills provided the basis for many of the revisions throughout this version of the book—your efforts and continued friendship mean the world to me.

And last, but certainly not least, to John Blyler my friend and co-author—thank you for letting me share this journey with you.

Cheryl Ajluni

CHAPTER 1

COMPONENTS and Systems

Components, those bits and pieces which make up a radio frequency (RF) circuit, seem at times to be taken for granted. A capacitor is, after all, a capacitor—isn't it? A 1-megohm resistor presents an impedance of at least 1 megohm—doesn't it? The reactance of an inductor always increases with frequency, right? Well, as we shall see later in this discussion, things aren't always as they seem. Capacitors at certain frequencies may not be capacitors at all, but may look inductive, while inductors may look like capacitors, and resistors may tend to be a little of both.

In this chapter, we will discuss the properties of resistors, capacitors, and inductors at radio frequencies as they relate to circuit design. But, first, let's take a look at the most simple component of any system and examine its problems at radio frequencies.

WIRE

Wire in an RF circuit can take many forms. Wirewound resistors, inductors, and axial- and radial-leaded capacitors all use a wire of some size and length either in their leads, or in the actual body of the component, or both. Wire is also used in many interconnect applications in the lower RF spectrum. The behavior of a wire in the RF spectrum depends to a large extent on the wire's diameter and length. Table 1-1 lists, in the American Wire Gauge (AWG) system, each gauge of wire, its corresponding diameter, and other characteristics of interest to the RF circuit designer. In the AWG system, the diameter of a wire will roughly double every six wire gauges. Thus, if the last six gauges and their corresponding diameters are memorized from the chart, all other wire diameters can be determined without the aid of a chart (Example 1-1).

Skin Effect

A conductor, at low frequencies, utilizes its entire cross-sectional area as a transport medium for charge carriers. As the frequency is increased, an increased magnetic field at the center of the conductor presents an impedance to the charge carriers, thus decreasing the current density at the center of the conductor and increasing the current density around its perimeter. This increased current density near the edge of the conductor is known as *skin effect*. It occurs in all conductors including resistor leads, capacitor leads, and inductor leads.

EXAMPLE 1-1

Given that the diameter of AWG 50 wire is 1.0 mil (0.001 inch), what is the diameter of AWG 14 wire?

Solution

AWG 50 = 1 mil

AWG 44 = 2 × 1 mil = 2 mils

AWG 38 = 2 × 2 mils = 4 mils

AWG 32 = 2 × 4 mils = 8 mils

AWG 26 = 2× 8 mils = 16 mils

AWG 20 = 2 × 16 mils = 32 mils

AWG 14 = 2 × 32 mils = 64 mils (0.064 inch)

The depth into the conductor at which the charge-carrier current density falls to $1/e$, or 37% of its value along the surface, is known as the *skin depth* and is a function of the frequency and the permeability and conductivity of the medium. Thus, different conductors, such as silver, aluminum, and copper, all have different skin depths.

The net result of skin effect is an effective decrease in the cross-sectional area of the conductor and, therefore, a net increase in the ac resistance of the wire as shown in Fig. 1-1. For copper, the skin depth is approximately 0.85 cm at 60 Hz and 0.007 cm at 1 MHz. Or, to state it another way: 63% of the RF current flowing in a copper wire will flow within a distance of 0.007 cm of the outer edge of the wire.

Straight-Wire Inductors

In the medium surrounding any current-carrying conductor, there exists a magnetic field. If the current in the conductor is an alternating current, this magnetic field is alternately expanding and contracting and, thus, producing a voltage on the wire which opposes any change in current flow. This opposition to change is called *self-inductance* and we call anything that possesses this quality an *inductor*. Straight-wire inductance might seem trivial, but as will be seen later in the chapter, the higher we go in frequency, the more important it becomes.

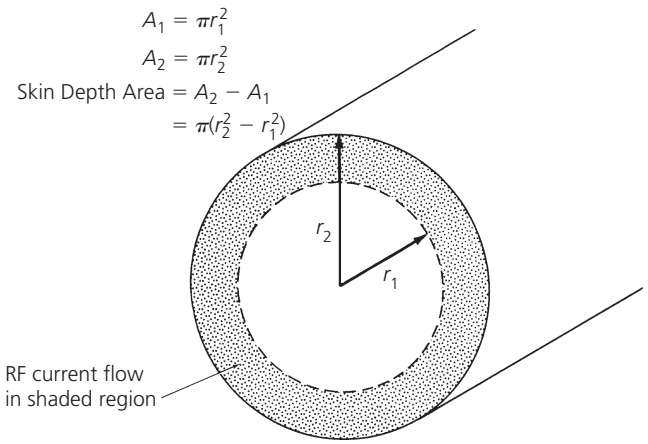


FIG. 1-1. Skin depth area of a conductor.

The inductance of a straight wire depends on both its length and its diameter, and is found by:

$$L = 0.002l\left[2.3\log\left(\frac{4l}{d}\right) - 0.75\right]\mu\text{H} \qquad \text{(Eq. 1-1)}$$

where,

L = the inductance in μH,
l = the length of the wire in cm,
d = the diameter of the wire in cm.

This is shown in calculations of Example 1-2.

EXAMPLE 1-2

Find the inductance of 5 centimeters of No. 22 copper wire.

Solution

From Table 1-1, the diameter of No. 22 copper wire is 25.3 mils. Since 1 mil equals 2.54×10^{-3} cm, this equals 0.0643 cm. Substituting into Equation 1-1 gives

$$L = (0.002)(5)\left[2.3\log\left(\frac{4(5)}{0.0643}\right) - 0.75\right]$$
$$= 50 \text{ nanohenries}$$

The concept of inductance is important because any and all conductors at radio frequencies (including hookup wire, capacitor leads, etc.) tend to exhibit the property of inductance. Inductors will be discussed in greater detail later in this chapter.

RESISTORS

Resistance is the property of a material that determines the rate at which electrical energy is converted into heat energy for a given electric current. By definition:

$$1 \text{ volt across } 1 \text{ ohm} = 1 \text{ coulomb per second}$$
$$= 1 \text{ ampere}$$

The thermal dissipation in this circumstance is 1 watt.

$$P = EI$$
$$= 1 \text{ volt} \times 1 \text{ ampere}$$
$$= 1 \text{ watt}$$

Resistors are used everywhere in circuits, as transistor bias networks, pads, and signal combiners. However, very rarely is there any thought given to how a resistor actually behaves once we depart from the world of direct current (DC). In some instances, such as in transistor biasing networks, the resistor will still perform its DC circuit function, but it may also disrupt the circuit's RF operating point.

Resistor Equivalent Circuit

The equivalent circuit of a resistor at radio frequencies is shown in Fig. 1-2. R is the resistor value itself, L is the lead inductance, and C is a combination of parasitic capacitances which varies from resistor to resistor depending on the resistor's structure. Carbon-composition resistors are notoriously poor high-frequency performers. A carbon-composition resistor consists of densely packed dielectric particulates or carbon granules. Between each pair of carbon granules is a very small parasitic capacitor. These parasitics, in aggregate, are not insignificant, however, and are the major component of the device's equivalent circuit.

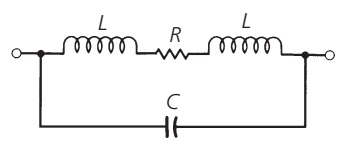


FIG. 1-2. Resistor equivalent circuit.

Wirewound resistors have problems at radio frequencies too. As may be expected, these resistors tend to exhibit widely varying impedances over various frequencies. This is particularly true of the low resistance values in the frequency range of 10 MHz to 200 MHz. The inductor L, shown in the equivalent circuit of Fig. 1-2, is much larger for a wirewound resistor than for a carbon-composition resistor. Its value can be calculated using the single-layer air-core inductance approximation formula. This formula is discussed later in this chapter. Because wirewound resistors look like inductors, their impedances will first increase as the frequency increases. At some frequency (F_r), however, the inductance (L) will resonate with the shunt capacitance (C), producing an impedance peak. Any further increase in frequency will cause the resistor's impedance to decrease as shown in Fig. 1-3.

A metal-film resistor seems to exhibit the best characteristics over frequency. Its equivalent circuit is the same as the

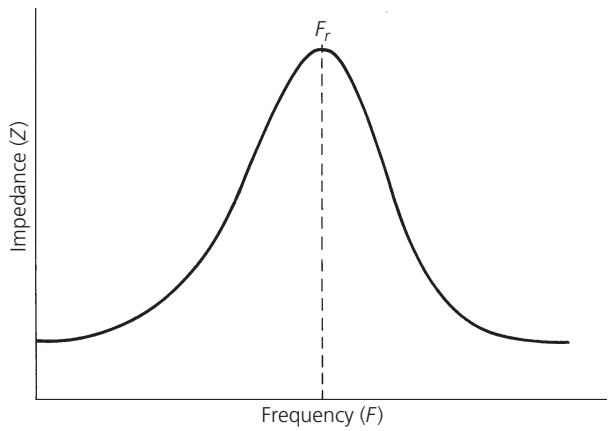


FIG. 1-3. Impedance characteristic of a wirewound resistor.

carbon-composition and wirewound resistor, but the values of the individual parasitic elements in the equivalent circuit decrease.

The impedance of a metal-film resistor tends to decrease with frequency above about 10 MHz, as shown in Fig. 1-4. This is due to the shunt capacitance in the equivalent circuit. At very high frequencies, and with low-value resistors (under 50 Ω), lead inductance and skin effect may become noticeable. The lead inductance produces a resonance peak, as shown for the 5 Ω resistance in Fig. 1-4, and skin effect decreases the slope of the curve as it falls off with frequency.

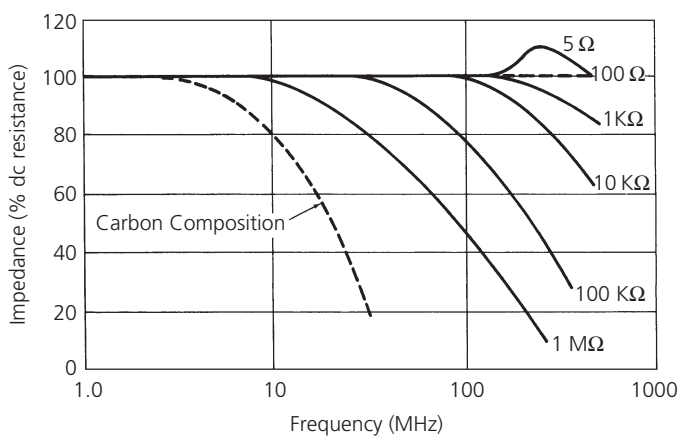


FIG. 1-4. Frequency characteristics of metal-film vs. carbon-composition resistors. (Adapted from *Handbook of Components for Electronics*, McGraw-Hill)

Many manufacturers will supply data on resistor behavior at radio frequencies but it can often be misleading. Once you understand the mechanisms involved in resistor behavior, however, it will not matter in what form the data is supplied. Example 1-3 illustrates that fact.

The recent trend in resistor technology has been to eliminate or greatly reduce the stray reactances associated with resistors. This has led to the development of thin-film chip resistors, such as

EXAMPLE 1-3

In Fig. 1-2, the lead lengths on the metal-film resistor are 1.27 cm (0.5 inch), and are made up of No. 14 wire. The total stray shunt capacitance (C) is 0.3 pF. If the resistor value is 10,000 ohms, what is its equivalent RF impedance at 200 MHz?

Solution

From Table 1-1, the diameter of No. 14 AWG wire is 64.1 mils (0.1628 cm). Therefore, using Equation 1-1:

$$L = (0.002)(1.27)\left[2.3 \log\left(\frac{4(1.27)}{0.1628} - 0.75\right)\right]$$
$$= 8.7 \text{ nanohenries}$$

This presents an equivalent reactance at 200 MHz of:

$$X_L = \omega L$$
$$= 2\pi(200 \times 10^6)(8.7 \times 10^{-9})$$
$$= 10.93 \text{ ohms}$$

The capacitor (C) presents an equivalent reactance of:

$$X_c = \frac{1}{\omega C}$$
$$= \frac{1}{2\pi(200 \times 10^6)(0.3 \times 10^{-12})}$$
$$= 2653$$

The combined equivalent circuit for this resistor, at 200 MHz, is shown in Fig. 1-5.

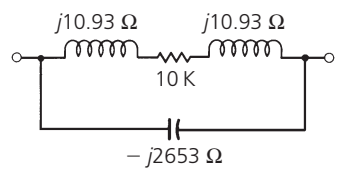


FIG. 1-5. Equivalent circuit values for Example 1-3.

From this sketch, we can see that, in this case, the lead inductance is insignificant when compared with the 10K series resistance and it may be neglected. The parasitic capacitance, on the other hand, cannot be neglected. What we now have, in effect, is a 2653 Ω reactance in parallel with a 10,000 Ω resistance. The magnitude of the combined impedance is:

$$Z = \frac{RX_e}{\sqrt{R^2 + X_e^2}}$$
$$= \frac{(10K)(2653)}{\sqrt{(10K)^2 + (2653)^2}}$$
$$= 2564.3 \text{ ohms}$$

Thus, our 10K resistor looks like 2564 ohms at 200 MHz.

those shown in Fig. 1-6. They are typically produced on alumina or beryllia substrates and offer very little parasitic reactance at frequencies from DC to 2 GHz.

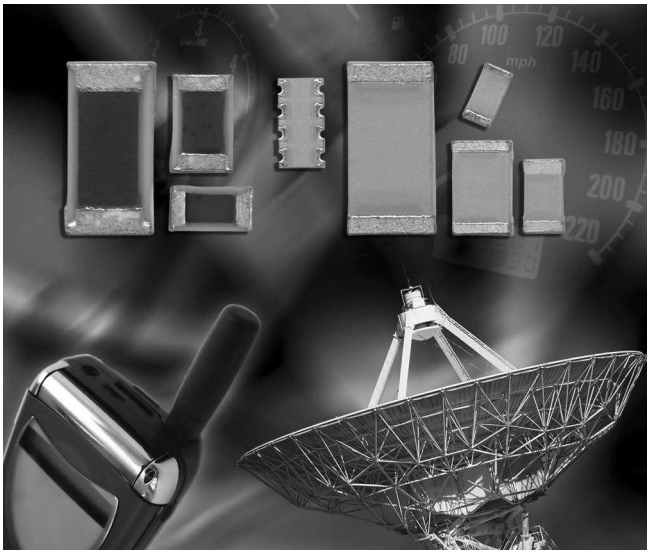


FIG. 1-6. Thin-film resistors. (*Courtesy of Vishay Intertechnology)*

CAPACITORS

Capacitors are used extensively in RF applications, such as bypassing, interstage coupling, and in resonant circuits and filters. It is important to remember, however, that not all capacitors lend themselves equally well to each of the above-mentioned applications. The primary task of the RF circuit designer, with regard to capacitors, is to choose the best capacitor for his particular application. Cost effectiveness is usually a major factor in the selection process and, thus, many trade-offs occur. In this section, we'll take a look at the capacitor's equivalent circuit and we will examine a few of the various types of capacitors used at radio frequencies to see which are best suited for certain applications. But first, a little review.

Parallel-Plate Capacitor

A capacitor is any device which consists of two conducting surfaces separated by an insulating material or dielectric. The dielectric is usually ceramic, air, paper, mica, plastic, film, glass, or oil. The capacitance of a capacitor is that property which permits the storage of a charge when a potential difference exists between the conductors. Capacitance is measured in units of farads. A 1-farad capacitor's potential is raised by 1 volt when it receives a charge of 1 coulomb.

$$C = \frac{Q}{V}$$

where,

C = capacitance in farads,
Q = charge in coulombs,
V = voltage in volts.

However, the farad is much too impractical to work with, so smaller units were devised.

$$1 \text{ microfarad} = 1\ \mu\text{F} = 1 \times 10^{-6} \text{ farad}$$
$$1 \text{ picofarad} = 1 \text{ pF} = 1 \times 10^{-12} \text{ farad}$$

As stated previously, a capacitor in its fundamental form consists of two metal plates separated by a dielectric material of some sort. If we know the area (A) of each metal plate, the distance (d) between the plate (in inches), and the permittivity (ε) of the dielectric material in farads/meter (f/m), the capacitance of a parallel-plate capacitor can be found by:

$$C = \frac{0.2249\varepsilon A}{d\varepsilon_0} \text{ picofarads} \qquad \text{(Eq. 1-2)}$$

where

$$\varepsilon_0 = \text{free-space permittivity} = 8.854 \times 10^{-12} \text{ f/m.}$$

In Equation 1-2, the area (A) must be large with respect to the distance (d). The ratio of ε to ε_0 is known as the dielectric constant (k) of the material. The dielectric constant is a number that provides a comparison of the given dielectric with air (see Fig. 1-7). The ratio of $\varepsilon/\varepsilon_0$ for air is, of course, 1. If the dielectric constant of a material is greater than 1, its use in a capacitor as a dielectric will permit a greater amount of capacitance for the same dielectric thickness as air. Thus, if a material's dielectric constant is 3, it will produce a capacitor having three times the capacitance of one that has air as its dielectric. For a given value of capacitance, then, higher dielectric-constant materials will produce physically smaller capacitors. But, because the dielectric plays such a major role in determining the capacitance of a capacitor, it follows that the influence of a dielectric on capacitor operation, over frequency and temperature, is often important.

Dielectric	K
Air	1
Polystrene	2.5
Paper	4
Mica	5
Ceramic (low K)	10
Ceramic (high K)	100−10,000

FIG. 1-7. Dielectric constants of some common materials.

Real-World Capacitors

The usage of a capacitor is primarily dependent upon the characteristics of its dielectric. The dielectric's characteristics also determine the voltage levels and the temperature extremes at which the device may be used. Thus, any losses or imperfections in the dielectric have an enormous effect on circuit operation.

The equivalent circuit of a capacitor is shown in Fig. 1-8, where C equals the capacitance, R_s is the heat-dissipation loss expressed either as a power factor (PF) or as a dissipation factor (DF), R_p

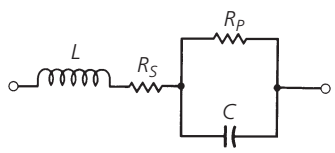


FIG. 1-8. Capacitor equivalent circuit.

is the insulation resistance, and L is the inductance of the leads and plates. Some definitions are needed now.

Power Factor—In a *perfect* capacitor, the alternating current will lead the applied voltage by 90°. This phase angle (ϕ) will be smaller in a real capacitor due to the total series resistance ($R_s + R_p$) that is shown in the equivalent circuit. Thus,

$$PF = \cos\phi$$

The power factor is a function of temperature, frequency, and the dielectric material.

Insulation Resistance—This is a measure of the amount of DC current that flows through the dielectric of a capacitor with a voltage applied. No material is a perfect insulator; thus, some leakage current must flow. This current path is represented by R_p in the equivalent circuit and, typically, it has a value of 100,000 megohms or more.

Effective Series Resistance—Abbreviated ESR, this resistance is the combined equivalent of R_s and R_p, and is the AC resistance of a capacitor.

$$ESR = \frac{PF}{\omega C}(1 \times 10^6)$$

where

$$\omega = 2\pi f$$

Dissipation Factor – The *DF* is the ratio of AC resistance to the reactance of a capacitor and is given by the formula:

$$DF = \frac{ESR}{X_c} \times 100\%$$

Q – The *Q* of a circuit is the reciprocal of *DF* and is defined as the quality factor of a capacitor.

$$Q = \frac{1}{DF} = \frac{X_c}{ESR}$$

Thus, the larger the *Q*, the better the capacitor.

The effect of these imperfections in the capacitor can be seen in the graph of Fig. 1-9. Here, the impedance characteristic of an ideal capacitor is plotted against that of a real-world capacitor. As shown, as the frequency of operation increases, the lead inductance becomes important. Finally, at F_r, the inductance becomes series resonant with the capacitor. Then, above F_r, the capacitor acts like an inductor. In general, larger-value capacitors tend to exhibit more internal inductance than smaller-value capacitors. Therefore, depending upon its internal structure, a

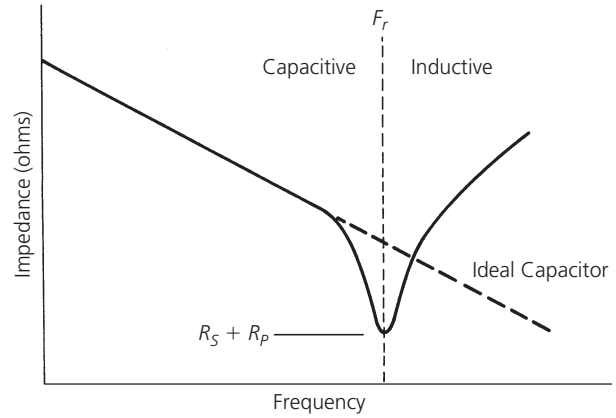


FIG. 1-9. Impedance characteristic vs. frequency.

0.1-μF capacitor may not be as good as a 300-pF capacitor in a bypass application at 250 MHz. In other words, the classic formula for capacitive reactance, $X_e = \frac{1}{\omega C}$, might seem to indicate that larger-value capacitors have less reactance than smaller-value capacitors at a given frequency. At RF frequencies, however, the opposite may be true. At certain higher frequencies, a 0.1-μF capacitor might present a higher impedance to the signal than would a 330-pF capacitor. This is something that must be considered when designing circuits at frequencies above 100 MHz. Ideally, each component that is to be used in any VHF, or higher frequency, design should be examined on a network analyzer similar to the one shown in Fig. 1-10. This will allow the designer to know exactly what he is working with before it goes into the circuit.

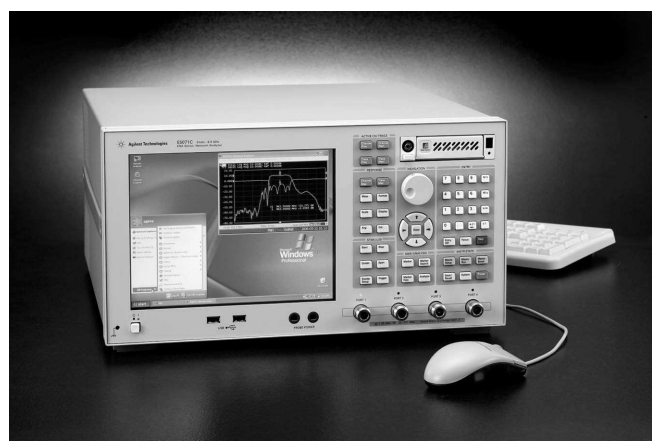


FIG. 1-10. Agilent E5071C Network Analyzer.

Capacitor Types

There are many different dielectric materials used in the fabrication of capacitors, such as paper, plastic, ceramic, mica, polystyrene, polycarbonate, teflon, oil, glass, and air. Each material has its advantages and disadvantages. The RF designer

is left with a myriad of capacitor types that he could use in any particular application and the ultimate decision to use a particular capacitor is often based on convenience rather than good sound judgment. In many applications, this approach simply cannot be tolerated. This is especially true in manufacturing environments where more than just one unit is to be built and where they must operate reliably over varying temperature extremes. It is often said in the engineering world that anyone can design something and make it work *once*, but it takes a good designer to develop a unit that can be produced in quantity and still operate as it should in different temperature environments.

Ceramic Capacitors

Ceramic dielectric capacitors vary widely in both dielectric constant (k = 5 to 10,000) and temperature characteristics. A good rule of thumb to use is: "The higher the k, the worse is its temperature characteristic." This is shown quite clearly in Fig. 1-11.

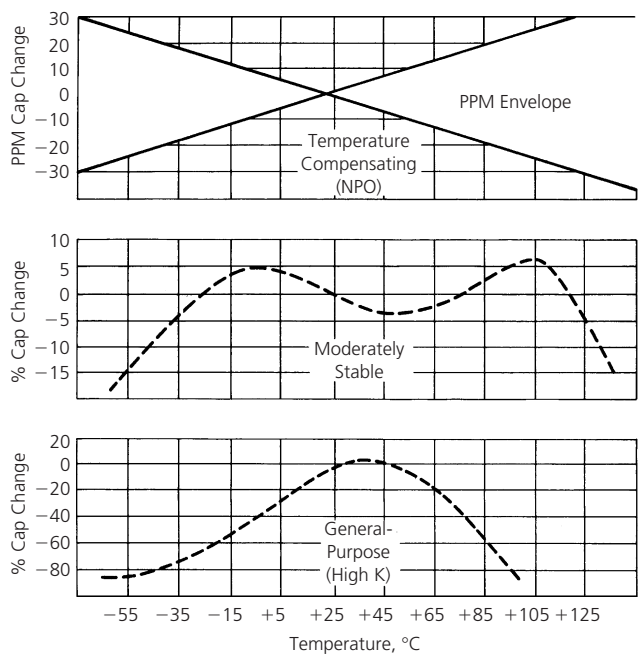


FIG. 1-11. Temperature characteristics for ceramic dielectric capacitors.

As illustrated, low-k ceramic capacitors tend to have linear temperature characteristics. These capacitors are generally manufactured using both magnesium titanate, which has a positive temperature coefficient (TC), and calcium titanate which has a negative TC. By combining the two materials in varying proportions, a range of controlled temperature coefficients can be generated. These capacitors are sometimes called temperature compensating capacitors, or NPO (negative positive zero) ceramics. They can have TCs that range anywhere from +150 to −4700 ppm/°C (parts-per-million-per-degree-Celsius) with tolerances as small as ±15 ppm/°C. Because of their excellent temperature stability, NPO ceramics are well suited for oscillator, resonant circuit, or filter applications.

Moderately stable ceramic capacitors (Fig. 1-11) typically vary ±15% of their rated capacitance over their temperature range. This variation is typically nonlinear, however, and care should be taken in their use in resonant circuits or filters where stability is important. These ceramics are generally used in switching circuits. Their main advantage is that they are generally smaller than the NPO ceramic capacitors and, of course, cost less.

High-K ceramic capacitors are typically termed general-purpose capacitors. Their temperature characteristics are very poor and their capacitance may vary as much as 80% over various temperature ranges (Fig. 1-11). They are commonly used only in bypass applications at radio frequencies.

There are ceramic capacitors available on the market which are specifically intended for RF applications. These capacitors are typically high-Q (low ESR) devices with flat ribbon leads or with no leads at all. The lead material is usually solid silver or silver plated and, thus, contains very low resistive losses. At VHF frequencies and above, these capacitors exhibit very low lead inductance due to the flat ribbon leads. These devices are, of course, more expensive and require special printed-circuit board areas for mounting. The capacitors that have no leads are called chip capacitors. These capacitors are typically used above 500 MHz where lead inductance cannot be tolerated. Chip capacitors and flat ribbon capacitors are shown in Fig. 1-12.

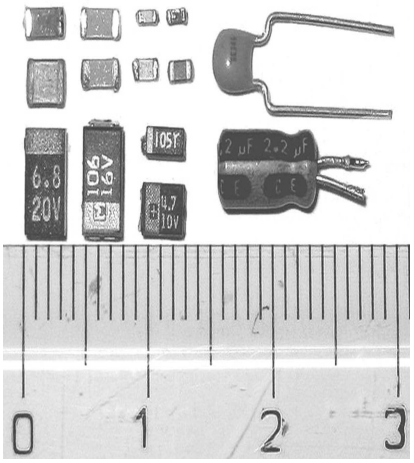


FIG. 1-12. Chip and ceramic capacitors. (*Courtesy of Wikipedia*)

Mica Capacitors

Mica capacitors typically have a dielectric constant of about 6, which indicates that for a particular capacitance value, mica capacitors are typically large. Their low k, however, also produces an extremely good temperature characteristic. Thus, mica capacitors are used extensively in resonant circuits and in filters where PC board area is of no concern.

Silvered mica capacitors are even more stable. Ordinary mica capacitors have plates of foil pressed against the mica dielectric. In silvered micas, the silver plates are applied by a process called *vacuum evaporation* which is a much more exacting process. This produces an even better stability with very tight and reproducible tolerances of typically +20 ppm/°C over a range −60°C to +89°C.

The problem with micas, however, is that they are becoming increasingly less cost effective than ceramic types. Therefore, if you have an application in which a mica capacitor would seem to work well, chances are you can find a less expensive NPO ceramic capacitor that will work just as well.

Metalized-Film Capacitors

"Metalized-film" is a broad category of capacitor encompassing most of the other capacitors listed previously and which we have not yet discussed. This includes teflon, polystyrene, polycarbonate, and paper dielectrics.

Metalized-film capacitors are used in a number of applications, including filtering, bypassing, and coupling. Most of the polycarbonate, polystyrene, and teflon styles are available in very tight (±2%) capacitance tolerances over their entire temperature range. Polystyrene, however, typically cannot be used over +85°C as it is very temperature sensitive above this point. Most of the capacitors in this category are typically larger than the equivalent-value ceramic types and are used in applications where space is not a constraint.

INDUCTORS

An inductor is nothing more than a wire wound or coiled in such a manner as to increase the magnetic flux linkage between the turns of the coil (see Fig. 1-13). This increased flux linkage increases the wire's self-inductance (or just plain inductance) beyond that which it would otherwise have been. Inductors are used extensively in RF design in resonant circuits, filters, phase shift and delay networks, and as RF chokes used to prevent, or at least reduce, the flow of RF energy along a certain path.

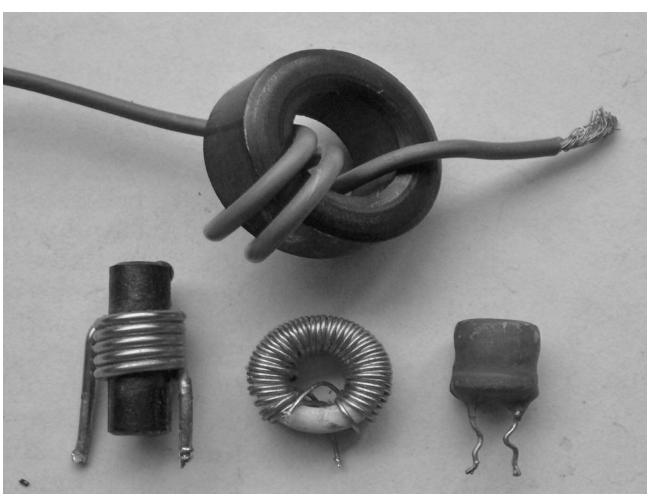


FIG. 1-13. Simple inductors. (*Courtesy of Wikipedia*)

Real-World Inductors

As we have discovered in previous sections of this chapter, there is no "perfect" component, and inductors are certainly no exception. As a matter of fact, of the components we have discussed, the inductor is probably the component most prone to very drastic changes over frequency.

Fig. 1-14 shows what an inductor really looks like at RF frequencies. As previously discussed, whenever we bring two conductors into close proximity but separated by a dielectric, and place a voltage differential between the two, we form a capacitor. Thus, if any wire resistance at all exists, a voltage drop (even though very minute) will occur between the windings, and small capacitors will be formed. This effect is shown in Fig. 1-14 and is called distributed capacitance (C_d). Then, in Fig. 1-15, the capacitance (C_d) is an aggregate of the individual parasitic distributed capacitances of the coil shown in Fig. 1-14.

The effect of C_d upon the reactance of an inductor is shown in Fig. 1-16. Initially, at lower frequencies, the inductor's reactance parallels that of an ideal inductor. Soon, however, its reactance departs from the ideal curve and increases at a much faster rate until it reaches a peak at the inductor's parallel resonant frequency (F_r). Above F_r, the inductor's reactance begins to

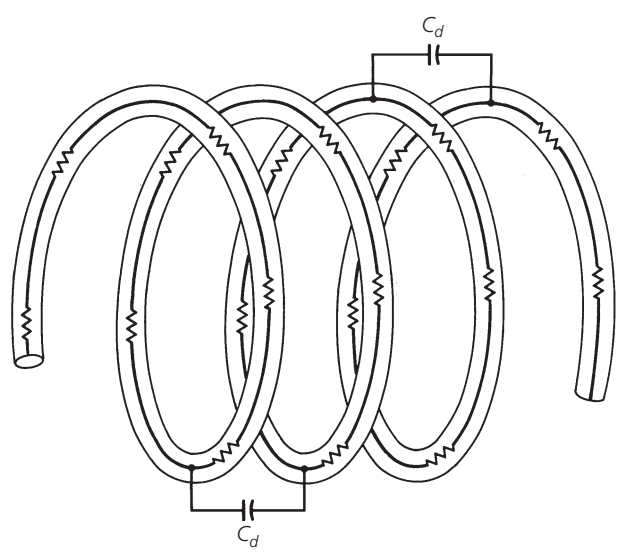


FIG. 1-14. Distributed capacitance and series resistance in an inductor.

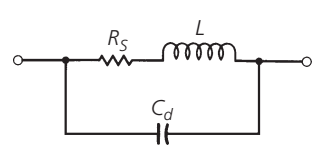


FIG. 1-15. Inductor equivalent circuit.

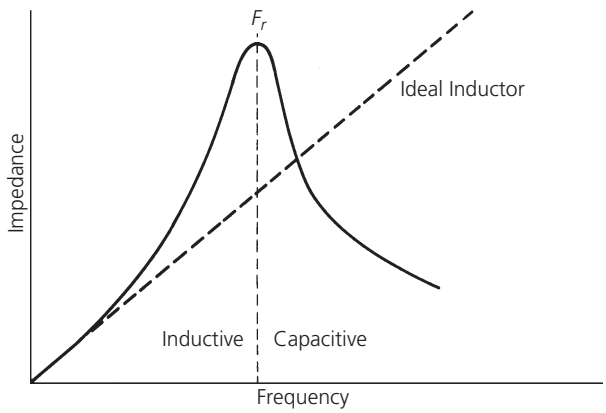


FIG. 1-16. Impedance characteristic vs. frequency for a practical and an ideal inductor.

EXAMPLE 1-4

To show that the impedance of a lossless inductor at resonance is infinite, we can write the following:

$$Z = \frac{X_L X_C}{X_L + X_C} \qquad \text{(Eq.1-3)}$$

where

Z = the impedance of the parallel circuit,
X_L = the inductive reactance ($j\omega L$),
X_C = the capacitive reactance $\left(\frac{1}{j\omega C}\right)$.

Therefore,

$$Z = \frac{j\omega L\left(\frac{1}{j\omega C}\right)}{j\omega L + \frac{1}{j\omega C}} \qquad \text{(Eq.1-4)}$$

Multiplying numerator and denominator by $j\omega C$, we get:

$$Z = \frac{j\omega L}{(j\omega L)(j\omega C) + 1}$$

$$= \frac{j\omega L}{j^2\omega^2 LC + 1} \qquad \text{(Eq.1-5)}$$

From algebra, $j^2 = -1$; then, rearranging:

$$Z = \frac{j\omega L}{1 - \omega^2 LC} \qquad \text{(Eq.1-6)}$$

If the term $\omega^2 LC$, in Equation 1-6, should ever become equal to 1, then the denominator will be equal to zero and impedance Z will become infinite. The frequency at which $\omega^2 LC$ becomes equal to 1 is:

$$\omega^2 LC = 1$$

$$LC = \frac{1}{\omega^2}$$

$$\sqrt{LC} = \frac{1}{\omega}$$

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

EXAMPLE 4-4—Cont

$$\frac{1}{2\pi\sqrt{LC}} = f \qquad \text{(Eq.1-7)}$$

which is the familiar equation for the resonant frequency of a tuned circuit.

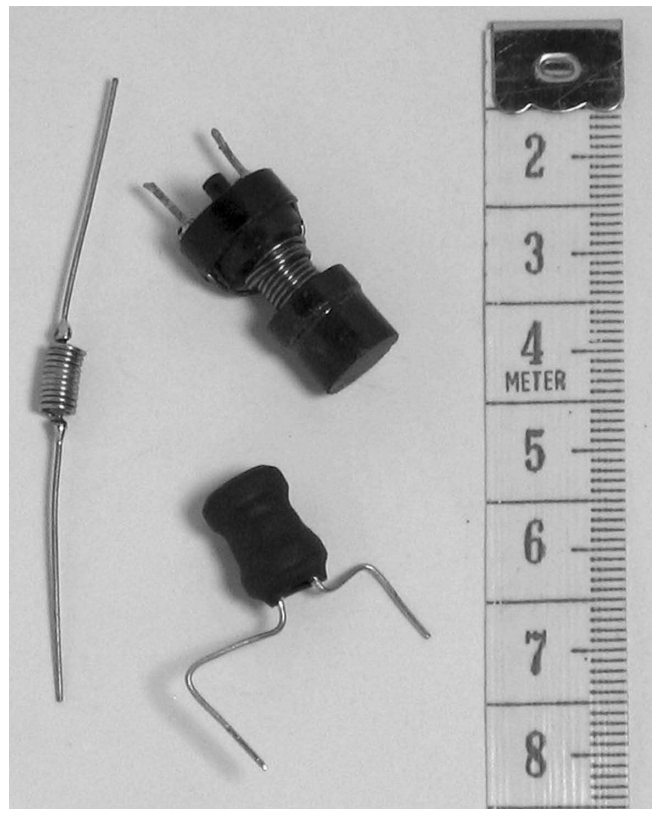


FIG. 1-17. Chip inductors. (*Courtesy of Wikipedia*)

decrease with frequency and, thus, the inductor begins to look like a capacitor. Theoretically, the resonance peak would occur at infinite reactance (see Example 1-4). However, due to the series resistance of the coil, some finite impedance is seen at resonance.

Recent advances in inductor technology have led to the development of microminiature fixed-chip inductors. One type is shown in Fig. 1-17. These inductors feature a ceramic substrate with gold-plated solderable wrap-around bottom connections. They come in values from 0.01 μH to 1.0 mH, with typical Qs that range from 40 to 60 at 200 MHz.

It was mentioned earlier that the series resistance of a coil is the mechanism that keeps the impedance of the coil finite at resonance. Another effect it has is to broaden the resonance peak of the impedance curve of the coil. This characteristic of resonant circuits is an important one and will be discussed in detail in Chapter 3.

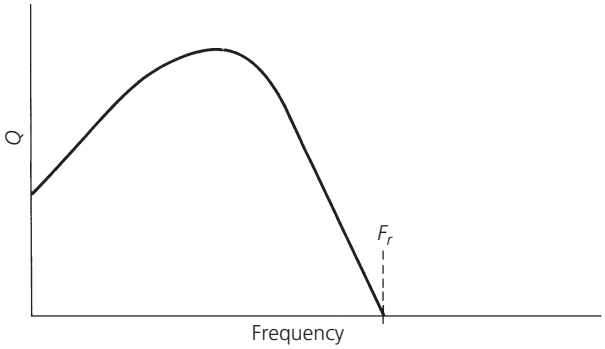


FIG. 1-18. The Q variation of an inductor vs. frequency.

The ratio of an inductor's reactance to its series resistance is often used as a measure of the quality of the inductor. The larger the ratio, the better is the inductor. This quality factor is referred to as the Q of the inductor.

$$Q = \frac{X}{R_s}$$

If the inductor were wound with a perfect conductor, its Q would be infinite and we would have a lossless inductor. Of course, there is no perfect conductor and, thus, an inductor always has some finite Q.

At low frequencies, the Q of an inductor is very good because the only resistance in the windings is the dc resistance of the wire—which is very small. But as the frequency increases, skin effect and winding capacitance begin to degrade the quality of the inductor. This is shown in the graph of Fig. 1-18. At low frequencies, Q will increase directly with frequency because its reactance is increasing and skin effect has not yet become noticeable. Soon, however, skin effect does become a factor. The Q still rises, but at a lesser rate, and we get a gradually decreasing slope in the curve. The flat portion of the curve in Fig. 1-18 occurs as the series resistance and the reactance are changing at the same rate. Above this point, the shunt capacitance and skin effect of the windings combine to decrease the Q of the inductor to zero at its resonant frequency.

Some methods of increasing the Q of an inductor and extending its useful frequency range are:

1. Use a larger diameter wire. This decreases the AC and DC resistance of the windings.
2. Spread the windings apart. Air has a lower dielectric constant than most insulators. Thus, an air gap between the windings decreases the interwinding capacitance.
3. Increase the permeability of the flux linkage path. This is most often done by winding the inductor around a magnetic-core material, such as iron or ferrite. A coil made in this manner will also consist of fewer turns for a given inductance. This will be discussed in a later section of this chapter.

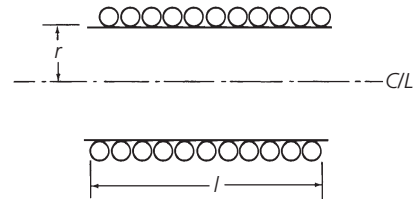


FIG. 1-19. Single-layer air-core inductor requirements.

Single-Layer Air-Core Inductor Design

Every RF circuit designer needs to know how to design inductors. It may be tedious at times, but it's well worth the effort. The formula that is generally used to design single-layer air-core inductors is given in Equation 1-8 and diagrammed in Fig. 1-19.

$$L = \frac{0.394r^2N^2}{9r + 10l} \qquad \text{(Eq. 1-8)}$$

where

r = the coil radius in cm,

l = the coil length in cm,

L = the inductance in microhenries.

However, coil length l must be greater than $0.67r$. This formula is accurate to within one percent. See Example 1-5.

EXAMPLE 1-5

Design a 100 nH (0.1 μH) air-core inductor on a 1/4-inch (0.635 cm) coil form.

Solution

For optimum Q, the length of the coil should be equal to its diameter. Thus, $l = 0.635$ cm, $r = 0.317$ cm, and $L = 0.1\mu$H.

Using Equation 1-8 and solving for N gives:

$$N = \sqrt{\frac{29L}{0.394r}}$$

where we have taken $l = 2r$, for optimum Q.

Substituting and solving:

$$N = \sqrt{\frac{29(0.1)}{(0.394)(0.317)}}$$
$$= 4.8 \text{ turns}$$

Thus, we need 4.8 turns of wire within a length of 0.635 cm. A look at Table 1-1 reveals that the largest diameter enamel-coated wire that will allow 4.8 turns in a length of 0.635 cm is No. 18 AWG wire which has a diameter of 42.4 mils (0.107 cm).

Wire Size (AWG)	Dia in Mils* (Bare)	Dia in Mils (Coated)	Ohms/ 1000 ft.	Area Circular Mils	Wire Size (AWG)	Dia in Mils* (Bare)	Dia in Mils (Coated)	Ohms/ 1000 ft.	Area Circular Mils
1	289.3		0.124	83690	26	15.9	17.2	41.0	253
2	257.6		0.156	66360	27	14.2	15.4	51.4	202
3	229.4		0.197	52620	28	12.6	13.8	65.3	159
4	204.3		0.249	41740	29	11.3	12.3	81.2	123
5	181.9		0.313	33090	30	10.0	11.0	104.0	100
6	162.0		0.395	26240	31	8.9	9.9	131	79.2
7	144.3		0.498	20820	32	8.0	8.8	162	64.0
8	128.5	131.6	0.628	16510	33	7.1	7.9	206	50.4
9	114.4	116.3	0.793	13090	34	6.3	7.0	261	39.7
10	101.9	104.2	0.999	10380	35	5.6	6.3	331	31.4
11	90.7	93.5	1.26	8230	36	5.0	5.7	415	25.0
12	80.8	83.3	1.59	6530	37	4.5	5.1	512	20.2
13	72.0	74.1	2.00	5180	38	4.0	4.5	648	16.0
14	64.1	66.7	2.52	4110	39	3.5	4.0	847	12.2
15	57.1	59.5	3.18	3260	40	3.1	3.5	1080	9.61
16	50.8	52.9	4.02	2580	41	2.8	3.1	1320	7.84
17	45.3	47.2	5.05	2050	42	2.5	2.8	1660	6.25
18	40.3	42.4	6.39	1620	43	2.2	2.5	2140	4.84
19	35.9	37.9	8.05	1290	44	2.0	2.3	2590	4.00
20	32.0	34.0	10.1	1020	45	1.76	1.9	3350	3.10
21	28.5	30.2	12.8	812	46	1.57	1.7	4210	2.46
22	25.3	27.0	16.2	640	47	1.40	1.6	5290	1.96
23	22.6	24.2	20.3	511	48	1.24	1.4	6750	1.54
24	20.1	21.6	25.7	404	49	1.11	1.3	8420	1.23
25	17.9	19.3	32.4	320	50	.99	1.1	10600	0.98

* 1 mil = 2.54×10^{-3} cm

TABLE 1-1. AWG Wire Chart

Keep in mind that even though optimum Q is attained when the length of the coil (l) is equal to its diameter ($2r$), this is sometimes not practical and, in many cases, the length is much greater than the diameter. In Example 1-5, we calculated the need for 4.8 turns of wire in a length of 0.635 cm and decided that No. 18 AWG wire would fit. The only problem with this approach is that when the design is finished, we end up with a very tightly wound coil. This increases the distributed capacitance between the turns and, thus, lowers the useful frequency range of the inductor by lowering its resonant frequency. We could take either one of the following compromise solutions to this dilemma:

1. Use the next smallest AWG wire size to wind the inductor while keeping the length (l) the same. This approach will allow a small air gap between windings and, thus, decrease the interwinding capacitance. It also, however, increases the resistance of the windings by decreasing the diameter of the conductor and, thus, it lowers the Q.
2. Extend the length of the inductor (while retaining the use of No. 18 AWG wire) just enough to leave a small air gap between the windings. This method will produce the same effect as Method No. 1. It reduces the Q somewhat but it decreases the interwinding capacitance considerably.

Magnetic-Core Materials

In many RF applications, where large values of inductance are needed in small areas, air-core inductors cannot be used because of their size. One method of decreasing the size of a coil

while maintaining a given inductance is to decrease the number of turns while at the same time increasing its magnetic flux density. The flux density can be increased by decreasing the "reluctance" or magnetic resistance path that links the windings of the inductor. We do this by adding a magnetic-core material, such as iron or ferrite, to the inductor. The permeability (μ) of this material is much greater than that of air and, thus, the magnetic flux isn't as "reluctant" to flow between the windings. The net result of adding a high permeability core to an inductor is the gaining of the capability to wind a given inductance with fewer turns than what would be required for an air-core inductor. Thus, several advantages can be realized.

1. Smaller size—due to the fewer number of turns needed for a given inductance.
2. Increased Q—fewer turns means less wire resistance.
3. Variability—obtained by moving the magnetic core in and out of the windings.

There are some major problems that are introduced by the use of magnetic cores, however, and care must be taken to ensure that the core that is chosen is the right one for the job. Some of the problems are:

1. Each core tends to introduce its own losses. Thus, adding a magnetic core to an air-core inductor could possibly *decrease* the Q of the inductor, depending on the material used and the frequency of operation.
2. The permeability of all magnetic cores changes with frequency and usually decreases to a very small value at the upper end of their operating range. It eventually approaches the permeability of air and becomes "invisible" to the circuit.
3. The higher the permeability of the core, the more sensitive it is to temperature variation. Thus, over wide temperature ranges, the inductance of the coil may vary appreciably.
4. The permeability of the magnetic core changes with applied signal level. If too large an excitation is applied, saturation of the core will result.

These problems can be overcome if care is taken, in the design process, to choose cores wisely. Manufacturers now supply excellent literature on available sizes and types of cores, complete with their important characteristics.

TOROIDS

A toroid, very simply, is a ring or doughnut-shaped magnetic material that is widely used to wind RF inductors and transformers. Toroids are usually made of iron or ferrite. They come in various shapes and sizes (Fig. 1-20) with widely varying characteristics. When used as cores for inductors, they can typically yield very high Qs. They are self-shielding, compact, and best of all, easy to use.

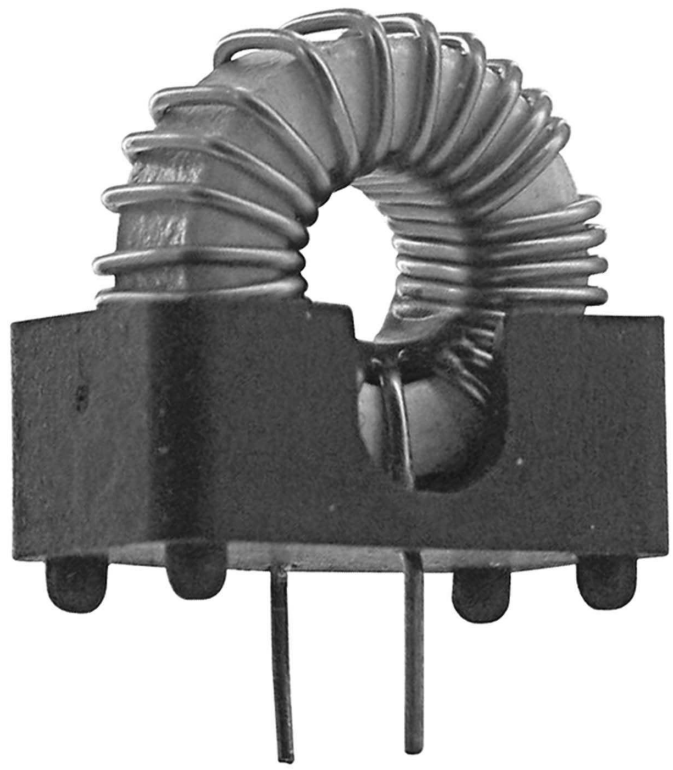


FIG. 1-20. Toroidal core inductor. (*Courtesy of Allied Electronics*)

The Q of a toroidal inductor is typically high because the toroid can be made with an extremely high permeability. As was discussed in an earlier section, high permeability cores allow the designer to construct an inductor with a given inductance (for example, 35 μH) with fewer turns than is possible with an air-core design. Fig. 1-21 indicates the potential savings obtained in number of turns of wire when coil design is changed from air-core to toroidal-core inductors. The air-core inductor, if wound for optimum Q, would take 90 turns of a very small wire (in order to fit all turns within a 1/4-inch length) to reach 35 μH; however, the toroidal inductor would only need 8 turns to reach the design goal. Obviously, this is an extreme case but it serves a useful purpose and illustrates the point. The toroidal core does require fewer turns for a given inductance than does an air-core design. Thus, there is less AC resistance and the Q *can be* increased dramatically.

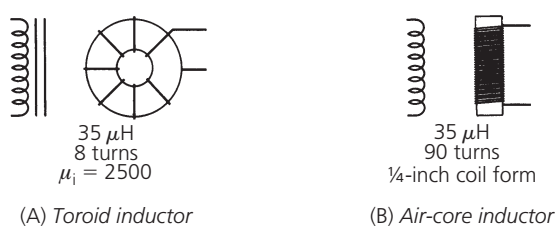


FIG. 1-21. Turns comparison between inductors for the same inductance.

The self-shielding properties of a toroid become evident when Fig. 1-22 is examined. In a typical aircore inductor, the magnetic-flux lines linking the turns of the inductor take the shape shown in Fig. 1-22A. The sketch clearly indicates that the air surrounding the inductor is definitely part of the magnetic-flux path. Thus, this inductor tends to radiate the RF signals flowing within. A toroid, on the other hand (Fig. 1-22B), completely contains the magnetic flux within the material itself; thus, no radiation occurs. In actual practice, of course, some radiation will occur but it is minimized. This characteristic of toroids eliminates the need for bulky shields surrounding the inductor. The shields not only tend to reduce available space, but they also reduce the Q of the inductor that they are shielding.

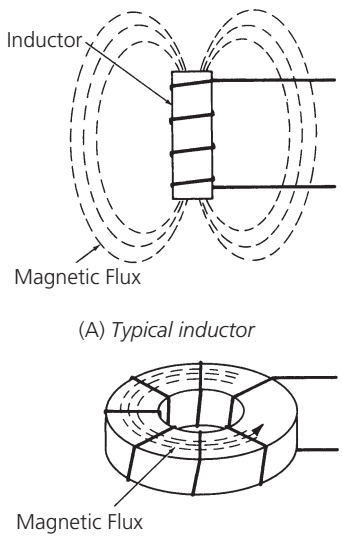


FIG. 1-22. Shielding effect of a toroidal inductor.

Core Characteristics

Earlier, we discussed, in general terms, the relative advantages and disadvantages of using magnetic cores. The following discussion of typical toroidal-core characteristics will aid you in specifying the core that you need for your particular application.

Fig. 1-23 is a typical magnetization curve for a magnetic core. The curve simply indicates the magnetic-flux density (B) that occurs in the inductor with a specific magnetic-field intensity (H) applied. As the magnetic-field intensity is increased from zero (by increasing the applied signal voltage), the magnetic-flux density that links the turns of the inductor increases quite linearly. The ratio of the magnetic-flux density to the magnetic-field intensity is called the permeability of the material. This has already been mentioned on numerous occasions.

$$\mu = B/H(\text{webers/ampere-turn}) \qquad \text{(Eq. 1-9)}$$

Thus, the permeability of a material is simply a measure of how well it transforms an electrical excitation into a magnetic flux. The better it is at this transformation, the higher is its permeability.

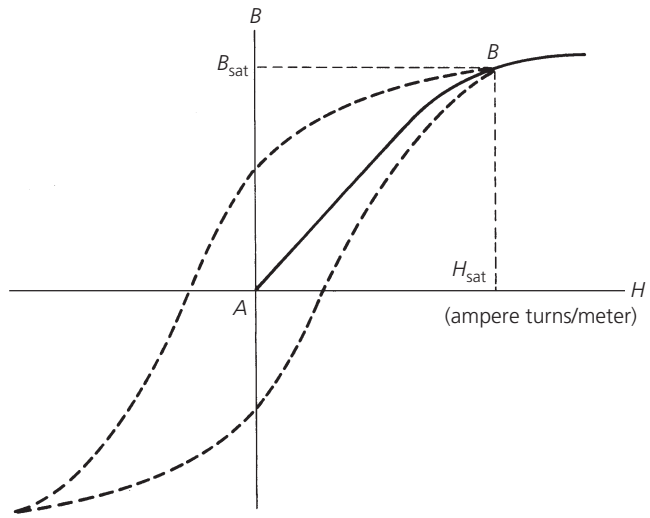


FIG. 1-23. Magnetization curve for a typical core.

As mentioned previously, initially the magnetization curve is linear. It is during this linear portion of the curve that permeability is usually specified and, thus, it is sometimes called initial permeability (μ_i) in various core literature. As the electrical excitation increases, however, a point is reached at which the magnetic-flux intensity does not continue to increase at the same rate as the excitation and the slope of the curve begins to decrease. Any further increase in excitation may cause *saturation* to occur. H_{sat} is the excitation point above which no further increase in magnetic-flux density occurs (B_{sat}). The incremental permeability above this point is the same as air. Typically, in RF circuit applications, we keep the excitation small enough to maintain linear operation.

B_{sat} varies substantially from core to core, depending upon the size and shape of the material. Thus, it is necessary to read and understand the manufacturer's literature that describes the particular core you are using. Once B_{sat} is known for the core, it is a very simple matter to determine whether or not its use in a particular circuit application will cause it to saturate. The in-circuit operational flux density (B_{op}) of the core is given by the formula:

$$B_{op} = \frac{E \times 10^8}{(4.44) f N A_e} \qquad \text{(Eq. 1-10)}$$

where,

B_{op} = the magnetic-flux density in gauss,
E = the maximum rms voltage across the inductor in volts,
f = the frequency in hertz,
N = the number of turns,
A_e = the effective cross-sectional area of the core in cm^2.

Thus, if the calculated B_{op} for a particular application is less than the published specification for B_{sat}, then the core will not saturate and its operation will be somewhat linear.

Another characteristic of magnetic cores that is very important to understand is that of internal loss. It has previously been mentioned that the careless addition of a magnetic core to an air-core inductor could possibly *reduce* the Q of the inductor. This concept might seem contrary to what we have studied so far, so let's examine it a bit more closely.

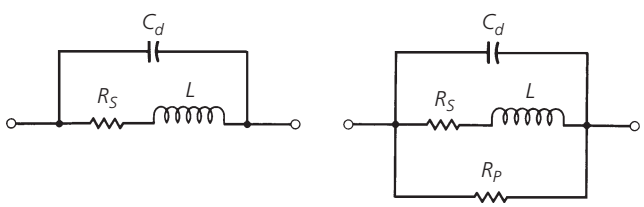
The equivalent circuit of an air-core inductor (Fig. 1-15) is reproduced in Fig. 1-24A for your convenience. The Q of this inductor is

$$Q = \frac{X_L}{R_s} \qquad \text{(Eq. 1-11)}$$

where

$X_L = \omega L$,

$R_s =$ the resistance of the windings.



(A) *Air core* (B) *Magnetic core*

FIG. 1-24. Equivalent circuits for air-core and magnetic-core inductors.

If we add a magnetic core to the inductor, the equivalent circuit becomes like that shown in Fig. 1-24B. We have added resistance R_p to represent the losses which take place in the core itself. These losses are in the form of *hysteresis*. Hysteresis is the power lost in the core due to the realignment of the magnetic particles within the material with changes in excitation, and the eddy currents that flow in the core due to the voltages induced within. These two types of internal loss, which are inherent to some degree in every magnetic core and are thus unavoidable, combine to reduce the efficiency of the inductor and, thus, increase its loss. But what about the new Q for the magnetic-core inductor? This question isn't as easily answered. Remember, when a magnetic core is inserted into an existing inductor, the value of the inductance is increased. Therefore, at any given frequency, its reactance increases proportionally. The question that must be answered then, in order to determine the new Q of the inductor, is: By what factors did the inductance and loss increase? Obviously, if by adding a toroidal core, the inductance were increased by a factor of two and its total loss was also increased by a factor of two, the Q would remain unchanged. If, however, the total coil loss were increased to four times its previous value while only doubling the inductance, the Q of the inductor would be reduced by a factor of two.

Now, as if all of this isn't confusing enough, we must also keep in mind that the additional loss introduced by the core is not constant, but varies (usually increases) with frequency. Therefore, the designer must have a complete set of manufacturer's data sheets for every core he is working with.

Toroid manufacturers typically publish data sheets which contain all the information needed to design inductors and transformers with a particular core. (Some typical specification and data sheets are given in Figs. 1-25 and 1-26.) In most cases, however, each manufacturer presents the information in a unique manner and care must be taken in order to extract the information that is needed without error, and in a form that can be used in the ensuing design process. This is not always as simple as it sounds. Later in this chapter, we will use the data presented in Figs. 1-25 and 1-26 to design a couple of toroidal inductors so that we may see some of those differences. Table 1-2 lists some of the commonly used terms along with their symbols and units.

Powdered Iron vs. Ferrite

In general, there are no hard and fast rules governing the use of ferrite cores versus powdered-iron cores in RF circuit-design applications. In many instances, given the same permeability and type, either core could be used without much change in performance of the actual circuit. There are, however, special applications in which one core might outperform another, and it is those applications which we will address here.

Powdered-iron cores, for instance, can typically handle more RF power without saturation or damage than the same size ferrite core. For example, ferrite, if driven with a large amount of RF power, tends to retain its magnetism permanently. This ruins the core by changing its permeability permanently. Powdered iron, on the other hand, if overdriven will eventually return to its initial permeability (μ_i). Thus, in any application where high RF power levels are involved, iron cores might seem to be the best choice.

In general, powdered-iron cores tend to yield higher-Q inductors, at higher frequencies, than an equivalent size ferrite core. This is due to the inherent core characteristics of powdered iron cores which produce much less internal loss than ferrite cores. This characteristic of powdered iron makes it very useful in narrow-band or tuned-circuit applications. Table 1-3 lists a few of the common powdered-iron core materials along with their typical applications.

At very low frequencies, or in broadband circuits which span the spectrum from VLF up through VHF, ferrite seems to be the general choice. This is true because, for a given core size, ferrite cores have a much higher permeability. The higher permeability is needed at the low end of the frequency range where, for a given inductance, fewer windings would be needed with the ferrite core. This brings up another point. Since ferrite cores, in general, have a higher permeability than the same size powdered-iron core, a coil of a given inductance can usually be wound on a much smaller ferrite core and with fewer turns. Thus, we can save circuit board area.

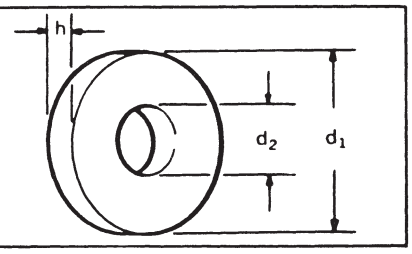
BROAD BAND-RATED FERRAMIC COMPONENTS

- Values measured at 100 KHz, T = 25°C.
- Temperature Coefficient (TC) = 0 to +0.75% /°C max., -40 to +70°C.
- Disaccommodation (D) = 3.0% max., 10-100 min., 25°C.
- Hysteresis Core Constant (η_i) measured at 20 KHz to 30 gauss (3 milli Tesla).
- For mm dimensions and core constants, see page 30.

Nom. μ_i 2500

MECHANICAL SPECIFICATIONS

	PART NUMBER				TOL	UNITS
	AA–01	AA–02	AA–03	AA–04		
d_1	0.135	0.155	0.230	0.100	±0.005	in.
d_2	0.065	0.088	0.120	0.050	±0.005	in.
h	0.055	0.051	0.060	0.050	±0.005	in.



ELECTRICAL SPECIFICATIONS

	PART NUMBER				TOL	UNITS
	AA–01	AA–02	AA–03	AA–04		
A_L	510	365	495	440	±20%	nH/turn^2
X_p/N^2	0.320	0.229	0.310	0.276	±20%	ohm/turn^2
R_p/N^2	10.4	7.5	10.0	8.9	min.	ohm/turn^2
Q	54	54	54	54	min.	
V_{rms}	7.9	7.1	13.6	5.1	max.	mv
η_i	1,480	1,400	0,920	2,150	max.	$VSA^{-2}\ H^{-3/2}$

TYPICAL CHARACTERISTIC CURVES – Part Numbers: AA–01, AA–02, AA–03 and AA–04

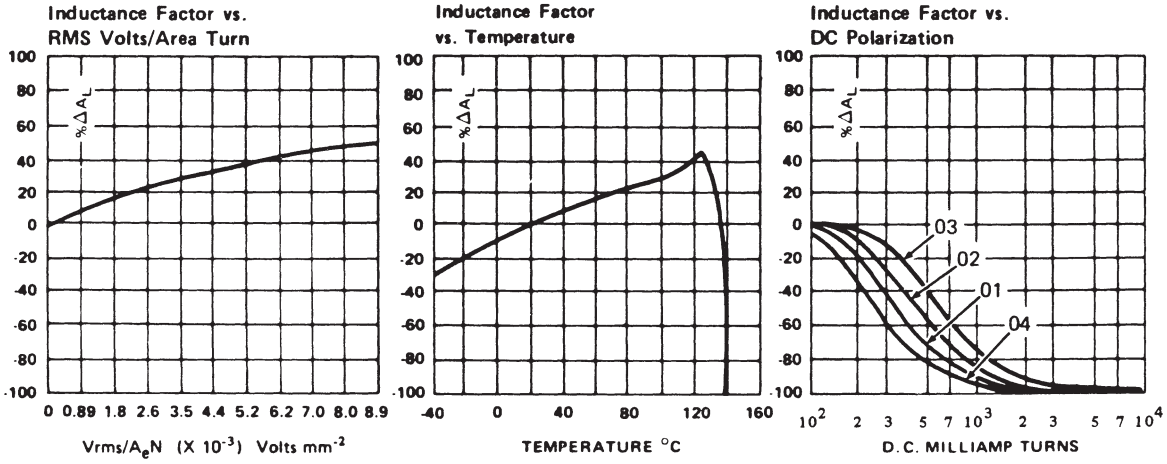


FIG. 1-25. Typical data sheet – with generic part numbers – for ferrite toroidal cores. (*Courtesy of Indiana General*)

IRON-POWDER TOROIDAL CORES

Core size	PHYSICAL DIMENSIONS Outer Diam. (in)	Inner Diam. (in)	Height (in)	Cross Sect. Area $(cm)^2$	Mean Length (cm)
T-225A - -	2.250	1.400	1.000	2.742	14.56
T-225 - -	2.250	1.400	.550	1.508	14.56
T-200 - -	2.000	1.250	.550	1.330	12.97
T-184 - -	1.840	.960	.710	2.040	11.12
T-157 - -	1.570	.950	.570	1.140	10.05
T-130 - -	1.300	.780	.437	.930	8.29
T-106 - -	1.060	.560	.437	.706	6.47
T- 94 - -	.942	.560	.312	.385	6.00
T- 80 - -	.795	.495	.250	.242	5.15
T- 68 - -	.690	.370	.190	.196	4.24
T- 50 - -	.500	.303	.190	.121	3.20
T- 44 - -	.440	.229	.159	.107	2.67
T- 37 - -	.375	.204	.128	.070	2.32
T- 30 - -	.307	.150	.128	.065	1.83
T- 25 - -	.255	.120	.096	.042	1.50
T- 20 - -	.200	.088	.070	.034	1.15
T- 16 - -	.160	.078	.060	.016	0.75
T- 12 - -	.120	.062	.050	.010	0.74

IRON - POWDER MATERIAL vs. FREQUENCY RANGE

Higher Q will be obtained in the upper portion of a materials frequency range when smaller cores are used. Likewise, in the lower portion of a materials frequency range, higher Q can be achieved when using the larger cores.

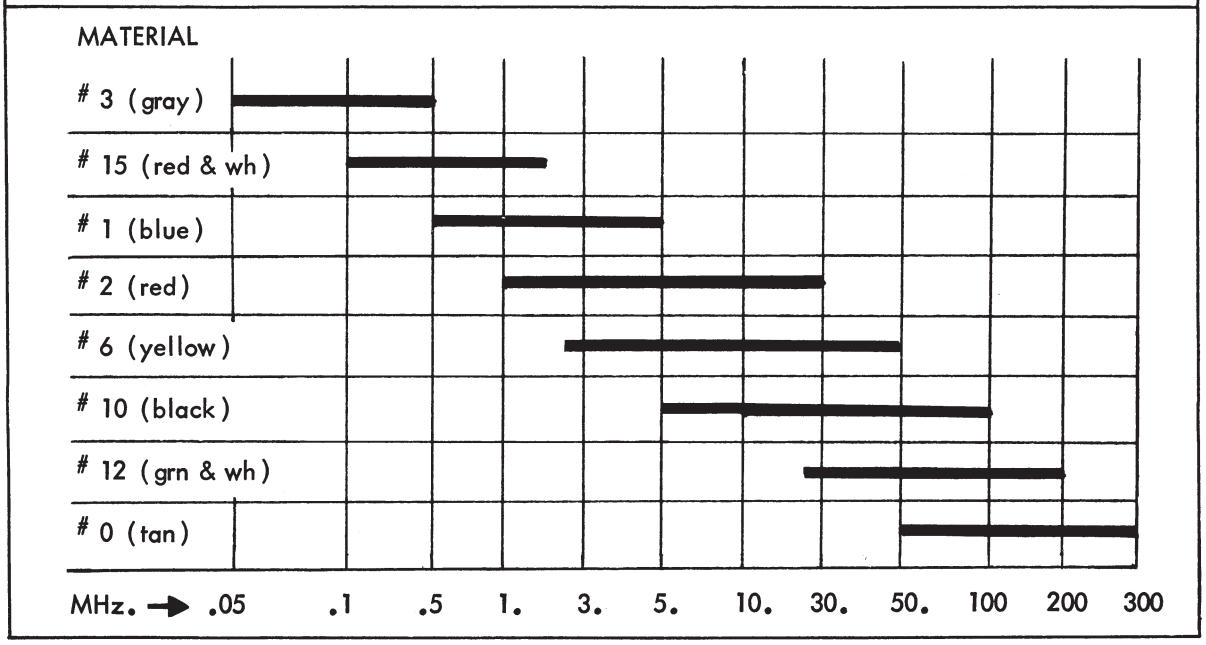


FIG. 1-26. Data sheet for powdered-iron toroidal cores. (*Courtesy Amidon Associates*)

IRON-POWDER TOROIDAL CORES

FOR RESONANT CIRCUITS

MATERIAL # 0 permeability 1 50 MHz to 300 MHz Tan

Core number	Outer diam. (in.)	Inner diam. (in.)	Height (in.)	A_L value uh / 100 t
T-130-0	1.300	.780	.437	15.0
T-106-0	1.060	.560	.437	19.2
T- 94-0	.942	.560	.312	10.6
T- 80-0	.795	.495	.250	8.5
T- 68-0	.690	.370	.190	7.5
T- 50-0	.500	.303	.190	6.4
T- 44-0	.440	.229	.159	6.5
T- 37-0	.375	.205	.128	4.9
T- 30-0	.307	.151	.128	6.0
T- 25-0	.255	.120	.096	4.5
T- 20-0	.200	.088	.067	3.5
T- 16-0	.160	.078	.060	3.0
T- 12-0	.125	.062	.050	3.0

MATERIAL # 12 permeability 3 20 MHz to 200 MHz Green & White

Core number	Outer diam. (in.)	Inner diam. (in.)	Height (in.)	A_L value uh / 100 t
T-80-12	.795	.495	.250	22
T-68-12	.690	.370	.190	21
T-50-12	.500	.300	.190	18
T-44-12	.440	.229	.159	18
T-37-12	.375	.205	.128	15
T-30-12	.307	.151	.128	16
T-25-12	.255	.120	.096	12
T-20-12	.200	.088	.067	10
T-16-12	.160	.078	.060	8
T-12-12	.125	.062	.050	7

Key to part numbers for :
IRON POWDER TOROIDAL CORES

T ------- 200 -------- 2

Toroid Outer diameter Material

Number of turns = 100

A_L values ± 5%

$$\sqrt{\frac{\text{desired inductance (uh)}}{A_L \text{ value (uh per 100 turns)}}}$$

FIG. 1-26. (Continued)

IRON-POWDER TOROIDAL CORES

FOR RESONANT CIRCUITS

MATERIAL # 10 permeability 6 10 MHz to 100 MHz Black

Core number	Outer diam. (in.)	Inner diam. (in.)	Height (in.)	A_L value uh / 100 t
T-94-10	.942	.560	.312	58
T-80-10	.795	.495	.250	32
T-68-10	.690	.370	.190	32
T-50-10	.500	.303	.190	31
T-44-10	.440	.229	.159	33
T-37-10	.375	.205	.128	25
T-30-10	.307	.151	.128	25
T-25-10	.255	.120	.096	19
T-20-10	.200	.088	.067	16
T-16-10	.160	.078	.060	13
T-12-10	.125	.062	.050	12

NUMBER OF TURNS vs. WIRE SIZE and CORE SIZE
Approximate number of turns of wire - single layer wound - single insulation

Core Size	wire size 40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10
T-12	47	37	29	21	15	11	8	5	4	2	1	1	1	0	0	0
T-16	63	49	38	29	21	16	11	8	5	3	3	1	1	1	0	0
T-20	72	56	43	33	25	18	14	9	6	5	4	3	1	1	1	0
T-25	101	79	62	48	37	28	21	15	11	7	5	4	3	1	1	1
T-30	129	101	79	62	48	37	28	21	15	11	7	5	4	3	1	1
T-37	177	140	110	87	67	53	41	31	23	17	12	9	7	5	3	1
T-44	199	157	124	97	76	60	46	35	27	20	15	10	7	6	5	3
T-50	265	210	166	131	103	81	63	49	37	28	21	16	11	8	6	5
T-68	325	257	205	162	127	101	79	61	47	36	28	21	15	11	9	7
T-80	438	347	276	219	172	137	108	84	66	51	39	30	23	17	12	8
T-94	496	393	313	248	195	156	123	96	75	58	45	35	27	20	14	10
T-106	496	393	313	248	195	156	123	96	75	58	45	35	27	20	14	10
T-130	693	550	439	348	275	220	173	137	107	83	66	51	40	30	23	17
T-157	846	672	536	426	336	270	213	168	132	104	82	64	50	38	29	22
T-184	846	672	536	426	336	270	213	168	132	104	82	64	50	38	29	22
T-200	1115	886	707	562	445	357	282	223	176	139	109	86	68	53	41	31
T-225	1250	993	793	631	499	400	317	250	198	156	123	98	77	60	46	36

FIG. 1-26. (Continued)

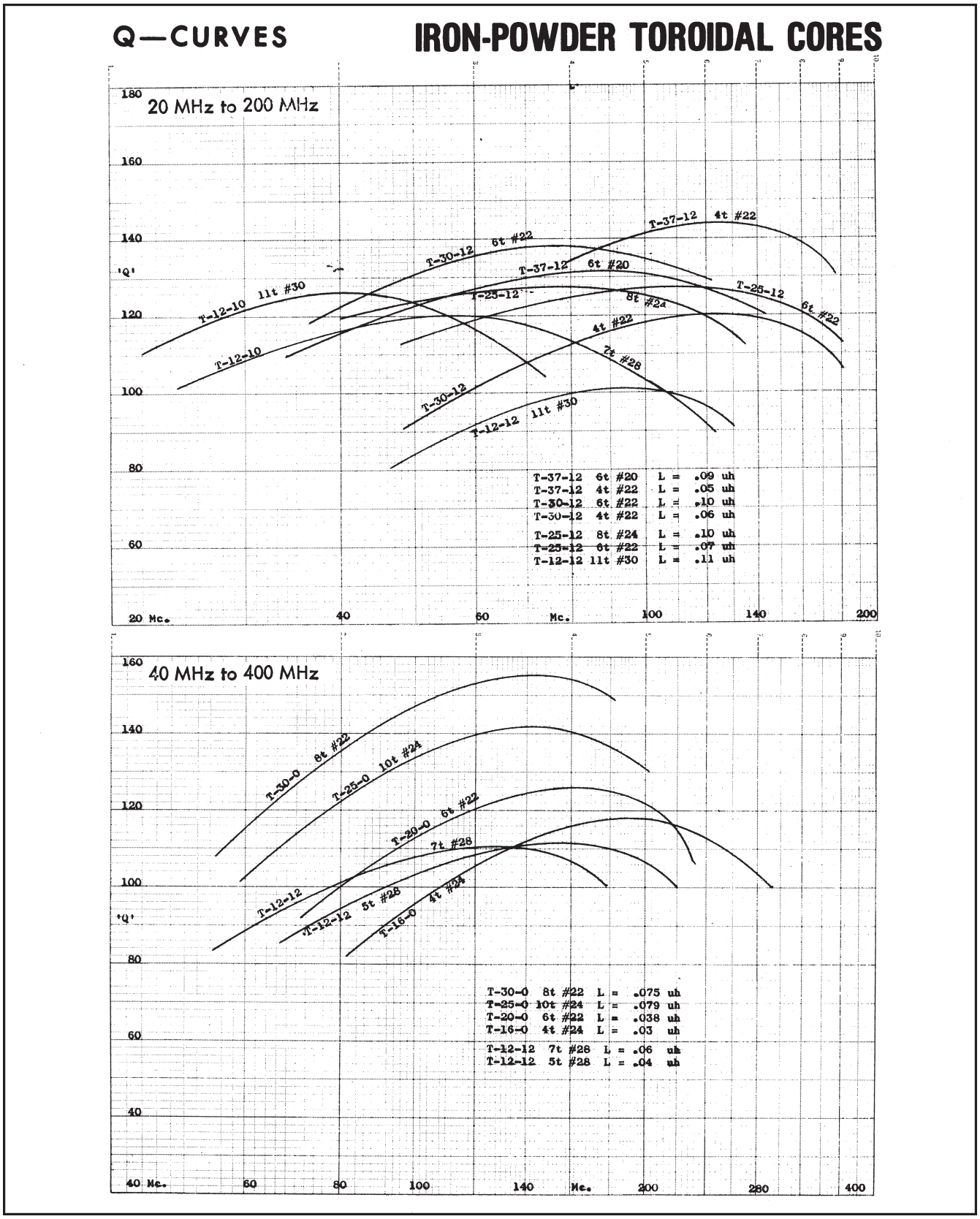


FIG. 1-26. (Continued)

Symbol	Description	Units
A_c	Available cross-sectional area. The area (perpendicular to the direction of the wire) for winding turns on a particular core.	cm^2
A_e	Effective area of core. The cross-sectional area that an equivalent gapless core would have.	cm^2
A_L	Inductive index. This relates the inductance to the number of turns for a particular core.	$nH/turn^2$
B_{sat}	Saturation flux density of the core.	gauss
B_{op}	Operating flux density of the core. This is with an applied voltage.	gauss
l_e	Effective length of the flux path.	cm
μ_i	Initial permeability. This is the effective permeability of the core at low excitation in the linear region.	numeric

TABLE 1-2. Toroidal Core Symbols and Definitions

TOROIDAL INDUCTOR DESIGN

For a toroidal inductor operating on the linear (nonsaturating) portion of its magnetization curve, its inductance is given by the following formula:

$$L = \frac{0.4\pi N^2 \mu_1 A_c \times 10^{-2}}{l_e} \qquad \text{(Eq. 1-12)}$$

where

L = the inductance in nanohenries,
N = the number of turns,
μ_i = initial permeability,
A_c = the cross-sectional area of the core in cm^2,
l_e = the effective length of the core in cm.

In order to make calculations easier, most manufacturers have combined μ_i, A_c, l_e, and other constants for a given core into a single quantity called the *inductance index*, A_L. The inductance index relates the inductance to the number of turns for a particular core.

This simplification reduces Equation 1-12 to:

$$L = N^2 A_L \text{ nanohenries} \qquad \text{(Eq. 1-13)}$$

where

L = the inductance in nanohenries,
N = the number of turns,
A_L = the inductance index in $nanohenries/turn^2$.

Material	Application/Classification
Carbonyl C	A medium-*Q* powdered-iron material at 150 kHz. A high-cost material for AM tuning applications and low-frequency IF transformers.
Carbonyl E	The most widely used of all powdered-iron materials. Offers high-*Q* and medium permeability in the 1 MHz to 30 MHz frequency range. A medium-cost material for use in IF transformers, antenna coils, and general-purpose designs.
Carbonyl J	A high-*Q* powdered-iron material at 40 to 100 MHz, with a medium permeability. A high-cost material for FM and TV applications.
Carbonyl SF	Similar to carbonyl E, but with a better *Q* up through 50 MHz. Costs more than carbonyl E.
Carbonyl TH	A powdered-iron material with a higher *Q* than carbonyl E up to 30 MHz, but less than carbonyl SF. Higher cost than carbonyl E.
Carbonyl W	The highest cost powdered-iron material. Offers a high *Q* to 100 MHz, with medium permeability.
Carbonyl HP	Excellent stability and a good *Q* for lower frequency operation—to 50 kHz. A powdered-iron material.
Carbonyl GS6	For commercial broadcast frequencies. Offers good stability and a high *Q*.
IRN-8	A synthetic oxide hydrogen-reduced material with a good *Q* from 50 to 150 MHz. Medium priced for use in FM and TV applications.

TABLE 1-3. Powdered-Iron Materials

Thus, the number of turns to be wound on a given core for a specific inductance is given by:

$$N = \sqrt{\frac{L}{A_L}} \qquad \text{(Eq. 1-14)}$$

This is shown in Example 1-6.

The Q of the inductor cannot be *calculated* with the information given in Fig. 1-25. If we look at the X_p/N^2, R_p/N^2 vs. Frequency curves given for the BBR-7403, however, we can make a calculated guess. At low frequencies (100 kHz), the Q of the coil would be approximately 54, where,

$$Q = \frac{R_p/N^2}{X_p/N^2}$$

$$= \frac{R_p}{X_p} \qquad \text{(Eq. 1-15)}$$

EXAMPLE 1-6

Using the data given in Fig. 1-25, design a toroidal inductor with an inductance of 50 μH. What is the largest AWG wire that we could possibly use while still maintaining a single-layer winding? What is the inductor's Q at 100 MHz?

Solution

There are numerous possibilities in this particular design since no constraints were placed on us. Fig. 1-25 is a data sheet for the Indiana General Series of ferrite toroidal cores. This type of core would normally be used in broadband or low-Q transformer applications rather than in narrow-band tuned circuits. This exercise will reveal why.

The mechanical specifications for this series of cores indicate a fairly typical size for toroids used in small-signal RF circuit design. The largest core for this series is just under a quarter of an inch in diameter. Since no size constraints were placed on us in the problem statement, we will use the AA-03 which has an outside diameter of 0.0230 inch. This will allow us to use a larger diameter wire to wind the inductor.

The published value for A_L for the given core is 495 nH/turn^2. Using Equation 1-14, the number of turns required for this core is:

$$N = \sqrt{\frac{50{,}000\ \text{nH}}{495\ \text{nH/turn}^2}}$$
$$= 10\ \text{turns}$$

Note that the inductance of 50 μH was replaced with its equivalent of 50,000 nH. The next step is to determine the largest diameter wire that can be used to wind the transformer while still maintaining a single-layer winding. In some cases, the data supplied by the manufacturer will include this type of winding information. Thus, in those cases, the designer need only look in a table to determine the maximum wire size that can be used. In our case, this information was not given, so a simple calculation must be made. Fig. 1-27 illustrates the geometry of the problem. It is obvious from the diagram that the inner radius (r_1) of the toroid is the limiting factor in determining the maximum number of turns for a given wire diameter.

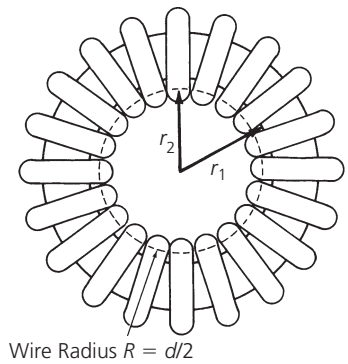


FIG. 1-27. Toroid coil winding geometry.

The exact maximum diameter wire for a given number of turns can be found by:

$$d = \frac{2\pi r_1}{N + \pi} \qquad \text{(Eq. 1-15)}$$

where

d = the diameter of the wire in inches,

r_1 = the inner radius of the core in inches,

N = the number of turns.

For this example, we obtain the value of r_1 from Fig. 1-25 ($d_2 = 0.120$ inch).

$$d = \frac{2\pi \frac{0.120}{2}}{10 + \pi}$$
$$= 28.69 \times 10^{-3}\ \text{inches}$$
$$= 28.69\ \text{mils}$$

As a practical rule of thumb, however, taking into account the insulation thickness variation among manufacturers, it is best to add a "fudge factor" and take 90% of the calculated value, or 25.82 mils. Thus, the largest diameter wire used would be the next size below 25.82 mils, which is AWG No. 22 wire.

As the frequency increases, resistance R_p decreases while reactance X_p increases. At about 3 MHz, X_p equals R_p and the Q becomes unity. The Q then falls below unity until about 100 MHz where resistance R_p begins to increase dramatically and causes the Q to again pass through unity. Thus, due to losses in the core itself, the Q of the coil at 100 MHz is probably very close to 1. Since the Q is so low, this coil would not be a very good choice for use in a narrow-band tuned circuit. See Example 1-7.

PRACTICAL WINDING HINTS

Fig. 1-28 depicts the correct method for winding a toroid. Using the technique of Fig. 1-28A, the interwinding capacitance is minimized, a good portion of the available winding area is utilized, and the resonant frequency of the inductor is increased, thus extending the useful frequency range of the device. Note that by using the methods shown in Figs. 1-28B and 1-28C, both lead capacitance and interwinding capacitance will affect the toroid.

EXAMPLE 1-7

Using the information provided in the data sheet of Fig. 1-26, design a high-Q ($Q > 80$), 300 nH, toroidal inductor for use at 100 MHz. Due to PC board space available, the toroid may not be any larger than 0.3 inch in diameter.

Solution

Fig. 1-26 is an excerpt from an Amidon Associates iron-powder toroidal-core data sheet. The recommended operating frequencies for various materials are shown in the Iron-Powder Material vs. Frequency Range graph. Either material No. 12 or material No. 10 seems to be well suited for operation at 100 MHz. Elsewhere on the data sheet, material No. 12 is listed as IRN-8. (IRN-8 is described in Table 1–3.) Material No. 10 is not described, so choose material No. 12.

Then, under a heading of Iron-Powder Toroidal Cores, the data sheet lists the physical dimensions of the toroids along with the value of A_L for each. Note, however, that this particular company chooses to specify A_L in μH/100 turns rather than μH/100 turns^2. The conversion factor between their value of A_L and A_L in nH/turn^2 is to divide their value of A_L by 10. Thus, the T-80-12 core with an A_L of 22 μH/100 turns is equal to 2.2 nH/turn^2.

Next, the data sheet lists a set of Q-curves for the cores listed in the preceding charts. Note that all of the curves shown indicate Qs that are greater than 80 at 100 MHz.

Choose the largest core available that will fit in the allotted PC board area. The core you should have chosen is the number T-25-12, with an outer diameter of 0.255 inch.

$$A_L = 12\ \mu\text{H}/100\ \text{t}$$
$$= 1.2\ \text{nH/turn}^2$$

Therefore, using Equation 1-14, the number of turns required is

$$N = \sqrt{\frac{L}{A_L}}$$
$$= \sqrt{\frac{300}{1.2}}$$
$$= 15.81$$
$$= 16\ \text{turns}$$

Finally, the chart of Number of Turns vs. Wire Size and Core Size on the data sheet clearly indicates that, for a T-25 size core, the largest size wire we can use to wind this particular toroid is No. 28 AWG wire.

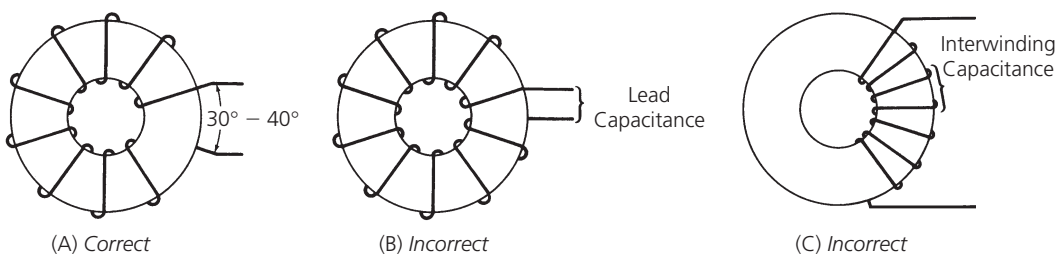


FIG. 1-28. Practical winding hints.

CHAPTER 2

RESONANT Circuits

In this chapter, we will explore the parallel resonant circuit and its characteristics at radio frequencies. We will examine the concept of loaded-Q and how it relates to source and load impedances. We will also see the effects of component losses and how they affect circuit operation. Finally, we will investigate some methods of coupling resonant circuits to increase their selectivity.

SOME DEFINITIONS

The resonant circuit is certainly nothing new in RF circuitry. It is used in practically every transmitter, receiver, or piece of test equipment in existence, to selectively pass a certain frequency or group of frequencies from a source to a load while attenuating all other frequencies outside of this passband. The perfect resonant-circuit *passband* would appear as shown in Fig. 2-1. Here we have a perfect rectangular-shaped passband with infinite attenuation above and below the frequency band of interest, while allowing the desired signal to pass undisturbed. The realization of this filter is, of course, impossible due to the physical characteristics of the components that make up a filter. As we learned in Chapter 1, there is no perfect component and, thus, there can be no perfect filter. If we understand the mechanics of resonant circuits, however, we can certainly tailor an imperfect circuit to suit our needs just perfectly.

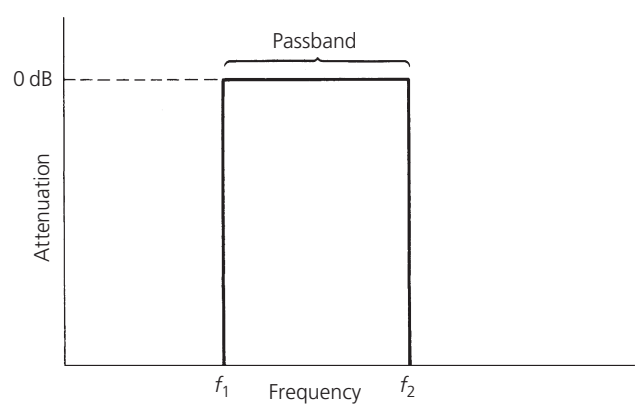


FIG. 2-1. The perfect filter response.

Fig. 2-2 is a diagram of what a practical filter response might resemble. Appropriate definitions are presented below:

1. *Decibel*—In radio electronics and telecommunications, the decibel (dB) is used to describe the ratio between two measurements of electrical power. It can also be combined with a suffix to create an absolute unit of electrical power. For example, it can be combined with "m" for "milliwatt" to produce the "dBm". Zero dBm is one milliwatt, and 1 dBm is one decibel greater than 0 dBm, or about 1.259 mW.

   Decibels are used to account for the gains and losses of a signal from a transmitter to a receiver through some medium (e.g., free space, wave guides, coax, fiber optics, etc.) using a link budget.

2. *Decibel Watts*—The decibel watt (dBw) is a unit for the measurement of the strength of a signal, expressed in decibels relative to one watt. This absolute measurement of electric power is used because of its capability to express both very large and very small values of power in a short range of number, e.g., 10 watts = 10 dBw, and 1,000,000 W = 60 dBw.

3. *Bandwidth*—The bandwidth of any resonant circuit is most commonly defined as being the difference between the upper and lower frequency ($f_2 - f_1$) of the circuit at which its amplitude response is 3 dB below the passband response. It is often called the half-power bandwidth.

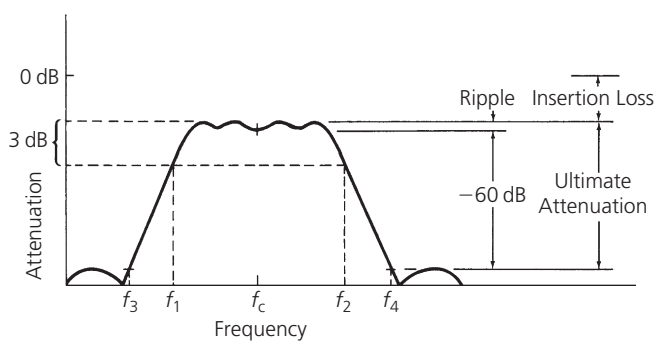


FIG. 2-2. A practical filter response.

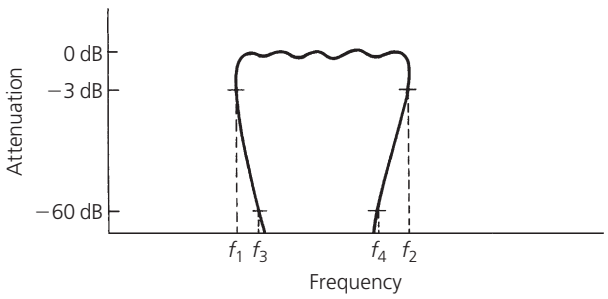


FIG. 2-3. An impossible shape factor.

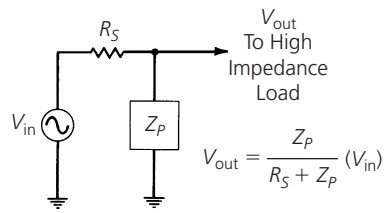


FIG. 2-4. Voltage division rule.

4. *Q*—The ratio of the center frequency of the resonant circuit to its bandwidth is defined as the circuit *Q*.

$$Q = \frac{f_e}{f_2 - f_1} \qquad \text{(Eq. 2-1)}$$

This *Q* should not be confused with component *Q* which was defined in Chapter 1. Component *Q* does have an effect on circuit *Q*, but the reverse is not true. Circuit *Q* is a measure of the selectivity of a resonant circuit. The higher its *Q*, the narrower its bandwidth, the higher is the selectivity of a resonant circuit.

5. *Shape Factor*—The shape factor of a resonant circuit is typically defined as being the ratio of the 60-dB bandwidth to the 3-dB bandwidth of the resonant circuit. Thus, if the 60-dB bandwidth ($f_4 - f_3$) were 3 MHz and the 3-dB bandwidth ($f_2 - f_1$) were 1.5 MHz, then the shape factor would be:

$$SF = \frac{3\,\text{MHz}}{1.5\,\text{MHz}} = 2$$

Shape factor is simply a degree of measure of the steepness of the skirts. The smaller the number, the steeper are the response skirts. Notice that our perfect filter in Fig. 2-1 has a shape factor of 1, which is the ultimate. The passband for a filter with a shape factor smaller than 1 would have to look similar to the one shown in Fig. 2-3. Obviously, this is a physical impossibility.

6. *Ultimate Attenuation*—Ultimate attenuation, as the name implies, is the final *minimum* attenuation that the resonant circuit presents outside of the specified passband. A perfect resonant circuit would provide infinite attenuation outside of its passband. However, due to component imperfections, infinite attenuation is infinitely impossible to get. Keep in mind also, that if the circuit presents response peaks outside of the passband, as shown in Fig. 2-2, then this of course detracts from the ultimate attenuation specification of that resonant circuit.

7. *Insertion Loss*—Whenever a component or group of components is inserted between a generator and its load, some of the signal from the generator is absorbed in those components due to their inherent resistive losses. Thus, not as much of the transmitted signal is transferred to the load as when the load is connected directly to the generator. (I am assuming here that no impedance matching function is being performed.) The attenuation that results is called *insertion loss* and it is a very important characteristic of resonant circuits. It is usually expressed in decibels (dB).

8. *Ripple*—Ripple is a measure of the flatness of the passband of a resonant circuit and it is also expressed in decibels. Physically, it is measured in the response characteristics as the difference between the maximum attenuation *in the passband* and the minimum attenuation in the passband. In Chapter 3, we will actually design filters for a specific passband ripple.

RESONANCE (LOSSLESS COMPONENTS)

In Chapter 1, the concept of resonance was briefly mentioned when we studied the parasitics associated with individual component elements. We will now examine the subject of resonance in detail. We will determine what causes resonance to occur and how we can use it to our best advantage.

The voltage division rule (illustrated in Fig. 2-4) states that whenever a shunt element of impedance Z_p is placed across the output of a generator with an internal resistance R_s, the maximum output voltage available from this circuit is

$$V_{out} = \frac{Z_p}{R_s + Z_p}(V_{in}) \qquad \text{(Eq. 2-2)}$$

Thus, V_{out} will always be less than V_{in}. If Z_p is a frequency-dependent impedance, such as a capacitive or inductive reactance, then V_{out} will also be frequency dependent and the ratio of V_{out} to V_{in}, which is the gain (or, in this case, loss) of the circuit, will also be frequency dependent. Let's take, for example, a 25-pF capacitor as the shunt element (Fig. 2-5A) and plot the function of V_{out}/V_{in} in dB versus frequency, where we have:

$$\frac{V_{out}}{V_{in}} = 20\log_{10}\frac{X_C}{R_s + X_C} \qquad \text{(Eq. 2-3)}$$

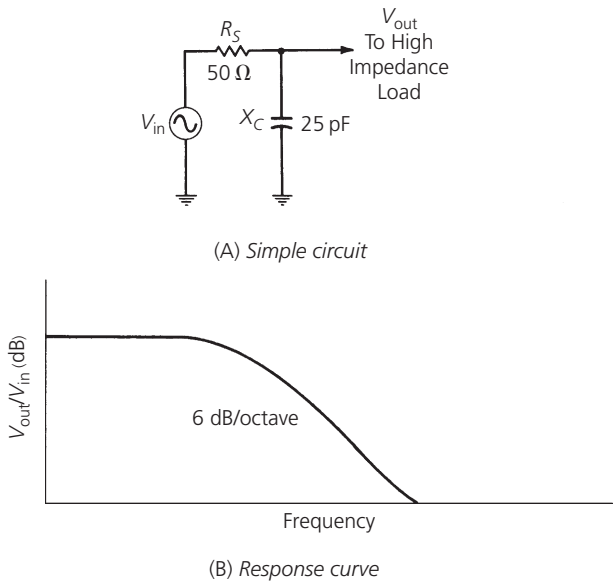


FIG. 2-5. Frequency response of a simple RC (resistor-capacitor) low-pass filter.

where

$\frac{V_{out}}{V_{in}}$ = the loss in dB,

R_s = the source resistance,

X_c = the reactance of the capacitor.

and, where

$$X_C = \frac{1}{j\omega C}.$$

The plot of this equation is shown in the graph of Fig. 2-5B. Notice that the loss of this RC (resistor-capacitor) circuit increases as the frequency increases; thus, we have formed a simple *low-pass filter* (e.g., a filter that passes low frequency signals but attenuates (or reduces the amplitude of) signals with frequencies higher than the cutoff frequency). Notice, also, that the attenuation slope eventually settles down to the rate of 6 dB for every octave (doubling) increase in frequency. This is due to the single reactive element in the circuit. As we will see later, this attenuation slope will increase an additional 6 dB for each *significant* reactive element that we insert into the circuit.

If we now delete the capacitor from the circuit and insert a 0.05-μH inductor in its place, we obtain the circuit of Fig. 2-6A and the plot of Fig. 2-6B, where we are plotting:

$$\frac{V_{out}}{V_{in}} = 20\log_{10}\frac{X_L}{R_s + X_L} \qquad \text{(Eq. 2-4)}$$

where

$\frac{V_{out}}{V_{in}}$ = the loss in dB,

R_s = the source resistance,

X_L = the reactance of the coil.

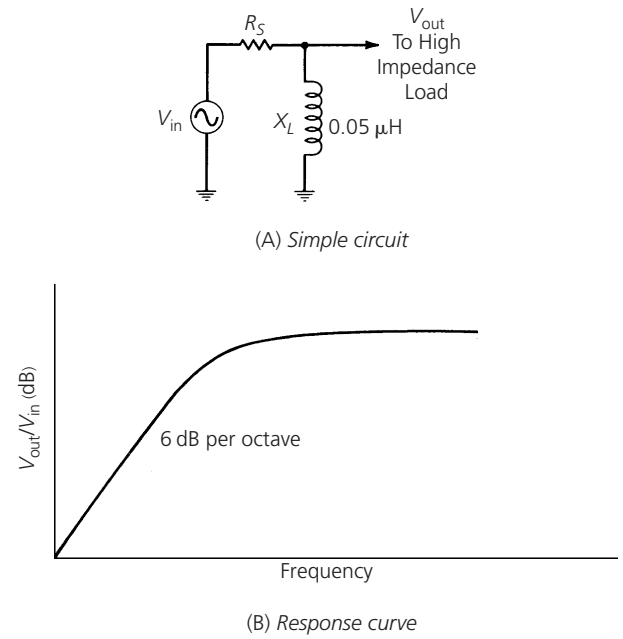


FIG. 2-6. Simple high-pass filter.

and, where

$$X_L = j\omega L.$$

Here, we have formed a simple *high-pass filter* (e.g., a filter that passes high frequencies well, but attenuates (or reduces) frequencies lower than the cutoff frequency) with a final attenuation slope of 6 dB/octave.

Thus, through simple calculations involving the basic voltage division formula (Equation 2-2), we were able to plot the frequency response of two separate and opposite reactive components. But what happens if we place both the inductor and capacitor across the generator simultaneously, thereby creating an LC (inductor-capacitor) circuit? Actually, this case is no more difficult to analyze than the previous two circuits. In fact, at any frequency, we can simply apply the basic voltage division rule as before. The only difference here is that we now have two reactive components to deal with instead of one and these components are in parallel (Fig. 2-7). If we make the calculation for all frequencies of interest, we will obtain the plot shown in Fig. 2-8.

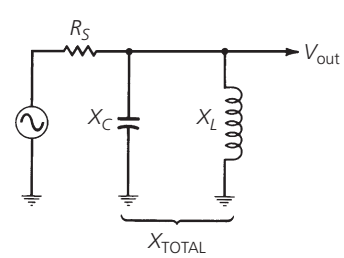


FIG. 2-7. Resonant circuit with two reactive components.

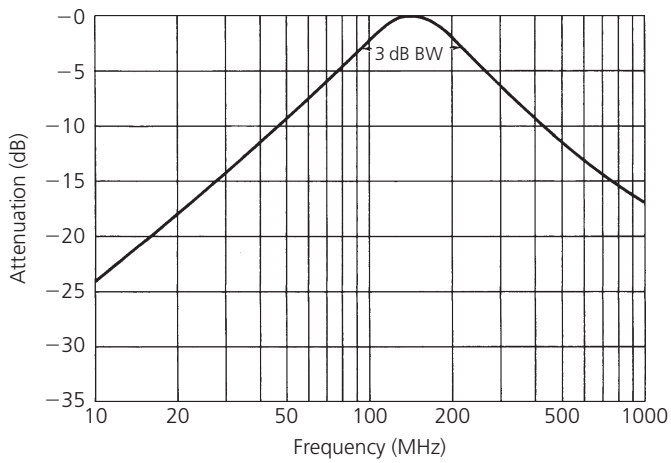


FIG. 2-8. Frequency response of an LC (inductor–capacitor) resonant circuit.

The mathematics behind this calculation are as follows:

$$V_{\text{out}} = \frac{X_{\text{total}}}{R_s + X_{\text{total}}}(V_{\text{in}}) \qquad \text{(Eq. 2-5)}$$

where

$$X_{\text{total}} = \frac{X_C X_L}{X_C + X_L}.$$

and, where

$$X_C = \frac{1}{j\omega C},$$
$$X_L = j\omega C$$

Therefore, we have:

$$X_{\text{total}} = \frac{\frac{1}{j\omega C}(j\omega L)}{\frac{1}{j\omega C} + j\omega L}$$
$$= \frac{\frac{L}{C}}{\frac{1}{j\omega C} + j\omega L}$$

Multiply the numerator and the denominator by $j\omega C$. (Remember that $j^2 = -1$.)

$$X_{\text{total}} = \frac{j\omega L}{1 + (j\omega L)\,(j\omega C)}$$
$$= \frac{j\omega L}{1 - \omega^2 LC}$$

Thus, substituting and transposing in Equation 2-5, we have:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{\frac{j\omega L}{1-\omega^2 LC}}{R_s + \frac{j\omega L}{1-\omega^2 LC}}$$

Multiplying the numerator and the denominator through by $1 - \omega^2 LC$ yields:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{j\omega L}{(R_s - \omega^2 R_s LC) + j\omega L}$$

Thus, the loss at any frequency may be calculated from the above equation or, if needed, in dB.

$$\frac{V_{\text{out}}}{V_{\text{in}}} = 20\log_{10}\left|\frac{j\omega L}{(R_s - \omega^2 R_s LC) + j\omega L}\right|$$

where | | represents the magnitude of the quantity within the brackets.

Notice, in Fig. 2-8, that as we near the resonant frequency of the tuned circuit, the slope of the resonance curve increases to 12 dB/octave. This is due to the fact that we now have two *significant* reactances present and each one is changing at the rate of 6 dB/octave and sloping in opposite directions. As we move away from resonance in either direction, however, the curve again settles to a 6-dB/octave slope because, again, only one reactance becomes significant. The other reactance presents a very high impedance to the circuit at these frequencies and the circuit behaves as if the reactance were no longer there.

Unlike the high-pass or low-pass filters discussed here, the RLC circuit (also known as a resonant or tuned circuit) does something different. As an electrical circuit consisting of a resistor (R), an inductor (L), and a capacitor (C), connected in series or in parallel, the RLC circuit has many applications, particularly in radio and communications engineering. They can be used, for example, to select a certain narrow range of frequencies from the total spectrum of ambient radio waves. In this next section, we will take a closer look at what the RLC circuit can do for the RF engineer.

LOADED *Q*

The Q of a resonant circuit was defined earlier to be equal to the ratio of the center frequency of the circuit to its 3-dB bandwidth (Equation 2-1). This "circuit Q," as it was called, is often given the label *loaded Q* because it describes the passband characteristics of the resonant circuit under actual in-circuit or *loaded* conditions. The loaded Q of a resonant circuit is dependent upon three main factors. (These are illustrated in Fig. 2-9.)

1. The source resistance (R_s).
2. The load resistance (R_L).
3. The component Q as defined in Chapter 1.

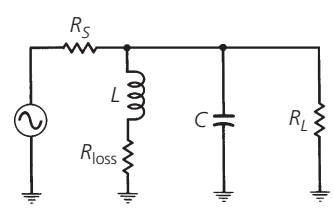


FIG. 2-9. Circuit for loaded-Q calculations.

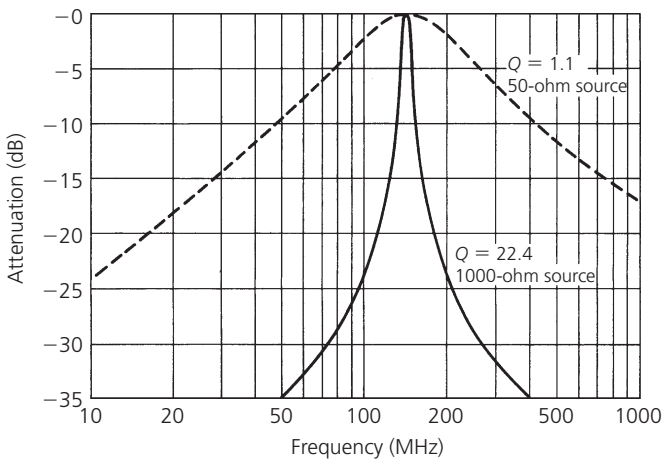


FIG. 2-10. The effect of R_s and R_L on loaded Q.

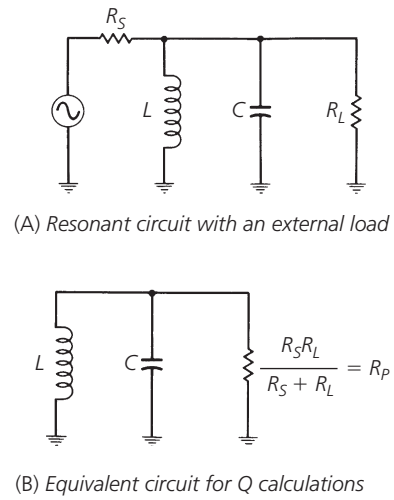


FIG. 2-11. The equivalent parallel impedance across a resonant circuit.

Effect of R_s and R_L on the Loaded Q

Let's discuss briefly the role that source and load impedances play in determining the loaded Q of a resonant circuit. This role is probably best illustrated through an example. In Fig. 2-8, we plotted a resonance curve for a circuit consisting of a 50-ohm source, a 0.05-μH lossless inductor, and a 25-pF lossless capacitor. The loaded Q of this circuit, as defined by Eq. 2-1 and determined from the graph, is approximately 1.1. Obviously, this is not a very narrow-band or high-Q design. But now, let's replace the 50-ohm source with a 1000-ohm source and again plot our results using the equation derived in Fig. 2-7 (Equation 2-5). This new plot is shown in Fig. 2-10. (The resonance curve for the 50-ohm source circuit is shown with dashed lines for comparison purposes.) Notice that the Q, or selectivity of the resonant circuit, has been increased dramatically to about 22. Thus, by raising the source impedance, we have increased the Q of our resonant circuit.

Neither of these plots addresses the effect of a load impedance on the resonance curve. If an external load of some sort were attached to the resonant circuit, as shown in Fig. 2-11A, the effect would be to broaden or "de-Q" the response curve to a degree that depends on the value of the load resistance. The equivalent circuit, for resonance calculations, is shown in Fig. 2-11B. The resonant circuit sees an equivalent resistance of R_s in parallel with R_L, as its true load. This total external resistance is, by definition, smaller in value than either R_s or R_L, and the loaded Q must decrease. If we put this observation in equation form, it becomes (assuming lossless components):

$$Q = \frac{R_p}{X_p} \qquad \text{(Eq. 2-6)}$$

where

R_p = the equivalent parallel resistance of R_s and R_L,

X_p = either the inductive or capacitive reactance. (They are equal at resonance.)

Equation 2-6 illustrates that a decrease in R_p will decrease the Q of the resonant circuit and an increase in R_p will increase the circuit Q, and it also illustrates another very important point. The same effect can be obtained by keeping R_p constant and varying X_p. Thus, for a given source and load impedance, the optimum Q of a resonant circuit is obtained when the inductor is a small value and the capacitor is a large value. Therefore, in either case, X_p is decreased. This effect is shown using the circuits in Fig. 2-12 and the characteristics curves in Fig. 2-13.

The circuit designer, therefore, has two approaches he can follow in designing a resonant circuit with a particular Q (Example 2-1).

1. He can select an optimum value of source and load impedance.
2. He can select component values of L and C that optimize Q.

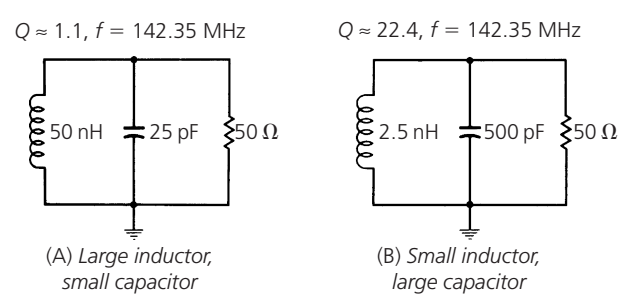


FIG. 2-12. Effect of Q vs. X_p at 142.35 MHz.

Often there is no real choice in the matter because, in many instances, the source and load are defined and we have no control over them. When this occurs, X_p is automatically defined for

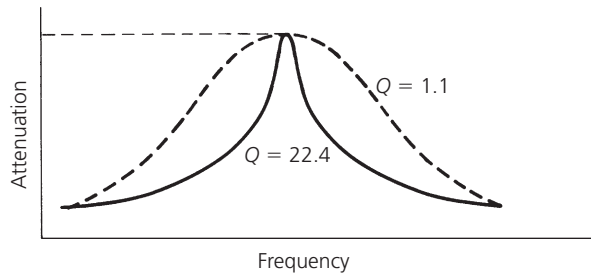


FIG. 2-13. Plot of loaded-Q curves for circuits in Fig. 2-12.

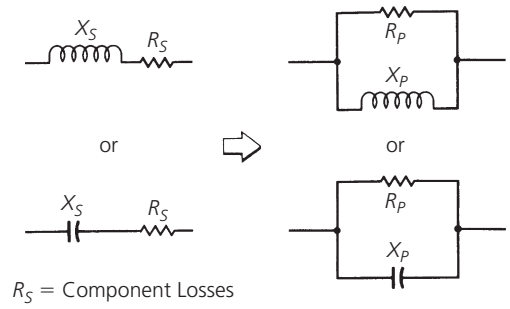


FIG. 2-14. A series-to-parallel transformation.

EXAMPLE 2-1

Design a resonant circuit to operate between a source resistance of 150 ohms and a load resistance of 1000 ohms. The loaded Q must be equal to 20 at the resonant frequency of 50 MHz. Assume lossless components and no impedance matching.

Solution

The effective parallel resistance across the resonant circuit is 150 ohms in parallel with 1000 ohms, or

$$R_p = 130 \text{ ohms}$$

Thus, using Eq. 2-6:

$$X_p = \frac{R_p}{Q} = \frac{130}{20} = 6.5 \text{ ohms}$$

and

$$X_p = \omega L = \frac{1}{\omega C}$$

Therefore, $L = 20.7$ nH, and $C = 489.7$ pF.

a given Q and we usually end up with component values that are impractical at best. Later in this chapter, we will study some methods of eliminating this problem.

The Effect of Component Q on Loaded Q

Thus far in this chapter, we have assumed that the components used in the resonant circuits are lossless and, thus, produce no degradation in loaded Q. In reality, however, such is not the case and the individual component Q's must be taken into account. In a lossless resonant circuit, the impedance seen across the circuit's terminals at resonance is infinite. In a practical circuit, however, due to component losses, there exists some finite equivalent parallel resistance. This is illustrated in Fig. 2-14. The resistance (R_p) and its associated shunt reactance (X_p) can be found from the following transformation equations:

$$R_p = (Q^2 + 1)R_s \qquad \text{(Eq. 2-7)}$$

where

R_p = the equivalent parallel resistance,
R_s = the series resistance of the component,
$Q = Q_s$ which equals Q_p which equals the Q of the component.

and

$$X_p = \frac{R_p}{Q_p} \qquad \text{(Eq. 2-8)}$$

If the Q of the component is greater than 10, then,

$$R_p \approx Q^2 R_s \qquad \text{(Eq. 2-9)}$$

and

$$X_p \approx X_s \qquad \text{(Eq. 2-10)}$$

These transformations are valid at only one frequency because they involve the component reactance which is frequency dependent (Example 2-2).

Example 2-2 vividly illustrates the potential drastic effects that can occur if poor-quality (low Q) components are used in highly selective resonant circuit designs. The net result of this action is that we effectively place a low-value shunt resistor directly across the circuit. As was shown earlier, any low-value resistance that shunts a resonant circuit drastically reduces its loaded Q and, thus, increases its bandwidth.

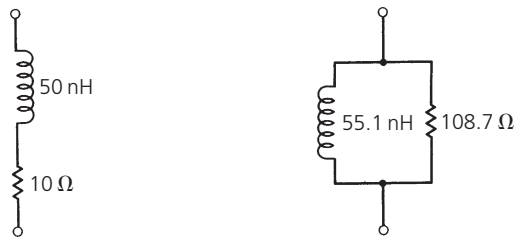
In most cases, we only need to involve the Q of the inductor in loaded-Q calculations. The Q of most capacitors is quite high over their useful frequency range, and the equivalent shunt resistance they present to the circuit is also quite high and can usually be neglected. Care must be taken, however, to ensure that this is indeed the case.

INSERTION LOSS

Insertion loss (defined earlier in this chapter) is another direct effect of component Q. If inductors and capacitors were perfect and contained no internal resistive losses, then insertion loss for LC resonant circuits and filters would not exist. This is, of course,

EXAMPLE 2-2

Given a 50-nH coil as shown in Fig. 2-15A, compute its Q at 100 MHz. Then, transform the series circuit of Fig. 2-15A into the equivalent parallel inductance and resistance circuit of Fig. 2-15B.



(A) *Series circuit* (B) *Equivalent parallel circuit*

FIG. 2-15. Example of a series-to-parallel transformation.

Solution

The Q of this coil at 100 MHz is, from Chapter 1,

$$Q = \frac{X_s}{R_s} = \frac{2\pi(100 \times 10^6)(50 \times 10^{-9})}{10} = 3.14$$

Then, since the Q is less than 10, use Equation 2-7 to find R_p.

$$R_p = (Q^2 + 1)R_s = [(3.14)^2 + 1]10 = 108.7 \text{ ohms}$$

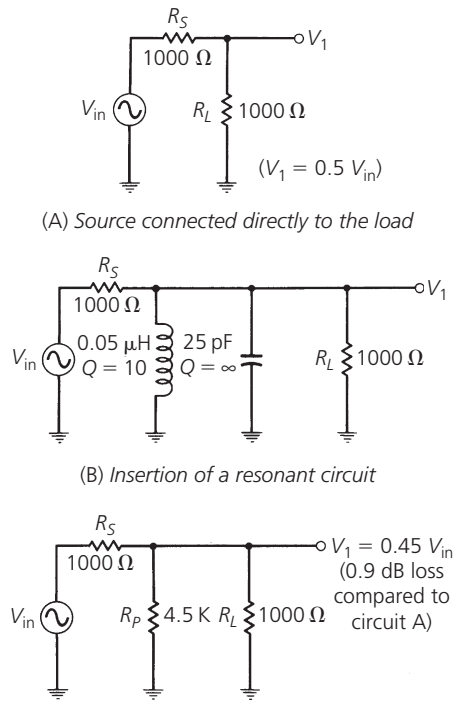
Next, we find X_p using Equation 2-8:

$$X_p = \frac{R_p}{Q_p} = \frac{108.7}{3.14} = 34.62$$

Thus, the parallel inductance becomes:

$$L_p = \frac{X_p}{\omega} = \frac{34.62}{2\pi(100 \times 10^6)} = 55.1 \text{ nH}$$

These values are shown in the equivalent circuit of Fig. 2-15B.



(A) *Source connected directly to the load*

(B) *Insertion of a resonant circuit*

(C) *Equivalent circuit at resonance*

FIG. 2-16. The effect of component Q on insertion loss.

not the case and, as it turns out, insertion loss is a very critical parameter in the specification of any resonant circuit.

Fig. 2-16 illustrates the effect of inserting a resonant circuit between a source and its load. In Fig. 2-16A, the source is connected directly to the load. Using the voltage division rule, we find that:

$$V_1 = 0.5\, V_{in}$$

Fig. 2-16B shows that a resonant circuit has been placed between the source and the load. Then, Fig. 2-16C illustrates the equivalent circuit at resonance. Notice that the use of an inductor with a Q of 10 at the resonant frequency creates an effective shunt resistance of 4500 ohms at resonance. This resistance, combined with R_L, produces an 0.9-dB voltage loss at V_1 when compared to the equivalent point in the circuit of Fig. 2-16A.

An insertion loss of 0.9 dB doesn't sound like much, but it can add up very quickly if we cascade several resonant circuits. We will see some very good examples of this later in Chapter 3. For now, examine the problem given in Example 2-3.

IMPEDANCE TRANSFORMATION

As we have seen in earlier sections of this chapter, low values of source and load impedance tend to load a given resonant circuit down and, thus, tend to decrease its loaded Q and increase its bandwidth. This makes it very difficult to design a simple LC high-Q resonant circuit for use between two very low values of

EXAMPLE 2-3

Design a simple parallel resonant circuit to provide a 3-dB bandwidth of 10 MHz at a center frequency of 100 MHz. The source and load impedances are each 1000 ohms. Assume the capacitor to be lossless. The Q of the inductor (that is available to us) is 85. What is the insertion loss of the network?

Solution

From Equation 2-1, the required loaded Q of the resonant circuit is:

$$Q = \frac{f_e}{f_2 - f_1} = \frac{100\,\text{MHz}}{10\,\text{MHz}} = 10$$

To find the inductor and capacitor values needed to complete the design, it is necessary that we know the equivalent shunt resistance and reactance of the components at resonance. Thus, from Equation 2-8:

$$X_p = \frac{R_p}{Q_p}$$

where

X_p = the reactance of the inductor and capacitor at resonance,
R_p = the equivalent shunt resistance of the inductor,
Q_p = the Q of the inductor.

Thus,

$$R_p = (85)X_p \qquad \text{(Eq. 2-11)}$$

The loaded Q of the resonant circuit is equal to:

$$Q = \frac{R_{\text{total}}}{X_p}$$

$$10 = \frac{R_{\text{total}}}{X_p}$$

where

R_{total} = the shunt resistance, which equals $R_p \| R_s \| R_L$.
Therefore, we have:

$$10 = \frac{\frac{R_p(500)}{R_p + 500}}{X_p} \qquad \text{(Eq. 2-12)}$$

We now have two equations and two unknowns (X_p, R_p). If we substitute Equation 2-11 into Equation 2-12 and solve for X_p, we get:

$$X_p = 44.1 \text{ ohms}$$

Plugging this value back into Equation 2-11 gives:

$$R_p = 3.75\text{K}$$

Thus, our component values must be

$$L = \frac{X_p}{\omega} = 70\,\text{nH}$$

$$C = \frac{1}{\omega X_p} = 36\,\text{pF}$$

The final circuit is shown in Fig. 2-17.

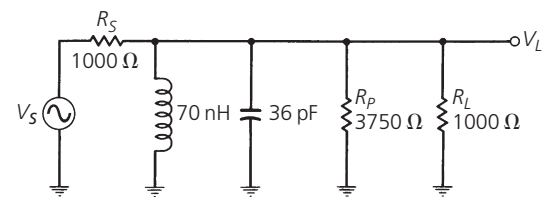


FIG. 2-17. Resonant circuit design for Example 2-3.

The insertion-loss calculation, at center frequency, is now very straightforward and can be found by applying the voltage division rule as follows. Resistance R_p in parallel with resistance R_L is equal to 789.5 ohms. The voltage at V_L is, therefore,

$$V_L = \frac{789.5}{789.5 + 1000}(V_s) = .44\,V_s$$

The voltage at V_L, without the resonant circuit in place, is equal to 0.5 V_s due to the 1000-ohm load. Thus, we have:

$$\text{Insertion Loss} = 20\log_{10}\frac{0.44\,V_s}{0.5\,V_s} = 1.1\,\text{dB}$$

source and load resistance. In fact, even if we were able to come up with a design on paper, it most likely would be impossible to build due to the extremely small (or negative) inductor values that would be required.

One method of getting around this potential design problem is to make use of one of the impedance transforming circuits shown in Fig. 2-18. These handy circuits fool the resonant circuit into seeing a source or load resistance that is much larger than what is actually present. For example, an impedance transformer could present an impedance (R_s') of 500 ohms to the resonant circuit, when in reality there is an impedance (R_s) of 50 ohms. Consequently, by utilizing these transformers, both the Q of the resonant tank and its selectivity can be increased. In many cases, these methods can make a previously unworkable problem workable again, complete with realistic values for the coils and capacitors involved.

The design equations for each of the transformers are presented in the following equations and are useful for designs that need

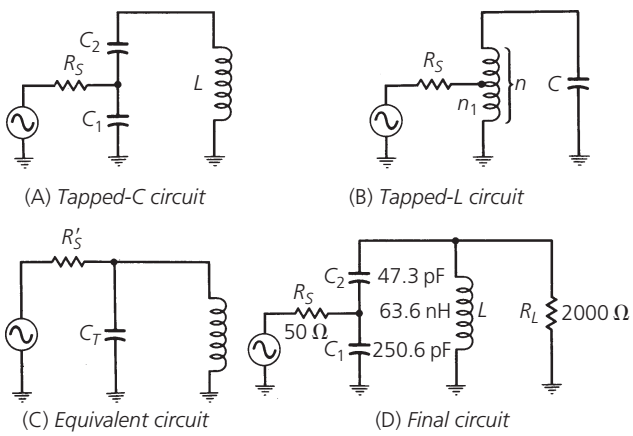


FIG. 2-18. Two methods used to perform an impedance transformation.

loaded Q's that are greater than 10 (Example 2-4). For the tapped-C transformer (Fig. 2-18A), we use the formula:

$$R'_s = R_s\left(1 + \frac{C_1}{C_2}\right)^2 \qquad \text{(Eq. 2-13)}$$

The equivalent capacitance (C_T) that will resonate with the inductor is equal to C_1 in series with C_2, or:

$$C_T = \frac{C_1 C_2}{C_1 + C_2} \qquad \text{(Eq. 2-14)}$$

For the tapped-L network of Fig. 2-18B, we use the following formula:

$$R'_s = R_s\left(\frac{n}{n_1}\right)^2 \qquad \text{(Eq. 2-15)}$$

As an exercise, you might want to rework Example 2-4 without the aid of an impedance transformer. You will find that the

EXAMPLE 2-4

Design a resonant circuit with a loaded Q of 20 at a center frequency of 100 MHz that will operate between a source resistance of 50 ohms and a load resistance of 2000 ohms. Use the tapped-C approach and assume that inductor Q is 100 at 100 MHz.

Solution

We will use the tapped-C transformer to step the source resistance up to 2000 ohms to match the load resistance for optimum power transfer. (Impedance matching will be covered in detail in Chapter 4.) Thus,

$$R'_s = 2000 \text{ ohms}$$

and from Equation 2-13, we have:

$$\frac{C_1}{C_2} = \sqrt{\frac{R'_s}{R_s}} - 1$$
$$= 5.3$$

or,

$$C_1 = 5.3C_2 \qquad \text{(Eq. 2-16)}$$

Proceeding as we did in Example 2-3, we know that for the inductor:

$$Q_p = \frac{R_p}{X_p} = 100$$

Therefore,

$$R_p = 100\,X_p \qquad \text{(Eq. 2-17)}$$

We also know that the loaded Q of the resonant circuit is equal to:

$$Q = \frac{R_{\text{total}}}{X_p}$$

where

R_{total} = the total equivalent shunt resistance,

$$= R'_s \| R_p \| R_L$$
$$= 1000 \| R_p$$

and, where we have taken R'_s and R_L to each be 2000 ohms, in parallel. Hence, the loaded Q is

$$Q = \frac{1000R_p}{(1000 + R_p)X_p} \qquad \text{(Eq. 2-18)}$$

Substituting Equation 2-17 (and the value of the desired loaded Q) into Equation 2-18, and solving for X_p, yields:

$$X_p = 40 \text{ ohms}$$

And, substituting this result back into Equation 2-17 gives

$$R_p = 4000 \text{ ohms}$$

and

$$L = \frac{X_p}{\omega}$$
$$= 63.6 \text{ nH}$$
$$C_T = \frac{1}{X_p \omega}$$
$$= 39.78 \text{ pF}$$

We now know what the total capacitance must be to resonate with the inductor. We also know from Equation 2-16 that C_1 is 5.3 times larger than C_2. Thus, if we substitute Equation 2-16 into Equation 2-14, and solve the equations simultaneously, we get:

$$C_2 = 47.3 \text{ pF}$$
$$C_1 = 250.6 \text{ pF}$$

The final circuit is shown in Fig. 2-18D.

inductor value which results is much more difficult to obtain and control physically because it is so small.

COUPLING OF RESONANT CIRCUITS

In many applications where steep passband skirts and small shape factors are needed, a single resonant circuit might not be sufficient. In situations such as this, individual resonant circuits are often coupled together to produce more attenuation at certain frequencies than would normally be available with a single resonator. The coupling mechanism that is used is generally chosen specifically for each application, as each type of coupling has its own peculiar characteristics that must be dealt with. The most common forms of coupling are: capacitive, inductive, transformer (mutual), and active (transistor).

Capacitive Coupling

Capacitive coupling is probably the most frequently used method of linking two or more resonant circuits. This is true mainly due to the simplicity of the arrangement but another reason is that it is relatively inexpensive. Fig. 2-19 indicates the circuit arrangement for a two-resonator capacitively coupled filter.

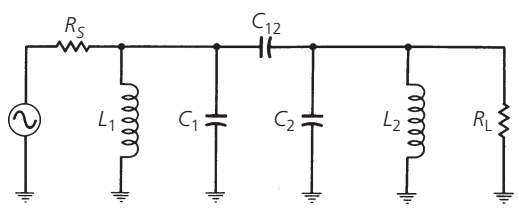


FIG. 2-19. Capacitive coupling.

The value of the capacitor that is used to couple each resonator cannot be just chosen at random, as Fig. 2-20 indicates. If capacitor C_{12} of Fig. 2-19 is too large, too much coupling occurs and the frequency response broadens drastically with two response peaks in the filter's passband. If capacitor C_{12} is too small, not enough signal energy is passed from one resonant circuit to the other and the insertion loss can increase to an unacceptable level. The compromise solution to these two extremes is the point of *critical coupling*, where we obtain a reasonable bandwidth and the lowest possible insertion loss and, consequently, a maximum transfer of signal power. There are instances in which overcoupling or undercoupling might serve a useful purpose in a design, such as in tailoring a specific frequency response that a critically coupled filter cannot provide. But these applications are generally left to the multiple resonator filter. The multiple resonator filter is covered in Chapter 3. In this section, we will only concern ourselves with critical coupling as it pertains to resonant circuit design.

The loaded Q of a critically coupled two-resonator circuit is approximately equal to 0.707 times the loaded Q of one of its resonators. Therefore, the 3-dB bandwidth of a two-resonator circuit is actually wider than that of one of its resonators. This might

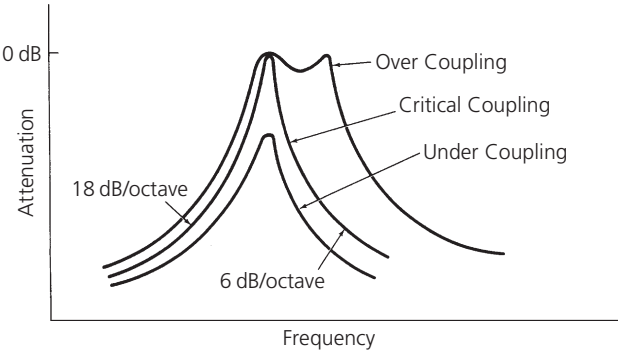


FIG. 2-20. The effects of various values of capacitive coupling on passband response.

seem contrary to what we have studied so far, but remember, the main purpose of the two-resonator passively coupled filter is not to provide a narrower 3-dB bandwidth, but to increase the steepness of the stopband skirts and, thus, to reach an ultimate attenuation much faster than a single resonator could. This characteristic is shown in Fig. 2-21. Notice that the shape factor has decreased for the two-resonator design. Perhaps one way to get an intuitive feel for how this occurs is to consider that each resonator is itself a load for the other resonator, and each decreases the loaded Q of the other. But as we move away from the passband and into the stopband, the response tends to fall much more quickly due to the combined response of each resonator.

The value of the capacitor used to couple two identical resonant circuits is given by

$$C_{12} = \frac{C}{Q} \qquad \text{(Eq. 2-19)}$$

where

C_{12} = the coupling capacitance.
C = the resonant circuit capacitance.
Q = the loaded Q of a single resonator.

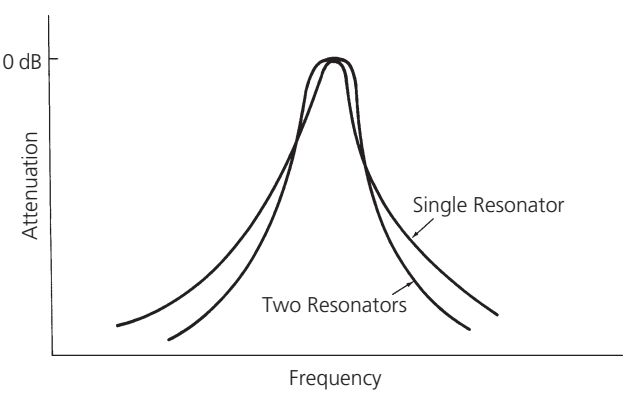


FIG. 2-21. Selectivity of single- and two-resonator designs.

One other important characteristic of a capacitively coupled resonant circuit can be seen if we take another look at Fig. 2-20. Notice that even for the critically coupled case, the response curve is not symmetric around the center frequency but is skewed somewhat. The lower frequency portion of the response plummets down at the rate of 18-dB/octave while the upper slope decreases at only 6-dB/octave. This can be explained if we take a look at the equivalent circuit both above and below resonance. Below resonance, we have the circuit of Fig. 2-22A. The reactance of the two resonant-circuit capacitors (Fig. 2-19) has increased, and the reactance of the two inductors has decreased to the point that only the inductor is seen as a shunt element and the capacitors can be ignored. This leaves three reactive components and each contributes 6-dB/octave to the response.

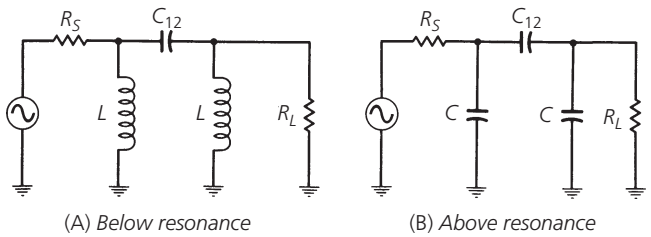


FIG. 2-22. Equivalent circuit of capacitively coupled resonant circuits.

On the high side of resonance, the equivalent circuit approaches the configuration of Fig. 2-22B. Here the inductive reactance has increased above the capacitive reactance to the point where the inductive reactance can be ignored as a shunt element. We now have an arrangement of three capacitors that effectively looks like a single shunt capacitor and yields a slope of 6-dB/octave.

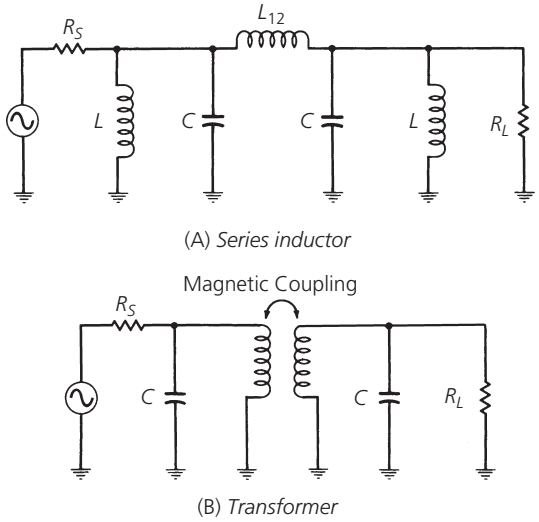


FIG. 2-23. Inductive coupling.

Inductive Coupling

Two types of inductively coupled resonant circuits are shown in Fig. 2-23. One type (Fig. 2-23A) uses a series inductor or coil to transfer energy from the first resonator to the next, and the other type (Fig. 2-23B) uses transformer coupling for the same purpose. In either case, the frequency response curves will resemble those of Fig. 2-24 depending on the amount of coupling. If we compare Fig. 2-24A with Fig. 2-20, we see that the two are actually mirror images of each other. The response of die inductively coupled resonator is skewed toward the higher end of the frequency spectrum, while the capacitively coupled response is skewed toward the low frequency side. An examination of the equivalent circuit reveals why. Fig. 2-25A indicates that below resonance, the capacitors drop out of the equivalent circuit very quickly because their reactance becomes much greater than the shunt inductive reactance. This leaves an arrangement of three inductors which can be thought of as a single tapped inductor and which produces a 6-dB/octave rolloff. Above resonance, the shunt inductors can be ignored for the same reasons, and you have the circuit of Fig. 2-25B. We now have three effective elements in the equivalent circuit with each contributing 6-dB/octave to the response for a combined slope of 18-dB/octave.

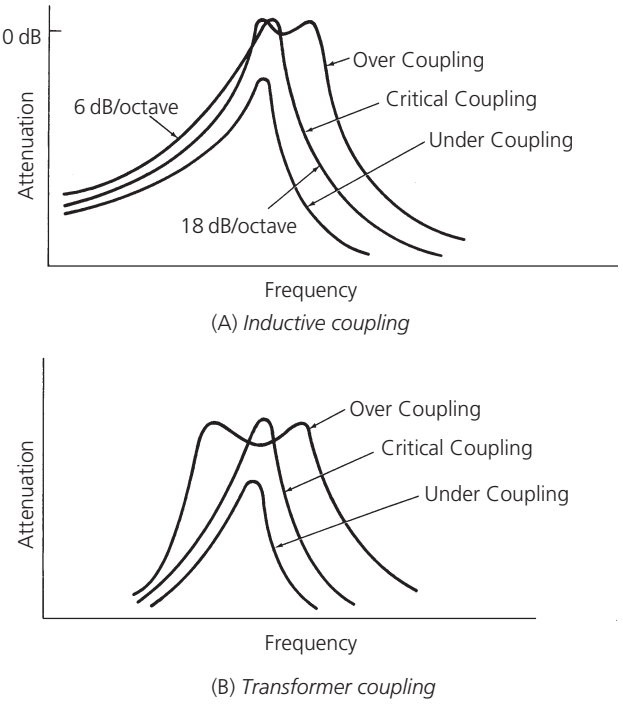
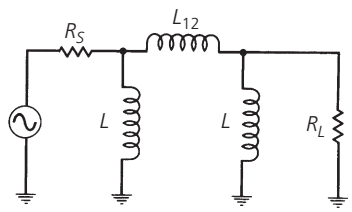
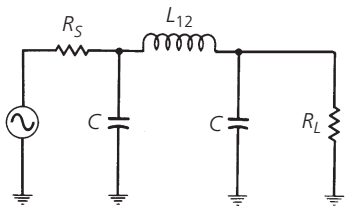


FIG. 2-24. The effects of various values of inductive coupling on passband response.

The mirror-image characteristic of inductively and capacitively coupled resonant circuits is a very useful concept. This is especially true in applications that require symmetrical response curves. For example, suppose that a capacitively coupled design exhibited too much skew for your application. One very simple



(A) *Below resonance*



(B) *Above resonance*

FIG. 2-25. Equivalent circuit of inductively coupled resonant circuits.

way to correct the problem would be to add a "top-L" coupled section to the existing network. The top-L coupling would attempt to skew the response in the opposite direction and would, therefore, tend to counteract any skew caused by the capacitive coupling. The net result is a more symmetric response shape.

The value of the inductor used to couple two identical resonant circuits can be found by

$$L_{12} = QL \qquad \text{(Eq. 2-20)}$$

where

L_{12} = the inductance of the coupling inductor,

Q = the loaded Q of a single resonator,

L = the resonant circuit inductance.

A little manipulation of Equations 2-19 and 2-20 will reveal a very interesting point. The reactance of C_{12} calculated with Equation 2-19 will equal the reactance of L_{12} calculated with Equation 2-20 for the same operating Q and resonant frequency. The designer now has the option of changing any "top-C" coupled resonator to a top-L design simply by replacing the coupling capacitor with an inductor of equal reactance at the resonant frequency. When this is done, the degree of coupling, Q, and resonant frequency of the design will remain unchanged while the slope of the stopband skirts will flip-flop from one side to the other. For obvious reasons, top-L coupled designs work best in applications where the primary objective is a certain ultimate attenuation that must be met above the passband. Likewise, top-C designs are best for meeting ultimate attenuation specifications below the passband.

Transformer coupling does not lend itself well to an exact design procedure because there are so many factors that influence the degree of coupling. The geometry of the coils, the spacing between them, the core materials used, and the shielding, all have a pronounced effect on the degree of coupling attained in any design.

Probably the best way to design your own transformer is to use the old trial-and-error method, but do it in an orderly fashion and be consistent. It's a very sad day when one forgets how he got from point A to point B, especially if point B is an improvement in the design. Remember:

1. Decreasing the spacing between the primary and secondary increases the coupling.
2. Increasing the permeability of the magnetic path increases the coupling.
3. Shielding a transformer decreases its loaded Q and has the effect of increasing the coupling.

Begin the design by setting the loaded Q of each resonator to about twice what will be needed in the actual design. Then, slowly decrease the spacing between the primary and secondary until the response broadens to the loaded Q that is actually needed. If that response can't be met, try changing the geometry of the windings or the permeability of the magnetic path. Then, vary the spacing again. Use this as an iterative process to zero-in on the response that is needed. Granted, this is not an exact process, but it works and, if documented, can be reproduced.

There are literally thousands of commercially available transformers on the market that just might suit your needs perfectly. So before the trial-and-error method is put into practice, try a little research—it just might save a lot of time and money.

Active Coupling

It is possible to achieve very narrow 3-dB bandwidths in cascaded resonant circuits through the use of active coupling. Active coupling, for this purpose, is defined as a transistor, at least theoretically, which allows signal flow in only one direction (Fig. 2-26). If each of the tuned circuits is the same and if each has the same loaded Q, the total loaded Q of the cascaded circuit is approximately equal to

$$Q_{\text{total}} = \frac{Q}{\sqrt{2^{1/n} - 1}} \qquad \text{(Eq. 2-21)}$$

where

Q_{total} = the total Q of the cascaded circuit,

Q = the Q of each individual resonant circuit,

n = the number of resonant circuits.

The first step in any design procedure must be to relate the required Q_{total} of the network back to the individual loaded Q of each resonator. This is done by rearranging Equation 2-21 to solve for Q. As an example, with $n = 4$ resonators, and given that Q_{total} of the cascaded circuit must be 50, Equation 2-21 tells us that the Q of the individual resonator need only be about 22—a fairly simple and realizable design task.

Active coupling is obviously more expensive than passive coupling due to the added cost of each active device. But, in some

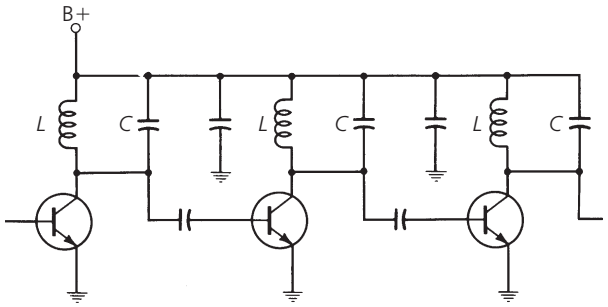


FIG. 2-26. Active coupling.

applications, there is no real trade-off involved because passive coupling just might not yield the required loaded Q. Example 2-5 illustrates some of the factors you must deal with.

SUMMARY

This chapter was meant to provide an insight into how resonant circuits actually perform their function as well as to provide you with the capability for designing one to operate at a certain value of loaded Q. In Chapter 3, we will carry our study one step further to include low-pass, high-pass, and bandpass filters of various shapes and sizes.

EXAMPLE 2-5

Design a top-L coupled two-resonator tuned circuit to meet the following requirements:

1. Center frequency = 75 MHz
2. 3-dB bandwidth = 3.75 MHz
3. Source resistance = 100 ohms
4. Load resistance = 1000 ohms

Assume that inductors are available that have an unloaded Q of 85 at the frequency of interest. Finally, use a tapped-C transformer to present an effective source resistance (R_s') of 1000 ohms to the filter.

Solution

The solution to this design problem is not a very difficult one, but it does involve quite a few separate and distinct calculations which might tend to make you lose sight of our goal. For this reason, we will walk through the solution in a very orderly fashion with a complete explanation of each calculation.

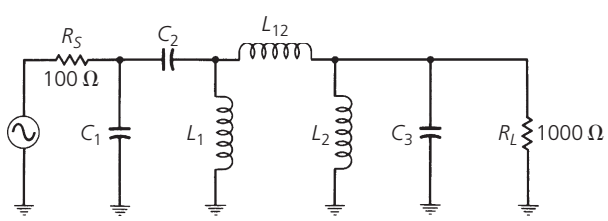


FIG. 2-27. Circuit for Example 2-5.

The circuit we are designing is shown in Fig. 2-27. Let's begin with a few definitions.

Q_{total} = the loaded Q of the entire circuit

Q_p = the Q of the inductor

Q_R = the loaded Q of each resonator

From our discussion on coupling and its effects on bandwidth, we know that

$$Q_R = \frac{Q_{total}}{0.707}$$

and,

$$Q_{total} = \frac{f_e}{B} = \frac{75\text{ MHz}}{3.75\text{ MHz}} = 20$$

so,

$$Q_R = \frac{20}{0.707} = 28.3$$

Thus, to provide a total loaded Q of 20, it is necessary that the loaded Q of each resonator be 28.3. For the inductor,

$$Q_p = \frac{R_p}{X_p} = 85$$

or

$$R_p = 85\,X_p \qquad \text{(Eq. 2-22)}$$

The loaded Q of each resonant circuit is

$$Q_R = \frac{R_{total}}{X_p} \qquad \text{(Eq. 2-23)}$$

where,

R_{total} = the total equivalent shunt resistance for each resonator and
$= R_s' \| R_p$
$= R_L \| R_p$

since both circuits are identical. Remember, we have already taken into account the loading effect that each resonant circuit has on the other through the factor 0.707, which was used at the beginning of the example. Now, we have:

$$R_{total} = \frac{R_s' R_p}{R_s' + R_p}$$

Continued on next page

EXAMPLE 2-5—Cont

Substituting into Equation 2-23:

$$Q_R = \frac{R_s' R_p}{(R_s' + R_p)X_p}$$

and,

$$X_p = \frac{R_s' R_p}{(R_s' + R_p)Q_R}$$

$$= \frac{1000R_p}{(1000 + R_p)28.3} \qquad \text{(Eq. 2-24)}$$

We can now substitute Equation 2-22 into Equation 2-24 and solve for X_p.

$$X_p = \frac{(1000)(85X_p)}{(1000 + 85X_p)28.3}$$

$$= 23.57 \text{ ohms}$$

and,

$$R_p = 85X_p$$

$$= 2003 \text{ ohms}$$

To find the component values

$$L_1 = L_2 = \frac{X_p}{\omega}$$

$$= 50 \text{ nH}$$

and,

$$C_s = \frac{1}{\omega X_p}$$

$$90 \text{ pF}$$

Now all that remains is to design the tapped-C transformer and the coupling inductor. From Equation 2-12:

$$R_s' = R_s\left(1 + \frac{C_1}{C_2}\right)^2$$

or,

$$\frac{C_1}{C_2} = \sqrt{\frac{R_s'}{R_s}} - 1$$

$$= 2.16$$

and,

$$C_1 = 2.16C_2 \qquad \text{(Eq. 2-25)}$$

We know that the total capacitance that must be used to resonate with the inductor is 90 pF and

$$C_{\text{total}} = \frac{C_1 C_2}{C_1 + C_2} \qquad \text{(Eq. 2-26)}$$

Substituting Equation 2-25 into Equation 2-26 and taking C_{total} to be 90 pF yields:

$$90 \text{ pF} = \frac{2.16C_2^2}{3.16C_2}$$

and,

$$C_2 = 132 \text{ pF}$$

$$C_1 = 285 \text{ pF}$$

To solve for the coupling inductance from Equation 2-20:

$$L_{12} = Q_R L$$

$$= (28.3)(50 \text{ nH})$$

$$= 1.415\ \mu\text{H}$$

The design is now complete. Notice that the tapped-C transformer is actually serving a dual purpose. It provides a DC block between the source and load in addition to its transformation properties.

CHAPTER 3

FILTER Design

Filters occur so frequently in the instrumentation and communications industries that no book covering the field of RF circuit design could be complete without at least one chapter devoted to the subject. Indeed, entire books have been written on the art of filter design alone, so this single chapter cannot possibly cover all aspects of all types of filters. But it will familiarize you with the characteristics of four of the most commonly used filters and will enable you to design very quickly and easily a filter that will meet, or exceed, most of the common filter requirements that you will encounter.

We will cover Butterworth, Chebyshev, and Bessel filters in all of their common configurations: *low-pass, high-pass, bandpass,* and *bandstop*. We will learn how to take advantage of the attenuation characteristics unique to each type of filter. Finally, we will learn how to design some very powerful filters in as little as 5 minutes by merely looking through a catalog to choose a design to suit your needs.

BACKGROUND

In Chapter 2, the concept of resonance was explored and we determined the effects that component value changes had on resonant circuit operation. You should now be somewhat familiar with the methods that are used in analyzing passive resonant circuits to find quantities, such as loaded Q, insertion loss, and bandwidth. You should also be capable of designing one- or two-resonator circuits for any loaded Q desired (or, at least, determine why you cannot). Quite a few of the filter applications that you will encounter, however, cannot be satisfied with the simple bandpass arrangement given in Chapter 2. There are occasions when, instead of passing a certain band of frequencies while rejecting frequencies above and below (bandpass), we would like to attenuate a small band of frequencies while passing all others. This type of filter is called, appropriately enough, a *bandstop filter*. Still other requirements call for a low-pass or high-pass response. The characteristic curves for these responses are shown in Fig. 3-1. The low-pass filter will allow all signals below a certain cutoff frequency to pass while attenuating all others. A high-pass filter's response is the mirror-image of the low-pass response and attenuates all signals below a certain cutoff frequency while allowing those above cutoff to pass. These

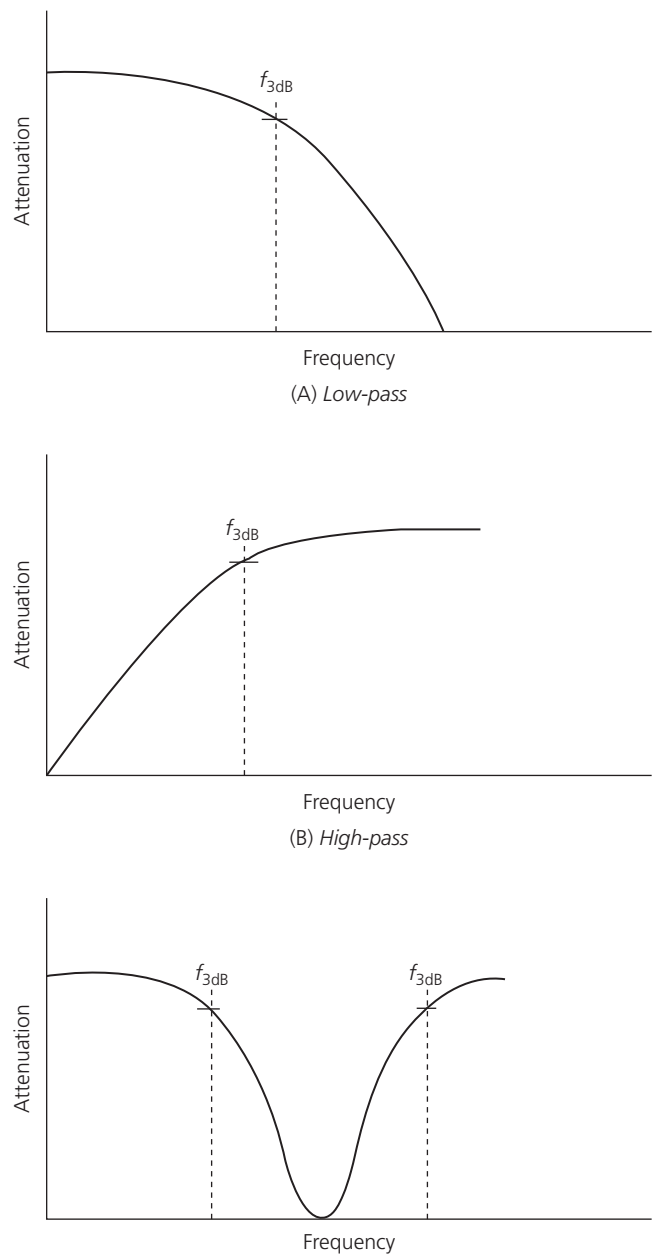


FIG. 3-1. Typical filter response curves.

types of response simply cannot be handled very well with the two-resonator bandpass designs of Chapter 2.

In this chapter, we will use the low-pass filter as our workhorse, as all other responses will be derived from it. So let's take a quick look at a simple low-pass filter and examine its characteristics. Fig. 3-2 is an example of a very simple *two-pole*, or *second-order* low-pass filter. The order of a filter is determined by the slope of the attenuation curve it presents in the stopband. A second-order filter is one whose rolloff is a function of the frequency squared, or 12 dB per octave. A third-order filter causes a rolloff that is proportional to frequency cubed, or 18 dB per octave. Thus, the order of a filter can be equated with the number of significant reactive elements that it presents to the source as the signal deviates from the passband.

The circuit of Fig. 3-2 can be analyzed in much the same manner as was done in Chapter 2. For instance, an examination of the effects of loaded Q on the response would yield the family of curves shown in Fig. 3-3. Surprisingly, even this circuit configuration can cause a peak in the response. This is due to the fact that at some frequency, the inductor and capacitor will become resonant and, thus, peak the response if the loaded Q is high enough. The resonant frequency can be determined from

$$F_r = \frac{1}{2\pi\sqrt{LC}} \qquad \text{(Eq. 3-1)}$$

For low values of loaded Q, however, no response peak will be noticed.

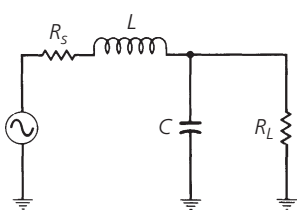


FIG. 3-2. A simple low-pass filter.

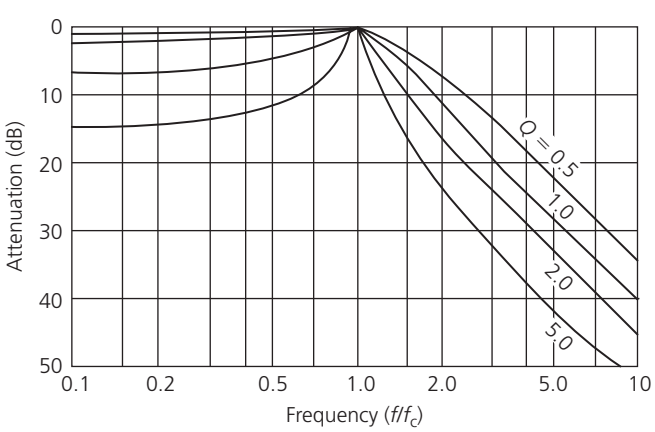


FIG. 3-3. Typical two-pole filter response curves.

The loaded Q of this filter is dependent upon the individual Q's of the series leg and the shunt leg where, assuming perfect components,

$$Q_1 = \frac{X_L}{R_s} \qquad \text{(Eq. 3-2)}$$

and,

$$Q_2 = \frac{R_L}{X_c} \qquad \text{(Eq. 3-3)}$$

and the total Q is:

$$Q_{\text{total}} = \frac{Q_1 Q_2}{Q_1 + Q_2} \qquad \text{(Eq. 3-4)}$$

If the total Q of the circuit is greater than about 0.5, then for optimum transfer of power from the source to the load, Q_1 should equal Q_2. In this case, at the peak frequency, the response will approach 0-dB insertion loss. If the total Q of the network is less than about 0.5, there will be no peak in the response and, for optimum transfer of power, R_s should equal R_L. The peaking of the filter's response is commonly called ripple (defined in Chapter 2) and can vary considerably from one filter design to the next depending on the application. As shown, the two-element filter exhibits only one response peak at the edge of the passband.

It can be shown that the number of peaks within the passband is directly related to the number of elements in the filter by:

$$\text{Number of Peaks} = N - 1$$

where

N = the number of elements.

Thus, the three-element low-pass filter of Fig. 3-4 should exhibit two response peaks as shown in Fig. 3-5. This is true only if the

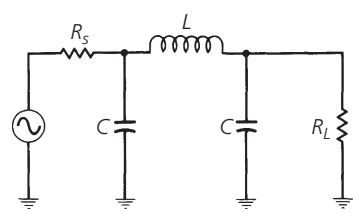


FIG. 3-4. Three-element low-pass filter.

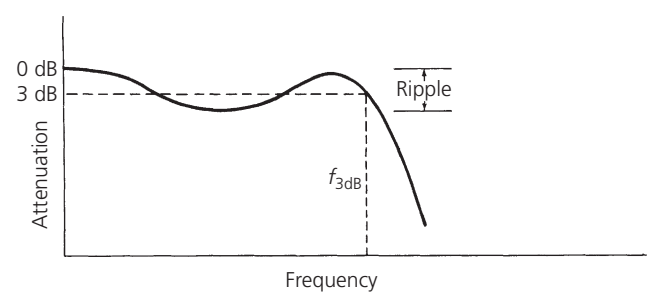


FIG. 3-5. Typical response of a three-element low-pass filter.

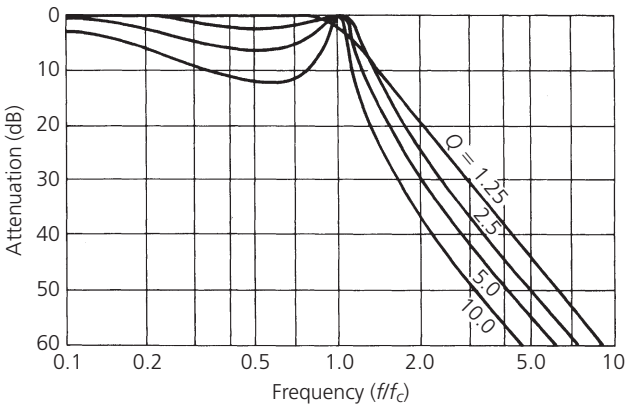


FIG. 3-6. Curves showing frequency response vs. loaded *Q* for three-element low-pass filters.

loaded *Q* is greater than one. Typical response curves for various values of loaded *Q* for the circuit given in Fig. 3-4 are shown in Fig. 3-6. For all odd-order networks, the response at DC and at the upper edge of the passband approaches 0 dB with dips in the response between the two frequencies. All even-order networks will produce an insertion loss at DC equal to the amount of passband ripple in dB. Keep in mind, however, that either of these two networks, if designed for low values of loaded *Q*, can be made to exhibit little or no passband ripple. But, as you can see from Figs. 3-3 and 3-6, the elimination of passband ripple can be made only at the expense of bandwidth. The smaller the ripple that is allowed, the wider the bandwidth becomes and, therefore, selectivity suffers. Optimum flatness in the passband occurs when the loaded *Q* of the three-element circuit is equal to one (1). Any value of loaded *Q* that is less than one will cause the response to roll off noticeably even at very low frequencies, within the defined passband. Thus, not only is the selectivity poorer but the passband insertion loss is too. In an application where there is not much signal to begin with, an even further decrease in signal strength could be disastrous.

Now that we have taken a quick look at two representative low-pass filters and their associated responses, let's discuss filters in general:

1. High-*Q* filters tend to exhibit a far greater initial slope toward the stopband than their low-*Q* counterparts with the same number of elements. Thus, at any frequency in the stopband, the attenuation will be greater for a high-*Q* filter than for one with a lower *Q*. The penalty for this improvement is the increase in passband ripple that must occur as a result.
2. Low-*Q* filters tend to have the flattest passband response but their initial attenuation slope at the band edge is small. Thus, the penalty for the reduced passband ripple is a decrease in the *initial* stopband attenuation.
3. As with the resonant circuits discussed in Chapter 2, the source and load resistors loading a filter will have a profound effect on the *Q* of the filter and, therefore, on the passband ripple and shape factor of the filter. If a filter is inserted between two resistance values for which it was not designed, the performance will suffer to an extent, depending upon the degree of error in the terminating impedance values.
4. The *final* attenuation slope of the response is dependent upon *the order of the network*. The order of the network is equal to the number of reactive elements *in the low-pass filter*. Thus, a second-order network (2 elements) falls off at a final attenuation slope of 12 dB per octave, a third-order network (3 elements) at the rate of 18 dB per octave, and so on, with the addition of 6 dB per octave per element.

MODERN FILTER DESIGN

Modern filter design has evolved through the years from a subject known only to specialists in the field (because of the advanced mathematics involved) to a practical well-organized catalog of ready-to-use circuits available to anyone with a knowledge of eighth grade level math. In fact, an average individual with absolutely no prior practical filter design experience should be able to sit down, read this chapter, and within 30 minutes be able to design a practical high-pass, low-pass, bandpass, or bandstop filter to his specifications. It sounds simple and it is—once a few basic rules are memorized.

The approach we will take in all of the designs in this chapter will be to make use of the myriad of normalized *low-pass prototypes* that are now available to the designer. The actual design procedure is, therefore, nothing more than determining your requirements and then finding a filter in a catalog that satisfies these requirements. Each normalized element value is then scaled to the frequency and impedance you desire, and then transformed to the type of response (bandpass, high-pass, bandstop) that you wish. With practice, the procedure becomes very simple and soon you will be defining and designing filters.

The concept of normalization may at first seem foreign to the person who is a newcomer to the field of filter design, and the idea of transforming a low-pass filter into one that will give one of the other three types of responses might seem absurd. The best advice I can give (to anyone not familiar with these practices and who might feel a bit skeptical at this point) is to press on. The only way to truly realize the beauty and simplicity of this approach is to try a few actual designs. Once you try a few, you will be hooked, and any other approach to filter design will suddenly seem tedious and unnecessarily complicated.

NORMALIZATION AND THE LOW-PASS PROTOTYPE

In order to offer a catalog of useful filter circuits to the electronic filter designer, it became necessary to standardize the presentation of the material. Obviously, in practice, it would

be extremely difficult to compare the performance and evaluate the usefulness of two filter networks if they were operating under two totally different sets of circumstances. Similarly, the presentation of any comparative design information for filters, if not standardized, would be totally useless. This concept of standardization or *normalization*, then, is merely a tool used by filter experts to present all filter design and performance information in a manner useful to circuit designers. Normalization assures the designer of the capability of comparing the performance of any two filter types when given the same operating conditions.

All of the catalogued filters in this chapter are low-pass filters normalized for a cutoff frequency of one radian per second (0.159 Hz) and for source and load resistors of one ohm. A characteristic response of such a filter is shown in Fig. 3-7. The circuit used to generate this response is called the *low-pass prototype*.

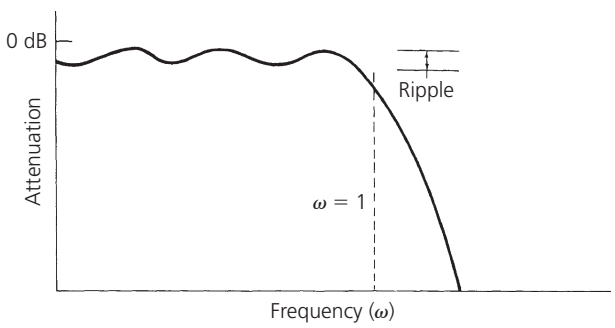


FIG. 3-7. Normalized low-pass response.

Obviously, the design of a filter with such a low cutoff frequency would require component values much larger than those we are accustomed to working with; capacitor values would be in farads rather than microfarads and picofarads, and the inductor values would be in henries rather than in microhenries and nanohenries. But once we choose a suitable low-pass prototype from the catalog, we can change the impedance level and cutoff frequency of the filter to any value we wish through a simple process called *scaling*. The net result of this process is a practical filter design with realizable component values.

FILTER TYPES

Many of the filters used today bear the names of the men who developed them. In this section, we will take a look at three such filters and examine their attenuation characteristics. Their relative merits will be discussed and their low-pass prototypes presented. The three filter types discussed will be the Butterworth, Chebyshev, and Bessel responses.

The Butterworth Response

The Butterworth filter is a medium-Q filter that is used in designs that require the amplitude response of the filter to be as flat as possible. The Butterworth response is the flattest passband response available and contains no ripple. The typical response of such a filter might look like that of Fig. 3-8.

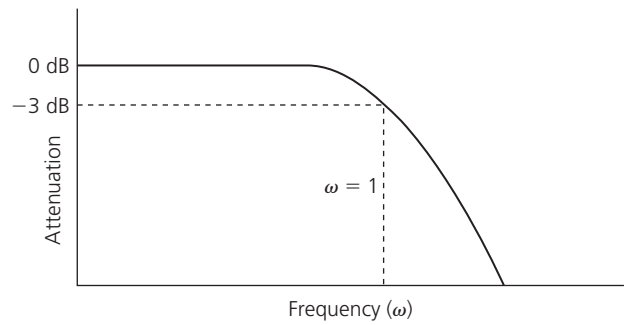


FIG. 3-8. The Butterworth response.

Since the Butterworth response is only a medium-Q filter, its initial attenuation steepness is not as good as some filters but it is better than others. This characteristic often causes the Butterworth response to be called a middle-of-the-road design.

The attenuation of a Butterworth filter is given by

$$A_{dB} = 10\log\left[1 + \left(\frac{\omega}{\omega_c}\right)^{2n}\right] \qquad \text{(Eq. 3-5)}$$

where

ω = the frequency at which the attenuation is desired,
ω_c = the cutoff frequency ($\omega_{3\text{dB}}$) of the filter,
n = the number of elements in the filter.

If Equation 3-5 is evaluated at various frequencies for various numbers of elements, a family of curves is generated which will give a very good graphical representation of the attenuation provided by any order of filter at any frequency. This information is illustrated in Fig. 3-9. Thus, from Fig. 3-9, a 5-element (fifth order) Butterworth filter will provide an attenuation of approximately 30 dB at a frequency equal to twice the cutoff frequency

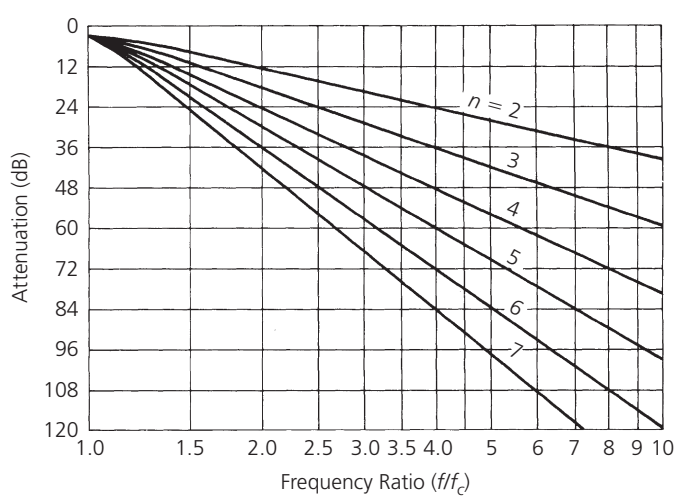


FIG. 3-9. Attenuation characteristics for Butterworth filters.

of the filter. Notice here that the frequency axis is normalized to ω/ω_c and the graph begins at the cutoff (-3 dB) point. This graph is extremely useful as it provides you with a method of determining, at a glance, the order of a filter needed to meet a given attenuation specification. A brief example should illustrate this point (Example 3-1).

EXAMPLE 3-1

How many elements are required to design a Butterworth filter with a cutoff frequency of 50 MHz, if the filter must provide at least 50 dB of attenuation at 150 MHz?

Solution

The first step in the solution is to find the ratio of $\omega/\omega_c = f/f_c$.

$$\frac{f}{f_c} = \frac{150\,\text{MHz}}{50\,\text{MHz}} = 3$$

Thus, at three times the cutoff frequency, the response must be down by at least 50 dB. Referring to Fig. 3-9, it is seen very quickly that a minimum of 6 elements is required to meet this design goal. At an f/f_c of 3, a 6-element design would provide approximately 57 dB of attenuation, while a 5-element design would provide only about 47 dB, which is not quite good enough.

The element values for a normalized Butterworth low-pass filter operating between equal 1-ohm terminations (source and load) can be found by

$$A_k = 2\sin\frac{(2k-1)\pi}{2n}, \quad k = 1, 2, \ldots n \qquad \text{(Eq. 3-6)}$$

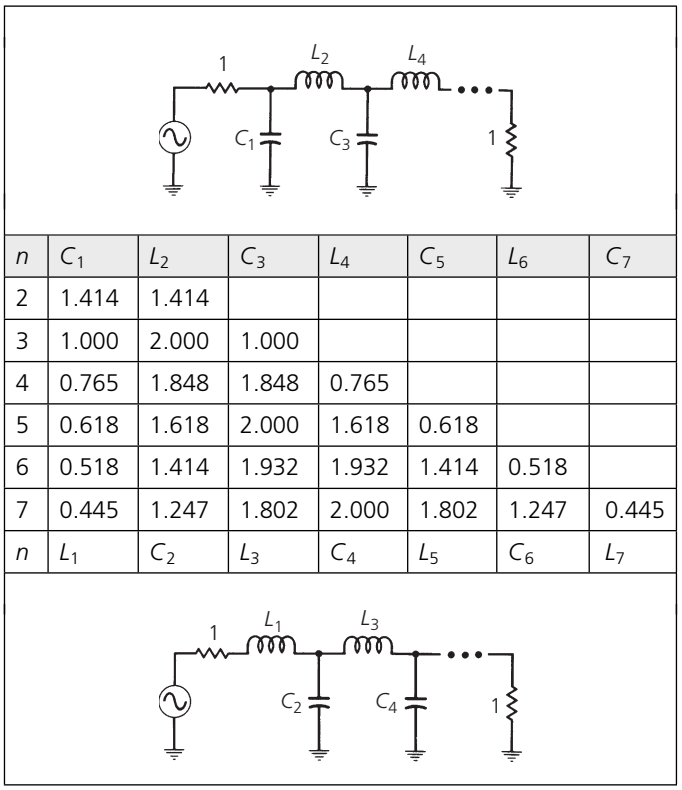
where

n is the number of elements,

A_k is the kth reactance in the ladder and may be either an inductor or capacitor.

The term $(2k-1)\pi/2n$ is in radians. We can use Equation 3-6 to generate our first entry into the catalog of low-pass prototypes shown in Table 3-1. The placement of each component of the filter is shown immediately above and below the table.

The rules for interpreting Butterworth tables are simple. The schematic shown above the table is used whenever the ratio R_s/R_L is calculated as the design criteria. The table is read from the top down. Alternately, when R_L/R_s is calculated, the schematic below the table is used. Then, the element designators in the table are read from the bottom up. Thus, a four-element low-pass prototype could appear as shown in Fig. 3-10. Note here that the element values not given in Table 3-1 are simply left out



n	C_1	L_2	C_3	L_4	C_5	L_6	C_7
2	1.414	1.414					
3	1.000	2.000	1.000				
4	0.765	1.848	1.848	0.765			
5	0.618	1.618	2.000	1.618	0.618		
6	0.518	1.414	1.932	1.932	1.414	0.518	
7	0.445	1.247	1.802	2.000	1.802	1.247	0.445
n	L_1	C_2	L_3	C_4	L_5	C_6	L_7

TABLE 3-1. Butterworth Equal Termination Low-Pass Prototype Element Values ($R_s = R_L$)

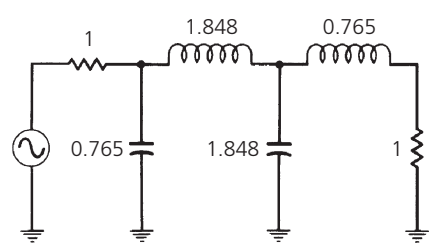


FIG. 3-10. A four-element Butterworth low-pass prototype circuit.

of the prototype ladder network. The 1-ohm load resistor is then placed directly across the output of the filter.

Remember that the cutoff frequency of each filter is 1 radian per second, or 0.159 Hz. Each capacitor value given is in farads, and each inductor value is in henries. The network will later be scaled to the impedance and frequency that is desired through a simple multiplication and division process. The component values will then appear much more realistic.

Occasionally, we have the need to design a filter that will operate between two unequal terminations as shown in Fig. 3-11. In this case, the circuit is normalized for a load resistance of 1 ohm, while taking what we get for the source resistance. Dividing both the load and source resistor by 10 will yield a load resistance of 1 ohm and a source resistance of 5 ohms as shown in Fig. 3-12.

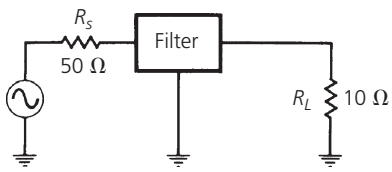


FIG. 3-11. Unequal terminations.

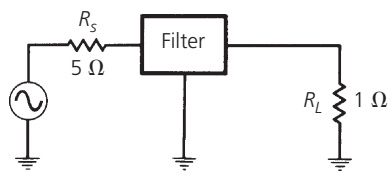
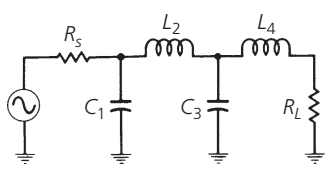


FIG. 3-12. Normalized unequal terminations.

We can use the normalized terminating resistors to help us find a low-pass prototype circuit.

Table 3-2 is a list of Butterworth low-pass prototype values for various ratios of source to load impedance (R_s/R_L). The schematic shown above the table is used when R_s/R_L is calculated, and the element values are read down from the top of the table. Alternately, when R_L/R_s is calculated, the schematic below the table is used while reading up from the bottom of the table to get the element values (Example 3-2).



n	R_s/R_L	C_1	L_2	C_3	L_4
2	1.111	1.035	1.835		
	1.250	0.849	2.121		
	1.429	0.697	2.439		
	1.667	0.566	2.828		
	2.000	0.448	3.346		
	2.500	0.342	4.095		
	3.333	0.245	5.313		
	5.000	0.156	7.707		
	10.000	0.074	14.814		
	∞	1.414	0.707		
3	0.900	0.808	1.633	1.599	
	0.800	0.844	1.384	1.926	
	0.700	0.915	1.165	2.277	
	0.600	1.023	0.965	2.702	
	0.500	1.181	0.779	3.261	
	0.400	1.425	0.604	4.064	
	0.300	1.838	0.440	5.363	
	0.200	2.669	0.284	7.910	
	0.100	5.167	0.138	15.455	
	∞	1.500	1.333	0.500	
4	1.111	0.466	1.592	1.744	1.469
	1.250	0.388	1.695	1.511	1.811
	1.429	0.325	1.862	1.291	2.175
	1.667	0.269	2.103	1.082	2.613
	2.000	0.218	2.452	0.883	3.187
	2.500	0.169	2.986	0.691	4.009
	3.333	0.124	3.883	0.507	5.338
	5.000	0.080	5.684	0.331	7.940
	10.000	0.039	11.094	0.162	15.642
	∞	1.531	1.577	1.082	0.383
n	R_L/R_s	L_1	C_2	L_3	C_4

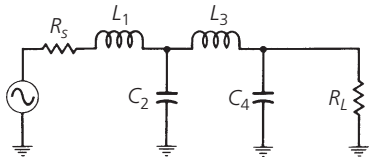


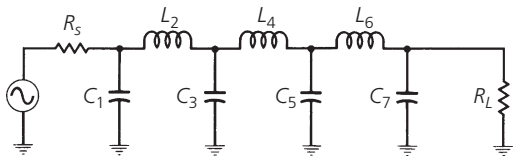
TABLE 3-2A. Butterworth Low-Pass Prototype Element Values

TABLE 3-2A. (Continued)

Obviously, all possible ratios of source to load resistance could not possibly fit on a chart of this size. This, of course, leaves the potential problem of not being able to find the ratio that you need for a particular design task. The solution to this dilemma is to simply choose a ratio that most closely matches the ratio you need to complete the design. For ratios of 100:1 or so, the best results are obtained if you assume this value to be so high for practical purposes as to be infinite. Since, in these instances, you are only approximating the ratio of source to load resistance, the filter derived will only approximate the response that was originally intended. This is usually not too much of a problem.

The Chebyshev Response

The Chebyshev filter is a high-Q filter that is used when: (1) a steeper initial descent into the stopband is required, and (2) the passband response is no longer required to be flat. With this type of requirement, ripple can be allowed in the passband. As more ripple is introduced, the initial slope at the beginning of the stopband is increased and produces a more rectangular attenuation



n	R_s/R_L	C_1	L_2	C_3	L_4	C_5	L_6	C_7
5	0.900	0.442	1.027	1.910	1.756	1.389		
	0.800	0.470	0.866	2.061	1.544	1.738		
	0.700	0.517	0.731	2.285	1.333	2.108		
	0.600	0.586	0.609	2.600	1.126	2.552		
	0.500	0.686	0.496	3.051	0.924	3.133		
	0.400	0.838	0.388	3.736	0.727	3.965		
	0.300	1.094	0.285	4.884	0.537	5.307		
	0.200	1.608	0.186	7.185	0.352	7.935		
	0.100	3.512	0.091	14.095	0.173	15.710		
	∞	1.545	1.694	1.382	0.894	0.309		
6	1.111	0.289	1.040	1.322	2.054	1.744	1.335	
	1.250	0.245	1.116	1.126	2.239	1.550	1.688	
	1.429	0.207	1.236	0.957	2.499	1.346	2.062	
	1.667	0.173	1.407	0.801	2.858	1.143	2.509	
	2.000	0.141	1.653	0.654	3.369	0.942	3.094	
	2.500	0.111	2.028	0.514	4.141	0.745	3.931	
	3.333	0.082	2.656	0.379	5.433	0.552	5.280	
	5.000	0.054	3.917	0.248	8.020	0.363	7.922	
	10.000	0.026	7.705	0.122	15.786	0.179	15.738	
	∞	1.553	1.759	1.553	1.202	0.758	0.259	
7	0.900	0.299	0.711	1.404	1.489	2.125	1.727	1.296
	0.800	0.322	0.606	1.517	1.278	2.334	1.546	1.652
	0.700	0.357	0.515	1.688	1.091	2.618	1.350	2.028
	0.600	0.408	0.432	1.928	0.917	3.005	1.150	2.477
	0.500	0.480	0.354	2.273	0.751	3.553	0.951	3.064
	0.400	0.590	0.278	2.795	0.592	4.380	0.754	3.904
	0.300	0.775	0.206	3.671	0.437	5.761	0.560	5.258
	0.200	1.145	0.135	5.427	0.287	8.526	0.369	7.908
	0.100	2.257	0.067	10.700	0.142	16.822	0.182	15.748
	∞	1.558	1.799	1.659	1.397	1.055	0.656	0.223
n	R_L/R_s	L_1	C_2	L_3	C_4	L_5	C_6	L_7

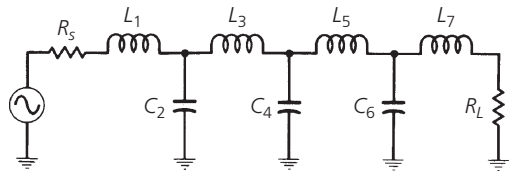


TABLE 3-2B. Butterworth Low-Pass Prototype Element Values

EXAMPLE 3-2

Find the low-pass prototype value for an $n = 4$ Butterworth filter with unequal terminations: $R_S = 50$ ohms, $R_L = 100$ ohms.

Solution

Normalizing the two terminations for $R_L = 1$ ohm will yield a value of $R_s = 0.5$. Reading down from the top of Table 3-2, for an $n = 4$ low-pass prototype value, we see that there is no $R_s/R_L = 0.5$ ratio listed. Our second choice, then, is to take the value of $R_L/R_s = 2$, and read up from the bottom of the table while using the schematic below the table as the form for the low-pass prototype values. This approach results in the low-pass prototype circuit of Fig. 3-13.

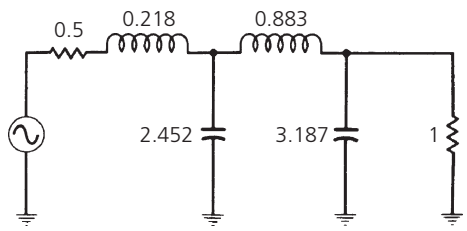


FIG. 3-13. Low-pass prototype circuit for Example 3-2.

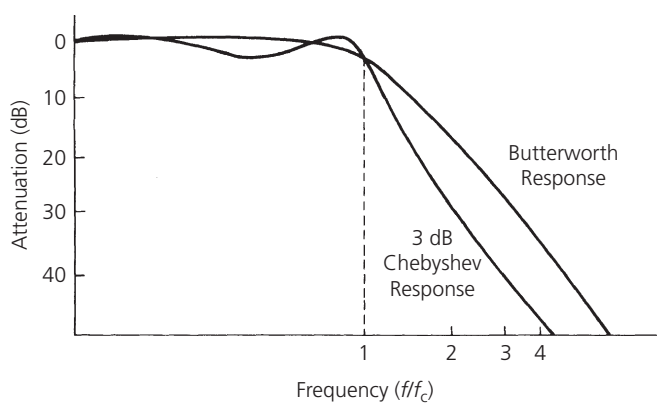


FIG. 3-14. Comparison of three-element Chebyshev and Butterworth responses.

curve when compared to the rounded Butterworth response. This comparison is made in Fig. 3-14. Both curves are for $n = 3$ filters. The Chebyshev response shown has 3 dB of passband ripple and produces a 10-dB improvement in stopband attenuation over the Butterworth filter.

n	Chebyshev Polynomial
1	$\left(\frac{\omega}{\omega_c}\right)$
2	$2\left(\frac{\omega}{\omega_c}\right)^2 - 1$
3	$4\left(\frac{\omega}{\omega_c}\right)^3 - 3\left(\frac{\omega}{\omega_c}\right)$
4	$8\left(\frac{\omega}{\omega_c}\right)^4 - 8\left(\frac{\omega}{\omega_c}\right)^2 + 1$
5	$16\left(\frac{\omega}{\omega_c}\right)^5 - 20\left(\frac{\omega}{\omega_c}\right)^3 + 5\left(\frac{\omega}{\omega_c}\right)$
6	$32\left(\frac{\omega}{\omega_c}\right)^6 - 48\left(\frac{\omega}{\omega_c}\right)^4 + 18\left(\frac{\omega}{\omega_c}\right)^2 - 1$
7	$64\left(\frac{\omega}{\omega_c}\right)^7 - 112\left(\frac{\omega}{\omega_c}\right)^5 + 58\left(\frac{\omega}{\omega_c}\right)^3 - 7\left(\frac{\omega}{\omega_c}\right)$

TABLE 3-3. Chebyshev Polynomials to the Order n

The attenuation of a Chebyshev filter can be found by making a few simple but tiresome calculations, and can be expressed as:

$$A_{\text{dB}} = 10 \log \left[1 + \varepsilon^2 C_n^2 \left(\frac{\omega}{\omega_c} \right)' \right] \qquad \text{(Eq. 3-7)}$$

where

$C_n^2 \left(\frac{\omega}{\omega_c} \right)'$ is the Chebyshev polynomial to the order n evaluated at $\left(\frac{\omega}{\omega_c} \right)'$.

The Chebyshev polynomials for the first seven orders are given in Table 3-3. The parameter ε is given by:

$$\varepsilon = \sqrt{10^{R_{dB}/10} - 1} \qquad \text{(Eq. 3-8)}$$

where

R_{dB} is the passband ripple in decibels.

Note that $\left(\frac{\omega}{\omega_c} \right)'$ is not the same as $\left(\frac{\omega}{\omega_c} \right)$. The quantity $\left(\frac{\omega}{\omega_c} \right)'$ can be found by defining another parameter:

$$B = \frac{1}{n} \cosh^{-1} \left(\frac{1}{\varepsilon} \right) \qquad \text{(Eq. 3-9)}$$

where

n = the order of the filter,
ε = the parameter defined in Equation 3-8,
$\cosh^{-1}$ = the inverse hyperbolic cosine of the quantity in parentheses.

Finally, we have:

$$\left(\frac{\omega}{\omega_c} \right)' = \left(\frac{\omega}{\omega_c} \right) \cosh B \qquad \text{(Eq. 3-10)}$$

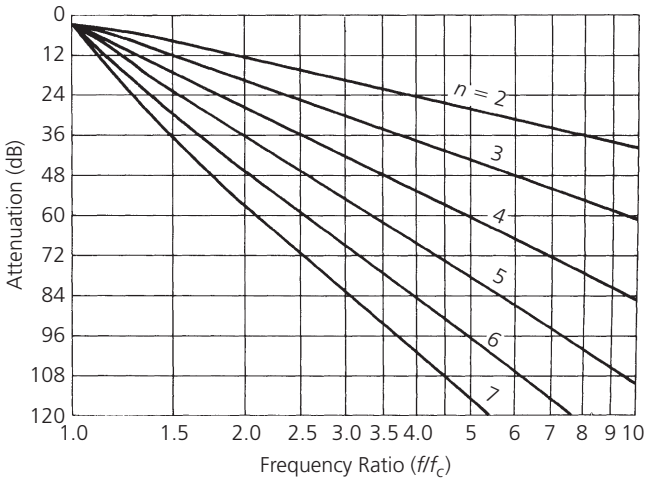


FIG. 3-15. Attenuation characteristics for a Chebyshev filter with 0.01-dB ripple

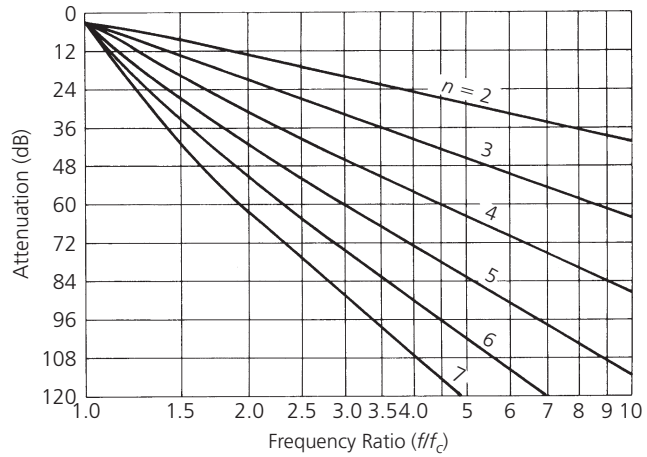


FIG. 3-16. Attenuation characteristics for a Chebyshev filter with 0.1-dB ripple.

where

$\left(\dfrac{\omega}{\omega_c}\right)$ = the ratio of the frequency of interest to the cutoff frequency,

cosh = the hyperbolic cosine.

If your calculator does not have hyperbolic and inverse hyperbolic functions, they can be manually determined from the following relations:

$$\cosh x = 0.5(e^x + e^{-x})$$

and

$$\cosh^{-1} x = \ln (x \pm \sqrt{x^2 - 1})$$

The preceding equations yield families of attenuation curves, each classified according to the amount of ripple allowed in the passband. Several of these families of curves are shown in Figs. 3-15 through 3-18, and include 0.01-dB, 0.1-dB, 0.5-dB, and 1.0-dB ripple. Each curve begins at $\omega/\omega_c = 1$, which is the normalized cutoff, or 3-dB frequency. The passband ripple is, therefore, not shown.

If other families of attenuation curves are needed with different values of passband ripple, the preceding Chebyshev equations can be used to derive them. The problem in Example 3-3 illustrates this.

Obviously, performing the calculations of Example 3-3 for various values of ω/ω_c, ripple, and filter order is a very time-consuming chore unless a programmable calculator or computer is available to do most of the work for you.

The low-pass prototype element values corresponding to the Chebyshev responses of Figs. 3-15 through 3-18 are given in Tables 3-4 through 3-7. Note that the Chebyshev prototype values could not be separated into two distinct sets of tables covering the equal and unequal termination cases, as was done

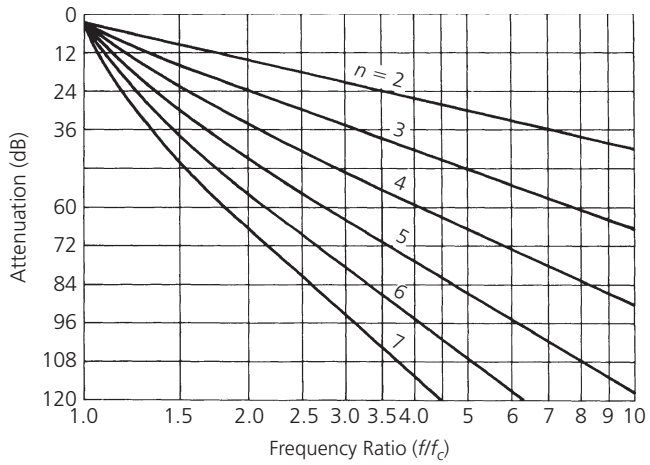


FIG. 3-17. Attenuation characteristics for a Chebyshev filter with 0.5-dB ripple.

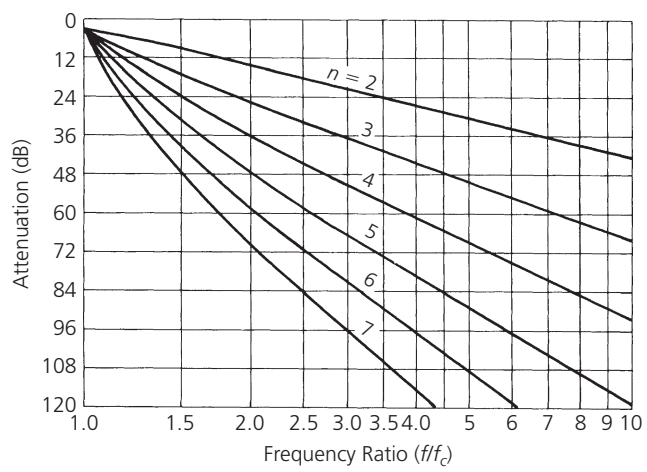


FIG. 3-18. Attenuation characteristics for a Chebyshev filter with 1-dB ripple.

EXAMPLE 3-3

Find the attenuation of a 4-element, 2.5-dB ripple, low-pass Chebyshev filter at $\omega/\omega_c = 2.5$.

Solution

First evaluate the parameter:

$$\varepsilon = \sqrt{10^{2.5/10} - 1}$$
$$= 0.882$$

Next, find B.

$$B = \frac{1}{4}\left[\cosh^{-1}\left(\frac{1}{0.882}\right)\right]$$
$$= 0.1279$$

Then, $(\omega/\omega_c)'$ is:

$$\left(\frac{\omega}{\omega_c}\right)' = 2.5\cosh.1279$$
$$= 2.5204$$

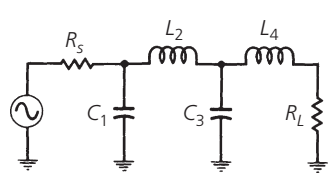
Finally, we evaluate the fourth order ($n = 4$) Chebyshev polynomial at $(\omega/\omega_c)' = 2.52$.

$$C_n^2\left(\frac{\omega}{\omega_c}\right)' = 8\left(\frac{\omega}{\omega_c}\right)^4 - 8\left(\frac{\omega}{\omega_c}\right)^2 + 1$$
$$= 8(2.5204)^4 - 8(2.5204)^2 + 1$$
$$= 273.05$$

We can now evaluate the final equation.

$$A_{dB} = 10\log_{10}\left[1 + \varepsilon^2 C_n^2\left(\frac{\omega}{\omega_c}\right)'\right]$$
$$= 10\log_{10}[1 + (0.882)^2(273.05)^2]$$
$$= 47.63\text{ dB}$$

Thus, at an ω/ω_c, of 2.5, you can expect 47.63 dB of attenuation for this filter.



n	R_S/R_L	C_1	L_2	C_3	L_4
2	1.101	1.347	1.483		
	1.111	1.247	1.595		
	1.250	0.943	1.997		
	1.429	0.759	2.344		
	1.667	0.609	2.750		
	2.000	0.479	3.277		
	2.500	0.363	4.033		
	3.333	0.259	5.255		
	5.000	0.164	7.650		
	10.000	0.078	14.749		
	∞	1.412	0.742		
3	1.000	1.181	1.821	1.181	
	0.900	1.092	1.660	1.480	
	0.800	1.097	1.443	1.806	
	0.700	1.160	1.228	2.165	
	0.600	1.274	1.024	2.598	
	0.500	1.452	0.829	3.164	
	0.400	1.734	0.645	3.974	
	0.300	2.216	0.470	5.280	
	0.200	3.193	0.305	7.834	
	0.100	6.141	0.148	15.390	
	∞	1.501	1.433	0.591	
4	1.100	0.950	1.938	1.761	1.046
	1.111	0.854	1.946	1.744	1.165
	1.250	0.618	2.075	1.542	1.617
	1.429	0.495	2.279	1.334	2.008
	1.667	0.398	2.571	1.128	2.461
	2.000	0.316	2.994	0.926	3.045
	2.500	0.242	3.641	0.729	3.875
	3.333	0.174	4.727	0.538	5.209
	5.000	0.112	6.910	0.352	7.813
	10.000	0.054	13.469	0.173	15.510
	∞	1.529	1.694	1.312	0.523
n	R_L/R_s	L_1	C_2	L_3	C_4

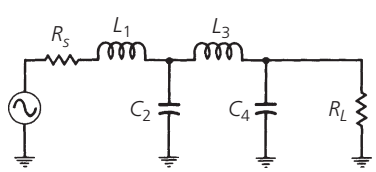
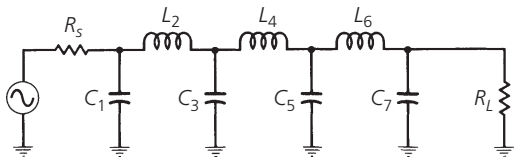


TABLE 3-4A. Chebyshev Low-Pass Element Values for 0.01-dB Ripple.

TABLE 3-4A. (Continued)

for the Butterworth prototypes. This is because the even order ($n = 2, 4, 6, \ldots$) Chebyshev filters cannot have equal terminations. The source and load must always be different for proper operation as shown in the tables.



n	R_s/R_L	C_1	L_2	C_3	L_4	C_5	L_6	C_7
5	1.000	0.977	1.685	2.037	1.685	0.977		
	0.900	0.880	1.456	2.174	1.641	1.274		
	0.800	0.877	1.235	2.379	1.499	1.607		
	0.700	0.926	1.040	2.658	1.323	1.977		
	0.600	1.019	0.863	3.041	1.135	2.424		
	0.500	1.166	0.699	3.584	0.942	3.009		
	0.400	1.398	0.544	4.403	0.749	3.845		
	0.300	1.797	0.398	5.772	0.557	5.193		
	0.200	2.604	0.259	8.514	0.368	7.826		
	0.100	5.041	0.127	16.741	0.182	15.613		
	∞	1.547	1.795	1.645	1.237	0.488		
6	1.101	0.851	1.796	1.841	2.027	1.631	0.937	
	1.111	0.760	1.782	1.775	2.094	1.638	1.053	
	1.250	0.545	1.864	1.489	2.403	1.507	1.504	
	1.429	0.436	2.038	1.266	2.735	1.332	1.899	
	1.667	0.351	2.298	1.061	3.167	1.145	2.357	
	2.000	0.279	2.678	0.867	3.768	0.954	2.948	
	2.500	0.214	3.261	0.682	4.667	0.761	3.790	
	3.333	0.155	4.245	0.503	6.163	0.568	5.143	
	5.000	0.100	6.223	0.330	9.151	0.376	7.785	
	10.000	0.048	12.171	0.162	18.105	0.187	15.595	
	∞	1.551	1.847	1.790	1.598	1.190	0.469	
7	1.000	0.913	1.595	2.002	1.870	2.002	1.595	0.913
	0.900	0.816	1.362	2.089	1.722	2.202	1.581	1.206
	0.800	0.811	1.150	2.262	1.525	2.465	1.464	1.538
	0.700	0.857	0.967	2.516	1.323	2.802	1.307	1.910
	0.600	0.943	0.803	2.872	1.124	3.250	1.131	2.359
	0.500	1.080	0.650	3.382	0.928	3.875	0.947	2.948
	0.400	1.297	0.507	4.156	0.735	4.812	0.758	3.790
	0.300	1.669	0.372	5.454	0.546	6.370	0.568	5.148
	0.200	2.242	0.242	8.057	0.360	9.484	0.378	7.802
	0.100	4.701	0.119	15.872	0.178	18.818	0.188	15.652
	∞	1.559	1.867	1.866	1.765	1.563	1.161	0.456
n	R_L/R_s	L_1	C_2	L_3	C_4	L_5	C_6	L_7

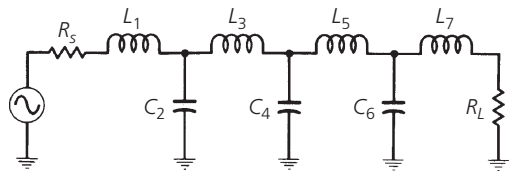
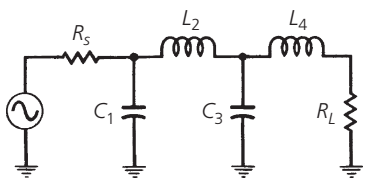


TABLE 3-4B. Chebyshev Low-Pass Element Values for 0.01-dB Ripple.



n	R_s/R_L	C_1	L_2	C_3	L_4
2	1.355	1.209	1.638		
	1.429	0.977	1.982		
	1.667	0.733	2.489		
	2.000	0.560	3.054		
	2.500	0.417	3.827		
	3.333	0.293	5.050		
	5.000	0.184	7.426		
	10.000	0.087	14.433		
	∞	1.391	0.819		
3	1.000	1.433	1.594	1.433	
	0.900	1.426	1.494	1.622	
	0.800	1.451	1.356	1.871	
	0.700	1.521	1.193	2.190	
	0.600	1.648	1.017	2.603	
	0.500	1.853	0.838	3.159	
	0.400	2.186	0.660	3.968	
	0.300	2.763	0.486	5.279	
	0.200	3.942	0.317	7.850	
	1.100	7.512	0.155	15.466	
	∞	1.513	1.510	0.716	
4	1.355	0.992	2.148	1.585	1.341
	1.429	0.779	2.348	1.429	1.700
	1.667	0.576	2.730	1.185	2.243
	2.000	0.440	3.227	0.967	2.856
	2.500	0.329	3.961	0.760	3.698
	3.333	0.233	5.178	0.560	5.030
	5.000	0.148	7.607	0.367	7.614
	10.000	0.070	14.887	0.180	15.230
	∞	1.511	1.768	1.455	0.673
n	R_L/R_s	L_1	C_2	L_3	C_4

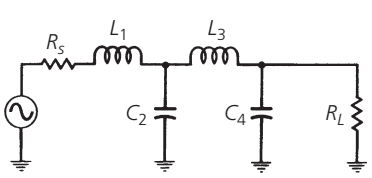


TABLE 3-5A. Chebyshev Low-Pass Prototype Element Values for 0.1-dB Ripple

The rules used for interpreting the Butterworth tables apply here also. The schematic shown above the table is used, and the element designators are read down from the top, when the ratio R_s/R_L is calculated as a design criteria. Alternately, with R_L/R_s calculations, use the schematic given below the table and read the element designators upwards from the bottom of the table. Example 3-4 is a practice problem for use in understanding the procedure.

EXAMPLE 3-4

Find the low-pass prototype values for an $n = 5$, 0.1-dB ripple, Chebyshev filter if the source resistance you are designing for is 50 ohms and the load resistance is 250 ohms.

Solution

Normalization of the source and load resistors yields an $R_s/R_L = 0.2$. A look at Table 3-5, for a 0.1-dB ripple filter with an $n = 5$ and an $R_s/R_L = 0.2$, yields the circuit values shown in Fig. 3-19.

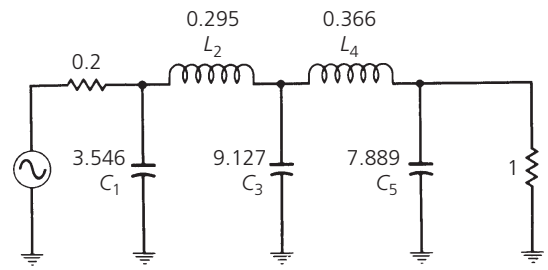


FIG. 3-19. Low-pass prototype circuit for Example 3-4.

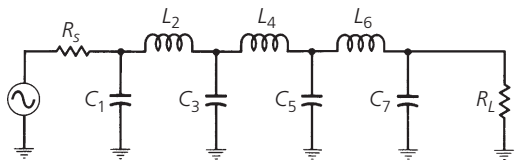
It should be mentioned here that equations could have been presented in this section for deriving the element values for the Chebyshev low-pass prototypes. The equations are extremely long and tedious, however, and there would be little to be gained from their presentation.

The Bessel Filter

The initial stopband attenuation of the Bessel filter is very poor and can be approximated by:

$$A_{\text{dB}} = 3\left(\frac{\omega}{\omega_c}\right)^2 \qquad \text{(Eq. 3-11)}$$

This expression, however, is not very accurate above an ω/ω_c that is equal to about 2. For values of ω/ω_c greater than 2, a straight-line approximation of 6 dB per octave per element can be made. This yields the family of curves shown in Fig. 3-20.



n	R_S/R_L	C_1	L_2	C_3	L_4	C_5	L_6	C_7
5	1.000	1.301	1.556	2.241	1.556	1.301		
	0.900	1.285	1.433	2.380	1.488	1.488		
	0.800	1.300	1.282	2.582	1.382	1.738		
	0.700	1.358	1.117	2.868	1.244	2.062		
	0.600	1.470	0.947	3.269	1.085	2.484		
	0.500	1.654	0.778	3.845	0.913	3.055		
	0.400	1.954	0.612	4.720	0.733	3.886		
	0.300	2.477	0.451	6.196	0.550	5.237		
	0.200	3.546	0.295	9.127	0.366	7.889		
	0.100	6.787	0.115	17.957	0.182	15.745		
	∞	1.561	1.807	1.766	1.417	0.651		
6	1.355	0.942	2.080	1.659	2.247	1.534	1.277	
	1.429	0.735	2.249	1.454	2.544	1.405	1.629	
	1.667	0.542	2.600	1.183	3.064	1.185	2.174	
	2.000	0.414	3.068	0.958	3.712	0.979	2.794	
	2.500	0.310	3.765	0.749	4.651	0.778	3.645	
	3.333	0.220	4.927	0.551	6.195	0.580	4.996	
	5.000	0.139	7.250	0.361	9.261	0.384	7.618	
	10.000	0.067	14.220	0.178	18.427	0.190	15.350	
	∞	1.534	1.884	1.831	1.749	1.394	0.638	
7	1.000	1.262	1.520	2.239	1.680	2.239	1.520	1.262
	0.900	1.242	1.395	2.361	1.578	2.397	1.459	1.447
	0.800	1.255	1.245	2.548	1.443	2.624	1.362	1.697
	0.700	1.310	1.083	2.819	1.283	2.942	1.233	2.021
	0.600	1.417	0.917	3.205	1.209	3.384	1.081	2.444
	0.500	1.595	0.753	3.764	0.928	4.015	0.914	3.018
	0.400	1.885	0.593	4.618	0.742	4.970	0.738	3.855
	0.300	2.392	0.437	6.054	0.556	6.569	0.557	5.217
	0.200	3.428	0.286	8.937	0.369	9.770	0.372	7.890
	0.100	6.570	0.141	17.603	0.184	19.376	0.186	15.813
	∞	1.575	1.858	1.921	1.827	1.734	1.379	0.631
n	R_L/R_S	L_1	C_2	L_3	C_4	L_5	C_6	L_7

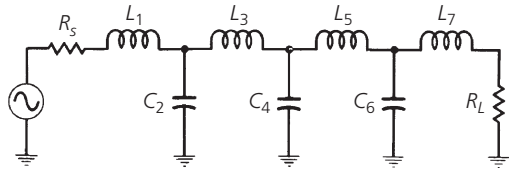
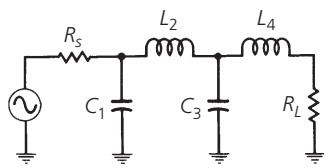


TABLE 3-5B. Chebyshev Low-Pass Prototype Element Values for 0.1-dB Ripple



n	R_S/R_L	C_1	L_2	C_3	L_4
2	1.984	0.983	1.950		
	2.000	0.909	2.103		
	2.500	0.564	3.165		
	3.333	0.375	4.411		
	5.000	0.228	6.700		
	10.000	0.105	13.322		
	∞	1.307	0.975		
3	1.000	1.864	1.280	1.834	
	0.900	1.918	1.209	2.026	
	0.800	1.997	1.120	2.237	
	0.700	2.114	1.015	2.517	
	0.500	2.557	0.759	3.436	
	0.400	2.985	0.615	4.242	
	0.300	3.729	0.463	5.576	
	0.200	5.254	0.309	8.225	
	0.100	9.890	0.153	16.118	
	∞	1.572	1.518	0.932	
4	1.984	0.920	2.586	1.304	1.826
	2.000	0.845	2.720	1.238	1.985
	2.500	0.516	3.766	0.869	3.121
	3.333	0.344	5.120	0.621	4.480
	5.000	0.210	7.708	0.400	6.987
	10.000	0.098	15.352	0.194	14.262
	∞	1.436	1.889	1.521	0.913
n	R_L/R_S	L_1	C_2	L_3	C_4

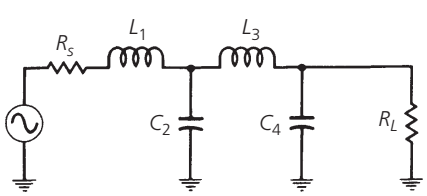


TABLE 3-6A. Chebyshev Low-Pass Prototype Element Values for 0.5-dB Ripple

But why would anyone deliberately design a filter with very poor initial stopband attenuation characteristics? The Bessel filter was originally optimized to obtain a *maximally flat group delay* or *linear phase* characteristic in the filter's passband. Thus, selectivity or stopband attenuation is not a primary concern when dealing with the Bessel filter. In high- and medium-Q filters, such as the Chebyshev and Butterworth filters, the phase response is extremely nonlinear over the filter's passband. This

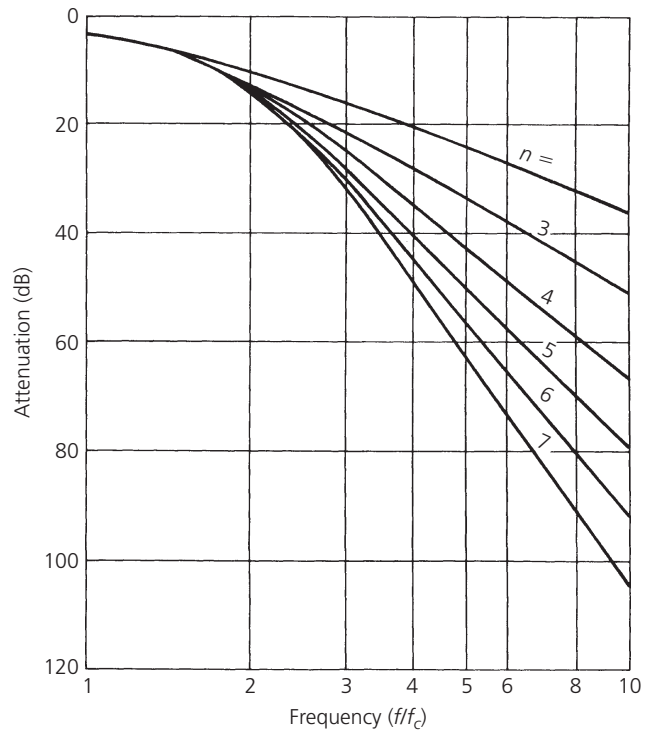


FIG. 3-20. Attenuation characteristics of Bessel filters.

phase nonlinearity results in distortion of wideband signals due to the widely varying time delays associated with the different spectral components of the signal. Bessel filters, on the other hand, with their maximally flat (constant) group delay are able to pass wideband signals with a minimum of distortion, while still providing *some* selectivity.

The low-pass prototype element values for the Bessel filter are given in Table 3-8. Table 3-8 tabulates the prototype element values for various ratios of source to load resistance.

FREQUENCY AND IMPEDANCE SCALING

Once you specify the filter, choose the appropriate attenuation response, and write down the low-pass prototype values, the next step is to transform the prototype circuit into a usable filter. Remember, the cutoff frequency of the prototype circuit is 0.159 Hz ($\omega = 1$ rad/sec), and it operates between a source and a load resistance that are normalized so that $R_L = 1$ ohm.

The transformation is effected through the following formulas:

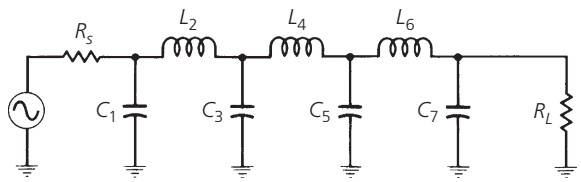
$$C = \frac{C_n}{2\pi f_c R} \qquad \text{(Eq. 3-12)}$$

and

$$L = \frac{RL_n}{2\pi f_c} \qquad \text{(Eq. 3-13)}$$

where

$C =$ the final capacitor value,
$L =$ the final inductor value,



n	R_s/R_L	C_1	L_2	C_3	L_4	C_5	L_6	C_7
5	1.000	1.807	1.303	2.691	1.303	1.807		
	0.900	1.854	1.222	2.849	1.238	1.970		
	0.800	1.926	1.126	3.060	1.157	2.185		
	0.700	2.035	1.015	3.353	1.058	2.470		
	0.600	2.200	0.890	3.765	0.942	2.861		
	0.500	2.457	0.754	4.367	0.810	3.414		
	0.400	2.870	0.609	5.296	0.664	4.245		
	0.300	3.588	0.459	6.871	0.508	5.625		
	0.200	5.064	0.306	10.054	0.343	8.367		
	0.100	9.556	0.153	19.647	0.173	16.574		
	∞	1.630	1.740	1.922	1.514	0.903		
6	1.984	0.905	2.577	1.368	2.713	1.299	1.796	
	2.000	0.830	2.704	1.291	2.872	1.237	1.956	
	2.500	0.506	3.722	0.890	4.109	0.881	3.103	
	3.333	0.337	5.055	0.632	5.699	0.635	4.481	
	5.000	0.206	7.615	0.406	8.732	0.412	7.031	
	10.000	0.096	15.186	0.197	17.681	0.202	14.433	
7	1.000	1.790	1.296	2.718	1.385	2.718	1.296	1.790
	0.900	1.835	1.215	2.869	1.308	2.883	1.234	1.953
	0.800	1.905	1.118	3.076	1.215	3.107	1.155	2.168
	0.700	2.011	1.007	3.364	1.105	3.416	1.058	2.455
	0.600	2.174	0.882	3.772	0.979	3.852	0.944	2.848
	0.500	2.428	0.747	4.370	0.838	2.289	0.814	3.405
	0.400	2.835	0.604	5.295	0.685	5.470	0.669	4.243
	0.300	3.546	0.455	6.867	0.522	7.134	0.513	5.635
	0.200	5.007	0.303	10.049	0.352	10.496	0.348	8.404
	0.100	9.456	0.151	19.649	0.178	20.631	0.176	16.665
	∞	1.646	1.777	2.031	1.789	1.924	1.503	0.895
n	R_L/R_s	L_1	C_2	L_3	C_4	L_5	C_6	L_7

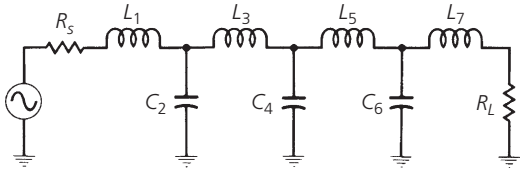
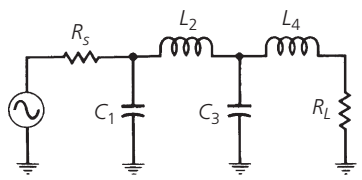


TABLE 3-6B. Chebyshev Low-Pass Prototype Element Values for 0.5-dB Ripple



n	R_s/R_L	C_1	L_2	C_3	L_4
2	3.000	0.572	3.132		
	4.000	0.365	4.600		
	8.000	0.157	9.658		
	∞	1.213	1.109		
3	1.000	2.216	1.088	2.216	
	0.500	4.431	0.817	2.216	
	0.333	6.647	0.726	2.216	
	0.250	8.862	0.680	2.216	
	0.125	17.725	0.612	2.216	
	∞	1.652	1.460	1.108	
4	3.000	0.653	4.411	0.814	2.535
	4.000	0.452	7.083	0.612	2.848
	8.000	0.209	17.164	0.428	3.281
	∞	1.350	2.010	1.488	1.106
n	R_L/R_s	L_1	C_2	L_3	C_4

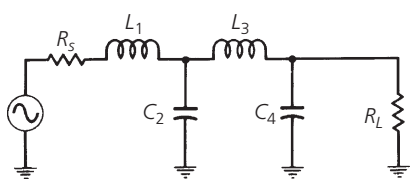


TABLE 3-7A. Chebyshev Low-Pass Prototype Element Values for 1.0-dB Ripple

C_n = a low-pass prototype element value,

L_n = a low-pass prototype element value,

R = the final load resistor value,

f_c = the final cutoff frequency.

The normalized low-pass prototype source resistor must also be transformed to its final value by multiplying it by the final value of the load resistor (Example 3-5). Thus, the ratio of the two always remains the same.

The process for designing a low-pass filter is a very simple one which involves the following procedure:

1. Define the response you need by specifying the required attenuation characteristics at selected frequencies.

EXAMPLE 3-5

Scale the low-pass prototype values of Fig. 3-19 (Example 3-4) to a cutoff frequency of 50 MHz and a load resistance of 250 ohms.

Solution

Use Equations 3-12 and 3-13 to scale each component as follows:

$$C_1 = \frac{3.546}{2\pi(50 \times 10^6)(250)} = 45\,\text{pF}$$

$$C_3 = \frac{9.127}{2\pi(50 \times 10^6)(250)} = 116\,\text{pF}$$

$$C_5 = \frac{7.889}{2\pi(50 \times 10^6)(250)} = 100\,\text{pF}$$

$$L_2 = \frac{(250)(0.295)}{2\pi(50 \times 10^6)} = 235\,\text{nH}$$

$$L_4 = \frac{(250)(0.366)}{2\pi(50 \times 10^6)} = 291\,\text{nH}$$

The source resistance is scaled by multiplying its normalized value by the final value of the load resistor.

$$R_{s(\text{final})} = 0.2(250) = 50\,\text{ohms}$$

The final circuit appears in Fig. 3-21.

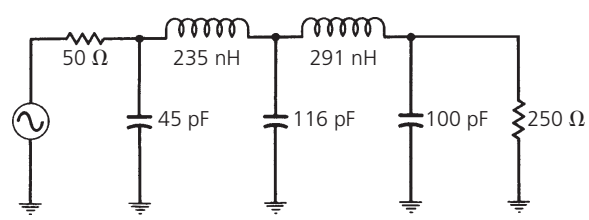
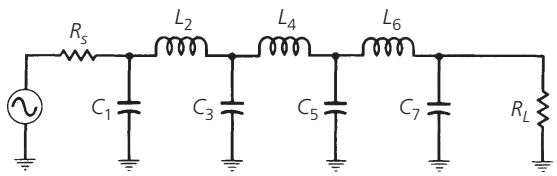


FIG. 3-21. Low-pass filter circuit for Example 3-5.

2. Normalize the frequencies of interest by dividing them by the cutoff frequency of the filter. This step forces your data to be in the same form as that of the attenuation curves of this chapter, where the 3-dB point on the curve is:

$$\frac{f}{f_c} = 1$$

3. Determine the maximum amount of ripple that you can allow in the passband. Remember, the greater the amount



n	R_s/R_L	C_1	L_2	C_3	L_4	C_5	L_6	C_7
5	1.000	2.207	1.128	3.103	1.128	2.207		
	0.500	4.414	0.565	4.653	1.128	2.207		
	0.333	6.622	0.376	6.205	1.128	2.207		
	0.250	8.829	0.282	7.756	1.128	2.207		
	0.125	17.657	0.141	13.961	1.128	2.207		
	∞	1.721	1.645	2.061	1.493	1.103		
6	3.000	0.679	3.873	0.771	4.711	0.969	2.406	
	4.000	0.481	5.644	0.476	7.351	0.849	2.582	
	8.000	0.227	12.310	0.198	16.740	0.726	2.800	
	∞	1.378	2.097	1.690	2.074	1.494	1.102	
7	1.000	2.204	1.131	3.147	1.194	3.147	1.131	2.204
	0.500	4.408	0.566	6.293	0.895	3.147	1.131	2.204
	0.333	6.612	0.377	9.441	0.796	3.147	1.131	2.204
	0.250	8.815	0.283	12.588	0.747	3.147	1.131	2.204
	0.125	17.631	0.141	25.175	0.671	3.147	1.131	2.204
	∞	1.741	1.677	2.155	1.703	2.079	1.494	1.102
n	R_L/R_s	L_1	C_2	L_3	C_4	L_5	C_6	L_7

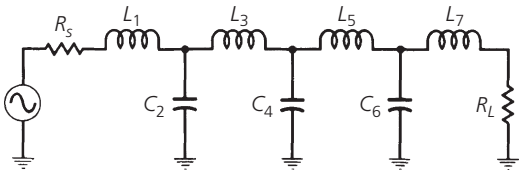


TABLE 3-7B. Chebyshev Low-Pass Prototype Element Values for 1.0-dB Ripple

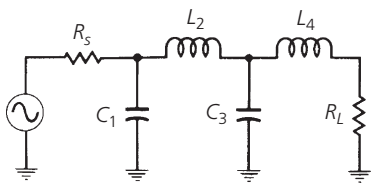
of ripple allowed, the more selective the filter is. Higher values of ripple may allow you to eliminate a few components.

4. Match the normalized attenuation characteristics (Steps 1 and 2) with the attenuation curves provided in this chapter. Allow yourself a small "fudge-factor" for good measure. This step reveals the minimum number of circuit elements that you can get away with–given a certain filter type.
5. Find the low-pass prototype values in the tables.
6. Scale all elements to the frequency and impedance of the final design.

Example 3-6 diagrams the process of designing a low-pass filter using the preceding steps.

HIGH-PASS FILTER DESIGN

Once you have learned the mechanics of low-pass filter design, high-pass design becomes a snap. You can use all of the attenuation response curves presented, thus far, for the low-pass filters by simply inverting the will produce an attenuation of about 60 dB at an f/f_c of 3 (Fig. 3-16). If you were working instead with a *high-pass filter* of the same size and type, you could still use Fig. 3-16 to tell you that at an f/f_c of 1/3 (or,



n	R_s/R_L	C_1	L_2	C_3	L_4
2	1.000	0.576	2.148		
	1.111	0.508	2.310		
	1.250	0.443	2.510		
	1.429	0.380	2.764		
	1.667	0.319	3.099		
	2.000	0.260	3.565		
	2.500	0.203	4.258		
	3.333	0.149	5.405		
	5.000	0.097	7.688		
	10.000	0.047	14.510		
	∞	1.362	0.454		
3	1.000	0.337	0.971	2.203	
	0.900	0.371	0.865	2.375	
	0.800	0.412	0.761	2.587	
	0.700	0.466	0.658	2.858	
	0.600	0.537	0.558	3.216	
	0.500	0.635	0.459	3.714	
	0.400	0.783	0.362	4.457	
	0.300	1.028	0.267	5.689	
	0.200	1.518	0.175	8.140	
	0.100	2.983	0.086	15.470	
	∞	1.463	0.843	0.293	
4	1.000	0.233	0.673	1.082	2.240
	1.111	0.209	0.742	0.967	2.414
	1.250	0.184	0.829	0.853	2.630
	1.429	0.160	0.941	0.741	2.907
	1.667	0.136	1.089	0.630	3.273
	2.000	0.112	1.295	0.520	3.782
	2.500	0.089	1.604	0.412	4.543
	3.333	0.066	2.117	0.306	5.805
	5.000	0.043	3.142	0.201	8.319
	10.000	0.021	6.209	0.099	15.837

TABLE 3-8A. Bessel Low-Pass Prototype Element Values

n	R_s/R_L	C_1	L_2	C_3	L_4
	∞	1.501	0.978	0.613	0.211
n	R_L/R_s	L_1	C_2	L_3	C_4

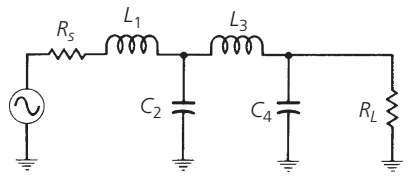


TABLE 3-8A. (Continued)

EXAMPLE 3-6

Design a low-pass filter to meet the following specifications:

$$f_c = 35\,\text{MHz},$$

Response greater than 60 dB down at 105 MHz,
Maximally flat passband—no ripple,

$$R_s = 50\,\text{ohms},$$

$$R_L = 500\,\text{ohms}.$$

Solution

The need for a maximally flat passband automatically indicates that the design must be a Butterworth response. The first step in the design process is to normalize everything. Thus,

$$\frac{R_s}{R_L} = \frac{50}{500} = 0.1$$

Next, normalize the frequencies of interest so that they may be found in the graph of Fig. 3-9. Thus, we have:

$$\frac{f_{60\text{dB}}}{f_{3\text{dB}}} = \frac{105\,\text{MHz}}{35\,\text{MHz}} = 3$$

We next look at Fig. 3-9 and find a response that is down at least 60 dB at a frequency ratio of $f/f_c = 3$. Fig. 3-9 indicates that it will take a minimum of 7 elements to provide the attenuation specified. Referring to the catalog of Butterworth low-pass prototype values given in Table 3-2 yields the prototype circuit of Fig. 3-22.

Continued on next page

Example 3-6.—Cont

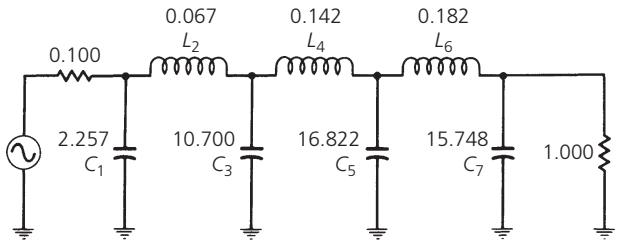


FIG. 3-22. Low-pass prototype circuit for Example 3-6.

We then scale these values using Equations 3-12 and 3-13. The first two values are worked out for you.

$$C_1 = \frac{2.257}{2\pi(35 \times 10^6)500}$$
$$= 21\text{ pF}$$

$$L_2 = \frac{(500)(0.067)}{2\pi(35 \times 10^6)}$$
$$= 152\text{ nH}$$

Similarly,

$$C_3 = 97\text{ pF},$$
$$C_5 = 153\text{ pF},$$
$$C_7 = 143\text{ pF},$$
$$L_4 = 323\text{ nH},$$
$$L_6 = 414\text{ nH},$$
$$R_S = 50\text{ ohms},$$
$$R_L = 500\text{ ohms}.$$

The final circuit is shown in Fig. 3-23.

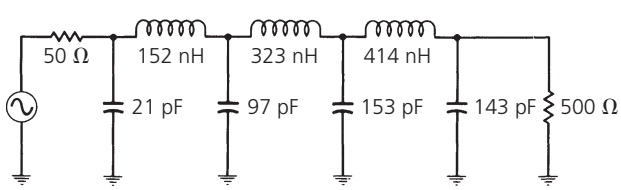
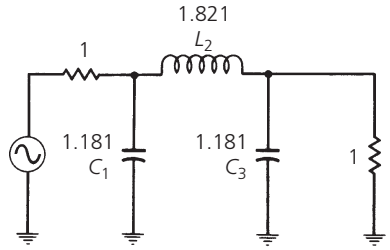


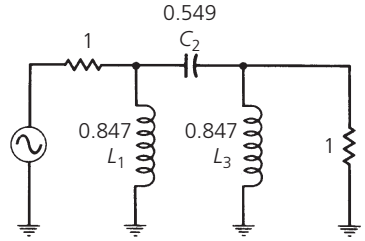
FIG. 3-23. Low-pass filter circuit for Example 3-6.

$f_c/f = 3$) a 5-element, 0.1-dB-ripple Chebyshev *high-pass filter* will also produce an attenuation of 60 dB. This is obviously more convenient than having to refer to more than one set of curves.

After finding the response that satisfies all of the requirements, the next step is to simply refer to the tables of low-pass prototype values and copy down the prototype values that are called for. High-pass values for the elements are then obtained directly from the low-pass prototype values as follows (refer to Fig. 3-24):



(A) *Low-pass prototype circuit*



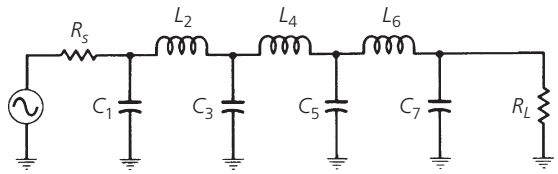
(B) *Equivalent high-pass prototype circuit*

FIG 3-24. Low-pass to high-pass filter transformation.

Simply replace each filter element with an element of the opposite type and with a reciprocal value. Thus, L_1 of Fig. 3-24B is equal to $1/C_1$ of Fig. 3-24A. Likewise, $C_2 = 1/L_2$ and $L_3 = 1/C_3$. Stated another way, if the low-pass prototype indicates a capacitor of 1.181 farads, then use an inductor with a value of $1/1.181 = 0.847$ henry, instead, for a high-pass design. However, the source and load resistors should not be altered.

The transformation process results in an attenuation characteristic for the high-pass filter that is an exact mirror image of the low-pass attenuation characteristic. The ripple, if there is any, remains the same and the magnitude of the slope of the stopband (or passband) skirts remains the same. Example 3-7 illustrates the design of high-pass filters.

A closer look at the filter designed in Example 3-7 reveals that it is symmetric. Indeed, all filters given for the equal termination class are symmetric. The equal termination class of filter thus yields a circuit that is easier to design (fewer calculations) and, in most cases, cheaper to build for a high-volume product, due to the number of equal valued components.



n	R_S/R_L	C_1	L_2	C_3	L_4	C_5	L_6	C_7
5	1.000	0.174	0.507	0.804	1.111	2.258		
	0.900	0.193	0.454	0.889	0.995	2.433		
	0.800	0.215	0.402	0.996	0.879	2.650		
	0.700	0.245	0.349	1.132	0.764	2.927		
	0.600	0.284	0.298	1.314	0.651	3.295		
	0.500	0.338	0.247	1.567	0.538	3.808		
	0.400	0.419	0.196	1.946	0.427	4.573		
	0.300	0.555	0.146	2.577	0.317	5.843		
	0.200	0.825	0.096	3.835	0.210	8.375		
	0.100	1.635	0.048	7.604	0.104	15.949		
	∞	1.513	1.023	0.753	0.473	0.162		
6	1.000	0.137	0.400	0.639	0.854	1.113	2.265	
	1.111	0.122	0.443	0.573	0.946	0.996	2.439	
	1.250	0.108	0.496	0.508	1.060	0.881	2.655	
	1.429	0.094	0.564	0.442	1.207	0.767	2.933	
	1.667	0.080	0.655	0.378	1.402	0.653	3.300	
	2.000	0.067	0.782	0.313	1.675	0.541	3.812	
	2.500	0.053	0.973	0.249	2.084	0.429	4.577	
	3.333	0.040	1.289	0.186	2.763	0.319	5.847	
	5.000	0.026	1.289	0.123	4.120	0.211	8.378	
	10.000	0.013	3.815	0.061	8.186	0.105	15.951	
	∞	1.512	1.033	0.813	0.607	0.379	0.129	
7	1.000	0.111	0.326	0.525	0.702	0.869	1.105	2.266
	0.900	0.122	0.292	0.582	0.630	0.963	0.990	2.440
	0.800	0.137	0.259	6.652	0.559	1.080	0.875	2.656
	0.700	0.156	0.226	0.743	0.487	1.231	0.762	2.932
	0.600	0.182	0.193	0.863	0.416	1.431	0.649	3.298
	0.500	0.217	0.160	1.032	0.346	1.711	0.537	3.809
	0.400	0.270	0.127	1.285	0.276	2.130	0.427	4.572
	0.300	0.358	0.095	1.705	0.206	2.828	0.318	5.838
	0.200	0.534	0.063	2.545	0.137	4.221	0.210	8.362
	0.100	1.061	0.031	5.062	0.068	8.397	0.104	15.917
	∞	1.509	1.029	0.835	0.675	0.503	0.311	0.105
n	R_L/R_s	L_1	C_2	L_3	C_4	L_5	C_6	L_7

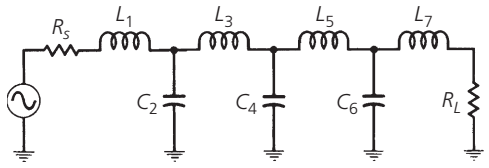


TABLE 3-8B. Bessel Low-Pass Prototype Element Values

THE DUAL NETWORK

Thus far, we have been referring to the group of low-pass prototype element value tables presented, and then we choose the schematic that is located either above or below the tables for the form of the filter that we are designing, depending on the value of R_L/R_s. Either form of the filter will produce exactly the same attenuation, phase, and group-delay characteristics, and each form is called the *dual* of the other.

Any filter network in a ladder arrangement, such as the ones presented in this chapter, can be changed into its dual form by application of the following rules:

1. Change all inductors to capacitors, and vice-versa, without changing element values. Thus, 3 henries becomes 3 farads.
2. Change all resistances into conductances, and vice-versa, with the value unchanged. Thus, 3 ohms becomes 3 mhos, or $\frac{1}{3}$ ohm.
3. Change all shunt branches to series branches, and vice versa.
4. Change all elements in series with each other into elements that are in parallel with each other.
5. Change all voltage sources into current sources, and vice versa.

Fig. 3-26 shows a ladder network and its dual representation.

Dual networks are convenient, in the case of equal terminations, if you desire to change the topology of the filter without changing the response. It is most often used, as shown in Example 3-7, to eliminate an unnecessary inductor which might have crept into the design through some other transformation process. Inductors are typically more lower-Q devices than capacitors and, therefore, exhibit higher losses. These losses tend to cause insertion loss, in addition to generally degrading the overall performance of the filter. The number of inductors in any network should, therefore, be reduced whenever possible.

A little experimentation with dual networks having unequal terminations will reveal that you can quickly get yourself into trouble if you are not careful. This is especially true if the load and source resistance are a design criteria and cannot be changed to suit the needs of your filter. Remember, when the dual of a network with unequal terminations is taken, then the terminations *must*, by definition, change value as shown in Fig. 3-26.

BANDPASS FILTER DESIGN

The low-pass prototype circuits and response curves given in this chapter can also be used in the design of bandpass filters. This is done through a simple transformation process similar to what was done in the high-pass case.

The most difficult task awaiting the designer of a bandpass filter, if the design is to be derived from the low-pass prototype, is in

EXAMPLE 3-7

Design an LC high-pass filter with an f_c of 60 MHz and a minimum attenuation of 40 dB at 30 MHz. The source and load resistance are equal at 300 ohms. Assume that a 0.5-dB passband ripple is tolerable.

Solution

First, normalize the attenuation requirements so that the low-pass attenuation curves may be used.

$$\frac{f}{f_c} = \frac{30\,\text{MHz}}{60\,\text{MHz}} = 0.5$$

Inverting, we get:

$$\frac{f_c}{f} = 2$$

Now, select a normalized low-pass filter that offers at least 40-dB attenuation at a ratio of $f_c/f = 2$. Reference to Fig. 3.17 (attenuation response of 0.5-dB-ripple Chebyshev filters) indicates that a normalized $n = 5$ Chebyshev will provide the needed attenuation. Table 3-6 contains the element values for the corresponding network. The normalized low-pass prototype circuit is shown in Fig. 3-25A. Note that the schematic below Table 3-6B was chosen as the low-pass prototype circuit rather than the schematic above the table. The reason for doing this will become obvious after the next step. Keep in mind, however, that the ratio of R_s/R_L is the same as the ratio of R_L/R_s, and is unity. Therefore, it does not matter which form is used for the prototype circuit.

Next, transform the low-pass circuit to a high-pass network by replacing each inductor with a capacitor, and vice versa, using reciprocal element values as shown in Fig. 3-25B. Note here that, had we begun with the low-pass prototype circuit shown above Table 3-6B, this transformation would have yielded a filter containing three inductors rather than the two shown in Fig. 3-25B. The object in any of these filter designs is to reduce the number of inductors in the final design. More on this later. The final step in the design process is to scale the network in both impedance and frequency using Equations 3-12 and 3-13. The first two calculations are done for you.

$$C_1 = \frac{\dfrac{1}{1.807}}{2\pi(60 \times 10^6)(300)} = 4.9\,\text{pF}$$

$$L_2 = \frac{300\left(\dfrac{1}{1.303}\right)}{2\pi(60 \times 10^6)} = 611\,\text{nH}$$

Example 3-7.—Cont

The remaining values are:

$$C_3 = 3.3\,\text{pF}$$

$$C_5 = 4.9\,\text{pF}$$

$$L_4 = 611\,\text{nH}$$

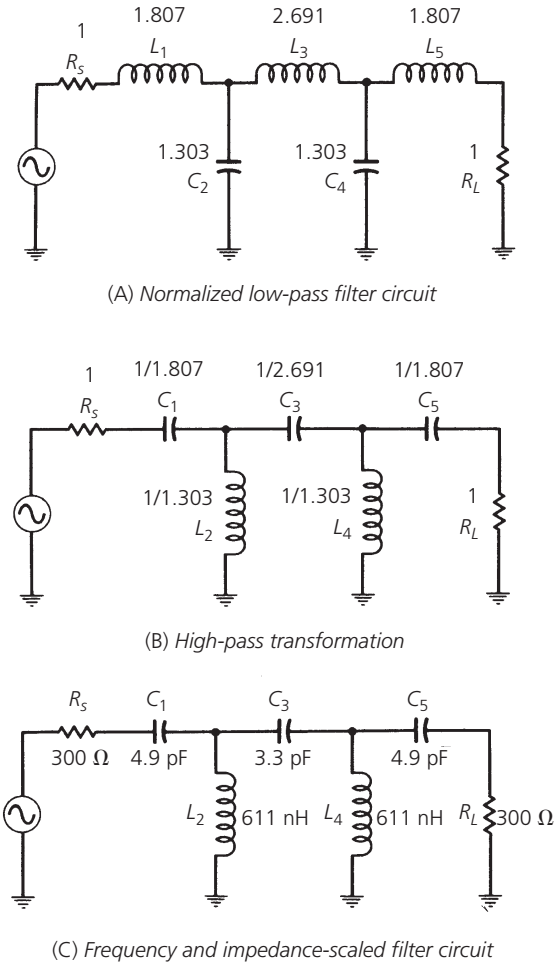


FIG. 3-25. High-pass filter design for Example 3-7.

The final filter circuit is given in Fig. 3-25C.

specifying the bandpass attenuation characteristics in terms of the low-pass response curves. A method for doing this is shown by the curves in Fig. 3-27. As you can see, when a low-pass design is transformed into a bandpass design, the attenuation bandwidth ratios remain the same. This means that a low-pass filter with a 3-dB cutoff frequency, or a bandwidth of 2 kHz, would transform into a bandpass filter with a 3-dB bandwidth of 2 kHz. If the response of the low-pass network were down 30 dB

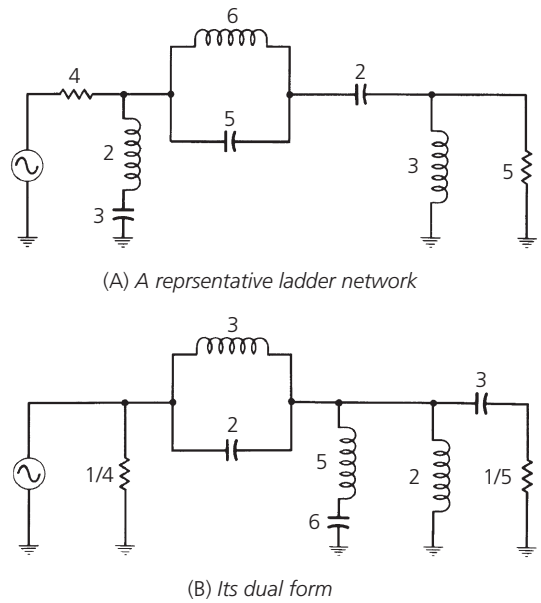


FIG. 3-26. Duality.

at a frequency or bandwidth of 4 kHz ($f/f_c = 2$), then the response of the bandpass network would be down 30 dB at a bandwidth of 4 kHz. Thus, the normalized f/f_c axis of the low-pass attenuation curves becomes a ratio of bandwidths rather than frequencies,

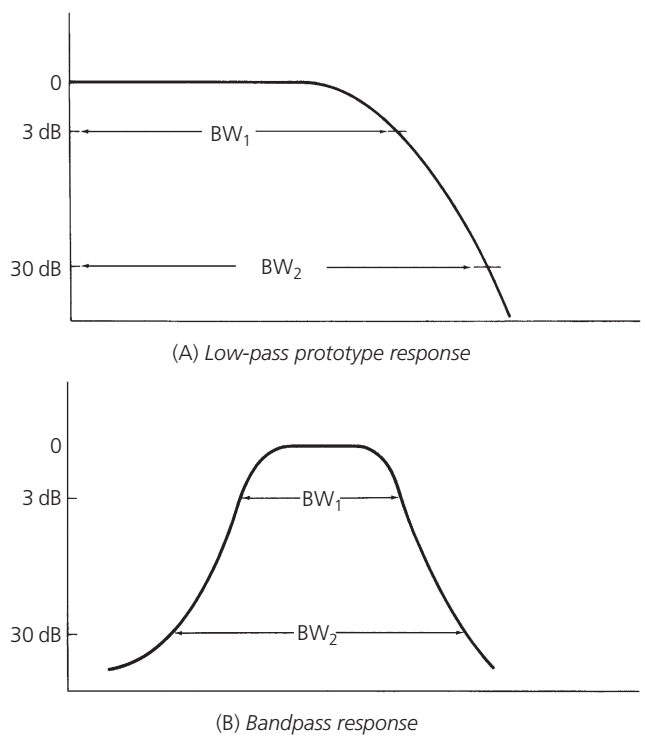


FIG. 3-27. Low-pass to bandpass transformation bandwidths.

such that:

$$\frac{BW}{BW_c} = \frac{f}{f_c} \qquad \text{(Eq. 3-14)}$$

where

BW = the bandwidth at the required value of attenuation,

BW_c = the 3-dB bandwidth of the bandpass filter.

Often a bandpass response is not specified, as in Example 3-8. Instead, the requirements are often given as attenuation values at specified frequencies as shown by the curve in Fig. 3-28. In this case, you must transform the stated requirements into information that takes the form of Equation 3-14. As an example, consider Fig. 3-28. How do we convert the data that is given into the bandwidth ratios we need? Before we can answer that, we have to find f_3. Use the following method.

EXAMPLE 3-8

Find the Butterworth low-pass prototype circuit which, when transformed, would satisfy the following bandpass filter requirements:

$$BW_{3dB} = 2\,\text{MHz}$$
$$BW_{40dB} = 6\,\text{MHz}$$

Solution

Note that we are not concerned with the center frequency of the bandpass response just yet. We are only concerned with the relationship between the above requirements and the low-pass response curves. Using Equation 3-14, we have:

$$\frac{BW}{BW_c} = \frac{f}{f_c} = \frac{BW_{40dB}}{BW_{3dB}}$$
$$= \frac{6\,\text{MHz}}{2\,\text{MHz}}$$
$$= 3$$

Therefore, turn to the Butterworth response curves shown in Fig. 3-9 and find a prototype value that will provide 40 dB of attenuation at an $f/f_c = 3$. The curves indicate a 5-element Butterworth filter will provide the needed attenuation.

The frequency response of a bandpass filter exhibits geometric symmetry. That is, it is only symmetric when plotted on a logarithmic scale. The center frequency of a geometrically symmetric filter is given by the formula:

$$f_o = \sqrt{f_a f_b} \qquad \text{(Eq. 3-15)}$$

where f_a and f_b are any two frequencies (one above and one below the passband) having equal attenuation. Therefore, the center

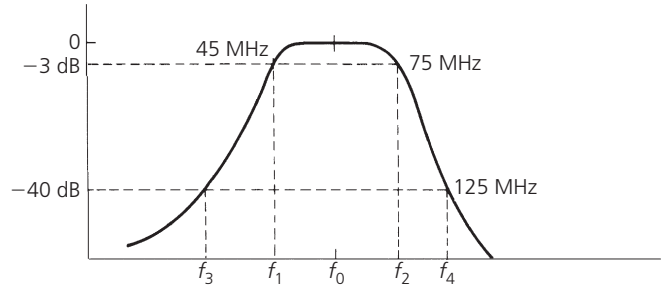


FIG. 3-28. Typical bandpass specifications.

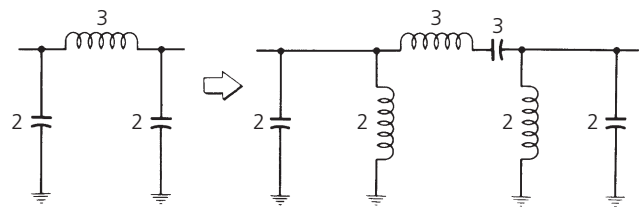


FIG. 3-29. Low-pass to bandpass circuit transformation.

frequency of the response curve shown in Fig. 3-28 must be

$$f_o = \sqrt{(45)(75)\,\text{MHz}}$$
$$= 58.1\,\text{MHz}$$

We can use Equation 3-15 again to find f_3.

$$58.1 = \sqrt{f_3(125)}$$

or,

$$f_3 = 27\,\text{MHz}$$

Now that f_3 is known, the data of Fig. 3-28 can be put into the form of Equation 3-14.

$$\frac{BW_{40dB}}{BW_{3dB}} = \frac{125\,\text{MHz} - 27\,\text{MHz}}{75\,\text{MHz} - 45\,\text{MHz}}$$
$$= 3.27$$

To find a low-pass prototype curve that will satisfy these requirements, simply refer to any of the pertinent graphs presented in this chapter and find a response that will provide 40 dB of attenuation at an f/f_c of 3.27. (A fourth-order or better Butterworth filter will do quite nicely.)

The actual transformation from the low-pass to the bandpass configuration is accomplished by resonating each low-pass element with an element of the opposite type and of the same value. All shunt elements of the low-pass prototype circuit become parallel-resonant circuits, and all series elements become series-resonant circuits. This process is illustrated in Fig. 3-30.

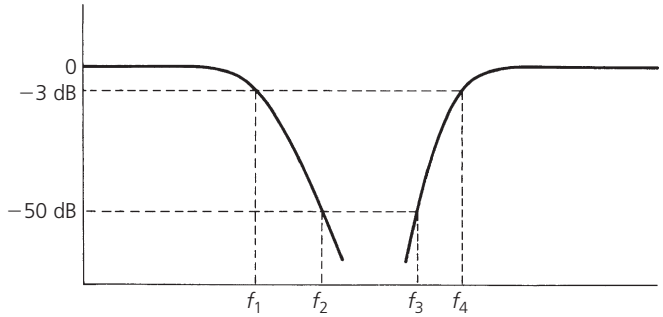


FIG. 3-30. Typical band-rejection filter curves.

To complete the design, the transformed filter is then frequency- and impedance-scaled using the following formulas. For the parallel-resonant branches,

$$C = \frac{C_n}{2\pi RB} \qquad \text{(Eq. 3-16)}$$

$$L = \frac{RB}{2\pi f_0^2 L_n} \qquad \text{(Eq. 3-17)}$$

and, for the series-resonant branches,

$$C = \frac{B}{2\pi f_0^2 C_n R} \qquad \text{(Eq. 3-18)}$$

$$L = \frac{RL_n}{2\pi B} \qquad \text{(Eq. 3-19)}$$

where, in all cases,

R = the final load impedance,
B = the 3-dB bandwidth of the final design,
f_o = the geometric center frequency of the final design,
L_n = the normalized inductor *bandpass* element values,
C_n = the normalized capacitor *bandpass* element values.

Example 3-9 furnishes one final example of the procedure for designing a bandpass filter.

SUMMARY OF THE BANDPASS FILTER DESIGN PROCEDURE

1. Transform the bandpass requirements into an equivalent low-pass requirement using Equation 3-14.
2. Refer to the low-pass attenuation curves provided in order to find a response that meets the requirements of Step 1.
3. Find the corresponding low-pass prototype and write it down.
4. Transform the low-pass network into a bandpass configuration.
5. Scale the bandpass configuration in both impedance and frequency using Equations 3-16 through 3-19.

BAND-REJECTION FILTER DESIGN

Band-rejection filters are very similar in design approach to the bandpass filter of the last section. Only, in this case, we want to *reject* a certain group of frequencies as shown by the curves in Fig. 3-30.

The band-reject filter lends itself well to the low-pass prototype design approach using the same procedures as were used for the bandpass design. First, define the bandstop requirements in terms of the low-pass attenuation curves. This is done by using the inverse of Equation 3-14. Thus, referring to Fig. 3-30, we have:

$$\frac{BW_c}{BW} = \frac{f_4 - f_1}{f_3 - f_2}$$

This sets the attenuation characteristic that is needed and allows you to read directly off the low-pass attenuation curves by substituting BW_c/BW for f_c/f on the normalized frequency axis. Once the number of elements that are required in the low-pass prototype circuit is determined, the low-pass network is transformed into a band-reject configuration as follows:

Each shunt element in the low-pass prototype circuit is replaced by a shunt *series-resonant circuit*, and each series-element is replaced by a series *parallel-resonant circuit*.

This is shown in Fig. 3-31. Note that both elements in each of the resonant circuits have the same normalized value.

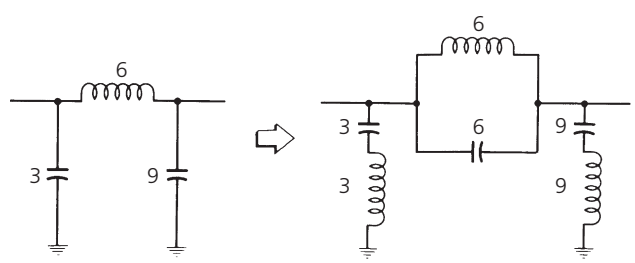


FIG. 3-31. Low-pass to band-reject transformation.

Once the prototype circuit has been transformed into its band-reject configuration, it is then scaled in impedance and frequency using the following formulas. For all series-resonant circuits:

$$C = \frac{C_n}{2\pi RB} \qquad \text{(Eq. 3-20)}$$

$$L = \frac{RB}{2\pi f_0^2 L_n} \qquad \text{(Eq. 3-21)}$$

For all parallel-resonant circuits:

$$C = \frac{B}{2\pi f_0^2 RC_n} \qquad \text{(Eq. 3-22)}$$

$$L = \frac{RL_n}{2\pi B} \qquad \text{(Eq. 3-23)}$$

Example 3-9

Design a bandpass filter with the following requirements:

$$f_o = 75\,\text{MHz} \quad \text{Passband Ripple} = 1\,\text{dB}$$

$$BW_{3\text{dB}} = 7\,\text{MHz} \quad R_s = 50\,\text{ohms}$$

$$BW_{45\text{dB}} = 35\,\text{MHz} \quad R_L = 100\,\text{ohms}$$

Solution

Using Equation 3-14:

$$\frac{BW_{45\text{dB}}}{\text{BW}_{3\text{dB}}} = \frac{35}{7} = 5$$

Substitute this value for f/f_c in the low-pass attenuation curves for the 1-dB-ripple Chebyshev response shown in Fig. 3-18. This reveals that a 3-element filter will provide about 50 dB of attenuation at an $f/f_c = 5$, which is more than adequate. The corresponding element values for this filter can be found in Table 3-7 for an $R_s/R_L = 0.5$ and an $n = 3$. This yields the low-pass prototype circuit of Fig. 3-32A which is transformed into the bandpass prototype circuit of Fig. 3-32B. Finally, using Equations 3-16 through 3-19, we obtain the final circuit that is shown in Fig. 3-32C. The calculations follow. Using Equations 3-16 and 3-17:

$$C_1 = \frac{4.431}{2\pi(100)(7 \times 10^6)} = 1007\,\text{pF}$$

$$L_1 = \frac{(100)(7 \times 10^6)}{2\pi(75 \times 10^6)^2(4.431)} = 4.47\,\text{nH}$$

Using Equations 3-18 and 3-19:

$$C_2 = \frac{7 \times 10^6}{2\pi(75 \times 10^6)^2(0.817)100} = 2.4\,\text{pF}$$

$$L_2 = \frac{(100)(0.817)}{2\pi(7 \times 10^6)} = 1.86\,\mu\text{H}$$

Similarly,

$$C_3 = 504\,\text{pF}$$

$$L_3 = 8.93\,\text{nH}$$

where, in all cases,

B = the 3-dB bandwidth,
R = the final load resistance,
f_o = the geometric center frequency,
C_n = the normalized capacitor band-reject element value,
L_n = the normalized inductor band-reject element value.

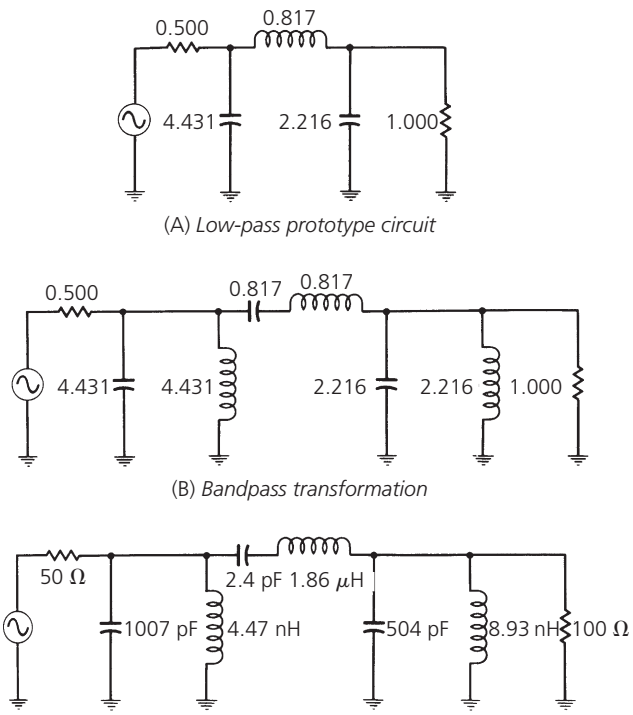


FIG. 3-32. Bandpass filter design for Example 3-9.

THE EFFECTS OF FINITE *Q*

Thus far in this chapter, we have assumed the inductors and capacitors used in the designs to be lossless. Indeed, all of the response curves presented in this chapter are based on that assumption. But we know from our previous study of Chapters 1 and 2 that even though capacitors can be approximated as having infinite Q, inductors cannot, and the effects of the finite-Q inductor must be taken into account in any filter design.

The use of finite element Q in a design intended for lossless elements causes the following unwanted effects (refer to Fig. 3-33):

1. Insertion loss of the filter is increased whereas the final stopband attenuation does not change. The relative attenuation between the two is decreased.
2. At frequencies in the vicinity of cutoff (f_c), the response becomes more rounded and usually results in an attenuation greater than the 3 dB that was originally intended.
3. Ripple that was designed into the passband will be reduced. If the element Q is sufficiently low, ripple will be totally eliminated.
4. For band-reject filters, the attenuation in the stop-band becomes finite. This, coupled with an increase in *passband* insertion loss, decreases the relative attenuation significantly.

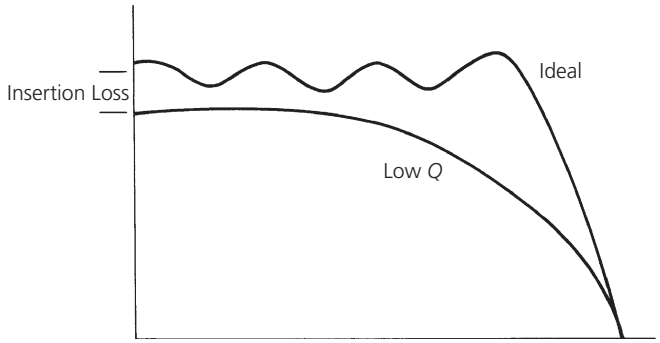


FIG. 3-33. The effect of finite-*Q* elements on filter response.

Filter Type	Minimum Element *Q* Required
Bessel	3
Butterworth	15
0.01-dB Chebyshev	24
0.1-dB Chebyshev	39
0.5-dB Chebyshev	57
1-dB Chebyshev	75

TABLE 3-9. Filter Elemental-*Q* Requirements

Regardless of the gloomy predictions outlined above, however, it is possible to design filters, using the approach outlined in this chapter, that very closely resemble the ideal response of each network. The key is to use the highest-Q inductors available for the given task. Table 3-9 outlines the recommended minimum element-Q requirements for the filters presented in this chapter. Keep in mind, however, that any time a low-Q component is used, the actual attenuation response of the network strays from the ideal response to a degree depending upon the element Q. It is, therefore, highly recommended that you make it a habit to use only the highest-Q components available.

The insertion loss of the filters presented in this chapter can be calculated in the same manner as was used in Chapter 2. Simply replace each reactive element with resistor values corresponding to the Q of the element and, then, exercise the voltage division rule from source to load.

CHAPTER 4

IMPEDANCE Matching

Impedance matching is often necessary in the design of RF circuitry to provide the maximum possible transfer of power between a source and its load. Probably the most vivid example of the need for such a transfer of power occurs in the front end of any sensitive receiver. Obviously, any *unnecessary* loss in a circuit that is already carrying extremely small signal levels simply cannot be tolerated. Therefore, in most instances, extreme care is taken during the initial design of such a front end to make sure that each device in the chain is matched to its load.

In this chapter, then, we will study several methods of matching a given source to a given load. This will be done numerically, with the aid of the Smith Chart, and by using software design tools. In all cases, exact step-by-step procedures will be presented, making any calculations as painless as possible.

BACKGROUND

There is a well-known theorem which states that, for *DC circuits*, maximum power will be transferred from a source to its load if the *load resistance* equals the *source resistance*. A simple proof of this theorem is given by the calculations and the sketches shown in Fig. 4-1. In the calculation, for convenience, the source is normalized for a resistance of one ohm and a source voltage of one volt.

In dealing with AC or time-varying waveforms, however, that same theorem states that the maximum transfer of power, from a source to its load, occurs when the *load impedance* (Z_L) is equal to the *complex conjugate* of the *source impedance*. Complex conjugate simply refers to a complex impedance having the same *real part* with an opposite reactance. Thus, if the source

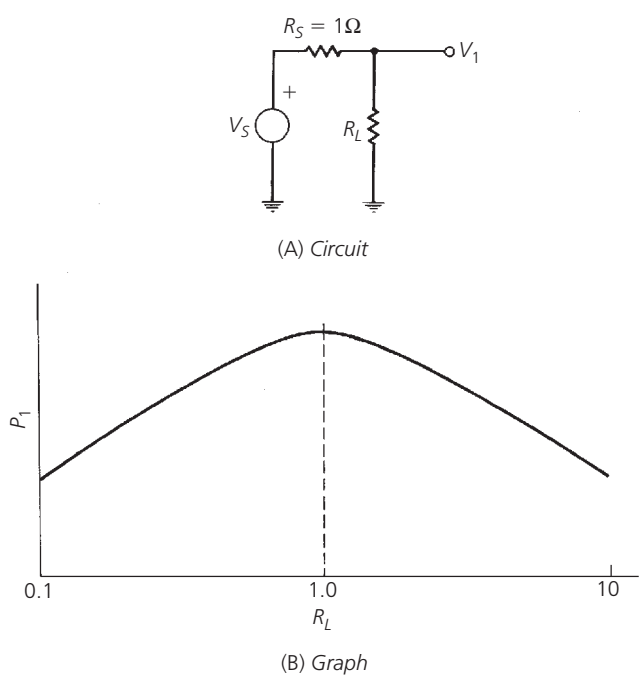


FIG. 4-1. The power theorem.

Proof that P_{out} *MAX* occurs when $R_L = R_s$, in the circuit of Fig. 4-1A, is given by the formula:

$$V_1 = \frac{R_L}{R_S + R_L}(V_S)$$

Set $V_s = 1$ and $R_s = 1$, for convenience. Therefore,

$$V_1 = \frac{R_L}{1 + R_L}$$

Then, the power into R_L is:

$$P_1 = \frac{V_1^2}{R_L} = \frac{\left(\frac{R_L}{1 + R_L}\right)^2}{R_L} = \frac{R_L}{(1 + R_L)^2}$$

If you plot P_1 versus R_L, as in the preceding equation, the result is shown by the curve of the graph in Fig. 4-1B.

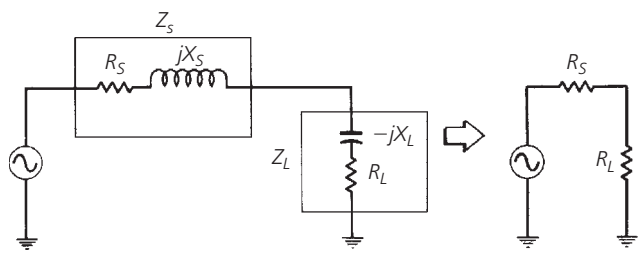


FIG. 4-2. Source impedance driving its complex conjugate and the resulting equivalent circuit.

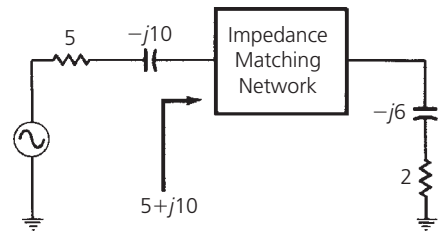


FIG. 4-3. Impedance transformation.

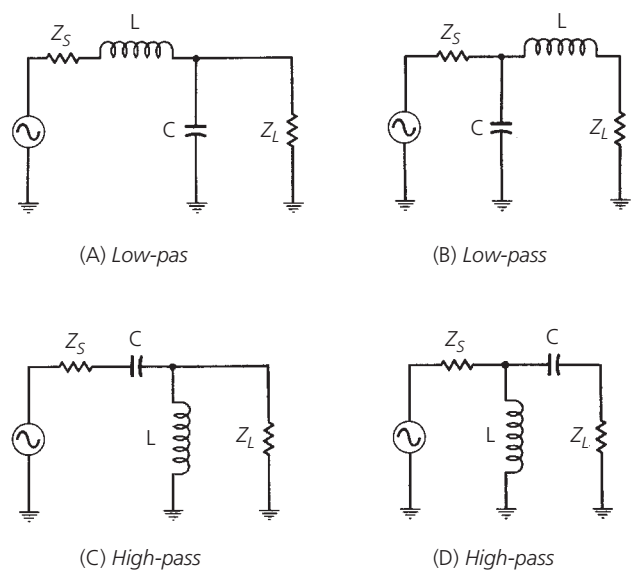


FIG. 4-4. The L network.

impedance were $Z_s = R + jX$, then its complex conjugate would be $Z_s = R - jX$.

If you followed the mathematics associated with Fig. 4-1, then it should be obvious why maximum transfer of power does occur when the load impedance is the complex conjugate of the source. This is shown schematically in Fig. 4-2. The source (Z_s), with a series reactive component of $+jX$ (an inductor), is driving its complex conjugate load impedance consisting of a $-jX$ reactance (capacitor) in series with R_L. The $+jX$ component of the source and the $-jX$ component of the load are in series and, thus, cancel each other, leaving only R_s and R_L, which are equal by definition. Since R_s and R_L are equal, maximum power transfer will occur. So when we speak of a source driving its complex conjugate, we are simply referring to a condition in which any *source* reactance is resonated with an equal and opposite *load* reactance, thus leaving only equal resistor values for the source and the load terminations.

The primary objective in any impedance *matching* scheme, then, is to force a load impedance to "look like" the complex conjugate of the source impedance so that maximum power may be transferred to the load. This is shown in Fig. 4-3 where a load impedance of $2 - j6$ ohms is transformed by the impedance matching network to a value of $5 + j10$ ohms. Therefore, the source "sees" a load impedance of $5 + j10$ ohms, which just happens to be its complex conjugate. It should be noted here that because we are dealing with *reactances*, which are frequency dependent, the *perfect* impedance match can occur only at one frequency. That is the frequency at which the $+jX$ component exactly equals the $-jX$ component and, thus, cancellation or resonance occurs. At all other frequencies removed from the matching center frequency, the impedance match becomes progressively worse and eventually nonexistent. This can be a problem in broadband circuits where we would ideally like to provide a perfect match everywhere within the broad passband. There are methods, however, of increasing the bandwidth of the match and a few of these methods will be presented later in this chapter.

There are an infinite number of possible networks that could be used to perform the impedance matching function of Fig. 4-3. Something as simple as a 2-element LC network or as elaborate as a 7-element filter, depending on the application, would work equally well. The remainder of this chapter is devoted to providing you with an insight into a few of those infinite possibilities. After studying this chapter, you should be able to match almost any two complex loads with a minimum of effort.

THE L NETWORK

Probably the simplest and most widely used matching circuit is the L network shown in Fig. 4-4. This circuit receives its name because of the component orientation, which resembles the shape of an L. As shown in the sketches, there are four possible arrangements of the two L and C components. Two of the arrangements (Figs. 4-4A and 4-4B) are in a low-pass configuration while the other two (Figs. 4-4C and 4-4D) are in a high-pass configuration. Both of these circuits should be recognized from Chapter 3.

Before we introduce equations which can be used to design the matching networks of Fig. 4-4, let's first analyze an existing matching network so that we can understand exactly how the impedance match occurs. Once this analysis is made, a little of the "black magic" surrounding impedance matching should subside.

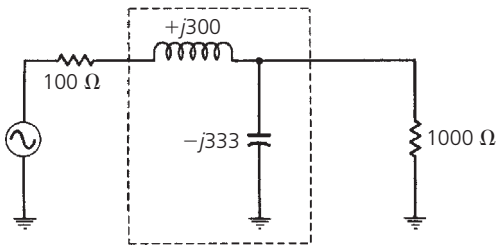


FIG. 4-5. Simple impedance-match network between a 100-ohm source and a 1000-ohm load.

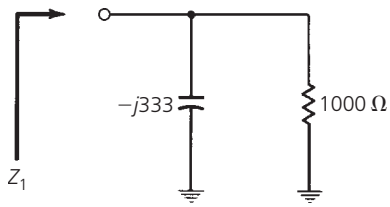


FIG. 4-6. Impedance looking into the parallel combination of R_L and X_c.

Fig. 4-5 shows a simple L network impedance-matching circuit between a 100-ohm source and a 1000-ohm load. Without the impedance-matching network installed, and with the 100-ohm source driving the 1000-ohm load directly, about 4.8 dB of the available power from the source would be lost. Thus, roughly one-third of the signal *available* from the source is gone before we even get started. The impedance-matching network eliminates this loss and allows for maximum power transfer to the load. This is done by forcing the 100-ohm source to see 100 ohms when it looks into the impedance-matching network. But how?

If you analyze Fig. 4-5, the simplicity of how the match occurs will amaze you. Take a look at Fig. 4-6. The first step in the analysis is to determine what the load impedance actually looks like when the $-j333$-ohm capacitor is placed across the 1000-ohm load resistor. This is easily calculated by:

$$\begin{aligned} Z &= \frac{X_c R_L}{X_c + R_L} \\ &= \frac{-j333(1000)}{-j333 + 1000} \\ &= 315\angle{-71.58^\circ} \\ &= 100 - j300 \text{ ohms} \end{aligned}$$

Thus, the parallel combination of the $-j333$-ohm capacitor and the 1000-ohm resistor *looks like* an impedance of $100 - j300$ ohms. This is a *series* combination of a 100-ohm resistor and a $-j300$-ohm capacitor as shown in Fig. 4-7. Indeed, if you hooked a signal generator up to circuits that are similar to Figs. 4-6 and 4-7, you would not be able to tell the difference between the two as they would exhibit the same characteristics (except at DC, obviously).

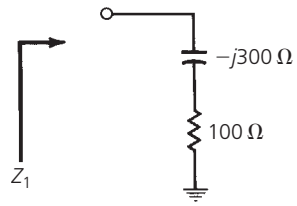


FIG. 4-7. Equivalent circuit of Fig. 4-6.

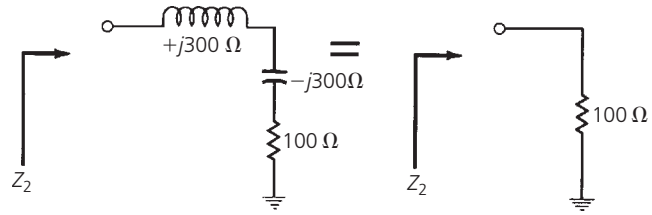


FIG. 4-8. Completing the match.

Now that we have an *apparent* series $100 - j300$-ohm impedance for a load, all we must do to complete the impedance match to the 100-ohm source is to add an equal and opposite ($+j300$ ohm) reactance in series with the network of Fig. 4-7. The addition of the $+j300$-ohm inductor causes cancellation of the $-j300$-ohm capacitor leaving only an apparent 100-ohm load resistor. This is shown in Fig. 4-8. Keep in mind here that the actual network topology of Fig. 4-5 has not changed. All we have done is to analyze small portions of the network so that we can understand the function of each component.

To summarize then, the function of the *shunt* component of the impedance-matching network is to transform a larger impedance down to a smaller value with a real part equal to the real part of the other terminating impedance (in our case, the 100-ohm source). The series impedance-matching element then resonates with or cancels any reactive component present, thus leaving the source driving an apparently equal load for optimum power transfer. So you see, the impedance match isn't "black magic" at all but can be completely explained every step of the way.

Now, back to the *design* of the impedance-matching networks of Fig. 4-4. These circuits can be very easily designed using the following equations:

$$Q_s = Q_p = \sqrt{\frac{R_p}{R_s} - 1} \qquad \text{(Eq. 4-1)}$$

$$Q_s = \frac{X_s}{R_s} \qquad \text{(Eq. 4-2)}$$

$$Q_p = \frac{R_p}{X_p} \qquad \text{(Eq. 4-3)}$$

where, as shown in Fig. 4-9:

Q_s = the Q of the series leg,
Q_p = the Q of the shunt leg,
R_p = the shunt resistance,

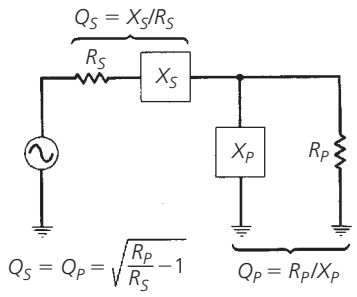


FIG. 4-9. Summary of the L-network design.

$X_p =$ the shunt reactance,
$R_s =$ the series resistance,
$X_s =$ the series reactance.

The quantities X_p and X_s may be either capacitive or inductive reactance but each must be of the opposite type. Once X_p is chosen as a capacitor, for example, X_s must be an inductor, and vice versa. Example 4-1 illustrates the procedure.

DEALING WITH COMPLEX LOADS

The design of Example 4-1 was used for the simple case of matching two real impedances (pure resistances). It is very rare when such an occurrence actually exists in the real world. Transistor input and output impedances are almost always *complex*; that is they contain both resistive and reactive components ($R \pm jX$). Transmission lines, mixers, antennas, and most other sources and loads are no different in that respect. Most will always have some reactive component which must be dealt with. It is, therefore, necessary to know how to handle these stray reactances and, in some instances, to actually put them to work for you.

There are two basic approaches in handling complex impedances:

1. Absorption—To actually absorb any stray reactances into the impedance-matching network itself. This can be done through prudent placement of each matching element such that element capacitors are placed in parallel with stray capacitances, and element inductors are placed in series with any stray inductances. The *stray* component values are then subtracted from the *calculated* element values, leaving new element values (C', L'), which are smaller than the calculated element values.
2. Resonance—To resonate any stray reactance with an equal and opposite reactance at the frequency of interest. Once this is done the matching network design can proceed as shown for two pure resistances in Example 4-1.

Of course, it is possible to use both of the approaches outlined above at the same time. In fact, the majority of

EXAMPLE 4-1

Design a circuit to match a 100-ohm source to a 1000-ohm load at 100 MHz. Assume that a DC voltage must also be transferred from the source to the load.

Solution

The need for a DC path between the source and load dictates the need for an inductor in the series leg, as shown in Fig. 4-4A. From Equation 4-1, we have:

$$Q_s = Q_p = \sqrt{\frac{1000}{100} - 1} = \sqrt{9} = 3$$

From Equation 4-2, we get:

$$X_s = Q_s R_s = (3)(100) = 300 \text{ ohms (inductive)}$$

Then, from Equation 4-3,

$$X_p = \frac{R_p}{Q_p} = \frac{1000}{3} = 333 \text{ ohms (capacitive)}$$

Thus, the component values at 100 MHz are:

$$L = \frac{X_s}{\omega} = \frac{300}{2\pi(100 \times 10^6)} = 477 \text{ nH}$$

$$C = \frac{1}{\omega X_p} = \frac{1}{2\pi(100 \times 10^6)(333)} = 4.8 \text{ pF}$$

This yields the circuit shown in Fig. 4-10. Notice that what you have done is to design the circuit that was previously given in Fig. 4-5 and then analyzed.

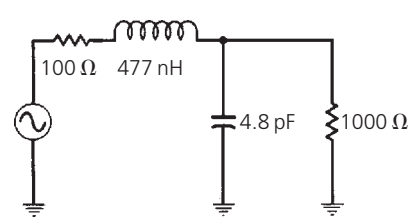


FIG. 4-10. Final circuit for Example 4-1.

impedance-matching designs probably do utilize a little of both. Let's take a look at two simple examples to help clarify matters.

Notice that nowhere in Example 4-2 was a *conjugate* match even mentioned. However, you can rest assured that if you perform the simple analysis outlined in the previous section of this chapter, the impedance looking into the matching network, as seen by the source, will be $100 - j126$ ohms, which is indeed the complex conjugate of $100 + j126$ ohms.

Obviously, if the *stray* element values are larger than the calculated element values, absorption cannot take place. If, for instance, the *stray* capacitance of Fig. 4-11 were 20 pF, we could not have added a *shunt* element capacitor to give us the total needed shunt capacitance of 4.8 pF. In a situation such as this, when absorption is not possible, the concept of resonance coupled with absorption will often do the trick.

Examples 4-2 and 4-3 detail some very important concepts in the design of impedance-matching networks. With a little planning and preparation, the design of simple impedance-matching networks between complex loads becomes a simple number-crunching task using elementary algebra. Any stray reactances present in the source and load can usually be absorbed in the matching network (Example 4-2), or they can be resonated with an equal and opposite reactance, which is then absorbed into the network (Example 4-3).

THREE-ELEMENT MATCHING

Equation 4-1 reveals a potential disadvantage of the 2-element L networks described in the previous sections. It is a fact that once R_s and R_p, or the source and load impedance, are determined, the Q of the network is defined. In other words, with the L network, the designer does not have a choice of circuit Q and simply must take what he gets. This is, of course, usually the case because the source and load impedance are typically given in any design and, thus, R_p and R_s cannot be changed.

The lack of circuit-Q versatility in a matching network can be a hindrance, however, especially if a *narrow* bandwidth is required. The 3-element network overcomes this disadvantage and can be used for narrow-band high-Q applications. Furthermore, the designer can *select* any practical circuit Q that he wishes as long as it is *greater than* that Q which is possible with the L-matching network alone. In other words, the circuit Q established with an L-matching network is the *minimum* circuit Q available in the 3-element matching arrangement.

The 3-element network (shown in Fig. 4-17) is called a *Pi network* because it closely resembles the Greek letter π. Its companion network (shown in Fig. 4-18) is called a *T network* for equally obvious reasons.

The Pi Network

The Pi network can best be described as two "back-to-back" L networks that are both configured to match the load and the source to an invisible or "virtual" resistance located at

EXAMPLE 4-2

Use the absorption approach to match the source and load shown in Fig. 4-11 (at 100 MHz).

Solution

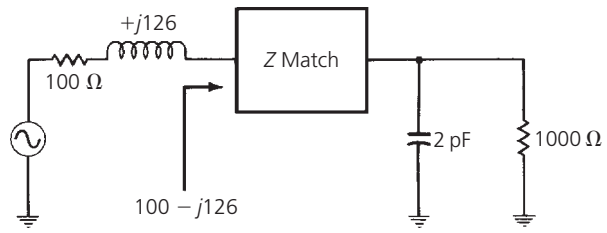


FIG. 4-11. Complex source and load circuit for Example 4-2.

The first step in the design process is to totally ignore the reactances and simply match the 100-ohm real part of the source to the 1000-ohm real part of the load (at 100 MHz). Keep in mind that you would like to use a matching network that will place element inductances in series with stray inductance and element capacitances in parallel with stray capacitances. Thus, conveniently, the network circuit shown in Fig. 4-4A is again chosen for the design and, again, Example 4-1 is used to provide the details of the procedure. Thus, the calculated values for the network, if we ignore stray reactances, are shown in the circuit of Fig. 4-10. But, since the stray reactances really do exist, the design is not yet finished as we must now somehow absorb the stray reactances into the matching network. This is done as follows. At the load end, we need 4.8 pF of capacitance for the matching network. We already have a stray 2 pF available at the load, so why not use it? Thus, if we use a 2.8-pF *element* capacitor, the *total* shunt capacitance becomes 4.8 pF, the design value. Similarly, at the source, the matching network calls for a series 477-nH inductor. We already have a $+j126$-ohm, or 200-nH, inductor available in the source. Thus, if we use an actual element inductance of $477\text{ nH} - 200\text{ nH} = 277\text{ nH}$, then the *total* series inductance will be 477 nH, which is the calculated design value. The final design circuit is shown in Fig. 4-12.

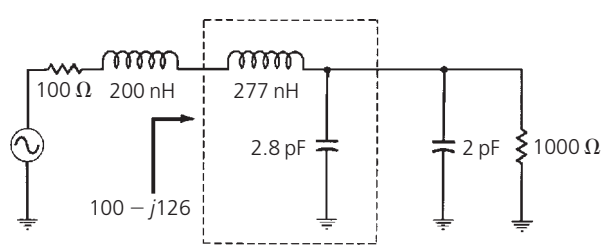


FIG. 4-12. Final design circuit for Example 4-2.

EXAMPLE 4-3

Design an impedance matching network that will block the flow of DC from the source to the load in Fig. 4-13. The frequency of operation is 75 MHz. Try the resonant approach.

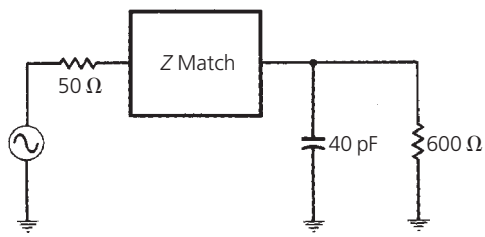


FIG. 4-13. Complex load circuit for Example 4-3.

Solution

The need to block the flow of DC from the source to the load dictates the use of the matching network of Fig. 4-4C. But, first, let's get rid of the stray 40-pF capacitor by resonating it with a shunt inductor at 75 MHz.

$$L = \frac{1}{\omega^2 C_{\text{stray}}} = \frac{1}{[2\pi(75 \times 10^6)^2(40 \times 10^{-12})]} = 112.6\,\text{nH}$$

This leaves us with the circuit shown in Fig. 4-14. Now that we have eliminated the stray capacitance, we can proceed with matching the network between the 50-ohm load and the apparent 600-ohm load. Thus,

$$Q_s = Q_p = \sqrt{\frac{R_p}{R_s} - 1} = \sqrt{\frac{600}{50} - 1} = 3.32$$

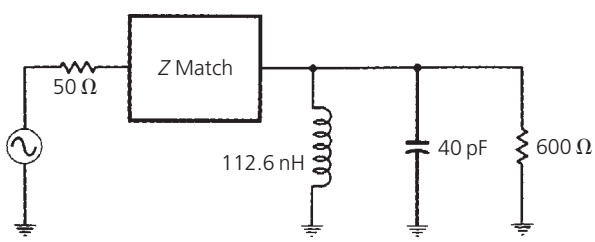


FIG. 4-14. Resonating the stray load capacitance.

$$X_s = Q_s R_s = (3.32)(50) = 166\,\text{ohms}$$

$$X_p = \frac{R_p}{Q_p} = \frac{600}{3.32} = 181\,\text{ohms}$$

Therefore, the element values are:

$$C = \frac{1}{\omega X_s} = \frac{1}{2\pi(75 \times 10^6)(166)} = 12.78\,\text{pF}$$

$$L = \frac{X_p}{\omega} = \frac{181}{2\pi(75 \times 10^6)} = 384\,\text{nH}$$

These values, then, yield the circuit of Fig. 4-15. But notice that this circuit can be further simplified by simply replacing the two shunt inductors with a single inductor. Therefore,

$$L_{\text{new}} = \frac{L_1 L_2}{L_1 + L_2} = \frac{(384)(112.6)}{384 + 112.6} = 87\,\text{nH}$$

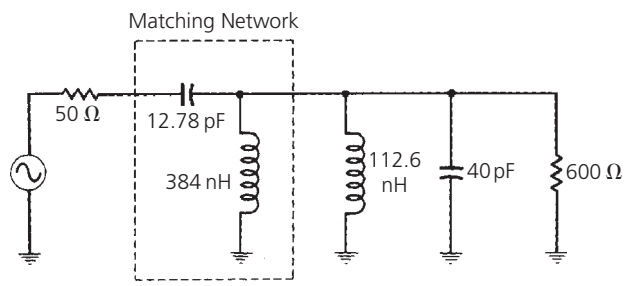


FIG. 4-15. The circuit of Fig. 4-14 after impedance matching.

The final circuit design appears in Fig. 4-16.

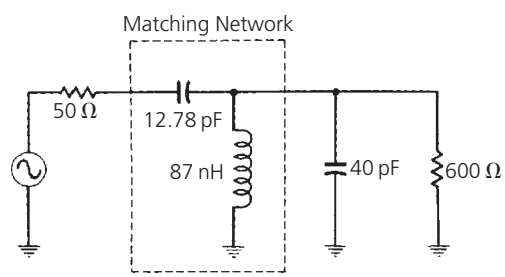


FIG. 4-16. Final design circuit for Example 4-3.

the junction between the two networks. This is illustrated in Fig. 4-19. The significance of the negative signs for $-X_{s1}$ and $-X_{s2}$ is symbolic. They are used merely to indicate that the X_s values are the opposite type of reactance from X_{p1} and X_{p2}, respectively. Thus, if X_{p1} is a capacitor, X_{s1} must be an inductor, and vice versa. Similarly, if X_{p2} is an inductor, X_{s2} must be a capacitor, and vice versa. They do *not* indicate negative reactances (capacitors).

The design of each section of the Pi network proceeds exactly as was done for the L networks in the previous sections. The virtual resistance (R) must be smaller than either R_s or R_L because it is connected to the series arm of each L section but, otherwise, it can be any value you wish. Most of the time, however, R is defined by the desired loaded Q of the circuit that you specify at the beginning of the design process. For our purposes, the loaded Q of this network will

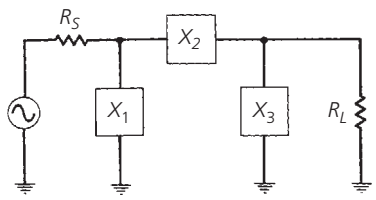


FIG. 4-17. The three-element Pi network.

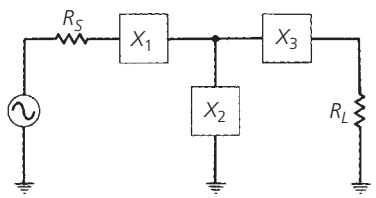


FIG. 4-18. The three-element T network.

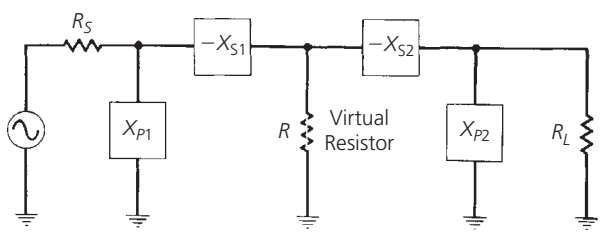


FIG. 4-19. The Pi network shown as two back-to-back L networks.

be defined as:

$$Q = \sqrt{\frac{R_H}{R} - 1} \qquad \text{(Eq. 4-4)}$$

where

R_H = the largest terminating impedance of R_s or R_L,
R = the virtual resistance.

Although this is not entirely accurate, it is a widely accepted Q-determining formula for this circuit, and is certainly close enough for most practical work. Example 4-4 illustrates the procedure.

Any of the networks in Fig. 4-21 will perform the impedance match between the 100-ohm source and the 1000-ohm load. The one that you choose for each particular application will depend on any number of factors including:

1. The elimination of stray reactances.
2. The need for harmonic filtering.
3. The need to pass or block DC voltage.

The T network

The design of the 3-element T network is exactly the same as for the Pi network except that with the T, you match the load and the source, through two L-type networks, to a virtual resistance that is *larger* than either the load or source resistance. This means that the two L-type networks will then have their shunt legs connected together as shown in Fig. 4-22.

The T network is often used to match two low-valued impedances when a high-Q arrangement is needed. The loaded Q of the T network is determined by the L section that has the highest Q. By definition, the L section with the highest Q will occur on the end with the *smallest* terminating resistor. Remember, too, that each terminating resistor is in the *series* leg of each network. Therefore, the formula for determining the loaded Q of the T network is:

$$Q = \sqrt{\frac{R}{R_{small}} - 1} \qquad \text{(Eq. 4-5)}$$

where

R = the virtual resistance,
R_{small} = the smallest terminating resistance.

This formula is exactly the same as the Q formula that was previously given for the Pi-type networks. However, since we have reversed or "flip-flopped" the L sections to produce the T network, we must also make sure that we redefine the Q formula to account for the new resistor placement, in relation to those L networks. In other words, Equations 4-4 and 4-5 are only special applications of the general formula that is given in Equation 4-1 (and repeated here for convenience).

$$Q = \sqrt{\frac{R_p}{R_s} - 1} \qquad \text{(Eq. 4-1)}$$

where

R_p = the resistance in the shunt branch of the L network,
R_s = the resistance in the series branch of the L network.

So, try not to get confused with the different definitions of circuit Q. They are all the same.

Each L network is calculated in exactly the same manner as was given in the previous examples and, as we shall soon see, we will also end up with four possible configurations for the T network (Example 4-5).

LOW-*Q* OR WIDEBAND MATCHING NETWORKS

Thus far in this chapter we have studied: (1) the L network, which has a circuit Q that is automatically defined when the source and load impedances are set, and (2) the Pi and T networks, which allow us to select a circuit Q independent of the source and load impedances *as long as the Q chosen is larger than that which is available with the L network*. This seems to indicate, and rightfully so, that the Pi and T networks are great for narrow-band matching networks. But what if an impedance match is required over a fairly broad range of frequencies? How do we handle that? The answer is to simply use two L sections in

EXAMPLE 4-4

Using Fig. 4-19 as a reference, design four different Pi networks to match a 100-ohm source to a 1000-ohm load. Each network must have a loaded Q of 15.

Solution

From Equation 4-4, we can find the virtual resistance we will be matching.

$$R = \frac{R_H}{Q^2 + 1} = \frac{1000}{226} = 4.42 \text{ ohms}$$

To find X_{p2} we have:

$$X_{p2} = \frac{R_p}{Q_p} = \frac{R_L}{Q} = \frac{1000}{15} = 66.7 \text{ ohms}$$

Similarly, to find X_{s2}:

$$X_{s2} = QR_{series} = 15\,(R) = (15)(4.42) = 66.3 \text{ ohms}$$

This completes the design of the L section on the *load* side of the network. Note that R_{series} in the above equation was substituted for the virtual resistor R, which by definition is in the series arm of the L section.

The Q for the other L network is now defined by the ratio of R_s to R, as per Equation 4-1, where:

$$Q_1 = \sqrt{\frac{R_s}{R} - 1} = \sqrt{\frac{100}{4.42} - 1} = 4.6$$

Notice here that the source resistor is now considered to be in the shunt leg of the L network. Therefore, R_s is defined as R_p, and

$$X_{p1} = \frac{R_p}{Q_1} = \frac{100}{4.6} = 21.7 \text{ ohms}$$

Similarly,

$$X_{s2} = Q_1 R_{series} = Q_1 R = (4.6)(4.46) = 20.51 \text{ ohms}$$

The actual network design is now complete and is shown in Fig. 4-20. Remember that the virtual resistor (R) is not really in the circuit and, therefore, is not shown. Reactances $-X_{s1}$ and $-X_{s2}$ are now in series and can simply be added together to form a single component.

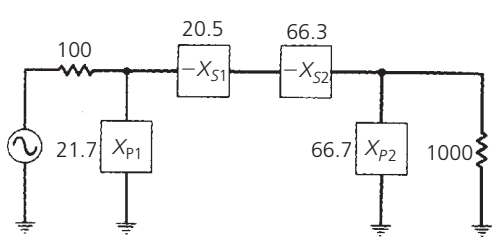


FIG. 4-20. Calculated reactances for Example 4-4.

So far in the design, we have dealt only with reactances and have not yet computed actual component values. This is because of the need to maintain a general design approach so that four final networks can be generated quickly as per the problem statement.

Notice that X_{p1}, X_{s1}, X_{p2}, and X_{s2} can all be either capacitive or inductive reactances. The only constraint is that X_{p1} and X_{s1} are of opposite types, and X_{p2} and X_{s2} are of opposite types. This yields the four networks of Fig. 4-21 (the source and load have been omitted). Each component in Fig. 4-21 is shown as a reactance (in ohms). Therefore, to perform the transformation from the dual-L to the Pi network, the two series components are merely added if they are alike, and subtracted if the reactances are of opposite type. The final step, of course, is to change each reactance into a component value of capacitance and inductance at the frequency of operation.

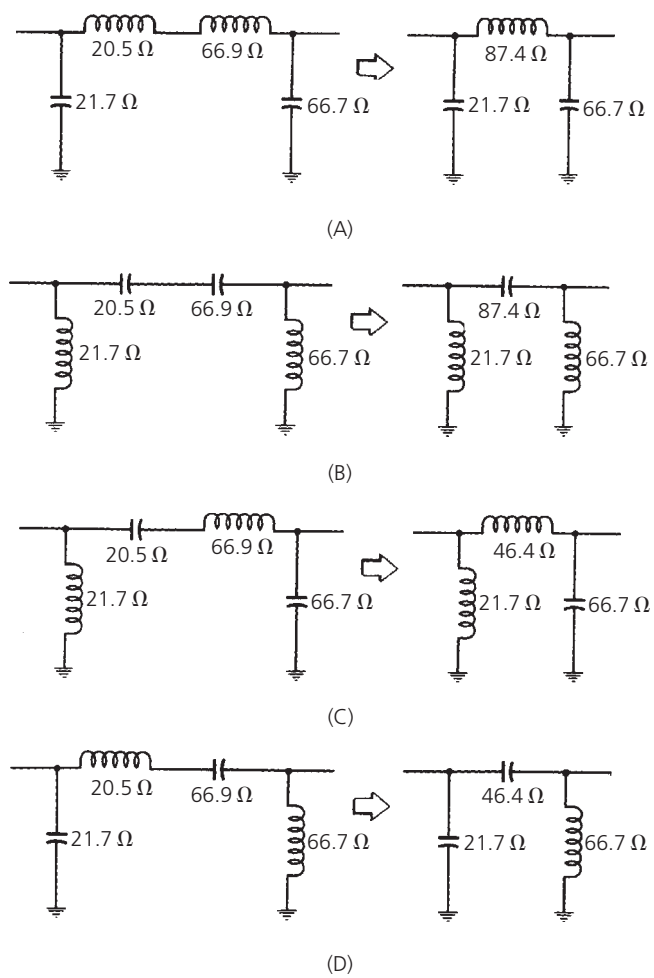


FIG. 4-21. The transformation from double-L to Pi networks.

still another configuration, as shown in Fig. 4-25. Notice here that the virtual resistor is in the shunt leg of one L section and in the series leg of the other L section. We, therefore, have two *series-connected* L sections rather than the back-to-back configuration of the Pi and T networks. In this new configuration, the value of the virtual resistor (R) must be larger than the smallest termination impedance and, also, smaller than the largest termination impedance. Of course, any virtual resistance that satisfies these criteria may be chosen. The net result is a range of loaded-*Q* values that is *less than* the range of *Q* values obtainable from either a single L section, or the Pi and T networks previously described.

EXAMPLE 4-5

Using Fig. 4-22 as a reference, design four different networks to match a 10-ohm source to a 50-ohm load. Each network is to have a loaded *Q* of 10.

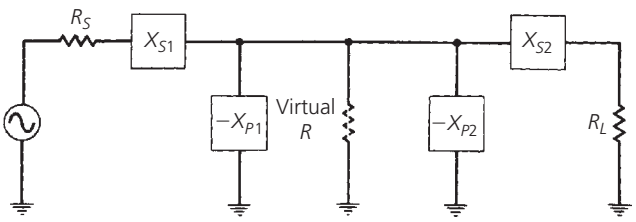


FIG.4-22. The T network shown as two back-to-back L networks.

Solution

Using Equation 4-5, we can find the virtual resistance we need for the match.

$$R = R_{\text{small}}(Q^2 + 1) = 10(101) = 1010\,\text{ohms}$$

From Equation 4-2:

$$X_{s1} = QR_s = 10(10) = 100\,\text{ohms}$$

From Equation 4-3:

$$X_{p1} = \frac{R}{Q} = \frac{1010}{10} = 101\,\text{ohms}$$

Now, for the L network on the load end, the *Q* is defined by the virtual resistor and the load resistor. Thus,

$$Q_2 = \sqrt{\frac{R}{R_L} - 1} = \sqrt{\frac{1010}{50} - 1} = 4.4$$

Therefore,

$$X_{p2} = \frac{R}{Q_2} = \frac{1010}{4.4} = 230\,\text{ohms}$$

$$X_{s2} = Q_2 R_L = (4.4)(50) = 220\,\text{ohms}$$

The network is now complete and is shown in Fig. 4-23 without the virtual resistor.

The two shunt reactances of Fig. 4-23 can again be combined to form a single element by simply substituting a value that is equal to the combined equivalent parallel reactance of the two.

The four possible T-type networks that can be used for matching the 10-ohm source to the 50-ohm load are shown in Fig. 4-24.

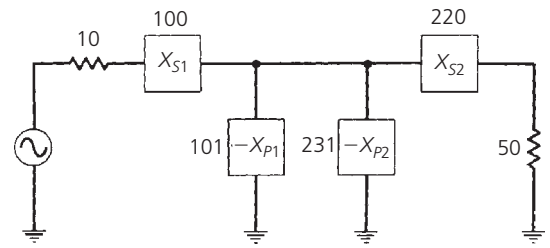


FIG. 4-23. The calculated reactances of Example 4-5.

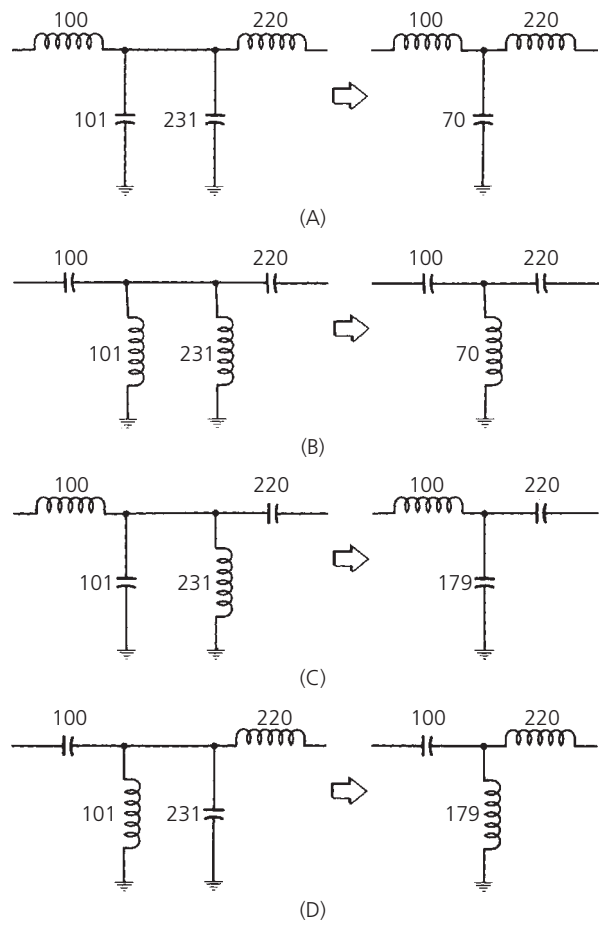


FIG. 4-24. The transformation of circuits from double-L to T-type networks.

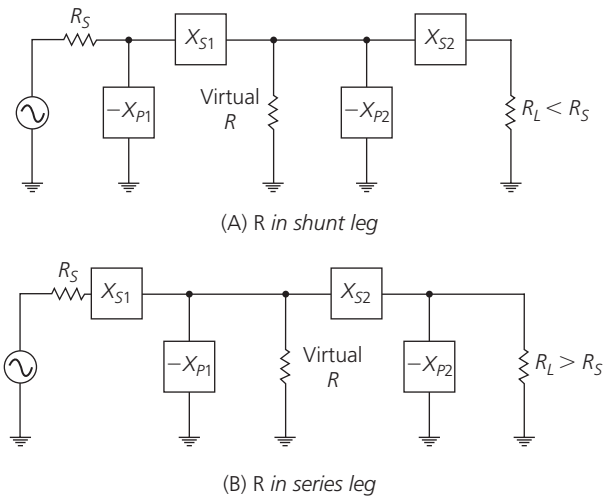


FIG. 4-25. Two series-connected L networks for lower Q applications.

The maximum bandwidth (minimum Q) available from this network is obtained when the virtual resistor (R) is made equal to the geometric mean of the two impedances being matched.

$$R = \sqrt{R_S R_L} \qquad \text{(Eq. 4-6)}$$

The loaded Q of the network, for our purposes, is defined as:

$$Q = \sqrt{\frac{R}{R_{\text{smaller}}} - 1} = \sqrt{\frac{R_{\text{larger}}}{R} - 1} \qquad \text{(Eq. 4-7)}$$

where

R = the virtual resistance,
R_{smaller} = the smallest terminating resistance,
R_{larger} = the largest terminating resistance.

If even wider bandwidths are needed, more L networks may be cascaded with virtual resistances between each network. Optimum bandwidths in these cases are obtained if the ratios of each of the two succeeding resistances are equal:

$$\frac{R_1}{R_{\text{smaller}}} = \frac{R_2}{R_1} = \frac{R_3}{R_2} \ldots = \frac{R_{\text{larger}}}{R_n} \qquad \text{(Eq. 4-8)}$$

where

R_{smaller} = the smallest terminating resistance,
R_{larger} = the largest terminating resistance,
$R_1, R_2, \ldots R_n$ = virtual resistors.

This is shown in Fig. 4-26.

The design procedure for these wideband matching networks is precisely the same as was given for the previous examples. To design for a specific low Q, simply solve Equation 4-7 for R to find the virtual resistance needed. Or, to design for an optimally wide bandwidth, solve Equation 4-6 for R. Once R is known, the design is straightforward

THE SMITH CHART

Perhaps one of the most useful graphical tools available to the RF circuit designer today is the Smith Chart, shown in Fig. 4-27.

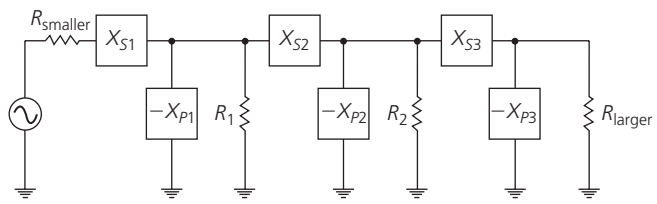


FIG. 4-26. Expanded version of Fig. 4-25 for even wider bandwidths.

The chart was originally conceived back in the 1930s by a Bell Laboratories engineer named Phillip Smith, who wanted an easier method of solving the tedious repetitive equations that often appear in RF theory. His solution, appropriately named the Smith Chart, is still widely in use.

At first glance, a Smith Chart appears to be quite complex. Indeed, why would anyone of sound mind even care to look at such a chart? The answer is really quite simple; once the Smith Chart and its uses are understood, the RF circuit designer's job becomes much less tedious and time consuming. Very lengthy complex equations can be solved graphically on the chart in seconds, thus lessening the possibility of errors creeping into the calculations.

Smith Chart Construction

The mathematics behind the construction of a Smith Chart are given here for those who are interested. It is important to note, however, that you do not *need* to know or understand the mathematics surrounding the actual construction of a chart as long as you understand what the chart represents and how it can be used to your advantage. Indeed, there are so many uses for the chart that an entire volume has been written on the subject. In this chapter, we will concentrate mainly on the Smith Chart as an impedance matching tool and other uses will be covered in later chapters. The mathematics follow.

The reflection coefficient of a load impedance when given a source impedance can be found by the formula:

$$\rho = \frac{Z_s - Z_L}{Z_s + Z_L} \qquad \text{(Step 1)}$$

In normalized form, this equation becomes:

$$\rho = \frac{Z_o - 1}{Z_o + 1} \qquad \text{(Step 2)}$$

where Z_o is a complex impedance of the form $R + jX$.

The polar form of the reflection coefficient can also be represented in rectangular coordinates:

$$\rho = p + jq$$

Substituting into Step 2, we have:

$$p + jq = \frac{R + jX - 1}{R + jX + 1} \qquad \text{(Step 3)}$$

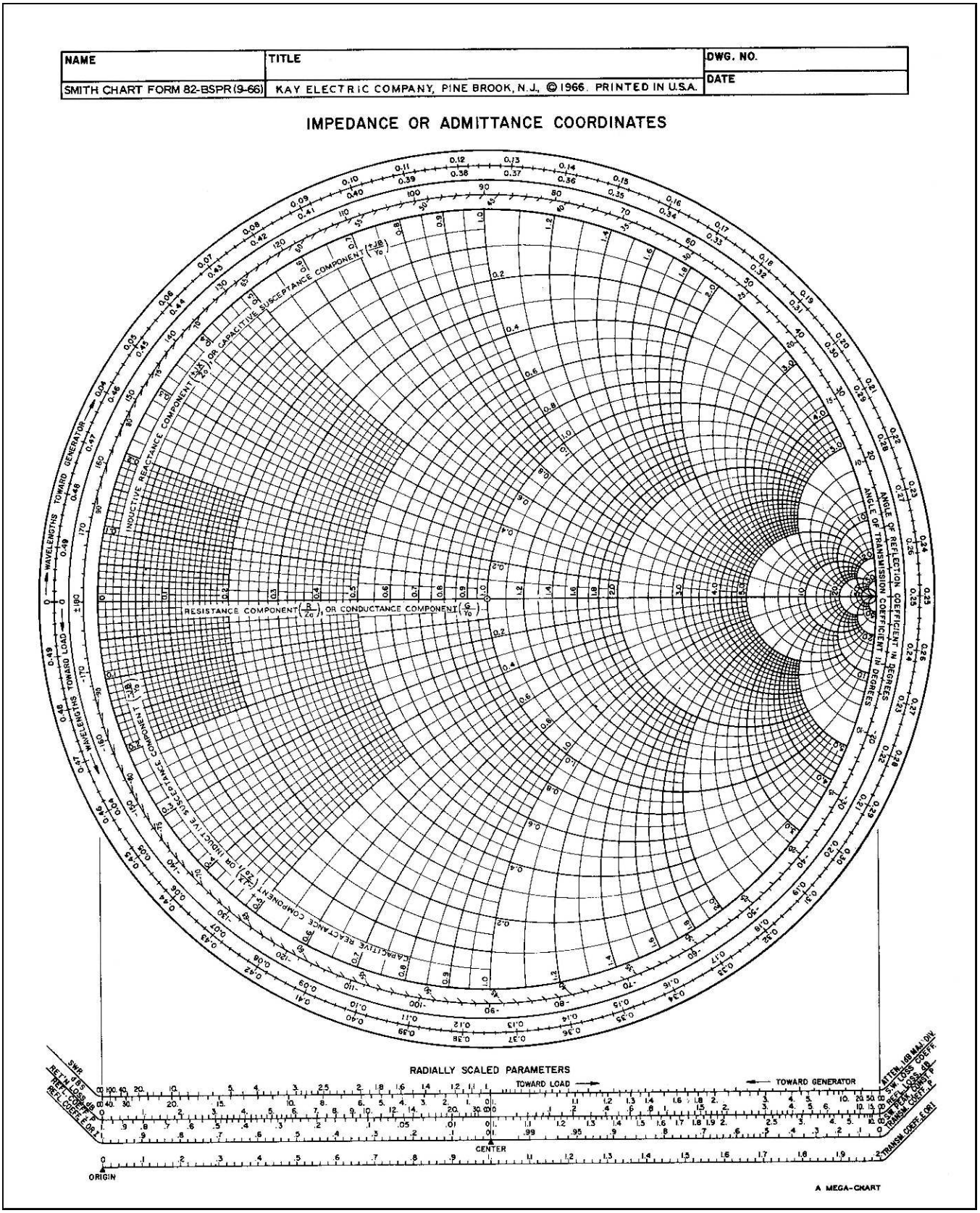


FIG. 4-27. The Smith Chart. (***Courtesy Analog Instruments Co.***) For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

If we solve for the real and imaginary parts of $p + jq$, we get:

$$p = \frac{R^2 - 1 + X^2}{(R+1)^2 + X^2} \qquad \text{(Step 4)}$$

and

$$q = \frac{2x}{(R+1)^2 + X^2} \qquad \text{(Step 5)}$$

Solve Step 5 for X:

$$X = \left(\frac{p(R+1)^2 - R^2 + 1}{1-p}\right)^{1/2} \qquad \text{(Step 6)}$$

Then, substitute Step 6 into Step 5 to obtain:

$$\left(p - \frac{R}{R+1}\right)^2 + q^2 = \left(\frac{1}{R+1}\right)^2 \qquad \text{(Step 7)}$$

Step 7 is the equation for a family of circles whose centers are at:

$$p = \frac{R}{R+1}$$
$$q = 0$$

and whose radii are equal to:

$$\frac{1}{R+1}$$

These are the constant resistance circles, some of which are shown in Fig. 4-28A.

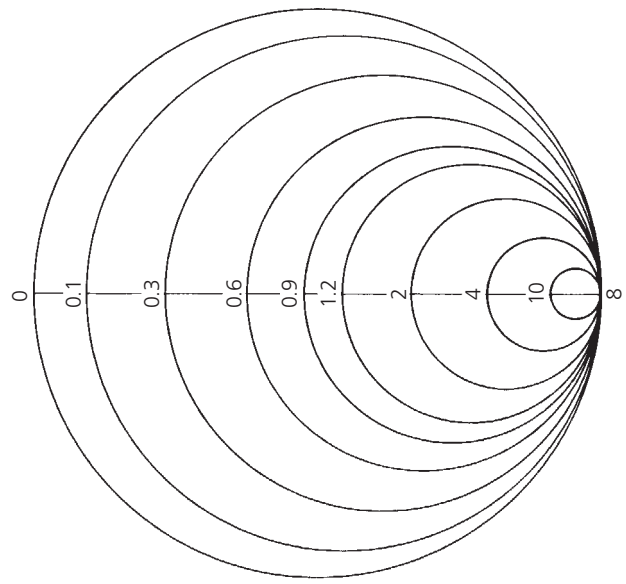
Similarly, we can eliminate R from Steps 4 and 5 to obtain:

$$(p-1)^2 + \left(q - \frac{1}{X}\right)^2 = \left(\frac{1}{X}\right)^2 \qquad \text{(Step 8)}$$

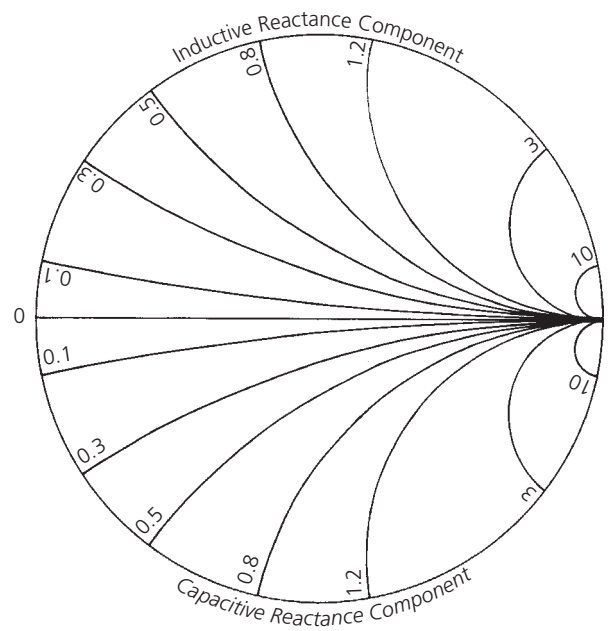
which represents a family of circles with centers at $p = 1$, $V = 1/X$, and radii of $1/X$. These circles are shown plotted on the p, jq axis in Fig. 4-28B.

As the preceding mathematics indicate, the Smith Chart is basically a combination of a family of circles and a family of arcs of circles, the centers and radii of which can be calculated using the equations given (Steps 1 through 8). Fig. 4-28 shows the chart broken down into these two families. The circles of Fig. 4-28A are known as *constant resistance circles*. Each point on a constant resistance circle has the same *resistance* as any other point on the circle. The arcs of circles shown in Fig. 4-28B are known as *constant reactance circles*, as each point on a circle has the same *reactance* as any other point on that circle. These circles are centered off of the chart and, therefore, only a small portion of each is contained within the boundary of the chart. All arcs above the centerline of the chart represent $+jX$, or inductive reactances, and all arcs below the centerline represent $-jX$, or capacitive reactances. The centerline must, therefore, represent an axis where $X = 0$ and is, therefore, called the *real axis.*

Notice in Fig. 4-28A that the "constant resistance = 0" circle defines the outer boundary of the chart. As the resistive component increases, the radius of each circle decreases and the center of each circle moves toward the right on the chart. Then, at infinite resistance, you end up with an infinitely small circle that is located at the extreme right-hand side of the chart. A similar thing happens for the constant reactance circles shown in Fig. 4-28B. As the magnitude of the reactive component increases ($-jX$ or $+jX$), the radius of each circle decreases, and the center of each circle moves closer and closer to the extreme right side of the chart. Infinite resistance and infinite reactance are thus represented by the same point on the chart.



(A) *Constant resistance circles*



(B) *Constant reactance circles*

FIG. 4-28. Smith Chart construction.

Since the outer boundary of the chart is defined as the "$R = 0$" circle, with higher values of R being contained within the chart, it follows then that any point outside of the chart must contain a negative resistance. The concept of negative resistance is useful in the study of oscillators and it is mentioned here only to state that the concept does exist, and if needed, the Smith Chart can be expanded to deal with it.

When the two charts of Fig. 4-28 are incorporated into a single version, the Smith Chart of Fig. 4-29 is born. If we add a few peripheral scales to aid us in other RF design tasks, such as determining *standing wave ratio (SWR), reflection coefficient*, and *transmission loss* along a transmission line, the basic chart of Fig. 4-27 is completed.

Basic Smith Chart Tips

When developing the Smith Chart, there are certain precautions that should be noted. These are among the most important:

- All the circles have one same, unique intersecting point at the coordinate (1, 0).
- The zero circle where there is no resistance ($R = 0$) is the largest one.
- The infinite resistor circle is reduced to one point at (1, 0).
- There should be no negative resistance. If one (or more) should occur, you will be faced with the possibility of oscillatory conditions.
- Another resistance value can be chosen by simply selecting another circle corresponding to the new value.

Plotting Impedance Values

Any point on the Smith Chart represents a *series* combination of resistance and reactance of the form $Z = R + jX$. Thus, to locate the impedance $Z = 1 + j1$, you would find the $R = 1$ constant resistance circle and follow it until it crossed the $X = 1$ constant reactance circle. The junction of these two circles would then represent the needed impedance value. This particular point, shown in Fig. 4-30, is located in the upper half of the chart because X is a positive reactance or an inductor. On the other hand, the point $1 - j1$ is located in the *lower* half of the chart because, in this instance, X is a negative quantity and represents a capacitor. Thus, the junction of the $R = 1$ constant resistance circle and the $X = -1$ constant reactance circle defines that point.

In general, then, to find any *series* impedance of the form $R \pm jX$ on a Smith Chart, you simply find the junction of the $R =$ constant and $X =$ constant circles. In many cases, the actual circles will not be present on the chart and you will have to interpolate between two that are shown. Thus, plotting impedances and, therefore, any manipulation of those impedances must be considered an inexact procedure which is subject to "pilot error." Most of the time, however, the error introduced by subjective judgements on the part of the user, in plotting impedances on the chart, is so small as to be negligible for practical work. Fig. 4-31 shows a few more impedances plotted on the chart.

Notice that all of the impedance values plotted in Fig. 4-31 are very small numbers. Indeed, if you try to plot an impedance of $Z = 100 + j150$ ohms, you will not be able to do it accurately because the $R = 100$ and $X = 150$ ohm circles would be (if they were drawn) on the extreme right edge of the chart—very close to infinity. In order to facilitate the plotting of larger impedances, *normalization* must be used. That is, each impedance to be plotted is divided by a convenient number that will place the new *normalized* impedance near the center of the chart where increased accuracy in plotting is obtained. Thus, for the preceding example, where $Z = 100 + j150$ ohms, it would be convenient to divide Z by 100, which yields the value $Z = 1 + j1.5$. This is very easily found on the chart. Once a chart is normalized in this manner, all impedances plotted on that chart *must be* divided by the *same* number in the normalization process. Otherwise, you will be left with a bunch of impedances with which nothing can be done.

Impedance Manipulation on the Chart

Fig. 4-32 graphically indicates what happens when a series capacitive reactance of $-j1.0$ ohm is added to an impedance of $Z = 0.5 + j0.7$ ohm. Mathematically, the result is

$$\begin{aligned} Z &= 0.5 + j0.7 - j1.0 \\ &= 0.5 - j0.3 \text{ ohms} \end{aligned}$$

which represents a series RC quantity. Graphically, what we have done is move *downward* along the $R = 0.5$-ohm constant resistance circle for a distance of $X = -j1.0$ ohm. This is the plotted impedance point of $Z = 0.5 - j0.3$ ohm, as shown. In a similar manner, as shown in Fig. 4-33, adding a series inductance to a plotted impedance value simply causes a move *upward* along a constant resistance circle to the new impedance value. This type of construction is very important in the design of impedance-matching networks using the Smith Chart and must be understood. In general then, the addition of a series capacitor to an impedance moves that impedance *downward* (counterclockwise) along a constant resistance circle for a distance that is equal to the reactance of the capacitor. The addition of any series inductor to a plotted impedance moves that impedance *upward* (clockwise) along a constant resistance circle for a distance that is equal to the reactance of the inductor.

Conversion of Impedance to Admittance

The Smith Chart, although described thus far as a family of impedance coordinates, can easily be used to convert any impedance (Z) to an admittance (Y), and vice versa. This can be accomplished by simply flipping the Smith Chart over. Note that if both the impedance and admittance charts are plotted together, overlaid, one upon the other, the new chart is called an immittance chart. While this may sound complicated, it can be extremely useful in designing match networks with components like series or shunt inductors and capacitors.

As previously pointed out, a series inductor, when added to a load, causes a rotation clockwise along a circle of constant resistance on the chart, while a shunt inductor causes rotation

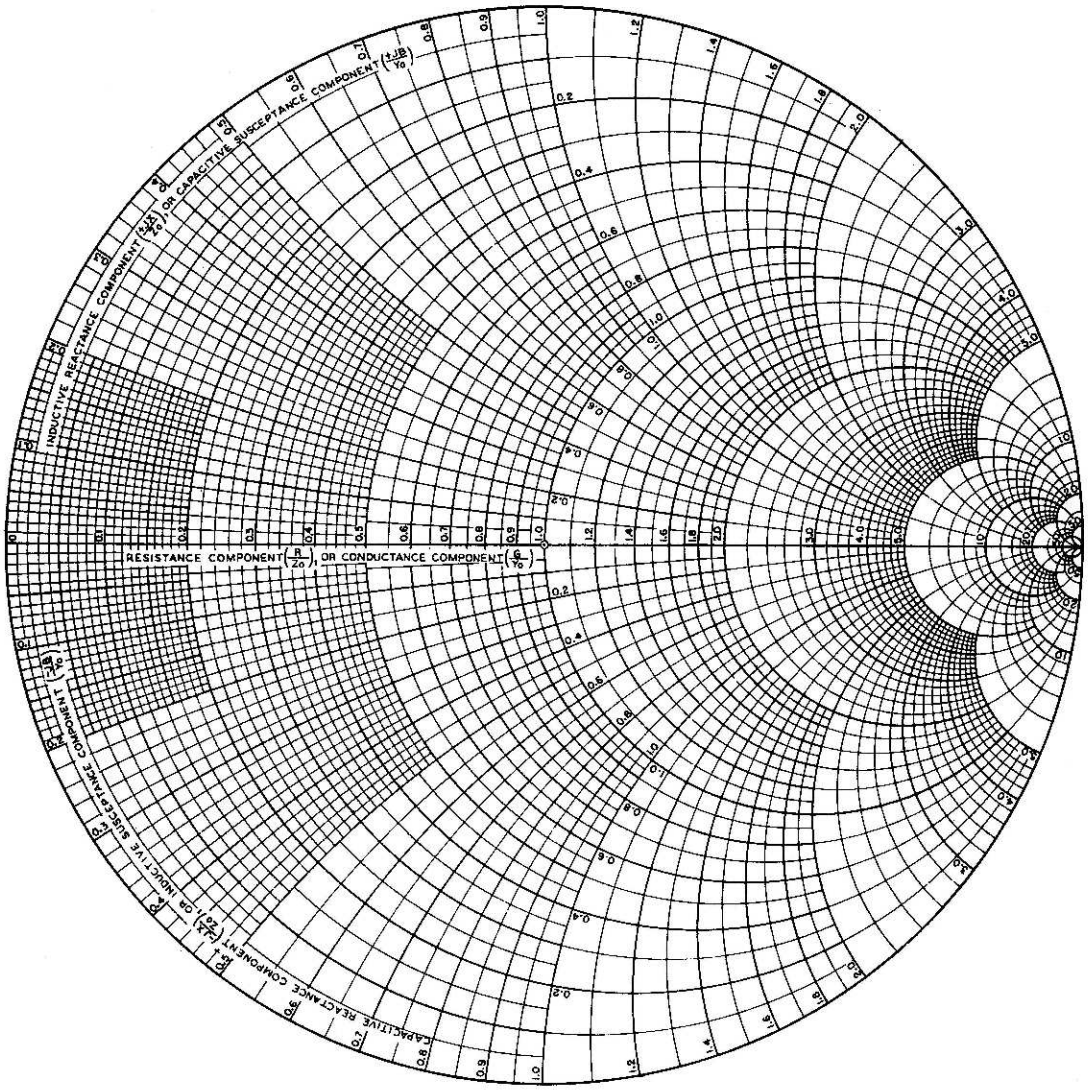


FIG. 4-29. The basic Smith Chart. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

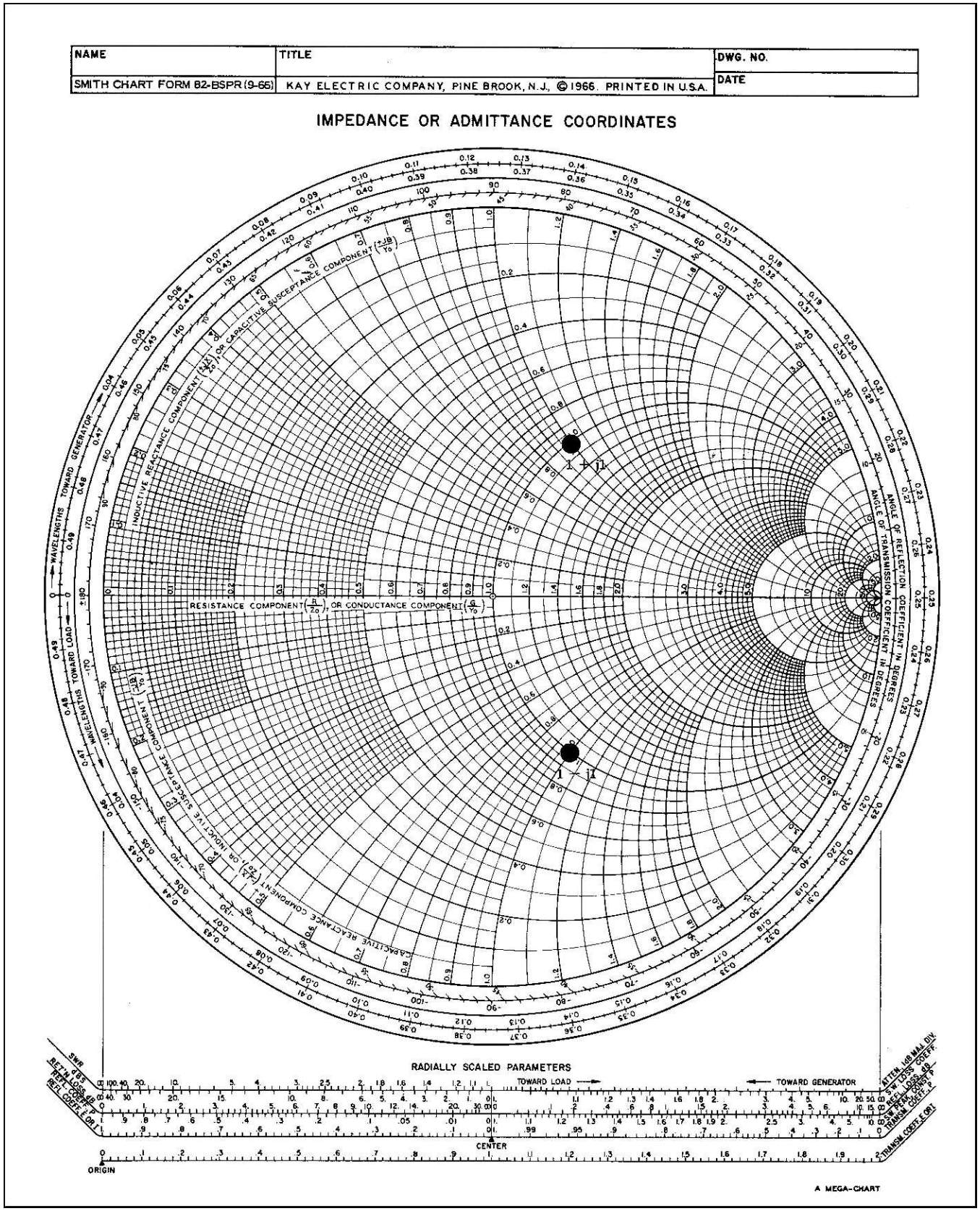


FIG. 4-30. Plotting impedances on the chart. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

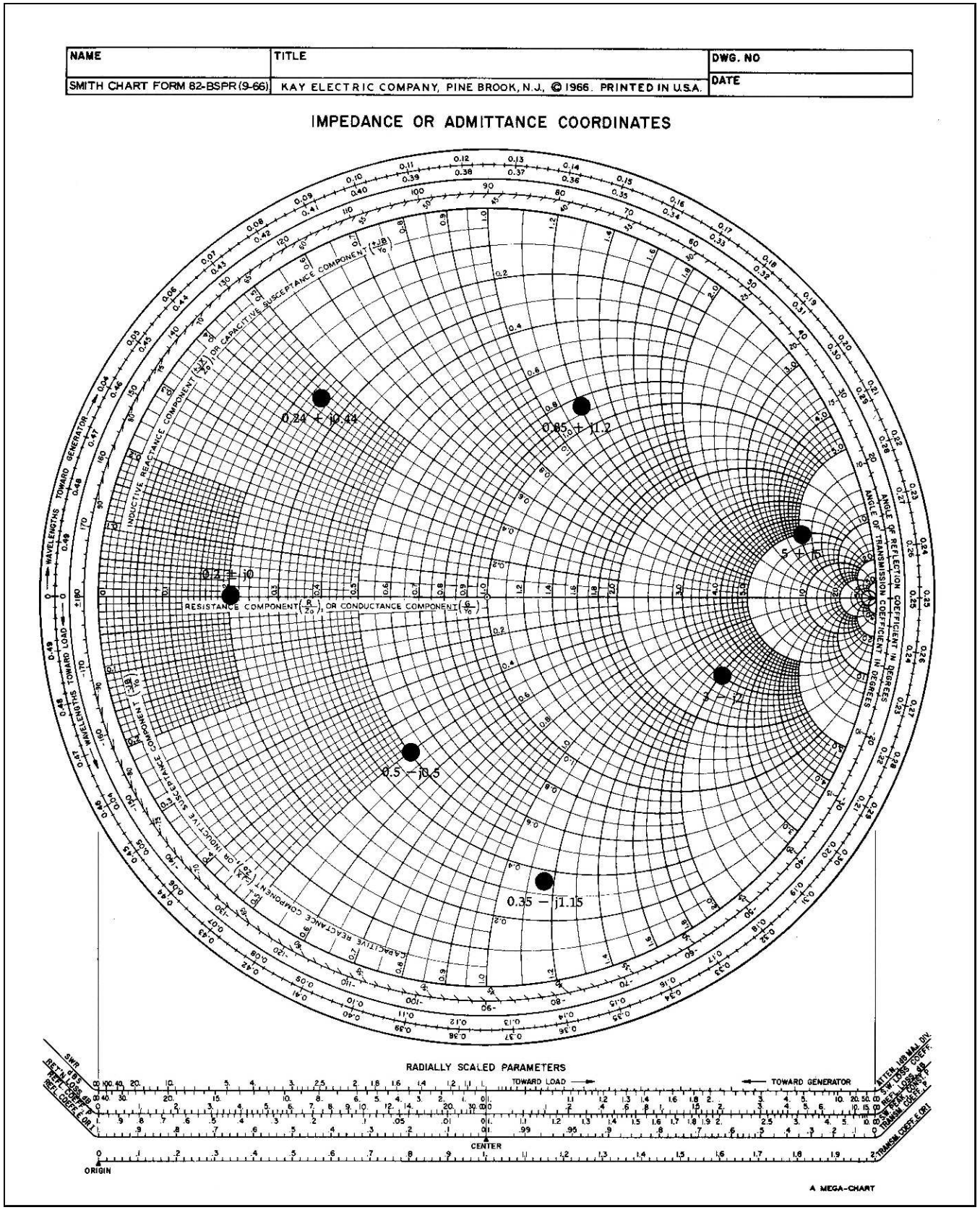


FIG. 4-31. More impedances are plotted on the chart. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

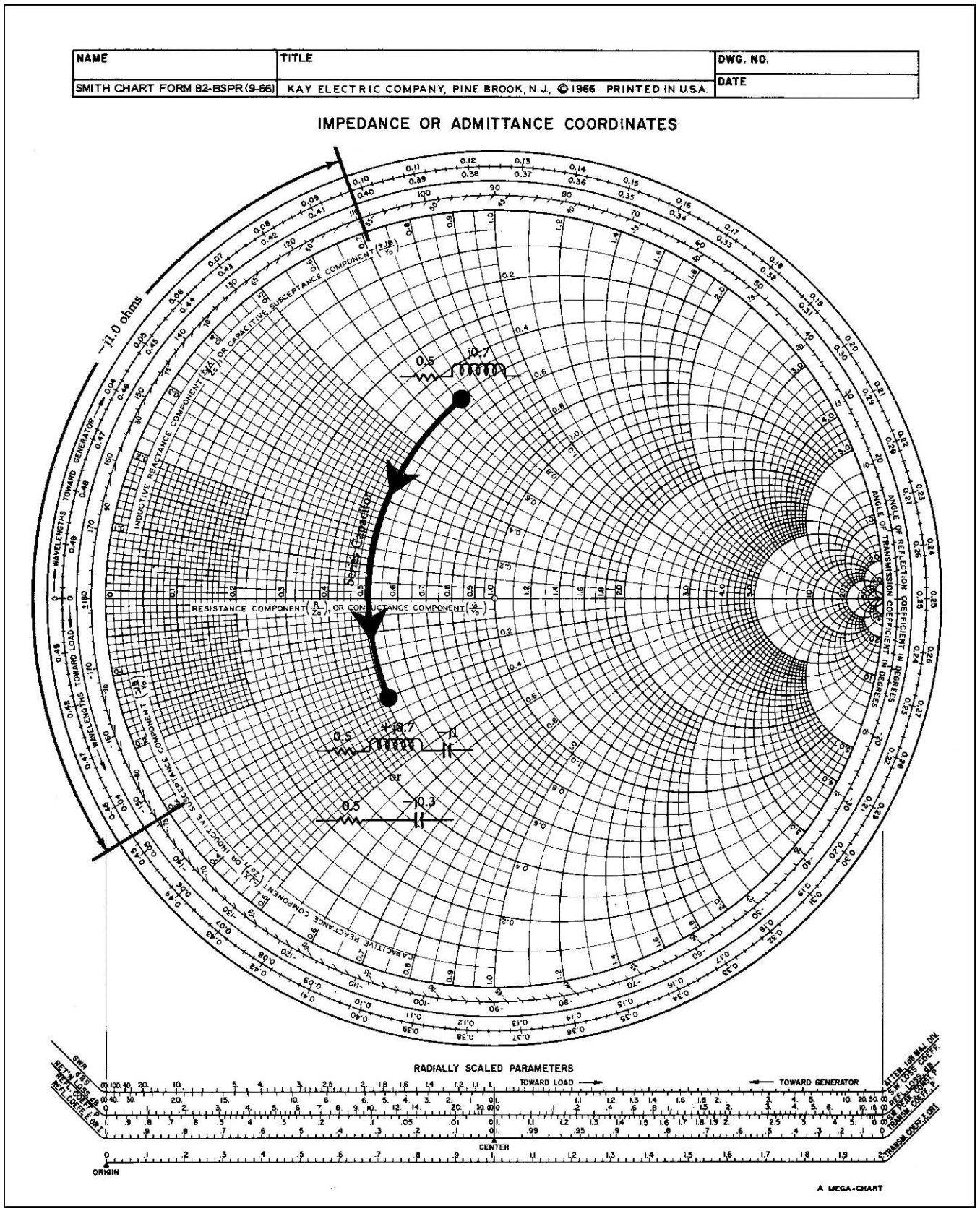


FIG. 4-32. Addition of a series capacitor. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

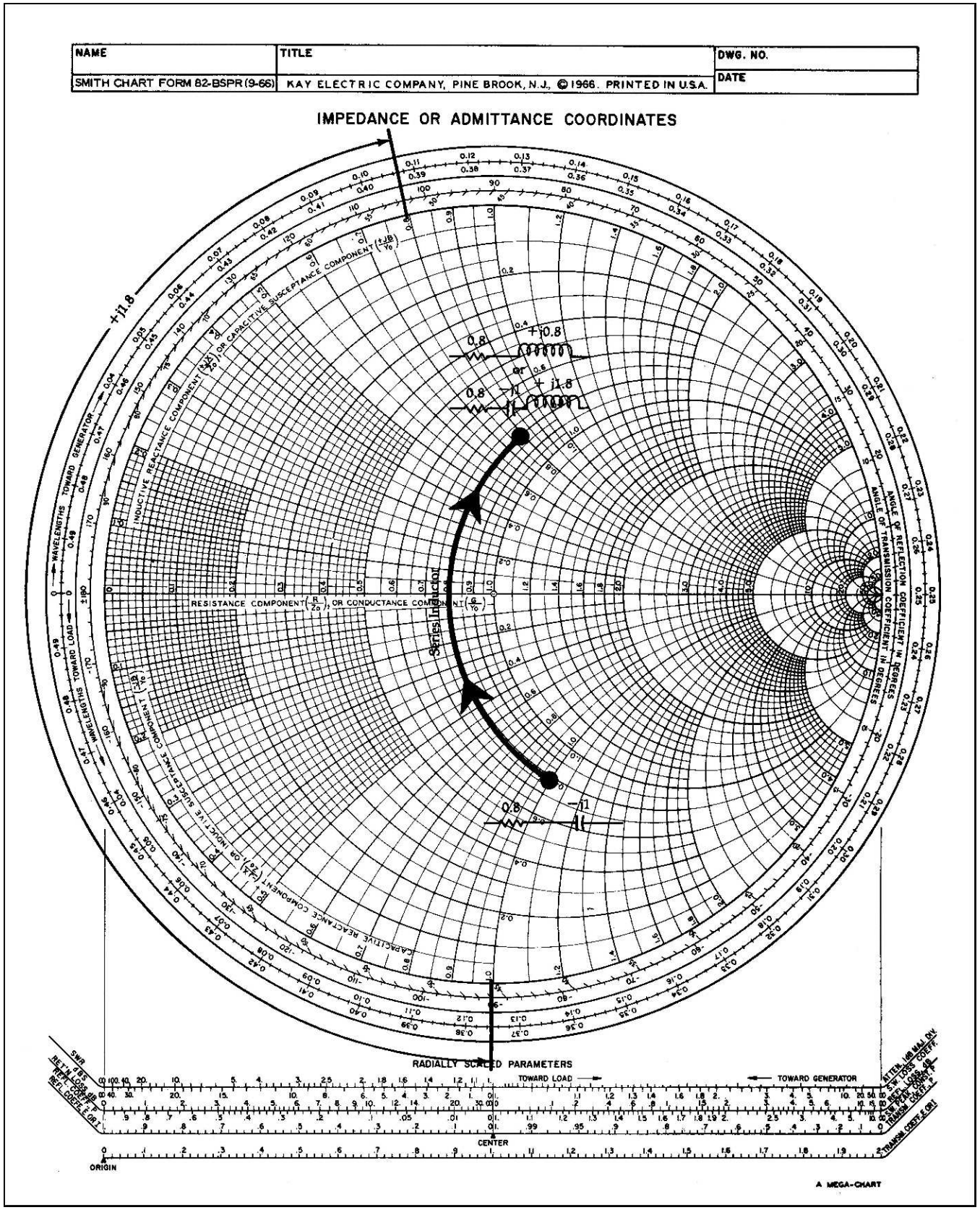


FIG. 4-33. Addition of a series inductor. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

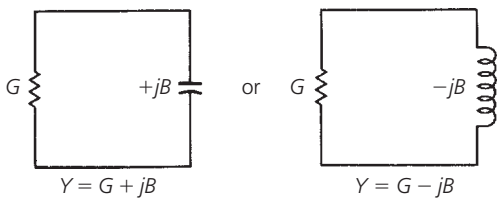


FIG. 4-34. Circuit representation for admittance.

counter-clockwise along a circle of constant admittance. In a similar manner, a series capacitor, added to a load, causes rotation counter-clockwise along a circle of constant resistance, while a shunt capacitor causes rotation clockwise along a circle of constant admittance.

In mathematical terms, an admittance is simply the inverse of an impedance, or

$$Y = \frac{1}{Z} \qquad \text{(Eq. 4-9)}$$

where the admittance (Y) contains both a real and an imaginary part, similar to the impedance (Z). Thus,

$$Y = G \pm jB \qquad \text{(Eq. 4-10)}$$

where

$G =$ the conductance in mhos,
$B =$ the susceptance in mhos.

The circuit representation is shown in Fig. 4-34. Notice that the *susceptance is positive for a capacitor and negative for an inductor*, whereas, for reactance, the opposite is true.

To find the inverse of a series impedance of the form $Z = R + jX$ mathematically, you would simply use Equation 4-9 and perform the resulting calculation. But, how can you use the Smith Chart to perform the calculation for you without the need for a calculator? The easiest way of describing the use of the chart in performing this function is to first work a problem out mathematically and, then, plot the results on the chart to see how the two functions are related. Take, for example, the series impedance $Z = 1 + j1$. The inverse of Z is:

$$\begin{aligned} Y &= \frac{1}{1 + j1} \\ &= \frac{1}{1.414\angle 45^\circ} \\ &= 0.7071\angle -45^\circ \\ &= 0.5 - j0.5 \text{ mho} \end{aligned}$$

If we plot the points $1 + j1$ and $0.5 - j0.5$ on the Smith Chart, we can easily see the graphical relationship between the two. This construction is shown in Fig. 4-35. Notice that the two points are located at exactly the same distance (d) from the center of the chart but in opposite directions (180°) from each other. Indeed, the same relationship holds true for *any* impedance and its inverse. Therefore, without the aid of a calculator, you can find the reciprocal of an impedance or an admittance by simply plotting the point on the chart, measuring the distance (d) from the center of the chart to that point, and, then, plotting the measured result the same distance from the center but in the opposite direction (180°) from the original point. This is a very simple construction technique that can be done in seconds.

Another approach that we could take to achieve the same result involves the manipulation of the actual chart rather than the performing of a construction on the chart. For instance, rather than locating a point 180° away from our original starting point, why not just rotate the chart itself 180° while fixing the starting point in space? The result is the same, and it can be read directly off of the rotated chart without performing a single construction. This is shown in Fig. 4-36 (Smith Chart Form ZY-01-N)* where the rotated chart is shown in black. Notice that the impedance plotted (solid lines on the red coordinates) is located at $Z = 1 + j1$ ohms, and the reciprocal of that (the admittance) is shown by dotted lines on the black coordinates as $Y = 0.5 - j0.5$. Keep in mind that because we have rotated the chart 180° to obtain the admittance coordinates, the upper half of the admittance chart represents *negative susceptance* ($-jB$) which is *inductive*, while the lower half of the admittance chart represents a *positive susceptance* ($+jB$) which is *capacitive*. Therefore, nothing has been lost in the rotation process.

The chart shown in Fig. 4-36, containing the superimposed impedance and admittance coordinates, is an extremely useful version of the Smith Chart and is the one that we will use throughout the remainder of the book. But first, let's take a closer look at the admittance coordinates alone.

Admittance Manipulation on the Chart

Just as the impedance coordinates of Figs. 4-32 and 4-33 were used to obtain a visual indication of what occurs when a *series* reactance is added to an *impedance*, the admittance coordinates provide a visual indication of what occurs when a *shunt* element is added to an *admittance*. The addition of a shunt capacitor is shown in Fig. 4-37. Here we begin with an admittance of $Y = 0.2 - j0.5$ mho and add a shunt capacitor with a susceptance (reciprocal of reactance) of $+j0.8$ mho. Mathematically, we know that parallel susceptances are simply added together to find the equivalent susceptance. When this is done, the result becomes:

$$\begin{aligned} Y &= 0.2 - j0.5 + j0.8 \\ &= 0.2 + j0.3 \text{ mho} \end{aligned}$$

If this point is plotted on the admittance chart, we quickly recognize that all we have done is to move along a constant conductance circle (G) *downward* (clockwise) a distance of $jB = 0.8$ mho. In other words, the real part of the admittance has not changed, only the imaginary part has. Similarly, as Fig. 4-38

*Smith Chart Form ZY-01-N is a copyright of Analog Instruments Company, P.O. Box 808, New Providence, NJ 07974. It and other Smith Chart accessories are available from the company.

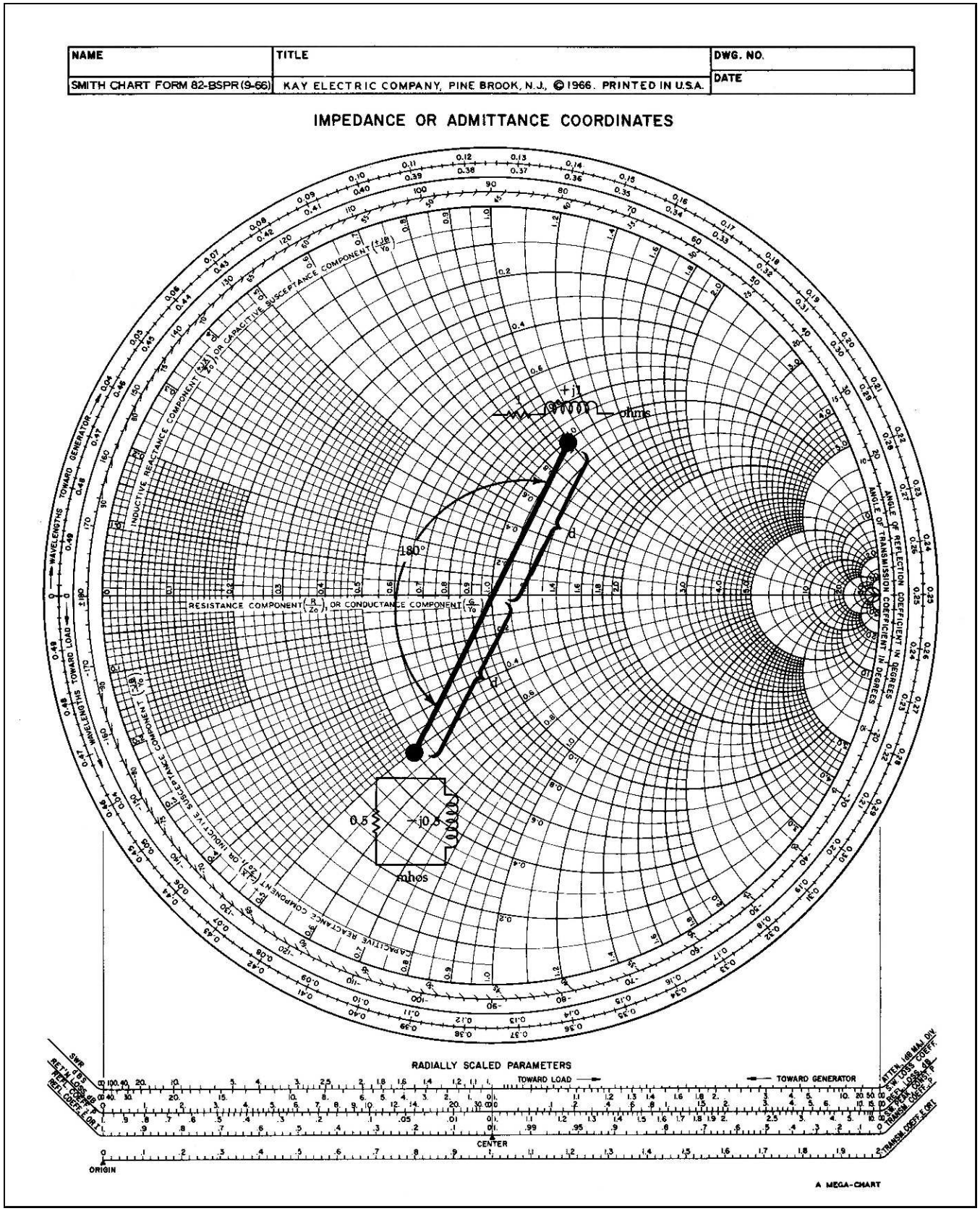


FIG. 4-35. Impedance-admittance conversion on the Smith Chart. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

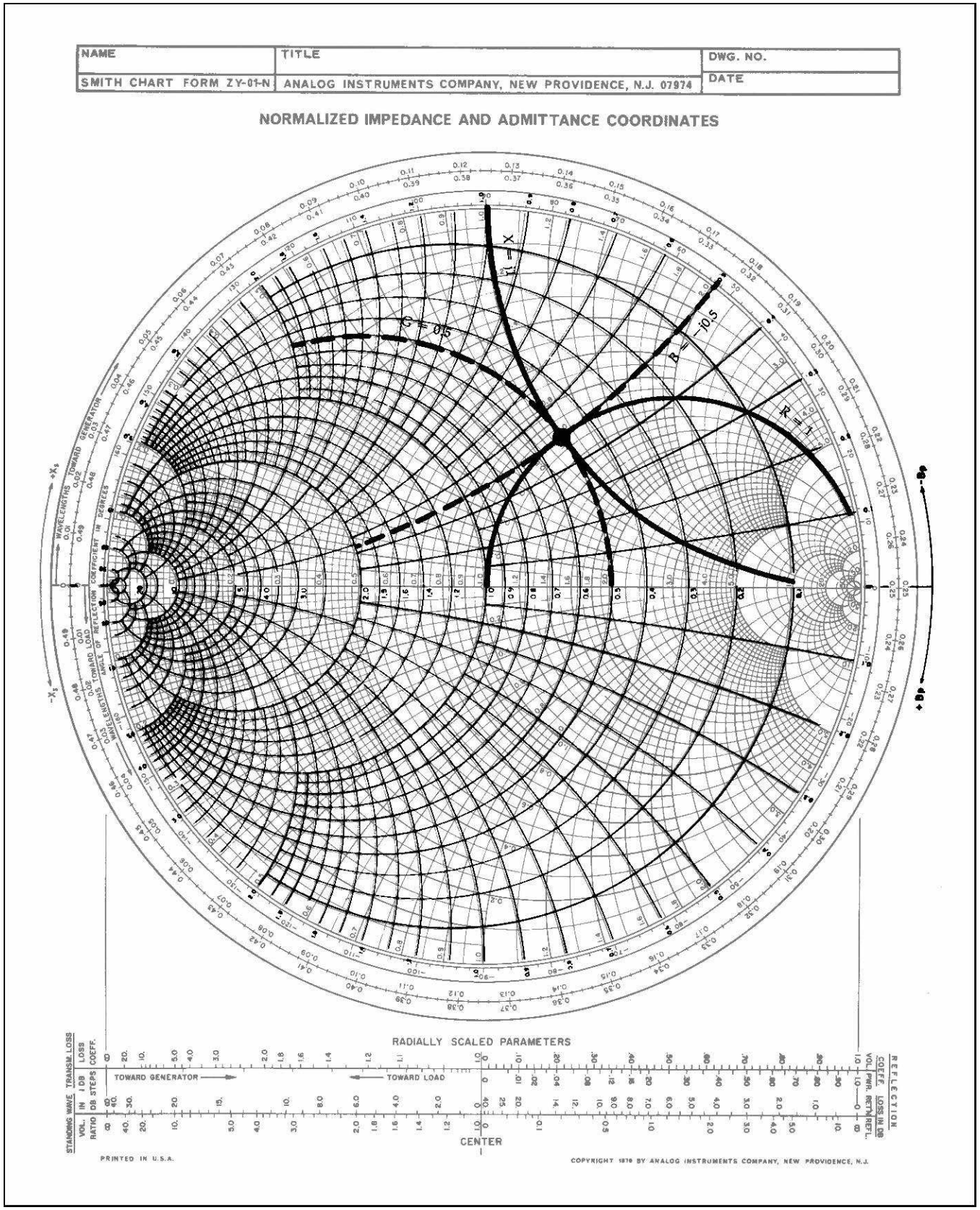


FIG. 4-36. Superimposed admittance coordinates. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

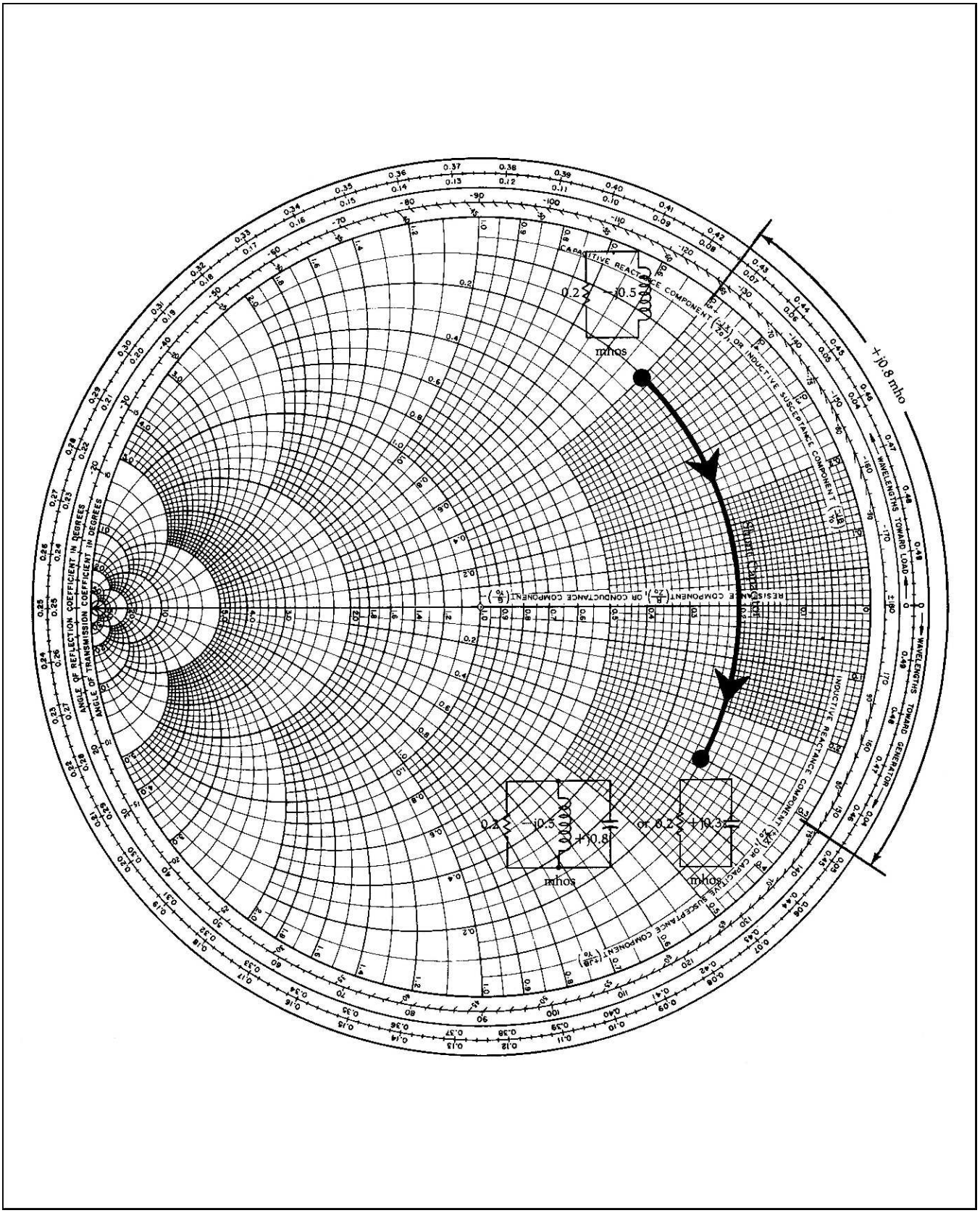


FIG. 4-37. Addition of a shunt capacitor. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

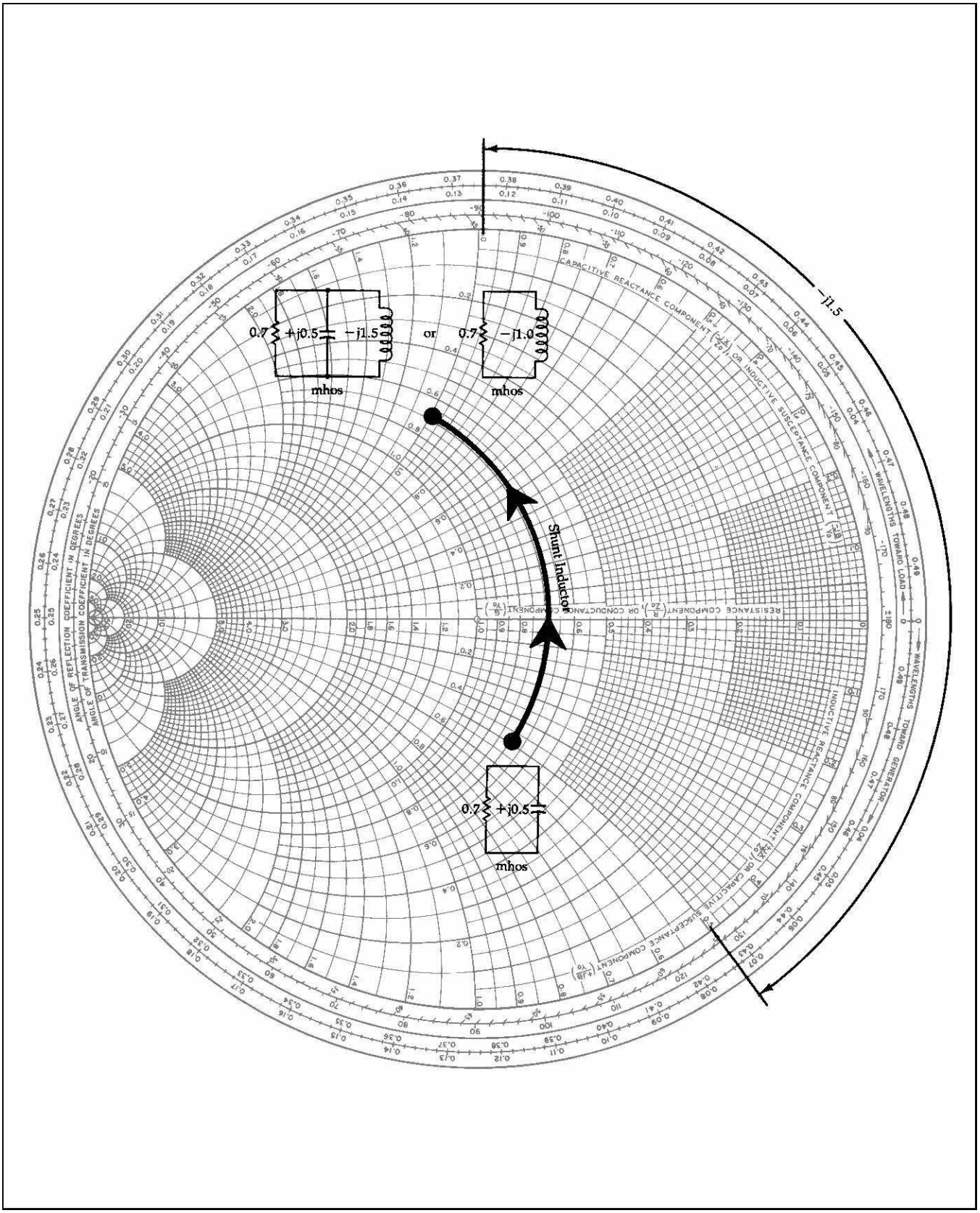


FIG. 4-38. Addition of a shunt inductor. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

indicates, adding a shunt inductor to an admittance moves the point along a constant conductance circle upward (counterclockwise) a distance ($-jB$) equal to the value of its susceptance.

If we again superimpose the impedance and admittance coordinates and combine Figs. 4-32, 4-33, 4-37, and 4-38 for the general case, we obtain the useful chart shown in Fig. 4-39. This chart graphically illustrates the direction of travel, along the impedance and admittance coordinates, which results when the particular type of component that is indicated is added to an existing impedance or admittance. A simple example should illustrate the point (Example 4-6).

IMPEDANCE MATCHING ON THE SMITH CHART

Because of the ease with which series and shunt components can be added in ladder-type arrangements on the Smith Chart, while easily keeping track of the impedance as seen at the input terminals of the structure, the chart seems to be an excellent candidate for an impedance-matching tool. The idea here is simple. Given a load impedance and given the impedance that the source would like to see, simply plot the load impedance and, then, begin adding series and shunt elements on the chart until the desired impedance is achieved–just as was done in Example 4-6.

Two-Element Matching

Two-element matching networks are mathematically very easy to design using the formulas provided in earlier sections of this chapter. For the purpose of illustration, however, let's begin our study of a Smith Chart impedance-matching procedure with the simple network given in Example 4-7.

To make life much easier for you as a Smith Chart user, the following equations may be used. For a series-C component:

$$C = \frac{1}{\omega XN} \qquad \text{(Eq. 4-11)}$$

For a series-L component:

$$L = \frac{XN}{\omega} \qquad \text{(Eq. 4-12)}$$

For a shunt-C component:

$$C = \frac{B}{\omega N} \qquad \text{(Eq. 4-13)}$$

For a shunt-L component:

$$L = \frac{N}{\omega B} \qquad \text{(Eq. 4-14)}$$

where

$\omega = 2\pi f$,
$X =$ the reactance as read from the chart,
$B =$ the susceptance as read from the chart,
$N =$ the number used to normalize the original impedances that are to be matched.

If you use the preceding equations, you will never have to worry about changing susceptances into reactances before unnormalizing the impedances. The equations take care of both operations. The only thing you have to do is read the value of susceptance (for shunt components) or reactance (for series components) directly off of the chart, plug this value into the equation used, and wait for your actual component values to pop out.

Three-Element Matching

In earlier sections of this chapter, you learned that the only real difference between two-element and three-element matching is that with three-element matching, you are able to choose the loaded Q for the network. That was easy enough to do in a mathematical-design approach due to the virtual resistance concept. But how can circuit Q be represented on a Smith Chart?

As you have seen before, in earlier chapters, the Q of a series-impedance circuit is simply equal to the ratio of its reactance to its resistance. Thus, any point on a Smith Chart has a Q associated with it. Alternately, if you were to specify a certain Q, you could find an infinite number of points on the chart that could satisfy that Q requirement. For example, the following impedances located on a Smith Chart have a Q of 5:

$$\begin{aligned} R + jX &= 1 \pm j5 \\ &= 0.5 \pm j2.5 \\ &= 0.2 \pm j1 \\ &= 0.1 \pm j0.5 \\ &= 0.05 \pm j0.25 \end{aligned}$$

These values are plotted in Fig. 4-45 and form the arcs shown. Thus, any impedance located on these arcs must have a Q of 5. Similar arcs for other values of Q can be drawn with the arc of infinite Q being located along the perimeter of the chart and the $Q = 0$ arc (actually a straight line) lying along the pure resistance line located at the center of the chart.

The design of high-Q three-element matching networks on a Smith Chart is approached in much the same manner as in the mathematical methods presented earlier in this chapter. Namely, one branch of the network will determine the loaded Q of the circuit, and it is this branch that will set the characteristics of the rest of the circuit.

The procedure for designing a three-element impedance- matching network for a specified Q is summarized as follows:

1. Plot the constant-Q arcs for the specified Q.
2. Plot the load impedance and the complex conjugate of the source impedance.
3. Determine the end of the network that will be used to establish the loaded Q of the design. For T networks, the end with the *smaller* terminating resistance determines the Q. For Pi networks, the end with the *larger* terminating resistor sets the Q.
4. For T networks:

$$R_s > R_L$$

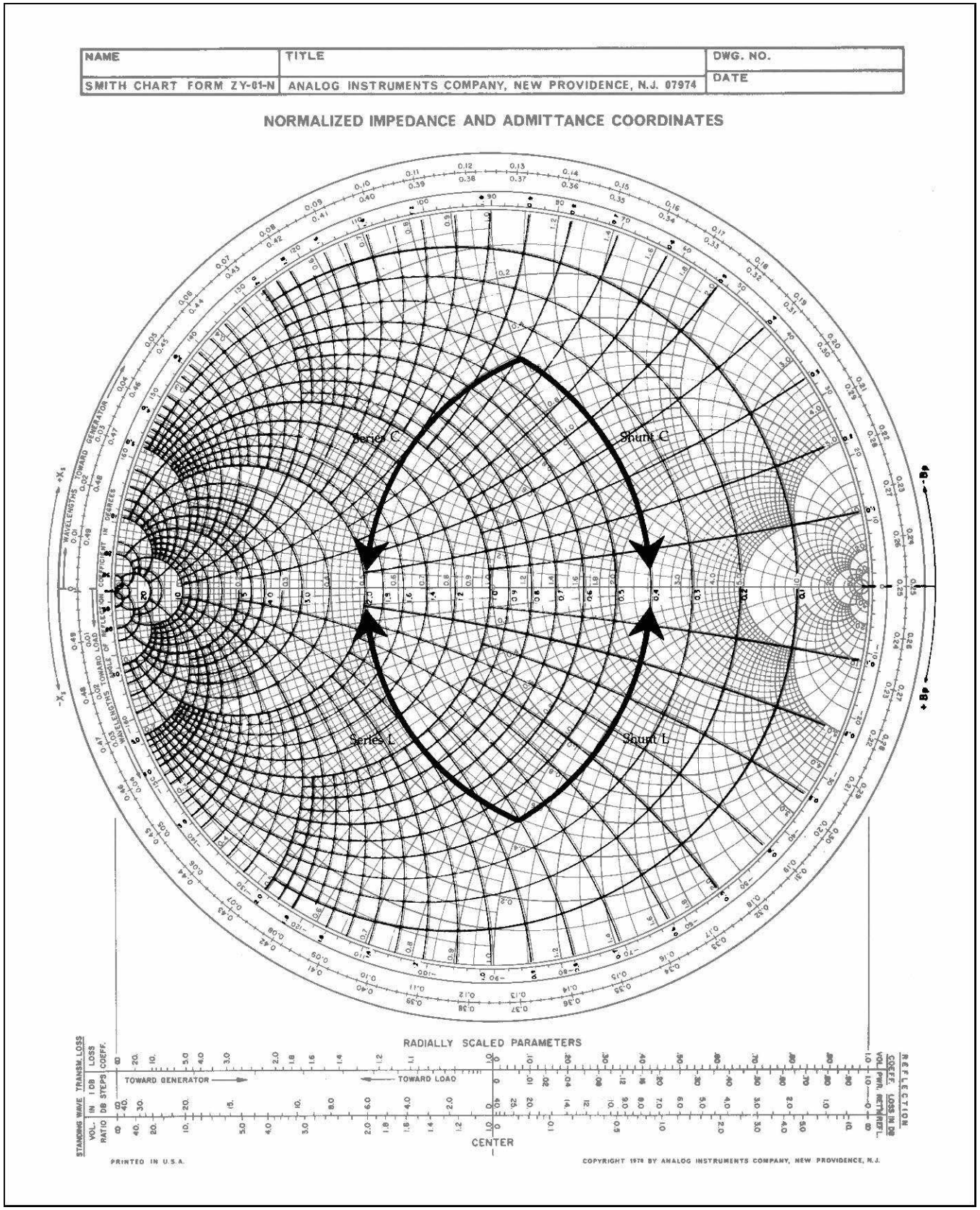


FIG. 4-39. Summary of component addition on a Smith Chart. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

Move from the load along a constant-R circle (series element) and intersect the Q curve. The length of this move determines your first element. Then, proceed from this point to Z_s^* ($Z_s^* = ZB$ conjugate) in two moves–first with a shunt and, then, with a series element.

$$R_s < R_L$$

Find the intersection (I) of the Q curve and the source impedance's $R =$ constant circle, and plot that point. Move *from the load impedance* to point I with two elements—first, a series element and, then, a shunt element. Move from point I to Z_s^* along the $R =$ constant circle with another series element.

3. For Pi networks:

$$R_s > R_L$$

Find the intersection (I) of the Q curve and the source impedance's $G =$ constant circle, and plot that point. Move from the load impedance to point I with two elements—first, a shunt element and, then, a series element. Move from point I to Z_s^* along the $G =$ constant circle with another shunt element.

$$R_s < R_L$$

Move from the load along a constant G circle (shunt element) and intersect the Q curve. The length of this move determines your first element. Then, proceed from this point to Z_s^* in two moves—first, with a series element and, then, with a shunt element.

The above procedures might seem complicated to the neophyte but remember that we are only forcing the constant-resistance or constant-conductance arc, located between the Q-determining termination and the specified-Q curve, to be one of our matching elements. An example may help to clarify matters (Example 4-8).

Multi-Element Matching

In multi-element matching networks where there is no Q constraint, the Smith Chart becomes a veritable treasure trove containing an infinite number of possible solutions. To get from point A to point B on a Smith Chart, there is, of course, an optimum solution. However, the optimum solution is not the only solution. The two-element network gets you from point A to point B with the least number of components and the three-element network can provide a specified Q by following a different route. If you do not care about Q, however, there are 3-, 4-, 5-, 10-, and 20-element (and more) impedance-matching networks that are easily designed on a Smith Chart by simply following the constant-conductance and constant-resistance circles until you eventually arrive at point B, which, in our case, is usually the complex conjugate of the source impedance. Fig. 4-48 illustrates this point. In the lower right-hand corner of the chart is point A. In the upper left-hand corner is point B. Three of the infinite number of possible solutions that can be used to get from point A to point B, by adding series and shunt inductances and capacitances, are

EXAMPLE 4-6

What is the impedance looking into the network shown in Fig. 4-40? Note that the task has been simplified due to the fact that shunt susceptances are shown rather than shunt reactances.

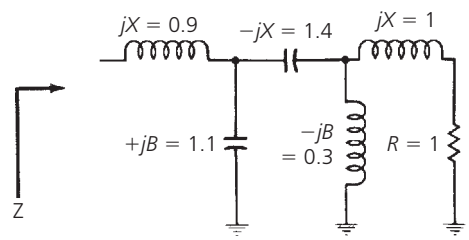


FIG. 4-40. Circuit for Example 4-6.

Solution

This problem is very easily handled on a Smith Chart and not a single calculation needs to be performed. The solution is shown in Fig. 4-42. It is accomplished as follows.

First, break the circuit down into individual branches as shown in Fig. 4-41. Plot the impedance of the series RL branch where $Z = 1 + j1$ ohm. This is point A in Fig. 4-42. Next, following the rules diagrammed in Fig. 4-39, begin adding each component back into the circuit, one at a time. Thus, the following constructions (Fig. 4-42) should be noted:

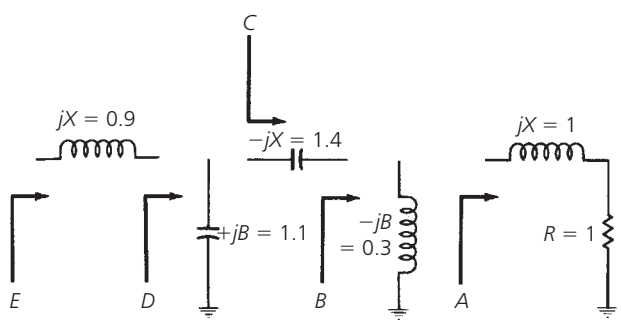


FIG. 4-41. Circuit is broken down into individual branch elements.

Arc AB = shunt L = $-jB = 0.3$ mho
Arc BC = series C = $-jX = 1.4$ ohms
Arc CD = shunt C = $+jB = 1.1$ mhos
Arc DE = series L = $+jX = 0.9$ ohm

The impedance at point E (Fig. 4-42) can then be read directly off of the chart as $Z = 0.2 + j0.5$ ohm.

Continued on next page

EXAMPLE 4-6—Cont

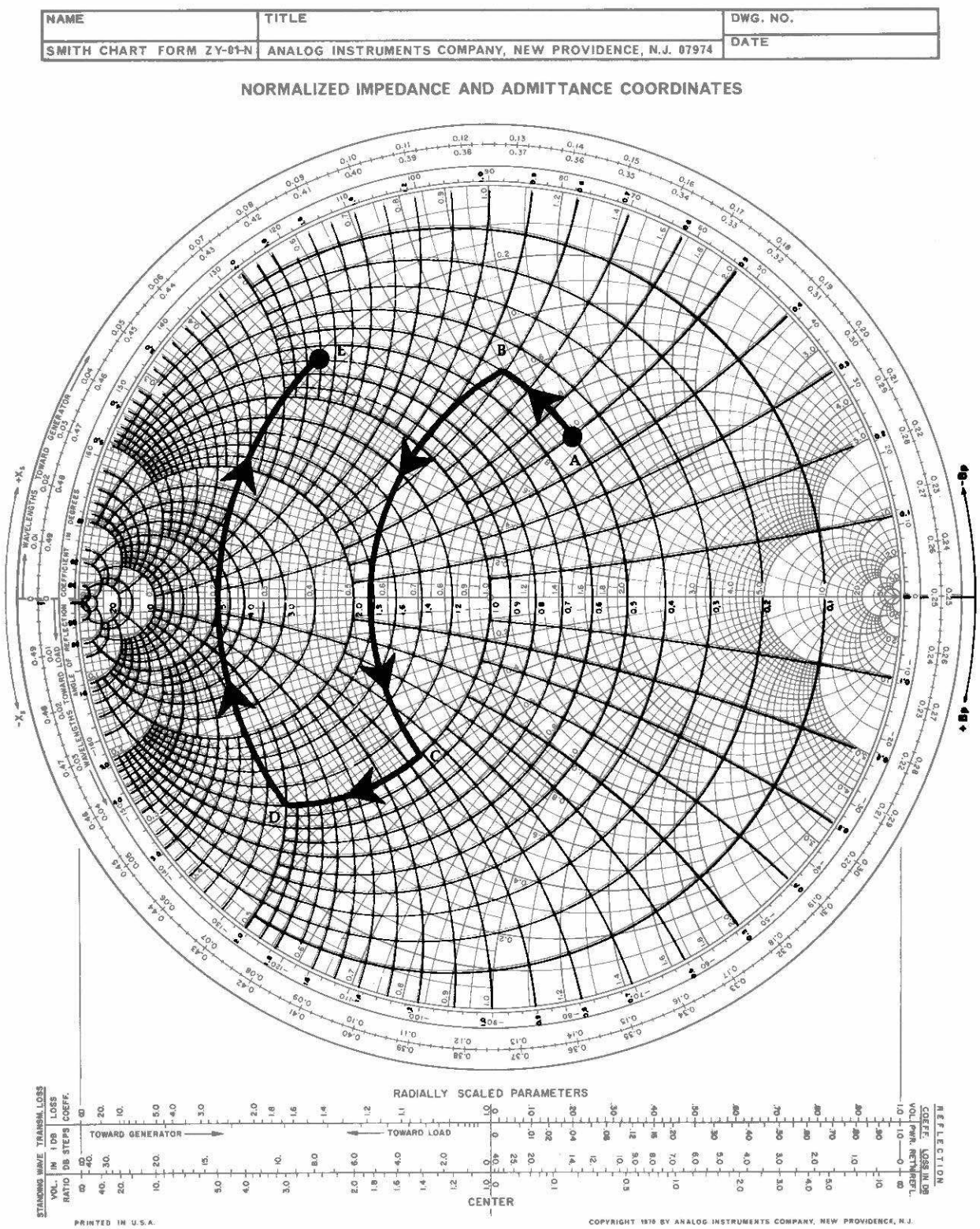


FIG. 4-42. Smith Chart solution for Example 4-6. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 4-7

Design a two-element impedance-matching network on a Smith Chart so as to match a $25 - j15$-ohm source to a $100 - j25$-ohm load at 60 MHz. The matching network must also act as a low-pass filter between the source and the load.

Solution

Since the source is a complex impedance, it wants to "see" a load impedance that is equal to its complex conjugate (as discussed in earlier sections of this chapter). Thus, the task before us is to force the $100 - j25$-ohm load to look like an impedance of $25 + j15$ ohms.

Obviously, the source and load impedances are both too large to plot on the chart, so normalization is necessary. Let's choose a convenient number ($N = 50$) and divide all impedances by this number. The results are $0.5 + j0.3$ ohm for the impedance the source would like to see and $2 - j0.5$ ohms for the actual load impedance. These two values are easily plotted on the Smith Chart, as shown in Fig. 4-44, where, at point A, Z_L is the *normalized* load impedance and, at point C, Z_s^* is the *normalized* complex conjugate of the source impedance.

The requirement that the matching network also be a low-pass filter forces us to use some form of series-L, shunt-C arrangement. The only way we can get from the impedance at point A to the impedance at point C and still fulfill this requirement is along the path shown in Fig. 4-44. Thus, following the rules of Fig. 4-39, the arc AB of Fig. 4-44 is a shunt capacitor with a value of $+jB = 0.73$ mho. The arc BC is a series inductor with a value of $+jX = 1.2$ ohms.

The shunt capacitor as read from the Smith Chart is a susceptance and can be changed into an equivalent reactance by simply taking the reciprocal.

$$X_o = \frac{1}{+jB} = \frac{1}{j0.73\text{ mho}} = -j1.37\text{ ohms}$$

To complete the network, we must now unnormalize all impedance values by *multiplying* them by the number $N = 50$—the value originally used in the normalization process. Therefore:

$$X_L = 60\text{ ohms}$$
$$X_C = 68.5\text{ ohms}$$

The component values are:

$$L = \frac{X_L}{\omega} = \frac{60}{2\pi(60 \times 10^6)} = 159\text{ nH}$$

$$C = \frac{1}{\omega X_c} = \frac{1}{2\pi(60 \times 10^6)(68.5)} = 38.7\text{ pF}$$

The final circuit is shown in Fig. 4-43.

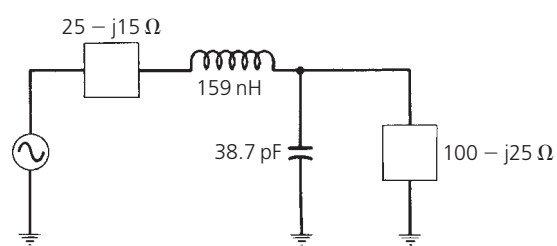


FIG. 4-43. Final circuit for Example 4-7.

Continued on next page

shown. Solution 1 starts with a series-L configuration and takes 9 elements to get to point B. Solution 2 starts with a shunt-L procedure and takes 8 elements, while Solution 3 starts with a shunt-C arrangement and takes 5 elements. The element reactances and susceptances can be read directly from the chart, and Equations 4-11 through 4-14 can be used to calculate the actual component values within minutes.

SOFTWARE DESIGN TOOLS

Another method for matching a given source to a given load is to use one of a number of easily available software design tools. Options range from inexpensive, web-based Smith Chart tools to more comprehensive, integrated design tool environments with impedance matching capabilities.

Smith Chart Tools

When the Smith Chart was first created, it was quickly adopted as a standard, required skill for microwave engineers. These days, it is being employed by designers of high-speed circuits as well as newly degreed RF engineers. While the process of using a Smith Chart remains the same, it no longer has to be manual in nature. Instead, today's computerized Smith Chart software tools provide the engineer real visual insight into the process of mapping the impedance plan onto the reflection coefficient plane.

Rather than on a piece of paper, computerized Smith Chart tools put the entire process on the screen, including a clearly labeled chart, and tabular display of frequency, impedance and VSWR data. Even the circuit that is being designed appears on

EXAMPLE 4-7—Cont

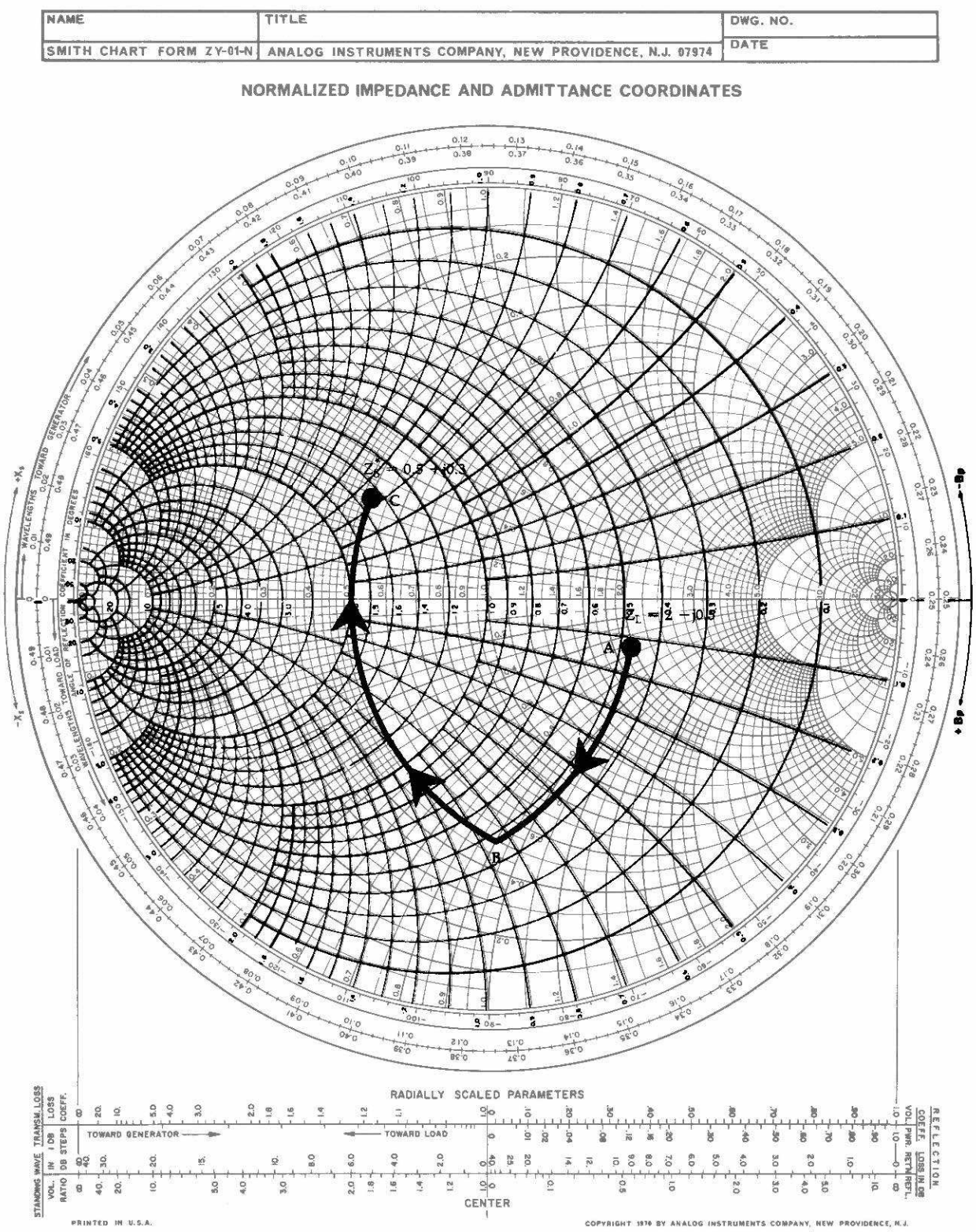


FIG. 4-44. Solution of Example 4-7. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 4-7—Cont

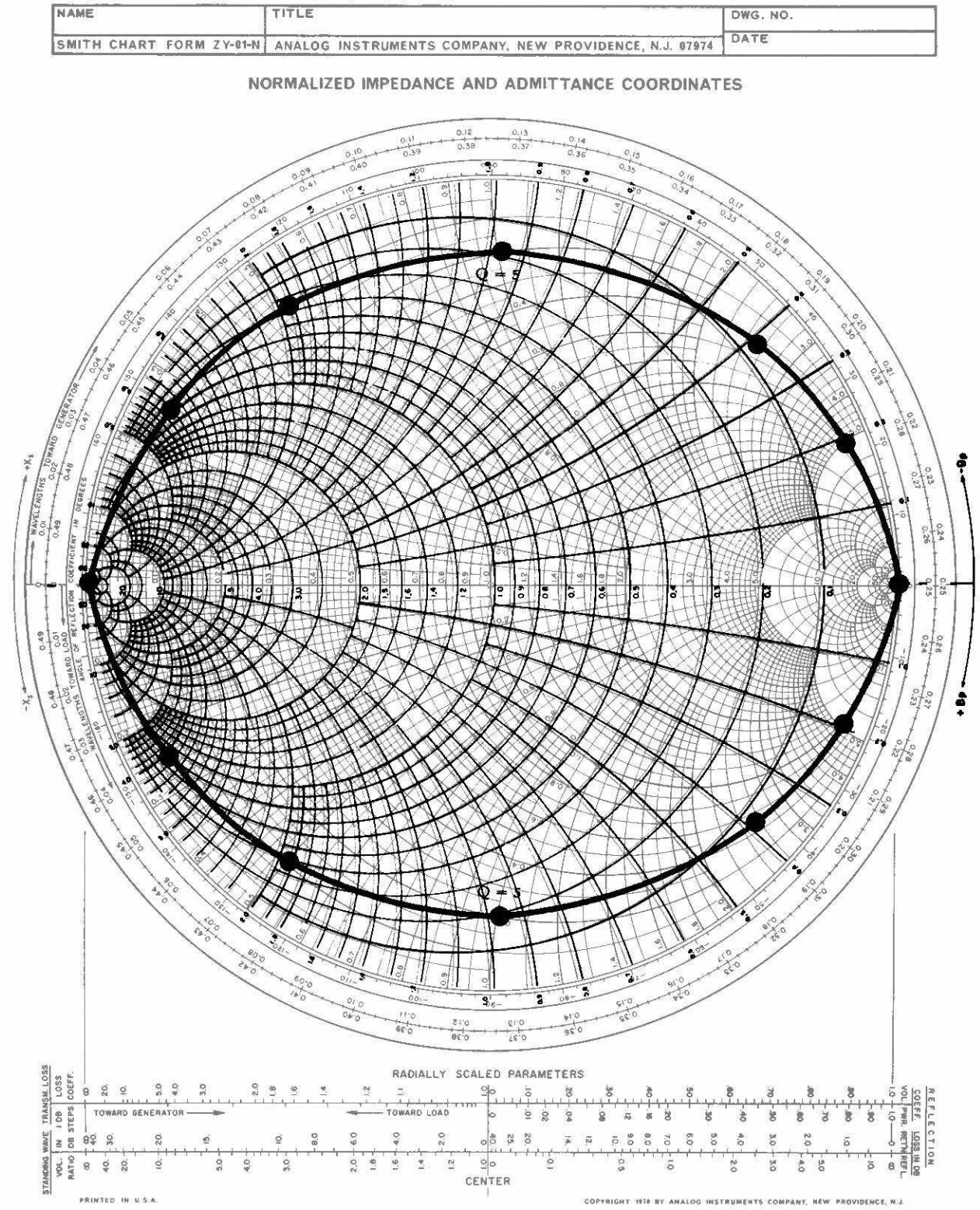


FIG. 4-45. Lines of constant Q. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 4-7—Cont

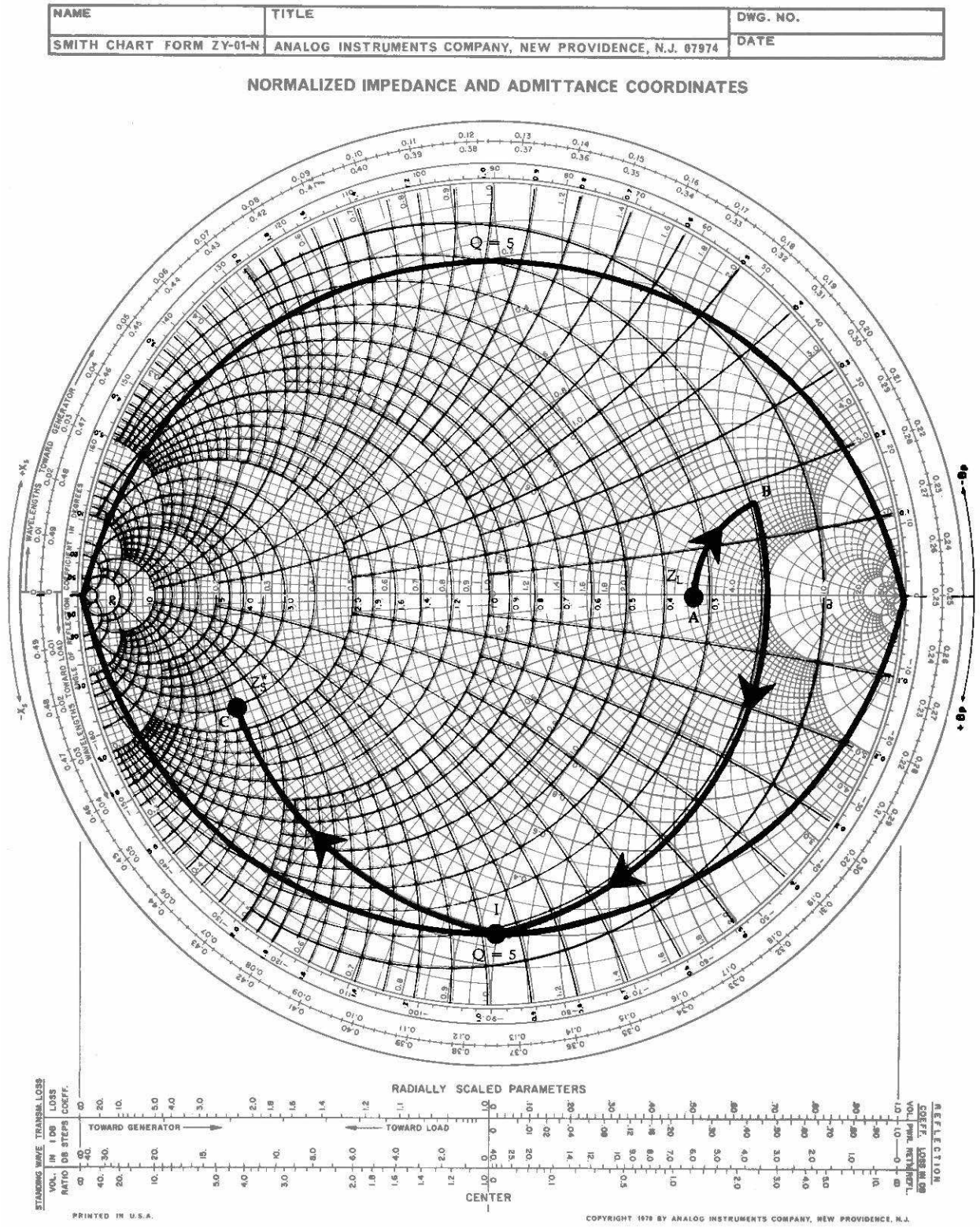


FIG. 4-46. Smith Chart solution for Example 4-8. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 4-8

Design a T network to match a $Z = 15 + j15$-ohm source to a 225-ohm load at 30 MHz with a loaded Q of 5.

Solution

Following the procedures previously outlined, draw the arcs for $Q = 5$ first and, then, plot the load impedance and the complex conjugate of the source impedance. Obviously, normalization is necessary as the impedances are too large to be located on the chart. Divide by a convenient value (choose $N = 75$) for normalization. Therefore:

$$Z_S^* = 0.2 - j0.2 \text{ ohms}$$

$$Z_L = 3 \text{ ohms}$$

The construction details for the design are shown in Fig. 4-46.

The design statement specifies a T network. Thus, the source termination will determine the network Q because $R_S < R_L$.

Following the procedure for $R_S < R_L$ (Step 4, above), first plot point I, which is the intersection of the $Q = 5$ curve and the $R =$ constant circuit that passes through Z_S^*. Then, move from the load impedance to point I with two elements.

Element 1 = arc AB = series L = $j2.5$ ohms

Element 2 = arc BI = shunt C = $j1.15$ mhos

Then, move from point I to Z_S^* along the $R =$ constant circle.

Element 3 = arc IC = series L = $j0.8$ ohm

Use Equations 4-11 through 4-14 to find the actual element values.

Element 1 = series L:

$$L = \frac{(2.5)75}{2\pi(30 \times 10^6)} = 995 \text{ nH}$$

Element 2 = shunt C:

$$C = \frac{1.15}{2\pi(30 \times 10^6)75} = 81 \text{ pF}$$

Element 3 = series L:

$$L = \frac{(0.8)75}{2\pi(30 \times 10^6)} = 318 \text{ nH}$$

The final network is shown in Fig. 4-47.

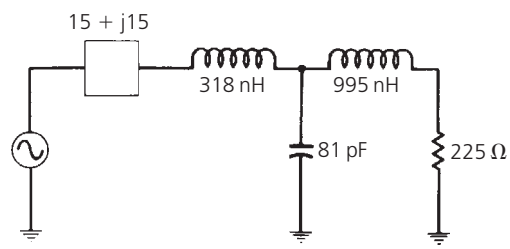


FIG. 4-47. Final circuit for Example 4-8.

the screen. The engineer simply uses a library of lumped and distributed elements to formulate a matched design. Most tools even come with their own time-saving macro commands and tutorials with plenty of examples to guide the engineer through the process. These types of tools generally cost on the order of a hundred dollars or less.

Design Example

As an example, consider that the SmithMatch impedance match network design utility from Microwave Software (www.microwavesoftware.com) is used to analyze a three-element distributed line network for use over the range of 2000 to 3000 MHz. The intention is to match a 10-ohm fixed load to a 50-ohm source (Fig. 4-49).

Note that although SmithMatch is an inexpensive web-based solution, that does not imply that it is not extremely powerful. The Internet simply offers a lower-cost vehicle for delivering the solution to engineers.

The load impedance file for this example is named **"TRL3."** It contains a 10-ohm fixed resistor at four frequencies in the 2000-MHz to 3000-MHz range. To begin, call the SmithMatch module by choosing "(1) SmithMatch" from the tool's Main Menu. You will see:

System Z0 [< Enter >= Quit]?_

Enter **"50"** as the system Z0 characteristic impedance, and then press **<Enter>**.

Filename?_

Enter **"TRL3"** as the name of the .IMP load impedance file and then press **<Enter>**.

The 10-ohm load file will appear on the screen. A small circle marks the low end of the band as illustrated in Fig. 4-50.

A VSWR = 1.5 circle is added to the plot in Fig. 4-50. When this is done, the **"Command ?"** prompt will appear. Next, three transmission line circuit elements (TRLs) are added—one at a time—and the first cut match analyzed.

The SmithMatch tool defines the element code for a "TRL" as 16. Note that a distributed line has two degrees of freedom, its Z0 (characteristic impedance), in ohms, and its electrical line length, theta, in degrees. An important convention in SmithMatch is that when you enter theta—the electrical length of a distributed

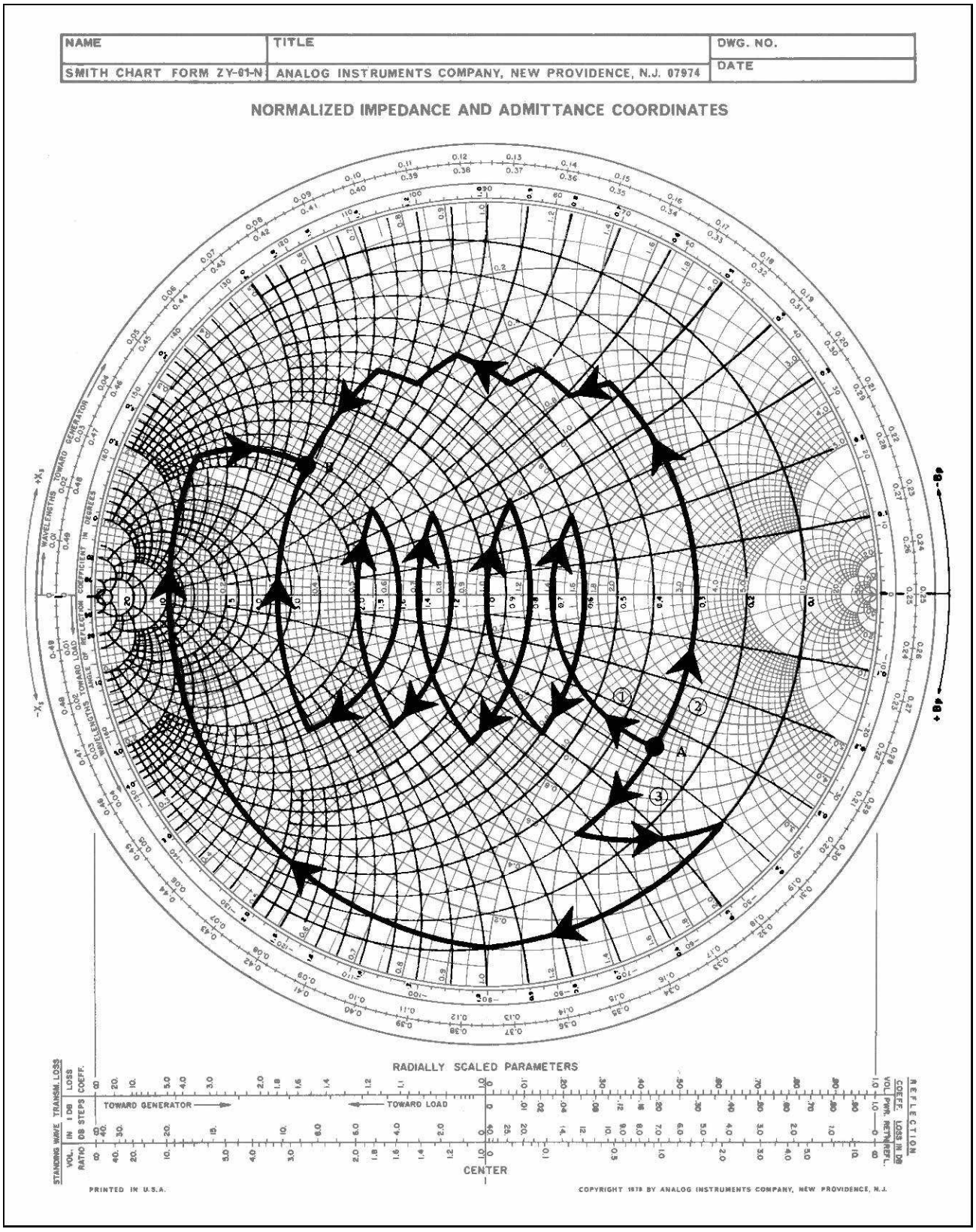


FIG. 4-48. Multi-element matching. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

element, like a 16, 17, or 18 in the tool's Element Library—you specify the length in degrees at the low end of the band.

In this instance, since the band is 2 to 3 GHz, the line lengths must be specified at a frequency of 2 GHz. The computation is simple. Just multiply the given length in degrees by the fraction 2/3. The "2" is the low band edge, and the "3" is the reference frequency for the 90-degree length lines.

When you type **"16"** at the **"Command ?"** prompt, to enter the Z0 = 20-ohm line closest to the load, you'll be asked to enter values for the two TRL parameters.

TRL Z0,Theta ? _

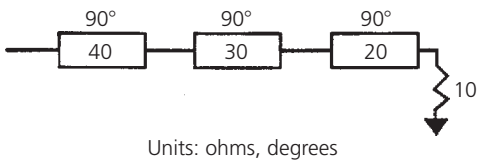
There is a comma separating the two line parameters above, so type **"20, 60"** and press **<Enter>**. For distributed elements, always type the two values separated by a comma. Fig. 4-51 illustrates the screen display after the first TRL has been added.

Type **"Y"** in response to the question:

Save Element (Y/N) ? _

Note that the on-screen circuit file has once again been updated. Now enter the second line, Z0 = 30 ohm TRL. You see the screen in Fig. 4-52.

The Smith Chart is now getting a bit cluttered, so it is time to clean it. Type **"C"** for clean at the **"Command ?"** prompt. What will be left is the last trace drawn. Finally, add the last TRL and look at the final results. Type **"40,60"** when asked; then, after cleaning the screen once again, you should see the plot in Fig. 4-53.



Units: ohms, degrees
Note : Line length specified at 3000 MHz

FIG. 4-49. First cut network using the SmithMatch solution available at www.microwavesoftware.com.

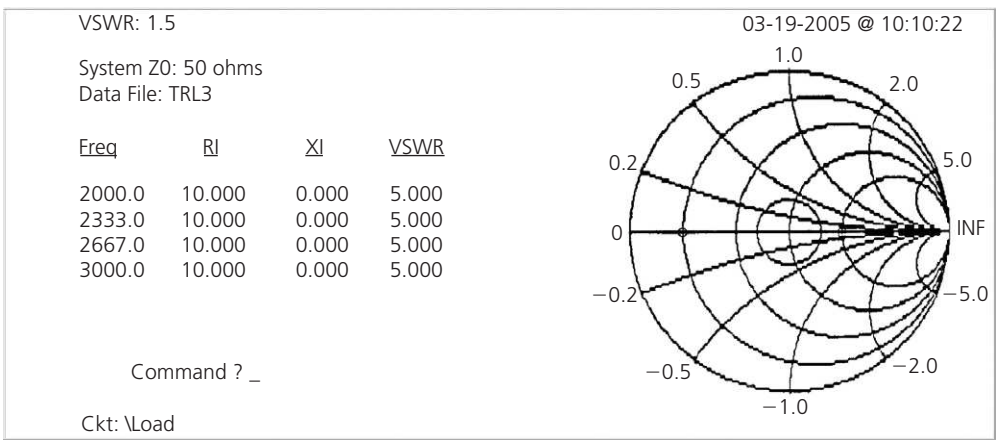


FIG. 4-50. The low end of the band is denoted by the small circle over on the left half of the Smith Chart on the axis of reals.

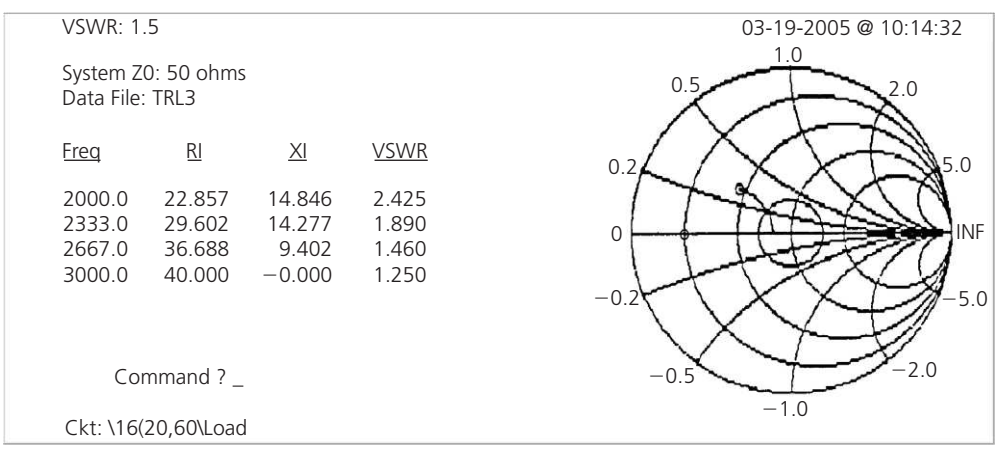


FIG. 4-51. In this display notice that the impedance plot has spread out, and is moving in towards the center of the chart.

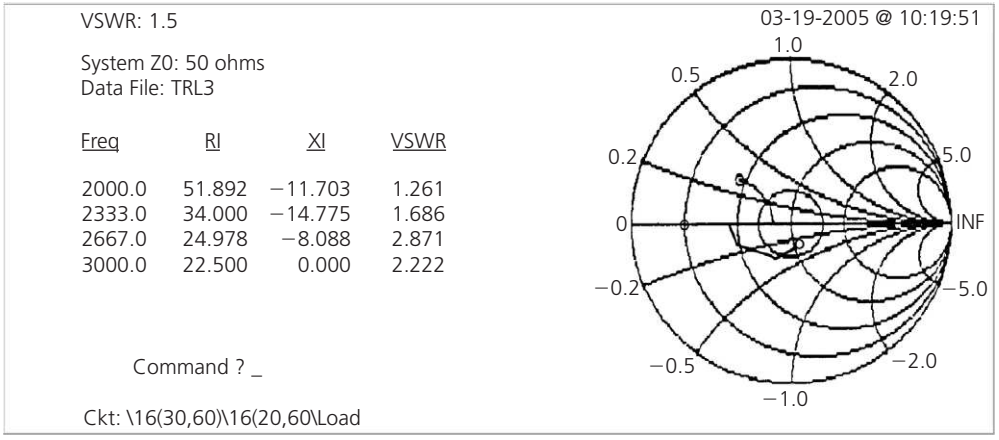


FIG. 4-52. The second quarter-wave TRL caused the trace to flip, end for end, and you are now closer to chart center at the low end of the band.

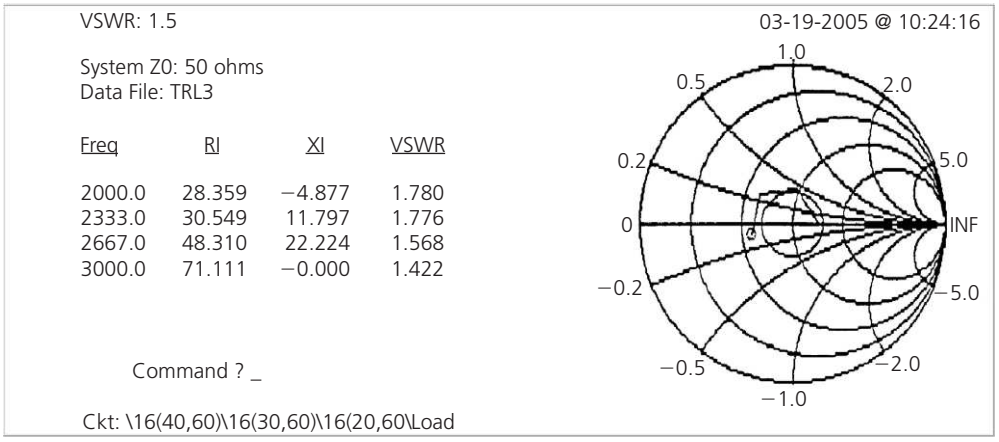


FIG. 4-53. Note in this figure that the last quarter-wave section again flipped the trace end for end. Save this last element so that the on-screen circuit file will be updated and look like the plot shown here.

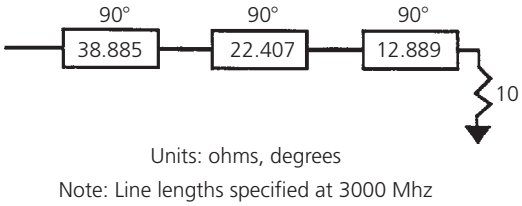


FIG. 4-54. 2–3 GHz broadband TRL match.

The result is a pretty good first-cut match. It is not all that great at the low end of the band, but it is good enough for tweaking, either by hand, or via the use of the Microwave Software OptiMatch program. The 2–3 GHz Broadband TRL Match circuit with the final optimized values is shown in Fig. 4-54.

Integrated Design Tools

Integrated Electronic Design Automation (EDA) software for designing RF and microwave components and subsystems provides designers with state-of-the-art performance in a single design environment that is fast, powerful and accurate. In other words, the engineer has access not just to a solution for impedance matching, but to an entire design environment that supports a range of functionality to assist the engineer throughout the RF circuit design process, from initial system architecture through final documentation.

It is important to note that, while these solutions can be quite effective and easy to use, they do require familiarity with the multiple data inputs that need to be entered and the correct formats. You will also need some expertise to find the useful data among the tons of results coming out. These types of tools generally cost on the order of a few thousand dollars and up.

Design Example

As an example, consider the Genesys software platform from Agilent Technologies (www.agilent.com/find/eesof). With five different configurations, it accommodates the range of RF

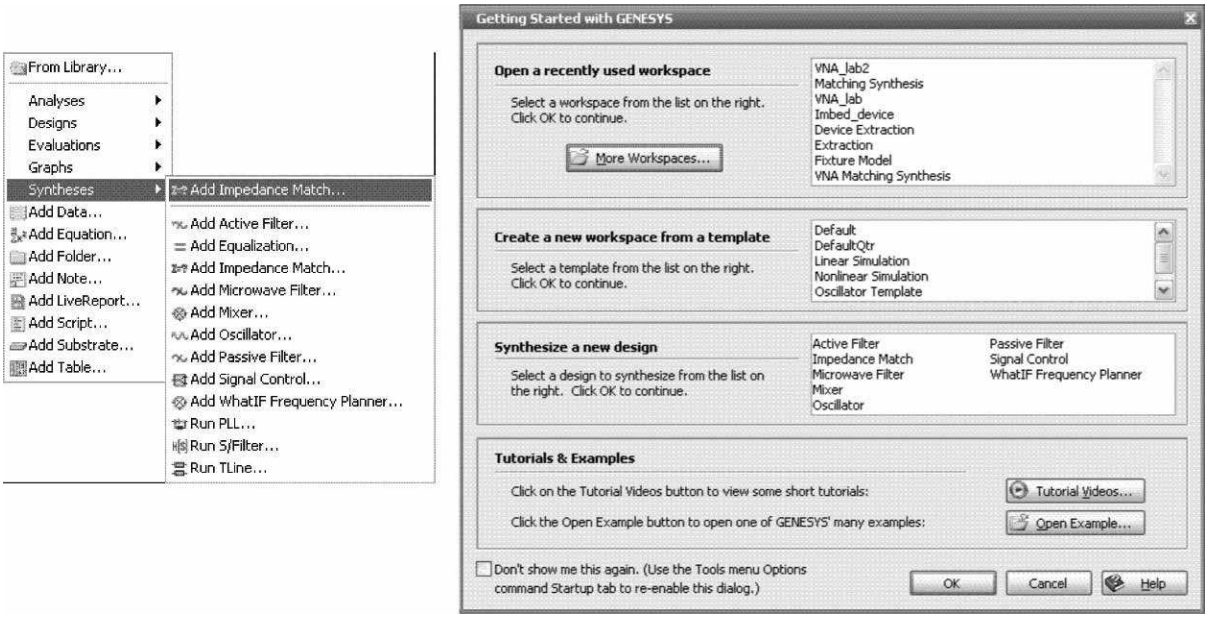


FIG. 4-55. Start-up dialog showing how to launch MATCH.

tasks that today's engineers perform. Its impedance matching synthesis tool, known as MATCH, synthesizes simple quarter-wave, Pi and Tee networks through general-order Chebyshev networks. The tool works with single stage or multiple stages, non-unilateral devices, and arbitrary terminations and is well-equipped to find solutions for broadband, ill-behaved terminations.

To better understand how this solution works in comparison to a point tool like SmithMatch, consider the following example in which Genesys first captures measured data from a vector network analyzer (VNA) and then performs de-embedding to extract the actual device parameters. Next, the corrected device data is incorporated into an amplifier design and MATCH is used to provide a simultaneous input and output matching structure. This example focuses primarily on the use of MATCH.

To begin, select the **"new file"** icon from the top menu bar. Then launch the MATCH synthesis tool from the start-up dialog shown in Fig. 4-55.

After accepting the default naming, the MATCH dialog and associated windows like those shown in Fig. 4-56 will appear. The last configuration used will be displayed as the default menu. On the Settings tab set the frequency of analysis from 2200 MHz to 2600 MHz and the number of points to 50.

Under the **Sections** tab define the termination, input and output matching structures and the de-embedded device s-parameters (Fig. 4-57). For the purposes of this example, the input and output terminations are 50 ohms (default). From the **Sections** tab click on the output port to activate it. Use the **Add Device** to place the symbol for our two port de-embedded part. If you select the **Type** drop down, you will notice that you can reference both a file and a sub-network in MATCH. Using the Browse button select the file named "Device.s2p" from the working directory. Click on the default matching structure "LCPi" and select the **LC Bandpass** as the input matching structure from the **Type** drop-down list in Fig. 4-58. Select the **"no transformer"** option.

Highlighting the output port again, select the **Add Section** button to add the output matching section. Using the same procedure as before, select the **LC Bandpass** with **"no transformer"** for the output matching network. Press the **Calculate** button to enable Genesys to determine the matching topologies component values. The initial response and topology is shown in Fig. 4-59. The optimization default goals are set at −30 dB for input and output matching. They can be modified by double clicking on the optimization icon in the **workspace tree** and selecting the **Goals** tab to modify the matching goals. Additional goals can be added for noise figure, gain, etc. For this exercise, however, we will accept the default goals.

The initial component values have provided a −16 dB (worst case) match over the 400-MHz bandwidth. To help improve this response, click on the **Optimization** button on the **Sections** tab. After a few seconds, the optimized match provides the goal stated −30 dB across the 400-MHz bandwidth (Fig. 4-60).

It should be noted that only devices whose stability factor K is greater than or equal to one will be successfully matched at both input and output. When a device shows "conditional" stability, a trade-off is often required between input or output matching.

An additional feature of the MATCH tool is its ability to set limits on component parameters. Select the **Defaults** tab to view the values for inductor and capacitor *Q*s as well as the limits on distributed elements. These settings can also aide you in limiting

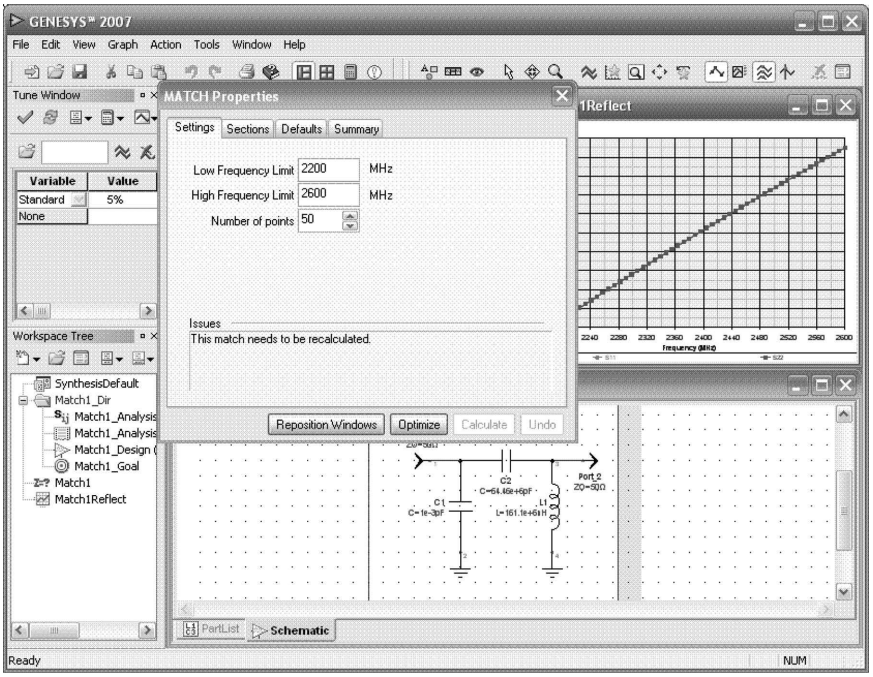


FIG. 4-56. MATCH dialog window.

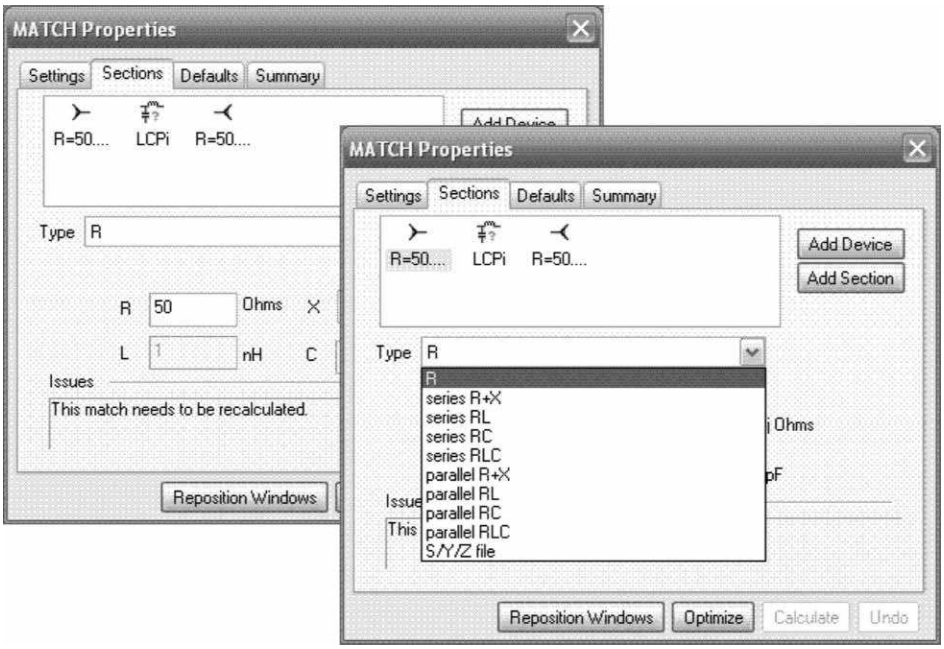


FIG. 4-57. In this figure note that the input and output terminations can be a fixed real impedance, an S/Y/Z parameter file, or modeled as a complex termination made up of passive components.

the realizable transmission structures and in placing a limit on the loss from the lumped components as in Fig. 4-61.

Invariably, optimization provides fractional component values. Proceeding with the next logical step then, you will need to tune the component values to the closest standard value. During this process it may be necessary to make multiple passes in order to optimize the final match.

Select the schematic with the amplifier and matching structure from this example. Press both **"crtl+A"** keys to select all the passive components. Alternately, you can select the

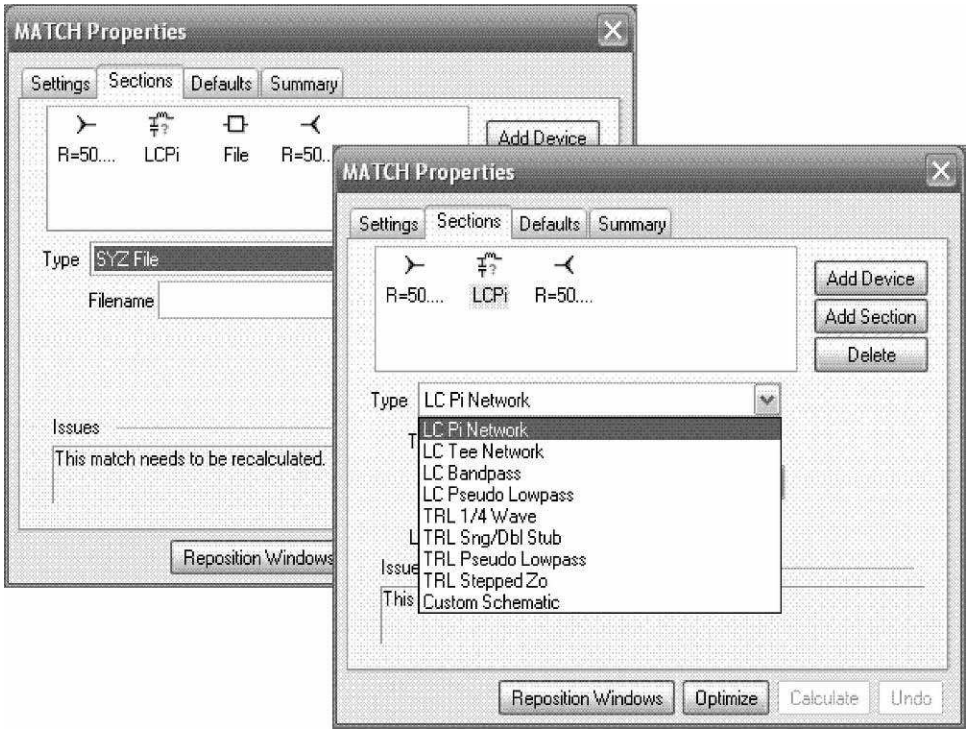


FIG. 4-58. Match properties screen shot.

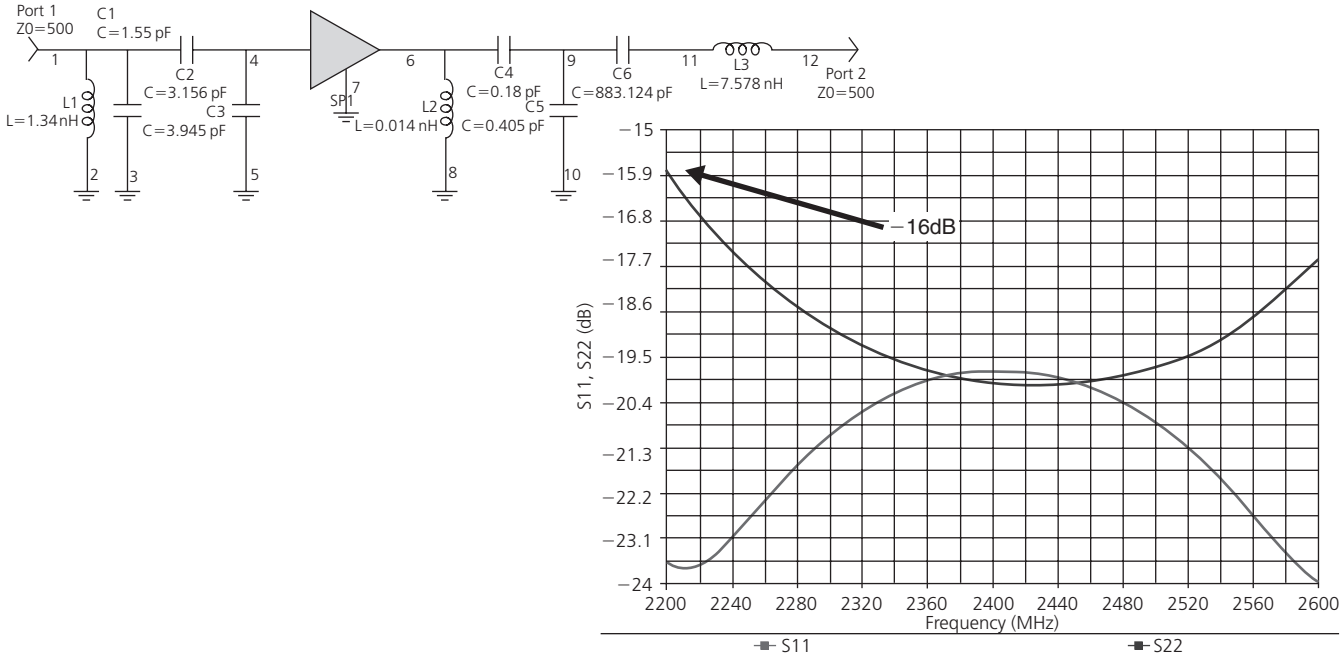


FIG. 4-59. Matching topologies component values calculated by Genesys.

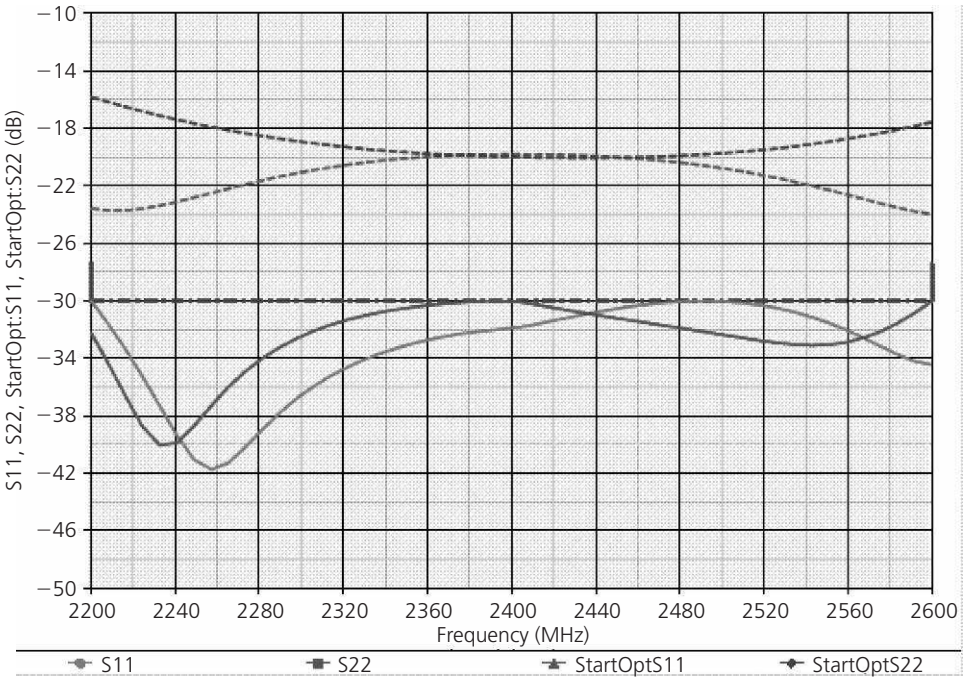


FIG. 4-60. Optimized match.

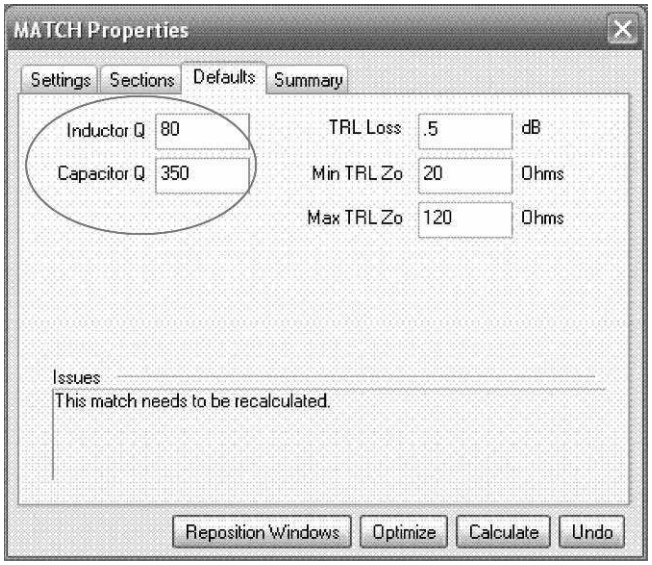


FIG. 4-61. Limiting the loss from lumped components.

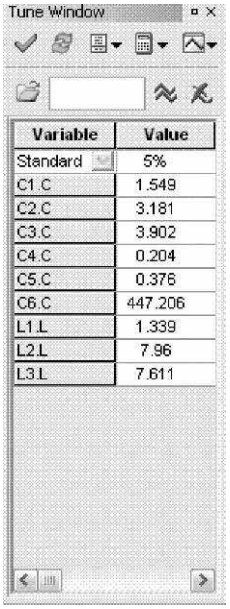


FIG. 4-62. A listing of the tunable components.

components from the top menu **Edit/Select/All** selection. From the **Schematic** menu selection, follow the menu pick to **Make Components Tunable** which will enable tuning for all the components on the schematic. The tunable components will now appear in the **Tune** window in the workspace as shown in Fig. 4-62. Ensure the **Standard 5%** setting is in the tune window and then attempt to optimize the match to meet the original requirements. You may tune each of the components using the mouse wheel when the value is selected or alternately use the **Page Up or Page Down** keys.

The results in Fig. 4-63 represent two to three passes of component tuning. In this figure the minimum match was −26 dB across the band. Note that two of the capacitors are in fractional picofarad values. For small values of capacitance, an interdigital

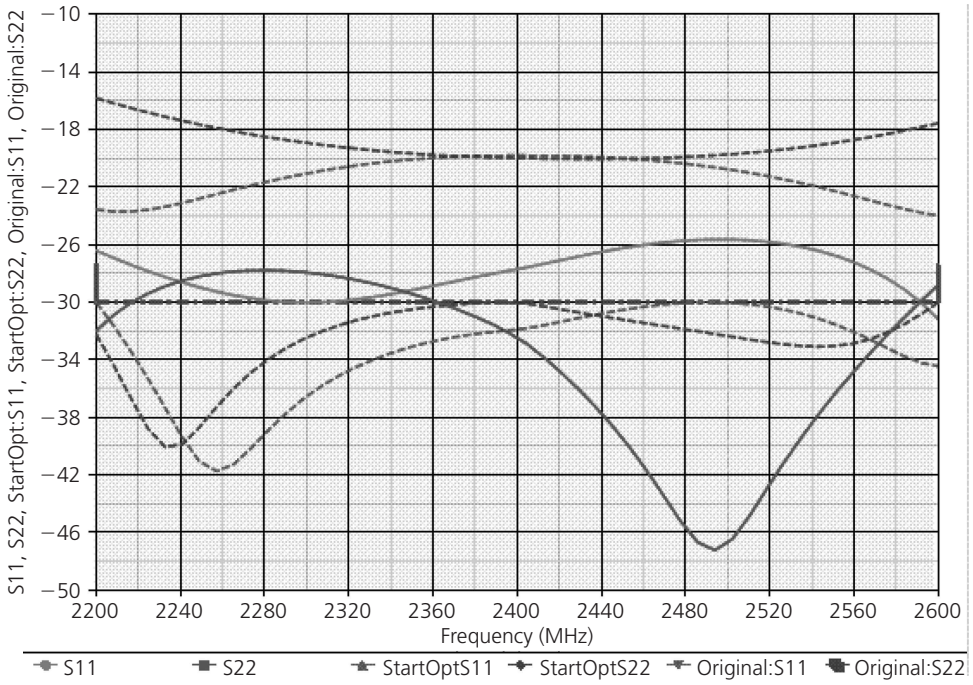


FIG. 4-63. Results of optimization.

capacitor might offer a cost-effective alternative. Going back to the **Sections** tab of MATCH will allow you to try different matching structures to try to further improve on the design.

SUMMARY

Impedance matching is not a form of "black magic" but is a step-by-step, well-understood process that is used to help transfer maximum power from a source to its load. The impedance-matching networks can be designed either mathematically, graphically with the aid of a Smith Chart or via the use of a range of software design tools. Simpler networks of two and three elements are usually handled best mathematically, while networks of four or more elements are very easily handled using the Smith Chart. Design tools can easily accommodate either scenario.

THE TRANSISTOR at Radio Frequencies

CHAPTER 5

In Chapter 1, we discussed resistors, capacitors, and inductors, and their behavior at radio frequencies. We found that, when working at higher frequencies, we could no longer think of a capacitor as just a capacitor, or an inductor as a perfect inductor. In fact, each of these components can be represented by an equivalent circuit that indicates just how imperfect that component really is.

In this chapter, we will find that the transistor, too, is an imperfect device whose characteristics also vary with frequency. Therefore, the equivalent circuit for a typical transistor is introduced and analyzed. Then, with the aid of the equivalent circuit, the input, output, feedback, and gain characteristics are described. We will then examine Y and S parameters and take a look at how manufacturers typically present the transistor's characteristics on their data sheets.

Transistor-level design remains an important part of RF design, even though today's analog and RF engineers live in a world of "connect the IC boxes." There are a number of key reasons why transistor level design—while it may be more difficult and require more time than merely connected IC modules—can be well worth the extra effort. Some of these reasons include:

1. Exploration of design options—Working at the transistor level allows the engineer to explore all possible design options at a powerful device level.
2. Methodologies—The engineer can take advantage of emerging automated design tools and methodologies to speed the design process and make it much more effective.
3. Results—Putting in the higher level of effort required to work at the transistor level is compensated for by allowing the engineer to really push the envelope of speed and power and to take advantage of design reuse.

The topics of design and design tools will be addressed in future chapters.

RF TRANSISTOR MATERIALS

Transistors are used to make all the necessary RF components, from low noise amplifiers (LNA) to high power amplifiers; from mixers and oscillators to switches and attenuators, and more. Further, these RF device transistors are fabricated in a variety of different ways and with different materials. Each different material causes the performance of the transistor to change in subtle—and not so subtle—ways.

There are two basic fabrication methods for transistors used in the RF space: bipolar junction transistors (BJTs) or simply bipolars, and field effect transistors (FETs). The main difference is that bipolars have abrupt junctions in the semiconductor material and FET's do not (see Fig. 5-1).

Instead, FETs have a gate element that creates an electromagnetic field when charged. This gate changes the conductivity of

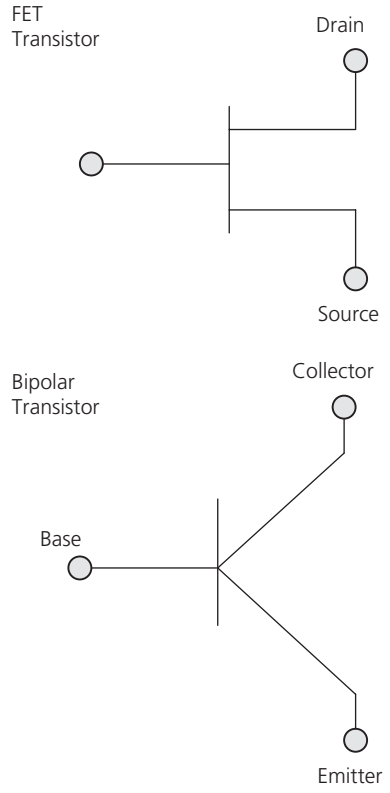


FIG. 5-1. Schematic representation of FET and BJT.

the channel, thus turning the transistor off or on. Once the gate on a FET has been charged, no additional power (really, current) is needed to keep the transistor on (or closed). This means that FETs are voltage-controlled devices.

By comparison, a bipolar transistor is current controlled and requires a small amount of current flowing through the transistor to keep it on. Though small, this current for a single bipolar transistor can quickly add up if you have millions of such transistors on a chip. For this reason, FET-based silicon chips use less power than bipolars. The main disadvantage of FETs is that they switch a little slower than bipolar devices.

RF designers use a variety of power transistors, including: lateral-diffused (LD) MOSFETs, gallium-arsenide (GaAs) metal-semiconductor FETs (MESFETs), GaAs/InGaP heterojunction bipolar transistors (HBTs), gallium-nitride (GaN) high-electron-mobility transistors (HEMTs), and silicon-carbide (SiC) FETs.

THE TRANSISTOR EQUIVALENT CIRCUIT

Just as resistors, capacitors, and inductors can be modeled by an equivalent circuit at radio frequencies, transistor behavior can also be best described by such a circuit as shown in Fig. 5-2. This is a common-emitter configuration of the equivalent circuit known as the hybrid-π model. At first glance, the hybrid-π model looks to be quite formidable for analysis purposes. After defining each component of the model, however, some simplifying assumptions will be made to aid in the analysis process.

$r_{bb'}$—*Base spreading resistance*. This is an inevitable resistance that occurs at the junction between the base terminal or contact and the semiconductor material that composes the base. Its value is usually in the tens of ohms. Smaller transistors tend to exhibit larger values of $r_{bb'}$.

$r_{b'e}$—*Input resistance*. The resistance that occurs at the base-emitter junction of a forward-biased transistor. Typical values range around 1000 ohms.

$r_{b'c}$—*Feedback resistance*. This is a very large ($\approx$5 megohm) resistance appearing from the base to the collector of the transistor.

r_{ce}—*Output resistance*. As the name implies, this is simply the resistance seen looking back into the collector of the transistor. A value for a typical transistor would be about 100K.

C_e—*Emitter diffusion capacitance*. This capacitance is really the sum of the emitter diffusion capacitance and the emitter junction capacitance, both of which are associated with the physics of the semiconductor junction itself and which is beyond the scope of this book. It does exist, however, and since the junction capacitance is so small, C_e is usually called diffusion capacitance with a typical value of 100 pF.

C_c—*Feedback capacitance*. This component is formed at the reverse-biased collector-to-base junction of the transistor. As the frequency of operation for the transistor increases, C_c can begin to have a very pronounced effect on transistor operation. A typical value for this component might be 3 pF.

Also shown in Fig. 5-2 is a *current source* of value $\beta I_{B'}$. Beta (β) is, of course, the small-signal ac current gain of the transistor while $I_{B'}$ is the current through $r_{b'e}$. The current source can be thought of as simply an indication of current flow in the collector that is dependent upon the current that flows in the base of the transistor. Therefore, the collector current is equal to the base current times the β of the transistor, or $I_c = \beta I_{B'}$.

Keep in mind that Fig. 5-2 depicts only those inherent parasitic elements that are *internal* to the semiconductor material itself. Somehow, however, a connection has to be made from the semiconductor material to the transistor leads. This is done with a minute piece of wire called a bonding wire, which, at high frequencies, adds a bit of inductance to the equivalent circuit. The transistor leads themselves tend to exhibit additional series inductance and the equivalent circuit begins to resemble that of Fig. 5-3, where L_B, L_E, and L_C are the base, emitter, and collector lead and bonding inductance, respectively.

It certainly should be obvious now that the equivalent circuit for a typical transistor is not trivial, but contains numerous components, all of which will affect the device's operation at high

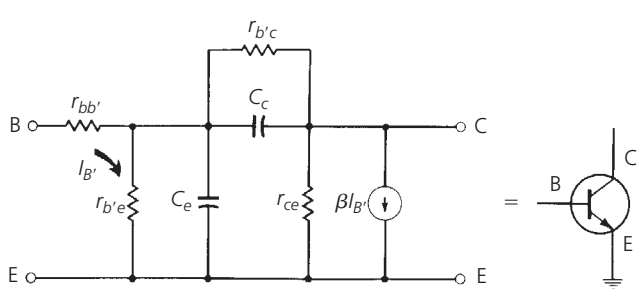


FIG. 5-2. Transistor equivalent circuit—common-emitter configuration.

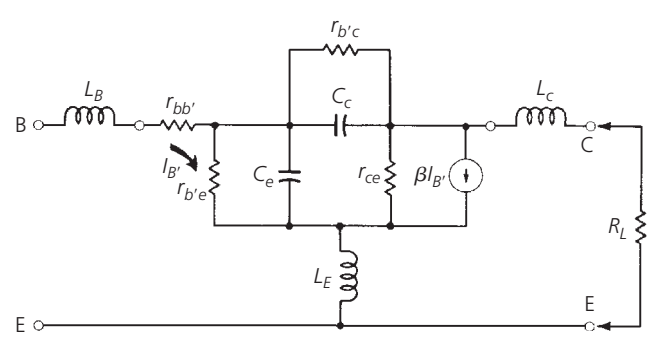


FIG. 5-3. An equivalent circuit including lead inductance.

frequency *to a certain degree*. If some simplifying assumptions are made, however, we should be able to use the equivalent circuit to determine how the transistor behaves at radio frequencies.

Input Impedance

One of the first simplifications that can be made to the circuit of Fig. 5-3 is to eliminate $r_{b'c}$. Five megohms is, after all, a rather large resistance and, for our purposes, looks like an open circuit. The next step is to use a principle called the *Miller effect* to transpose C_c from its series base-to-collector connection to a position that is in parallel with C_e, with a new value of $(C_c)(1 - \beta R_L)$, where R_L is the load resistance. This capacitance is then combined with C_e to form a new total capacitance, C_T. These changes are shown in Fig. 5-4.

The input impedance variation over frequency for a transistor is very easily found by analyzing the circuit of Fig. 5-5. Here we have included only the elements of the equivalent circuit that have an effect on the transistor's input impedance. Notice that the *primary* contributors are $r_{b'e}$ and C_T—neither of which the designer has any control over. The quantity $r_{bb'}$, on the other hand, is a very small resistance while L_B and L_E can vary in size depending on circuit layout. If you are very careful, L_B and L_E can be limited practically to the bonding inductance that was mentioned previously. If this is the case, these elements will have practically no effect on input impedance until well above very high frequencies (VHF).

If we begin our analysis at DC, the circuit of Fig. 5-5 reduces to $r_{bb'}$ in series with $r_{b'e}$ and the input impedance is a pure resistance and is at its maximum value. As the frequency of operation increases, however, C_T begins to play an increasingly important role. Its shunting effect (around $r_{b'e}$) tends to reduce the impedance considerably, until at high frequencies, it effectively eliminates $r_{b'e}$ from the circuit. When this occurs, $r_{bb'}$, L_B, and L_E become the major contributors to the transistor's input impedance.

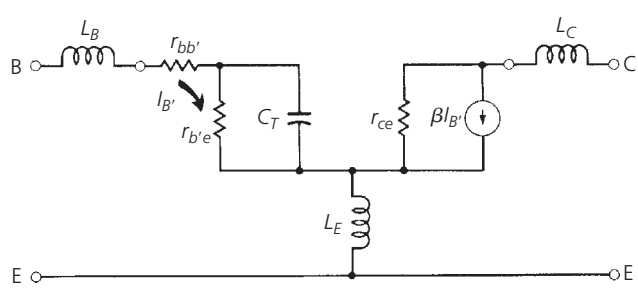


FIG. 5-4. An equivalent circuit using the Miller effect.

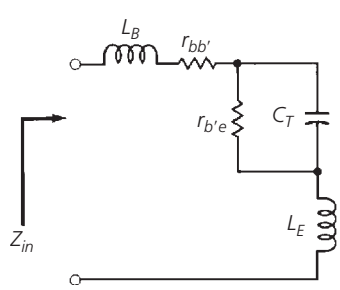


FIG. 5-5. Equivalent input impedance.

The impedance looking into the terminals of Fig. 5-5 can be described as follows:

$$
\begin{aligned}
Z_{\text{in}} &= j\omega L_B + r_{bb'} + \frac{\frac{1}{j\omega C_T}(r_{b'e})}{\frac{1}{j\omega C_T + r_{b'e}}} + j\omega L_E \\
&= j\omega(L_B + L_E) + r_{bb'} + \frac{r_{b'e}}{1 + r_{b'e} j\omega C_T} \\
&= j\omega L_T + r_{bb'} + \frac{r_{b'e}}{1 + j\omega r_{b'e} C_T}
\end{aligned}
$$

This equation is plotted on the Smith Chart shown in Fig. 5-6 with the following values inserted into the equation.

$$
\begin{aligned}
L_T &= 20\,\text{nH} \qquad & r_{b'e} &= 1000\,\text{ohms} \\
r_{bb'} &= 50\,\text{ohms} \qquad & C_T &= 100\,\text{pF}
\end{aligned}
$$

Notice that the chart is normalized for convenience. The actual input impedance of the hypothetical transistor is 1050 ohms at DC and 50 ohms at 112 MHz. Therefore, to find the *actual impedance* of this transistor at any frequency, simply multiply the value found on the chart by 100.

The impedance is presented on the Smith Chart for two reasons. First, and most obvious, is for practice and, second, because of the ease with which both impedance and admittance can be read from the chart at a glance. Most manufacturers, as you will see, use *admittance* parameters rather than *impedance* parameters to describe transistor characteristics on their data sheets. This can sometimes be confusing to the designer who is not used to working with admittances. However, you will soon be handling both impedance and admittance information equally well.

Output Impedance

The output impedance of a transistor typically decreases with frequency. Let's go back to the original circuit of Fig. 5-2 to see why. We can manipulate Fig. 5-2 in much the same manner as was done in the last section, and can arrive at a convenient circuit that will be useful for an output impedance analysis. Looking into the collector terminal, the first component quantity that we see is r_{ce}, which has a typical value of 100K. This resistance is very large in comparison to the other components in the network and can usually be ignored. The same thing can be said for $r_{b'c}$. This leaves us with the circuit of Fig. 5-7.

The first inclination in an analysis of this circuit would be to assume that C_c and C_e are the determining factors in any output impedance calculation and that they alone cause the output impedance to decrease with frequency. Although C_c and C_e do have an effect on the output impedance of the device, there is

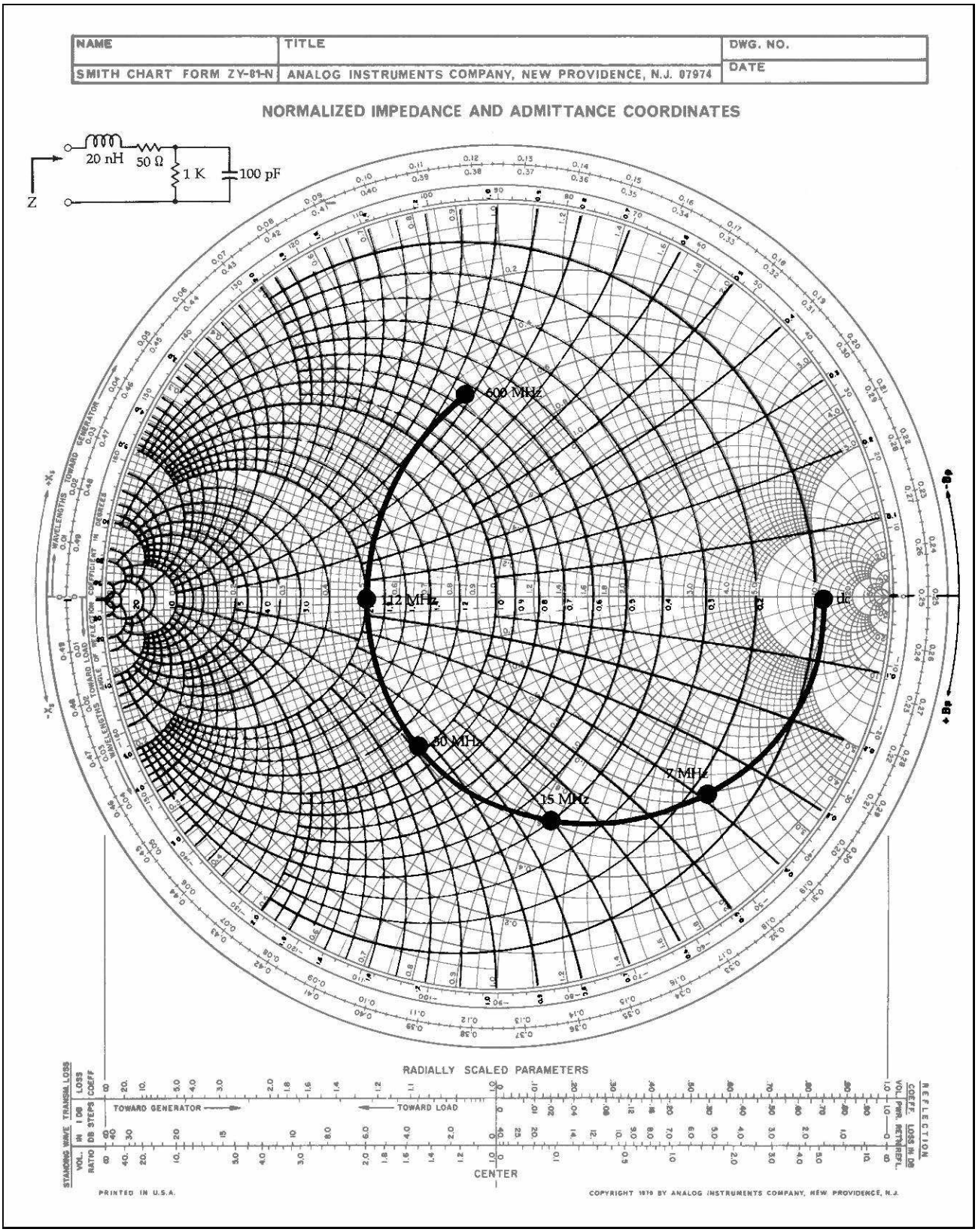


FIG. 5-6. Input impedance vs. frequency. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

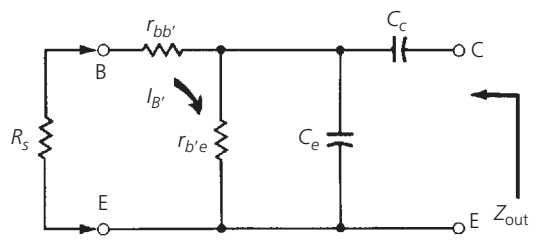


FIG. 5-7. Equivalent output impedance.

another mechanism that is not so obvious that also has quite an effect. This can best be understood if we assume that the transistor is in operation and that some of the collector signal is being fed back to the base through C_c. When this occurs, some of the signal voltage being fed back appears across $r_{b'e}$ causing current to flow in the resistor. This current flow in the base region is amplified by the β of the transistor, thus increasing the collector current. The increase in collector current appears as a decrease in collector impedance. Therefore, even though C_c and C_e act to reduce the output impedance level of the transistor through a decrease in their capacitive reactances, there is also a *hidden* element which tends to further decrease the impedance level beyond that which you would ordinarily expect to find by just looking at the equivalent circuit. Any changes in an external source resistance (R_s) will also change Z_{out}. Increasing R_s decreases Z_{out} because more of the signal current being fed back is forced through $r_{b'e}$.

Feedback Characteristics

The feedback components of the transistor equivalent circuit that is shown in Fig. 5-2 are $r_{b'e}$ and C_c. Of the two, C_c is the most important since it is the element whose value changes with frequency. The quantity $r_{b'e}$, on the other hand, is very large and constant and contributes very little to the feedback characteristics of the device.

As the frequency of operation for a transistor increases, C_c becomes more and more important to the circuit designer because, of course, its reactance is decreasing. Thus, more and more of the collector signal is fed back to the base. At low frequencies, the feedback is usually not much of a problem because C_c, coupled with other stray capacitances located in and around the circuit or circuit-board area, is usually not enough to cause instability. At high frequencies, however, stray reactances coupled with C_c could act to produce a 180° phase shift from collector to base in the fed-back signal. This 180° phase shift, when added to the 180° phase shift that is produced in the normal signal inversion from base-to-collector during amplification, could turn an amplifier into an oscillator very quickly.

Another problem associated with the internal feedback of the transistor is the fact that the collector circuitry is not truly isolated from the base circuitry. Thus, any change in the load resistance of the collector circuitry directly affects the input impedance of the transistor. Or, similarly, any change in the source resistance

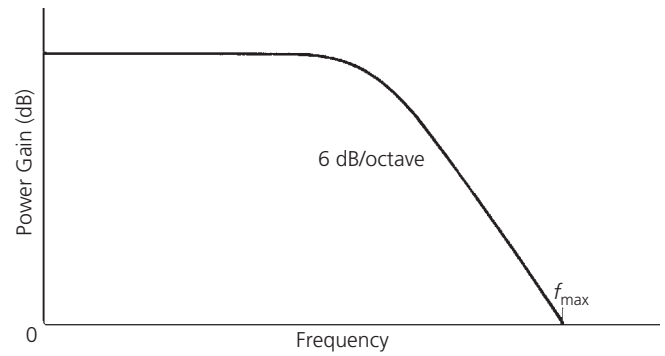


FIG. 5-8. Typical power gain vs. frequency curve.

in the base circuitry directly affects the output impedance of the transistor. This malady is especially important to consider when you are trying to perform an impedance match on both the input and the output of the transistor simultaneously. If, for example, you first match the transistor's input impedance to the source and then match the load to the transistor's output impedance, the output matching network will cause the transistor's input impedance to change from its original value. Therefore, the input matching network is no longer valid and must be redesigned. Once you redesign the input matching network, however, this impedance change will reflect through to the collector causing an output impedance change which invalidates the output matching network. Therefore, if you totally ignore the feedback components in the transistor's equivalent circuit when designing impedance matching networks, you will not obtain a *perfect* match for the transistor. Nevertheless, if C_c is small, the match at both the input and the output might be tolerable in many cases.

It should be pointed out that there is a method for performing a *simultaneous conjugate match* on a transistor while taking into account the effects of C_c. This method is covered in detail in Chapter 6.

Gain

The gain that we are normally interested in for RF transistors is the *power gain* of the device, rather than just the voltage or current gain. It is power gain that is important because of the myriad of impedance levels which abound in RF circuitry. When an impedance level changes in a circuit, voltage and current gains alone no longer mean anything. Even a passive device can produce a voltage or current gain but it cannot produce both simultaneously. That is what transistors are for—to produce real gain.

The power gain of a transistor typically resembles a curve similar to that shown in Fig. 5-8. This curve is not at all surprising if you again consider the equivalent transistor circuit of Fig. 5-4. Notice that what we have, in effect, is an RC low-pass filter with a gain which must fall off (neglecting lead inductance) at the rate of 6 dB per octave. The maximum frequency at which the transistor provides a power gain is labeled as f_{max} in the diagram. The gain curve passes through f_{max} at 0 dB (gain = 1), and at the rate of 6 dB per octave.

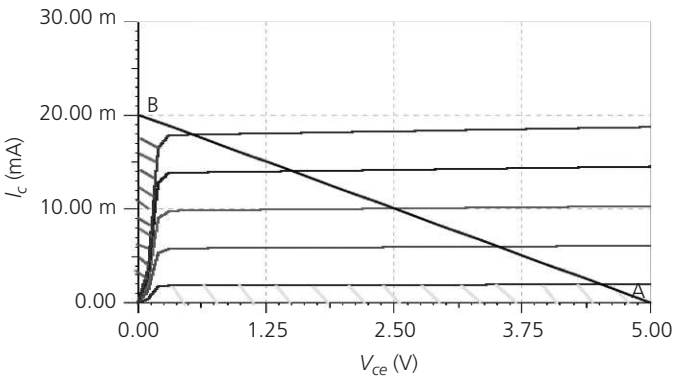


FIG. 5-9. Transistor's characteristic curves.

Gain is usually classified as either unilateralized, neutralized, or unneutralized. *Unilateralized power gain* is defined as the gain available from the transistor when the effects of both feedback components ($r_{b'c}$ and C_c) are negated. Remember that $r_{b'c}$ and C_c are providing negative feedback internal to the transistor and, thus, decrease the gain of the device. Eliminating the negative feedback increases the gain of the transistor. *Neutralized power gain* is that gain which occurs when only the feedback capacitance (C_c) is negated or neutralized. *Unneutralized gain*, on the other hand, occurs when *neither* feedback component is compensated for. Of the three, the unilateralized amplifier produces the most gain and the unneutralized amplifier produces the least. The difference in power gains between the unilateralized case and the neutralized case is usually so small as to be negligible. Thus, neutralization is usually sufficient.

Neutralization is accomplished by providing *external* feedback from the collector to the base of the transistor at just the right amplitude and phase to exactly cancel the internal negative feedback. Further details on this are provided in Chapter 6.

Transistor as a Switch

Before we go further, a quick word is in order about the use of transistors as switches, since several other chapters deal with transistors as amplifiers. Most electronic components and devices can't source enough power to operate. Transistors, when used as a switch, have the unique capability of allowing large amounts of current flow, all of which is controlled by a very small change in junction voltage—so long as the ground between the transistor and the rest of the circuit is the same.

When a bipolar transistor is used in any circuit, be it as a switch or amplifier, its function is determined by the device's set of characteristic curves (see Figure 5-9). The output characteristic curves determine the range of output voltage (collector-emitter voltage or V_{ce}) for different collector current, I_c. If the transistor is to be used as an amplifier, then biasing is selected so that the linear part of the output curves—the almost horizontal sections) are used. But if the transistor is to be used as a switch, then biasing is selected to operate on regions of the output curve known as saturation and cut-off.

In the cut-off region, the operating conditions of the transistor are zero input base current, zero output collector current and maximum (determined by the supply rail) collector voltage. In this state, the transistor acts like a switch in the off position, with no current flowing through the output. In the saturation region, the transistor is biased so that the maximum amount of base current is applied. This operating condition results in a maximum collector current flow and minimum collector emitter voltage. Here, the transistor is operating as a switch in the on position. In both cut-off and saturation, minimum power is dissipated in the transistor.

Put simply, the base (B) of the transistor (see Figure 5-10) acts like the on/off switch. If a current is flowing to the base, then a path will exist for current to flow from the collector (C) to the emitter (E)—in other words, the "switch" is on. If there is no current flowing to the base, then no current path exists between the collector and emitter—or the "switch" is off.

MEMs as a Switch

Another way to make a switch is with old fashion relays, as opposed to transistors. But using today's latest semiconductor materials and technology, these old fashion relays take on a whole new meaning.

Microelectromechanical systems (MEMS) technology is starting to impact the design of RF front ends through the increasing availability of practical MEMS-based components, such as RF switches. MEMS technology essentially involves the fabrication of mechanical structures, such as a variable capacitor or a switch, by means of a semiconductor process. In a switch, for example, these resulting nanostructures can achieve extremely high levels of off-state isolation through the physical separation of switch ports. In a variable capacitor, the use of MEMS technology makes it possible to achieve the wide capacitance tuning range of a mechanically trimmed capacitor, but with the programmable control of an electronically adjustable device.

MEMS technology is still at a fairly early stage of adoption within RF front-end designs. The technology is perhaps most closely associated with the accelerometer sensors found in many automotive air-bag electronic systems. Heavily financed by military research dollars, MEMS is viewed as a possible compact solution for the switched channel filters common to military tactical radios, allowing a considerably reduction in size and weight in manpack designs compared to current switched-filter solutions.

Availability of MEMS-based variable capacitors and inductors, for example, also opens up the possibility of new approaches to RF front-end designs, with opportunities for programmable rather than fixed preselector and IF filters. In addition, the capability of electronically changing the impedance at different circuit junctions with a MEMS device suggests the possibility of electronically tunable antenna matching circuits and filter/amplifier matching circuits—in essence allowing an RF front

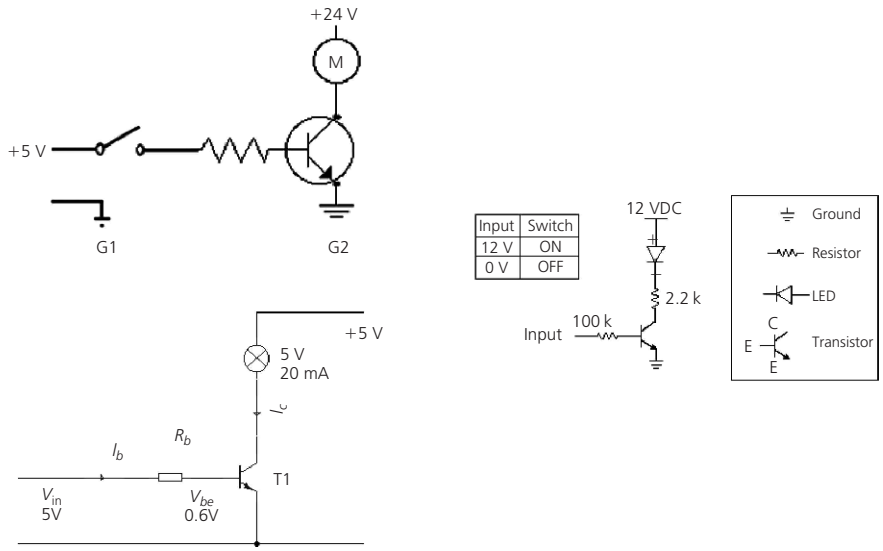


FIG. 5-10. Transistor as a switch.

end to be optimized for a specific set of operating conditions in the field.

Y PARAMETERS

In Chapter 4, admittance was introduced, with the help of the Smith Chart, as the reciprocal of impedance. It is expressed in the form of $Y = G \pm jB$, where G is conductance or the reciprocal of resistance and B is susceptance or the reciprocal of reactance. Both G and jB are taken to be parallel components as opposed to the series representation ($Z = R \pm jX$) for impedance.

The admittance parameters of a transistor are simply a tool to aid in the unambiguous presentation of the characteristics of the device at a certain frequency and bias point. Or, put another way, they are a method of indicating to a potential user what the transistor "looks like" to something connected to its terminals under certain conditions. Admittance parameters can be used to design impedance-matching networks for the transistor, to determine its maximum available gain, and to determine its stability—or lack thereof. In short, they present a model of the transistor to the designer so that he may best utilize the device in his particular application.

The Transistor as a Two-Port Network

The transistor is obviously a three-terminal device consisting of an emitter, base, and collector. In most applications, however, one of the terminals is common to both the input and the output network as shown in Fig. 5-11. In the common-emitter configuration of Fig. 5-11A, for instance, the emitter is grounded and is thus common to both the input and the output network. So,

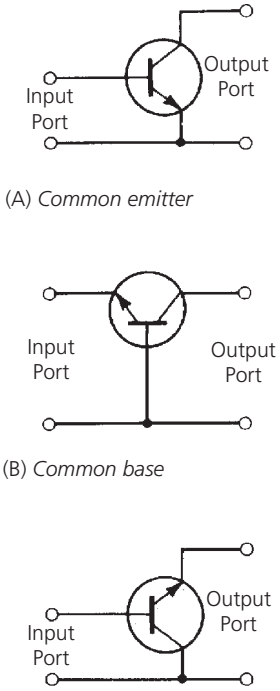


FIG. 5-11. The three-terminal transistor as a two-port network.

rather than describe the device as a three-terminal network, it is convenient to describe the transistor in a *black-box* fashion by calling it a two-port network. One port is described as the input port and the other as the output port. This is shown in

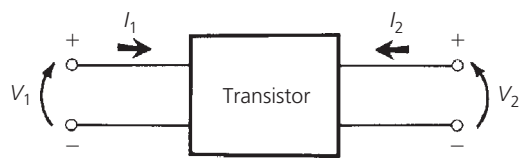


FIG. 5-12. The transistor as a two-port "black box."

Fig. 5-11. Once the two-port realization is made, the transistor can be completely characterized by observing its behavior at the two ports.

Two-Port Y Parameters

Two-port admittance parameters can be used to completely characterize the behavior of a transistor at a certain frequency and bias point, and are considered to be independent of applied signal level as long as linear operation is maintained. The *black-box* configuration used to create the Y-parameter characterization is shown in Fig. 5-12.

The short-circuit Y parameters for the two-port configuration of Fig. 5-12 are given by:

$$y_i = \frac{I_1}{V_1}\bigg|_{V_2 = 0} \qquad \text{(Eq. 5-1)}$$

$$y_r = \frac{I_1}{V_2}\bigg|_{V_1 = 0} \qquad \text{(Eq. 5-2)}$$

$$y_f = \frac{I_2}{V_1}\bigg|_{V_2 = 0} \qquad \text{(Eq. 5-3)}$$

$$y_o = \frac{I_2}{V_2}\bigg|_{V_1 = 0} \qquad \text{(Eq. 5-4)}$$

where

y_i = the short-circuit input admittance,
y_r = the short-circuit reverse-transfer admittance,
y_f = the short-circuit forward-transfer admittance,
y_o = the short-circuit output admittance.

The short circuit that is used to make V_1 and V_2 equal to zero is not a DC short circuit, but a short circuit presented at the signal or test frequency. This is usually accomplished by placing a large capacitor across the terminal which requires a short. An examination of Equation 5-1, for instance, reveals that in order to measure y_i in a laboratory, you would first have to connect a large capacitor across the output terminals of the device. This will set V_2 equal to zero. Then, a known signal voltage (V_1) is injected into the input port and a measurement of I_1 is made. The ratio of I_1 to V_1, with their appropriate phase relationships accounted for, is the short-circuit input admittance of the device, which is usually a complex number in the form of $G \pm jB$. Similarly, to measure y_f, simply leave the short circuit in place, inject signal voltage V_1, and measure I_2. The complex ratio of I_2 to V_1 is the short-circuit transfer admittance. Similar methods are used to measure y_o and y_r.

The short circuit is used in the measurement of Y parameters because of the definition of the two-port model. Referring to Fig. 5-12, it is obvious that the current I_1 is dependent upon the voltage (V_1) at the input terminals of the device. But, what might not be so obvious is the fact that I_1 is also dependent upon V_2. This is due to the internal feedback ($R_{b'c}$ and C_c) of the transistor and it must be accounted for. Stated mathematically:

$$I_1 = y_i V_1 + y_r V_2 \qquad \text{(Eq. 5-5)}$$

which simply states that I_1 is dependent upon the input admittance, the reverse-transfer (feedback) admittance, V_1, and V_2. Notice, however, that if we force V_2 equal to zero, I_1 is totally dependent upon V_1 and the input admittance of the device. Or stated another way, the input admittance can be found by injecting V_1 and measuring I_1. A similar argument can be made for y_r if V_1 is forced equal to zero in Equation 5-5. The equation for the output port is

$$I_2 = y_f V_1 + y_o V_2 \qquad \text{(Eq. 5-6)}$$

Equations 5-3 and 5-4 can be derived from Equation 5-6 by alternately setting V_1 and V_2 equal to zero.

Transistor admittance parameter variations with frequency are often published by manufacturers to aid the designer in his design efforts. They are extremely useful, but often very difficult to measure, especially at high frequency. The difficulty arises at high frequencies mainly due to the fact that a good short circuit is difficult to obtain. As we learned in Chapter 1, a capacitor at high frequencies is not a short circuit at all, but presents some reactance at the operating frequency. Obviously, if any reactance creeps into the "short circuit," the voltage at the port in question is no longer zero and our measurement is no longer valid. The higher the impedance at the "shorted" port, the worse our measurement error becomes. There are, of course, other methods besides capacitors for producing short circuits at the test frequency. But they are generally cumbersome, tedious, and time-consuming, and, as such, leave a lot to be desired. Because of the problems associated with finding a true short circuit at high frequencies, the trend in recent years has been to characterize higher frequency transistors in terms of their scattering or S parameters.

S PARAMETERS

Scattering, or S, parameters are another extremely useful design aid that most manufacturers supply for their higher frequency transistors. S parameters are becoming more and more widely used because they are much easier to measure and work with than Y parameters. They are easy to understand, convenient, and provide a wealth of information at a glance.

While Y parameters utilize input and output voltages and currents to characterize the operation of the two-port network, S parameters use normalized incident and reflected traveling waves at each network port. Furthermore, with S parameters, there is no need to present a short circuit to the two-port device. Instead, the network is always terminated in the characteristic impedance of the measuring system. In the majority of measuring systems, this impedance is 50 ohms (purely resistive).

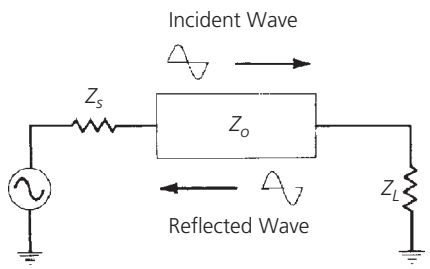


FIG. 5-13. Incident and reflected waves on a transmission line.

The 50-ohm termination requirement is much easier to control than the short-circuit Y-parameter requirement, thus facilitating measurement. In addition, the 50-ohm source and load seen by the two-port network generally forces the device under test, if active, to be stable and not oscillate. This is not always true in a short-circuit measuring system where an active device often does not want to see a short circuit applied to one of its ports. Often such a termination would cause an active device, such as a transistor, to become unstable, thus making measurements impossible. S parameters, therefore, are usually much easier for the manufacturer to measure and, because they are also conceptually easy to understand, are widely used in the design of transistor amplifiers and oscillators.

Transmission Line Background

In order to understand the concept of S parameters, it is necessary to first have a working knowledge of some very simplified transmission line theory. The mathematics have been extensively discussed in the many references cited at the end of the book (Bibliography) and will not be covered here. Instead, you should try to gain an intuitive feel for the incident and reflected traveling waves in a transmission line system.

As shown in Fig. 5-13, voltage, current, or power emanating from a source impedance (Z_s) and delivered to a load (Z_L) can be considered to be in the form of incident and reflected waves traveling in opposite directions along a transmission line of characteristic impedance (Z_o). If the load impedance (Z_L) is exactly equal to Z_o, the incident wave is totally absorbed in the load and there is no reflected wave. If, on the other hand, Z_L differs from Z_o, some of the incident wave is not absorbed in the load but is reflected back toward the source. If the source impedance Z_s were equal to Z_o, the reflected wave from the load would be absorbed in the source and no further reflections would occur. Of course, for a Z_s not equal to Z_o, a portion of the reflected wave from the load is re-reflected from the source back toward the load and the entire process repeats itself perpetually (for a lossless transmission line). The degree of mismatch between Z_o, and Z_L, or Z_s, determines the amount of the incident wave that is reflected. The ratio of the reflected wave to the incident wave is known as the *reflection coefficient* and is simply a measure of the quality of the match between the transmission line and the terminating impedances. The reflection coefficient is a complex quantity expressed as a magnitude and an angle in polar form.

$$\begin{aligned}\Gamma &= \text{reflection coefficient}\\ &= \frac{V_{\text{reflected}}}{V_{\text{incident}}}\\ &= \rho\angle\theta \qquad \text{(Eq. 5-7)}\end{aligned}$$

As the match between the characteristic impedance of the transmission line and the terminating impedances improves, the reflected wave becomes smaller. Therefore, using Equation 5-7, the reflection coefficient decreases. When a perfect match exists, there is no reflected wave and the reflection coefficient is zero. If the load impedance, on the other hand, is an open or short circuit, none of the incident power can be absorbed in the load and all of it must be reflected back toward the source. In this case, the reflection coefficient is equal to 1, or a perfect *mismatch*. Thus, the *normal* range of values for the *magnitude* of the reflection coefficient is between zero and one. The reason *normal* is stressed is that in order for the reflection coefficient to be *greater than one*, the magnitude of the reflected wave from a load impedance must be greater than the magnitude of the incident wave to that load. In order for that to occur, it follows that the load in question must be a source of power. This concept is useful in the design of oscillators, but reflection coefficients that are greater than unity, in the input networks of amplifiers, are very bad news.

As we learned in Chapter 4, the reflection coefficient can be expressed in terms of the impedances under consideration. For example, the reflection coefficient at the load of the circuit shown in Fig. 5-13 can be expressed as:

$$\Gamma = \frac{Z_L - Z_o}{Z_L + Z_o} \qquad \text{(Eq. 5-8)}$$

Notice that if Z_L is set equal to Z_o in Equation 5-8, the reflection coefficient becomes zero. Conversely, setting Z_L equal to zero (a short circuit), the magnitude of the reflection coefficient goes to unity. Thus, Equation 5-8 holds true for the concepts we have discussed thus far.

Often Equation 5-8 is normalized to the characteristic impedance of the transmission line. Thus, dividing the numerator and denominator of Equation 5-8 by Z_o, we have:

$$\begin{aligned}\Gamma &= \frac{\frac{Z_L}{Z_o} - 1}{\frac{Z_L}{Z_o} + 1}\\ &= \frac{Z_n - 1}{Z_n + 1} \qquad \text{(Eq. 5-9)}\end{aligned}$$

where
Z_n is the normalized load impedance.

Equation 5-9 is the same equation that was used in Chapter 4 to develop the Smith Chart. In fact, you will find that reflection coefficients may be plotted directly on the Smith Chart, and the corresponding load impedance read off of the chart immediately—without the need for any calculation using Equations 5-8 or 5-9. The converse is also true. Given a specific characteristic impedance of a transmission line and a load impedance,

the reflection coefficient can be read directly from the chart. No calculation is necessary. Example 5-1 illustrates this fact.

The construction performed in Example 5-1 should take less than 30 seconds once you become familiar with the chart. Obviously, an alternate solution would have been to use Equation 5-8 or 5-9 and perform the computation mathematically. Without the aid of a good scientific calculator to perform the complex number manipulation, however, the numerical computation becomes both tedious and time-consuming. That is the reason Mr. Smith developed the chart in the first place—to perform the transformation between impedance and reflection coefficient, and vice versa, without the need for complex number manipulation.

To find the impedance that gives a certain value of reflection coefficient in a normalized system, you would simply perform the reverse of the construction given in Example 5-1. These procedures are outlined as follows:

1. Draw a line from the center of the chart to the angle of the given reflection coefficient. The normalized impedance is located *somewhere* along that line.
2. From the voltage-reflection coefficient scale located at the bottom of the chart, transfer the value for distance (d), corresponding to the magnitude of the reflection coefficient, to the line drawn in Step 1. Plot this point for a distance (d) from the center of the chart along the line drawn in Step 1.
3. The *normalized* impedance at the point plotted in Step 2 is then read directly from the chart just as any other impedance would be read.

S Parameters and the Two-Port Network

Let us now insert a two-port network between the source and the load in the circuit of Fig. 5-13. This yields the circuit of Fig. 5-16. The following may be said for any traveling wave that originates at the source:

1. A portion of the wave originating from the source and incident upon the two-port device (a_1) will be reflected (b_1) and another portion will be transmitted through the two-port device.
2. A fraction of the *transmitted* signal is then reflected from the load and becomes incident upon the *output* of the two-port device (a_2).
3. A portion of the signal (a_2) is then reflected from the output port back toward the load (b_2), while a fraction is transmitted through the two-port device back to the source.

It is obvious from the above discussion that any traveling wave present in the circuit of Fig. 5-16 is made up of two components. For instance, the total traveling-wave component flowing from the output of the two-port device to the load is actually made up of that portion of a_2 which is reflected from the output of the two-port device, *plus* that portion of a_1 that is transmitted through the two-port device. Similarly, the total traveling wave flowing from the input of the two-port device back toward the source is made up of that portion of a_1 that is reflected from

EXAMPLE 5-1

What is the load reflection coefficient for the circuit shown in Fig. 5-14?

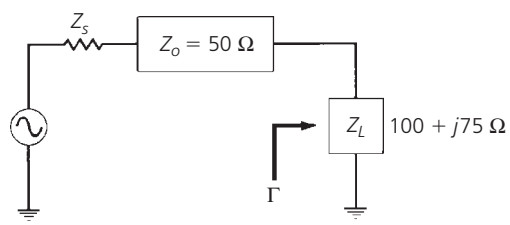


FIG. 5-14. Transmission line circuit for Example 5-1.

Solution

The first step is to normalize the load impedance so that you may plot it on a Smith Chart as was done in Chapter 4. In this case, however, since we are dealing with transmission lines, you must normalize the chart to the characteristic impedance of the line rather than just some convenient number.

$$Z_L = \frac{100 + j75}{50}$$
$$= 2 + j1.5$$

Plot this point on the chart as shown in Fig. 5-15. Draw a line from the center of the chart through the point $2 + j1.5$ and extend this line to the outside edge of the chart (which is calibrated in degrees). Note that for clarity, all extraneous scales normally shown around the periphery of the chart have been eliminated. The reflection coefficient can now be read directly from the chart. The distance from the center of the chart to the point $Z = 2 + j1.5$ is equal to the *magnitude* of the reflection coefficient. To find its numerical value, simply transfer this distance (d) to the *voltage-reflection coefficient* scale located at the bottom of the Smith Chart. This yields a value of 0.54 for the magnitude. To find the angle in degrees, simply read the angle at the intersection of the previously constructed line and the outside edge of the chart. This angle is approximately 29.7°. Thus, the load-reflection coefficient for a load impedance of $100 + j75$ ohms in a 50-ohm system is:

$$0.54\angle 29.7^\circ$$

Continued on next page

EXAMPLE 5-1—Cont

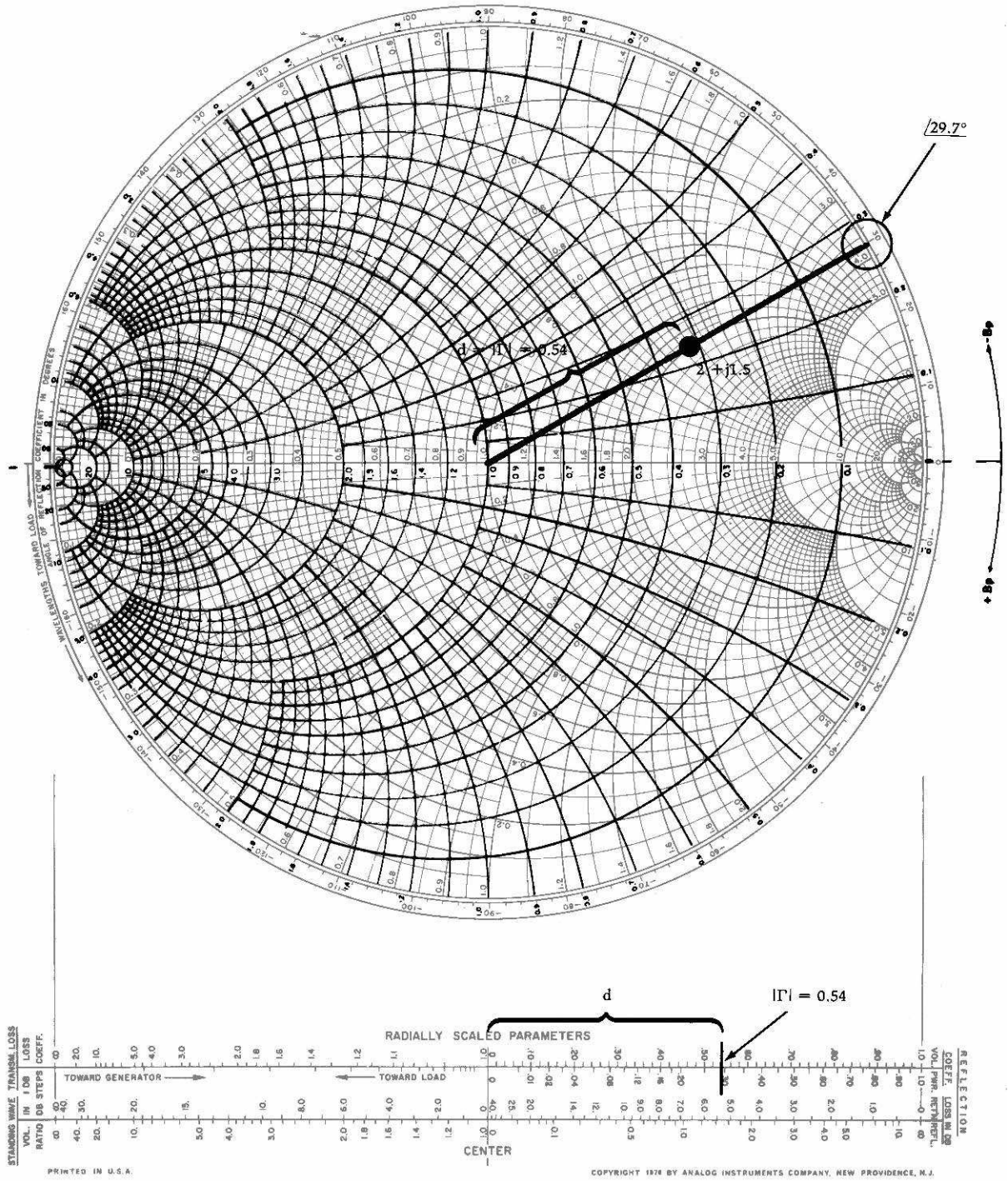


FIG. 5-15. Solution to Example 5-1. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

the input port *plus* that fraction of a_2 that is transmitted through the two-port device.

If we set these observations in equation form, just as was done for the Y parameters, we get the following:

$$b_1 = S_{11}a_1 + S_{12}a_2 \qquad \text{(Eq. 5-10)}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 \qquad \text{(Eq. 5-11)}$$

where,

S_{11} = the input reflection coefficient,
S_{12} = the reverse transmission coefficient,
S_{21} = the forward transmission coefficient,
S_{22} = the output reflection coefficient.

Notice, in Equation 5-10, that if we set a_2 equal to zero, then,

$$S_{11} = \frac{b_1}{a_1}\bigg|a_2 = 0 \qquad \text{(Eq. 5-12)}$$

which is a reflected wave divided by an incident wave and, therefore, by definition, is equal to the input reflection coefficient studied previously. Thus, S_{11} can be plotted on a Smith Chart and the input impedance of the two-port device can be found immediately.

Similarly, using Equation 5-11,

$$S_{22} = \frac{b_2}{a_2}\bigg|a_1 = 0 \qquad \text{(Eq. 5-13)}$$

This is also a reflection coefficient and can be plotted on a Smith Chart. Thus, the output impedance of the two-port device can also be found immediately.

The other two S parameters are found as follows:

$$S_{21} = \frac{b_2}{a_1}\bigg|a_2 = 0 \qquad \text{(Eq. 5-14)}$$

$$S_{12} = \frac{b_1}{a_2}\bigg|a_1 = 0 \qquad \text{(Eq. 5-15)}$$

Notice that Equations 5-12 through 5-15 all require that a_1 or a_2 be set equal to zero in order to measure the individual S parameters. This is easily done by forcing Z_s and Z_L to be equal to the characteristic impedance of the measuring system. Therefore, any wave that is incident upon Z_s or Z_L is totally absorbed and none is reflected back toward the two-port device. For example, let's consider the measurement of the input reflection coefficient, S_{11}. Ideally, we would like to provide an input signal to the two-port device and measure only that fraction of the input signal that is *reflected* back toward the source. In a practical situation, however, some of the incident signal is transmitted through the two-port device, reflected (a_2) from load impedance, and, then, reverse transmitted through the two-port device back to the source. The measured reflected signal is then an aggregate consisting of that portion of a_1, which is reflected, and that portion of a_2, which is transmitted. Obviously, this is not what we need. If Z_L is set equal to Z_o, however, then there is no reflection from the load and the *measured* reflected signal from the input port, divided by the signal incident upon that port, is truly the input reflection coefficient, S_{11}. Similar arguments can be made for the other S parameters to be measured. Therefore, to measure the S parameters of a two-port network, the network is always terminated (source and load) in the characteristic impedance of the measuring system, thus eliminating all reflections from the terminations.

The significance of S_{21} and S_{12}, as shown in Equations 5-14 and 5-15, is that they are simply the *forward* and *reverse gain* (or *loss*) of the two-port network, respectively, when the two-port device is terminated in the characteristic impedance of the measuring system. These are more meaningful than the equivalent Y parameters y_f and y_r, which were previously studied. Parameter y_f, for instance, is a forward transadmittance and y_r is a reverse transadmittance, neither of which can be intuitively related to an insertion gain or loss for the two-port network.

S parameters, like Y parameters, are simply a convenient method of presenting the characteristics of a device to a potential user. Often, a manufacturer will publish both sets of parameters, along with their variation over frequency, to give the designer the flexibility of working with the parameters with which he feels more comfortable. There will come a time, however, when you will be given only one set of parameters and, as things usually go, it will be the wrong set. If you ever run into this problem, simply refer to the following conversion formulas:

1. $S_{11} = \dfrac{(1-y_i)(1+y_o)+y_r y_f}{(1-y_i)(1+y_o)-y_r y_f}$

2. $S_{12} = \dfrac{-2y_r}{(1+y_i)(1+y_o)-y_r y_f}$

3. $S_{21} = \dfrac{-2y_f}{(1+y_i)(1+y_o)-y_r y_f}$

4. $S_{22} = \dfrac{(1+y_i)(1-y_o)+y_f y_r}{(1+y_i)(1+y_o)-y_r y_f}$

5. $y_i = \dfrac{(1+S_{22})(1-S_{11})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}} \times \dfrac{1}{Z_o}$

6. $y_r = \dfrac{-2S_{12}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}} \times \dfrac{1}{Z_o}$

7. $y_f = \dfrac{-2S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}} \times \dfrac{1}{Z_o}$

8. $y_o = \dfrac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1+S_{22})(1+S_{11})-S_{12}S_{21}} \times \dfrac{1}{Z_o}$

Notice that if you are converting from Y to S parameters, as in the first four formulas, each individual Y parameter must first be multiplied by Z_o before being substituted into the equations.

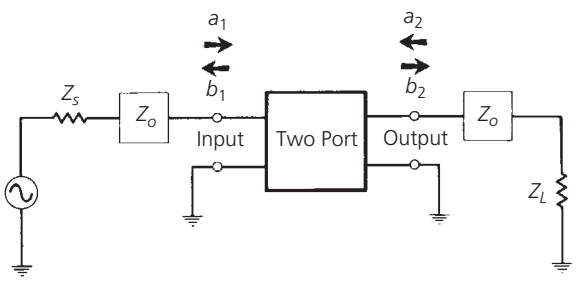


FIG. 5-16. Incident and reflected waves for a two-port device.

UNDERSTANDING RF TRANSISTOR DATA SHEETS

The RF transistor data sheet is only a bit more complex than that of its low-frequency counterpart. In addition to all of the low-frequency information normally provided on a transistor data sheet, the transistor's RF characteristics, in the form of Y parameters and S parameters and other related information, are included as well. It is assumed that you are already familiar with the low-frequency portion of the data sheet. Therefore, we will concern ourselves only with that information that is typically added specifically for RF transistors.

Fig. 5-17 is a data sheet for the Freescale (formerly the Semiconductor Product Section of Motorola) 2N5179 npn silicon RF high-frequency transistor. This particular transistor was chosen simply because the data sheet provides both Y and S parameters and is, therefore, very good for instructional use.

One of the first things you might notice about this particular transistor is that it has four leads! The extra lead is not connected internally to the device itself, but is connected to the case which just happens to be a metal can. In normal circuit operation, the extra pin is grounded, thus providing a shield around the device to help reduce unwanted stray fields.

The first page of the data sheet is fairly straightforward and provides the *never-exceed* ratings for the transistor. This is a common practice even for low-frequency transistors and is nothing new. Notice that this manufacturer does list those applications for which the transistor is well-suited. This particular device was designed for high-gain, low-noise amplifier, oscillator, and mixer applications.

On page 2 of the data sheet, under the heading *Dynamic Characteristics*, several parameters of interest to the RF designer are listed.

f_T—This is called the *transition frequency*, or more commonly, the *gain bandwidth product* of the device. f_T is the theoretical frequency at which the common-emitter current gain (h_{fe}) of the transistor is unity or 0 dB.

Very rarely is f_T used in the RF amplifier-design process except to verify how close you might be to the transistor's upper frequency limit. Keep in mind that f_T is only an indication of the frequency at which the *current* gain of the device drops to 0 dB. Power gain *may* still be possible depending upon the available voltage gain from the device at the frequency in question. Usually f_T is not measured directly for very high-frequency transistors, but is extrapolated from data taken at lower frequencies. The accuracy of the measurement is, therefore, somewhat questionable and, as one manufacturer has stated, f_T is provided on the data sheets simply for historical reasons.

C_{cb}—This is the *collector-to-base capacitance* of the transistor as measured at 1 MHz with a collector-to-base voltage of 10 volts and the emitter open-circuited.

The smaller this capacitance is, the better off you will be, if you are using the transistor in an amplifier configuration. This capacitance can be equated to C_c in the transistor equivalent circuit of Fig. 5-1 at the beginning of this chapter.

h_{fe}—This is the *common-emitter current gain* or *beta* of the transistor at the specified low frequency of 1 kHz.

For an RF circuit design, h_{fe} will not do you much good either. The *DC beta* of the transistor (h_{FE}), however, will provide you with needed information in controlling the DC collector or bias current. This parameter is listed under the *On Characteristics* heading for the device (on the second page).

$r_{b'}C_c$—The *collector-to-base time constant* for the transistor is another measure of its feedback characteristics.

The smaller this number is, the better off you will be. This is another bit of information that is often ignored.

NF—The *noise figure* of the transistor is simply a measure of how much noise the transistor adds to the signal during the amplification process (see Noise Calculations at: http://books.elsevier.com/companions/9780750685184). Notice that for these data sheets, a maximum noise figure of 4.5 dB was measured for the device under a very rigid set of conditions.

Figure 1, on page 3 of the data sheet, was used for the noise figure measurement with the transistor biased at a V_{CE} of 6 volts, $I_C = 1.5$ mA, and the source resistance set equal to 50 ohms. This method of presenting the NF for a transistor, as you can well imagine, is practically useless. Very rarely will the circuit designer ever see the transistor under this exact set of operating conditions. Any variation from these conditions changes the measured noise figure drastically. For this reason, the manufacturer often provides a few noise figure contours which present NF graphically under a wide variety of operating conditions. These contours are shown in Figures 3, 4, and 5 of the data sheet. Figure 3 is a graph of noise figure versus frequency. The NF is measured at various frequencies under the same *bias* conditions. Notice, however, that this measurement was taken with a variable source resistance, where R_s was made equal to its optimum value for a minimum noise figure. The concept of an optimum source resistance for a minimum noise figure is presented in Chapter 6 of this book. Notice that the minimum noise figure increases as the frequency is increased. This is typical of RF transistors.

NPN RF Transistor

2N5179

The RF Line

4.5 dB @ 200 MHz

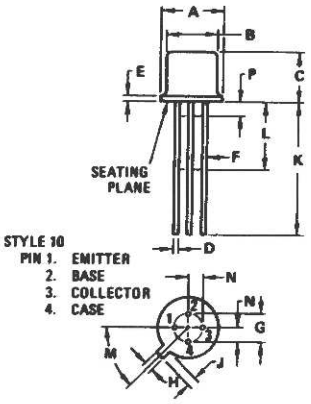
HIGH FREQUENCY TRANSISTOR

NPN SILICON

NPN SILICON RF HIGH FREQUENCY TRANSISTOR

. . . designed primarily for use in high-gain, low-noise amplifier, oscillator, and mixer applications. Can also be used in UHF converter applications.

- High Current-Gain — Bandwidth Product —
  f_T = 1.4 GHz (Typ) @ I_C = 10 mAdc
- Low Collector-Base Time Constant —
  $r_b'C_c$ = 14 ps (Max) @ I_E = 2.0 mAdc
- Characterized with Scattering Parameters
- Low Noise Figure —
  NF = 4.5 dB (Max) @ f = 200 MHz



DIM	MILLIMETERS MIN	MILLIMETERS MAX	INCHES MIN	INCHES MAX
A	5.31	5.84	0.209	0.230
B	4.52	4.95	0.178	0.195
C	4.32	5.33	0.170	0.210
D	0.41	0.53	0.016	0.021
E	–	0.76	–	0.030
F	0.41	0.48	0.016	0.019
G	2.54 BSC		0.100 BSC	
H	0.91	1.17	0.036	0.046
J	0.71	1.22	0.028	0.048
K	12.70	–	0.500	–
L	6.35	–	0.250	–
M	45° BSC		45° BSC	
N	1.27 BSC		0.050 BSC	
P	–	1.27	–	0.050

ALL JEDEC dimensions and notes apply

CASE 20-03
TO-72

*MAXIMUM RATINGS

Rating	Symbol	Value	Unit
Collector-Emitter Voltage Applicable 1.0 to 20 mAdc	V_{CEO}	12	Vdc
Collector-Base Voltage	V_{CB}	20	Vdc
Emitter-Base Voltage	V_{EB}	2.5	Vdc
Collector Current	I_C	50	mAdc
Total Device Dissipation @ T_A = 25°C Derate above 25°C	P_D	200 1.14	mW mW/°C
Total Device Dissipation @ T_C = 25°C Derate above 25°C	P_D	300 1.71	mW mW/°C
Storage Temperature Range	T_{stg}	-65 to +200	°C

*Indicates JEDEC Registered Data.

FIG. 5-17. Data sheet. (Courtesy Freescale (formerly Motorola Semiconductor Products Inc.)).

2N5179

*ELECTRICAL CHARACTERISTICS (T_A = 25°C unless otherwise noted)

Characteristic	Symbol	Min	Max	Unit
OFF CHARACTERISTICS				
Collector-Emitter Sustaining Voltage (I_C = 3.0 mAdc, I_B = 0)	$V_{CEO(sus)}$	12	–	Vdc
Collector-Base Breakdown Voltage (I_C = 0.001 mAdc, I_E = 0)	BV_{CBO}	20	–	Vdc
Emitter-Base Breakdown Voltage (I_E = 0.01 mAdc, I_C = 0)	BV_{EBO}	2.5	–	Vdc
Collector Cutoff Current (V_{CB} = 15 Vdc, I_E = 0) (V_{CB} = 15 Vdc, I_E = 0, T_A = 150°C)	I_{CBO}	– –	0.02 1.0	μAdc
ON CHARACTERISTICS				
DC Current Gain (I_C = 3.0 mAdc, V_{CE} = 1.0 Vdc)	h_{FE}	25	250	–
Collector-Emitter Saturation Voltage (I_C = 10 mAdc, I_B = 1.0 mAdc)	$V_{CE(sat)}$	–	0.4	Vdc
Base-Emitter Saturation Voltage (I_C = 10 mAdc, I_B = 1.0 mAdc)	$V_{BE(sat)}$	–	1.0	Vdc
DYNAMIC CHARACTERISTICS				
Current-Gain – Bandwidth Product ① (I_C = 5.0 mAdc, V_{CE} = 6.0 Vdc, f = 100 MHz)	f_T	900	2000	MHz
Collector-Base Capacitance (V_{CB} = 10 Vdc, I_E = 0, f = 0.1 to 1.0 MHz)	C_{cb}	–	1.0	pF
Small-Signal Current Gain (I_C = 2.0 mAdc, V_{CE} = 6.0 Vdc, f = 1.0 kHz)	h_{fe}	25	300	–
Collector-Base Time Constant (I_E = 2.0 mAdc, V_{CB} = 6.0 Vdc, f = 31.9 MHz)	$r_b' C_c$	3.0	14	ps
Noise Figure (See Figure 1) (I_C = 1.5 mAdc, V_{CE} = 6.0 Vdc, R_S = 50 ohms, f = 200 MHz)	NF	–	4.5	dB
FUNCTIONAL TEST				
Common-Emitter Amplifier Power Gain (See Figure 1) (V_{CE} = 6.0 Vdc, I_C = 5.0 mAdc, f = 200 MHz)	G_{pe}	15	–	dB
Power Output (See Figure 2) (V_{CB} = 10 Vdc, I_E = 12 mAdc, f≥500 MHz)	P_{out}	20	–	mW

*Indicates JEDEC Registered Values.

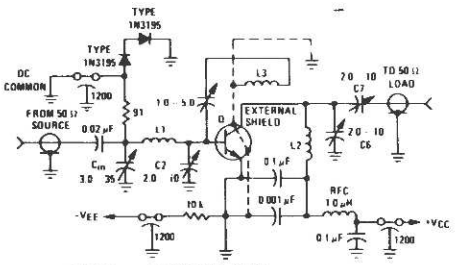
① f_T is defined as the frequency at which $|h_{fe}|$ extrapolates to unity.

NPN RF Transistor

FIG. 5-17. (Continued)

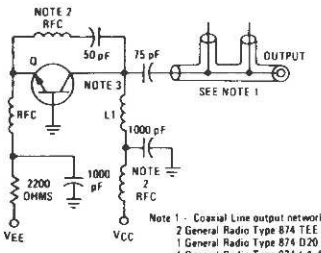
2N5179

FIGURE 1 – 200 MHz AMPLIFIER POWER GAIN AND NOISE FIGURE CIRCUIT



L1 1-3/4 Turns, #18 AWG, 0.5" L, 0.5" Diameter
L2 2 Turns, #16 AWG, 0.5" L, 0.5" Diameter
L3 2 Turns, #13 AWG, 0.25" L, 0.5" Diameter (Position 1/4" from L2)

FIGURE 2 – 500 MHz OSCILLATOR CIRCUIT



Note 1 - Coaxial Line output network consisting of
2 General Radio Type 874 TEE or equivalent
1 General Radio Type 874 D20 Adjustable Stub or equivalent
1 General Radio Type 874-LA Adjustable Line or equivalent
1 General Radio Type 874-WN3 Short-circuit termination or equivalent
Note 2 - RFC = 0.2 μH Ohmite #2460 or equivalent
Note 3 - Lead Number 4 (case) floating
L1 - 2 turns #16 AWG wire, 3/8 inch OD, 1 1/4 inch long
Q = 2N5179

FIGURE 3 – NOISE FIGURE versus FREQUENCY

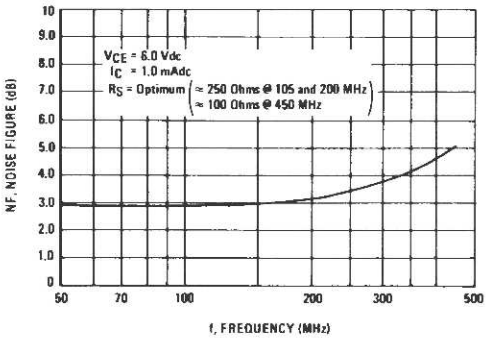


FIGURE 4 – NOISE FIGURE versus SOURCE RESISTANCE and COLLECTOR CURRENT

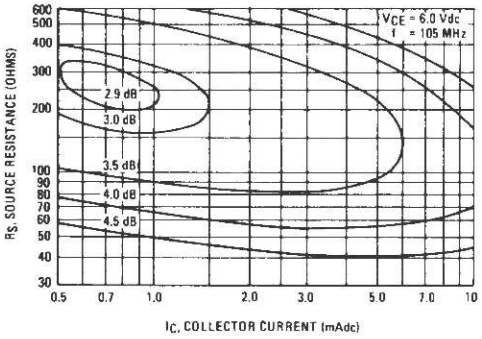
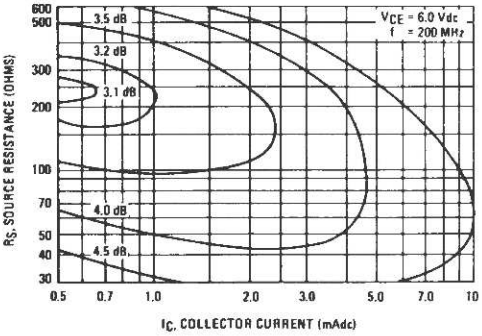


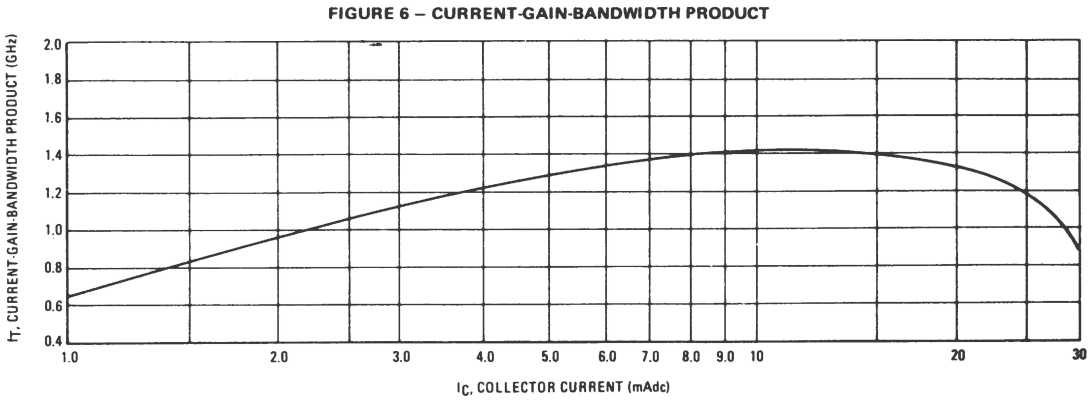
FIGURE 5 – NOISE FIGURE versus SOURCE RESISTANCE and COLLECTOR CURRENT

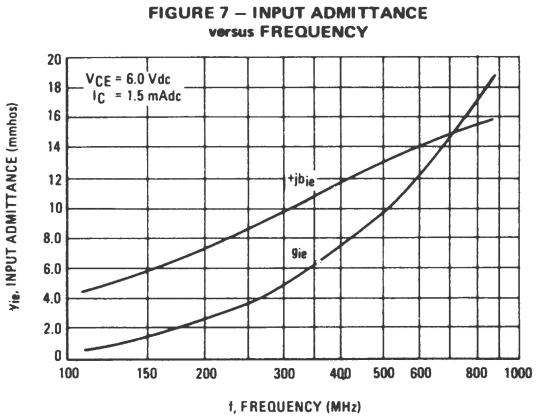


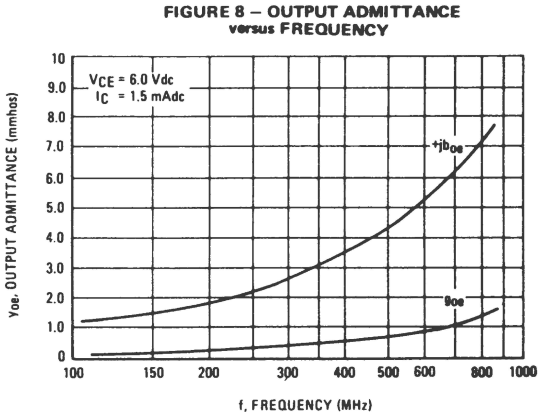
NPN RF Transistor

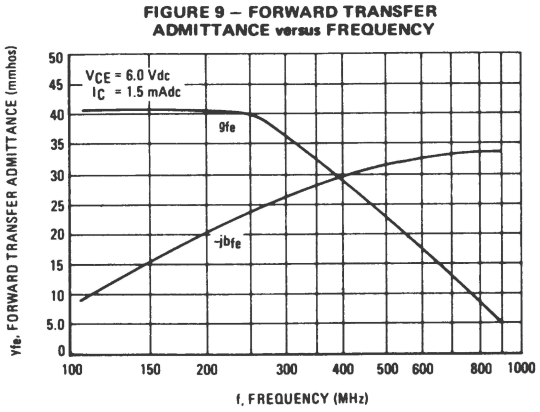
FIG. 5-17. (Continued)

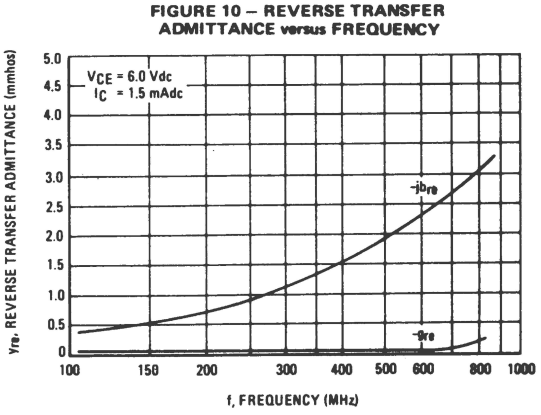
2N5179











NPN RF Transistor

FIG. 5-17. (Continued)

2N5179

FIGURE 11— S_{11}, INPUT REFLECTION COEFFICIENT

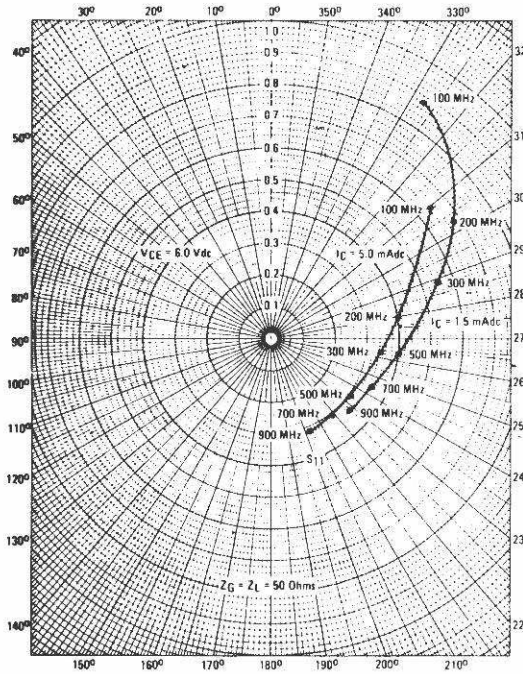


FIGURE 12—S_{22}, OUTPUT REFLECTION COEFFICIENT

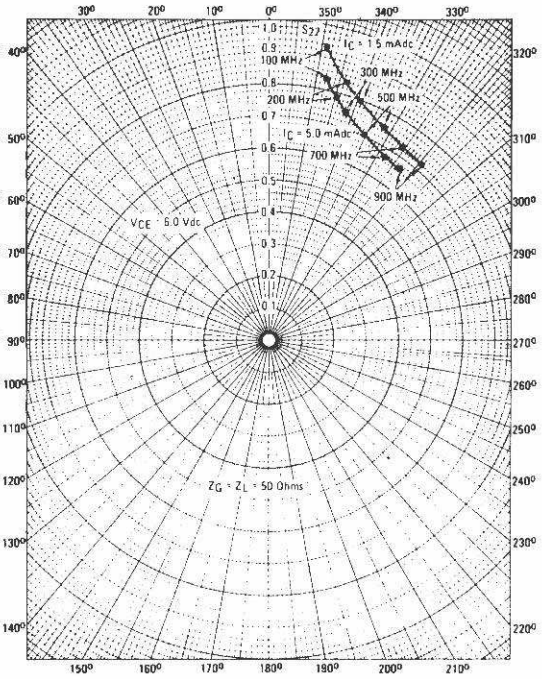


FIGURE 13—S_{12}, REVERSE TRANSMISSION COEFFICIENT

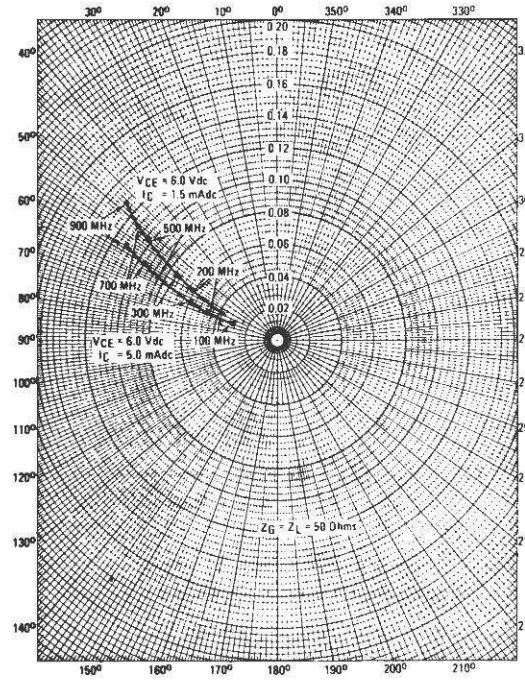
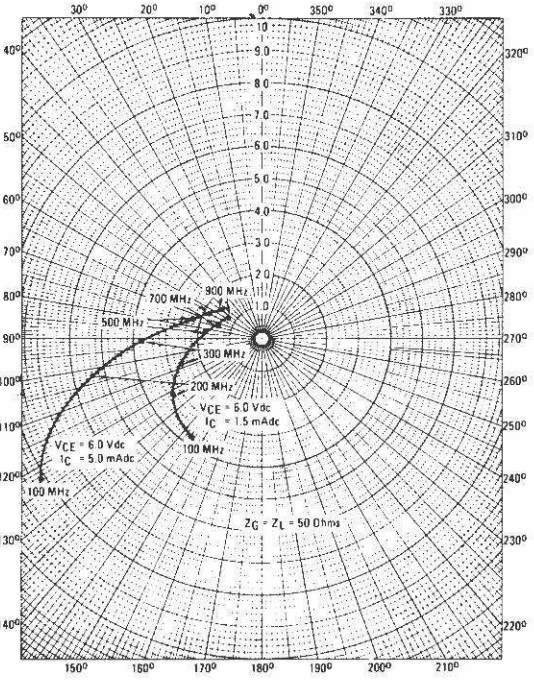


FIGURE 14—S_{21}, FORWARD TRANSMISSION COEFFICIENT

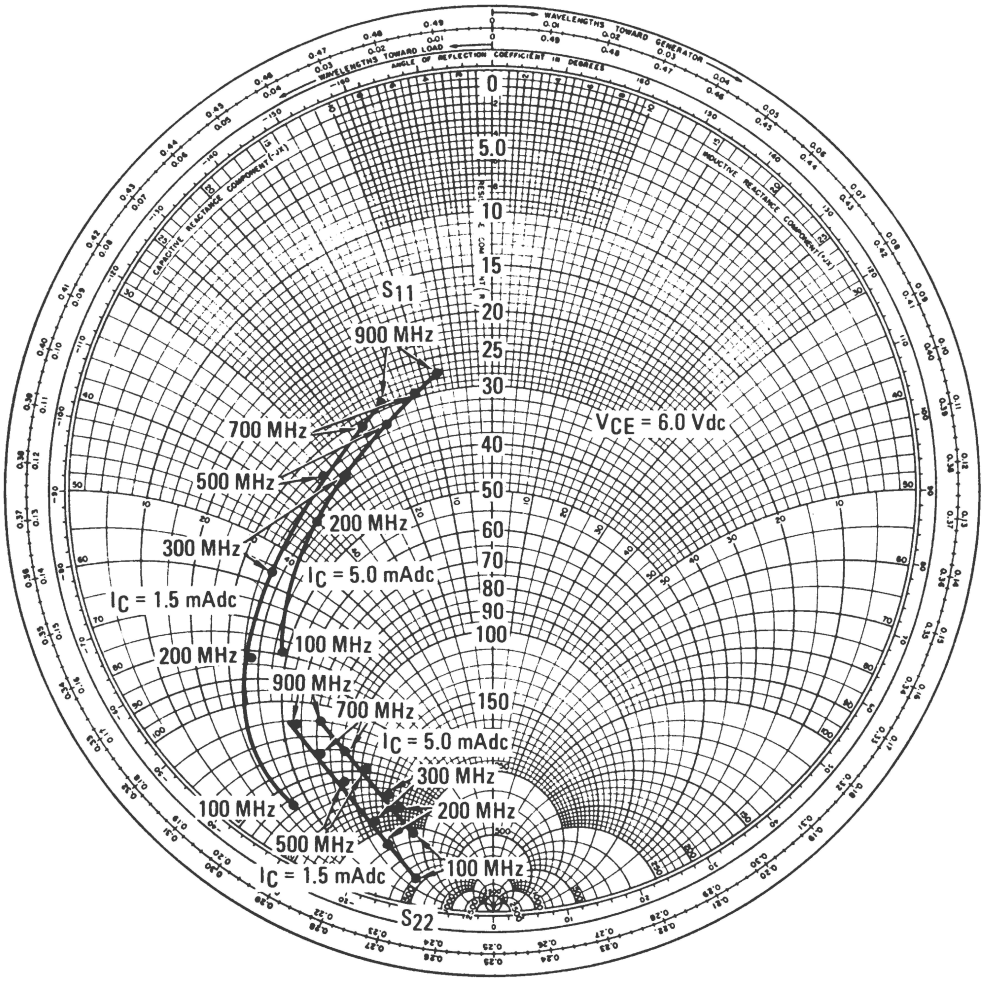


NPN RF Transistor

FIG. 5-17. (Continued)

2N5179

FIGURE 15 – S_{11}, INPUT REFLECTION COEFFICIENT AND S_{22}, OUTPUT REFLECTION COEFFICIENT



NPN RF Transistor

FIG. 5-17. (Continued)

Figure 4 on the data sheet is a plot of noise figure versus both collector current and source resistance for the transistor at a V_{CE} of 6 volts and an operating frequency of 105 MHz. It is obvious from the diagram that there are an infinite number of R_s, I_c combinations that will provide you with a specified noise figure. For example, the following combinations of R_s and I_c will provide you with a 3.5-dB noise figure:

I_c (mA)	R_s (ohms)
0.5	105 or 600
1.0	90 or 500
1.5	85 or 430
2.0	82 or 390
3.0	81 or 320
5.0	94 or 250

Notice that for each value of collector current there are two values of source resistance that will provide the specified noise figure. Obviously, any variation from the intended bias current or source resistance could change the noise figure drastically.

Figure 5 is simply another set of contours measured at the same bias levels but at a different frequency (200 MHz). If you intended to use the transistor at 300 MHz and wanted to know what bias current and source resistance to use for a specific noise figure, you would be out of luck. There are no noise contours provided for that frequency.

Figure 6, on page 4 of the data sheet, is a graph of f_T versus collector current. Optimum f_T is obtained at the peak of the curve which occurs at approximately 12 mA of collector current. This graph becomes more important at frequencies close to f_T, when you are trying to squeeze every last bit of gain out of the device that you possibly can. It will indicate the optimum collector current at which to operate the device. Once the value of collector current is defined, a sample device could then be biased accordingly and its Y or S parameters measured so that the design could proceed.

Figures 7, 8, 9, and 10 are a graphical presentation of the Y parameters versus frequency for the 2N5179. Measurements were plotted at $V_{CE} = 6$ volts and $I_c = 1.5$ mA. If you prefer a different set of bias conditions, you will have to measure your own Y parameters as no other data is provided.

The vertical axis of each diagram is calibrated in millimhos (mmhos). Therefore, the input admittance (Figure 7) of the 2N5179 at 200 MHz is approximately $y_i = 2.5 + j7.5$ mmhos, which can be represented by the circuit of Fig. 5-15A. Remember that positive susceptance ($+jB$) indicates a shunt capacitor while negative susceptance ($-jB$) indicates an inductor. Similarly, the *output* admittance of the transistor at 200 MHz is read in Figure 8 as $y_o = 0.25 + j1.8$ mmhos. The equivalent circuit for this output admittance is shown in Fig. 5-18B. The forward and reverse transfer admittances for the transistor are given in Figures 9 and 10, respectively, on the data sheet.

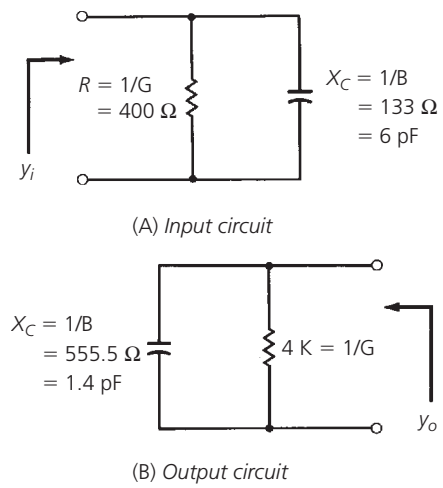


FIG. 5.18. Equivalent circuit for a 2N5179 (at 200 MHz).

In Chapter 6, you will learn how to apply the given Y parameter information from data sheets to the design of RF small-signal amplifiers.

Figures 11, 12, 13, 14, and 15 on the data sheet are plots of the transistor's S parameters versus frequency at two different bias levels: $V_{CE} = 6.0$ volts, $I_c = 1.5$ mA and $V_{CE} = 6.0$ volts, $I_c = 5.0$ mA. The four plots shown on page 5 of the data sheet provide S-parameter data in polar form. The radial distance outward from the center of the chart is equal to the magnitude; and the angle is read along the perimeter of the chart. For example, the S parameters for the 2N5179 with a bias of $V_{CE} = 6$ volts, $I_c = 5$ mA, at 100 MHz, are:

$$S_{11} = 0.65\angle 309^\circ$$
$$S_{22} = 0.84\angle 348^\circ$$
$$S_{12} = 0.03\angle 70^\circ$$
$$S_{21} = 8.2\angle 123^\circ$$

Parameters S_{21} and S_{12} are the forward and reverse gain of the device in magnitude form. To find the gain in dB, simply take the logarithm of the number and multiply it by 20.

$$S_{12}(\text{dB}) = 20\log_{10} 0.03 = -30.5\,\text{dB}$$
$$S_{21}(\text{dB}) = 20\log_{10} 8.2 = 18.3\,\text{dB}$$

From the preceding calculations, we can deduce that the output port to input port isolation (S_{12}) of the transistor is very good at -30.5 dB. Also, the gain of the transistor (S_{21}), when driven with a 50-ohm source and terminated in a 50-ohm load (even without impedance matching), is better than 18 dB. Notice that each gain was calculated as a voltage gain. In actuality, voltage and power gains are identical in this instance because the input and the output impedance levels are the same (50 ohms).

Figure 15 on the last page of the data sheet is another plot of the input and output reflection coefficients of the transistor. This time, however, a Smith Chart is used. As was stated previously in this chapter, S_{11} and S_{22} are simply reflection coefficients and can be plotted just like any other reflection coefficient. Once the information is plotted, the input and output impedance of the device can be read directly from the chart.

The chart shown in Figure 15 is a bit different from the charts we have used thus far in the book. It has been normalized to 50 ohms rather than the usual 1-ohm. Thus, the center of the chart now represents $50 \pm j0$ ohms rather than $1 \pm j0$. This type of chart normalization is often used when the impedances that the designer is working with tend to concentrate around a certain value—in this case, 50 ohms.

The input impedance of the 2N5179 as read from the Smith Chart, at 100 MHz with a collector current of 5 mA and $V_{CE} = 6$ volts, is:

$$Z_{\text{in}} = 48 - j79 \text{ ohms}$$

This agrees, within reading accuracy, with the polar plot of S_{11} (Figure 11) under the same operating conditions and can be verified numerically by plugging S_{11} in Equation 5-8 for Γ and solving for Z_L.

SUMMARY

The transistor is no different from any other component when it comes to misbehaving at radio frequencies. Like other components, the transistor, too, has stray inductance and capacitance which tends to limit its high-frequency performance. Y and S parameters were devised as a means of presenting this complex transistor behavior over frequency with a minimum of effort. Manufacturers typically present the Y- and S-parameter information in the form of data sheets, which should be considered only as a starting point when used in any RF design task. Manufacturers cannot possibly hope to provide Y- and S-parameter information at every conceivable bias point and in every possible circuit configuration. Instead, they usually try to provide a set of *typical* operating conditions for the device in question. Inevitably, however, the day will come when the data sheet provided with a device is of no use to you whatsoever, and you will find yourself measuring your own parameters and creating your own data sheets in order to complete a design task.

CHAPTER 6

SMALL-SIGNAL RF Amplifier Design

The design of RF small-signal amplifiers is a step-by-step logical procedure with an exact solution for each problem. There are many books available on the market today that offer schematics (complete with parts values) which are "adaptable to any of your circuit needs." That is, a circuit that the author may have designed for a specific set of operating conditions is offered and it may or may not meet our needs. Nonetheless, the design is presented without any design procedure attached, and the reader is left out in the cold when he tries to adapt the circuit to his particular set of operating conditions.

The chapter presented here, however, takes the opposite approach. Detailed step-by-step procedures are followed in the design process so that you can choose the transistor you want and use it under any (realistic) operating conditions that you desire. You will no longer have to adapt someone else's schematic to your needs. Rather you will create your own *homemade* RF amplifiers and optimize them for your personal application.

We will begin our discussion with a very brief overview of transistor biasing. We will discuss both the bipolar and the field-effect transistor (FET). As was shown in the last chapter, the quiescent bias point of a transistor has a great effect on its Y and S parameters. Biasing a transistor is, therefore, serious business and should not be taken lightly.

Next, we'll jump head first into the RF aspect of amplifiers by examining stability (tendency for oscillation), gain, impedance matching, and general amplifier design, with emphasis on the use of Y and S parameters as a design tool

SOME DEFINITIONS

To aid in this discussion, let's first take a closer look at the two types of transistors used in small signal design.

1. *Bipolar Transistor*—A bipolar or bipolar junction transistor (BJT) is a three-terminal semiconductor device commonly used for amplification of analog or digital signals. It is constructed of doped sections of semiconductor material sandwiched together. The center section is called the base of the transistor. By varying the current between the base and one terminal called the emitter, one can vary the current flow between the emitter and a third terminal known as the collector, causing amplification of the signal at that terminal.

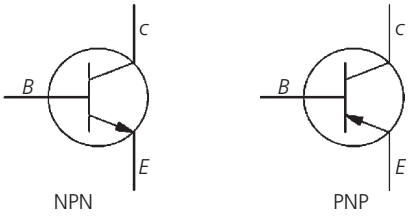


FIG. 6-1. The schematic symbols for PNP- and NPN-type BJTs, courtesy of Wikipedia (www.wikipedia.org/wiki/Bipolar_junction_transistor).

There are two major types of bipolar transistor: PNP and NPN (Fig. 6-1). A PNP transistor has a layer of N-type semiconductor between two layers of P-type material. An NPN transistor has a layer of P-type material between two layers of N-type material. In P-type material, electric charges are carried mainly in the form of electron deficiencies called holes. In N-type material, the charge carriers are primarily electrons. Of the two types of BJTs, the NPN is more commonly employed as it provides better performance (e.g., allowing greater currents and faster operation) in most circumstances.

As an example, consider Fig. 6-2. During typical operation of the transistor, the emitter–base junction is forward biased and the base–collector junction is reverse biased. When a positive voltage is applied to the base–emitter junction, the equilibrium between thermally generated carriers and the repelling electric field of the depletion region becomes unbalanced, allowing thermally excited electrons to inject into the base region. These electrons wander (or diffuse) through the base from the region of high concentration near the emitter towards the region of low concentration near the collector.

The collector–base junction is reverse-biased, so little electron injection occurs from the collector to the base.

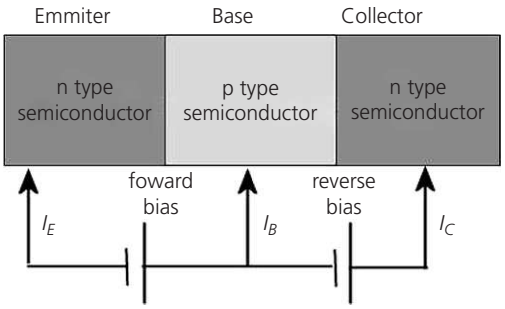


FIG. 6-2. NPN BJT with forward-biased E–B junction and reverse-biased B–C junction, courtesy of Wikipedia (www.wikipedia.org/wiki/Bipolar_junction_transistor).

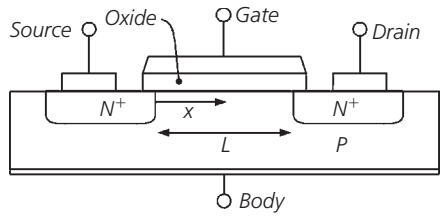


FIG. 6-3. Cross-section of an N-type MOSFET, courtesy of Wikipedia (www.wikipedia.org/wiki/Field_effect_transistor).

Electrons that diffuse through the base towards the collector are swept into the collector by the electric field in the depletion region of the collector–base junction.

The bipolar transistor has both advantages and disadvantages relative to the field-effect transistor. Bipolar devices can switch signals at high speeds. And, they can be manufactured to handle large currents in order to serve as high-power amplifiers in audio equipment and in wireless transmitters. On the other hand, they are not as effective as FETs for weak-signal amplification, or for applications requiring high circuit impedance.

Note that an improvement to the bipolar transistor is the heterojunction bipolar transistor (HBT). It can handle signals of very high frequencies up to several hundred gigahertz and is therefore commonly used nowadays in ultrafast circuits, mostly RF systems. One of the most popular such devices is the silicon–germanium (SiGe) HBT. Because it is compatible with standard silicon digital processes it allows integration of very high speed circuitry with complex lower speed digital circuitry.

2. *Field-effect Transistor*—The field-effect transistor or FET is a type of transistor commonly used for weak-signal amplification, such as for amplifying wireless signals. Like the BJT, it can amplify analog or digital signals. It can also switch DC or function as an oscillator.

The FET relies on an electric field to control the shape and therefore the conductivity of a path or channel in a semiconductor material. During operation, current flows along a semiconductor path called the channel. At one end of the channel, there is an electrode (source), and at the other end is another electrode (drain). The channel's physical diameter is fixed, but its effective electrical diameter can be varied by the application of a voltage to a control electrode known as the gate. This gate permits electrons to flow through or blocks their passage by creating or eliminating a channel between the source and drain. The conductivity of the FET depends, at any given instant in time, on the electrical diameter of the channel. A small change in gate voltage can cause a large variation in the current from the source to the drain. This is how the FET amplifies signals.

There are two major classes of FETs: junction FET (JFET) and the metal-oxide- semiconductor FET (MOSFET). All FETs (e.g., MOSFETs, MESFETs, MODFETs, and IGBTs) except J-FETs have four terminals (gate, drain, source and body/base/bulk). The JFET has no body terminal.

The JFET uses a reverse biased P-N junction to separate the gate from the body. Its channel consists of N-type semiconductor (N-channel) or P-type semiconductor (P-channel) material, while the gate is made of the opposite semiconductor type. In P-type material, electric charges are carried mainly in the form of electron deficiencies called holes. In N-type material, the charge carriers are primarily electrons. In a JFET, the junction is the boundary between the channel and the gate. Normally, this P-N junction is reverse-biased (a DC voltage is applied to it) so that no current flows between the channel and the gate. However, under some conditions there is a small current through the junction during part of the input signal cycle.

The most commonly used FET today is the MOSFET; it is used in everything from cellular handsets to wireless base stations. The MOSFET has a channel which can be either N-type or P-type semiconductor. Fig. 6-3 provides an example of an N-type MOSFET. The gate electrode is a piece of metal whose surface is oxidized. The thin oxide layer (typically SiO_2) electrically insulates the gate from the channel or body. Because the oxide layer acts as a dielectric, there is essentially never any current between the gate and the channel during any part of the signal cycle. As a result, the MOSFET has extremely large input impedance.

As compared to the BJT, FETs are preferable for use in circuits and systems requiring high impedance, as well as for weak-signal work in wireless communications and broadcast applications. In general, FETs are not used for high-power amplifications, such as is required in large wireless communications and broadcast transmitters.

TRANSISTOR BIASING

In most RF amplifier designs, unfortunately, very little thought is ever given to the design of bias networks for the individual transistors involved. Often, the lack of interest in bias networks may be justified. If, for instance, the amplifier is to be operated only at room temperature, there would be little need to spend much time developing an extremely temperature-stable DC operating point. If, on the other hand, the amplifier must operate reliably and maintain certain specifications (gain, noise figure, etc.) over large temperature extremes, the DC bias network *must* be carefully considered. Consider, for example, the 2N5179 data sheet from Freescale presented in the last chapter. A quick look at the Y- and S-parameter curves for the device will reveal that a change in the transistor's bias point does in fact change all of its RF operating characteristics. It only stands to reason, then, that the DC operating point must remain stable under your specified operating conditions or the RF characteristics may change drastically.

It has been shown that there are two basic internal transistor characteristics that have a profound effect upon the transistor's DC operating point over temperature; they are ΔV_{BE} and $\Delta\beta$. The object of a good temperature-stable bias design (see Fig. 6-4) is to minimize the effects of these parameters.

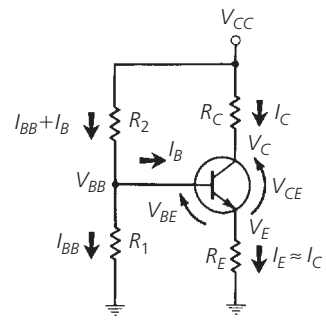
As the temperature *increases*, the base-to-emitter voltage (V_{BE}) of a transistor *decreases* at the rate of about 2.5 mV/°C from its nominal room-temperature value of 0.7 V (for a silicon as opposed to a CMOS or SiGe device). As V_{BE} decreases, more base current is allowed to flow which, in turn, produces more collector current and that is exactly what we would like to prevent. The total change in V_{BE} for a given temperature change is called ΔV_{BE}. The primary external circuit factor that the circuit designer has control over, and which tends to minimize the effects of ΔV_{BE}, is the emitter voltage (V_E) of the transistor. This is shown in Fig. 6-4. Here, a decrease in V_{BE} with temperature would cause an increase in emitter current and, hence, an increase in V_E. The increase in V_E is a form of negative feedback that tends to reverse bias the base-emitter junction and, therefore, *decrease* the collector current. A decrease in V_{BE}, therefore, tends to be counteracted by the increase in V_E, and the collector current does not increase as much with temperature. If these observations were put into equation form, we would have:

$$\Delta I_C \approx -\frac{\Delta V_{BE} I_C}{V_E} \qquad \text{(Eq. 6-1)}$$

where

ΔI_C = the change in collector current,
I_C = the quiescent collector current,
ΔV_{BE} = the change in base-to-emitter voltage,
V_E = the quiescent emitter voltage.

Thus, if V_E were made equal to 20 times ΔV_{BE}, the collector current would change only 5% over temperature due to ΔV_{BE}. It is important to note that it is the value of the emitter voltage (V_E) and not the value of the emitter resistor (R_E) that is the important bias-design criteria.



1. Choose the operating point for the transistor.
$$I_c = 10\,\text{mA},\ V_c = 10\,\text{V},\ V_{cc} = 20\,\text{V},\ \beta = 50$$
2. Assume a value for V_E that considers bias stability:
$$V_E = 2.5\,\text{volts}$$
3. Assume $I_E \approx I_c$ for high-beta transistors.
4. Knowing I_E and V_E, calculate R_E.
$$R_E = \frac{V_E}{I_E} = \frac{2.5}{10 \times 10^{-3}} = 250\,\text{ohms}$$
5. Knowing V_{cc}, V_c, and I_c, calculate R_c
$$R_c = \frac{V_{cc} - V_c}{I_c} = \frac{20 - 10}{10 \times 10^{-3}} = 1000\,\text{ohms}$$
6. Knowing I_c and β, calculate I_B.
$$I_B = \frac{I_c}{\beta} = 0.2\,\text{mA}$$
7. Knowing V_E and V_{BE}, calculate V_{BB}.
$$V_{BB} = V_E + V_{BE} = 2.5 + 0.7 = 3.2\,\text{volts}$$
8. Assume a value for I_{BB}, the larger the better (see text):
$$I_{BB} = 15\,\text{mA}$$
9. Knowing I_{BB} and V_{BB}, calculate R_1.
$$R_1 = \frac{V_{BB}}{I_{BB}} = \frac{3.2}{1.5 \times 10^{-3}}\,\text{ohms} = 2133\,\text{ohms}$$
10. Knowing V_{CC}, V_{BB}, I_{BB}, and I_B, calculate R_2.
$$R_2 = \frac{V_{CC} - V_{BB}}{I_{BB} + I_B} = \frac{20 - 3.2}{1.7 \times 10^{-3}} = 9882\,\text{ohms}$$

FIG. 6-4. Bias network design 1.

Equation 6-1 tends to imply that the higher V_E is, the better. This would be exactly true if we had nothing to worry about except biasing the transistor for the specified operating point. Obviously, there are other things that must be considered in the design. A high emitter voltage, for instance, does tend to waste power and decrease the AC signal gain. A bypass capacitor across R_E at the signal frequency is usually used to prevent the loss in gain, but the wasted power may still be a problem.

If we assume that the amplifier is to operate over a change in temperature of no more than ±50°C, then an emitter voltage of 2.5 V will provide a ±5% variation in I_C due to ΔV_{BE}. In fact, you will find that the majority of the transistor bias networks that are similar to Fig. 6-4 will provide a value of V_E from two to four volts depending upon the values of V_{CC} and V_C chosen. Higher values are, of course, possible depending upon the degree of stability you need.

The change in a transistor's DC current gain, or β, over temperature, is also of importance to the circuit designer. Any variation in β will produce a corresponding change in quiescent collector current and will, therefore, disrupt the transistor's designed operating point. The β of a silicon transistor typically *increases* with temperature at the rate of about 0.5% per °C. Thus, for a ±50°C temperature variation you can expect the β of the transistor and, hence, its collector current to vary as much as ±25%.

Not only does β vary with *temperature*, but the manufacturing tolerance for β among transistors of the same part number is typically very poor. It is not uncommon, for instance, for a manufacturer to specify a 10 to 1 range for β on the data sheet (such as 50 to 500). This, of course, makes it extremely difficult to design a bias network for the device in question when it is to be used in a production environment. Thus, a stable operating point with respect to β is difficult to obtain from a production standpoint as well as from a temperature standpoint.

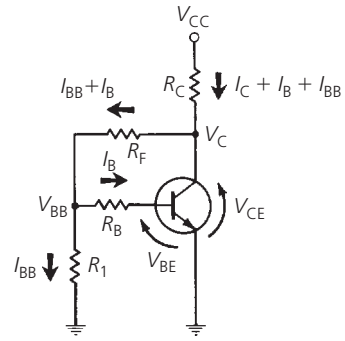
The change in collector current for a corresponding change in β can be approximated as:

$$\Delta I_C = I_{C1}\left(\frac{\Delta\beta}{\beta_1\beta_2}\right)\left(1+\frac{R_B}{R_E}\right) \qquad \text{(Eq. 6-2)}$$

where

I_{C1} = the collector current at $\beta = \beta_1$,
β_1 = the lowest value of β,
β_2 = the highest value of β,
$\Delta\beta = \beta_2 - \beta_1$
R_B = the parallel combination of R_1 and R_2 (in Fig. 6-4),
R_E = the emitter resistor.

This equation indicates that once a transistor is specified, the only control that the designer has over the effect of β changes on collector current is through the resistance ratio R_B/R_E. The *smaller this ratio, the less the collector current varies*. Again, however, some compromise is necessary. As you decrease the ratio R_B/R_E, you also produce the undesirable effect of decreasing the current gain of the amplifier. Also, as the ratio approaches unity, the improvement in operating-point stability rapidly decreases.



1. Choose the operating point for the transistor.

$$I_c = 10\,\text{mA},\ V_c = 10\,\text{V},\ V_{cc} = 20\,\text{V},\ \beta = 50$$

2. Assume values for V_{BB} and I_{BB} to supply a constant current, I_B.

$$V_{BB} = 2\ \text{volts}$$
$$I_{BB} = 1\ \text{mA}$$

3. Knowing I_c and β, calculate I_B.

$$I_B = \frac{I_C}{\beta} = 0.2\ \text{mA}$$

4. Knowing V_{BB}, $V_{BE} = 0.7$ V, and I_B calculate R_B.

$$R_B = \frac{V_{BB} - V_{BE}}{I_B} = \frac{2 - 0.7}{0.2 \times 10^{-3}} = 6500\ \text{ohms}$$

5. Knowing V_{BB} and I_{BB}, calculate R_1.

$$R_1 = \frac{V_{BB}}{I_{BB}} = \frac{2}{1 \times 10^{-3}} = 2000\ \text{ohms}$$

6. Knowing V_{BB}, I_{BB}, I_B, and V_C, calculate R_F.

$$R_F = \frac{V_C - V_{BB}}{I_{BB} + I_B} = \frac{10 - 2}{1.2 \times 10^{-3}} = 6667\ \text{ohms}$$

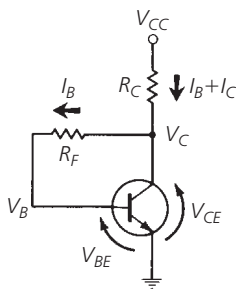
7. Knowing V_{CC}, V_C, I_C, I_B and I_{BB}, calculate R_C.

$$R_C = \frac{V_{CC} - V_C}{I_C + I_B + I_{BB}} = \frac{20 - 10}{11.2 \times 10^{-3}} = 893\ \text{ohms}$$

FIG. 6-5. Bias network design 2.

As a practical rule of thumb for stable designs, the ratio R_B/R_E should be less than 10.

Figs. 6-4, 6-5, and 6-6 indicate three possible bias configurations for bipolar transistors—in order of *decreasing* bias stability. Complete step-by-step design instructions using a typical



1. Choose the operating point for the transistor (V_C, I_C).

$$I_C - 10\,\text{mA},\ V_C = 10\,\text{V},\ V_{CC} = 20\,\text{V},\ \beta = 50$$

2. Knowing I_C and β, calculate I_B.

$$I_B = \frac{I_C}{\beta} = 0.2\,\text{mA}$$

3. Knowing V_C, $V_B = V_{BE} = 0.7\,\text{V}$, and I_B, calculate R_F.

$$R_F = \frac{V_C - V_B}{I_B} = \frac{10 - 0.7}{200 \times 10^{-6}} = 46.5\,\text{K}$$

4. Knowing I_B, I_C, V_{CC}, and V_C, calculate R_C.

$$R_C = \frac{V_{CC} - V_C}{I_B + I_C} = \frac{20 - 10}{10.2 \times 10^{-3}} = 980\,\text{ohms}$$

FIG. 6-6. Bias network design 3.

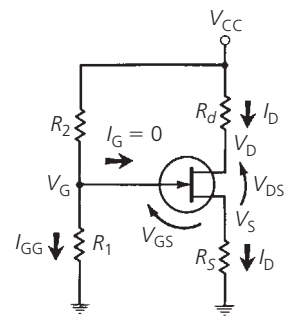
example are included with each circuit-configuration sketch. Note that the bias networks of Figs. 6-5 and 6-6 do not contain the emitter resistor (R_E) which provides the negative feedback needed to counteract collector-current variations over temperature. Instead, resistor R_F is connected from the collector to the base of the transistor to provide the negative feedback. Obviously, for these two designs, the designer has control over neither the ratio R_B/R_E nor the voltage V_E of Fig. 6-4. The designs are, therefore, of the "potluck" variety as far as DC stability is concerned. You basically take what you get. Surprisingly, however, R_F works quite well in minimizing the effects of transistor-parameter variations over temperature.

Figs. 6-7 and 6-8 show similar bias arrangements and design procedures for a field-effect transistor (FET). These are based on the well-known formula:

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_p}\right)^2 \qquad \text{(Eq. 6-3)}$$

where

I_D = the drain current,
I_{DSS} = the drain current with $V_{GS} = 0$,
V_{GS} = the gate-to-source voltage,
V_p = the pinch-off voltage.



1. Choose the operating point for the transistor.

$$I_D = 10\,\text{mA},\ V_D = 10\,\text{V}, V_{CC} = 20\,\text{V}$$

2. Knowing V_{CC}, V_D, and I_D, calculate R_d.

$$R_d = \frac{V_{CC} - V_D}{I_D} = \frac{10\,\text{V}}{10\,\text{mA}} = 1000\,\text{ohms}$$

3. Determine V_P and I_{DSS} from the data sheet.

$$V_P = -6\,\text{volts}$$
$$I_{DSS} = 5\,\text{mA}$$

4. Knowing I_D, I_{DSS}, and V_P, calculate V_{GS}.

$$V_{GS} = V_P\left(1 - \sqrt{\frac{I_D}{I_{DSS}}}\right) = -6\left(1 - \sqrt{\frac{10 \times 10^{-3}}{5 \times 10^{-3}}}\right) = 2.48\,\text{volts}$$

5. Assume a value for V_S in the 2- to 3-volt range.

$$V_S = 2.5\,\text{volts}$$

6. Knowing V_S and I_D, calculate R_S.

$$R_S = \frac{V_S}{I_D} = \frac{2.5}{10 \times 10^{-3}} = 250\,\text{ohms}$$

7. Knowing V_S and V_{GS}, calculate V_G.

$$V_G = V_{GS} + V_S = 2.48 + 2.5 = 4.98\,\text{volts}$$

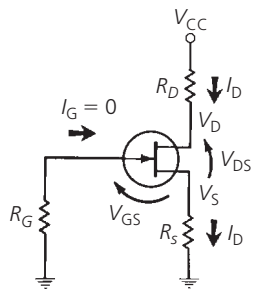
8. Assume a value for R_1 based upon dc input resistance needs.

$$R_1 = 220\,\text{K}$$

9. Knowing R_1, V_G, and V_{CC}, calculate R_2.

$$R_2 = \frac{R_1(V_{CC} - V_G)}{V_G} = \frac{220 \times 10^3(20 - 4.98)}{4.98} = 664\,\text{K}$$

FIG. 6-7. Bias network design 4.



1. Choose an operating point for the transistor.

$$I_D = 10\,\text{mA},\ V_D = 10\,\text{V},\ V_{CC} = 20\,\text{V}$$

2. Knowing V_{CC}, V_D, and I_D, calculate R_D.

$$R_D = \frac{V_{CC}}{V_D} = \frac{20 - 10}{10 \times 10^{-3}} = 1000\,\text{ohms}$$

3. Determine V_P and I_{DSS} from the transistor data sheet.

$$V_P = -6\,\text{volts}$$

$$I_{DSS} = 5\,\text{mA}$$

4. Knowing I_D, I_{DSS}, and V_P, calculate V_{GS}.

$$V_{GS} = V_P\left(1 - \sqrt{\frac{I_D}{I_{DSS}}}\right) = -6\left(1 - \sqrt{\frac{10 \times 10^{-3}}{5 \times 10^{-3}}}\right) = 2.48\,\text{volts}$$

5. Knowing $I_G = 0$, $V_{GS} = V_S$, and I_D, calculate R_S.

$$R_S = \frac{V_S}{I_D} = \frac{V_{GS}}{I_D} = \frac{2.48}{10 \times 10^{-3}} = 248\,\text{ohms}$$

6. Sine $I_G = 0$, R_G can be chosen to be any large value of resistor–approximately 1 megohm.

FIG. 6-8. Bias network design 5.

I_D is usually a value chosen by the user as part of the bias specifications, and I_{DSS} and V_p can be found on the data sheet for the transistor. Once these three values are known, Equation 6-3 can be used to solve for V_{GS}, and a suitable bias circuit can then be found.

DESIGN USING Y PARAMETERS

The RF small-signal performance of a transistor can be completely characterized by its two-port admittance parameters. Based on these parameters, equations can be written to aid you both in finding a transistor to suit your needs and in completing the design once the transistor is selected.

One of the first requirements in any amplifier design is to choose the transistor that is best suited for the job. Many RF amplifier designs are doomed from the beginning simply because the active device chosen for the job should have never been considered. Spend a little time shopping for the right device for your application. The more time you spend shopping prior to the start of the actual design, the less hair-pulling there will be later. Two of the most important considerations, in choosing a transistor for use in any amplifier design, are its stability and its maximum available gain (MAG). Stability, as it is used here, is a measure of the transistor's tendency toward oscillation. MAG is a type of figure-of-merit for the transistor, which indicates the maximum theoretical power gain you can expect to obtain from the device when it is *conjugately* matched to its source and load impedance. The MAG is never actually reached in practice; nevertheless, it is quite useful in gauging the capabilities of a transistor.

Stability Calculations

It has been said that one of the easiest methods of building an oscillator is to design an amplifier. Although experience has found this to be true, it really need not be the case. A bit of prior planning and basic *a priori* knowledge about the transistor that is to be used can go a long way toward preventing oscillations in any amplifier design.

It is possible to predict the degree of stability (or lack thereof) of a transistor before you actually place the device in a circuit. This is done through a calculation of the Linvill stability factor, C.

$$C = \frac{|y_r y_f|}{2g_i g_o - \text{Re}(y_r y_f)} \qquad \text{(Eq. 6-4)}$$

where

$|\ \ | =$ the magnitude of the product in brackets,
$y_r =$ the reverse-transfer admittance,
$y_f =$ the forward-transfer admittance,
$g_i =$ the input conductance,
$g_o =$ the output conductance,
Re $=$ the real part of the product in parentheses.

When C is *less than* 1, the transistor is *unconditionally stable at the bias point you have chosen*. This means that you could choose any possible combination of source and load impedance for the device, and the amplifier would remain stable *providing* that no external feedback paths exist which have not been accounted for.

If C is *greater than* 1, the transistor is *potentially unstable* and will oscillate for certain values of source and load impedance. A C-factor greater than 1 does not indicate, however, that the transistor cannot be used as an amplifier. It merely indicates that you must exercise extreme care in choosing your source and load impedances or oscillations may occur.

The Linvill stability factor is useful in predicting a *potential* stability problem. It does not indicate the actual impedance values

between which the transistor will go unstable. Obviously, if a transistor is chosen for a particular design problem, and the transistor's C-factor is less than 1 (unconditionally stable), it will be much easier to work with than a transistor which is potentially unstable. Keep in mind also that if C is less than *but very close to* 1 for any transistor, then any change in the bias point due to temperature variation could cause the transistor to become potentially unstable and most likely oscillate at some frequency. This is because Y parameters are specified *at a particular bias point* which varies with temperature. This is a very important concept to remember. The smaller C is, the better.

Y parameters can also be used to predict the stability of an amplifier given certain values of load and source impedance. This is called the Stern stability factor and is given by

$$K = \frac{2(g_i + G_S)(g_o + G_L)}{|y_r y_f| + \mathrm{Re}(y_r y_f)} \qquad \text{(Eq. 6-5)}$$

where

$G_S =$ the source conductance,
$G_L =$ the load conductance.

In this case, if K is *greater than* 1, the circuit will be stable for that value of source and load impedance. If K is *less than* 1, the circuit is potentially unstable and will most likely oscillate at some frequency. Note that the K-factor is a more definitive calculation for stability in that it predicts stability for a particular circuit. The C-factor, on the other hand, predicts a kind of nebulous possibility for instability without giving you an indication as to where the instability may occur.

The Linvill stability factor is, therefore, useful in finding stable transistors while the Stern stability factor predicts possible stability problems with circuits.

Maximum Available Gain

The MAG of a transistor can be found by using the following equation:

$$\mathrm{MAG} = \frac{|y_f|^2}{4 g_i g_o} \qquad \text{(Eq. 6-6)}$$

MAG is a useful calculation in the initial search for a transistor for any particular application. It will give you a good indication as to whether or not the transistor can provide enough gain for the task.

The maximum available gain for a transistor occurs when $y_r = 0$, and when Y_L and Y_S are the complex conjugates of y_o and y_i, respectively. The condition that y_r must equal zero for maximum gain to occur is due to the fact that under normal conditions, y_r acts as a *negative* feedback path internal to the transistor. With $y_r = 0$, no feedback is allowed and the gain is at a maximum.

In practical situations, it is physically impossible to reduce y_r to zero and, as a result, MAG can never truly be attained. It is possible, however, to *very nearly achieve* the MAG calculated in Equation 6-6 through a *simultaneous conjugate match* of the input and output admittance of the transistor. Thus, Equation 6-6 remains a valid tool in the search for a suitable transistor as long as you understand its limitations. For example, if an amplifier design calls for a minimum power gain of 18 dB at 200 MHz, don't choose a transistor with a calculated MAG of 19 dB. Allow yourself a small margin to cover for realistic values of y_r, component losses in the matching network, and variation in the bias point over temperature.

Simultaneous Conjugate Matching (Unconditionally Stable Transistors)

Optimum power gain is obtained from a transistor when y_i and y_o are conjugately matched to Y_S and Y_L, respectively. As was discussed in Chapter 5, however, the reverse-transfer admittance (y_r) associated with each transistor tends to reflect any impedance changes made at one port back toward the other port, causing a change in that port's impedance characteristics. This makes it very difficult to design good matching networks for a transistor while using only its input and output admittances and totally ignoring the contribution that y_r makes to the transistor's impedance characteristics. Even though Y_L affects the input admittance of the transistor and Y_S affects its output admittance, it is still possible to provide the transistor with a simultaneous conjugate match for maximum power transfer (from source to load) by using the following design equations:

$$G_S = \frac{\sqrt{[2 g_i g_o - \mathrm{Re}(y_f y_r)]^2 - |y_f y_r|^2}}{2 g_o} \qquad \text{(Eq. 6-7)}$$

$$B_S = -j b_i + \frac{\mathrm{Im}(y_f y_r)}{2 g_o} \qquad \text{(Eq. 6-8)}$$

$$G_L = \frac{\sqrt{[2 g_i g_o - \mathrm{Re}(y_f y_r)]^2 - |y_f y_r|^2}}{2 g_i} \qquad \text{(Eq. 6-9)}$$

$$= \frac{G_S g_o}{g_i} \qquad \text{(Eq. 6-10)}$$

$$B_L = -j b_o + \frac{\mathrm{Im}(y_f y_r)}{2 g_i} \qquad \text{(Eq. 6-11)}$$

where

$G_S =$ the source conductance,
$B_S =$ the source susceptance,
$G_L =$ the load conductance,
$B_L =$ the load susceptance,
Im = the imaginary part of the product in parentheses.

The above equations may look formidable but actually they are not—once you have used them a few times. Let's try an example of a simultaneous conjugate match for clarification (Example 6-1).

Note that a simultaneous conjugate match may also be performed using a number of easily available and convenient software design tools. As an example, consider the

following demo which utilizes The MathWorks' RF Toolbox (www.mathworks.com/products/rftoolbox), a specialized MATLAB toolbox for designing and analyzing networks of RF components. In particular, this example uses the Smith Chart plot to find the input and output matching networks that maximize the power delivered to a 50-ohm load. First the demo finds the required transmission line lengths for the single-stub matching networks. Then it cascades the matching networks with the amplifier and visualizes the results.

Step 1. Create an RFCKT.AMPLIFIER Object with the data in the samplebjt2.s2p data file

```
close all;
clear all;
amp = rfckt.amplifier;
read(amp, 'samplebjt2.s2p');
```

EXAMPLE 6-1

A transistor has the following Y parameters at 100 MHz, with $V_{CE} = 10$ volts and $I_C = 5$ mA.

$$y_i = 8 + j5.7 \text{ mmhos}$$
$$y_o = 0.4 + j1.5 \text{ mmhos}$$
$$y_f = 52 - j20 \text{ mmhos}$$
$$y_r = 0.01 - j0.1 \text{ mmho}$$

Design an amplifier which will provide maximum power gain between a 50-ohm source and a 50-ohm load at 100 MHz.

Solution

First, calculate the Linvill stability factor using Equation 6-4.

$$\begin{aligned} C &= \frac{|y_f y_r|}{2g_i g_o - \text{Re}(y_f y_r)} \\ &= \frac{|(52 - j20)(0.01 - j0.1)|}{2(8)(0.4) - \text{Re}[(52 - j20)(0.01 - j0.1)]} \\ &= \frac{5.57}{6.4 - (-1.47)} \\ &= 0.71 \end{aligned}$$

Since C is less than 1, the device is *unconditionally stable* and we may proceed with the design. Had C been greater than 1, however, we would have had to be extremely careful in matching the transistor to the source and load as instability could occur.

The MAG of this transistor is computed with Equation 6-6:

$$\begin{aligned} \text{MAG} &= \frac{|y_f|^2}{4g_i g_o} \\ &= \frac{|52 - j20|^2}{4(8)(0.4)} \\ &= 242.5 \\ &= 23.8 \text{ dB} \end{aligned}$$

The actual gain we can achieve will be somewhat less than this due to y_r and component losses.

Using Equations 6-7 through 6-11, calculate the source and load admittances for a simultaneous conjugate match. For the source, using Equation 6-7:

$$\begin{aligned} G_S &= \frac{\sqrt{[2g_i g_o - \text{Re}(y_f y_r)]^2 - |y_f y_r|^2}}{2g_o} \\ &= \frac{\sqrt{[6.4 + 1.47]^2 - (5.57)^2}}{2(.4)} \\ &= 6.95 \text{ mmhos} \end{aligned}$$

And, with Equation 6-8:

$$\begin{aligned} B_S &= -jb_i + \frac{\text{Im}(y_f y_r)}{2g_o} \\ &= -j5.7 + j\frac{-5.37}{2(.4)} \\ &= -j12.41 \text{ mmhos} \end{aligned}$$

Therefore, the source admittance that the transistor must "see" for optimum power transfer is $6.95 - j12.41$ mmhos. The transistor's actual input admittance is the conjugate of this number, or $6.95 + j12.41$ mmhos. For the load, using Equation 6-10:

$$\begin{aligned} G_L &= \frac{G_S g_o}{g_i} \\ &= \frac{(6.95)(0.4)}{8} \\ &= 0.347 \text{ mmho} \end{aligned}$$

And, with Equation 6-11:

$$\begin{aligned} B_L &= -jb_o + \frac{\text{Im}(y_f y_r)}{2g_i} \\ &= -j1.5 + j\frac{-5.37}{2(8)} \\ &= -j1.84 \text{ mmhos} \end{aligned}$$

Thus, for optimum power transfer, the load admittance must be $0.347 - j1.84$ mmhos. The actual output admittance of the transistor is the conjugate of the load admittance, or $0.347 + j1.84$ mmhos.

Continued on next page

EXAMPLE 6-1—Cont

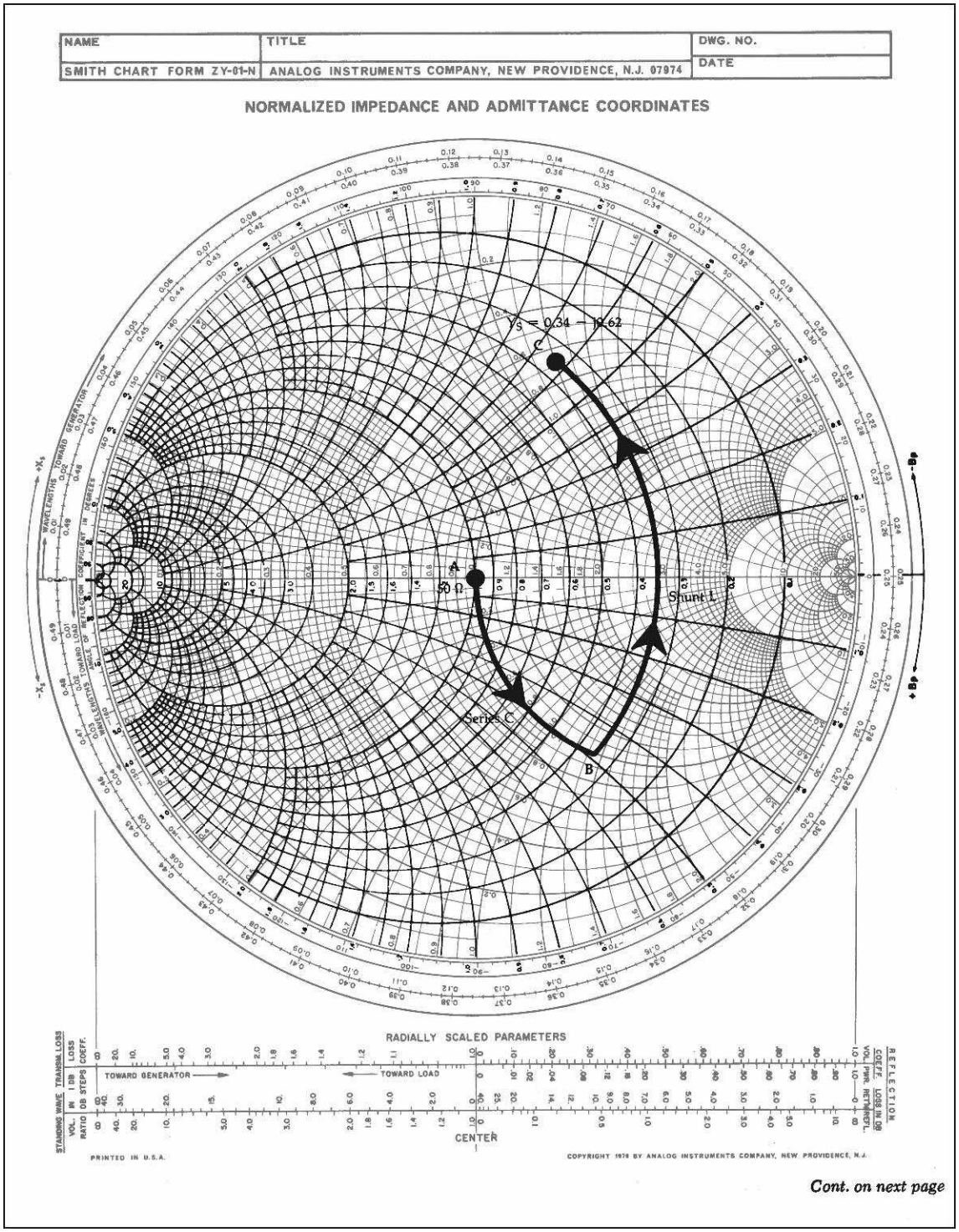


FIG. 6-9. Input network design for Example 6-1. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-1—Cont

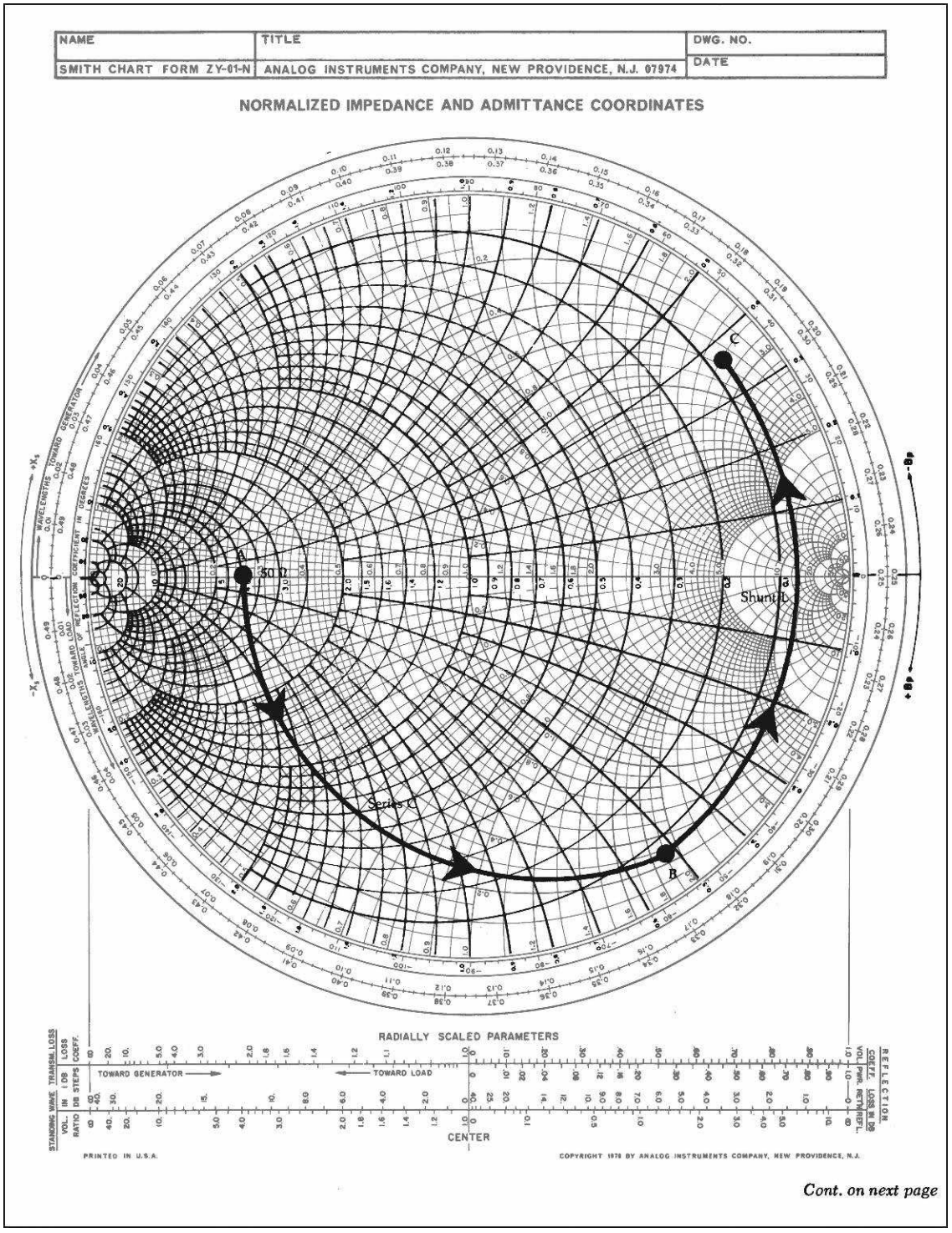


FIG. 6-10. Output network design for Example 6-1. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-1—Cont

The next step is to design the input and output impedance-matching networks that will transform the 50-ohm source and load to the impedance which the transistor would like to see for optimum power transfer. The input matching design is shown on the Smith Chart of Fig. 6-9. This chart is normalized so that the center of the chart represents 50 ohms or 20 mmhos. Thus, the point $Y_S = 6.95 - j12.42$ mmhos is normalized to:

$$Y_S = 50(6.95 - j12.41)\text{ mmhos}$$
$$= 0.34 - j0.62\text{ mho}$$

This normalized admittance is shown plotted in Fig. 6-9. Note that its corresponding impedance can be read directly from the chart as $Z_S = 0.69 + j1.2$ ohms. The input matching network must transform the 50-ohm source impedance to the impedance represented by this point. As was discussed in Chapter 4, there are numerous impedance-matching networks available to do the trick. The two-element L network was chosen here for simplicity and convenience.

$$\text{Arc AB} = \text{series C} = -j1.3\text{ ohms}$$
$$\text{Arc BC} = \text{shunt L} = -j1.1\text{ mhos}$$

The output circuit is designed and plotted in Fig. 6-10. Because the admittance values needed in the output network are so small, this chart had to be normalized to 200 ohms (5 mmhos). Thus, the normalized admittance plotted on the chart is:

$$Y_L = 200(0.347 - j1.84)\text{ mmhos}$$
$$= 0.069 - j0.368\text{ mho}$$

or,

$$Z_L = 0.495 + j2.62\text{ ohms}$$

The normalized 50-ohm load must be transformed to this impedance for maximum transfer of power. Again, the two-element L network was chosen to perform the match.

$$\text{Arc AB} = \text{series C} = -j1.9\text{ ohms}$$
$$\text{Arc BC} = \text{shunt L} = -j0.89\text{ mho}$$

The input and output matching networks are shown in Fig. 6-11. For clarity, the bias circuitry is not shown.

Actual component values are found using Equations 4-11 through 4-14. For the input network:

$$C_1 = \frac{1}{\omega XN}$$
$$= \frac{1}{2\pi(100 \times 10^6)(1.3)(50)}$$
$$= 24.5\text{ pF}$$

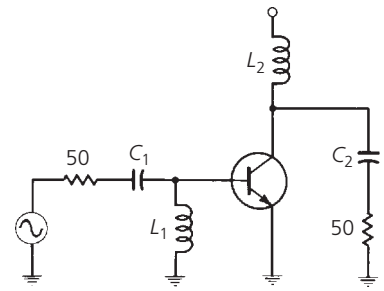


Fig. 6-11. Circuit topology for Example 6-1.

and

$$L_1 = \frac{N}{\omega B}$$
$$= \frac{50}{2\pi(100 \times 10^6)(1.1)}$$
$$= 72\text{ nH}$$

Similarly, for the output network:

$$C_2 = \frac{1}{2\pi(100 \times 10^6)(1.9)(200)}$$
$$= 4.18\text{ pF}$$

and

$$L_2 = \frac{200}{2\pi(100 \times 10^6)(0.89)}$$
$$= 358\text{ nH}$$

The final circuit, including the bias network, might appear as shown in Fig. 6-12. The 0.1-μF capacitors provide RF bypass at 100 MHz.

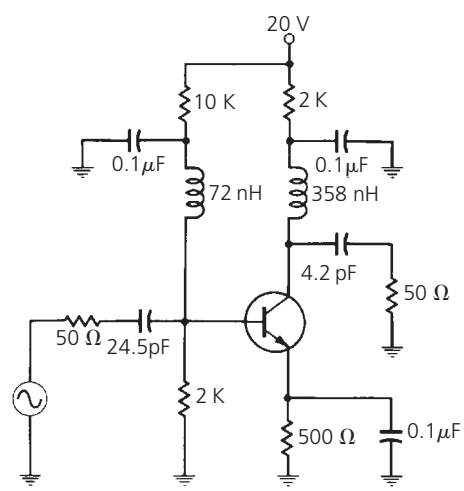


FIG. 6-12. Final circuit for Example 6-1.

Analyze the amplifier at the desired center frequency of 1.9 GHz and acquire the amplifier S-parameters, using the ANALYZE and CALCULATE methods of circuit object.

```
analyze(amp, 1.9e9);
data = calculate(amp, 'S11', 'S12',
'S21', 'S22', 'none');
[s11, s12, s21, s22] = deal(data{1},
data{2}, data{3}, data&#
123;4);
```

Step 2. Check for amplifier stability

Before proceeding with the design, check the stability of the amplifier. For unconditional stability, K must be greater than 1 and the absolute value of delta must be less than 1.

```
delta = s11*s22-s12*s21;
K = (1-abs(s11)^2-
abs(s22)^2+abs(delta)^2)/(2*abs(s12*s21))
abs_delta = abs(delta)
```

$$K = 1.0599$$

$$\text{abs\_delta} = 0.6776$$

Since both conditions are satisfied, the amplifier is unconditionally stable. Therefore, any passive source or load produces a stable condition.

Step 3. Define the simultaneous conjugate match value

To design the input and output matching networks, the demo calculates the required source and load reflection coefficients for a simultaneous conjugate match. It calculates the necessary load reflection coefficient for the design of the output matching network using the amplifier S-parameters.

```
B = 1+abs(s22)^2-abs(s11)^2-abs(delta)^2;
C = s22-delta*conj(s11);
gammaL = (B-sqrt(B^2-4*abs(C)^2))/2/C;
```

Step 4. Draw the SWR circle

Define the standing wave ratio (SWR) circle associated with the load reflection coefficient. The radius of this circle is given by the magnitude of the load reflection coefficient. The demo uses this radius (center is the origin) to calculate points on the circle.

```
theta = 0:pi/50:2*pi;
xin = abs(gammaL)*cos(theta);
yin = abs(gammaL)*sin(theta);
```

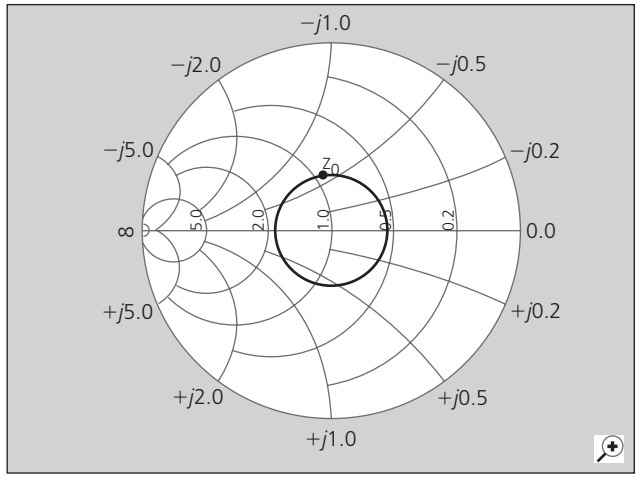


FIG. 6-13. YZ Smith Chart plot of desired input impedance at the output matching network.

Then it plots and labels the desired input impedance at the output matching network (based on the load reflection coefficient) along with the SWR circle on a YZ Smith Chart (Fig. 6-13).

```
hs = smithchart;
set(hs, 'Type', 'yz');
hold on
plot(xin, yin, '-', real(gammaL),
imag(gammaL), 'k.', ...
'LineWidth', 2, 'MarkerSize', 20);
text(-0.05, 0.35, 'z_{in}', 'FontSize',
12, 'FontUnits', 'normalized');
```

Step 5. Draw the constant conductance circle

To find the required susceptance to move the 50-ohm load admittance to the SWR circle, the demo defines the constant conductance circle. To do this, the demo calculates the normalized load impedance and the corresponding 50-ohm load admittance for the transmission lines.

```
zL = 50/50; % zL = 1
yL = 1/zL; % yL = 1
```

The demo calculates the diameter and center of the circle using the conductance value.

```
g = real(yL); % g = 1
d = -(g-1)/(g+1)+1; % d = 1
C = -1+d/2; % C = -1/2
```

Then it uses the radius and center of the constant conductance circle to calculate points on the circle.

```
xg = d/2*cos(theta)+C;
yg = d/2*sin(theta);
```

Next, the demo plots and labels the load impedance point (located at the center of the Smith Chart) along with the constant conductance circle associated with the load admittance on the Smith Chart (Fig. 6-14).

```
plot(xg, yg, 'r', 0, 0, 'k.',
'LineWidth', 2, 'MarkerSize', 20);
text(0.05, 0, 'z_L', 'FontSize', 12,
'FontUnits', 'normalized');
```

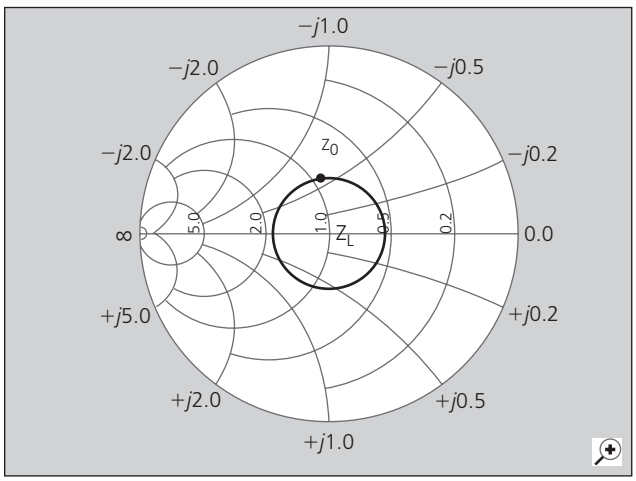


FIG. 6-14. Smith Chart with load impedance point and constant conductance circle.

Step 6. Find the intersection points

Now that the demo has drawn the SWR and constant conductance circles, you can find the points of intersection corresponding to the two possible solutions and the required susceptance values for the stub by visual inspection. Since only one solution is necessary, choose the lower-half intersection point, as shown in Fig. 6-15.

```
yA = 1+0.62j;
```

The demo plots and labels this intersection point on the Smith Chart using the reflection coefficient calculated from the admittance value.

```
gammaA = (1/yA-1)/(1/yA+1);
plot(real(gammaA), imag(gammaA), 'k.',
'MarkerSize', 20);
text(-0.09, -0.35, 'A', 'FontSize', 12,
'FontUnits', 'normalized');
hold off
```

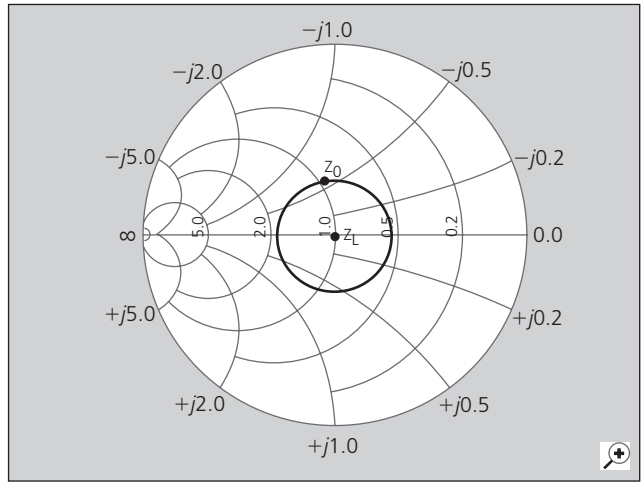


FIG. 6-15. Intersection point marked on Smith Chart.

Step 7. Calculate the required transmission line lengths

This demo finds the required lengths of the series transmission line and open-circuit stub (based on the intersection point) by calculating the required susceptance value for the stub and its corresponding reflection coefficient.

```
jbSA = yA-yL;
gammaSA = (1/jbSA-1)/(1/jbSA+1);
```

Then it finds the length of the stub by calculating the angle of rotation from the $y = 0$ (open-circuit) point to the calculated susceptance point.

```
stubAng = -angle(gammaSA)*180/pi;
stubLengthA = stubAng/360/2
```

stubLengthA = 0.0883

Finally, find the required length of the series transmission line based on the angle of rotation from point A to Zin.

```
seriesAng = 360-(angle(gammaL)-angle
(gammaA))*180/pi;
seriesLengthA = seriesAng/360/2
```

seriesLengthA = 0.2147

The required lengths (in terms of wavelength) for the transmission lines based on the solution from point A are given above. Following a similar approach, the line lengths for the input matching network are:

```
stubLengthin = 0.0763;
seriesLengthin = 0.2266;
```

Step 8. Verify the design

Build the circuit using microstrip transmission lines with a characteristic impedance of 50 ohms for the matching networks. To accomplish this, analyze a microstrip object, with default properties, at the design frequency of 1.9 GHz.

```
hstubOutput = rfckt.microstrip;
analyze(hstubOutput, 1.9e9);
Z0 = get(hstubOutput, 'Z0')
```

Z0 = 50.2561

This characteristic impedance is close to the desired impedance (50 ohms), and the demo can use it for this design. To appropriately set the required transmission line lengths in meters, the demo analyzes the microstrip to get a phase velocity value, which is necessary to calculate the wavelength.

```
phase_vel = get(hstubOutput, 'PV');
```

Set the appropriate transmission line lengths for the two series microstrip transmission lines necessary for the input and output matching networks.

```
hseriesOutput = rfckt.microstrip( ...
  'LineLength', phase_vel/1.9e9*series
  LengthA);
hseriesInput = rfckt.microstrip( ...
  'LineLength', phase_vel/1.9e9*series
  Lengthin);
```

Similarly, set the transmission line lengths and the stub mode for the two stubs necessary for the input and output matching networks.

```
set(hstubOutput, 'LineLength',
phase_vel/1.9e9*stubLengthA, ...
  'StubMode', 'shunt', 'Termination',
  'open');
hstubInput = rfckt.microstrip( ...
  'LineLength', phase_vel/1.9e9*stub
  Lengthin, ...
  'StubMode', 'shunt', 'Termination',
  'open');
```

Then cascade the circuit elements and analyze the amplifier with and without the matching networks over the frequency range of 1.5 to 2.3 GHz to visualize and compare the results.

```
matched_amp = rfckt.cascade('Ckts', ...
  {hstubInput, hseriesInput, amp,
hseriesOutput, hstubOutput});
analyze(matched_amp, 1.5e9:1e8:2.3e9);
analyze(amp, 1.5e9:1e8:2.3e9);
```

To verify the simultaneous conjugate match at the input and output of the amplifier, plot S11 and S22 parameters in dB for both circuits (see Figs. 6-16 and 6-17).

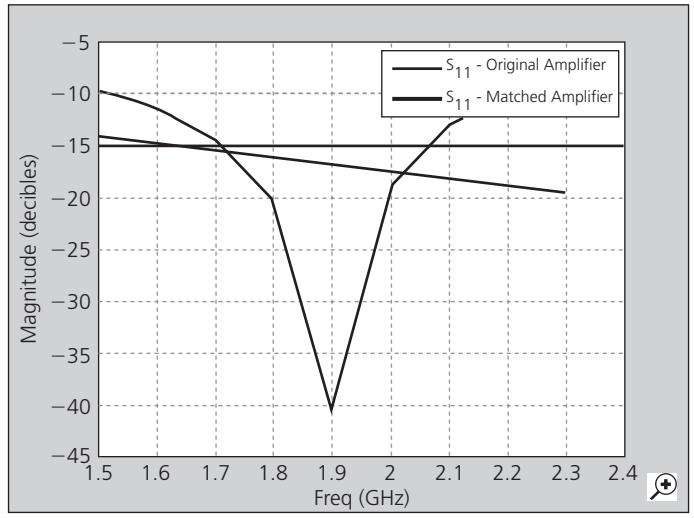


FIG. 6-16. Plot of S11 in dB.

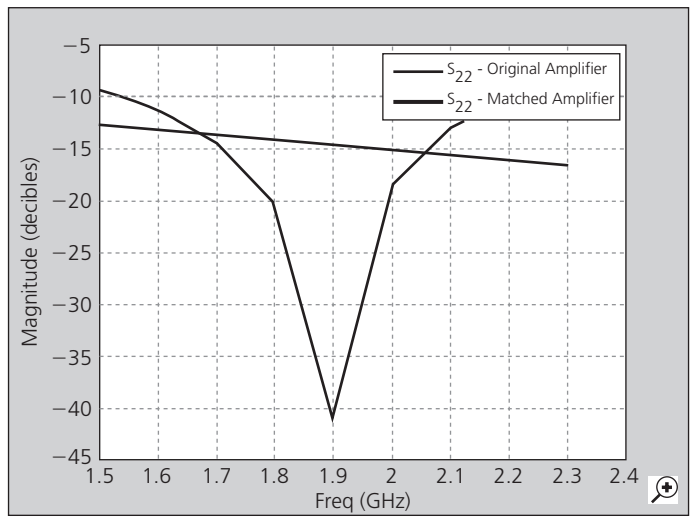


FIG. 6-17. Plot of S22 in dB.

```
hls = zeros(1,2);
hls(1) = plot(amp, 'S11', 'dB');
hold on;
hls(2) = plot(matched_amp, 'S11', 'dB');
set(hls(2), 'Color', [1 0 0]);
legend(hls,'S_{11} - Original Amplifier',
'S_{11} - Matched Ampli
fier');
hold off
```

```
hls(1) = plot(amp, 'S22', 'dB');
hold on;
hls(2) = plot(matched_amp, 'S22', 'dB');
set(hls(2), 'Color', [1 0 0]);
legend(hls,'S_{22} - Original Amplifier',
'S_{22} - Matched Ampli
fier');
hold off
```

Finally, plot S21 in dB for both circuits (Fig. 6-18).

```
hls(1) = plot(amp, 'S21', 'dB');
hold on;
hls(2) = plot(matched_amp, 'S21', 'dB');
set(hls(2), 'Color', [1 0 0]);
legend(hls,'S_{21} - Original Amplifier',
'S_{21} - Matched Ampli
fier');
hold off
```

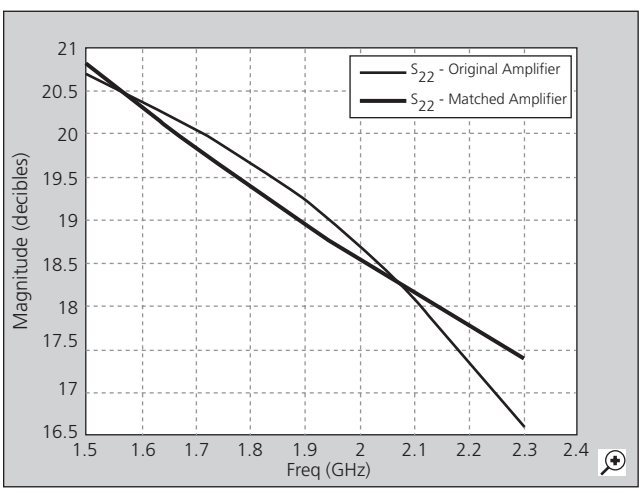


FIG. 6-18. Plot of S21 in dB.

You can compare the matched amplifier results with the expected transducer gain (in dB). From the S21 plot, you can see that the gain of the matched amplifier at 1.9 GHz is between 19 dB and 19.5 dB. The expected gain is:

```
Gt = 10*log10(abs(s21)/abs(s12)*
(K-sqrt(K^2-1)))
```

$$\text{Gt} = 19.2407$$

So, the matched amplifier's gain is very close to the expected transducer gain.

Transducer gain

Transducer gain is defined as the output power that is delivered to a load by a source, divided by the maximum power available from the source. This is the gain term most often referenced in RF amplifier design work. Transducer gain includes the effects of input and output impedance matching as well as the contribution that the transistor makes to the overall gain of the amplifier stage. Component resistive losses are neglected.

Given the source admittance (Y_S) and load admittance (Y_L) as seen by the transistor, the transducer gain is given by:

$$G_T = \frac{4G_S G_L |y_f|^2}{|(y_i + Y_S)(y_o + Y_L) - y_f y_r|^2} \qquad \text{(Eq. 6-12)}$$

EXAMPLE 6-2

Find the gain of the circuit that was designed in Example 6-1. Disregard any component losses.

Solution

The transducer gain for the amplifier is determined by substituting the values given in Example 6-1 into the Equation 6-12:

$$G_T = \frac{4(6.95)(0.347)|52 - j20|^2}{\left(\begin{array}{c}|(8 + j5.7 + 6.95 - j12.41)(0.4 + j1.5 + 0.347 \\ - j1.84) - (52 - j20)(0.01 - j0.1)|^2\end{array}\right)}$$

$$= \frac{29943}{|8.88 - j10.1 + 1.47 + j5.37|^2}$$

$$= 231.2$$

$$= 23.64\,\text{dB}$$

The transducer gain calculated in Example 6-2 is very close to the MAG that was calculated in Example 6-1. Therefore, in this case, the reverse-transfer admittance (y_r) of the transistor has very little effect on the overall gain of the stage. In many instances, however, y_r can take an appreciable toll on gain. For this reason, it is best to calculate G_T once the transistor's load and source admittances are determined. The calculation will provide you with a very good estimate of what the *actual* gain of the amplifier will be.

Designing with Potentially Unstable Transistors

If the Linvill stability factor (C) calculated with Equation 6-4 is greater than 1, the transistor you have chosen is *potentially*

unstable and may oscillate under certain conditions of source and load impedance. If this is the case, there are several options available that will enable you to use the transistor in a stable amplifier configuration:

1. Select a new bias point for the transistor.
2. Unilateralize or neutralize the transistor.
3. Selectively mismatch the input and output impedance of the transistor to reduce the gain of the stage.

The simplest solution to a stability problem is very often Option 1. This is especially true if C calculates to be very close to, but greater than, 1. Remember, any change in a transistor's operating point has a direct effect on its RF characteristics. Therefore, by simply changing the DC bias point, it is possible to change the Y parameters of the transistor and, hence, its stability. Of course, if this approach is taken, it is absolutely critical that the bias point be temperature-stable over the range of temperatures that the device must operate.

Since instability is generally caused by the feedback path, which consists of the reverse-transfer admittance (y_r) of the transistor, unilateralization or neutralization will often stabilize a design. *Unilateralization* consists of providing an external feedback path (Y_f) from the output to the input, such that $Y_f = -y_r$. Thus, Y_f cancels y_r leaving a composite reverse-transfer admittance (y_{rc}) equal to zero. With y_{rc} equal to zero, the device is unconditionally stable. This can be verified by substituting $y_{rc} = 0$ for y_r in Equation 6-4. The Linvill stability factor in this case becomes zero, thus indicating unconditional stability.

Often, when y_r is a complex admittance consisting of $g_r \pm jb_r$, it becomes very difficult to provide the correct *external* reverse admittance needed to totally eliminate the effect of y_r. In such cases, neutralization is often used. *Neutralization* is similar to unilateralization except that only the imaginary component of y_r is counteracted. An external feedback path is constructed from output to input such that $B_f = -b_r$. Thus, the composite reverse-transfer susceptance (b_{rc}) is equal to zero. Neutralization also tends to tame wild amplifiers because, in most transistors, g_r is negligible when compared to b_r. Thus, the elimination of b_r very nearly eliminates y_r. For this reason, neutralization is generally preferred over unilateralization. Two types of neutralizing circuits are shown in Fig. 6-19. In Fig. 6-19A, the series inductor and capacitor can be tuned to provide the correct amount of negative susceptance (inductance) necessary to cancel a positive reverse-transfer susceptance internal to the transistor. The circuit of Fig. 6-19B can be used to provide the correct amount of external positive susceptance necessary to cancel any $-jb$ that is internal to the transistor.

The addition of external components in order to neutralize an amplifier tends to increase the cost and complexity of the circuit.

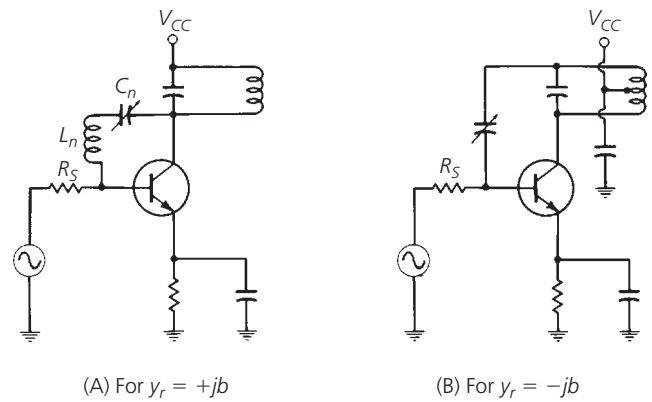


FIG. 6-19. Neutralization circuits.

Also, most neutralization circuits tend to neutralize the amplifier at the operating frequency only, and may cause problems (instability) at other frequencies. It is possible, however, to stabilize an amplifier without any form of external feedback. Another look at the *Stern stability factor* (K) in Equation 6-5 will reveal how.

If G_S and G_L are made sufficiently large enough to force K to be *greater than* 1, then the amplifier will remain stable for those terminations. This suggests *selectively mismatching* the transistor to achieve stability. Thus, the gain of the amplifier must be less than that which would be possible with a simultaneous conjugate match. The procedure for a design using unstable devices is as follows:

1. Choose G_S based on the optimum noise-figure information in the transistor's data sheet. Alternately, choose G_S based on some other criteria, such as convenience or input-network Q.
2. Select a value of K that will assure you of a stable amplifier ($K > 1$).
3. Substitute the above values for K and G_S into Equation 6-5 and solve for G_L.
4. Now that G_S and G_L are known, all that remains is to find B_S and B_L. Choose a value of B_L equal to the $-b_o$ of the transistor. The corresponding Y_L which results will then be *very close* to the true Y_L that is theoretically needed to complete the design.
5. Next, calculate the transistor's input admittance (Y_{in}) using the load chosen in Step 4 and the formula in Equation 6-13.

$$Y_{\text{in}} = y_i - \frac{y_r y_f}{y_o + Y_L} \qquad \text{(Eq. 6-13)}$$

where

$$Y_L = G_L \pm jB_L \text{ (found in Steps 3 and 4).}$$

6. Once Y_{in} is known, set B_S equal to the *negative* of the imaginary part of Y_{in}, or:

$$B_S = -B_{\text{in}}$$

7. Calculate the gain of the stage using Equation 6-12.

From this point forward, it is only necessary to provide input and output networks that will present the calculated Y_S and Y_L to the transistor. Example 6-3 illustrates the procedure.

DESIGN USING S PARAMETERS

As we discussed in Chapter 5, transistors can also be completely characterized by their scattering or S parameters. With these parameters, it is possible to calculate potential instabilities (tendency toward oscillation), maximum available gain, input and output impedances, and transducer gain. It is also possible to calculate optimum source and load impedances either for simultaneous conjugate matching or simply to help you choose specific source and load impedances for a specified transducer gain.

EXAMPLE 6-3

Consider a transistor with the following Y parameters at 200 MHz:

$$y_i = 2.25 + j7.2$$

$$y_o = 0.4 + j1.9$$

$$y_f = 40 - j20$$

$$y_r = 0.05 - j0.7$$

All of the above parameters are in mmhos. Find source and load admittances that will assure you of a stable design. Find the gain of the amplifier.

Solution

The *Linvill stability factor* (C) for the transistor is equal to 2.27 as calculated using Equation 6-4. Therefore, the device is potentially unstable and you must exercise extreme caution in choosing a source and load admittance for the transistor. Proceed as previously outlined in Steps 1 through 7.

The data sheet for the 2N5179 transistor states that the optimum source resistance for the best noise figure is 250 ohms. Thus, $G_S = 1/R_S = 4$ mmhos. Choose a Stern stability factor of $K = 3$ for an adequate safety margin.

Substitute G_S and K into Equation 6-5 and solve for G_L.

$$K = \frac{2(g_i + G_S)(g_o + G_L)}{|y_r y_f| + \text{Re}(y_r y_f)}$$

$$3 = \frac{2(2.25 + 4)(0.4 + G_L)}{31.35 + (-12)}$$

and

$$G_L = 4.24 \text{ mmhos}$$

Set B_L equal to $-b_o$ of the transistor,

$$B_L = -j1.9 \text{ mmhos}$$

The load admittance is now defined.

$$Y_L = 4.24 - j1.9 \text{ mmhos}$$

Calculate the input admittance of the transistor using Equation 6-13 and Y_L.

$$\begin{aligned} Y_{in} &= y_i - \frac{y_r y_f}{y_o + Y_L} \\ &= 2.25 + j7.2 - \frac{(0.701 \angle -85.9^\circ)(44.72 \angle -26.6^\circ)}{0.4 + j1.9 + 4.24 - j1.9} \\ &= 4.84 + j13.44 \text{ mmhos} \end{aligned}$$

Set B_S equal to the negative of the imaginary part of Y_{in}.

$$B_S = j13.44 \text{ mmhos}$$

The source admittance needed for the design is now defined as:

$$Y_S = 4.84 - j13.44 \text{ mmhos}$$

Now that Y_S and Y_L are known, you can calculate the expected gain of the amplifier using Equation 6-12.

$$\begin{aligned} G_T &= \frac{4(4.84)(4.24)|(44.72)|^2}{|(7.08 - j6.24)(4.64) - (-12 - j28.96)|^2} \\ &= \frac{135{,}671.7}{2011} \\ &= 67.61 \\ &= 18.3 \text{ dB} \end{aligned}$$

Therefore, even though the transistor is not conjugately matched, you can still realize a respectable amount of gain while maintaining a perfectly stable amplifier. Component values can be found by following the procedures outlined in Example 6-1.

Like Y parameters, S parameters vary with frequency and bias level. Therefore, you must first choose a transistor, select a stable operating point, and determine its S parameters at that operating point (either by measurement or from a data sheet) before following the procedures given in the following sections.

Stability

The tendency of a transistor toward oscillation can be gauged by its S-parameter data in much the same manner as was done in an earlier section with Y parameters. The calculation can be made even before an amplifier is built and, thus, it serves as a useful tool in finding a suitable transistor for your application.

To calculate the stability of a transistor with S parameters, you must first calculate the intermediate quantity D_S:

$$D_S = S_{11}S_{22} - S_{12}S_{21} \qquad \text{(Eq. 6-14)}$$

The Rollett Stability Factor (K) is then calculated as:

$$K = \frac{1 + |D_s|^2 - |S_{11}|^2 - |S_{22}|^2}{2 \cdot |S_{21}| \cdot |S_{12}|} \qquad \text{(Eq. 6-15)}$$

If K is *greater than* 1, then the device will be *unconditionally stable* for any combination of source and load impedance. If, on the other hand, K calculates to be *less than* 1, the device is *potentially unstable* and will most likely oscillate with certain combinations of source and load impedance. With K less than 1, you must be extremely careful in choosing source and load impedances for the transistor. It does not mean that the transistor cannot be used for your application; it merely indicates that the transistor will be *more difficult* to use.

If K calculates to be less than 1, there are several approaches that you can take to complete the design:

1. Select another bias point for the transistor.
2. Choose a different transistor.
3. Follow the procedures outlined later in this chapter.

Maximum Available Gain

The maximum gain you could ever hope to achieve from a transistor under conjugately matched conditions is called the Maximum Available Gain (MAG). To calculate MAG, first calculate the intermediate quantity B_1:

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |D_S|^2 \qquad \text{(Eq. 6-16)}$$

where D_S is the quantity calculated using Equation 6-14.

The MAG is then calculated:

$$\text{MAG} = 10 \log \frac{|S_{21}|}{|S_{12}|} + 10 \log |K \pm \sqrt{K^2 - 1}| \qquad \text{(Eq. 6-17)}$$

where

MAG is in dB,

K is the stability factor calculated using Equation 6-15.

The reason B_1 had to be calculated first is because its polarity determines which sign ($\pm$) to use before the radical in Equation 6-17. If B_1 is negative, use the plus sign. If B_1 is positive, use the minus sign.

Note that K must be greater than 1 (unconditionally stable) or Equation 6-17 will be undefined. That is, for a K less than 1, the radical in the equation will produce an *imaginary* number and the MAG calculation is no longer valid. Thus, MAG is undefined for unstable transistors.

Simultaneous Conjugate Match (Unconditionally Stable Transistors)

Once a suitable stable transistor has been found, and its gain capabilities have been found to match your requirements, you can proceed with the design.

The following design procedures will result in load and source reflection coefficients that will provide a conjugate match for the *actual* output and input impedances, respectively, of the transistor. Remember that the actual *output* impedance of a transistor is dependent upon the *source* impedance that the transistor "sees." Conversely, the actual *input* impedance of the transistor is dependent upon the *load* impedance that the transistor "sees." This dependency is, of course, caused by the reverse gain of the transistor (S_{12}). If S_{12} were equal to zero, then the load and source impedances would have no effect on the transistor's input and output impedances.

To find the desired *load reflection coefficient* for a conjugate match, perform the following calculations:

$$C_2 = S_{22} - (D_S S_{11}^*) \qquad \text{(Eq. 6-18)}$$

where the asterisk indicates the complex conjugate of S_{11} (same magnitude, but angle has the opposite sign). The quantity D_S is the intermediate quantity as calculated in Equation 6-14.

Next, calculate B_2.

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |D_S|^2 \qquad \text{(Eq. 6-19)}$$

The *magnitude* of the reflection coefficient is then found from the equation:

$$|\Gamma_L| = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2|C_2|} \qquad \text{(Eq. 6-20)}$$

The sign preceding the radical is the opposite of the sign of B_2 (which was previously calculated in Equation 6-19). The *angle* of the load-reflection coefficient is simply the negative of the angle of C_2 (found in Equation 6-18).

Once the desired load-reflection coefficient is found, it can be plotted on a Smith Chart, and the corresponding load impedance can be found directly. Or, if you prefer, you can substitute Γ_L into Equation 5-8, and solve for Z_L mathematically.

With the desired load-reflection coefficient specified, you can now calculate the source-reflection coefficient that is needed to properly terminate the transistor's input.

$$\Gamma_S = \left[S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-(\Gamma_L \cdot S_{22})}\right]^* \qquad \text{(Eq. 6-21)}$$

The asterisk again indicates that you should take the conjugate of the quantity in brackets (same magnitude, but opposite sign for the angle). In other words, once you complete the calculation (within the brackets) of Equation 6-21, the magnitude of the result will be correct, but the angle will have the wrong sign. Simply change the sign of the angle.

Once Γ_S is found, it can either be plotted on a Smith Chart or substituted into Equation 5-8 to find the corresponding source impedance. An example should help clarify matters (Example 6-4).

EXAMPLE 6-4

A transistor has the following S parameters at 200 MHz, with a $V_{CE} = 10$ V and an $I_C = 10$ mA:

$$S_{11} = 0.4\angle 162^\circ$$
$$S_{22} = 0.35\angle -39^\circ$$
$$S_{12} = 0.04\angle 60^\circ$$
$$S_{21} = 5.2\angle 63^\circ$$

The amplifier must operate between 50-ohm terminations. Design input and output matching networks to simultaneously conjugate match the transistor for maximum gain.

Solution

First use Equations 6-14 and 6-15 to see if the transistor is stable at the operating frequency and bias point:

$$\begin{aligned} D_s &= (0.4\angle 162^\circ)(0.35\angle -39^\circ) - (0.04\angle 60^\circ)(5.2\angle 63^\circ) \\ &= 0.14\angle 123^\circ - 0.208\angle 123^\circ \\ &= 0.068\angle -57^\circ \end{aligned}$$

Use the magnitude of D_S to calculate K.

$$\begin{aligned} K &= \frac{1 + (0.068)^2 - (0.4)^2 - (0.35)^2}{2(5.2)(0.04)} \\ &= 1.74 \end{aligned}$$

Since K is greater than 1, the transistor is unconditionally stable and we may proceed with the design. Next, calculate B_1 using Equation 6-16.

$$\begin{aligned} B_1 &= 1 + (0.4)^2 - (0.35)^2 - (0.068)^2 \\ &= 1.03 \end{aligned}$$

The Maximum Available Gain is then given by Equation 6-17:

$$\begin{aligned} \text{MAG} &= 10\log\frac{5.2}{0.04} + 10\log\left|1.74 - \sqrt{(1.74)^2 - 1}\right| \\ &= 21.14 + (-5) \\ &= 16.1\ \text{dB} \end{aligned}$$

The negative sign shown in front of the radical in the above equation results from B_1 being positive.

If the design specification had called out a minimum gain greater than 16.1 dB, a different transistor would be needed. We will consider 16.1 dB adequate for our purposes.

The next step is to find the load-reflection coefficient needed for a conjugate match. The two intermediate quantities (C_2 and B_2) must first be found. From Equation 6-18:

$$\begin{aligned} C_2 &= 0.35\angle -39^\circ) - [(0.068\angle -57^\circ)(0.4\angle -162^\circ)] \\ &= 0.272 - j0.22 - [-0.021 + j0.0017] \\ &= 0.377\angle -39^\circ \end{aligned}$$

and, from Equation 6-19:

$$\begin{aligned} B_2 &= 1 + (0.35)^2 - (0.4)^2 - (0.068)^2 \\ &= 0.958 \end{aligned}$$

Therefore, the magnitude of the load-reflection coefficient can now be found using Equation 6-20.

$$\begin{aligned} |\Gamma_L| &= \frac{0.958 - \sqrt{(0.958)^2 - 4(0.377)^2}}{2(0.377)} \\ &= 0.487 \end{aligned}$$

The angle of the load-reflection coefficient is simply equal to the negative of the angle of C_2, or $+39^\circ$. Thus,

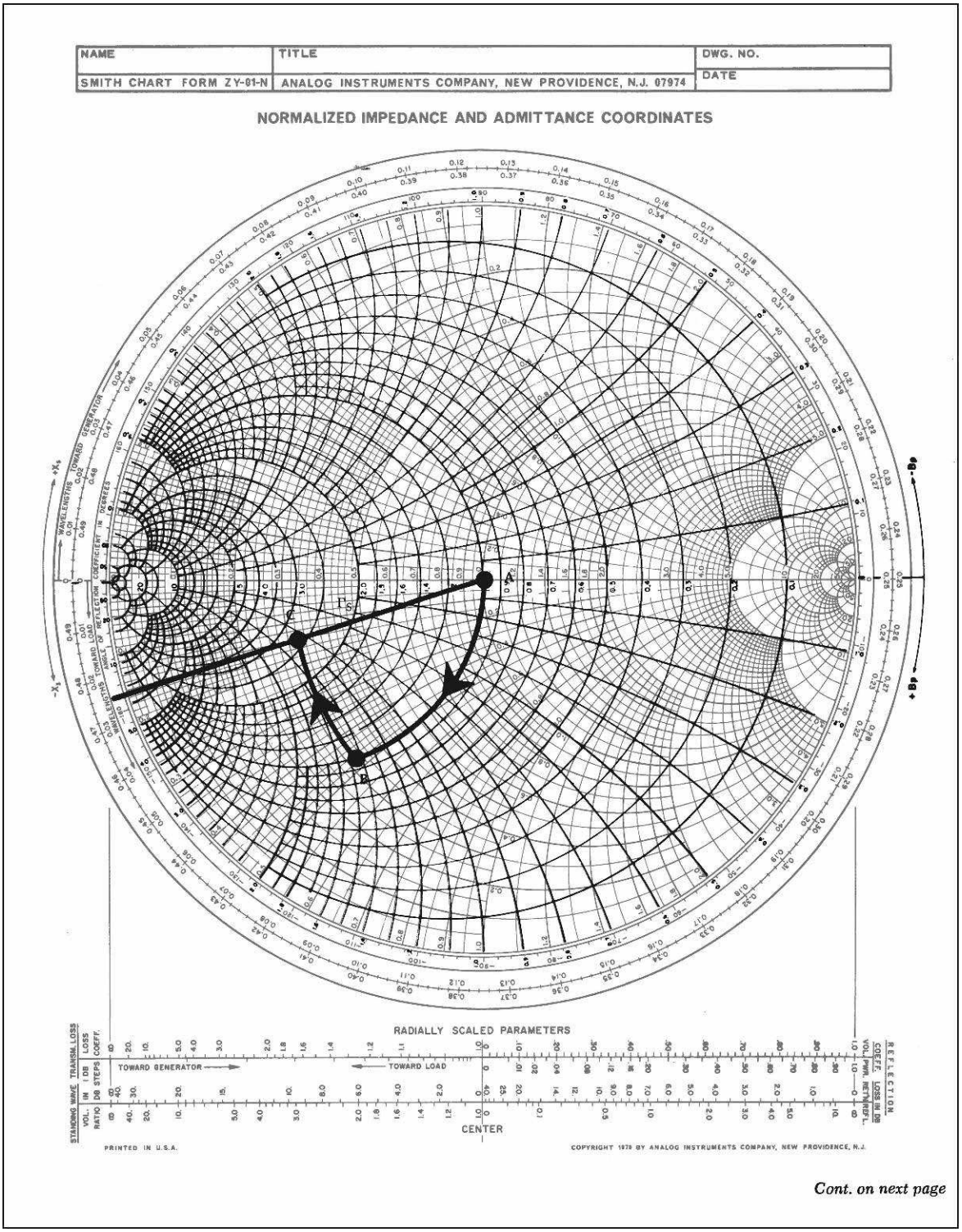
$$\Gamma_L = 0.487\angle 39^\circ$$

Using Γ_L, calculate Γ_S using Equation 6-21:

$$\begin{aligned} \Gamma_S &= \left[0.4\angle 162^\circ + \frac{(0.04\angle 60^\circ)(5.2\angle 63^\circ)(0.487\angle 39^\circ)}{1 - (0.487\angle 39^\circ)(0.35\angle -39^\circ)}\right]^* \\ &= [0.522\angle -162^\circ]^* \\ &= 0.522\angle 162^\circ \end{aligned}$$

Continued on next page

EXAMPLE 6-4—Cont



Cont. on next page

FIG. 6-20. Input network-design values for Example 6-4. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-4—Cont

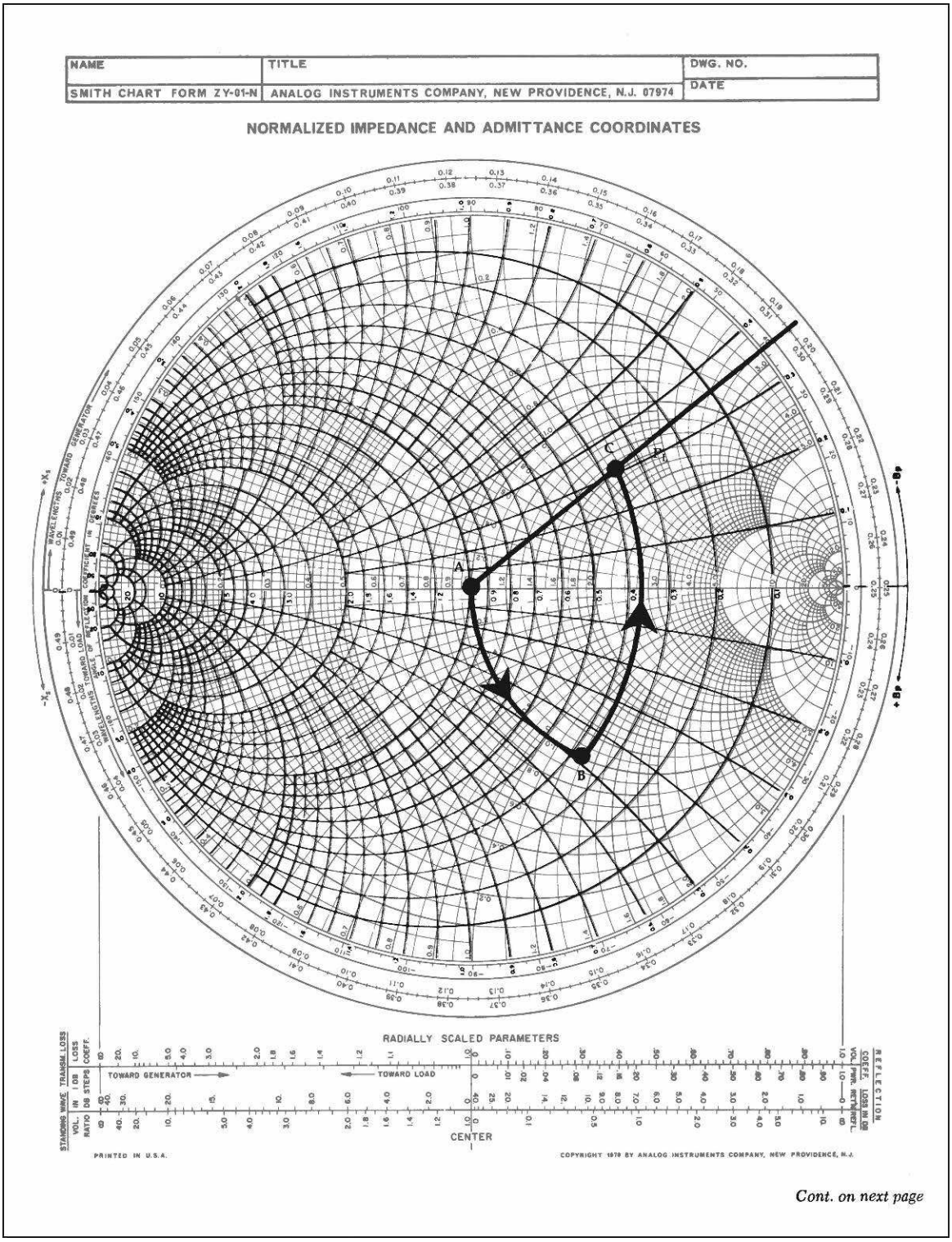


FIG. 6-21. Output network-design values for Example 6-4. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-4—Cont

Once the desired Γ_S and Γ_L are known, all that remains is to surround the transistor with components that provide it with source and load impedances which "look like" Γ_S and Γ_L.

The input matching-network design is shown on the Smith Chart of Fig. 6-20. The object of the design is to force the 50-ohm source to present a reflection coefficient of $0.522\angle{-162°}$. With Γ_S plotted as shown, the corresponding desired and normalized impedance is read directly from the chart as $Z_S = 0.32 - j0.14$ ohm. Remember, this is a *normalized* impedance because the chart has been normalized to 50 ohms. The actual impedance represented by Γ_S is equal to $50(0.32 - j0.14) = 16 - j7$ ohms. To force the 50-ohm source to actually appear as a $16 - j7$ ohm impedance to the transistor, we merely add a shunt and a series reactive component as shown on the chart of Fig. 6-20. Proceeding from the source, we have:

$$\text{Arc AB} = \text{Shunt C} = j1.45 \text{ mhos}$$

$$\text{Arc BC} = \text{Series L} = j0.33 \text{ ohm}$$

The actual component values are found using Equations 4-12 and 4-13.

$$C_1 = \frac{1.45}{2\pi(200 \times 10^6)50}$$

$$= 23 \text{ pF}$$

$$L_1 = \frac{(0.33)(50)}{2\pi(200 \times 10^6)}$$

$$= 13 \text{ nH}$$

This completes the input matching network.

The load-reflection coefficient is plotted in Fig. 6-21 and represents a desired load impedance (as read from the chart) of $Z_L - 50(1.6 + j1.28)$ ohms, or $80 + j64$ ohms. The matching network is designed as follows. Proceeding from the load:

$$\text{Arc AB} = \text{Series C} = -j1.3 \text{ ohms}$$

$$\text{Arc BC} = \text{Shunt L} = -j0.78 \text{ mho}$$

Component values are now found using Equations 4-11 and 4-14.

$$C_2 = \frac{1}{2\pi(200 \times 10^6)(1.3)(50)}$$

$$= 12 \text{ pF}$$

$$L_2 = \frac{50}{2\pi(200 \times 10^6)(0.78)}$$

$$= 51 \text{ nH}$$

The final design, excluding bias circuitry, is shown in Fig. 6-22.

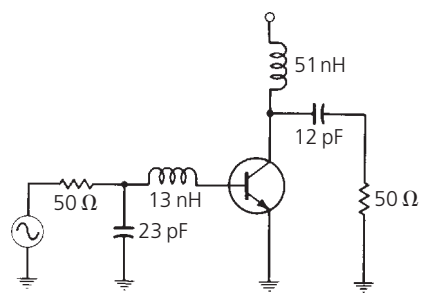


FIG. 6-22. Final circuit for Example 6-4.

Transducer Gain

The transducer gain, as defined earlier in this chapter, is the actual gain of an amplifier stage including the effects of input and output matching and device gain. It does not include losses attributed to power dissipation in imperfect components.

Transducer gain is found by

$$G_T = \frac{|S_{21}|^2(1 - |\Gamma_S|^2)(1 - |\Gamma_L|^2)}{|(1 - S_{11}\Gamma_S)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_L\Gamma_S|^2} \quad \text{(Eq. 6-22)}$$

where

Γ_S and Γ_L are the source- and load-reflection coefficients, respectively.

Calculation of G_T is a useful method of checking the power gain of an amplifier *before* it is built. This is shown by Example 6-5.

EXAMPLE 6-5

Calculate the transducer gain of the amplifier that was designed in Example 6-4.

Solution

Using Equation 6-22, we have:

$$G_T = \frac{(5.2)^2(1 - (0.522)^2)(1 - (0.487)^2)}{|(1 - 0.2088)(1 - 0.170) - (0.04\angle 60°)(5.2\angle 63°) \times (0.487\angle 39°)(0.522\angle{-162°})^2|}$$

$$= 41.15$$

$$= 16.1 \text{ dB}$$

Notice, again, that the transducer gain calculates to be very close to the MAG. If you carry the calculation out to several decimal places, you will find that G_T is still less than the MAG by a few hundredths of a dB. This is due to the fact that S_{12} is not equal to zero and is, therefore, providing a slight amount of negative feedback internal to the transistor.

Design for a Specified Gain

Often, when designing amplifiers, it is required that a single stage provide a certain amount of gain—no more and no less. In a situation such as this, a simultaneous conjugate match for the transistor would probably provide *too much* gain for the stage and would probably overdrive its load (or the succeeding stage). Obviously, if you so desired, you could search through mountains of manufacturer's literature hoping to find a transistor that, when conjugately matched, would provide exactly the amount of gain desired. This approach could take weeks or even months. Even if you did find a transistor with exactly the gain needed, you are now at the mercy of the manufacturer and are subject to any and all gain variations among transistors of the same type. There is a better way, however, and it alleviates the above problems very easily. It is called *selective mismatching*.

Selective mismatching is simply a controlled manageable way of decreasing gain by not matching the transistor to its load. This may sound like heresy to some, but it is a practical, logical, and well-accepted design procedure. There are still those who believe that at RF frequencies, a transistor *must* be matched to its source and load impedance. This is just not true. A transistor is simultaneously conjugate matched to its source and load only if maximum gain is desired, without regard for any other parameter, such as noise figure and bandwidth.

One of the easiest methods of selectively mismatching a transistor is through the use of a *constant-gain circle* as plotted on a Smith Chart. A constant-gain circle is simply a circle, the circumference of which represents a locus of points (load impedances) that will force the amplifier gain to a specified value. For instance, any of the infinite number of impedances located on the circumference of a 10-dB constant-gain circle would force the amplifier stage gain to 10 dB. Once the circle is drawn on a Smith Chart, either via manual techniques or the use of a computerized Smith Chart tool, you can see the load impedances that will provide a desired gain.

A constant-gain circle is plotted on a Smith Chart by performing a few calculations to determine:

1. Where the center of the circle is located.
2. The radius of the circle.

This information is calculated as follows:

1. Calculate D_S using Equation 6-14.
2. Calculate D_2.

$$D_2 = |S_{22}|^2 - |D_S|^2 \qquad \text{(Eq. 6-23)}$$

3. Calculate C_2.

$$C_2 = S_{22} - D_S S_{11}^* \qquad \text{(Eq. 6-24)}$$

4. Calculate G.

$$G = \frac{\text{Gain desired (absolute)}}{|S_{21}|^2} \qquad \text{(Eq. 6-25)}$$

   Note that the numerator in Equation 6-25 must be an absolute gain and not a gain in dB.

5. Calculate the location of the center of the circle.

$$r_o = \frac{GC_2^*}{1 + D_2 G} \qquad \text{(Eq. 6-26)}$$

6. Calculate the radius of the circle.

$$p_o = \frac{\sqrt{1 - 2K|S_{12}S_{21}|G + |S_{12}S_{21}|^2 G^2}}{1 + D_2 G} \qquad \text{(Eq. 6-27)}$$

Equation 6-26 produces a complex number in magnitude-angle format similar to a reflection coefficient. This number is plotted on the chart exactly as you would plot a value of reflection coefficient.

The radius of the circle that is calculated with Equation 6-27 is simply a fractional number between 0 and 1 which represents the size of that circle in relation to a Smith Chart. A circle with a radius of 1 has the same radius as a Smith Chart, a radius of 0.5 represents half the radius of a Smith Chart, and so on.

Once you choose the load-reflection coefficient and, hence, the load impedance that you will use, the next step is to determine the value of source-reflection coefficient that is needed to complete the design without producing any further decrease in gain. This value of source-reflection coefficient is the conjugate of the *actual input reflection coefficient* of the transistor with the specified load and is given by Equation 6-21. Example 6-6 outlines the procedure to follow.

Stability Circles

When the Rollett stability factor, as calculated with Equation 6-15, indicates a potential instability with the transistor, the chances are that with some combination of source and load impedance, the transistor will oscillate. Therefore, when K calculates to be less than 1, it is extremely important to choose source and load impedances very carefully. One of the best methods of determining those source and load impedances that will cause the transistor to go unstable is to plot *stability circles* on a Smith Chart. Again, this can be accomplished via manual techniques or through the use of computerized Smith Chart tools as discussed in Chapter 4.

A stability circle is simply a circle on a Smith Chart that represents the boundary between those values of source or load impedance that cause instability and those that do not. The perimeter of the circle thus represents the locus of points which forces $K = 1$. Either the inside *or* the outside of the circle may

EXAMPLE 6-6

A transistor has the following S parameters at 250 MHz, with a $V_{CE} = 5$ V and $I_C = 5$ mA.

$$S_{11} = 0.277\angle{-59^\circ}$$

$$S_{22} = 0.848\angle{-31^\circ}$$

$$S_{12} = 0.078\angle{93^\circ}$$

$$S_{21} = 1.92\angle{64^\circ}$$

Design an amplifier to provide 9 dB of gain at 250 MHz. The source impedance is $Z_S = 35 - j60$ ohms and the load impedance is $Z_L = 50 - j50$ ohms. The transistor is unconditionally stable with $K = 1.033$.

Solution

Using Equation 6-14 and Equations 6-23 through 6-27, and proceeding "by the numbers," we have:

$$\begin{aligned} D_S &= S_{11}S_{22} - S_{12}S_{21} \\ &= (0.277\angle{-59^\circ})(0.848\angle{-31^\circ}) \\ &\quad - (0.078\angle{93^\circ})(1.92\angle{64^\circ}) \\ &= 0.324\angle{-64.8^\circ} \end{aligned}$$

$$\begin{aligned} D_2 &= (0.848)^2 - (0.324)^2 \\ &= 0.614 \end{aligned}$$

$$\begin{aligned} C_2 &= 0.848\angle{-31^\circ} - (0.324\angle{-64.8^\circ})(0.277\angle{59^\circ}) \\ &= 0.768\angle{-33.9^\circ} \end{aligned}$$

$$\begin{aligned} G &= \frac{7.94}{(1.92)^2} \\ &= 2.15 \end{aligned}$$

The center of the circle is then located at the point:

$$\begin{aligned} r_o &= \frac{2.15(0.768\angle{33.9^\circ})}{1 + (0.614)(2.15)} \\ &= 0.712\angle{33.9^\circ} \end{aligned}$$

This point can now be plotted on the Smith Chart.

The radius of the 9-dB gain circle is calculated as:

$$\begin{aligned} p_o &= \frac{\sqrt{1 - 2(1.033)(0.078)(1.92)(2.15) + (0.150)^2(2.15)^2}}{1 + (0.614)(2.15)} \\ &= 0.285 \end{aligned}$$

The Smith Chart construction is shown in Fig. 6-23. Note that any load impedance located along the circumference of the circle will produce an amplifier gain of 9 dB *if the input impedance of the transistor is conjugate matched*.

The actual load impedance we have to work with is $50 - j50$ ohms, as given in the problem statement. Its normalized value ($1 - j1$) is shown in Fig. 6-23 (point A). The transistor's output network must transform the actual load impedance into a value that falls on the constant-gain circle. Obviously, there are numerous circuit configurations that will do the trick. The configuration shown was chosen for convenience. Proceeding *from* the load:

$$\text{Arc AB} = \text{Series C} = -j2 \text{ ohms}$$

$$\text{Arc BC} = \text{Shunt L} = -j0.425 \text{ mho}$$

Again, using Equations 4-11 through 4-14, the actual component values are:

$$\begin{aligned} C_1 &= \frac{1}{2\pi(250 \times 10^6)(2)(50)} \\ &= 6.4 \text{ pF} \end{aligned}$$

and

$$\begin{aligned} L_1 &= \frac{(50)}{2\pi(250 \times 10^6)(0.425)} \\ &= 75 \text{ nH} \end{aligned}$$

For a conjugate match at the input to the transistor with $\Gamma_L = 0.82\angle{14.2^\circ}$ (point C), the desired source-reflection coefficient must be (using Equation 6-21):

$$\begin{aligned} \Gamma_S &= \left[\frac{0.277\angle{-59^\circ} + (0.078\angle{93^\circ})(1.92\angle{64^\circ})(0.82\angle{14.2^\circ})}{1 - (0.82\angle{14.2^\circ})(0.848\angle{31^\circ})}\right]^* \\ &= 0.105\angle{160^\circ} \end{aligned}$$

This point is plotted as point D in Fig. 6-24. The actual normalized source impedance is plotted at point A ($0.7 - j1.2$ ohms). Thus, the input network must transform the actual impedance at point A to the desired impedance at point D. For practice, this was done with a three-element design as shown.

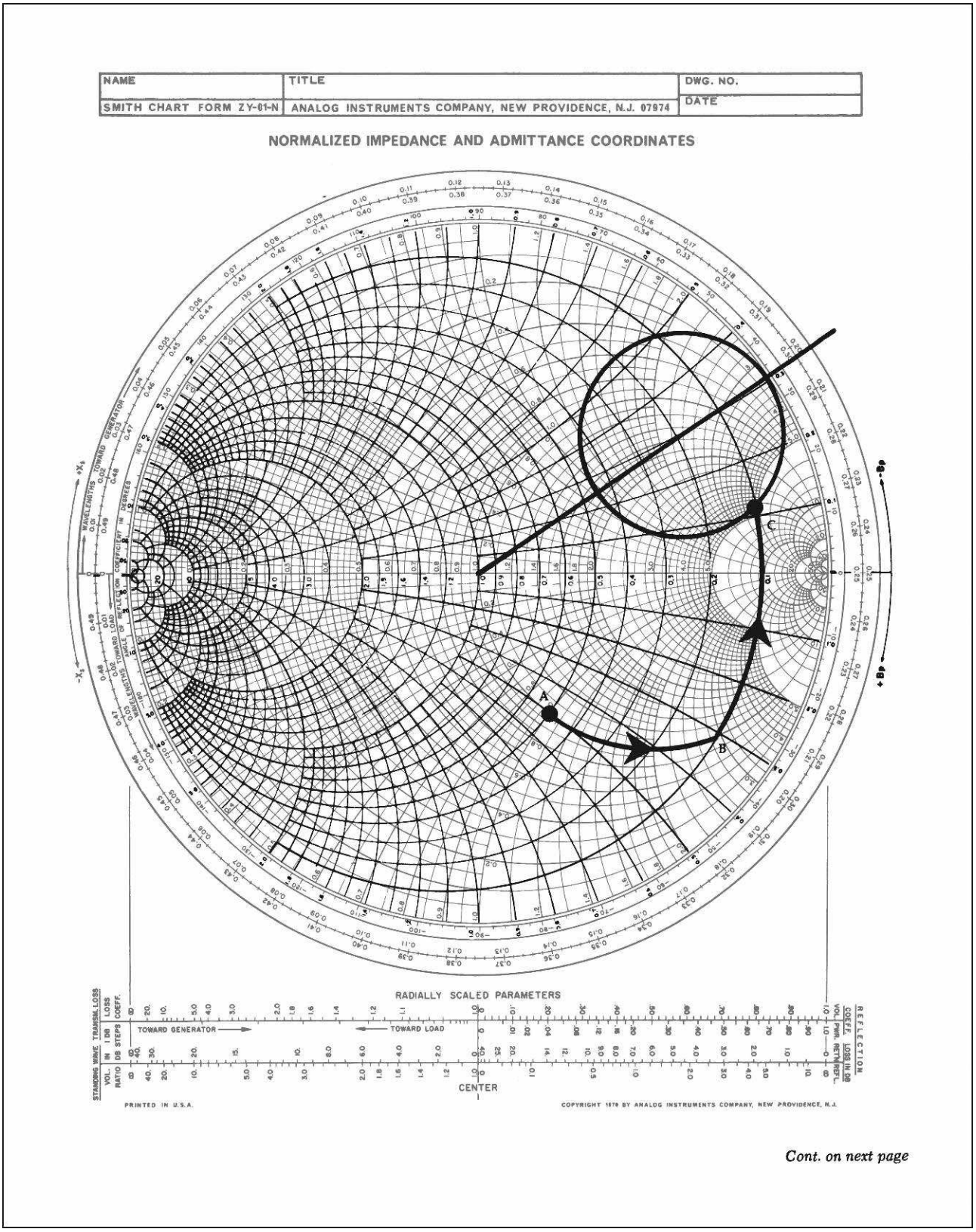
$$\text{Arc AB} = \text{Shunt C}_2 = j0.62 \text{ mho}$$

$$\text{Arc BC} = \text{Series L}_2 = j1.09 \text{ ohms}$$

$$\text{Arc CD} = \text{Shunt C}_3 = j2.1 \text{ mhos}$$

Continued on next page

EXAMPLE 6-6—Cont



Cont. on next page

FIG. 6-23. Output network-design values for Example 6-6. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-6—Cont

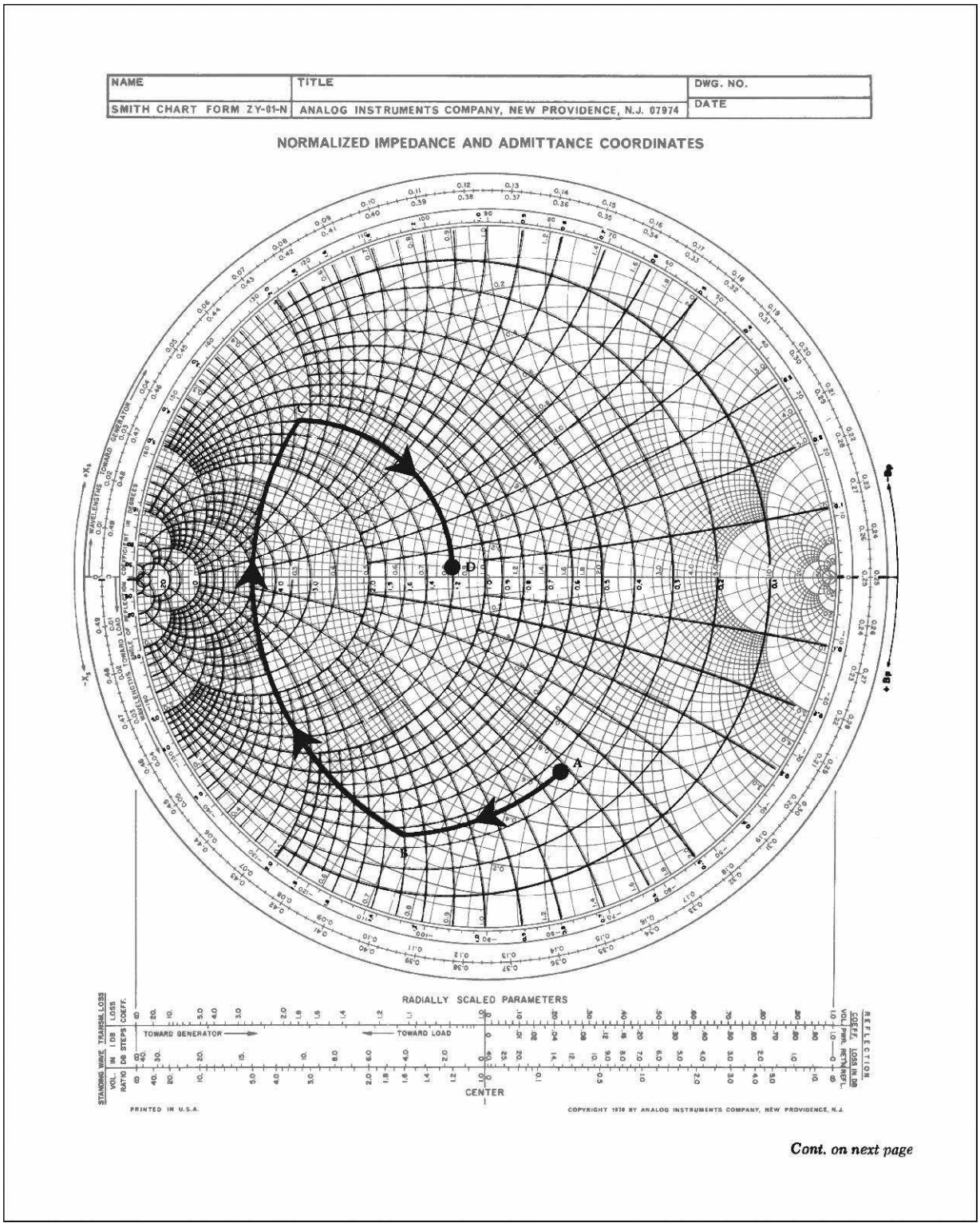


FIG. 6-24. Input network-design values for Example 6-6. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-6—Cont

From Equations 4-11 through 4-14:

$$C_2 = \frac{(0.62)}{2\pi(250 \times 10^6)(50)}$$
$$= 7.9\,\text{pF}$$
$$C_3 = \frac{2.1}{2\pi(250 \times 10^6)50}$$
$$= 27\,\text{pF}$$
$$L_2 = \frac{(1.09)(50)}{2\pi(250 \times 10^6)}$$
$$= 34.7\,\text{nH}$$

The completed design, excluding the bias network, is shown in Fig. 6-25.

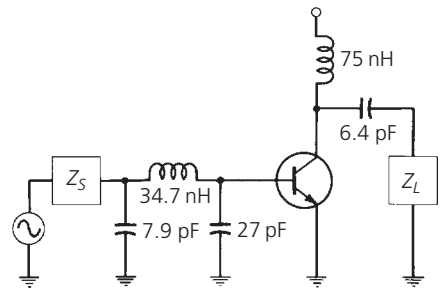


FIG. 6-25. Final circuit for Example 6-6.

represent the unstable region and that determination must be made after the circles are drawn.

The locations and radii of the input and output stability circles are found as follows:

1. Calculate D_S using Equation 6-14.
2. Calculate C_1.

$$C_1 = S_{11} - D_S S_{22}^* \qquad \text{(Eq. 6-28)}$$

3. Calculate C_2 using Equation 6-18.
4. Calculate the *center location* of the *input* stability circle.

$$r_{s1} = \frac{C_1^*}{|S_{11}|^2 - |D_S|^2} \qquad \text{(Eq. 6-29)}$$

5. Calculate the *radius* of the *input* stability circle.

$$p_{s1} = \left| \frac{S_{12}S_{21}}{|S_{11}|^2 - |D_S|^2} \right| \qquad \text{(Eq. 6-30)}$$

6. Calculate the *center location* of the *output* stability circle.

$$r_{s2} = \frac{C_2^*}{|S_{22}|^2 - |D_S|^2} \qquad \text{(Eq. 6-31)}$$

7. Calculate the *radius* of the *output* stability circle.

$$p_{s2} = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |D_S|^2} \right| \qquad \text{(Eq. 6-32)}$$

Once the calculations are made, the stability circles can be plotted directly on the Smith Chart. Note, however, that if you try to plot stability circles on the Smith Chart for an unconditionally stable transistor, you may never find them. This is because for an unconditionally stable amplifier the entire chart represents a stable operating region, as shown in Fig. 6-26.

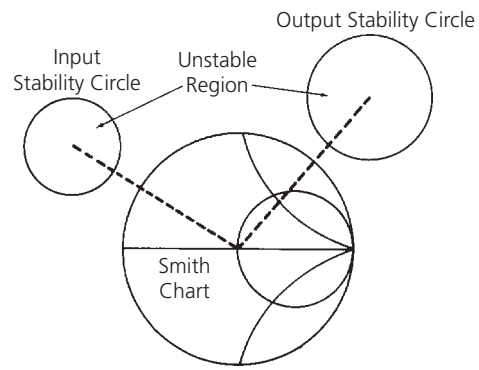


FIG. 6-26. Typical stability circles for an unconditionally stable amplifier.

For a *potentially unstable* transistor, the stability circles might resemble those shown in Fig. 6-27. Often, only a portion of the stability circle intersects the chart as shown.

After the stability circles are plotted on the chart, the next step is to determine which side of the circle (inside or outside) represents the *stable* region. This is very easily done if S_{11} and S_{22} for the transistor are less than 1. Since the S parameters were measured with a 50-ohm source and load, and since the transistor remained stable under these conditions (S_{11} or S_{22} would be greater than 1 for an unstable transistor), *then the center of the normalized Smith Chart must be part of the stable region* as described by the stability circles. Therefore, in this case, if one of the circles surrounds the center of the chart, the inside of that circle must represent the region of *stable* impedances for that port. If, on the other hand, the circle does not surround the center of the chart, then the entire area outside of that circle must represent the *stable* operating region for that port.

It is very rare that you will find a transistor that is unstable with a 50-ohm source and load and, if you do, it would probably be wise

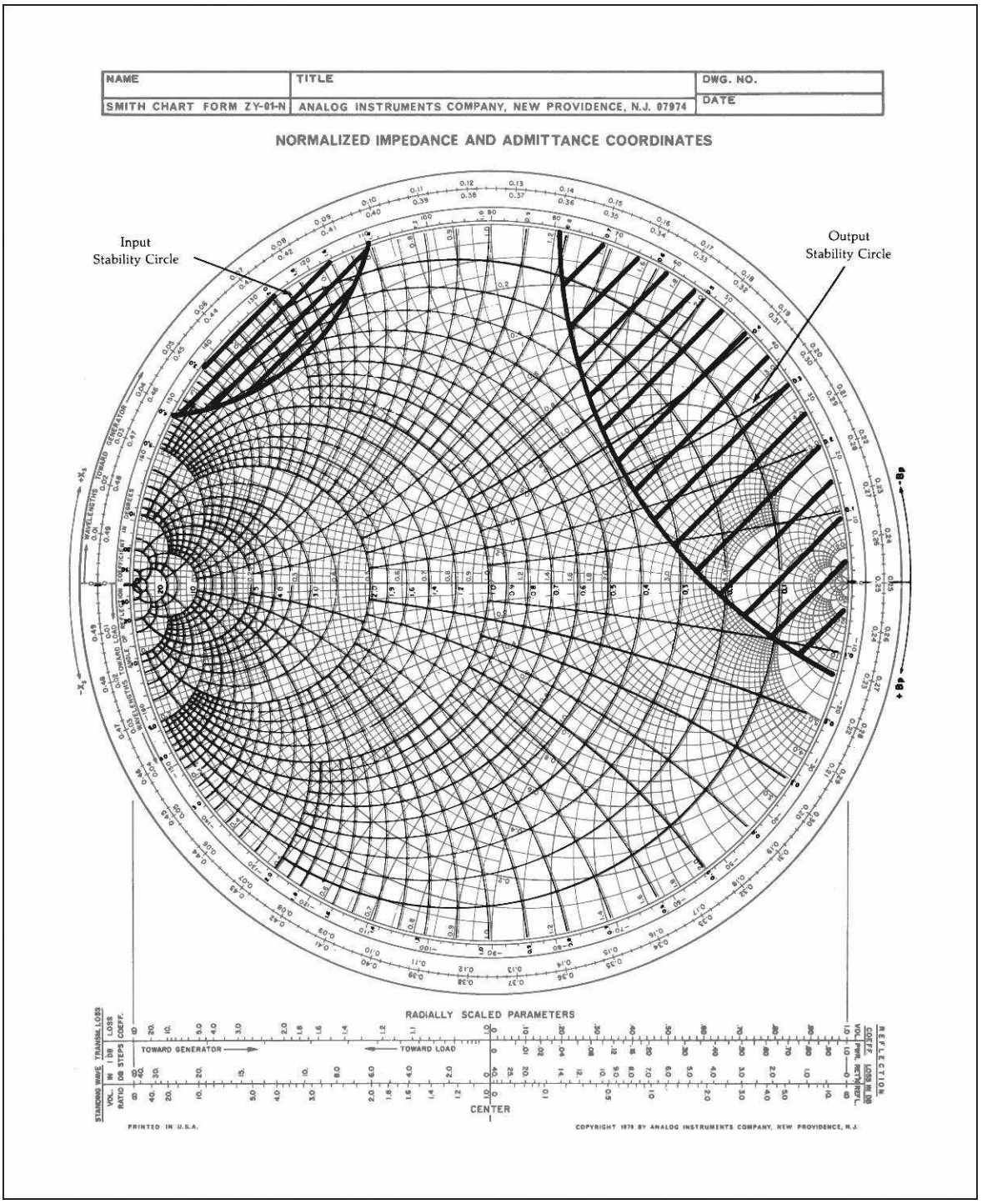


FIG. 6-27. Typical stability circles for a potentially unstable transistor. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

to try another device. Therefore, the procedure outlined above should be considered to be the most direct method of locating the stable operating regions on a Smith Chart. Example 6-7 diagrams the procedure.

Design for Optimum Noise Figure

The noise figure of any two-port network gives a measure of the amount of noise that is added to a signal that is transmitted through the network. For any practical circuit, the signal-to-noise

ratio at its output will be worse (smaller) than that at its input. In most circuit-design applications, however, it is possible to minimize the noise contribution of each two-port network through a judicious choice of operating point and source resistance.

In Chapter 5, it was briefly mentioned that for each transistor, indeed for each two-port network, there exists an optimum source resistance necessary to establish a minimum noise figure (see also Appendix B). Many manufacturers specify an optimum source resistance on the data sheet, such as in the case of the 2N5179 transistor presented in Chapter 5. Others will specify an optimum source-reflection coefficient. Such is the case for the Microwave Associates' MA-42120-Series transistor data sheet that is shown in Fig. 6-29. Note the Smith Chart, on page 3 of the data sheet, labeled "Typical Optimum Noise Source Impedance vs. Collector Current." Obviously, as shown on the chart, if you were planning to use the transistor at some frequency other than 60 MHz or 450 MHz, you would be out of luck as far as optimum noise-figure design is concerned. Typically, most data sheets are incomplete like this. There is just not enough space in a typical data book to provide the user with all of the information that he needs in order to design amplifiers at every possible frequency and bias point. The data sheet is meant only as a starting point in any design. Chances are you will end up making many of your own measurements on a device before it becomes a part of the design.

EXAMPLE 6-7

The S parameters for a 2N5179 transistor at 200 MHz, with a $V_{CE} = 6$ V and an $I_C = 5$ mA, are (see the data sheet in Chapter 5):

$$S_{11} = 0.4\angle 280^\circ$$
$$S_{22} = 0.78\angle 345^\circ$$
$$S_{12} = 0.048\angle 65^\circ$$
$$S_{21} = 5.4\angle 103^\circ$$

Choose a stable load- and source-reflection coefficient that will provide a power gain of 12 dB at 200 MHz.

Solution

A calculation of Rollett's stability factor (K) for the transistor indicates a potential instability with $K = 0.802$. Therefore, you must exercise extreme caution in choosing source and load impedances for the device or it may oscillate. To find the stable operating regions on the Smith Chart, plot the input and output stability circles. Proceeding with Step 1, above, we have:

$$D_S = (0.4\angle 280^\circ)(0.75\angle 345^\circ) - (0.048\angle 65^\circ)(5.4\angle 103^\circ)$$
$$= 0.429\angle -58.18^\circ$$
$$C_1 = 0.4\angle 280^\circ - (0.429\angle -58.2^\circ)(0.78\angle -345^\circ)$$
$$= 0.241\angle -136.6^\circ$$
$$C_2 = 0.78\angle 345^\circ - (0.429\angle -58.2^\circ)(0.4\angle -280^\circ)$$
$$= 0.65\angle -24^\circ$$

Then, the center of the input stability circle is located at the point:

$$r_{s1} = \frac{0.241\angle 136.6^\circ}{(0.4)^2 - (0.429)^2}$$
$$= 10\angle 136.6^\circ$$

The radius of the circle is calculated as:

$$p_{s1} = \left|\frac{(0.048\angle 65^\circ)(5.4\angle 103^\circ)}{(0.4)^2 - (0.429)^2}\right|$$
$$= 10.78$$

Similarly, for the output stability circle:

$$r_{s2} = \frac{0.65\angle 24^\circ}{(0.78)^2 - (0.429)^2}$$
$$= 1.53\angle 24^\circ$$
$$p_{s2} = \left|\frac{(0.048\angle 65^\circ)(5.4\angle 103^\circ)}{(0.78)^2 - (0.429)^2}\right|$$
$$= 0.610$$

These circles are shown in Fig. 6-28. Note that the input stability circle is actually drawn as a straight line because the radius of the circle is so large. Since S_{11} and S_{22} are both less than 1, we can deduce that the *inside* of the *input* stability circle represents the region of *stable* source impedances while the *outside* of the *output* stability circle represents the region of *stable* load impedances for the device.

The 12-dB gain circle is also shown plotted in Fig. 6-28. It is found using Equation 6-14 and Equations 6-23 through 6-27. Note that D_S and C_2 have already been calculated. The center location of the circle is found to be:

$$r_o = 0.287\angle 24^\circ$$

with a radius of:

$$p_o = 0.724$$

The only load impedances that we may not select for the transistor are located inside of the input stability circle. Any other load impedance located on the 12-dB gain circle will

Continued on next page

EXAMPLE 6-7—Cont

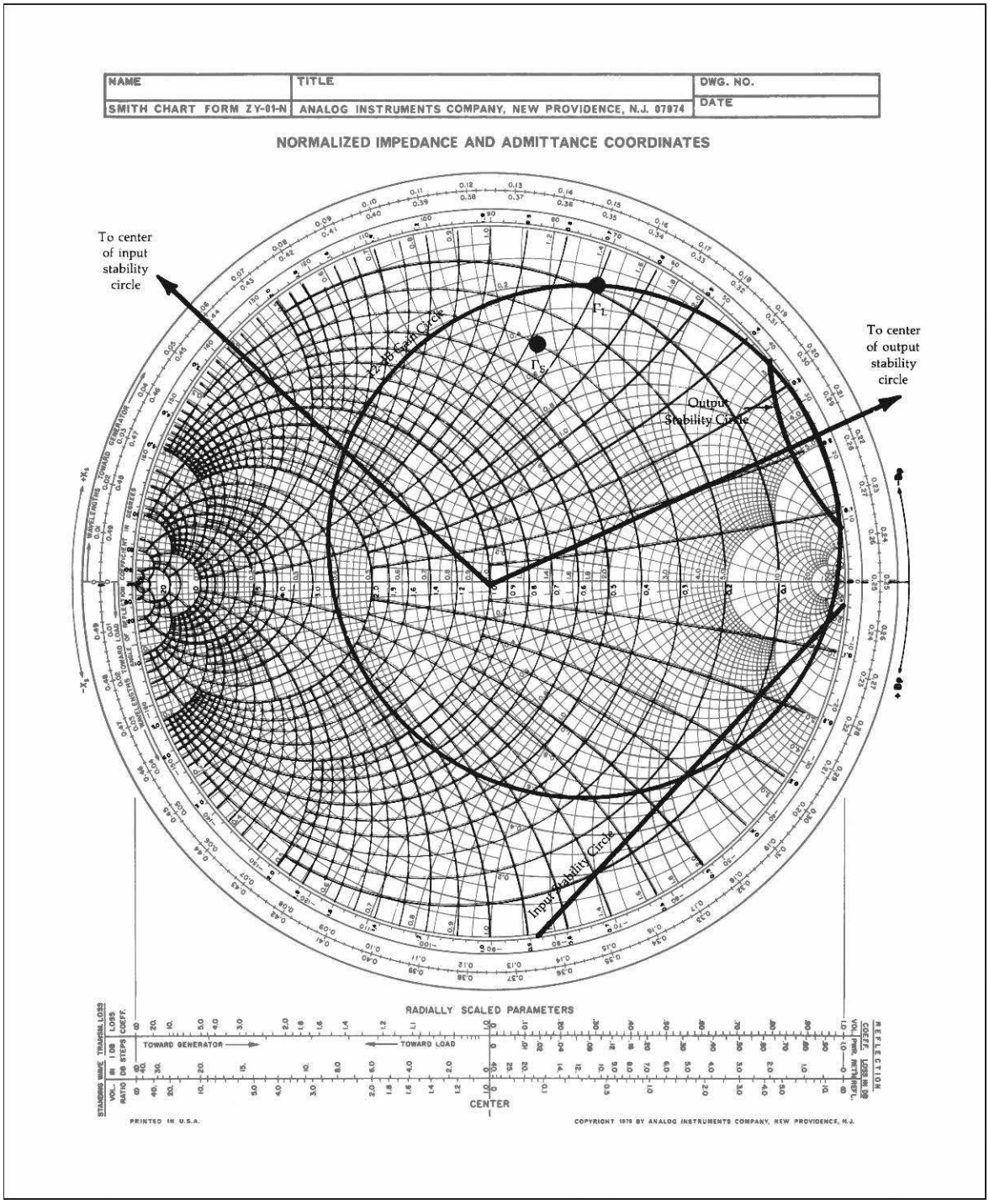


FIG. 6-28. Stability and gain circles for the transistor in Example 6-7. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-7—Cont

MA-42120 SERIES

specification sheet

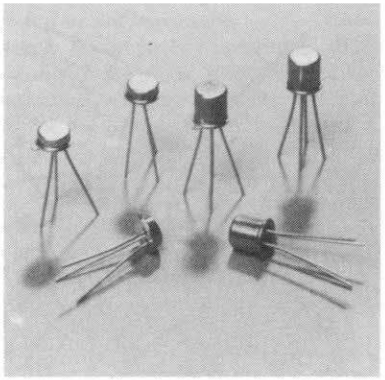
npn silicon planar transistors

FEATURES

- HIGH GAIN BANDWIDTH PRODUCT (1.5 GHz)
- HIGH f_{max} (4.2 GHz @ I_C = 20 mA)
- GOLD METALLIZATION
- HIGH RELIABILITY
- DIRECT INTERCHANGABILITY WITH THE FMT 1060 SERIES OF FAIRCHILD TRANSISTORS

DESCRIPTION

This series of NPN Epitaxial Silicon Planar Transistors is designed for VHF/UHF service. The performance of this series is comparable to the Fairchild FMT-1060 series. The high gain bandwidth products make the MA-42122 and MA-42123 useful to 1.0 GHz while the MA-42120 and MA-42121 have a maximum frequency of oscillation of 4.0 GHz. Two packages are offered, the TO-46 (ODS-508), for low power oscillator applications and the TO-72 (ODS-509) for small signal UHF amplifiers.



APPLICATIONS

VHF, UHF Low Level Oscillators
IF and RF Amplifiers

MA-42120 SERIES R.F. SPECIFICATIONS

MODEL NO.	MA-42120	MA-42121	MA-42122	MA-42123
CASE STYLE	508	508	509	509
Test Frequency (MHz)	450	450	450	450
Max Noise Fig. @ I_C (dB)	– –	– –	3.5	3.0
Gu (max) Typ. (dB)	13	13	14	14
I_C (mA)	– –	– –	1.5	1.5
1 dB Compression Point (dBm)	– –	– –	–12	–12
Fairchild Equivalent	FMT 1060	FMT 1060A	FMT 1061	FMT 1061A

MA-42120 SERIES HIGH FREQUENCY SPECIFICATIONS (25°C Ambient Temperature Unless Otherwise Noted)

Symbol	Characteristics	Type	Min.	Typ.	Max.	Units	Test Conditions
f_t	Gain Bandwidth Product	MA42121	1.3	1.5		GHz	V_{CE} = 10V, I_C = 20 mA
		MA42120	1.0	1.3		GHz	f = 500 MHz
		MA42123	1.3	1.5		GHz	
		MA42122	1.0	1.3		GHz	
$G_{A\,max}$	Maximum Available Gain	MA42121		12.8		dB	V_{CE} = 10V, I_C = 20 mA
		MA42123		13.8			f = 1.0 GHz
NF	Noise Figure	MA42123		2.3	3.0	dB	V_{CE} = 10V, I_C = 1.5 mA
		MA42122		2.7	3.5	dB	f = 450 MHz, R = 50 ohms
G_{pe}	Neutralized Power Gain	MA42123		17.0		dB	V_{CE} = 10V, I_E = 1.5 mA f = 450 MHz, R = 50 ohms
f_{max}	Maximum Frequency of Qscillation[1]	MA42121		4.2		GHz	V_{CE} = 10V, I_C = 20 mA
		MA42120		3.8		GHz	V_{CE} = 10V, I_C = 20 mA

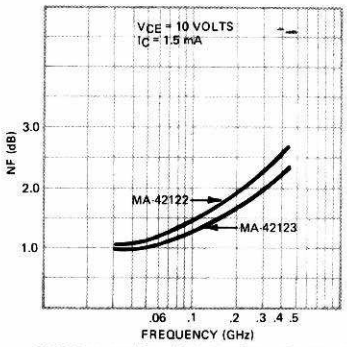
NOTE:
1. Calculated from S-Parameters, f_{max} is the frequency at which the extrapolated GA_{max} is 0 dB.

FIG. 6-29. Data sheet for Microwave Associates' MA-42120 series of transistors. (*Courtesy Microwave Associates*).

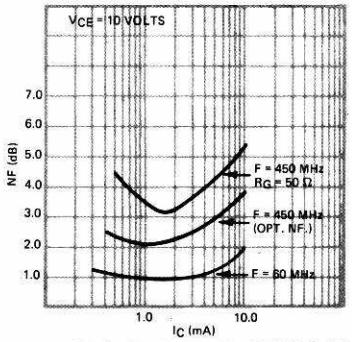
EXAMPLE 6-7—Cont

MA-42120 SERIES

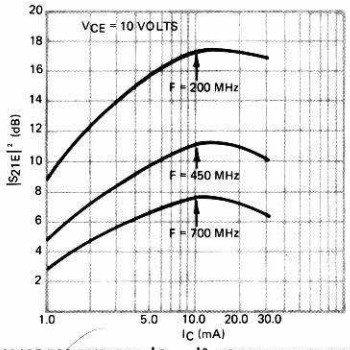
specification sheet



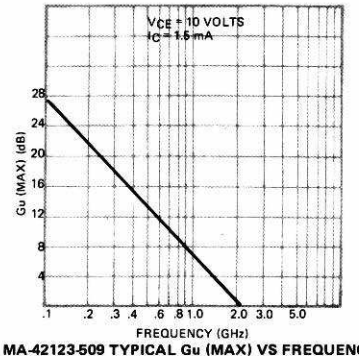
TYPICAL NOISE FIGURE VS FREQUENCY



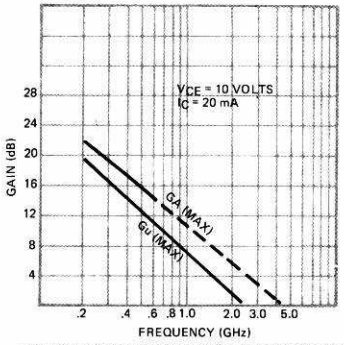
MA-42123-509 TYPICAL OPTIMUM N.F. VS COLLECTOR CURRENT



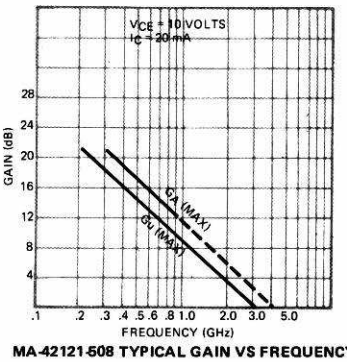
MA-42123-509 TYPICAL $|S_{21E}|^2$ VS COLLECTOR CURRENT



MA-42123-509 TYPICAL Gu (MAX) VS FREQUENCY



MA-42120-508 TYPICAL GAIN VS FREQUENCY



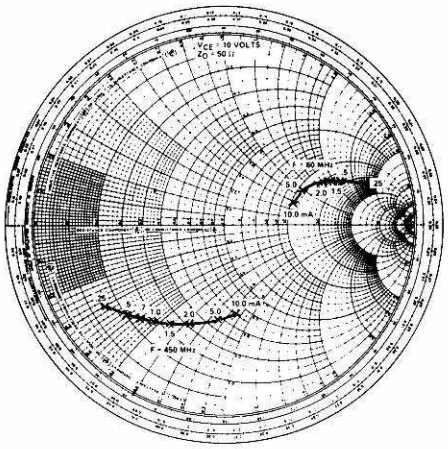
MA-42121-508 TYPICAL GAIN VS FREQUENCY

FIG. 6-29. (Continued).

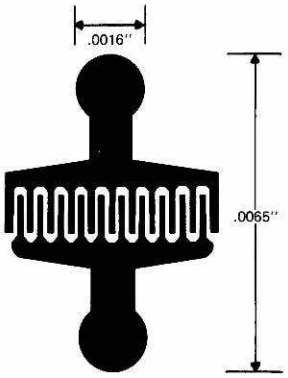
EXAMPLE 6-7—Cont

MA-42120 SERIES

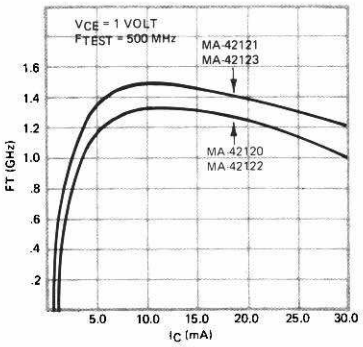
specification sheet



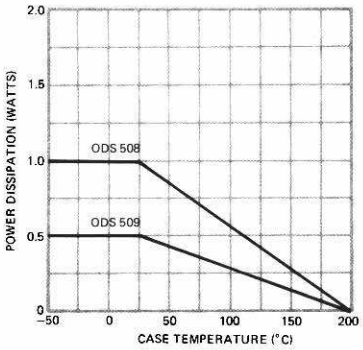
TYPICAL OPTIMUM NOISE SOURCE IMPEDANCE VS COLLECTOR CURRENT



GEOMETRY 70



MA-42120 SERIES TYPICAL F_T VS COLLECTOR CURRENT



MA-42120 SERIES POWER DISSIPATION VS. CASE TEMPERATURE

FIG. 6-29. (Continued).

EXAMPLE 6-7—Cont

MA-42120 SERIES

specification sheet

ELECTRICAL CHARACTERISTICS (25°C Ambient Temperature Unless Otherwise Noted)

Symbol	Characteristics	MA42120 MA42122 Min	Typ	Max	MA42121 MA42123 Min	Typ	Max	Units	Test Conditions
h_{FE}	DC Current Gain	20	45	110	40	75	185		I_C = 5.0 mA, V_{CE} = 5.0
V_{CE} (sat)	Pulsed Collector Saturation Voltage[1]		0.30	0.38		0.25	0.35	V	I_C = 80 mA, I_B = 8.0 mA
V_{BE} (sat)	Pulsed Base Saturation Voltage[1]		0.95	0.98		0.93	0.96	V	I_C = 40 mA, I_B = 20 mA
BV_{CBO}	Collector to Base Breakdown	30	35		30	35		V	I_C = 10 μA, I_E = 0
V_{CEO} (sus)	Collector to Emitter Sustaining Voltage	14	16.5		14	16.5		V	I_C = 1.0 mA, I_B = 0
I_{EBO}	Emitter Cutoff Current		20	100		20	100	μA	I_C = 0, V_{EB} = 4.0V
I_{CBO}	Collector Cutoff Current		0.01	50		0.01	50	nA	V_{CB} = 10V, I_E = 0
I_{CBO}	Collector Cutoff Current		0.3	1.0		0.3	1.0	μA	V_{CB} = 10V, I_E = 0 T_A = 125°C
C_{cb}	Collector to Base Capacitance (MA42120, MA42121)		1.0	1.4		1:0	1.4	pF	V_{CB} = 10V, I_E = 0
C_{cb}	Collector to Base Capacitance (MA42122, MA42123)		0.85	1.0		0.85	1.0	pF	V_{CB} = 10V, I_E = 0
C_{eb}	Emitter to Base Capacitance		1.5	3.0		1.5	3.0	pF	V_{EB} = 0.5V, I_C = 0
$\|h_{fe}\|$	Magnitude of High Frequency Current Gain	2.0	2.6		2.6	3.0			V_{CE} = 10V, I_C = 20 mA, f = 500 MHz

NOTE:
1. Pulse Conditions: Length = 300 μs; duty cycle = 1%

MAXIMUM RATING
(Case Temperature 25°C unless otherwise noted)

	Total Power Dissipation	508 case — 1.0 W 509 case — .5 W
V_{CBO}	Collector to Base Voltage	30 V
V_{EBO}	Emitter to Base Voltage	4.0 V
V_{CES}	Collector to Emitter Voltage	30 V
I_C	Collector Current	80 mA
	Storage Temperature	−65° to +200°C
	Operating Junction Temperature	+200°C
	Lead Temperature (Soldering — 10 seconds each lead)	+250°C
	Hermeticity	5 x $(10)^{-8}$ cc/sec of He

ENVIRONMENTAL RATINGS PER MIL-STD-750

	Method	Level
Storage Temperature	1031	−65 to +200°C
Temperature Cycle	1051	10 cycles −65 to +200°C
Shock	2016	500 g's
Vibration	2056	15 g's
Constant Acceleration	2006	20,000 g's
Humidity	1021	10 days

FIG. 6-29. (Continued).

EXAMPLE 6-7—Cont

MA-42120 SERIES

specification sheet

MA-42120-508
TYPICAL COMMON-EMITTER S-PARAMETERS AT 25°C LEAD TEMPERATURE
V_{CE} = 10 VOLTS, $Z_G = Z_L$ = 50 Ω

FREQUENCY (MHz)	COLLECTOR CURRENT (mA)	INPUT REFLECTION COEFFICIENT		FORWARD TRANSMISSION COEFFICIENT		REVERSE TRANSMISSION COEFFICIENT		OUTPUT REFLECTION COEFFICIENT	
		$\lvert S_{11E}\rvert$	ϕS_{11E}	$\lvert S_{21E}\rvert$	ϕS_{21E}	$\lvert S_{12E}\rvert$	ϕS_{12E}	$\lvert S_{22E}\rvert$	ϕS_{22E}
100	1.5	.89	−44°	3.8	146°	.05	65°	.94	−13°
	5.0	.69	−79°	8.6	126°	.03	55°	.81	−18°
	20.0	.50	−127°	12.0	107°	.02	52°	.67	−16°
200	1.5	.80	−72°	3.3	128°	.06	51°	.86	−20°
	5.0	.60	−113°	6.1	107°	.04	49°	.71	−21°
	20.0	.50	−150°	7.4	92°	.02	62°	.61	−16°
300	1.5	.72	−95°	2.7	112°	.07	43°	.82	−25°
	5.0	.55	−134°	4.3	96°	.04	50°	.68	−23°
	20.0	.51	−161°	5.0	83°	.03	68°	.61	−18°
400	1.5	.66	−115°	2.2	101°	.08	38°	.79	−29°
	5.0	.53	−148°	3.4	86°	.05	54°	.66	−25°
	20.0	.52	−169°	3.8	75°	.04	72°	.62	−20°
450	1.5	.64	−123°	2.1	95°	.07	38°	.77	−30°
	5.0	.53	−154°	3.1	81°	.05	58°	.66	−27°
	20.0	.53	−172°	3.4	71°	.04	77°	.63	−21°
500	1.5	.63	−130°	1.9	90°	.07	39°	.76	−32°
	5.0	.53	−159°	2.7	77°	.05	62°	.66	−28°
	20.0	.54	−175°	3.0	67°	.05	79°	.63	−23°
600	1.5	.61	−143°	1.6	83°	.07	40°	.73	−35°
	5.0	.53	−167°	2.3	72°	.06	66°	.65	−31°
	20.0	.55	178°	2.5	63°	.06	82°	.63	−26°
700	1.5	.61	−1[illegible]°	1.5	76°	.07	46°	.70	−39°
	5.0	.55	−174°	2.0	65°	.06	74°	.63	−34°
	20.0	.57	174°	2.2	56°	.06	87°	.62	−29°
800	1.5	.61	−162°	1.3	70°	.06	55°	.68	−44°
	5.0	.56	179°	1.8	61°	.07	79°	.62	−38°
	20.0	.59	170°	1.8	52°	.08	91°	.62	−34°
900	1.5	.61	−169°	1.2	66°	.07	64°	.68	−50°
	5.0	.56	174°	1.6	57°	.08	84°	.62	−44°
	20.0	.60	165°	1.6	49°	.09	94°	.62	−41°
1000	1.5	.61	−176°	1.1	60°	.07	77°	.68	−55°
	5.0	.56	169°	1.4	52°	.09	91°	.63	−50°
	20.0	.61	160°	1.5	44°	.10	99°	.63	−47°

MA 42121-508
TYPICAL COMMON-EMITTER S-PARAMETERS AT 25°C LEAD TEMPERATURE
V_{CE} = 10 VOLTS, $Z_G = Z_L$ = 50

FREQUENCY (MHz)	COLLECTOR CURRENT (mA)	INPUT REFLECTION COEFFICIENT		FORWARD TRANSMISSION COEFFICIENT		REVERSE TRANSMISSION COEFFICIENT		OUTPUT REFLECTION COEFFICIENT	
		$\lvert S_{11E}\rvert$	ϕS_{11E}	$\lvert S_{21E}\rvert$	ϕS_{21E}	$\lvert S_{12E}\rvert$	ϕS_{12E}	$\lvert S_{22E}\rvert$	ϕS_{22E}
100	1.5	.92	−36°	3.8	150°	.04	67°	.95	−13°
	5.0	.74	−67°	9.1	131°	.03	59°	.82	−19°
	20.0	.51	−118°	13	109°	.02	55°	.66	−18°
200	1.5	.83	−59°	3.4	133°	.07	55°	.86	−21°
	5.0	.61	−98°	6.7	111°	.04	48°	.69	−24°
	20.0	.49	−142°	8.0	93°	.03	59°	.59	−18°
300	1.5	.74	−81°	2.9	117°	.08	47°	.81	−27°
	5.0	.54	−120°	4.9	99°	.05	50°	.65	−26°
	20.0	.49	−156°	5.5	84°	.03	65°	.58	−19°
400	1.5	.67	−99°	2.5	106°	.09	41°	.77	−31°
	5.0	.51	−136°	3.9	89°	.06	51°	.62	−28°
	20.0	.50	−164°	4.2	77°	.04	67°	.59	−22°
450	1.5	.65	−106°	2.3	100°	.09	40°	.75	−33°
	5.0	.50	−142°	3.5	85°	.06	53°	.62	−29°
	20.0	.51	−167°	3.7	73°	.05	71°	.59	−23°
500	1.5	.62	−114°	2.1	95°	.09	40°	.74	−35°
	5.0	.50	−147°	3.1	81°	.06	55°	.62	−30°
	20.0	51	170°	3.3	69°	.05	74°	.60	−24°
600	1.5	.59	−127°	1.8	88°	.09	38°	.71	−37°
	5.0	.49	−157°	2.6	75°	.06	58°	.61	−32°
	20.0	.53	−176°	2.7	64°	.06	76°	.60	−26°
700	1.5	.58	−138°	1.6	80°	.09	39°	.68	−41°
	5.0	.50	−164°	2.3	68°	.07	63°	.59	−35°
	20.0	.55	179°	2.4	58°	.06	81°	59	−30°
800	1.5	.58	−147°	1.5	74°	.08	43°	.65	−45°
	5.0	.51	−170°	2.0	63°	.08	68°	.57	−39°
	20.0	.57	175°	2.0	54°	.07	84°	59	−35°
900	1.5	.58	−154°	1.3	69°	.08	47°	.65	−51°
	5.0	.52	−174°	1.8	60°	.08	71°	.57	−44°
	20.0	.58	171°	1.8	51°	.08	88°	.59	−41°
1000	1.5	.57	−161°	1.2	64°	.08	56°	.65	−56°
	5.0	.52	−179°	1.6	55°	.09	78°	.59	−49°
	20.0	.59	166°	1.6	46°	.09	93°	.61	−46°

FIG. 6-29. (Continued).

EXAMPLE 6-7—Cont

MA-42120 SERIES

specification sheet

MA 42122-509
TYPICAL COMMON-EMITTER S-PARAMETERS AT 25°C LEAD TEMPERATURE
V_{CE} = 10 VOLTS, $Z_G = Z_L = 50\ \Omega$

FREQUENCY (MHz)	COLLECTOR CURRENT (mA)	INPUT REFLECTION COEFFICIENT		FORWARD TRANSMISSION COEFFICIENT		REVERSE TRANSMISSION COEFFICIENT		OUTPUT REFLECTION COEFFICIENT	
		$\lvert S_{11E}\rvert$	ϕS_{11E}	$\lvert S_{21E}\rvert$	ϕS_{21E}	$\lvert S_{12E}\rvert$	ϕS_{12E}	$\lvert S_{22E}\rvert$	ϕS_{22E}
100	1.5	.87	−43°	3.7	145°	.03	63°	.96	−10°
	5.0	.66	−70°	8.1	127°	.02	53°	.87	−12°
	20.0	.43	−103°	11.2	109°	.02	55°	.78	−12°
200	1.5	.76	−69°	3.2	127°	.04	54°	.89	−15°
	5.0	.52	−102°	5.7	107°	.03	56°	.79	−16°
	20.0	.37	−132°	7.0	93°	.02	66°	.73	−13°
300	1.5	.66	−94°	2.6	110°	.05	47°	.86	−19°
	5.0	.45	−126°	4.1	94°	.03	58°	.77	−18°
	20.0	.36	−149°	4.8	83°	.03	70°	.73	−15°
400	1.5	.69	−115°	2.2	99°	.05	45°	.84	−23°
	5.0	.41	−144°	3.3	84°	.04	62°	.76	−21°
	20.0	.37	−162°	3.7	75°	.04	73°	.73	−17°
450	1.5	.56	−124°	2.0	93°	.05	47°	.83	−25°
	5.0	.40	−151°	2.9	79°	.04	66°	.76	−22°
	20.0	.37	−167°	3.3	70°	.04	76°	.73	−19°
500	1.5	.54	−132°	1.8	88°	.05	49°	.82	−26°
	5.0	.40	−157°	2.6	75°	.05	69°	.75	−23°
	20.0	.37	−172°	2.9	66°	.05	79°	.73	−21°
600	1.5	.61	−148°	1.6	80°	.05	53°	.80	−29°
	5.0	.40	−169°	2.2	69°	.05	74°	.74	−26°
	20.0	.39	179°	2.4	60°	.06	81°	.73	−23°
700	1.5	.50	−160°	1.4	72°	.05	63°	.77	−33°
	5.0	.40	−179°	2.0	62°	.06	79°	.72	−29°
	20.0	.41	172°	2.1	53°	.06	84°	.72	−27°
800	1.5	.50	−171°	1.3	66°	.08	72°	.74	−38°
	5.0	.41	173°	1.7	57°	.07	83°	.71	−34°
	20.0	.43	166°	1.8	49°	.08	87°	.70	−32°
900	1.5	.49	178°	1.2	61°	.07	80°	.74	−43°
	5.0	.41	166°	1.6	52°	.08	85°	.70	−39°
	20.0	.43	159°	1.6	45°	.09	88°	.70	−37°
1000	1.5	.48	168°	1.1	55°	.08	89°	.74	−48°
	5.0	.41	167°	1.4	47°	.10	90°	.71	−45°
	20.0	.44	151°	1.5	39°	.11	93°	.70	−42°

MA 42123-509
TYPICAL COMMON-EMITTER S-PARAMETERS AT 25°C LEAD TEMPERATURE
V_{CE} = 10 VOLTS, $Z_G = Z_L = 50\ \Omega$

FREQUENCY (MHz)	COLLECTOR CURRENT (mA)	INPUT REFLECTION COEFFICIENT		FORWARD TRANSMISSION COEFFICIENT		REVERSE TRANSMISSION COEFFICIENT		OUTPUT REFLECTION COEFFICIENT	
		$\lvert S_{11E}\rvert$	ϕS_{11E}	$\lvert S_{21E}\rvert$	ϕS_{21E}	$\lvert S_{12E}\rvert$	ϕS_{12E}	$\lvert S_{22E}\rvert$	ϕS_{22E}
100	1.5	.89	−39°	3.8	147°	.04	54°	.95	−11°
	5.0	.68	−85°	8.6	128°	.03	56°	.85	−15°
	20.0	.43	−121°	10.9	104°	.02	57°	.67	−16°
200	1.5	.79	−63°	3.3	129°	.05	56°	.87	−17°
	5.0	.53	−95°	6.1	108°	.04	55°	.75	−17°
	20.0	.40	−144°	6.6	89°	.03	66°	.61	−16°
300	1.5	.69	−86°	2.7	113°	.06	48°	.84	−21°
	5.0	.45	−118°	4.4	95°	.05	57°	.72	−20°
	20.0	.41	−158°	4.5	80°	.04	70°	.61	−18°
400	1.5	.61	−105°	2.3	101°	.07	44°	.81	−25°
	5.0	.41	−135°	3.5	86°	.05	58°	.71	−22°
	20.0	.41	−168°	3.5	72°	.06	72°	.61	−21°
450	1.5	.58	−114°	2.2	96°	.07	45°	.79	−27°
	5.0	.40	−141°	3.2	81°	.05	62°	.70	−23°
	20.0	.42	−172°	3.1	67°	.06	73°	.61	−22°
500	1.5	.55	−121°	2.0	91°	.07	46°	.79	−28°
	5.0	.39	−148°	2.8	77°	.06	65°	.70	−24°
	20.0	.43	−176°	2.7	64°	.07	76°	.62	−24°
600	1.5	.52	−136°	1.7	83°	.07	46°	.76	−31°
	5.0	.39	−159°	2.4	71°	.06	66°	.69	−27°
	20.0	.45	177°	2.3	58°	.08	77°	.62	−27°
700	1.5	.50	−148°	1.6	74°	.07	51°	.73	−34°
	5.0	.39	−168°	2.1	64°	.07	70°	.67	−29°
	20.0	.47	171°	2.0	51°	.09	80°	.61	−31°
800	1.5	.50	−159°	1.4	68°	.07	57°	.71	−39°
	5.0	.39	−176°	1.9	59°	.08	74°	.66	−33°
	20.0	.48	166°	1.7	47°	.11	83°	.60	−36°
900	1.5	.49	−167°	1.3	63°	.07	64°	.70	−43°
	5.0	.39	178°	1.7	55°	.09	76°	.65	−38°
	20.0	.49	160°	1.5	43°	.12	84°	.59	−43°
1000	1.5	.47	−176°	1.2	57°	.08	74°	.70	−49°
	5.0	.39	171°	1.5	49°	.10	81°	.66	−43°
	20.0	.50	154°	1.4	37°	.14	87°	.60	−49°

FIG. 6-29. (Continued).

EXAMPLE 6-7—Cont

provide the needed gain as long as the input of the device is conjugately matched and as long as the impedance required for a conjugate match falls inside of the input stability circle.

Choose Γ_L equal to a convenient value on the 12-dB gain circle.

$$\Gamma_L = 0.89\angle 70^\circ$$

Using Equation 6-21, calculate the source-reflection coefficient needed for a conjugate match and plot this point on the Smith Chart.

$$\Gamma_S = 0.678\angle 79.4^\circ$$

Notice that Γ_S falls within the stable region of the input stability circle and, therefore, represents a stable termination for the transistor.

On page 2 of the data sheet, you will find a set of curves labeled "Typical Optimum N.F. vs. Collector Current." Note that for this particular device, at 450 MHz, the optimum collector current for minimum noise figure is approximately 1.5 mA. This value of collector current should result in a noise figure of just above 2 dB. Again, the data is presented for only 60 MHz and 450 MHz.

Designing amplifiers for a minimum noise figure is simply a matter of determining, either experimentally or from the data sheet, the source resistance and the bias point that produce the minimum noise figure for the device (Example 6-8). Once determined, the actual source impedance is simply forced to "look like" the optimum value. Of course, all stability considerations still apply. If the Rollett stability factor (K) calculates to be less than 1, then you must be careful in your choice of source- and load-reflection coefficients. It is best, in this case, to draw the stability circles for an accurate graphical indication of where the unstable regions lie.

After providing the transistor with its optimum source impedance, the next step is to determine the optimum load-reflection coefficient needed to properly terminate the transistor's output. This is given by:

$$\Gamma_L = \left[S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}\right]^* \qquad \text{(Eq. 6-33)}$$

where

Γ_S is the source-reflection coefficient for minimum noise figure.

EXAMPLE 6-8

It has been determined that the optimum bias point for minimum noise figure for a transistor is $V_{CE} = 10$ V and $I_C = 5$ mA. Its optimum source-reflection coefficient, as given on the data sheet, is:

$$\Gamma_S = 0.7\angle 140^\circ$$

The S parameters for the transistor, under the given bias conditions at 200 MHz, are:

$$S_{11} = 0.4\angle 162^\circ$$

$$S_{22} = 0.35\angle -39^\circ$$

$$S_{12} = 0.04\angle 60^\circ$$

$$S_{21} = 5.2\angle 63^\circ$$

Design a low-noise amplifier to operate between a 75-ohm source and a 100-ohm load at 200 MHz. What gain can you expect from the amplifier when it is built?

Solution

The Rollett stability factor (K) calculates to be 1.74 which indicates unconditional stability (Equation 6-15). Therefore, we may proceed with the design. The design values of the input-matching network are shown in Fig. 6-30. Here the normalized 75-ohm source resistance is transformed to Γ_S using two components.

$$\text{Arc AB} = \text{Shunt C} = j1.7 \text{ mhos}$$

$$\text{Arc BC} = \text{Series L} = j0.86 \text{ ohm}$$

Using Equations 4-11 through 4-14, the component values are calculated to be:

$$C_1 = \frac{1.7}{(50)(2\pi)(200 \times 10^6)}$$

$$= 27 \text{ pF}$$

$$L_1 = \frac{(0.86)(50)}{2\pi(200 \times 10^6)}$$

$$= 34 \text{ nH}$$

The load-reflection coefficient needed to properly terminate the transistor is then found using Equation 6-33.

$$\Gamma_L = \left[0.35\angle -39^\circ + \frac{(0.04\angle 60^\circ)(5.2\angle 63^\circ)(0.7\angle 140^\circ)}{1 - (0.4\angle 162^\circ)(0.7\angle 140^\circ)}\right]^*$$

$$= 0.427\angle 60.7^\circ$$

Continued on next page

EXAMPLE 6-8—Cont

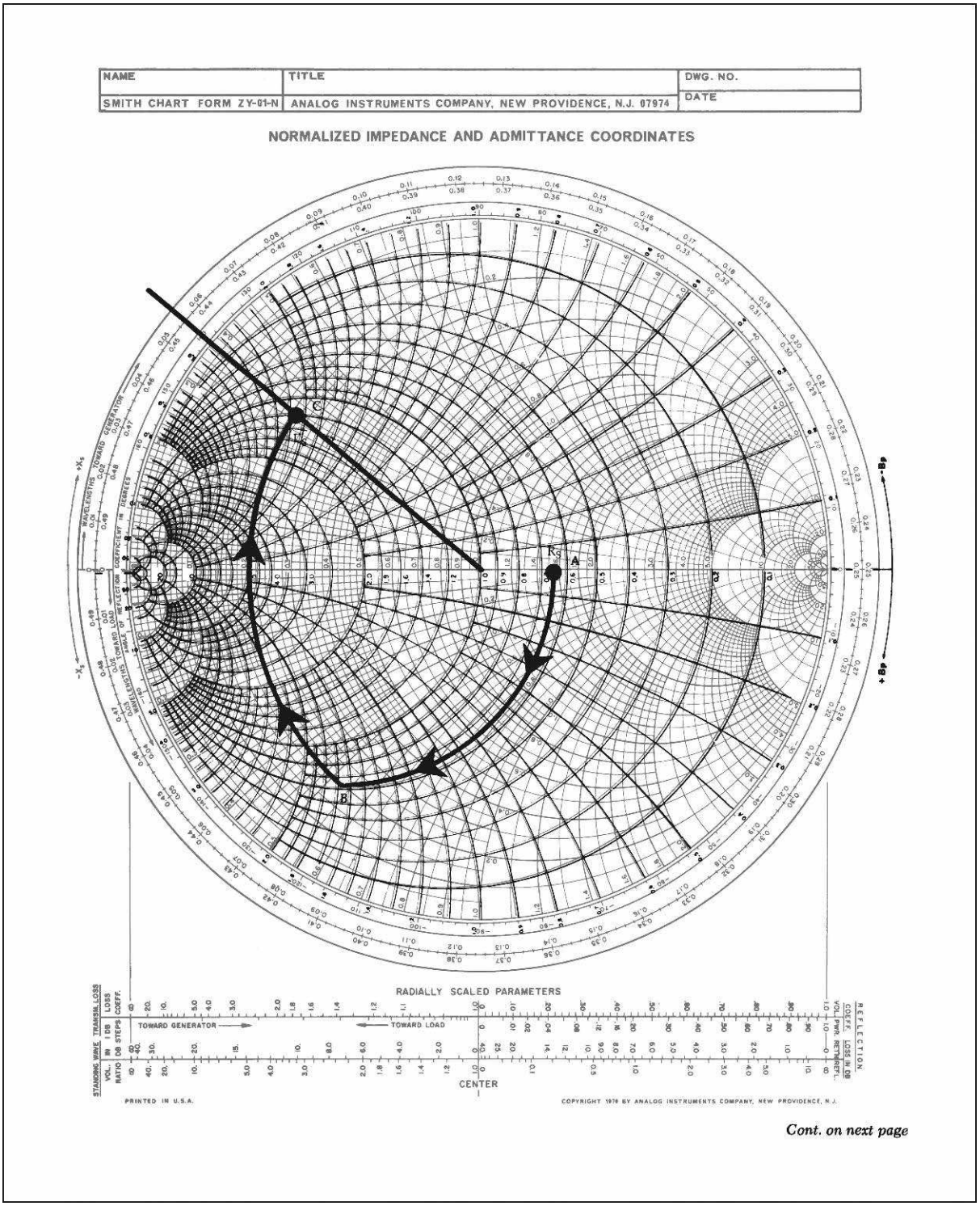


FIG. 6-30. Input network-design values for Example 6-8. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-8—Cont

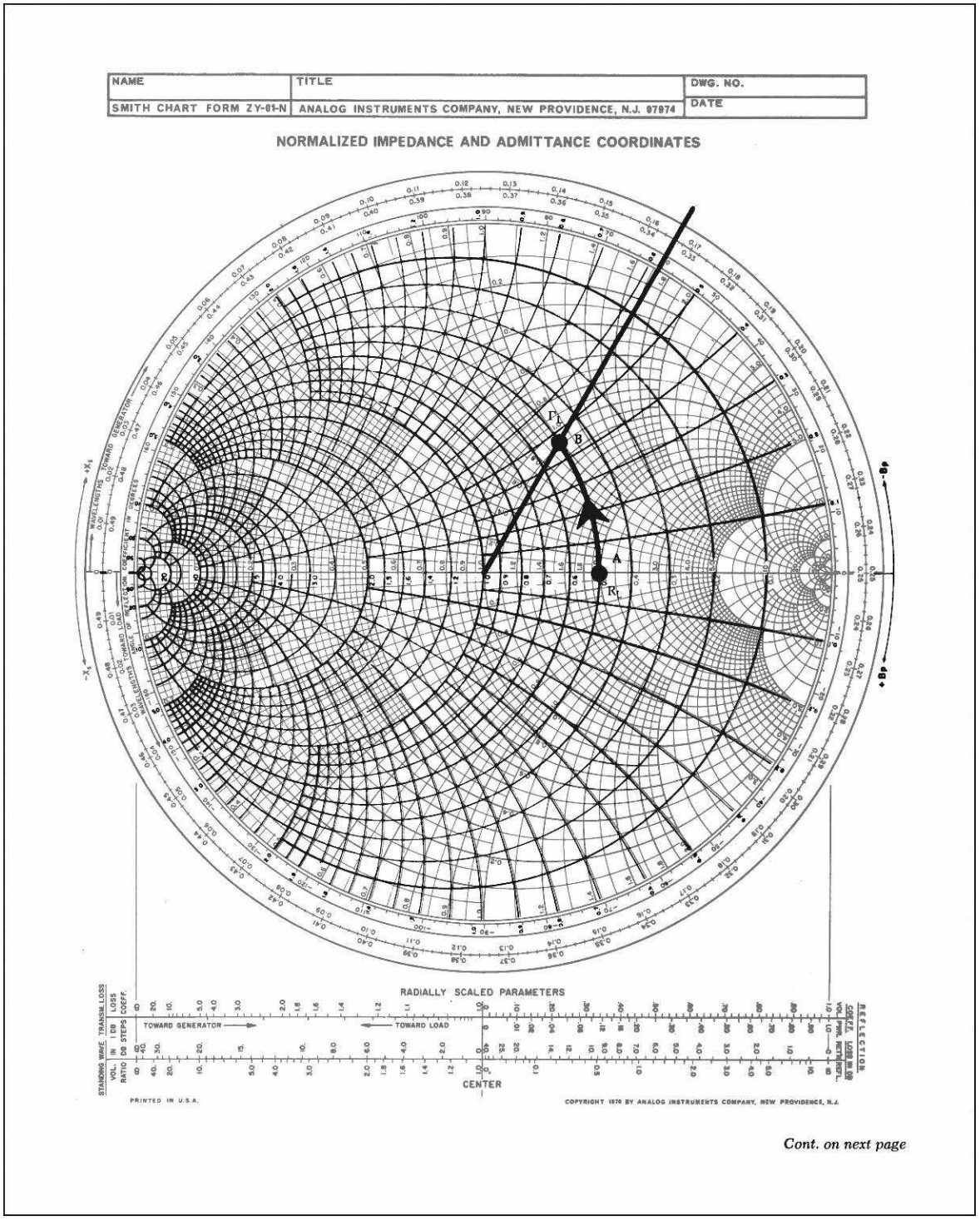


FIG. 6-31. Output network-design values for Example 6-8. For a more detailed full color view of this figure, please visit our companion site at http://books.elsevier.com/companions/9780750685184.

EXAMPLE 6-8

This value, along with the normalized load-resistance value, is plotted in Fig. 6-31. The 100-ohm load must be transformed into Γ_L. One possible method is shown in Fig. 6-31. Note that a single shunt inductor provides the necessary impedance transformation:

$$\text{Arc AB} = \text{Shunt L} = -j0.48\,\text{mho}$$

Again using Equations 4-11 through 4-14, the inductor's value is found to be:

$$L_2 = \frac{50}{2\pi(200 \times 10^6)(0.48)}$$
$$= 83\,\text{nH}$$

The final design, including a typical bias network, is shown in Fig. 6-32. The 0.1-μF capacitors are used only as bypass and coupling elements. The gain of the amplifier, as calculated with Equation 6-22, is 13.3 dB.

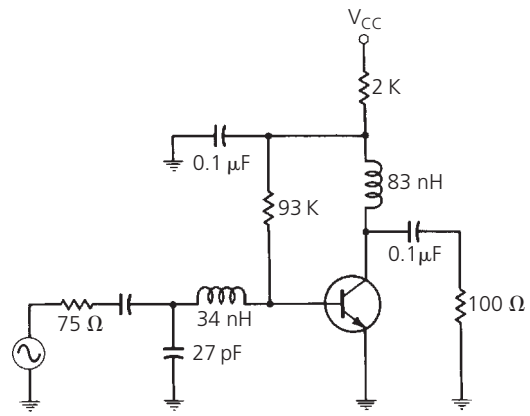


FIG. 6-32. Final circuit for Example 6-8.

Design Example

Many of the techniques discussed in this chapter can be accomplished using software design tools. To better illustrate this fact, consider the design of a low-noise amplifier block that will be used in a dual-band down converter. The design is completed using the Genesys environment from Agilent Technologies (www.agilent.com/find/eesof). Genesys software is an integrated electronic design automation (EDA) platform for RF and microwave design. It features a design flow that spans from initial system architecture through final documentation, and provides state-of-the-art performance in a single easy-to-use design environment that is fast, powerful, and accurate.

A summary of the design specifications and goals are shown in Table 6-1. Note that the required gain of 30 dB will exceed the capability of most active single devices; therefore our design will require a minimum of two devices with transducer gains equal to or greater than 15 dB each. The combination of gain, noise figure, power and match will prove to be challenging. NEC's NE23418 device was chosen because it will provide >17 dB gain at 2400 MHz with a noise figure of less than 1 dB.

Frequency Range	2200–2600 MHz 400 MHz BW
Gain	30 dB ± 1 dB
Noise Figure	<2 dB
Input/Output Match	−20 dB
P1dB	0 dBm
PSat	+3 dBm
TOI	+10 dBm

TABLE 6-1. Design specifications

To begin, open the Genesys workspace and select the NE23418 part from the NEC SPICE Parts library. Place the selected part on the schematic. Then place the bias components on the schematic along with the sources and ports as shown in Fig. 6-33. The component values are listed in Table 6-2.

The next step requires us to simulate the circuit to extract the bias values and linear S-parameters. To do this, select Analyses/Add Linear Analysis. When the linear analysis dialog appears, set the

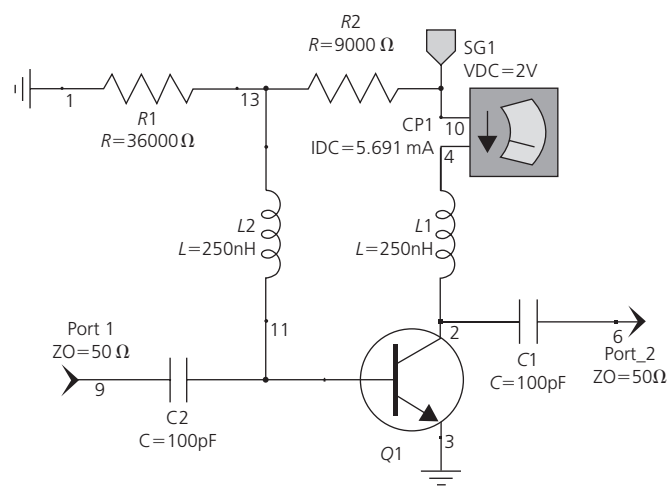


FIG. 6-33. Design schematic.

R1	36000
R2	9000
SG1	2V
L1	250 nH
L2	250 nH
C1	100 pF
C2	100 pF

TABLE 6-2. Component values

start and stop frequencies at 1500 MHz to 3000 MHz with 101 points. Accept the remaining default settings. DC analysis and linear circuit simulation will take place automatically.

When the simulation is complete, right click on the output port in the schematic and select S[2,1] as a measurement to graph. When the graph appears, double click anywhere on the graph's surface to bring up the graph's properties dialog. Select the Measurement Wizard button, then Linear1_Data which points to the dataset. This will bring up a third dialog, allowing the choice of measurement to plot. Select NFMIN for noise figure and complete the action by pressing the Finish button and accepting the subsequent selections. After closing the graph properties, your graph should resemble Fig. 6-34.

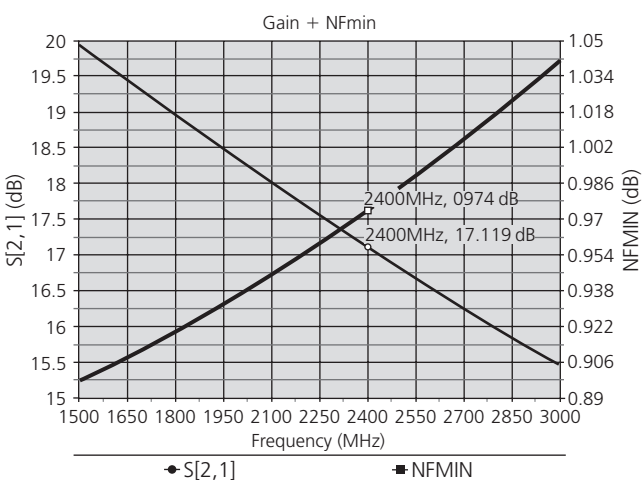


FIG. 6-34. Measurement plot.

Click on the graph line to attach a marker. The value of gain should be approximately 17 dB with less than 1 dB noise figure.

Next we will plot circles of constant noise figure as well as Gopt, S11, S22 and input and output stability circles. To do this, select Graphs/Add Smith Chart to launch the Smith Chart graph. Double click on the chart to bring up the graph properties. Next select the Measurement Wizard button, and the Linear1_Data set to bring up the measurements dialog. Scroll down the selection until the measurement NCI constant noise circles are found. Select this measurement and accept the results. Close the graph dialog and view the graph. What you are viewing is Gopt, which is the optimum reflection coefficient for minimum noise over the analysis frequency range. To view the group of constant noise circles, click on the plot. Where the marker falls will dictate the frequency for these circles. With the marker activated, use the cursor to tune across the band and note how the position and diameter of the circles vary. The circles represent the locus of constant noise figure. From Genesys help, you will read that the first circle represents a degradation of .25 dB, followed by .5 dB, then 1 dB, 1.5 dB and so forth. Note that if we choose to terminate the input of our amplifier with a 50-ohm termination, the noise figure would be degraded by less than .25 dB above the minimum noise figure.

Add another Smith Chart as before; however, this time we will plot Gopt, S11, S22 and input and output stability circles. To save time, enter the measurements to plot directly on the graph properties dialog screen. For completeness, the measurements are S11, S22, Gopt, SB1, and SB2 for the stability circles. Close the graph properties dialog. Use the Window/Tile function to place all four windows evenly on the screen. Your workspace should resemble Fig. 6-35.

Double click on the DC source in the schematic SG1, check the tune box and close. Using the Tune window, vary the source voltage from 1 to 3 V and note the change in linear parameters as well as stability circles. As the bias is increased, the stability circles approach the unit circle, allowing for easier matching but lower noise figure. Reset the voltage to 2 V, which provides the recommended gain and required noise figure. The fact that there are stability circles inside the unit circle and not enclosing the center tells us that we cannot present loads or sources in these regions; otherwise there is a potential for oscillation. In other words, we cannot provide a simultaneous conjugate match at both the input and output of our device. This is also evident by plotting the stability factor K from the measurements available in our graphs or tables. The implication is we cannot achieve Gmax for our device, but instead something close to Gmsg or maximum stable gain.

Having completed device selection, bias and linear evaluation, we will now use the basic biased stage for incorporation into a two-stage design. Select Synthesis/Add Impedance Match from the menu to bring up the MATCH dialog. Select the start and stop frequencies over which we wish to match our amplifier. Set the lower frequency to 2200 MHz and the upper to 2400 MHz with the number of points 50. Press the Sections tab to bring up the network/topology window and settings. Accept the default of 50 ohms for input and output terminations.

With the input port selected, click on the Add Device button. This will add a generic two port block to our topology. The Type drop-down selection allows us to determine where the block will get its data from. For this exercise we will use the sub circuit that was built previously. Select the Design option from the Type drop-down and then from the Design drop-down select "SCH1" which is the name of the biased network. Next, click on the output

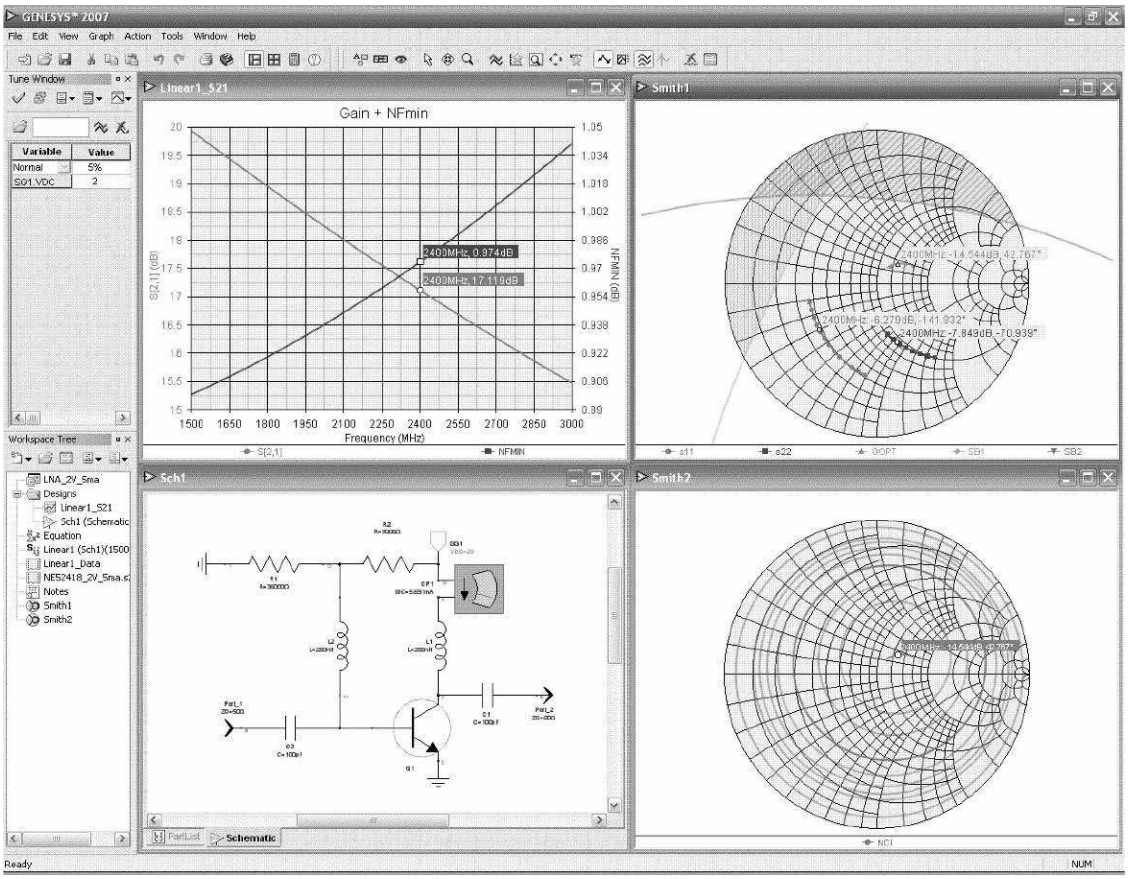


FIG. 6-35. Plots of stability circles.

port to activate it and then select the Add Device button to place our second stage in the topology. As above, select Design in the Type drop-down and point it to the same sub network "SCH1". Select the middle default matching network "LCPi". From the Type drop-down select "LC Bandpass". Insure that the Options drop-down shows "No Transformer" and the Order is set to 2. The topology should resemble that of Fig. 6-36.

Next, double click on the optimization icon. Select the Goals tab from the Optimization dialog box. Add an additional measurement of S21 to the goals properties and set its value to be equal to 30 dB. Also reset the goals for S11 and S22 to −15 dB along with their corresponding weights. Close the Optimization dialog.

Press the Calculate button on the MATCH properties dialog to have Genesys calculate the inter-stage matching structure. After a few seconds MATCH has generated a schematic with inter-stage topology and the associated devices in symbol form. In the matching circuit schematic, right click on the output port and select Add New Graph/Table/New Graph of S21. This will generate a rectangular plot of gain vs. frequency for our two-stage amplifier. Refer to Fig. 6-37.

Double click on the new graph to open its Properties dialog. Add a new measurement by typing in "NF" for noise figure in the field below "S[2,1]". Make sure that the NF measurement check box On Right is checked to display the noise figure on the right axis. Close the dialog; your graph should resemble Fig. 6-38.

The unoptimized gain varies from approximately 37.6 to 34.8 dB with a corresponding match of −.6 dB to approximately −2 dB worst case across our band. Press the Optimize button on the MATCH dialog to start the optimizer. After several seconds to a minute the error window should approach a value of 3–5 which will be the best to expect from the combination of goals and topology. The gain has been flattened to 30 dB ± .1 dB, well within the specification of ±1 dB. The noise figure has remained less than 2 dB across the band (approximately 1.07 maximum). Input and output match are still short of our goal of −20 dB across the band, but we have several options in this respect. We can add additional networks to the input and output or we can use a method involving couplers to improve broadband matching. Fig. 6-39 shows the gain, noise figure and matching achieved.

Reviewing the design goals at this point, we find that we have met the gain and noise figure requirements but fall short of the matching requirements by 4 dB. Unfortunately, for devices exhibiting marginal stability, improving match at one port degrades the match at the other. An obvious step would be to substitute our design into the dual-band receiver in which

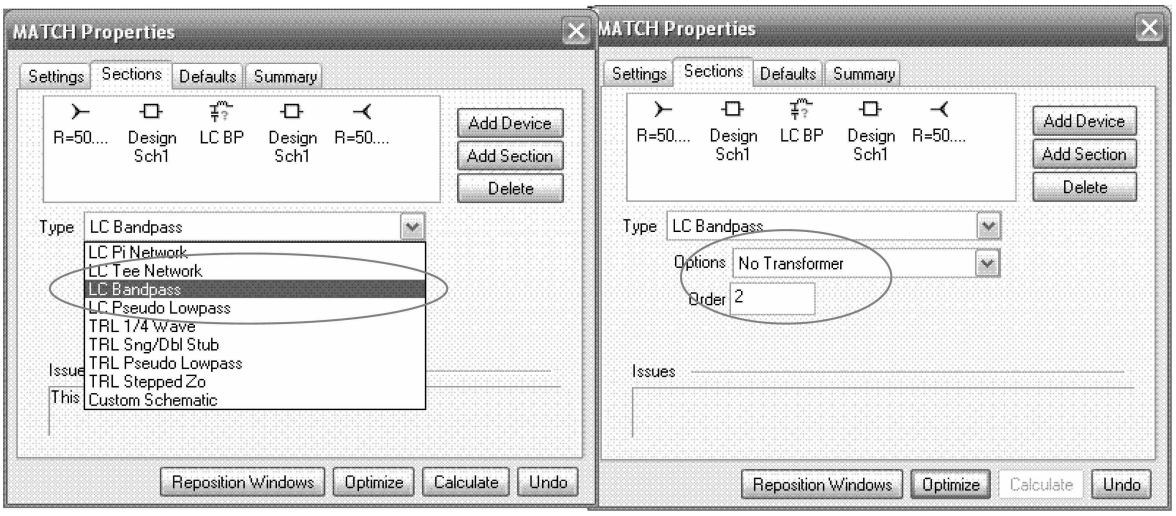


FIG. 6-36. Match properties screen.

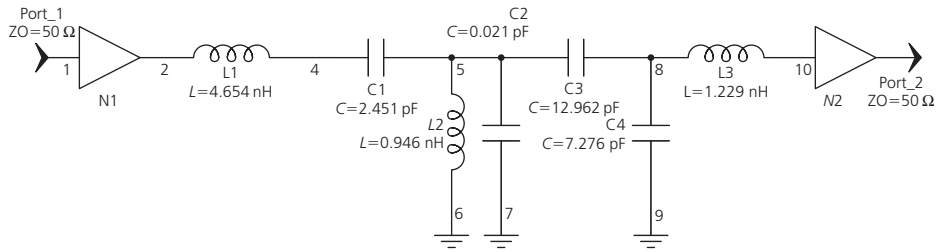


FIG. 6-37. Inter-stage topology schematic.

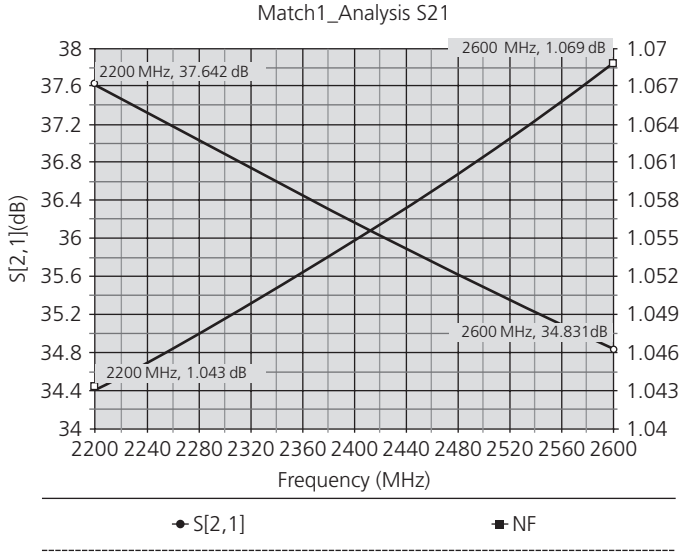


FIG. 6-38. Noise figure graph.

it was specified. If we find that using the amplifier "as is" in the system block diagram degrades the performance of the overall system then we must find another way to meet the return loss specifications. In addition we need to test the nonlinear performance of the amplifier to insure that power, saturation, etc., are met. In the Genesys environment, verification of nonlinear performance takes place via the harmonic balance simulator. Assume that this function is used to determine that our amplifier has met the +10 dBm TOI requirement. As a result, our design now meets or exceeds all requirements except the match specification.

The reason for requiring a good match is to ensure that the amplifier will not cause adverse interaction with connecting components. Since the amplifier is followed by a filter and mixer having a highly reflective output can reintroduce signals that will be remixed, and this may contribute to in-band spurious as well as degraded gain. Fortunately, in Genesys we have the ability to place the amplifier, as is, back into the system design to test the net performance prior to improving on the design.

To tackle the issue of input and output match we have several alternatives. We could attempt to revisit our matching structure

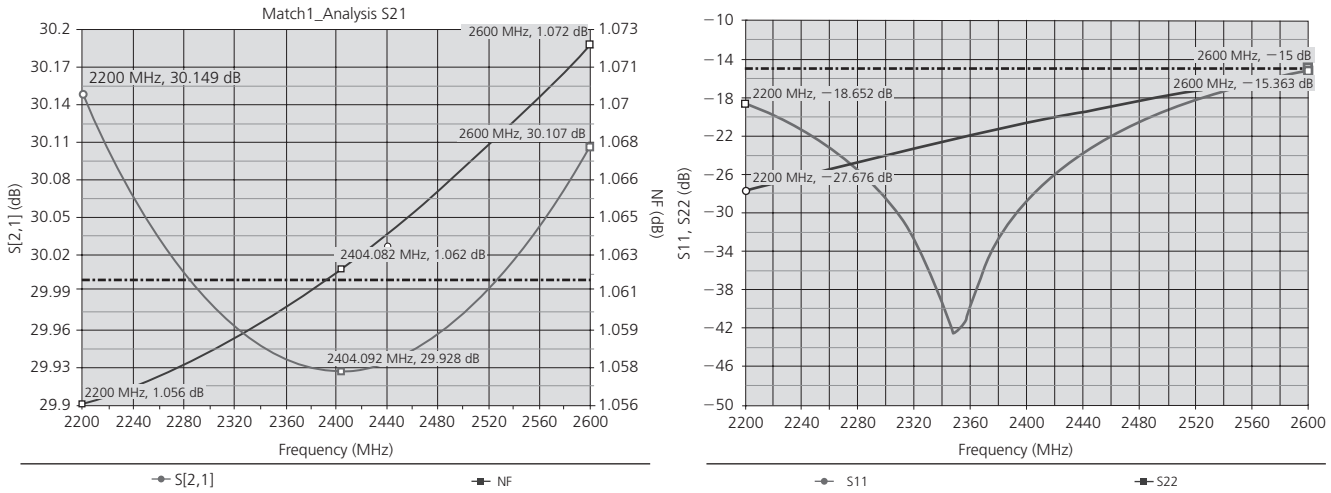


FIG. 6-39. Graphs of achieved gain, noise figure and matching.

and add additional elements to the input and output, but little would be gained if we still meet our other goals. Secondly, we could choose another device that might lend itself to a better overall match with sufficient gain. Third, a circulator or isolator component could be placed at the input and output. This is the least attractive because of size and cost. Another technique that has been used for some time involves a hybrid amplifier scheme where two identical amplifiers are driven in parallel by hybrid or 90° couplers. The benefit in this configuration is excellent match across the band, extended power output and suppression of even or odd harmonics, depending on the implementation. Also, if one amplifier fails, graceful degradation still provides gain.

CHAPTER 7

RF (LARGE SIGNAL) Power Amplifiers

In Chapters 5 and 6, we studied the transistor as a small-signal device. Y and S parameters were introduced as a means of facilitating amplifier design, and design equations were provided. When the transistor is used as a *large-signal device*, however, these equations are no longer valid. In fact, both Y and S parameters are called *small-signal* parameters, and should not be considered in the design of RF power amplifiers. After looking at basic power transistor characteristics in detail, we'll cover the leading semiconductor materials used in today's RF power ICs.

RF POWER TRANSISTOR CHARACTERISTICS

Instead of specifying the Y and S parameters for a power transistor, manufacturers will typically specify the *large-signal input impedance* and the *large-signal output impedance* for the device. These parameters are typically measured on the device when it is operating as a matched amplifier at the desired DC supply voltage and RF power output level. A matched amplifier, in this case, refers to a condition in which the input and output impedances are conjugately matched to the source and load, respectively.

The RF Power Transistor Data Sheet

Pertinent design information for RF power transistors is usually presented in the form of large-signal input and output impedances, as shown in Fig. 7-1. Fig. 7-1 is a data sheet for the Freescale (formerly Motorola) MRF233 RF power transistor. This particular data sheet was chosen for instructional purposes because it includes both series- and shunt-impedance information. This gives the circuit designer the opportunity of using an impedance format with which he is accustomed, without the need of converting from one format to the other.

Figure 5, on page 3 of the data sheet, is a Smith Chart representation of the *series* input and output impedance of the transistor (between 40 and 100 MHz). The information is also tabulated on the right side of the chart for your convenience. Note that the impedance is presented in the form $Z = R \pm jX$. Thus, at 100 MHz, the input impedance of the transistor is found to be $Z_{in} = 1.7 - j2.7$ ohms, while the output impedance is $Z_{out} = 5 - j5.6$ ohms. This equivalent-series representation for the transistor is shown in Fig. 7-2.

Figures 6, 7, 8, and 9, of the data sheet present the same impedance information in parallel form. The input and output impedance of the transistor are presented as a shunt resistance in parallel with a capacitor. Thus, referring to Figures 6 and 7 of the data sheet, the input impedance of the transistor is represented by a 6-ohm shunt resistor in parallel with a 422-pF capacitor. The curves of Figures 8 and 9 indicate an equivalent parallel output impedance for the transistor, which includes an 11.3-ohm resistor in parallel with a 158-pF capacitor, at 100 MHz. These shunt combinations are shown in the equivalent circuit of Fig. 7-3.

Note that you can perform your own transformation from series to shunt, and back again, by using Equations 2-6 and 2-7 and, then, following the procedure of Example 2-2.

Figure 2, on page 3 of the data sheet, is useful in helping you determine how much input signal power you will need to produce a given output power. Note that as the frequency of operation increases, the required input drive level increases. An input power to the transistor of 1 W will produce a 20-W output signal at 50 MHz (13-dB gain), while, at 90 MHz, that same input level will produce only 14 W out (11.5-dB gain).

Figure 3 presents the same basic information as Figure 2, but in a different format. Note that the output power decreases as the frequency of operation increases when given a constant input power level.

The remainder of the data sheet is straightforward and resembles that of any typical small-signal transistor.

TRANSISTOR BIASING

The type of bias applied to an RF power transistor is determined by the "class" of amplification that the designer wishes. There are many different classes of amplification available for the designer to choose from. The particular class chosen for a design will depend upon the application at hand.

The primary emphasis of this chapter will be on class-C amplifiers. However, class-A and class-B amplifier bias arrangements will also be covered.

NPN RF Power

MRF233

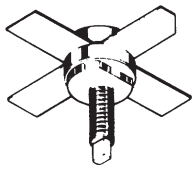
The RF Line

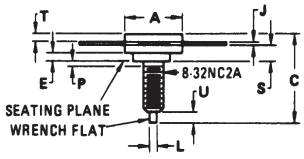
15 W – 90 MHz
RF POWER TRANSISTOR
NPN SILICON

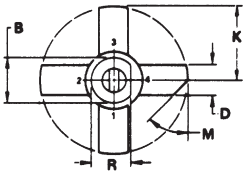
NPN SILICON RF POWER TRANSISTORS

. . . designed for 12.5 Volt, mid-band large-signal amplifier applications in industrial and commercial FM equipment operating in the 40 to 100 MHz range.

- Specified 12.5 Volt, 90 MHz Characteristics –
  Output Power = 15 Watts
  Minimum Gain = 10 dB
  Efficiency = 55%
- 100% Tested for Load Mismatch at all Phase Angles with 30:1 VSWR
- Characterized with Series Equivalent Large-Signal Impedance Parameters
- Characterized with Parallel Equivalent Large-Signal Impedance Parameters







STYLE 1:
PIN 1. EMITTER
2. BASE
3. EMITTER
4. COLLECTOR

DIM	MILLIMETERS MIN	MILLIMETERS MAX	INCHES MIN	INCHES MAX
A	9.40	9.78	0.370	0.385
B	8.13	8.38	0.320	0.330
C	17.02	20.07	0.670	0.790
D	5.46	5.97	0.215	0.235
E	1.78	–	0.070	–
J	0.08	0.18	0.003	0.007
K	12.45	–	0.490	–
L	1.40	1.78	0.055	0.070
M	45° NOM		45° NOM	
P	–	1.27	–	0.050
R	7.59	7.80	0.299	0.307
S	4.01	4.52	0.158	0.178
T	2.11	2.54	0.083	0.100
U	2.49	3.35	0.098	0.132

145A-09

MAXIMUM RATINGS

Rating	Symbol	Value	Unit
Collector-Emitter Voltage	V_{CEO}	18	Vdc
Collector-Base Voltage	V_{CBO}	36	Vdc
Emitter-Base Voltage	V_{EBO}	4.0	Vdc
Collector Current – Continuous	I_C	3.5	Adc
Total Device Dissipation @ $T_C = 25^oC$ (1) Derate Above 25^oC	P_D –	50 285	Watts $mW/^oC$
Storage Temperature Range	T_{stg}	-65 to +200	^oC
Stud Torque (2)	–	6.5	In-lb

THERMAL CHARACTERISTICS

Characteristic	Symbol	Max	Unit
Thermal Resistance, Junction to Case	$R_{\theta JC}$	3.5	$^oC/W$

(1) These devices are designed for RF operation. The total device dissipation rating applies only when the devices are operated as Class C RF amplifiers.
(2) For Repeated Assembly use 5 In. Lb.

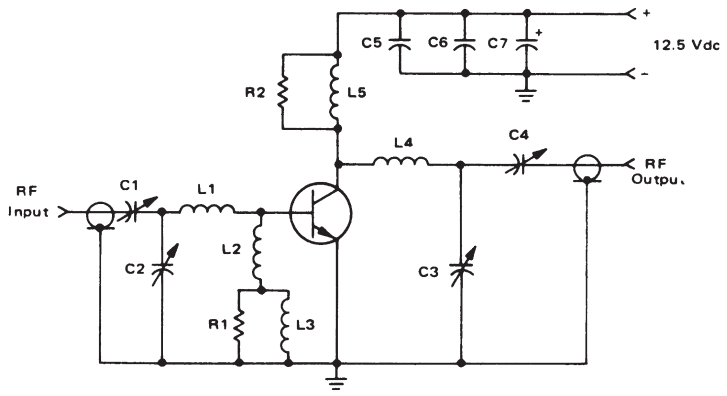
FIG. 7-1. Data sheet. (Courtesy Freescale (formerly Motorola Semiconductor Products Inc.))

MRF233

ELECTRICAL CHARACTERISTICS (T_C = 25°C unless otherwise noted).

Characteristic	Symbol	Min	Typ	Max	Unit
OFF CHARACTERISTICS					
Collector-Emitter Breakdown Voltage (I_C = 100 mAdc, I_B = 0)	BV_{CEO}	18	–	–	Vdc
Collector-Emitter Breakdown Voltage (I_C = 50 mAdc, V_{BE} = 0)	BV_{CES}	36	–	–	Vdc
Emitter-Base Breakdown Voltage (I_E = 5.0 mAdc, I_C = 0)	BV_{EBO}	4.0	–	–	Vdc
Collector Cutoff Current (V_{CB} = 15 Vdc, I_E = 0)	I_{CBO}	–	–	1.0	mAdc
ON CHARACTERISTICS					
DC Current Gain (I_C = 1.0 Adc, V_{CE} = 5.0 Vdc)	h_{FE}	5.0	–	–	–
DYNAMIC CHARACTERISTICS					
Output Capacitance (V_{CB} = 12.5 Vdc, I_E = 0, f = 1.0 MHz)	C_{ob}	–	100	120	pF
FUNCTIONAL TESTS (Figure 1)					
Common-Emitter Amplifier Power Gain (V_{CC} = 12.5 Vdc, P_{out} = 15 W, f = 90 MHz)	G_{PE}	10	–	–	dB
Collector Efficiency (V_{CC} = 12.5 Vdc, P_{out} = 15 W, f = 90 MHz)	η	55	–	–	%
Load Mismatch (V_{CC} = 12.5 Vdc, P_{out} = 15 W, f = 90 MHz, $T_C \leq 25°C$)	–	VSWR > 30:1 Through All Phase Angles in a 3 Second Interval After Which Devices Will Meet G_{PE} Test Limits			

FIGURE 1 – 90 MHz TEST CIRCUIT SCHEMATIC



C1,C3 9.0-180 pF, ARCO 463
C2,C4 25-280 pF ARCO 464
C5 1000 pF UNELCO
C6 0.01 μF ERIE Disc Ceramic
C7 1.0 μF, 35 Vdc TANTALUM
L1 2 Turns, #18 AWG, 3/8" I.D., 1/4" Long
L2 0.22 μH, 9230-04 MILLER Molded Choke
L3 2.2 μH, 9230-200 MILLER Molded Choke
L4 2 Turns, #18 AWG, 3/8" I.D., 3/8" Long
L5 10 Turns, #16 AWG, Wound On R2.
R1 15 Ohm, 1/2 W, 10% Carbon
R2 68 Ohm, 1 Watt, 10% Carbon
Input/Output Connectors – Type BNC

FIG. 7-1. (Continued)

MRF233

FIGURE 2 – OUTPUT POWER versus INPUT POWER

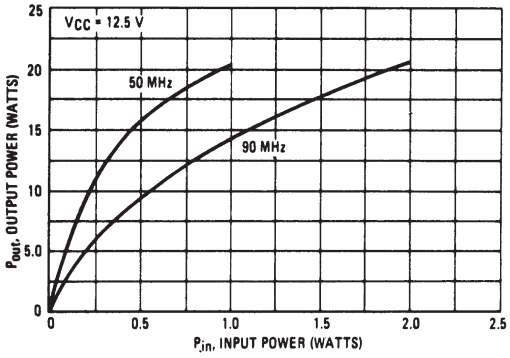


FIGURE 3 – OUTPUT POWER versus FREQUENCY

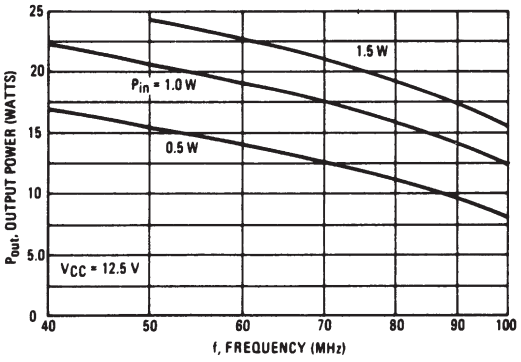


FIGURE 4 – OUTPUT POWER versus SUPPLY VOLTAGE

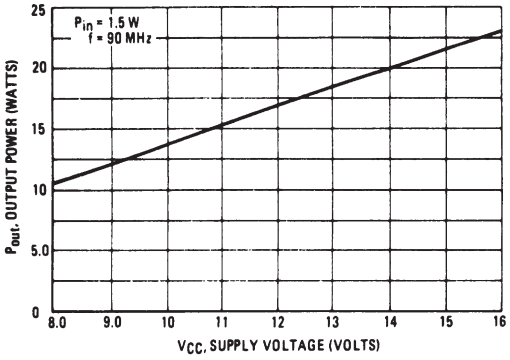


FIGURE 5 – SERIES EQUIVALENT IMPEDANCE

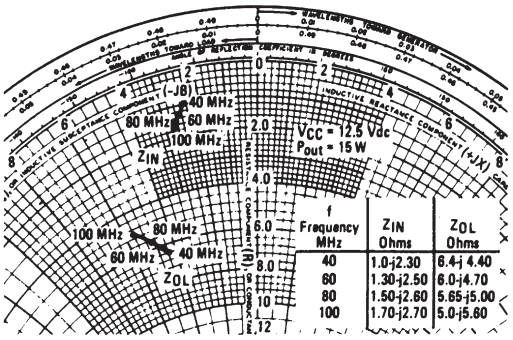


FIG. 7-1. (Continued)

MRF233

FIGURE 6 – PARALLEL EQUIVALENT INPUT RESISTANCE versus FREQUENCY

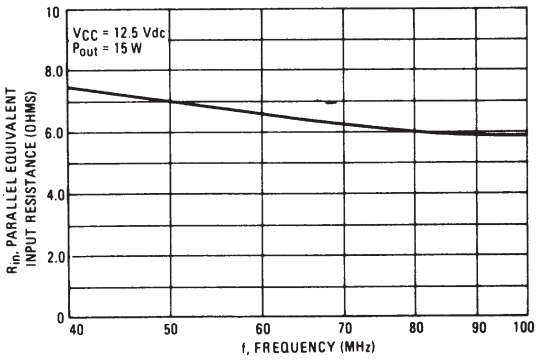


FIGURE 7 – PARALLEL EQUIVALENT INPUT CAPACITANCE versus FREQUENCY

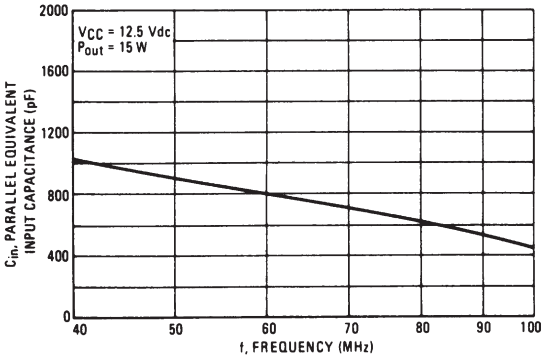


FIGURE 8 – PARALLEL EQUIVALENT OUTPUT RESISTANCE versus FREQUENCY

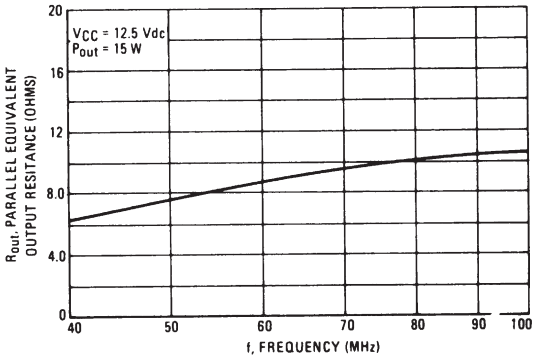


FIGURE 9 – PARALLEL EQUIVALENT OUTPUT CAPACITANCE versus FREQUENCY

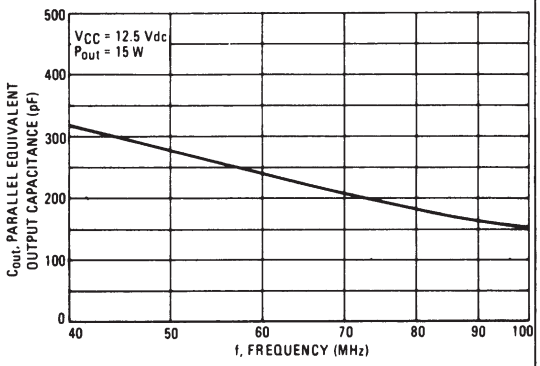


FIG. 7-1. (Continued)

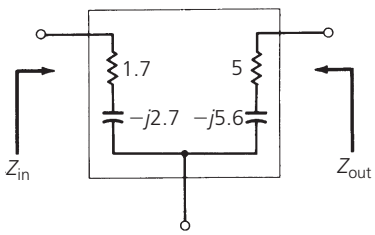


FIG. 7-2. Equivalent circuit for series input and output impedance at 100 MHz.

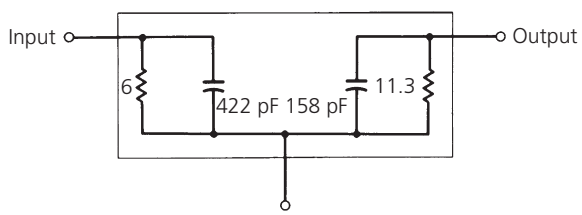


FIG. 7-3. Equivalent circuit for parallel input and output impedance at 100 MHz.

Class-A Amplifiers and Linearity

A class-A amplifier is defined as an amplifier that is biased so that the output current flows at all times. Thus, the input signal-drive level to the amplifier is kept small enough to avoid driving the transistor into cutoff. Another way of stating this is to say that the *conduction angle* of the transistor is 360°—meaning that the transistor conducts for the full cycle of the input signal.

The class-A amplifier is the most *linear* of all amplifier types. Linearity is simply a measure of how closely the output signal of the amplifier resembles the input signal. A linear amplifier is one in which the output signal is proportional to the input signal, as shown in Fig. 7-4. Notice that, in this case, the output signal level is equal to twice the input signal level, and the transfer function from input to output is a straight line.

No transistor is perfectly linear, however, and, therefore, the output signal of an amplifier is never an exact replica of the input signal. There are always spurious components added to a signal in the form of *harmonic* generation or *intermodulation distortion* (IMD). These types of nonlinearities in transistors produce amplifier transfer functions that no longer resemble straight lines. Instead, a curved characteristic appears, as shown in Fig. 7-5A. The distortion caused to an input signal of such an amplifier is shown in Fig. 7-5B. Notice the *flat topping* of the output signal that occurs due to the second-harmonic content generated by the amplifier. This type of distortion is called harmonic distortion and is expressed by the equation:

$$V_{out} = AV_{in} + BV_{in}^2 + CV_{in}^3 + \cdots\cdots \qquad \text{(Eq. 7-1)}$$

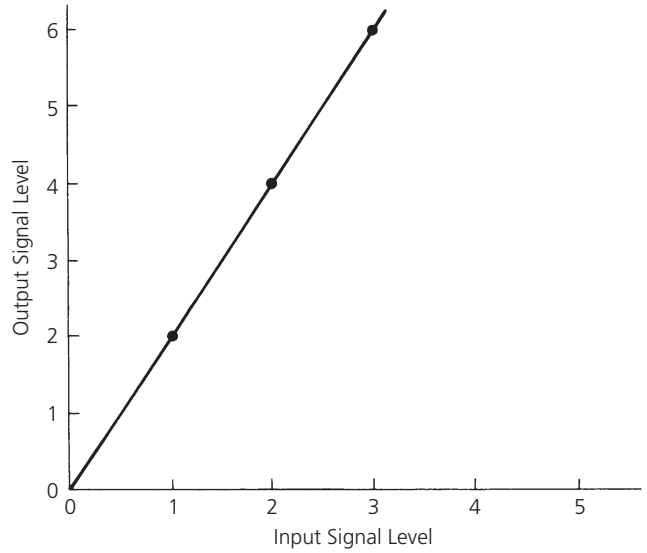


FIG. 7-4. Transfer characteristic for a linear amplifier.

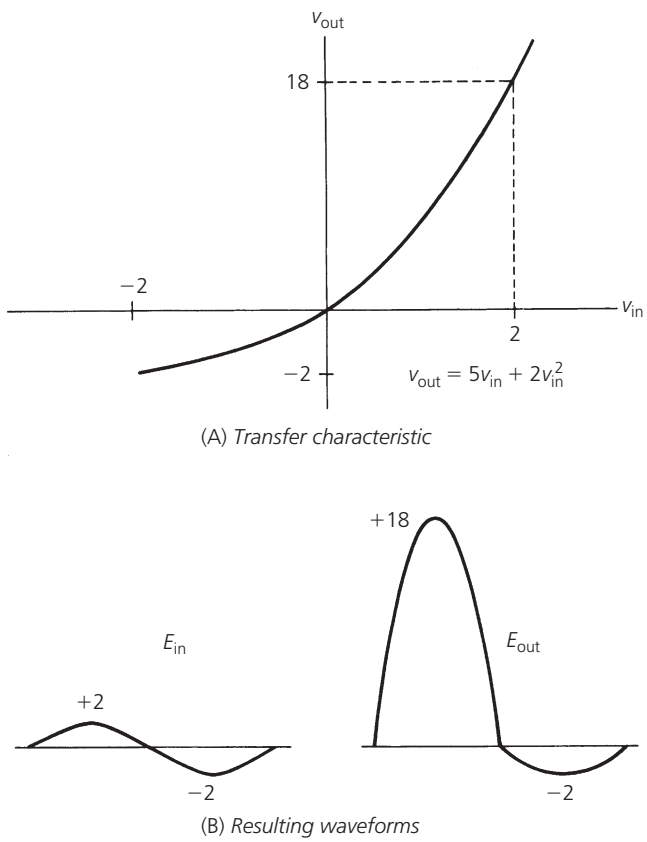


FIG. 7-5. Nonlinear amplifier characteristics.

The second term of Equation 7-1 is known as the *second harmonic* or *second-order distortion*. The third term is called the *third harmonic* or *third-order distortion*. Of course, a perfectly linear amplifier will produce no second, third, or higher order products to distort the signal.

Notice in Fig. 7-5, where the amplifier's transfer function is given as $V_{out} = 5V_{in} + 2V_{in}^2$, that the second-order distortion component increases as the square of the input signal. Thus, with increasing input-signal levels, the second-order component will increase much faster than the fundamental component in the output signal. Eventually, the second-order content in the output signal will equal the amplitude of the fundamental. This effect is shown graphically in Fig. 7-6. The point at which the second-order and first-order content of the output signal are equal is called the *second-order intercept* point. A similar graph may be drawn for an amplifier which exhibits a third-order distortion characteristic. In this case, the third-order term is plotted along with the fundamental gain term of the amplifier. In this manner, the third-order intercept may be determined. The second- and third-order intercept of an amplifier are often used as figures of merit. The higher the intercept point, the better the amplifier is at amplifying large signals.

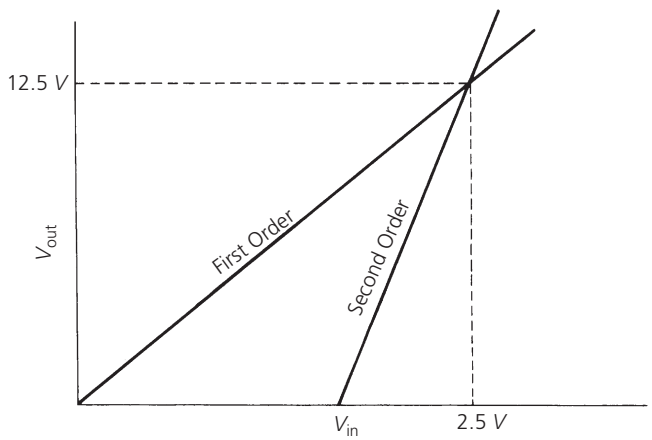


FIG. 7-6. Second-order intercept point.

When two or more signals are input to an amplifier simultaneously, the second-, third-, and higher-order *intermodulation* components are caused by the sum and difference products of each of the fundamental input signals and their associated harmonics. For example, when two perfect sinusoidal signals, at frequencies f_1 and f_2, are input to any nonlinear amplifier, the following output components will result:

fundamental: f_1, f_2
second order: $2f_1, 2f_2, f_1 + f_2, f_1 - f_2$
third order: $3f_1, 3f_2, 2f_1 \pm f_2, 2f_2 \pm f_1 +$ higher order terms

Under normal circuit operation, the second-, third-, and higher-order terms are usually at a much smaller signal level than the fundamental component and, in the time domain, this is seen as distortion. Note that, if f_1 and f_2 are very close in frequency, the $2f_1 - f_2$ and $2f_2 - f_1$ terms fall very close to the two fundamental terms. Third-order distortion products are, therefore, much more difficult to eliminate through filtering once they are generated within an amplifier.

The bias requirements for a class-A power amplifier are the same as those for the small-signal amplifiers presented in Chapter 6. In fact, the distinction between a class-A power amplifier and its small-signal counterpart is a hazy one at best. For all practical purposes, they are equivalent except for input and output signal levels.

Class-B Power Amplifiers

A class-B amplifier is one in which the conduction angle for the transistor is approximately 180°. Thus, the transistor conducts only half the time—either on the positive or negative half cycle of the input signal. Again, it is the DC bias applied to the transistor that determines the class-B operation.

Class-B amplifiers are more efficient than class-A amplifiers (70% vs. less than 50%). However, they are much less linear. Therefore, a typical class-B amplifier will produce quite a bit of harmonic distortion that must be filtered from the amplified signal.

Probably the most common configuration of a class-B amplifier is the push-pull arrangement shown in Fig. 7-7. In this configuration, transistor Q_1 conducts during the positive half cycles of the input signal while transistor Q_2 conducts during the negative half cycles. In this manner, the entire input signal is reproduced at the secondary of transformer T_2. Thus, neither device by itself produces an amplified replica of the input signal. Instead, the signal is actually split in half. Each half is then amplified and reassembled at the output.

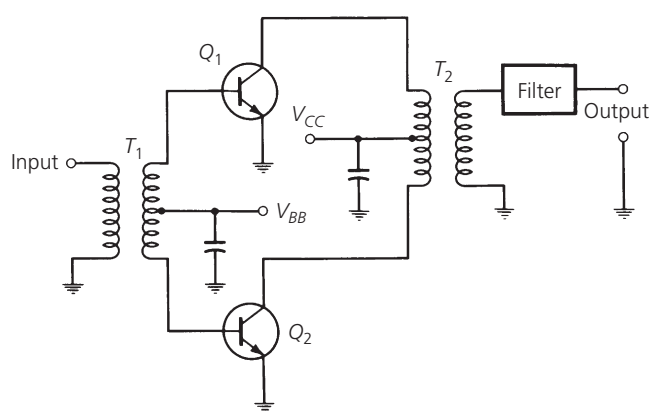


FIG. 7-7. Push-pull class-B amplifier.

Of course, a single transistor may be used in a class-B configuration. The only requirement is that a resonant circuit must be placed in the output network of the transistor in order to reproduce the "other" half of the input signal.

There are several methods of biasing a transistor for class-B operation. One of the most widely used methods is shown in Fig. 7-8. This method simply establishes a base voltage of 0.7 V on the transistor, using an external silicon diode. Often, this diode is mounted on the transistor itself to help prevent *thermal runaway*,

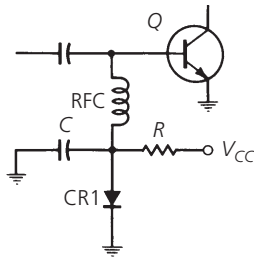


FIG. 7-8. Simple diode bias for class-B operation.

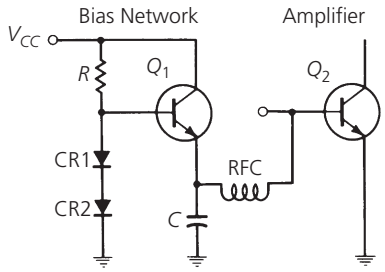


FIG. 7-9. Emitter-follower bias for class-B operation.

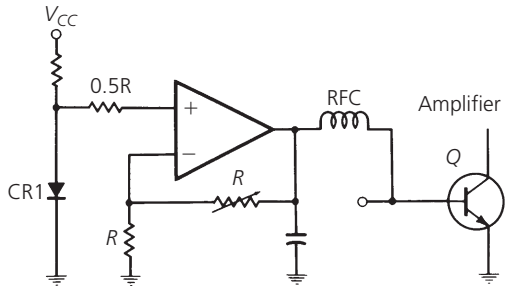


FIG. 7-10. Operational amplifier bias for class-B operation.

which is often a problem with incorrectly biased power amplifiers. Diode CR1 is usually of the heavy-duty variety because the value of resistor R is usually chosen so that the current through CR1 is rather high. This ensures that the bias to the transistor is stable. An alternative bias network is shown in Fig. 7-9. Here, two silicon diodes are used to forward bias an emitter-follower, which is used as a current amplifier. The voltage at the emitter of Q_1 and, hence, at the base of Q_2, is still 0.7 V due to the V_{BE} drop across transistor Q_1. The RF choke and capacitor shown in both Figs. 7-8 and 7-9 are there only to prevent the flow of RF into the bias network.

Still another bias arrangement for class-B operation is shown in Fig. 7-10. Here the bias voltage is made variable so that an optimum solution may be found for best IMD performance. Care must be taken in all three bias arrangements to ensure that the RFC is a low-Q choke for optimum operation.

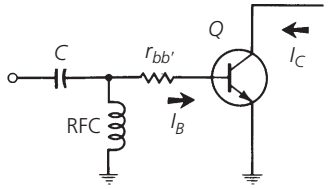


FIG. 7-11. Class-C self bias.

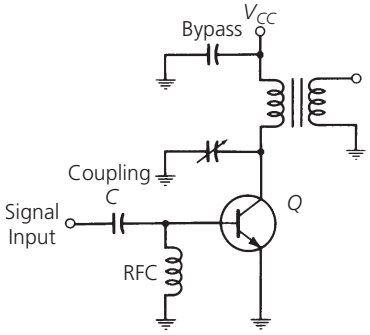


FIG. 7-12. Circuit for class-C self bias.

Class-C Power Amplifiers

A class-C amplifier is one in which the conduction angle for the transistor is significantly less than 180°. The transistor is biased such that under steady-state conditions no collector current flows. The transistor idles at cutoff. Linearity of the class-C amplifier is the poorest of the classes of amplifiers. Its efficiency can approach 85%, however, which is much better than either the class-B or the class-A amplifier.

In order to bias a transistor for class-C operation, it is necessary to *reverse bias* the base-emitter junction. External biasing is usually not needed, however, because it is possible to force the transistor to provide its own bias. This is shown in Fig. 7-11. If the base of the transistor is returned to ground through an RF choke (RFC), the base current flowing through the internal base-spreading resistance ($r_{bb'}$) tends to reverse bias the base-emitter junction. This is exactly the effect you would like to achieve. Of course, it is possible to provide an external DC voltage to reverse bias the junction, but why bother with the extra time and expense if the transistor will provide everything you need? Fig. 7-12 shows a typical class-C amplifier bias arrangement.

RF SEMICONDUCTOR DEVICES

The most popular semiconductor materials used in the manufacture of RF components—especially power amplifiers—are silicon (Si) and gallium arsenide (GaAs). Silicon devices are typically much cheaper to manufacture than gallium arsenide compounds. Unfortunately, silicon-based RF devices usually don't work as well as GaAs for most high-frequency or for high-power applications. Two exceptions include silicon-based lateral double-diffused MOSFET (LDMOS), a version of power

MOSFETs, which are used as high-powered amplifiers (100W and up) in wireless base station applications. Silicon germanium is another silicon-based device that can exceed the performance of GaAs devices, though only in low-power, high-frequency applications, such as in the front-end design of mobile phones (see Figure 7-13). SiGe power amplifiers provide better linear performance and better power efficiency over GaAs. Better efficiency, combined with the lower cost of silicon manufacturing, has made SiGe more popular in recent years.

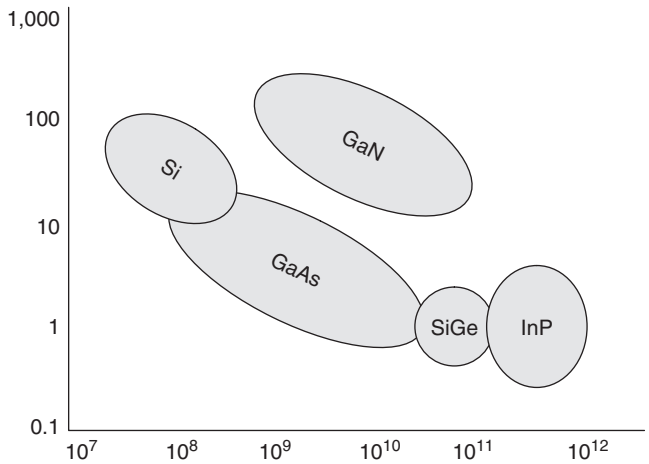


FIG. 7-13. RF Power (W) vs. Frequency (Hz) for RF Power Semiconductors.

In addition to silicon germanium (SiGe), several relatively newer semiconductor compounds are gaining acceptance by RF engineers (see Table 7-1). The first is indium phosphide (InP), which provides exceptional low noise performance at very high frequencies, especially in the millimeter wave range (>40 GHz). Also, InP power amplifiers work well at higher frequencies, though are more expensive to make than SiGe.

Semiconductor	Applications
Silicon (Si)	VLSI, Power amplifiers
Gallium Arsenide (GaAs)	RF, Microwave, MM-Wave
Silicon Germanium	Mixed signal, DSP, RF, microwave, MM-Wave
GaN	RF and microwave power amplifiers

TABLE 7-1. Newer Semiconductor Compounds.

Gallium Nitride (GaN) compounds hold great promise for high frequency, high power amplifiers (100W and up) for wireless transmitters. When combined with RF receivers in mobile phones, GaN amplifiers could enable the direct assessing of communication satellites. The advantages of GaN devices is its high power density, which is many times that of GaAs or InP. The main disadvantage of GaN—as with all gallium-based compounds—is one of high manufacturing expense.

Monolithic Microwave Integrated Circuits (MMIC)

An integrated circuit (IC) results when you add more than one device to a semiconductor substrate—for example, transistors, diodes and other electronic components. If the device operates at microwave frequencies (1 GHz to 300 GHz) and performs such functions as microwave mixing, power and low noise amplification, and switching, then the device is called a monolithic microwave IC, or MMIC (pronounced *mimic*). MMICs are most often made from GaAs, InP or SiGe (see Figure 7-14).

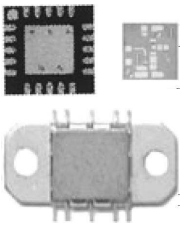


FIG. 7-14. GaAs MMICs are used in defense, space, and selected commercial markets. (Courtesy of M/A COM)

Like other mass-produced IC devices, MMICs enjoy the benefit of low cost in high volume and small chip size (from around 1 mm^2 to 10 mm^2). The main disadvantage of a MMIC is that they can have worse performance on certain parameters than the same devices made out of separate components. For example, if low noise is a critical performance requirement in a microwave low noise amplifier, then it may be best to use a discrete component amplifier or build one out of transistors, rather than use a multifunction MMIC. This follows from the fact that, since MMICs are integrated into a single semiconductor device, the separate parts of a MMIC cannot be easily tuned as with discrete components distributed on a PCB. Once the MMIC circuit has been designed, its performance characteristics are set. The design of integrated circuits using automated software tools is covered in Chapter 9.

Filters and MMICs

Technologies currently used to fabricate front-end RF and IF filters are diverse, although the general trend for all design approaches is to produce the smallest possible filter with the highest performance and power-handling capability. RF/microwave filters have been constructed with many materials and structures, including slabline, combline, and waveguide filters, in addition to filters based on different types of resonators, such as ceramic resonators, crystal resonators, dielectric resonators, film-bulk-acoustic resonators (FBARs), surface-acoustic-wave (SAW) resonators, and even the exotic yttrium-iron-garnet (YIG) resonators.

Monolithic IC filters fabricated on a chip with other semiconductor devices borrow from the traditional use of passive inductors (Ls) and capacitors (Cs) to form the resonant circuits at the basis of an RF/microwave filter. Because the values of on-chip Ls

and Cs are limited, and the fact that the size of these passive elements is determined by the wavelength or center frequency of the filter, the performance of IC-based filters is extremely limited, typically prompting designers to specify FBAR, SAW, or other types of filters in RF front ends. Similarly, active monolithic filters based on operational amplifiers (op-amps) that combine resistor (R) and capacitor (C) elements can achieve high rejection for typically IF signals, although they are limited in frequency range for other RF front-end applications.

POWER AMPLIFIER DESIGN

At the beginning of this chapter, you learned that the important design information for RF power transistors is presented in the form of large-signal impedance parameters. The formulas presented in Chapter 6 for small-signal transistor design, using Y and S parameters, are no longer valid. Instead, the designer must model with the help of the data sheet, deciding what the input and output impedance of the transistor looks like at the frequency of interest. With this information in hand, the designer needs only to match the input and output impedance of the device to the source and load, respectively. These two steps require only that the designer read the input and output impedances off of the data sheet, and then apply the principles of Chapter 4 to complete the matching network. Care must be taken to ensure that the information extracted from the data sheet is of the proper format–series or shunt information.

Often, instead of supplying complete output information for a transistor in the form of a series or shunt resistance and capacitance output, manufacturers will supply output capacitance information only. This is because the optimum load resistance for the transistor is easily calculated using a very simple formula, as we shall soon see.

Optimum Collector Load Resistance

In the absence of collector output *resistance* information on the data sheet, it becomes necessary for the designer to make a very simple calculation to determine the optimum load resistance for the transistor (Example 7-1). This value of load resistance is dependent upon the output power level required and is given by:

$$R_L = \frac{(V_{CC} - V_{\text{sat}})^2}{2P} \qquad \text{(Eq. 7-2)}$$

where
V_{CC} = the supply voltage,
V_{sat} = the saturation voltage of the transistor,
P = the output power level required.

Note that Equation 7-2 provides you with a value of load *resistance* only. It does not indicate anything about the reactive portion of the load. On the data sheet, however, the manufacturer typically provides values of shunt output capacitance versus frequency for the transistor. The designer's job is to provide a load for the transistor which absorbs this stray or parasitic capacitance so that the transistor may be matched to its load (see Chapter 4). Example 7-2 may illustrate this point.

EXAMPLE 7-1

What value of load resistance is required to obtain 2.0 W of RF output from a transistor if the supply voltage is 12 V and the saturation voltage of the transistor is 2 V?

Solution

Using Equation 7-2, we have

$$R_L = \frac{(12-2)^2}{2(2)} = 25 \text{ ohms}$$

Keep in mind that if the output *resistance* information had not been provided in the data sheet, it would have been a simple matter to calculate the required R_L using Equation 7-2. Once this calculation is made, the output matching network is designed in the same manner as was done in Example 7-2. The 50-ohm load is simply transformed into the impedances that the transistor would like to see for the specified power output.

Driver Amplifiers and Interstage Matching

Often it is required that power gain be distributed throughout several amplifier stages in order to produce a specified output power into a load. This is especially true in transmitter applications that require a substantial amount of power into an antenna.

The typical procedure for such a design involves finding, first, an output transistor that will handle the required *output* power and, then, designing driver amplifiers that will provide the necessary *drive* power to the final transistor. This type of gain distribution is shown in Fig. 7-19. Note that the required output-power level from the final amplifier is 15 W. A final transistor was chosen which will handle the required output power and which will provide a gain of 10 dB. The required drive level to the stage is, therefore, 1.5 W, and is supplied by a transistor with a gain of 15 dB. The signal source must, therefore, supply the driver with a signal level of 47 mW, which is within the capabilities of most oscillators.

Let's examine the interstage match between Stage A and Stage B in a little more detail. Often, in dealing with power amplifiers, it is unclear whether or not a true *impedance match* occurs between the power amplifier and its load. A true impedance match for an amplifier would involve providing a load for the transistor that is the complex conjugate of its output impedance. In designing power amplifiers, however, we speak of providing a load resistance (Equation 7-2) for the transistor in order to extract a specified power gain from the stage. This is simply a matter of semantics and, from a circuit-design viewpoint, it doesn't really matter how you look at it. Fig. 7-20 illustrates this point. Suppose the transistor of Stage B has an input impedance as shown ($Z_{\text{in}} = 1.7 - j2.7$ ohms). Also, suppose that Stage A, in order to

EXAMPLE 7-2

Using the data sheet of Fig. 7-1, design a class-C power amplifier that will deliver 15 W between a 50-ohm source and load at 100 MHz.

Solution

The data sheet for the MRF233 transistor provides input and output impedance information for the transistor in both series and parallel form. The designer is, therefore, left with a choice as to which he prefers. The input-matching design may proceed as in the following method.

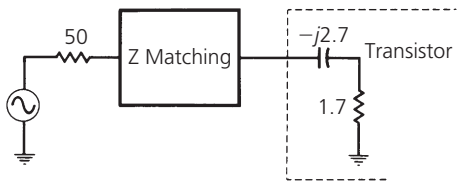


FIG. 7-15. Transistor input impedance for Example 7-2.

The input impedance of the transistor appears as shown in the diagram of Fig. 7-15. This information was taken from Figure 5 of the data sheet. Note that the object of the input-matching network is to transform the input impedance of the transistor up to 50 ohms to provide an optimum load for the source.

Using the techniques of Chapter 4, first resonate the series capacitance with an equal and opposite series inductance of $+j2.7$ ohms. Then, match the remaining 1.7-ohm resistive load to the source as follows. Using Equation 4-1 through 4-3, for the L-network:

$$Q_s = Q_p = \sqrt{\frac{R_p}{R_s} - 1} = \sqrt{\frac{50}{1.7} - 1} = 5.33 \qquad \text{(Eq. 4-1)}$$

$$X_s = R_s Q_s = (1.7)(5.33) = 9.06 \text{ ohms} \qquad \text{(Eq. 4-2)}$$

$$X_p = \frac{R_p}{Q_p} = \frac{50}{5.33} = 9.38 \text{ ohms} \qquad \text{(Eq. 4-3)}$$

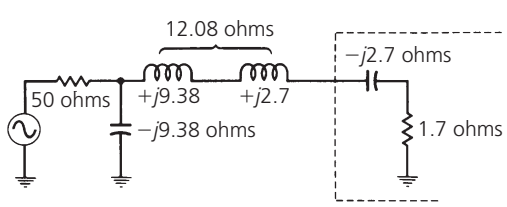


FIG. 7-16. Input matching network for Example 7-2.

This matching network is shown in Fig. 7-16. Note that it is convenient to use a shunt-C series-L matching network in order to easily absorb the $+j2.7$-ohm inductor that was needed earlier. Thus, the two inductors may be combined into a single component in the actual design.

Since complete output-impedance information is provided for the transistor (Figure 5 of the data sheet), it is only necessary to match this output impedance to the 50-ohm load. Proceeding as before, first resonate the $-j5.6$-ohm series capacitance with an inductor of equal value. Then, match the remaining 5-ohm resistive portion of the transistor output to the load as follows:

$$Q_s = Q_p = \sqrt{\frac{R_p}{R_s} - 1} = \sqrt{\frac{50}{5} - 1} = 3$$

$$X_s = Q_s R_s = (3)(5) = 15 \text{ ohms}$$

$$X_p = \frac{R_p}{Q_p} = \frac{50}{3} = 16.7 \text{ ohms}$$

The output network is shown in Fig. 7-17. Note that it is again convenient to use the series-L shunt-C arrangement so that we may absorb the $+j5.6$-ohm inductor used previously.

A practical circuit for this design might appear as shown in Fig. 7-18.

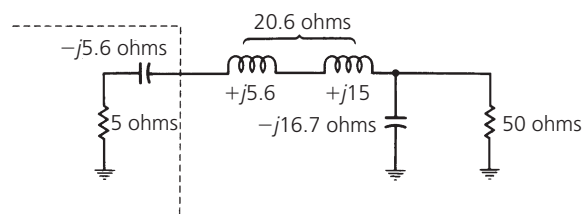


FIG. 7-17. Output matching network for Example 7-2.

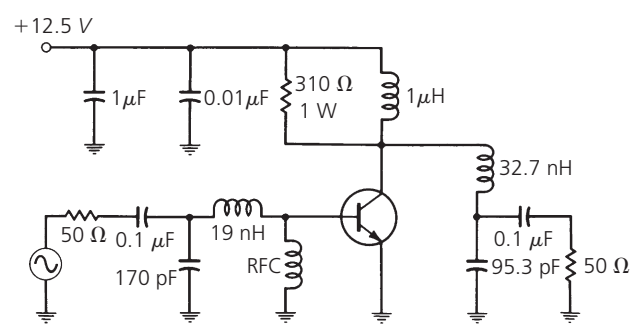


FIG. 7-18. Final circuit realization for Example 7-2.

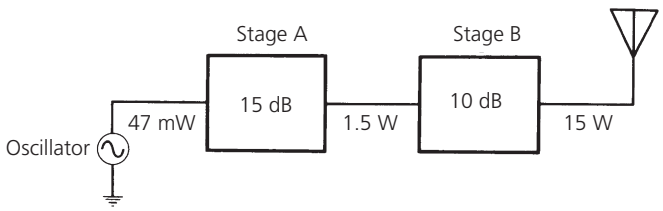


FIG. 7-19. System drive requirements for a 15-watt transmitter.

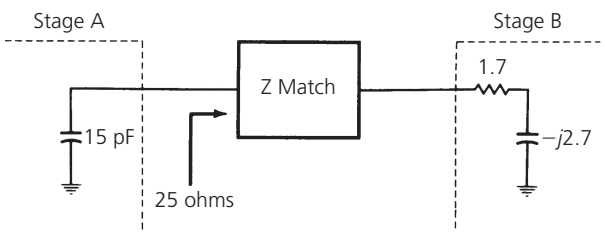


FIG. 7-20. Requirement for an interstage impedance-matching network.

supply the required 1.5 W of RF drive, requires a load resistance of 25 ohms. The role of the *impedance-matching* network, then, is to *transform* the low-input impedance of Stage B up to the 25-ohm level required by Stage A. In addition, the matching network must absorb or resonate out the 15-pF output capacitance of Stage A.

MATCHING TO COAXIAL FEEDLINES

The T and Pi networks studied in Chapter 4 are excellent candidates for use in matching coaxial feedlines to power amplifiers. Often such a network will serve a dual purpose, especially when configured as a low-pass filter, in providing harmonic suppression for a transmitter.

Fig. 7-21 is a diagram of a coaxial feed to an antenna at the antenna's resonant frequency. Resistance R_a is the antenna's radiation resistance. A quarter-wavelength vertical antenna operating against a very good ground plane has a *radiation resistance* of about 35 ohms while a half-wave center-fed dipole has a radiation resistance of about 70 ohms, at its resonant frequency. This is simply the resistance that the coaxial cable sees at the antenna terminals. Above and below the resonant frequency of the antenna, its radiation resistance begins to show a reactive component. This is illustrated in Fig. 7-22. Above resonance (Fig. 7-22A), the antenna looks inductive, and below resonance (Fig. 7-22B), the antenna looks capacitive.

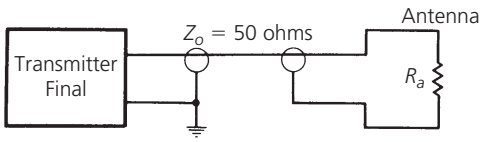


FIG. 7-21. Antenna radiation resistance at resonance.

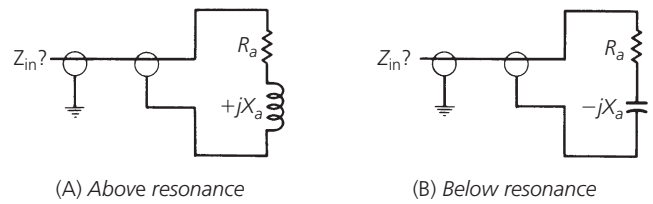


FIG. 7-22. Radiation resistance of an antenna.

At the transmitter end of the coaxial feedline, the impedance that the output transistor actually sees is not only a function of the antenna's radiation resistance, but also a function of the length of the coaxial feedline. The impedance along the line varies sinusoidally as you move away from the antenna. Thus, at a distance of one half-wavelength back from the antenna, the impedance looking into the coax would appear to be equal to the antenna's radiation resistance. At other distances removed from the antenna, however, the coax would appear to have a much different input impedance, depending upon the degree of mismatch between the antenna and the feedline. Therefore, it is extremely difficult to estimate the actual input impedance of any transmission line *unless the line is terminated in its characteristic impedance*. That is to say, a 50-ohm coaxial cable will not look like 50 ohms at its input unless there is a 50-ohm load at the other end of the cable. Since this is hardly ever the case when driving practical antenna systems, it is not very practical to design a matching network unless the network is tunable. In addition, many antenna installations operate over quite a range of frequencies. Since the radiation resistance of the antenna varies with frequency, the input impedance of the coaxial cable must also vary, and the matching network must be able to compensate for these variations.

Fig. 7-23 indicates two possible methods of providing a tunable impedance-matching network for a transmission line. The T network of Fig. 7-23A uses both tapped inductors and a tunable capacitor. The Pi network of Fig. 7-23B uses only tunable capacitors. Note that, in both cases, the low-pass configuration is used to aid in suppressing harmonics of the transmitted signal.

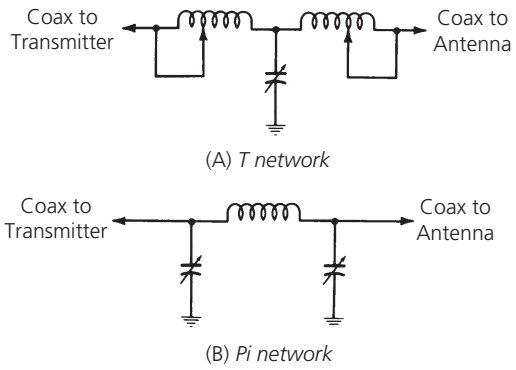


FIG. 7-23. Variable coaxial feedline matching networks.

The circuits of Fig. 7-23 are designed in the same manner as those shown in Chapter 4. Of course, if you had a requirement that the harmonics of the transmitted signal were required to be at a certain level below the fundamental, say 50 dB, then the filter-design approach used in Chapter 3 might be the best approach to take.

AUTOMATIC SHUTDOWN CIRCUITRY

Since power amplifiers are designed to supply a considerable amount of power to an antenna system, an impedance mismatch presented to the amplifier could cause very severe problems. As we learned in Chapter 4, an impedance mismatch between a source and its load causes reflection of some of the signal incident upon that load. This reflected signal will eventually make its way back to the source and, in high-power transmitters, can cause serious side-effects, such as transistor damage in the form of secondary breakdown. For this reason, many manufacturers of power amplifiers include a VSWR monitoring circuit in the transmitter, which monitors the *standing wave ratio* of the output circuit and, in the event that the VSWR becomes excessive, indicating a severe mismatch, the circuit automatically decreases the RF drive to the final amplifier, thereby reducing the transmitter's output power. The reduction in output power subsequently reduces the reflected power from the load, thus, protecting the output transistor. A simplified diagram of such a system appears in Fig. 7-24.

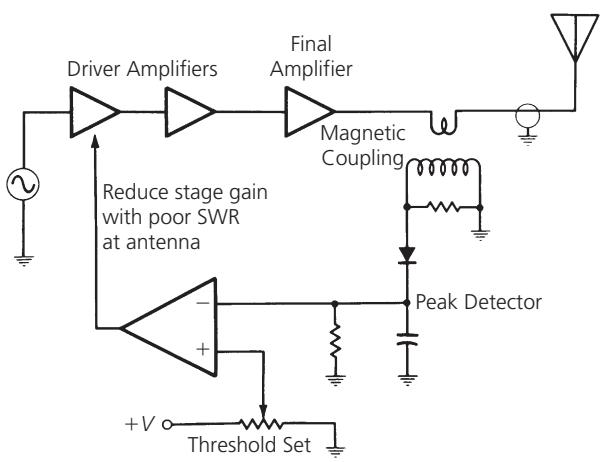


FIG. 7-24. Automatic shutdown circuitry.

BROADBAND TRANSFORMERS

Several types of broadband transformers, which are often used in power-amplifier design, are illustrated in Fig. 7-25. Fig. 7-25A is known as a 1:1 *balun*. It is used mainly to connect a *balanced* source to an *un*-balanced load, or vice versa, and it provides no impedance-transforming function.

Fig. 7-25B is a 4:1 transformer. That is to say, it will transform an impedance of $4R$ down to an impedance of R, or vice versa. The small dot located next to each winding indicates polarity.

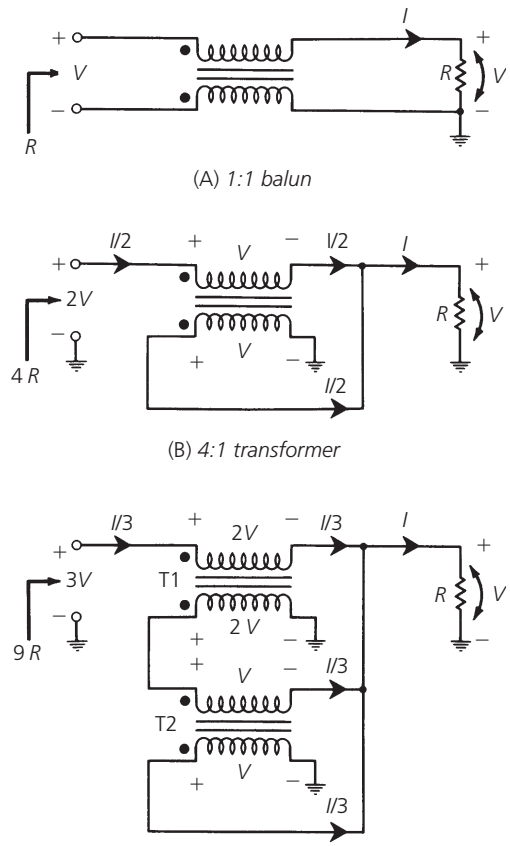


FIG. 7-25. Types of broadband transformers.

A detailed look at Fig. 7-25B will explain how the 4:1 transformation occurs. First, suppose that a voltage (V) has been impressed across the load resistor and the voltage across the same voltage must also be impressed across the lower winding of the transformer since the two are in parallel. The voltage on the lower winding impresses the same voltage (V) on the *upper* winding with the polarity indicated. This is true because each winding has the same number of turns. The voltage at the input terminals is, therefore, equal to the sum of the voltage across the load resistor and the voltage across the upper winding, or $2V$. Suppose now that a current of $I/2$ is injected at the input terminals of the transformer. This current flowing in the top winding of the transformer induces a current of $I/2$ in the bottom winding in the direction shown. The current in the load resistor is, therefore, equal to $I/2 + I/2$ or I. Therefore, if the load resistor is equal to 1 ohm, the resistance seen looking into the input terminals of the transformer, then, must be:

$$\text{Resistance} = \frac{\text{Voltage}}{\text{Current}} = \frac{2V}{I/2} = 4R$$

A 9:1 transformer is illustrated in Fig. 7-25C. Note that this transformer is actually made up of two separate transformers, T1 and T2. A similar analysis may be made of this transformer to confirm the 9:1 transformation. Voltages and currents are included in the diagrams to aid in the analysis process. Obviously, several transformers may be included in such a configuration to produce other transformation ratios. Fig. 7-26, for example, includes three separate transformers that are used to produce a 16:1 impedance transformation.

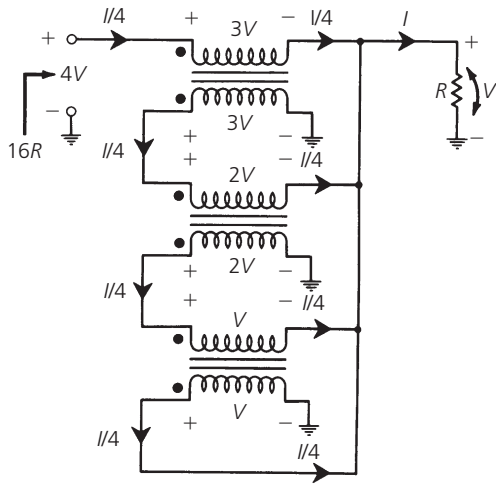


FIG. 7-26. A 16:1 broadband transformer.

Power Splitters

A basic power splitter is shown in Fig. 7-27. Ideally, the power into the primary of the transformer is split evenly between amplifier No. 1 and amplifier No. 2. However, due to input-impedance variations between the two amplifiers, this is rarely the case. Instead, one amplifier is usually provided with a bit more drive power than the other. To aid in equalizing the power split, resistor R at the center tap of the secondary is often left out of the circuit. (Once again, the small dots are used to indicate polarity.)

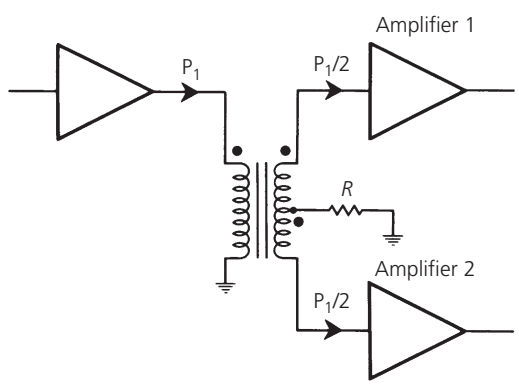


FIG. 7-27. A power splitter.

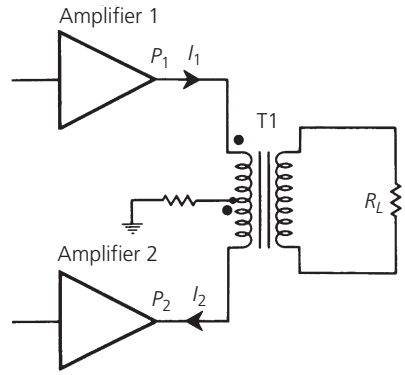


FIG. 7-28. A power combiner.

Power Combiners

A typical power combiner is shown in Fig. 7-28. Here, the power output of each amplifier is combined in transformer T1 to provide an output power of $P_1 + P_2$. Power combiners are often used in power-amplifier designs where it is impractical for a single stage to produce the necessary output power. In this case, two amplifiers, operating 180° out-of-phase, will each provide half the needed output power to the power combiner. The combined output is, therefore, equal to the power level required.

PRACTICAL WINDING HINTS

Broadband transformers are often called transmission-line transformers because they make use of the transmission-line properties of the windings. This is done by using *bifilar*- or *trifilar*-type windings rather than the conventional type of winding.

A conventional transformer usually has two entirely separate windings. That is, one of the windings is usually wound onto the core first, and then the other winding is wound on top of the first winding. Typically, the larger winding is wound first for convenience. This winding technique is shown in the toroidal transformer diagram of Fig. 7-29. Note that an impedance transformation occurs between the primary and secondary of the transformer. The value of the transformation is dependent upon the turns ratio from the primary to the secondary. Transmission-line transformers use an entirely different technique for the windings, as shown in Fig. 7-30. First, the primary and secondary windings are made by twisting the wires together for a

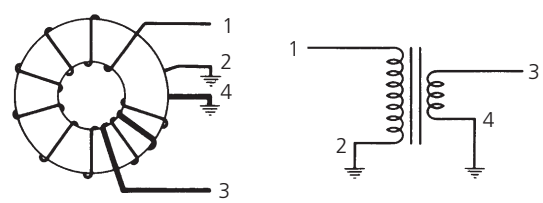


FIG. 7-29. A conventional transformer.

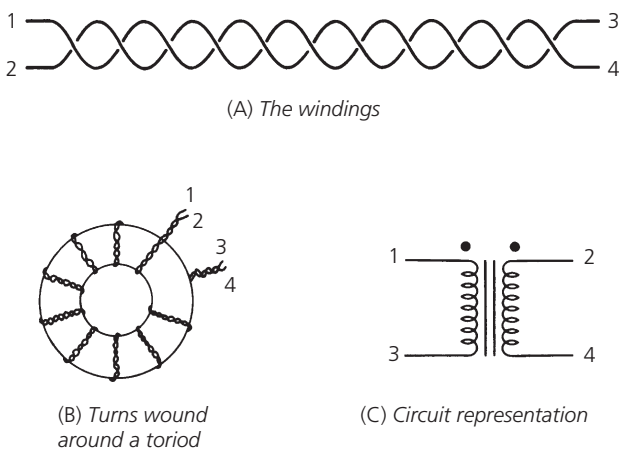


FIG. 7-30. A bifilar-wound broadband transformer.

certain number of turns per inch (Fig. 7-30A). This produces a certain characteristic impedance for the resulting "transmission line" in much the same manner that a coaxial cable exhibits a certain characteristic impedance which is dependent upon the spacing of its center conductor to its outer conductor. The actual characteristic impedance of the twisted pair is dependent upon the number of turns per inch, the shape of the windings, and the size wire used. For low-impedance lines, tight twists (many turns per inch) are used while high-impedance lines may not be twisted at all. Instead, the windings will be placed side-by-side around the core. For optimum operation, the characteristic impedance of the winding should be equal to:

$$Z_o = \sqrt{R_s R_L} \qquad \text{(Eq. 7-3)}$$

where

R_s = the primary impedance,
R_L = the secondary impedance.

Typically, Z_o must be found experimentally.

The transformer of Fig. 7-30 is called a *biflar-wound* transformer because it uses two conductors in the twisted-winding arrangement. A *trifilar* transformer is one that uses three conductors, and so on.

As shown, Fig. 7-30 is simply a 1:1 broadband transformer. If connected as shown in Fig. 7-25B, however, this arrangement could be used to produce a 4:1 broadband transformer. This may be done by simply connecting lines 2 and 3 together and using that junction as your output port. Line 4 is connected to ground and line 1 is then the input port.

Often, instead of using a twisted pair, the designer may instead use coaxial cable for the windings. The center conductor and outer conductor of the coax, then, take the place of each conductor in the twisted pair. This is done primarily because each type of coax has a very well-defined and consistent characteristic impedance and this eliminates the experimentation involved in defining the characteristic impedance of the twisted pair. Of course, the use of standard coaxial cable does not allow for trifilar-wound transformers. Typically, broadband transformers are wound on low-Q, high-permeability, ferrite toroidal cores (see Chapter 1). The high permeability is needed at the low end of the frequency spectrum where, for a given inductance, fewer turns would be needed.

SUMMARY

The power-amplifier design process is not as well defined as that of the small-signal amplifier. Thus, considerable experimentation may be necessary in order to optimize a design. Standard Y and S parameters are not used in power-amplifier design. Instead, large-signal impedance parameters are typically provided by the transistor manufacturers to aid in the design process. Following the impedance-matching procedures outlined in Chapter 4, the designer must match the source to the transistor's input impedance, and transform the load impedance to a value that is dependent upon the required output-power level from the stage. The source must be capable of providing the required RF drive level, or the calculated RF power output from the stage will never be achieved.

CHAPTER 8

RF FRONT-END Design

Throughout most of this book, we have studied the components that make up RF circuits. Now we put all these pieces together to form one of the most critical subsystems in any communication system—namely, the *RF front end*. We'll start by showing where the RF front end fits in today's modern applications, and then decompose or "tear down" this subsystem into its basic components and functions, many of which have already been covered in earlier chapters. This approach will also provide a logical way to introduce key receiver types and associated performance specifications, such as signal-to-noise ratio (SNR), receiver sensitivity, and selectivity.

The RF front end is part of an overall radio receiver-transmitter or transceiver system. It is generally defined as everything between the antenna and the digital baseband system. For a receiver, this "between" area includes all the filters, low-noise amplifiers (LNAs), and down-conversion mixer(s) needed to process the modulated signals received at the antenna into signals suitable for input into the baseband analog-to-digital converter (ADC). For this reason, the RF front end is often called the *analog-to-digital* or *RF-to-baseband* portion of a receiver.

Radios work by receiving RF waves containing previously modulated information sent by a RF transmitter. The receiver is basically a low noise amplifier that down converts the incoming signal. Hence, sensitivity and selectivity are the primary concerns in receiver design.

Conversely, a transmitter is an up converts an outgoing signal prior to passage through a high power amplifier. In this case, non-linearity of the amplifier is a primary concern. Yet, even with these differences, the design of the receiver front end and transmitter back end share many common elements—like local oscillators. In this chapter, we'll concentrate our efforts on understanding the receiver side.

Thanks to advances in the design and manufacture of integrated circuits (ICs), some of the traditional analog IF signal-processing tasks can be handled digitally. These traditional analog tasks, like filtering and up-down conversion, can now be handled by means of digital filters and digital signal processors (DSPs). Texas Instruments have coined the term digital radio processors for this type of circuit. This migration of analog into digital circuits means that the choice of what front-end functions are implemented by analog and digital means generally depends on such factors as required performance, cost, size, and power consumption. Because of the mix of analog and digital technologies, RF front end chips using mixed-signal technologies may also be referred to as RF-to-digital or RF-to-baseband (RF/D) chips.

Why is the front end so important? It turns out that this is arguably the most critical part of the whole receiver. Trade-offs in overall system performance, power consumption, and size are determined between the receiver front end and the ADCs in the baseband (middle end). In more detail, the analog front end sets the stage for what digital bit-error-rate (BER) performance is *possible* at final bit detection. It is here that the receiver can, within limits, be designed for the best potential SNR.

HIGHER LEVELS OF INTEGRATION

Look inside any modern mobile phone, multimedia device, or home-entertainment control system that relies on the reception and/or transmission of wireless signals and you'll find an RF front end. In the RIM Blackberry PDA, for example, the communication system consists of both a transceiver chip and RF front-end module (see Fig. 8-1). The front-end module incorporates several integrated circuits (ICs) that may be based on widely different semiconductor processes, such as conventional silicon CMOS and advanced silicon germanium (SiGe) technologies. Functionally, such multichip modules provide most if not all of the analog signal processing—filtering, detection, amplification and demodulation via a mixer. (The term "system-in-package" or SIP is a synonym for multichip module or MCM.)

Multichip front-end modules demonstrate an important trend in RF receiver design, namely, ever-increasing levels of system integration required to squeeze more functionality into a single chip. The reasons for this trend—especially in consumer electronics—come from the need for lower costs, lower power consumption (especially in mobile and portable products), and smaller product size. Still, regardless of the level of integration, the basic RF architecture remains unchanged: signal filtering, detection, amplification and demodulation. More specifically, a modulated RF carrier signal couples with an antenna designed for a specific band of frequencies. The antenna passes the modulated signals along to the RF

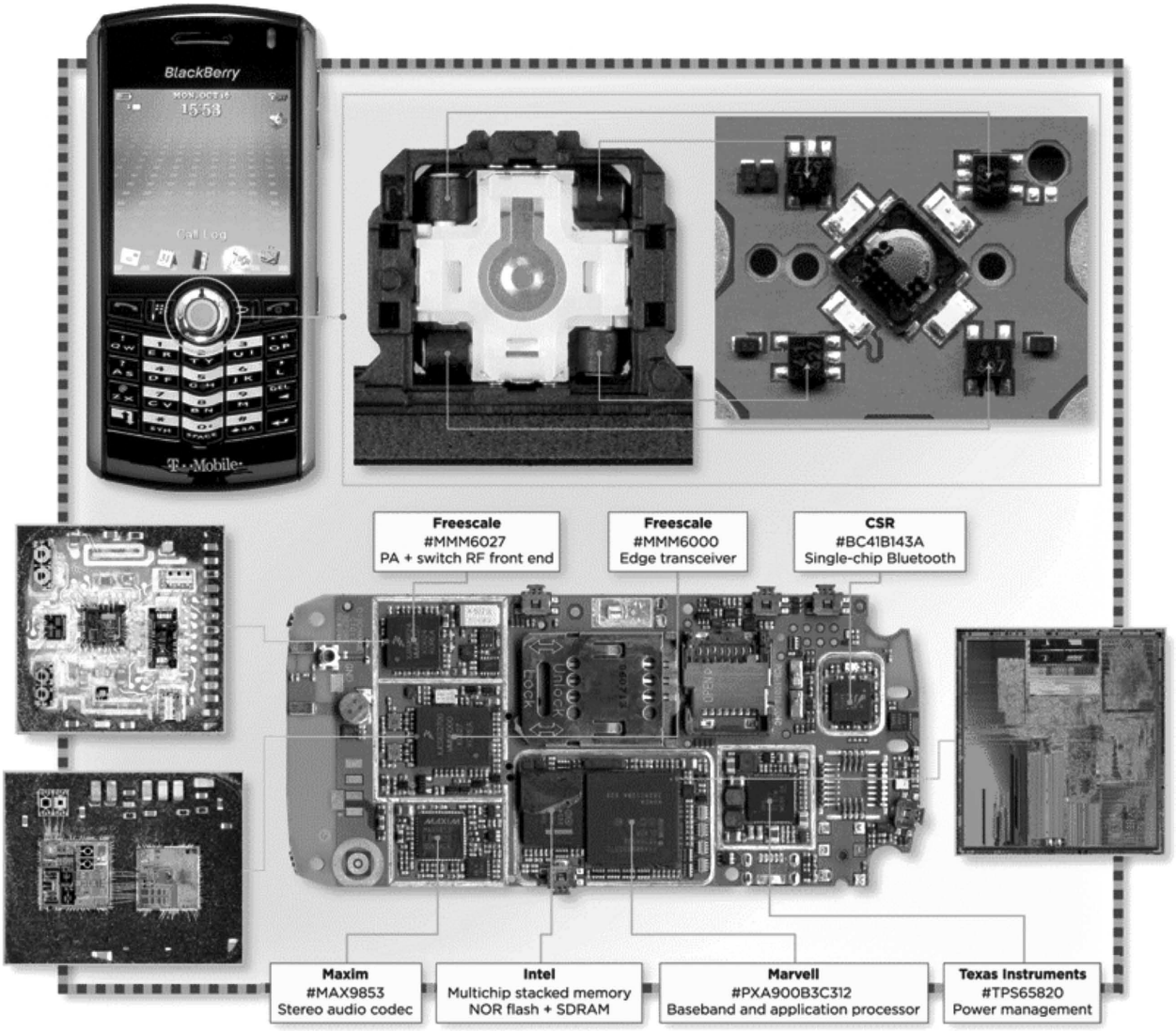


FIG. 8-1. Tear down of modern mobile device reveals several RF front-end chips. (Courtesy of iSuppli)

receiver's front end. After much conditioning in the front-end circuitry, the modulation or information portion of the signal—now in the form of an analog baseband signal—is ready for analog-to-digital conversion into the digital world. Once in the digital realm, the information can be extracted from the digitized carrier waveforms and made available as audio, video, or data.

Before the advent of such tightly integrated modules, each functional block of the RF front end was a separate component, designed separately. This means that there were separate components for the RF filter, detector, mixer-demodulator, and amplifier. More importantly, this meant that all of these physically independent blocks had to be connected together. To prevent signal attenuation and distortion and to minimize signal reflections due to impedance differences between function blocks, components were standardized for a characteristic impedance of 50 ohms, which was also the impedance of high-frequency test equipment. The 50-ohm coaxial cable [1] interface was a trade-off that minimized signal attenuation while maximizing power transfer—signal energy—between the independently designed RF filter, LNA, and mixer. Before higher levels of functional integration and thus lower costs could be achieved, it was necessary to design and manufacture these RF functional blocks using standard semiconductor processes, such as silicon CMOS IC processes.

Unfortunately, one of the drawbacks of CMOS technology can be the difficulty in achieving a 50-ohm input impedance. Still, it is only necessary to have the 50-ohm matched input and output impedances when the connection lines between the sub-circuits is long compared to the wavelength of the carrier wave. For ICs and MCMs at GHz frequencies, connections lines are short, so 50-ohm between sub-circuits isn't a problem. It is necessary to somehow get to 50 ohms to connect to the (longer) printed circuit board traces.

This is but one example of the changes that have taken place with modern integrated front ends. We will not cover all the changes here. Instead, we'll focus on the important design parameters that can affect the design of an RF front end, including the signal-to-noise ratio (SNR), receiver sensitivity, receiver and channel filter selectivity, and even the bit resolution of the ADC (covered later). This high-level description of the RF front end reveals not only the basic functioning but also the potential system trade-offs that must be considered.

As mentioned earlier, the basic stages of an RF front end include an antenna, filter, detector-demodulator, and amplifier. Each of these signal-conditioning stages contains unique circuit components, many of which have already been covered in earlier portions of this book. None of these components work in isolation and the performance of one component may well affect the performance of another. This is why we'll look at each of these key component function blocks in the context of several different radio architectures: detector, direct-conversion, and superheterodyne receiver configurations.

BASIC RECEIVER ARCHITECTURES

The fundamental operation of an RF front end is fairly straightforward: it detects and processes radio waves that have been transmitted with a specific known frequency or range of frequencies and known modulation format. The modulation carries the information of interest, be it voice, audio, data, or video. The receiver must be tuned to resonate with the transmitted frequency or frequencies in order to detect them. Those received signals are then filtered from all surrounding signals and noise and amplified prior to a process known as demodulation, which removes the desired information from the radio waves that carried it.

These three steps—filtering, amplification and demodulation—detail the overall process. But actual implementation of this process (i.e., designing the physical RF receiver printed-circuit board (PCB)) depends upon the type, complexity, and quantity of the data being transmitted. For example, designing an RF front end to handle a simple amplitude-modulated (AM) signal requires far less effort and hardware (and even software) than building an RF front end for the latest third-generation (3G) mobile telecommunications handset.

Because of the enhanced performance of analog components due to IC process improvements and decreasing costs of more powerful digital-signal-processing (DSP) hardware and software functions, the ways that different RF front-end architectures are realized has changed over the years. Still, the basic requirements for an RF front end, such as the frequency range and type of carrier to be received, the RF link budget, and the power, performance, and size restrictions of the front-end design, remain relatively the same in spite of the differences in radio architectures. Let's start by looking at the simplest of radio architectures or implementations.

AM Detector Receivers

One of the basic RF receiver architectures for detecting a modulated signal is the amplitude modulation (AM) detector receiver (see Fig. 8-2). The name comes from the fact that information like speech and music could be converted into amplitude (voltage) modulated signals riding on a carrier wave. Such an RF signal could be de-modulated at the receiving end by means of a simple diode detector. All that is needed for a basic AM receiver—like a simple crystal radio—is an antenna, RF filter, detector, and (optional) amplifier to boost the recovered information to a level suitable for a listening device, such as a speaker or headphones. The antenna, which is capacitive at the low frequencies used for AM broadcasting, is series matched with an inductor to maximize current through both, thus maximizing the voltage across the secondary coil (see Chapter 2). A variable capacitance filter may be used to select the designed frequency band (or channel) and to block any unwanted signals, such as noise. The filtered signal is then converted to demodulate the AM signal and recover the information. Fig. 8-3 represents a schematic version of the block diagram shown in Fig. 8-2.

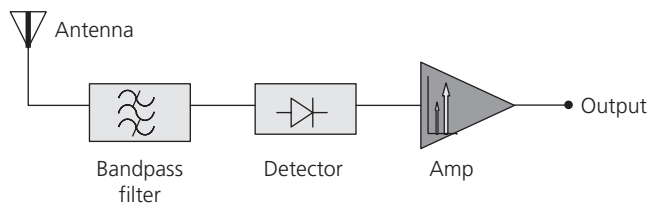


FIG. 8-2. Simple amplitude modulation (AM) radio architecture.

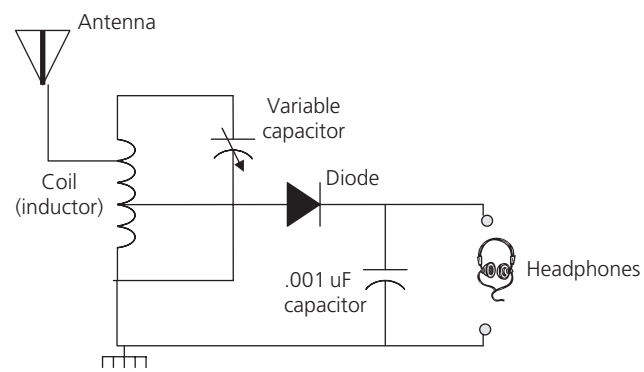


FIG. 8-3. Circuit schematic of a simple AM radio.

The heart of the AM architecture is the detector demodulator. In early crystal radios, the detector was simply a fine metal wire that contacted a crystal of galena (lead sulfide), thus creating a point contact rectifier or "crystal detector." In these early designs, the fine metal contact was often referred to as a "catwhisker." Although point-contact diodes are still in use today in communication receivers and radar, most have been

replaced by pn-junction diodes, which are more reliable and easier to manufacture.

For a simple AM receiver, the detector diode acts as a half-wave rectifier to convert or rectify a received AC signal to a DC signal by blocking the negative or positive portion of the waveform (see Fig. 8-4). A half-wave rectifier clips the input signal by allowing either the positive or negative half of the AC wave to pass easily through the rectifier, depending upon the polarity of the rectifier. A shunt inductor is typically placed in front of the detector to serve as an RF choke (see Chapter 2). The inductor maintains the input to the detector diode at DC ground while preserving a high impedance in parallel with the diode, thus maintaining the RF performance.

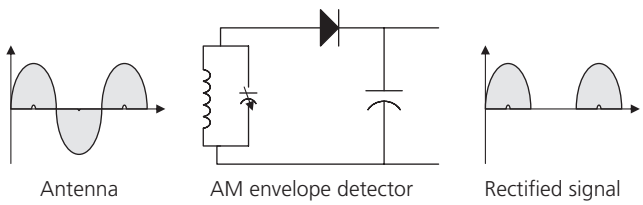


FIG. 8-4. Circuit schematic of a half-wave rectifier.

In a simple detector receiver, the AM carrier wave excites a resonance in the inductor/tuned capacitor (LC) tank subcircuit. The tank acts like a local oscillator (LO). to the current through the diode is proportional to the amplitude of the resonance and this gives thebaseband signal (typically analog audio). The baseband signal may be in either analog or digital format, depending upon the original format of the information used to modulate the AM carrier. As we shall see, this process of translating a signal down or up to the baseband level becomes a critical technique in most modern radios. The exception is time domain or pulse position modulation. Interestingly, this scheme dates back to the earliest (spark gap) radio transmitters. It's strange how history repeats itself. Another example is that the earliest radios were digital (Morse code), than analog was considered superior (analog voice transmission), now digital is back!

The final stage of a typical AM detector system is the amplifier, which is needed to provide adequate drive levels for an audio listening device, such as a headset or speaker. One of the disadvantages of the signal diode detector is its poor power transfer efficiency. But to understand this deficiency, you must first understand the limitation of the AM design that uses a half-wave rectifier at the receiver. At transmission from the source, the AM signal modulation process generates two copies of the information (voice or music) plus the carrier. For example, consider an AM radio station that broadcasts at a carrier frequency of 900 kHz. The transmission might be modulated by a 1000-Hz (1-kHz) signal or tone. The RF front end in an AM radio receiver will pick up the 900-kHz carrier signal along with the 1-kHz plus and minus modulation around the carrier, at frequencies of 901 and 899 kHz, respectively (see Fig. 8-5). The modulation frequencies are also known as the upper and lower

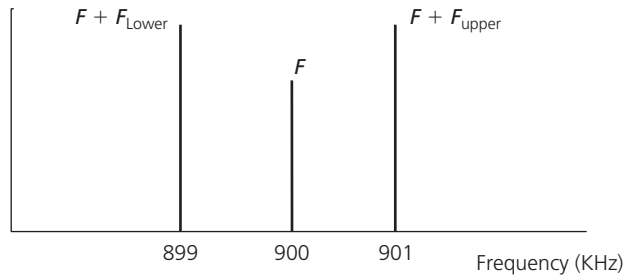


FIG. 8-5. Half-wave rectified AM radio produces upper and lower sidebands.

sideband frequencies, respectively. But only one of the sidebands is needed to completely demodulate the received signal. The other sideband contains duplicate information. Thus, the disadvantages of AM transmissions are twofold: (1) for a given information bandwidth, twice that bandwidth is needed to convey the information, and (2) the power used to transmit the unused sideband is wasted (typically, up to 50% of the total transmitted power).

Naturally, there are other ways to demodulate detector-based receiver architectures. We have just covered an approach used in popular AM receivers. Replacing the diode detector with another detector type would allow us to detect frequency-modulated (FM) or phase-modulated (PM) signals, this latter modulation commonly used in transmitting digital data. For example, many modern telecommunication receivers rely heavily on *phase-shift keying* (PSK), a form of phase (angle) modulation. The phrase "shift keying" is an older expression (from the Morse code era) for "digital."

All detector circuits are limited in their capability to differentiate between adjacent signal bands or channels. This capability is a measure of the *selectivity* of the receiver and is a function of the input RF filter to screen out unwanted signals and to pass (select) only the desired signals. As you might have guess from earlier discussions (see Chapter 2), selectivity is related to the *quality factor or Q* of the RF filter. A high Q means that the circuit provides sharp filtering and good differentiation between channels—a must for modern communication systems. Unfortunately, tuning the center carrier frequency of the filter across a large bandwidth while maintaining a high differentiation between adjacent channels is very difficult at the higher frequencies found in today's mobile devices. Selectivity across a large bandwidth is complicated by a receiver's *sensitivity* requirement, or the need to need to detect very small signals in the presence of system noise—noise that comes from the earth (thermal noise), not just the receiver system itself. The sensitivity of receiving systems is defined as the smallest signal that leads to an acceptable *signal-to-noise ratio* (SNR).

Receiver selectivity and sensitivity are key technical performance measures (TPMs) and will be covered in more detail in this chapter. At this point, it is sufficient to note that the AM diode detector architecture is limited in selectivity and sensitivity.

TRF Receiver

Moving up the scale in complexity, we come to the next evolutionary RF architecture: the tuned-radio-frequency (TRF) receiver (see Fig. 8-6). This early design was one of the first to use amplification techniques to enhance the quality of the signal reception. A TRF receiver consisted of several RF stages, all simultaneously tuned to the received frequency before detection and subsequent amplification of the audio signal. Each tuned stage consisted of a bandpass filter—which need not be an LC tank filter but could also be a Surface Acoustic Wave (SAW) filter or a dielectric cavity filter—with an amplifier to boost the desired signal while reducing unwanted signals such as interference. The final stage of the design is a combination of a diode rectifier and audio amplifier, collectively known as a grid-leak detector. In contrast to other radio architectures, there is no translation in frequency of the input signals, and no mixing of these input signals with those from a tunable LO. The original input signal is demodulated at the detector stage. On the positive side, this simple architecture does not generate the image signals that are common to other receiver formats using frequency mixers, such as superheterodynes (covered later in this chapter).

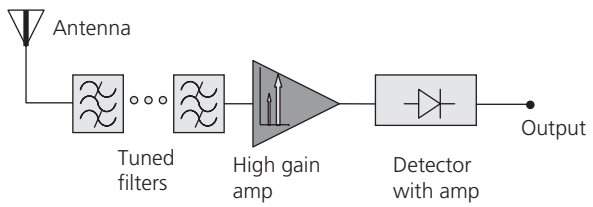


FIG. 8-6. Tuned Radio Frequency (TRF) architecture emphasizes multiple tuned filters.

The addition of each LC filter-amplifier stage in a TRF receiver increases the overall selectivity. On the downside, each such stage must be individually tuned to the desired frequency since each stage has to track the previous stage. Not only is this difficult to do physically, it also means that the received bandwidth increases with frequency. For example, if the circuit Q was 50 at the lower end of the AM band, say 550 kHz, then the receiver bandwidth would be 500/50 or 11 kHz—a reasonable value. However at the upper end of the AM spectrum, say 1650 kHz, the received bandwidth increases to 1650/50 or 33 kHz.

As a result, the selectivity in a TRF receiver is not constant, since the receiver is more selective at lower frequencies and less selective at higher frequencies. Such variations in selectivity can cause unwanted oscillations and modes in the tuned stages. In addition, amplification is not constant over the tuning range. Such shortcomings in the TRF receiver architecture have led to more widespread adoption of other receiver architectures, including direct-conversion and superheterodyne receivers, for many modern wireless applications.

Direct-Conversion Receiver

A way to overcome the need for several individually tuned RF filters in the TRF receiver is by directly converting the original signal to a much lower baseband frequency. In the direct-conversion receiver (DCR) architecture, frequency translation is used to change the high input frequency carrying the modulated information into a lower frequency that still carries the modulation but which is easier to detect and demodulate. This frequency translation is achieved by mixing the input RF signal with a reference signal of identical or near-identical frequency (see Fig. 8-7). The nonlinear mixing of the two signals results in a baseband signal prior to the detection or demodulating stage of the front-end receiver.

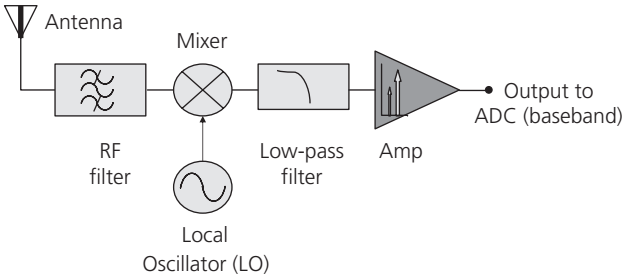


FIG. 8-7. Direct Conversion Receiver (DCR) architecture.

The reference signal is generated by a local oscillator (LO). When an input RF signal is combined in a nonlinear device, such as a diode or field-effect-transistor (FET) mixer, with an LO signal, the result is an intermediate-frequency (IF) signal that is the sum or difference of the RF and LO signals. When the LO signal is chosen to be the same as the RF input signal, the receiver is said to have a homodyne (or "same frequency") architecture and is also known as a zero-IF receiver. Conversely, if the reference signal is different from the frequency to be detected, then it's called a heterodyne (or "different frequency") receiver. The terms superheterodyne and heterodyne are synonyms ("super" means "higher" or "above" not "better").

In either homodyne or heterodyne approaches, new frequencies are generated by mixing two or more signals in a nonlinear device, such as a transistor or diode mixer. The mixing of two carefully chosen frequencies results in the creation of two new frequencies, one being the sum of the two mixed frequencies and the other being the difference between the two mixed signals. The lower frequency is called the beat frequency, in reference to the audio "beat" that can be produced by two signals close in frequency when the mixing product is an actual audio-frequency (AF) tone.

For example, if a frequency of 2000 Hz and another of 2100 Hz were beat together, then an audible beat frequency of 100 Hz would be produced. The end result is a frequency shifting from a higher frequency to lower—and in the case of RF receivers—baseband frequency.

Direct conversion or homodyne (zero-IF) receivers use an LO synchronized to the exact frequency of the carrier in order to directly translate the input signals to baseband frequencies. In theory, this simple approach eliminates the need for multiple frequency downconversion stages along with their associated filters, frequency mixers, and LOs. This means that a fixed RF filter can be used after the antenna, instead of multiple tuned RF filters as in the TRF receiver. The fixed RF filter can thus be designed to have a higher Q.

In direct-conversion design, the desired signal is obtained by tuning the local oscillator to the desired signal frequency. The remaining unwanted frequencies that appear after downconversion stay at the higher frequency bands and can be removed by a lowpass filter placed after the mixer stage.

If the incoming signal is digitally encoded, then the RF receiver uses digital filters within a DSP to perform the demodulation. Two mixers are needed to retain both the amplitude and phase of the original modulated signal: one for the in-phase (I) and another for a quadrature (Q) baseband output. Quadrature downconversion is needed since two sidebands generally form around any RF carrier frequency. As we have already seen, these sidebands are at different frequencies. Thus, using a single mixer, for a digitally encoded signal, would result in the loss of one of the sidebands. This is why an I/Q demodulator is typically used for demodulating the information contained in the I and Q signal components.

Unfortunately, many direct-conversion receivers are susceptible to spurious LO leakage, when LO energy is coupled to the I/Q demodulator by means of the system antenna or via another path. Any LO leakage can mix with the main LO signal to generate a DC offset, possibly imposing potentially large DC offset errors on the frequency-translated baseband signals. Through careful design, LO leakage in a direct-conversion receiver can be minimized by maintaining high isolation between the mixer's LO and RF ports.

Perhaps the biggest limitation of direct-conversion receivers is their susceptibility to various noise sources at DC, which creates a DC offset. The sources of unwanted signals typically are the impedance mismatches between the amplifier and mixer.

As noted earlier in this chapter, improvements in IC integration via better control of the semiconductor manufacturing process have mitigated many of the mismatch-related DC offset problems.

Still another way to solve DC offset problems is to downconvert to a center frequency near, but not at, zero. Near-zero IF receivers do just that, by downconverting to an intermediate frequency (IF) which preserves the modulation of the RF signal by keeping it above the noise floor and away from other unwanted signals. Unfortunately, this approach creates a new problem, namely that the image frequency and the baseband beat signals that arise from inherent signal distortion, can both fall within the intermediate band. The image frequencies, to be covered later, can be larger than the desired signal frequency, thus causing resolution challenges for the analog-to-digital converter.

Superheterodyne Receivers

In contrast to the simplicity of the direct-conversion receiver, the superheterodyne receiver architecture often incorporates multiple frequency translation stages along with their associated filters, amplifiers, mixers, and local oscillators (see Fig. 8-8). But in doing so, this receiver architecture can achieve unmatched selectivity and sensitivity. Unlike the direct-conversion receiver in which the LO frequencies are synchronized to the input RF signals, a superheterodyne receiver uses an LO frequency that is offset by a fixed amount from the desired signal. This fixed amount results in an intermediate frequency (IF) generated by mixing the LO and RF signals in a nonlinear device such as a diode or FET mixer.

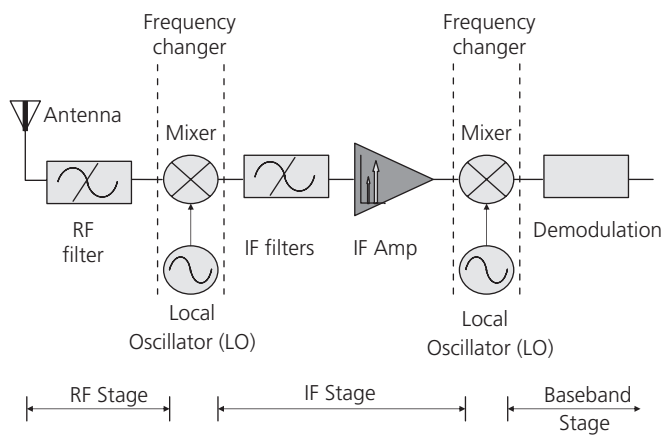


FIG. 8-8. Superheterodyne Architecture.

Generating local oscillators

The LO is often a phase-locked voltage-controlled oscillator (VCO) capable of covering the frequency range of interest for translating incoming RF signals to a desired IF range. In recent years, a number of other frequency-stabilization techniques, including analog fractional-N frequency synthesis and integer-N frequency synthesis as well as direct-digital-synthesis (DDS) approaches, have been used to generate the LO signals in wireless receiver architectures for frequency translation.

Any LO approach should provide signals over a frequency band of interest with the capability of tuning in frequency increments that support the system's channel bandwidths. For example, a system with 25-kHz channels is not well supported by a synthesized LO capable of tuning in minimum steps of only 1 MHz. In addition, the LO should provide acceptable single-sideband (SSB) phase-noise performance, specified at an offset frequency that coincides with the system's channel spacing. Referring to an LO's SSB phase noise offset 1 MHz from the carrier will not provide enough information about the phase noise that is closer to the carrier and that may affect communications systems

performance in closely spaced channels. Phase noise closer to the carrier is typically specified at offset frequencies of 1 kHz or less.

The LO source should also provide adequate drive power for the front-end mixers. In some cases, an LO buffer amplifier may be added to increase the signal source's output to the level required to achieve acceptable conversion loss in the mixer. And for portable applications, the power supply and power consumption of the LO become important considerations when planning for a power budget.

Mixers

Mixers are an integral component in any modern radio front end (see Fig. 8-9). Frequency mixers can be based on a number of different nonlinear semiconductor devices, including diodes and field-effect transistors (FETs). Because of their simplicity and capability of operation without DC bias, diode mixers have been prevalent in many wireless systems. Mixers based on diodes have been developed in several topologies, including single-ended, single-balanced, and double-balanced mixers. Additional variations on these configurations are also available, such as image-reject mixers and harmonic mixers which are typically employed at higher, often millimeter-wave, frequencies.

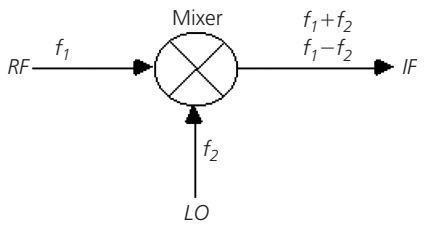


FIG. 8-9. Circuit symbol for radio mixer—or a diode multiplier in this example.

The simplest diode mixer is the single-ended mixer, which can be formed with an input balanced-unbalanced (balun) transformer, a single diode, an RF choke, and a lowpass filter. In a single-diode mixer, insertion loss results from conversion loss, diode loss, transformer loss. The mixer sideband conversion is nominally 3 dB, while the transformer losses (balun losses) are about 0.75 dB on each side, and there are diode losses because of the series resistances of the diodes. The equivalent circuit of a diode consists of a series resistor and a time-variable electronic resistor.

Moving up slightly in complexity, a single-ended mixer consists of a single diode, input matching circuitry, balanced-unbalanced (balun) transformer or some other means for injecting a mixing signal with the RF input signal, and a lowpass or bandpass filter to pass desired mixer products and reject unwanted signal components. Single-ended mixers are inexpensive and often used in low-cost detectors, such as motion detectors. The input balun must be highly selective to prevent radiation of the LO signal back into the RF port and out of the antenna.

Although the behavior of the diode changes with LO level, it can be matched for impedance at a particular frequency, such as the LO frequency, to achieve fairly consistent conversion-loss performance and flatness. The desired frequency converted signals are available at the IF port; the filter eliminates the unwanted high-frequency signal components generated by the mixing process. The LO drive level can be arbitrary, although different types of mixers and their diodes generally dictate an optimum LO drive level for mixer operation. The dimensions of the diode will dictate the frequency of operation, allowing use through millimeter-wave frequencies if the diode is made sufficiently small.

Some single-ended mixers use an antiparallel diode pair in place of the single diode to double the LO frequency and use the second harmonics of the LO's fundamental frequency, somewhat simplifying the IF filtering requirements. The trade-off involves having to supply higher LO power in order to achieve sufficient mixing power by means of the LO's second-harmonic signals.

A single-balanced mixer uses two diodes connected back to back. In the back-to-back configuration, noise components from the LO or RF that are fed into one diode are generated in the opposite sense in the other diode and tend to cancel at the IF port.

A double-balanced mixer is typically formed with four diodes in a quad configuration (see Fig. 8-10). The quad configuration provides excellent suppression of spurious mixing products and good isolation between all ports. Because of the symmetry, the LO voltage is sufficiently isolated from the RF input port and no RF voltage appears at the LO port. With a sufficiently large LO drive level, strong conduction occurs in alternate pairs of diodes, changing them from a low to high resistance state during each half of the LO's frequency cycle.

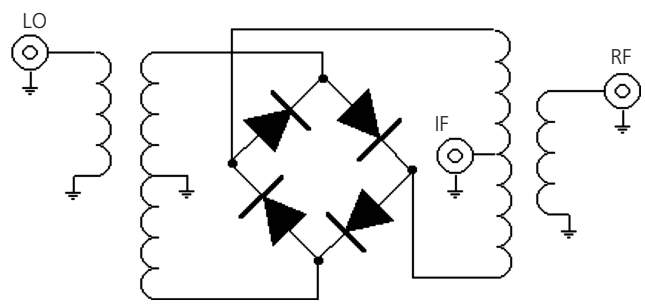


FIG. 8-10. Circuit schematic for a diode ring, double-balanced mixer.

Because the RF voltage is distributed across the four diodes, the 1-dB compression point is higher than that of a single-balanced mixer, although more LO power is needed for mixing. The conversion loss of a double-balanced mixer is similar to that of a single-balanced mixer, although the dynamic range of the double-balanced mixer is much greater due to the increase in the intercept point (recall IP discussion from earlier chapters).

By incorporating FET or bipolar transistors into monolithic IC mixer topologies, it is possible to produce active mixers with conversion gain rather than conversion loss. In general, this class of mixer can be operated with lower LO drive levels than passive FET or diode mixers, although active mixers will also distort when fed with excessive LO drive levels.

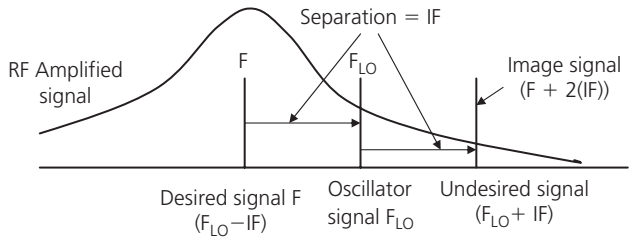


FIG. 8-11. Mixing results in unwanted frequencies, including image signals.

For RF front ends, wireless receivers, or even complete transceivers fabricated using monolithic IC semiconductor processes, the Gilbert cell mixer is a popular topology for its combination of low power consumption, high gain, and wide bandwidth. Originally designed as an analog four-quadrant multiplier for small-signal applications, the Gilbert-cell mixer can also be used in switching-mode operation for mixing purposes. Because it requires differential signals, the Gilbert-cell mixer is usually implemented with input and output transformers in the manner of double-balanced mixers.

Intermodulation and intercept points

The mixer generates IF signals that result from the sum and difference of the LO and RF signals combined in the mixer:

$$f_{IF} = f_{LO} \pm f_{RF} \qquad \text{(Eq. 8-1)}$$

These sum and difference signals at the IF port are of equal amplitude, but generally only the difference signal is desired for processing and demodulation so the sum frequency (also known as the image signal—see Fig. 8-11) must be removed, typically by means of IF bandpass or lowpass filtering. A secondary IF signal, which can be called f_{IF}^*, is also produced at the IF port as a result of the sum frequency reflecting back into the mixer and combining with the second harmonic of the LO signal. Mathematically, this secondary signal appears as:

$$f_{IF}^* = \pm[2f_{LO} - (f_{LO} - f_{RF})] \qquad \text{(Eq. 8-2)}$$

This secondary IF signal is at the same frequency as the primary IF signal. Unfortunately, differences in phase between the two signals typically result in uneven mixer conversion-loss response. But flat IF response can be achieved by maintaining constant impedance between the IF port and following component load (IF filter and amplifier) so that the sum frequency signals are prevented from re-entering the mixer. In terms of discrete components, some manufacturers offer constant-impedance IF bandpass filters that serve to minimize the disruptive reflection of these secondary IF signals. Such filters attenuate the unwanted sum frequency signals by absorption. Essentially, the return loss of the filter determines the level of the sum frequency signal that is reflected back into the mixer.

If a mixer's IF port is terminated with a conventional IF filter, such as a bandpass or lowpass type, the sum frequency signal will re-enter the mixer and generate intermodulation distortion. One of the main intermodulation products of concern is the two-tone, third-order product, which is separated from the IF by the same frequency spacing as the RF signal. These intermodulation frequencies are a result of the mixing of spurious and harmonic responses from the LO and the input RF signals:

$$f_{LO} = \pm(2f_{RF1} - f_{RF2}) \qquad \text{(Eq. 8-3)}$$

$$f_{LO} = \pm(2f_{RF2} - f_{RF1}) \qquad \text{(Eq. 8-4)}$$

But by careful impedance matching of the IF filter to the mixer's IF port, the effects of the sum frequency products and their intermodulation distortion can be minimized.

EXAMPLE: INTERMODULATION AND INTERCEPT POINTS

To get a better understanding of intermodulation products, let's consider the simple case of two frequencies, say f_1 and f_2. To define the products, we add the harmonic multiplying constants of the two frequencies. For example, the second order intermodulation products are $(f_1 + f_2)$; the third order are $(2f_1 - f_2)$; the fourth order are $(2f_1 + f_2)$; the fifth order are $(3f_1 - f_2)$; etc. If f_1 and f_2 are two frequencies of 100 kHz and 101 kHz—that is, 1 kHz apart—then we get the intermodulation products as shown in Table 8-1.

From the table it becomes apparent that only the odd order intermodulation products are close to the two fundamental frequencies of f_1 and f_2. Note that one third order product $(2f_1 - f_2)$ is only 1 kHz lower in frequency than f_1 and another

Order	Intermodulation Products		Ex: $f_1 = 100$ kHz, $f_2 = 101$ kHz	
1st Order	f_1	f_2	100 kHz	101 kHz
2nd Order	$f_1 + f_2$	$f_2 - f_1$	201 kHz	1 kHz
3rd Order	$2f_1 - fs_2$	$2f_2 - f_1$	99 kHz	102 kHz
	$2f_1 + f_2$	$2f_2 + f_1$	301 kHz	302 kHz
4th Order	$2f_2 + 2f_1$	$2f_2 - 2f_1$	402 kHz	2 kHz
5th Order	$3f_1 - 2f_2$	$3f_2 - 2f_1$	98 kHz	103 kHz
	$3f_1 + 2f_2$	$3f_2 + 2f_1$	502 kHz	503 kHz
Etc				

TABLE 8-1. Intermodulation Products

Continued on next page

EXAMPLE INTERMODULATION AND INTERCEPT POINTS—(Cont)

$(2f_2 - f_1)$ is only 1 kHz above f_2. The fifth order product is also closer to the fundamentals than corresponding even order products.

These odd order intermodulation products are of interest in the first mixer state of a superheterodyne receiver. As we have seen earlier, the very function of a mixer stage—namely, forming an intermediate lower frequency from the sum/difference of the input signal and a local oscillatory—results in the production of nonlinearity. Not surprisingly, the mixer stage is a primary source of unwanted intermodulation products. Consider this example: A receiver is tuned to a signal on 1000 kHz but there are also two strong signals, f_1 on 1020 kHz and f_2 on 1040 kHz. The closest signal is only 20 kHz away. Our IF stage filter is sharp with a 2.5-kHz bandwidth, which is quite capable of rejecting the unwanted 1020-kHz signal. However, the RF stages before the mixer are not so selective and the two signals f_1 and f_2 are seen at the mixer input. As such, intermodulation components are readily produced, including a third order intermodulation component $(2f_1 - f_2)$ at $(2 \times 1020 - 1040) = 1000$ kHz. This intermodulation product lies right on our input signal frequency! Such intermodulation components or out-of-band signals can easily cause interference within the working band of the receiver.

In terms of physical measurements, the two-tone, third-order intermodulation is the easiest to measure of the intermodulation interferences in an RF system. All that is needed is to have two carriers of equal power levels that are near the same frequency. The result of this measurement is used to determine the third-order intermodulation intercept point (IIP3), a theoretical level used to calculate third-order intermodulation levels at any total power level significantly lower than the intercept point.

The IP3 is the theoretical point on the RF input vs. IF output curve where the desired input signal and third-order products become equal in amplitude as the RF input is raised. Table 8-2 explains the meaning of the three critical intercept points in RF receiver design.

Order Intercept Points	Meaning
1 dB—1dB compression point	Compression is a measure of the linearity of a device. As the RF input signal level increases the IF output should follow for a mixing device. However, when the IF output cannot follow the RF input linearly and deviates by 1dB, this is referred to as the 1dB compression point.
IP2—Second-order intercept point	The second-order intercept point is the theoretical point on the RF input vs. IF output curve where the desired input signal and second order products become equal in amplitude as the RF input is raised.
IP3—Third-order intercept point	The third-order intercept point is the theoretical point on the RF input vs. IF output curve where the desired input signal and third-order products become equal in amplitude as the RF input is raised. Note: IIP3 is the input-referred IP3. IIP3 is just the output-referred IP3 (OIP3) divided by the small-signal gain.

TABLE 8-2. Intercept Points

Preselection filters

In a typical superheterodyne receiver architecture, the LO is offset by a precise fixed amount equal to the IF. The "image" version of the generated IF signal is separated from the desired signal frequency by a difference equal to twice the IF. In order to minimize the effects of unwanted signals entering the frequency-conversion mixer, a superheterodyne receiver employs a preselection filter in the front end to remove these unwanted signal products. An input matching network provides the impedance matching required to maximize energy flow from the receive antenna to the preselector filter and to the RF amplifier in a superheterodyne receiver. One of the goals of this matching circuitry is to minimize losses from the antenna coupling circuitry to the preselector filter, since such losses can decrease overall receiver sensitivity.

The impedance match between the receive antenna and the receiver's front-end circuitry may be subject to change. Some receivers, for example, may be used with different antennas, in which case the match would change with each antenna. And in some mobile applications, the effects of buildings, foliage, and other environmental factors can alter the impedance of the antenna and the consequent match with the front-end components. In addition, the impedance match between the antenna and front-end components is frequency-dependent and will be a function of the particular frequency bands for the communications system. In some cases, when the antenna must

cover a broad range of frequencies, it may be necessary to employ mechanical or electrical tuning to achieve an optimum impedance match for different receiver frequencies of interest.

Filter losses can impact overall superheterodyne receiver noise figure. Because high selectivity is needed to remove unwanted spurious and image signals from the receiver signal chain, preselection filtering is often broken into several stages with intervening LNAs between stages to minimize the effects of filter losses on receiver noise figure. In essence, this balance between filter losses and LNA gain in a superheterodyne receiver front end determines the overall selectivity and sensitivity of the receiver.

In a superheterodyne receiver, any filtering pertaining to the communications channel is performed at IF. Typically, fixed rather than tunable filters are used at IF since it is much more difficult to maintain constant bandwidth in a tunable filter than in a fixed filter. When different types of signals and modulation must be handled, the IF bandwidth can be readily changed by switching between different fixed filters. In general, a superheterodyne receiver enables high selectivity and high sensitivity because of the amount of processing that is possible at IF: filtering and amplification is simply easier and cheaper to achieve at IF than at microwave frequencies.

Although a superheterodyne receiver introduces problems with spurious responses that do not exist with the other receiver architectures, it is the ease and cost-effectiveness of filtering and boosting signals at IF that makes this configuration so attractive for wireless applications. In reviewing those two key parameters—sensitivity and selectivity—that characterize the performance of a superheterodyne front end, let's examine the factors that impact each parameter and how they affect overall receiver performance.

The sensitivity of a superheterodyne receiver is relative to an acceptable level of performance, such as the sensitivity for a given SNR in an analog radio front end or for a given bit-error rate (BER) in a digital radio front end. The sensitivity refers to the weakest possible signal level that a radio front end can process to achieve that agreed-upon SNR or BER. Sensitivity is affected by the level of noise external to the receiver and generated within the receiver itself. If the external noise is low enough, then the limit to receiver sensitivity will be established by the noise generated by the receiver's own components. Noise is contributed to a receiver by means of the receiving antenna, by contributions from the RF preselector filter, and from the active devices in the receiving system, including the front-end LNA.

System sensitivity and noise

The noise from each component in the front end adds to the receiver's noise floor, which sets the limit on the minimum signal level that can be detected. Noise can be characterized by its power spectral density (PSD), which is the power contained within a given bandwidth and is presented in units of watts per hertz. Every electronic component contributes some amount of noise to a receiving system, with the minimum amount of noise related to temperature known as the system's thermal noise, or kTB, where k is Boltzmann's constant 1.38×10^{-20} mW/K, T is the temperature in degrees Kelvin (K), and B is the noise bandwidth (in Hz).

At room temperature, the thermal noise generated in a 1-Hz bandwidth is

$$\begin{aligned} \mathrm{kTB} &= (1.38 \times 10^{-23} J/K)(293K)(1\,Hz) \\ &= 4.057 \times 10^{-21}\ \mathrm{W} = -174\,\mathrm{dBm} \\ &\text{or } -174\,\mathrm{dBm/Hz} \text{ in terms of PSD.} \end{aligned}$$

With an increase in bandwidth comes an increase in noise power and thus the importance of filtering in a superheterodyne receiver as a means of limiting the noise power. For this reason, the final IF filter in a superheterodyne receiver is made as narrow as possible to support the channel reception and to limit the amount of noise in the channel just prior to demodulation and detection. The final IF filter determines the noise bandwidth of the receiver, since it will be the most narrowband component in the front-end analog signal chain prior to detection.

Front-end receiver components are characterized in terms of noise by several parameters, including noise figure (NF) and noise factor (F). For the receiver as a whole, the noise factor is simply a ratio of the SNR at the output of the receiver compared to the SNR at the source of the receiver. For each component, similarly, the noise factor is the ratio of the SNR at the output to the SNR at the input. The noise figure is identical to the noise factor, except that it is given in dB. The noise factor is a pure ratio:

$$\text{Noise factor} = (\text{Output SNR}_2)/(\text{Input SNR}_1) \qquad \text{(Eq. 8-5)}$$

while the noise figure is presented in logarithmic form as

$$\mathrm{NF} = 10\log(\mathrm{SNR}_2/\mathrm{SNR}_1) \qquad \text{(Eq. 8-6)}$$

where SNR_2 is the output SNR of a component, device, or receiver and SNR_1 is the input SNR of the component, device, or receiver.

If an amplifier was ideal or a component completely without noise, its noise figure would equal 0 dB. In reality, the noise figure of an amplifier or component is always positive. For a passive device, the noise figure is equal to the insertion loss of the device. For example, the noise figure of a 1-dB attenuator without losses beyond the attenuation value is 1 dB.

In a superheterodyne front end, the noise power of the components that are connected or cascaded together rises from the input to the output as the noise from succeeding stages is added to the system. In a simple calculation of how the noise contributions of front-end stages add together, there is the well-known Friis's equation:

$$\mathrm{NFcascade} = 10\log\{[F_1 + (F_2 - 1)]/A_1\} \qquad \text{(Eq. 8-7)}$$

where $F =$ the noise factor, which is equivalent to $10^{NF/10}$ and A is the numerical power gain, which is equal to $10^{G/10}$ where G is the power gain is dB. From this equation, it can be seen how the noise factor of the first stage in the system (F_1) has a dominant effect on the overall noise performance of the receiver system.

Noise factor can be used in the calculation of the overall added noise of a series of cascaded components in a receiver, using the gain and noise factor values of the different components:

$$F = F_1 + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1 A_2} + \frac{F_n - 1}{A_1 A_2 \ldots A_{n-1}} \qquad \text{(Eq. 8-8)}$$

where the F parameters represent the noise factor values of the different front-end stages and the A parameters represent the numeric power gain levels of the different front-end stages. A quick look at this equation again shows the weight of the first noise stage on the overall noise factor. In a receiver with five noise-contributing stages ($n = 5$), for example, the noise of the final stages is greatly reduced by the combined gain of the components.

The noise floor of a receiver determines its sensitivity to low-level signals and its capability of detecting and demodulating those signals. The input referred noise level (noise at the antenna prior to the addition of noise by the other analog components in the receiver front end) is sometimes referred to as the minimum detectable signal (MDS). In some cases, a parameter known as *signal in noise and distortion* (SINAD) may also be used to characterize a receiver's noise performance, especially with a need to account for signals with noiselike distortion components. This parameter includes carrier-generated harmonics and other nonlinear distortion components in an evaluation of receiver sensitivity.

In a digital system, it is simpler to measure the bit-error rate (BER) induced by noise when a signal is weak. The BER affects the data rate so it is a more useful performance measure than the SNR for evaluating receiver sensitivity. With BER, the receiver's sensitivity can be referenced to a particular BER value. Typically a BER of 0.1%—e.g., in the GSM standard—is specified and the sensitivity of the receiver is measured by adjusting the level of the input signal until this BER is achieved at the output of the receiver.

A front end's noise floor is principally established by noise in components such as thermal noise, shot noise and flicker noise. At the same time, any decrease in gain will increase the noise floor. Thus, there must be enough margins in the system SNR to allow for a reduction in gain when making adjustments in gain for larger-level signals.

Front-End Amplifiers

The RF front-end component most commonly connected to an RF or IF filter is an RF or IF amplifier, respectively. Depending upon its function in the system, this amplifier may be designed for high output power (in the transmitter) or low-noise performance (in the receiver). At the receiver antenna, the receiver sensitivity will be a function of the ability of the preselector filter to limit incoming wideband noise and the front-end's low-noise amplifier (LNA) to provide enough gain to boost signal levels to an acceptable signal-to-noise ratio (SNR) for subsequent signal processing in the RF front end by mixers, demodulators, and/or ADCs.

As with the filters, an RF front-end's LNAs are specified depending on their location in the signal chain, either for relatively broadband use or for channelized use at the IF stages. An LNA is specified in terms of bandwidth, noise figure, small-signal gain, power supply and power consumption, output power at 1-dB compression, and linearity requirements. The linearity is usually judged in terms of third-order and second-order intercept points to determine the expected behavior of the amplifier when subjected to relatively large-level input signals. Ideally, an LNA can provide sufficient gain to render even low-level signals usable by the RF front-end's mixers and other components, while also handling high-level signals without excessive distortion.

At one time, LNAs fabricated with gallium arsenide (GaAs) process technology provided optimum performance in terms of noise figure and gain in RF and microwave communications systems. But ever-improving performance in silicon-germanium (SiGe) heterojunction-bipolar-transistor (HBT) now provides comparable or better noise-figure and gain performance in LNAs at frequencies through about 10 GHz.

In contrast to a superheterodyne receiver's noise, the other end of the dynamic range is the largest signal that the receiver can handle without distortion or, in the case of a digital receiver, degradation of the BER. In a receiver, excessively high signal levels can bring the onset of nonlinear behavior in the receiver's components, especially the mixers and LNAs. Such nonlinear effects are evidenced as gain compression, intermodulation distortion, and cross modulation, such as AM-to-PM conversion.

At large signal levels, harmonic and intermodulation distortion cause compression and interference that limit the largest signals that a receiver can handle. A receiver's dynamic range refers to the difference between the MDS and the maximum signal level. In a single-channel system, the dynamic range is essentially the difference between the 1-dB compressed output power and the output noise floor. The spurious-free dynamic range (SFDR) is defined as the range of input power levels from which the output signal just exceeds the output noise floor, and for which any distortion components remain buried below the noise floor.

IP3

The input third-order intercept point is often used as a measure of component and receiver power-handling capability. As mentioned earlier, it is defined as the extrapolated input power level per tone that would cause the output third-order intermodulation products to equal the single-tone linear fundamental output power. The output power at that point is the output third-order intercept point. The intercept point is fictitious in that it is necessary to extrapolate the fundamental component in a linear fashion and assume that the third-order intermodulation products increase forever with a 3:1 slope. In reality, the difference

between a component's actual output power at 1-dB compression and the third-order intercept point can be as little as 6 dB and as much as 20 dB. Along with the third-order intercept point, the second-order intercept point is also used as a measure of power-handling capability of dynamic range. It refers to the fictitious intersection of the second-harmonic output power with the fundamental-frequency output power.

In analyzing a receiver's dynamic range, it is important to note how the definitions of larger signals can vary. For example, for multiple-carrier communications systems, the peak power level will be much greater than the average power level because of the random phases of the multiple carriers and how they combine in phase. In a multicarrier system, the specified average power may be within the linear region of the system but the peaks may push the system into nonlinear behavior. This nonlinear behavior includes a phenomenon known as spectral regrowth and is characterized by such parameters as adjacent-channel power ratio (ACPR) where the power of a transmitted signal can literally leak into nearby channels because of intermodulation distortion.

Automatic gain control (AGC) can be used in a superheterodyne front end to decrease the gain when strong signals can cause overload or distortion, although there may be trade-offs for the SNR performance. If attenuation is added before the LNA in a receiver front end, for example, it can reduce the risk of nonlinearities caused by large signals at the cost of an increase in noise figure, as noted earlier with the 1-dB attenuator example. An AGC tends to sacrifice small-signal performance to achieve large-signal handling capability.

Selectivity

So far, this front-end discussion has covered sensitivity and dynamic range. Another important aspect of receiver design is achieving good selectivity, so that signals of interest are processed and unwanted spurious and interference signals are properly rejected or attenuated. Selectivity can be thought of as the capability of a receiver to separate a signal or signals at a given frequency or bandwidth from all other frequencies. Selectivity is generally a function of the matching networks between components, the filters, and the amplifiers in a front-end design. Achieving good selectivity with a filter, for example, requires that the filter achieve a wide enough passband to channel the designed signals and their modulation, but provide a sharp enough response with adequate rejection to eliminate unwanted signals not falling in the desired passband. Selectivity can also be thought of as the amount of rejection needed to reduce the level of an unwanted signal to some required amount at some nominal frequency from the desired passband. Selectivity can be achieved at different stages in a superheterodyne receiver, by using selective components and devices at the RF, IF, and baseband stages.

Receiver front-end selectivity should be as high as possible close to the antenna to remove large interfering signals before they enter the active devices in later stages and cause problems with distortion and overloading. For selective filters, impedance matching can be difficult and can adversely affect the performance of the other components in the front-end signal chain, notably the mixers.

Filters come in essentially four types: bandpass, band-reject, lowpass, and highpass filters. (For a more detailed discussion of filters, see Chapter 3.) A bandpass filter channels signals with minimal attenuation through a range of frequencies known as the passband, and rejects signals at frequencies above and below the passband. A band-reject filter (also known as a notch filter) is essentially the opposite of a bandpass filter. It rejects signals across one band (known as the stop band) and allows signals to pass with minimal attenuation at frequencies above and below the stop band. A lowpass filter channels signals with minimal attenuation below a specified cutoff frequency, while rejecting signals above that cutoff frequency. The cutoff frequency is commonly a point at which signal attenuation reaches 3 dB. A highpass filter is essentially the opposite of a lowpass filter, rejecting signals below the cutoff frequency and passing signals with minimal attenuation above the cutoff frequency.

Filters are judged in terms of a number of performance parameters, including insertion loss, return loss (or VSWR), rejection, ripple, selectivity (amplitude-versus-frequency response), group delay (how long a signal takes to propagate through a filter), phase response, and quality factor (Q). In a bandpass filter, insertion loss is the amount of signal attenuation above a 0-dB level that would be represented by an ideal transmission line in place of the filter. Insertion loss occurs due to a filter's dissipative elements (the resistors, inductors, capacitors, and transmission lines). Rejection is the amount of signal attenuation at specified points above and below the passband or center frequency, including the insertion loss.

Bandpass filters are defined in terms of their center frequency and the width of their passband. The center frequency of a bandpass filter can be defined arithmetically or geometrically. The geometric definition is usually employed in the filter design process, while the arithmetic definition is used to specify a filter. The arithmetic center frequency is simply the sum of the lower and upper bandedges divided by two. For example, for a bandpass filter with 3-dB frequencies of 900 and 1000 MHz, the arithmetic center frequency is $(900 + 1000)/2 = 950$ MHz.

The Q of a filter is the ratio of the midband frequency to the bandwidth. A narrowband filter, for example, with 3-dB band edges of 950 and 1000 MHz (center frequency of 975 MHz) has a Q of $975/50 = 19.5$. A bandpass filter with 3-dB bandwidth of 500 to 1000 MHz would have a much lower Q of $750/500 = 1.5$. Filter Q is related to the bandwidth, with narrower filter bandwidths resulting in higher filter Q values.

In fabricating filters, high-Q circuit elements (such as transmission lines, capacitors and inductors) are desirable for high-performance filter responses. Low-Q circuit elements tend to yield higher passband insertion loss and lower stopband attenuation. And lower-Q elements lead to a rounding of a filter's response, with poorly defined filter skirts.

Filter out-of-band rejection or attenuation can be increased by adding sections, although this also adds complexity and increases insertion loss due to the additional resonant elements. A wide range of filter responses, such as equiripple and linear-phase responses, are available in modern designs. The equiripple response, for example, provides minimum amplitude deviations across the passband. A linear-phase filter is suitable for preserving the content of pulsed waveforms typically used in modern communications systems, such as the Orthogonal Frequency Division Multiplex (OFDM) modulation used in WiMAX systems.

ADC'S EFFECT ON FRONT-END DESIGN

Analog-to-digital converters (ADCs) are commonly used in receivers for wireless applications for either IF or baseband signal sampling. The choice of ADC is generally determined by the rest of the receiver architecture, and can be affected by the selectivity of the filters, the dynamic range afforded by the front-end amplifiers, and the bandwidth and type of modulation to be processed. For example, the level or dynamic range of signals expected to be presented to the ADC will dictate the bit resolution needed for the converter. For example, in an example double-downconversion receiver architecture developed for broadband wireless access (BWA) applications using the IEEE 802.16 WiMAX standard, IF sampling can be performed with a 12-b ADC.

For cases where a single downconversion approach, with a subsequent higher IF, is used, a higher-resolution, 14-b converter is recommended in order to compensate for the less efficient selectivity of the single-conversion receiver and to avoid ADC saturation in the presence of high-level interference signals. Along with its input bandwidth (which should accommodate the highest IF of interest for a particular receiver design) and bit resolution, an ADC can also be specified in terms of its spurious-free dynamic range (SFDR). The ADC's sensitivity is influenced by wideband noise, including spurious noise, and often can be improved through the use of an antialiasing filter at the input of the ADC to eliminate sampling of noise and high-frequency spurious products.

To avoid aliasing when converting analog signals to the digital domain, the ADC sampling frequency must be at least twice the maximum frequency of the input analog signal. This minimum sampling condition—derived from Nyquist's theorem (covered in earlier chapters)—must be met in order to capture enough information about the input analog waveform to reconstruct it accurately.

In addition to selecting an ADC for IF or baseband sampling, the choice of buffer amplifier to feed the input of the converter can affect the performance possible with a given sampling scheme. The buffer amplifier should provide the rise/fall time and transient response to preserve the modulation information of the IF or baseband signals, while also providing the good amplitude accuracy and flatness needed to provide signal amplitudes at an optimum input level to the ADC for sampling.

Now let's consider an example using lowpass signals where the desired bandwidth goes from 0 (DC) to some maximum frequency (f_{MAX}). The Nyquist criterion states that the sampling frequency needs to be at least $2f_{MAX}$. So, if the ADC is sampling at a clock rate of 20 MHz, this would imply that the maximum frequency it can accept is 10 MHz. But then how could an FM radio broadcast signal (say, at 91.5 MHz) be converted using such a relatively low sampling rate? Here's where the design of the RF front end becomes critical. The RF receiver must support an intermediate frequency (IF) architecture, which translates a range of relatively high input frequencies to a lower-frequency range output (at the IF band). Using the example of the FM radio, with a tunable bandwidth of 88 to 108 MHz, then the receiver's front end must process signals over that tunable bandwidth to a lower IF range of no higher than 10 MHz. Such a design would ensure that the previously mentioned 20-MHz ADC could handle these IF signals without aliasing.

SOFTWARE DEFINED RADIOS

Up to this point, we have focused on the hardware implementation of RF front ends. But the capability of creating RF front-end architectures that are controlled by software has become a reality, thanks to the continuing migration of analog functionality to digital chips and the performance improvements in DSP technology. Today, a designer can use software defined radio (SDR) systems to process RF signals that were traditionally handled by analog and RF front-end circuitry. This new approach is being implemented in next generation, very high frequency (~6 GHz) applications that require flexible reconfiguration of the front end, like wireless base stations, mobile communication (often military) devices, and IPTV set-top boxes.

SDR technologies known as "frequency agile" systems allow the conversion of any analog wireless signal directly into digital baseband data, regardless of frequency. Semiconductor manufacturing processes have typically used gallium arsenide (GaAs) materials to demonstrate the RF-to-digital (RF/D) converter chips, though future production versions will convert the design to low-cost CMOS and BiCMOS technologies.

The great advantage of SDR systems is the on-the-fly interoperability among differing communications frequencies. In other words, a single RF receiver can communicate directly with different devices on multiple different frequency bands, using only software to receive, translate, and process the different signals.

CASE STUDY—MODERN COMMUNICATION RECEIVER

In this chapter we have introduced the design architectures common in most RF front-end receivers. We have defined a number of key parameters used to characterize the response of a receiver, including sensitivity and selectivity.

Now let's see how all of the concepts and parameters fit into the development of a typical modern communications transceiver. Such a communication front-end/back-end could be used to support a common US air interface like second generation

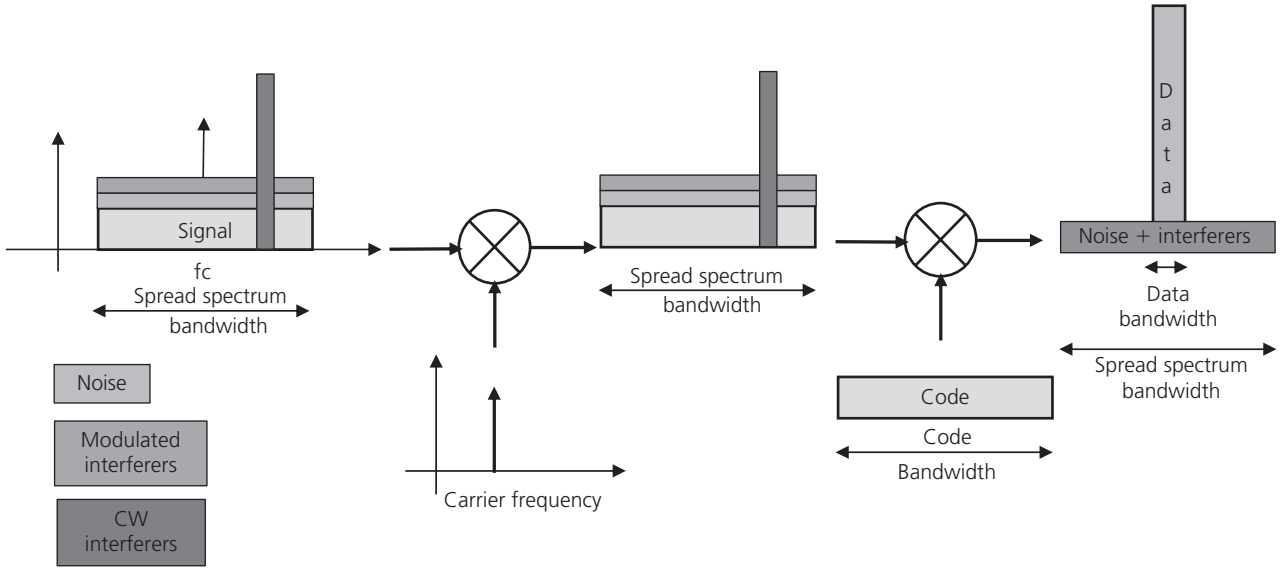


FIG. 8-12. CDMA reception process. (Courtesy of Agilent)

(2G), narrow-band Code Division Multiple Access (CDMA) or third-generation (3G), multimedia enabled wideband CDMA (W-CDMA) systems. By changing the RF tuning, this same architecture could be used for dual–band GSM (used in Europe) or TDMA systems in the same radio band, since the processing and demodulation is performed in the post-baseband, digital section.

This last point is important, since this chapter has focused on traditional analog receiver design as are used in TDMA designs. As the name implies, Time Division Multiple Access (TDMA) technology divides a radio channel into sequential time slices. Each channel user takes turns transmitting and receiving in a round-robin fashion. TDMA is a popular cellular phone technology since it provides greater channel capacity than its predecessor—Frequency Division Multiple Access (FDMA). Global System for Mobile Communications (GSM), an established cellular technology in Asia and Europe, uses a form of TDMA technology.

In this case study, though, we focus on Code Division Multiple Access (CDMA) designs for two reasons. First, the basic receiver architecture is similar to TDMA. Second, CDMA receiver designs are predominant in the US and are gaining global acceptance.

In CDMA systems, the received signal occupies a relatively narrow channel within a 60-MHz spectral allocation between 1930 MHz and 1990 MHz. W-CDMA channels operate on a wider bandwidth (3.84 MHz) than standard CDMA systems. All CDMA users can transmit at the same time while sharing the same carrier frequency. A user's signal appears to be noise for all except the correct receiver. Thus, the receiver circuit must decode one signal among many that are transmitted at the same time and at the same carrier frequency, based on correlation techniques.

The CDMA reception process is as shown in Fig. 8-12. Several mixer stages are required to separate the carrier frequency and the code bandwidth. Once complete, the desired data signal can be separated from the "noise" (other user channels) and interference.

In a modern receiver front-end communication system, the received signal is amplified, mixed down to IF, and filtered before being mixed down to baseband where it is digitized for demodulation (see Fig. 8-13). A double (multi-mixer) superheterodyne architecture is typically used in a CDMA receiver. The RF front-end consists of the typical duplexer and low-noise amplifier (LNA) to provide additional signal gain to compensate for signal losses from the subsequent image-reject filter and then the first mixer. Two downconverter stages are used between the RF and baseband subsystems. The first mixer downconverts the signal to a first IF stage of 183 MHz. The second mixer completes the downconversion from the IF stage to baseband. The I/Q outputs from the second mixer stage are digitally decoded and demodulated in the baseband DSP subsystem.

The receiver architecture contains an I/Q demodulator to separate the information contained in the I (in-phase) and Q (quadrature) signal components prior to the baseband input—recall earlier discussion on direct conversion techniques. Overall key receiver requirements (derived from the IS-95/IS-98 standards) for a CDMA system are defined by (see Fig. 8-14):

- Reference sensitivity is the minimum receiver input power, at the antenna, at which bit error rate (BER) $<= 10^{-3}$. This results in an acceptable noise power (P_n) within the channel bandwidth of -99 dBm. The acceptable noise power (-99 dBm) within the channel bandwidth results in a receiver noise figure (NF)

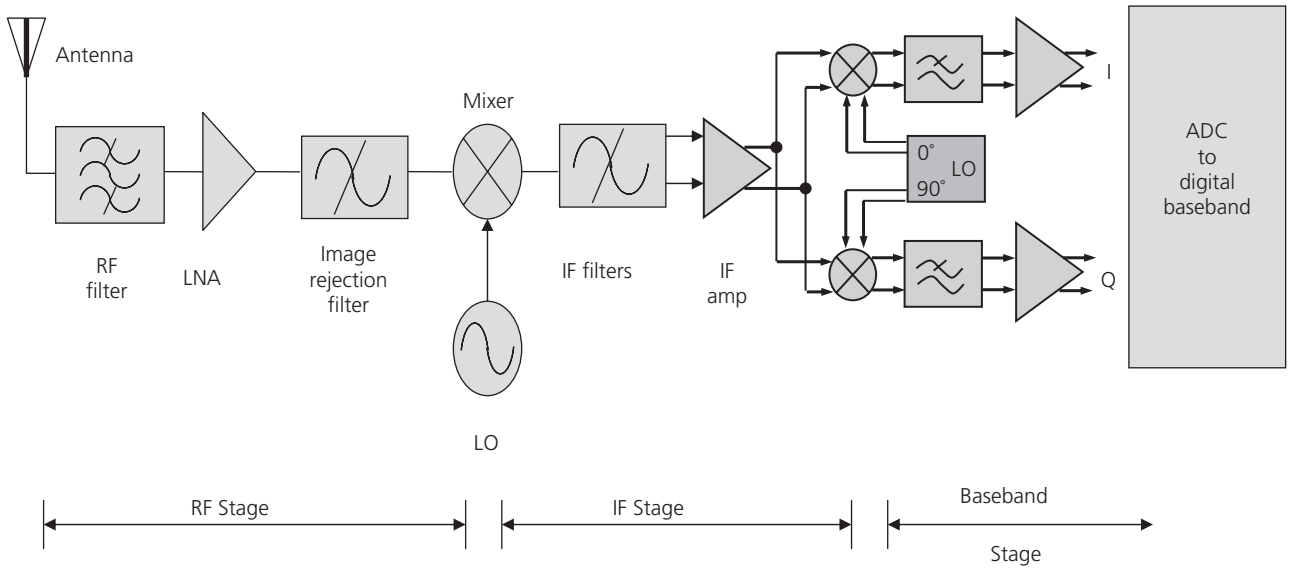


FIG. 8-13. Block diagram of a modern RF communication receiver.

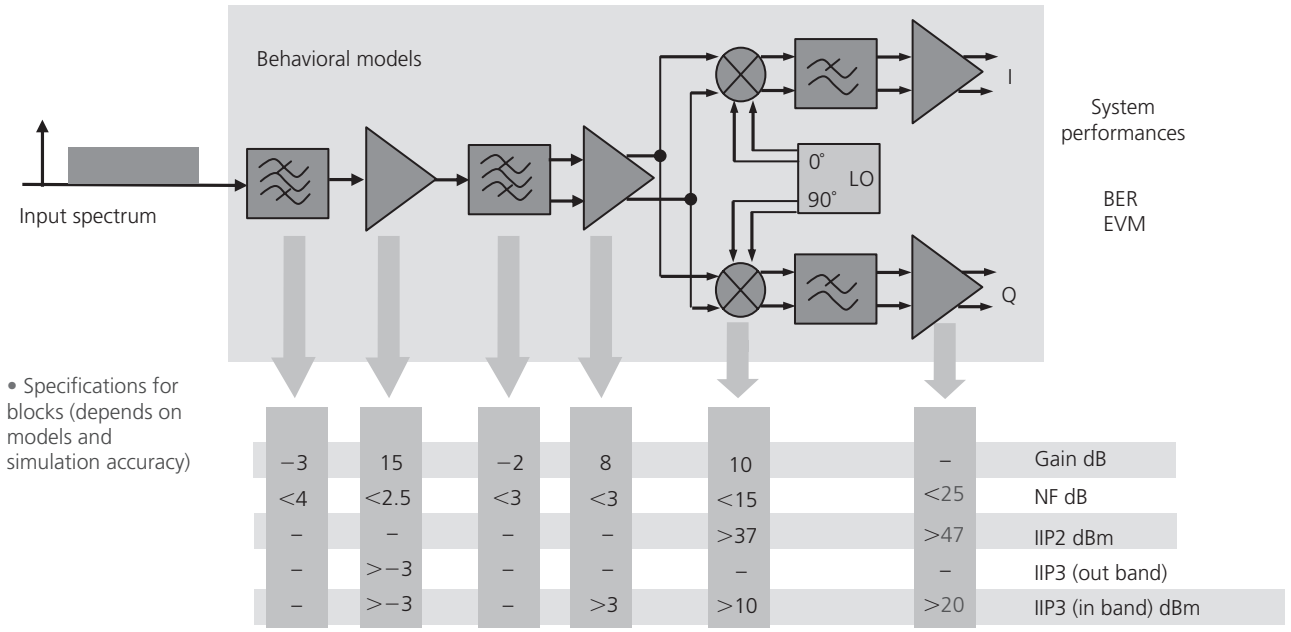


FIG. 8-14. Typical CDMA system specifications.

of 9 dB. Recall that the noise figure of a receiver is the ratio of the SNR at its input to the SNR at its output. It characterizes the degradation of the SNR by the receiver system.

- Adjacent channel selectivity (ACS) is the ratio of the receive filter attenuation on the assigned channel frequency to the receiver filter attenuation on the adjacent channel frequency.
- Intermodulation results from nonlinear modulation of two pure input signals. As mentioned in Chapter 7, when two or more signals are input to an amplifier simultaneously, the second-, third-, and higher-order intermodulation components are caused by the sum and difference products of each of the fundamental input signals and their associated harmonics. Of particular importance to CDMA receiver design is the third-order intercept point (IP3).

Now let's consider the issue of measuring and controlling the RF signal power. On the receive side, the input signal will generally vary over some dynamic range. This may be due to weather

conditions or to the source of the received signal moving away from the receiver (e.g., a mobile handset being operated in a fast car). But as explained earlier in this chapter, we want to present a constant signal level to the analog-to-digital converter (ADC) to maintain the proper resolution of the ADC. This will also maximize the signal-to-noise ratio (SNR). As a result, receive signal systems typically use one or more variable gain amplifiers (VGAs) that are controlled by power measurement devices that complete the automatic-gain-control (AGC) loop. Recall the signal processing on the receive side occurs after the IF and ADC stages.

An inaccurate received signal strength indication (RSSI) measurement can result in a poor leveling of the signal that is presented to the ADC. This will cause either overdrive of the ADC (input signal too large) or waste valuable dynamic range (input signal too small).

IF Amplifier Design

Several amplifiers are used in the IF stage of most receivers. Consider the architecture we've been examining, noting one of these amplifiers just prior to the two-stage *I*/*Q* mixer. This amplifier can be designed as an analog or digital AGC loop. Where fast regulation of gain is required, the inherent latency of a digitally controlled automatic gain control (AGC) loop may not be acceptable. In such situations, an analog AGC loop may be a good alternative (see Fig. 8-15).

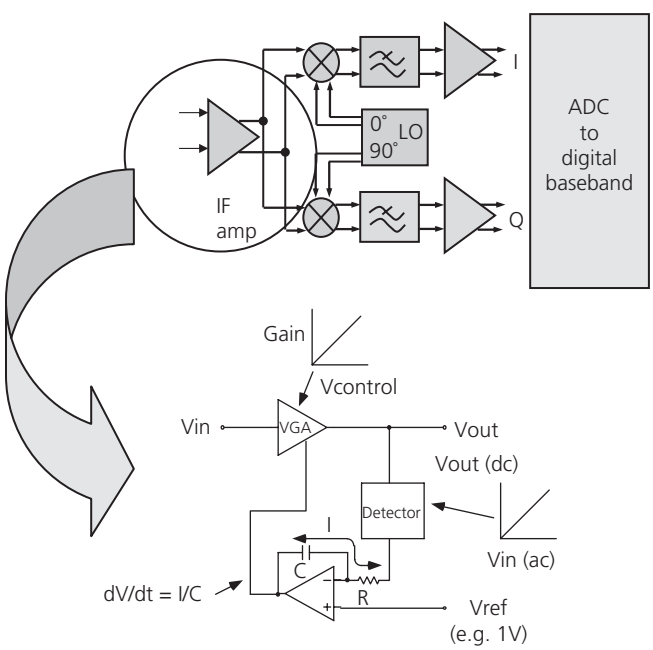


FIG. 8-15. Generic components of a AGC analog amplifier loop.

Beginning at the output of the variable gain amplifier (VGA), this signal is fed, usually via a directional coupler, to a detector. The output of the detector drives the input of an op amp, configured as an integrator. A reference voltage drives the non-inverting input of the op amp. Finally the output of the op-amp integrator drives the gain control input of the VGA. Now, let's examine how this circuit works.

We will assume initially that the output of the VGA is at some low level and that the reference voltage on the integrator is at 1 V. The low detector output results in a voltage drop across integrator resistor R. The resulting current through this resistor can only come from the integrator capacitor C. Current flow in this direction increases the output voltage of the integrator. This voltage, which drives the VGA, increases the gain (we are assuming that the VGA's gain control input has a positive sense, that is, increasing voltage increases gain). The gain will be increased, thereby increasing the amplifier's output level until the detector output equals 1 V. At that point, the current through the resistor/capacitor will decrease to zero and the integrator output will be held steady, thereby settling the loop. If capacitor charge is lost over time, the gain will begin to decrease. However, this leakage will be quickly corrected by additional integrator current from the newly reduced detector voltage.

The key usefulness of this circuit lies in its immunity to changes in the VGA gain control function. From a static perspective at least, the relationship between gain and gain control voltage is of no consequence to the overall transfer function. Based upon the value of V_{ref}, the integrator will set the gain control voltage to whatever level is necessary to produce the desired output level. Any temperature dependency in the gain control function will be eliminated. Also, nonlinearities in the gain transfer function of the VGA do not appear in the overall transfer function (V_{out} vs. V_{ref}). The only requirement is that the gain control function of the VGA be monotonic. It is crucial however that *detector* be temperature stable.

The circuit as we have described it has been designed to produce a constant output level for varying input levels. Because this results in a constant output level, it becomes clear that the detector does not require a wide dynamic range. We only require it to be temperature stable for input levels that correspond to the setpoint voltage V_{ref}. For example, the diode detector circuits previously discussed which have poor temperature stability a low levels but reasonable stability at high levels, might be a good choice in applications where the leveled output is quite high.

If, the detector we use has a higher dynamic range, we can now use this circuit to precisely set VGA output levels over a wide dynamic range. To do this, the integrator reference voltage, V_{ref}, is varied. The voltage range on V_{ref} follows directly from the detector's transfer function. For example, if the detector delivers 0.5 V for an input level of −20 dBV, a reference voltage of 0.5 V will cause the loop to settle when the detector input is −20 dBV (the VGA output will be greater than this amount by whatever coupling factor exists between VGA and detector).

The dynamic range for the variable V_{out} case will be determined by the device in the circuit with the least dynamic range (i.e., gain control range of VGA or linear dynamic range of detector). Again it should be noted that the VGA does not need a precise gain control function. The "dynamic range" of the VGA's

gain control in this case is defined as the range over which an increasing gain control voltage results in increasing gain.

The response time of this loop can be controlled by varying the RC time constant of the integrator. Setting this at a low level will result in fast output settling but can result in ringing in the output envelope. Setting the RC time constant high will give the loop good stability but will increase settling time.

It is interesting to note that use of the term AGC (automatic gain control) to describe this circuit architecture is fundamentally incorrect. The term AGC implies that the gain is being automatically set. In practice, it is the output level that is being automatically set, so the term ALC (automatic level control) would be more correct.

This case study has offered just a sample of the many issues that must be considered when design any communication receiver system. Numerous books and internet resources are available for those looking to understand more of the fascinating technology.

CHAPTER 9

RF DESIGN Tools

Electronic design automation (EDA) is a category of automated design tools used to develop and manufacture the hardware component of a range of electronic systems including field programmable gate arrays (FPGAs)/programmable logic devices (PLDs), integrated circuits (ICs), printed circuit boards (PCBs), and systems-on-a-chip (SoCs). These systems often contain a mix of analog, digital and RF circuitry. While the analog and RF sections of the system make up a much smaller portion of the design than their digital counterpart, they are by far the most difficult to design, requiring substantially more time and expertise. In fact, in systems design, roughly 75% of the design time is spent on RF design. Of that 75%, 25% is spent on actual design work and 75% on interface, library and integration issues.

Given these facts, it might be natural to assume that RF design is today aided by a wealth of highly advanced EDA tools, flows and methodologies. This is not the case. In fact, EDA tools and methodologies for the digital realm remain far more advanced then those which are typically used for analog and RF design. Even so, automated design tools—whether those of the inexpensive, web-based variety or those which are costly and full-featured—can play an important role in enabling you to create fully optimized RF designs more productively than was previously possible with more manual techniques.

In this chapter, we will discuss the specific types of tools which can aid both the expert and novice design engineer with RF design tasks. But first, we begin by taking a look at exactly what constitutes an RFIC design flow and the design languages you will need to know to make that flow work.

DESIGN TOOL BASICS

A wide range of EDA tools are used by RF engineers throughout the product lifecycle for the design, verification and test of RF circuits. They replace a process in which all design work and circuit layout was once done by hand in a sequential manner. The engineer would design a circuit and then hand it "over the wall" to another engineer who would verify the work. In turn, the verified design would then be handed off to an engineer who would lay out the circuit.

Design tools not only automate this process, but have transformed it from a sequential to a concurrent process as well. Today, via a methodology known as co-design, many different engineers can now work on various parts of a design at the same time. And the design process now begins at the system, as opposed to the circuit, level. The result is increased productivity and time to market, along with a reduction in costly, time-consuming errors. And, because EDA tools automate engineering best practices, acting as a repository of sorts for engineering expertise, even the novice engineer can be productive using them.

The most common types of EDA tools used today are:

- Schematic Capture tools—Captures a schematic representation of a design using components from a component library.
- Logic Synthesis tools—Translates high-level design languages into netlists.
- Place-and-Route (P&R) tools—Takes netlists and decides where to put each wire and transistor in the design layout.
- Simulation tools—Takes a description of a circuit (netlist) and simulates its behavior. These tools are critical to ensuring that you get your design right the first time.
- Verification tools—Verifies that a design will work as expected. Includes Design Rule Checkers (DRCs) and equivalence checkers which work to verify the design netlist and that the netlist matches the original design, respectively.

DESIGN LANGUAGES

Design languages are used to describe, or model, the circuit. Much of today's RF design typically begins either by hand with a textual description or with an algorithmic description of a data transformation, such as a model in MATLAB or a C/C++ object class. These procedural languages are especially useful in developing individual untimed algorithms, but they do not work well as system-level design tools because they lack the built-in constructs necessary for time and concurrency.

Consider, for example, that in a mobile communication system, many algorithms operate at different rates in a base station.

Algorithms also operate in the handset. Concurrently, RF energy bounces off objects such as a building or a tree to create a dynamically fading channel. Modeling the performance of an RF system realistically under these conditions requires the designer to capture the concurrency of all of this behavior, in addition to any significant physical properties. This is virtually an impossible task.

System-level design languages like SystemC or System Verilog offer a possible patch to this deficiency. SystemC, for example, provides transaction-level modeling constructs by adding a class library and a kernel to C++.

Whether the design starts life in the form of a system-level or procedural language, once its implementation is fleshed out, you will need to leverage a number of different easily accessible tools to assist in the hardware description language (HDL) coding process.

An HDL is a standard, text-based expression of the temporal behavior and/or spatial circuit structure of an electronic system. The two leading standard HDLs include Verilog and VHSIC Hardware Description Language (VHDL). Both have extensions (Verilog-AMS and VHDL-AMS) that allow them to work with analog and mixed-signal designs, including some RF. Generally, though, RF system design requires many more constructs than are present in these language extensions. Alternative language options exist that further extend the analog/mixed signal versions of Verilog and VHDL into the RF space, but because these additional constructs are vendor specific, they constrain the design database to a specific vendor.

It is important to note that whichever language you opt to use, the design description or model will eventually be used to perform mixed-level (e.g., analog, digital and RF) simulation. Therefore it may make sense to adopt a language that is broadly supported by a range of design tools and/or utilized by the rest of your design team. Many standard EDA tools offer support for a wide range of standard programming and HDLs.

What follows is a closer look at the languages in use today for RF design.

Verilog

Verilog is a hardware description language for describing electronic circuits and systems. It supports the design, verification and implementation of analog, digital, and mixed-signal circuits at various levels of abstraction and can be used for verification through simulation, for timing analysis, test analysis (e.g., testability analysis and fault grading) and logic synthesis.

Verilog-AMS

Verilog-AMS is an extension to Verilog that supports analog and mixed-signal modeling using both continuous-time and event-driven modeling semantics. It allows designers to use modules that encapsulate high-level behavioral descriptions as well as structural descriptions of systems and components.

Verilog-A

Verilog-A is an industry standard modeling language and is the continuous-time subset of Verilog-AMS. It is specifically designed for modeling behavior of analog components. The behavior of each module can be described mathematically in terms of its terminals and external parameters applied to the module. The structure of each component can be described in terms of interconnected sub-components.

As an example of Verilog A, consider the Advanced Design System (ADS) and RF Design Environment (RFDE) solutions from Agilent Technologies (www.agilent/com/find/eesof). Both feature a Verilog-A solution—Verilog-A Compiler—which is suitable for RF circuit and system design.

Verilog-A Compiler is based on technology from Tiburon Design Automation and enables simulation speed that is close to built-in C models. It can be used to create behavioral models of RF blocks such as mixers, amplifiers, etc., as well as transistor models for things like MOSFETs, BJTs and HBTs. Many transistor model examples are provided with the product, and they can be used as the basis for more complete device models.

A typical simulation status showing the cache creation message and a compilation/load message is illustrated in Fig. 9-1.

SystemVerilog

SystemVerilog is considered the industry's first unified hardware description and verification language (HDVL) standard (IEEE 1800), and is a major extension of the established IEEE 1364 Verilog language. Based on extensions to Verilog, it was developed originally by Accellera to dramatically improve productivity in the design of large gate-count, intellectual property (IP)-based, bus-intensive chips. SystemVerilog is targeted primarily at the chip implementation and verification flow, with powerful links to the system-level design flow.

VHDL

VHDL is a fairly general-purpose language, although it requires a simulator on which to run the code. It can read and write files on the host computer, so a VHDL program can be written that generates another VHDL program to be incorporated in the design being developed. As a result, it is possible to use VHDL to write a testbench that verifies the functionality of the design using files on the host computer to define stimuli, interacts with the user, and compares results with those expected.

VHDL-AMS

VHDL-Analog Mixed Signal (VHDL-AMS), the IEEE-endorsed standard modeling language (IEEE 1076.1), was created to provide a general-purpose, easily exchangeable and open language for modern analog-mixed-signal designs. Models can be exchanged between all simulation tools that adhere to the VHDL-AMS standard. The example in Fig. 9-2 illustrates the architecture basics of an amplifier in VHDL-AMS.

```
HPEESOFSIM (*) 2003C.day Dec 14 2003 (built: 12/14/03 21:28:57)
Copyright Agilent Technologies, 1989-2003.
A verilog-A compiled model library (CML) cache has been created at
    '/users/bobl/hpeesof/agilent-model-cache/cml/0.99/hpuxll'
Compiling Verilog-A file
'/rfdeExamples/VerilogA_Tutorial/sinRampSrc/veriloga/veriloga.va'
AGILENT-VACOMP (*) 2003C.day Dec 14 2003 (built: 12/15/03 01:49:33)
Tiburon Design Automation (R) Verilog-A Compiler Version 0.99.121203.
Copyright (C) Tiburon Design Automation, Inc. 2002-2003. All rights
reserved.
Loading Verilog-A module 'sinRampSrc' from
'/users/bobl/hpeesof/agilent-model-cache/cml/0.99/hpuxll/veriloga_09106/1
ib.hpuxll/veriloga.cml'.
TRAN Tranl[1] <input.ckt> time={0 s->100 ms)
Resource usage:
    Total stopwatch time: 23.21 seconds.
```

FIG. 9-1. When a Verilog-A cell view is simulated for the first time, the Agilent RFDE Verilog-A Compiler compiles the associated Verilog-A code and caches the result so that it may be used directly in all subsequent simulations of that view.

VHDL-AMS/FD

There are also extensions to VHDL-AMS that are specifically geared toward high-frequency design (Fig. 9-3). VHDL-AMS/Frequency Domain (FD) is an extension language that supports harmonic balance simulation for frequency domain analysis. It provides the RF designer with the capability to describe complex microwave and RF devices (e.g., wireless systems) from architecture to transistors. The RICON Harmonic Balance simulator is currently the only VHDL-AMS simulator to use the VHDL-AMS/FD extensions.

VHDL-RF/MW

VHDL-RF/MW is an extension to VHDL-AMS that supports design at radio and microwave frequencies. Extensions facilitate non-lumped terminals, finite-element modeling, parasitics, and frequency-domain modeling. VHDL-RF/MW is supported in the FTL Systems' Auriga modeling and verification solution which spans system level, register-transfer-level (RTL), analog/mixed signal, radio-frequency down to parasitic/microwave detail.

C/C++

C is a general-purpose, procedural, imperative computer programming language. C++ is an extension of C that provides object-oriented functionality with C-like syntax. It adds greater typing strength, scoping and other tools useful in object-oriented programming, and permits generic programming via templates to simplify high-level programming and offer a better approach to large-scale programming. C++ is also a larger language with more features and complexity than C, but C++ can improve productivity with its greater number of features. A list of features that C++ supports and C does not includes:

Classes	Inline functions
Member functions	Default arguments
Constructors and destructors	Function overloading
Derived classes	Namespaces
Virtual functions	Exception handling
Abstract classes	Run-time type identification
Access control (public, private, protected)	// comments
Friend functions	True const
Pointers to members	Declarations as statements
Static members	Automatically typedef struct tags
Mutable members	Type safe linkage
Operator overloading	New and delete
References	Bool keyword
Templates	Safer and more robust casting

SystemC

SystemC provides hardware-oriented constructs within the context of C++ as a class library implemented in standard C++. Its use spans design and verification from concept to implementation in hardware and software. The language provides an interoperable modeling platform, which enables the development and exchange of very fast system-level C++ models. It also provides a stable platform for development of system-level tools.

MATLAB/RF Toolbox/Simulink

MATLAB is a high-level programming language and interactive, integrated environment, developed by The MathWorks (www.mathworks.com/products/matlab/). It enables the engineer to perform computationally intensive tasks faster than with

```
-- This model contains simultaneous conditional use statements, but
-- the conditions are such that all quantities are continuous over the
-- switch between equations; thus no "break" is required.
--
library IEEE;
architecture basic of amplifier is
   terminal Cout: electrical;
   constant c1: real := imax/(sr*1.0e6);
   constant gmnom: real := 2.0 * IEEE.math_real.math_pi * ugf * c1;
   constant r1: real := gain/gmnom;
   constant dvmax: real := imax/gmnom;
--
-- The reader will find that diagraming these quantities as the edges of
-- a directed graph on the terminals will aid in understanding the
-- design.
--
   quantity Iinput through Vinput across vin to vinb;
   quantity Ivin through refer to vin;
   quantity Ivinb through refer to vinb;
   quantity Icout through Vcout above refer to cout;
   quantity Icout_vout through Vcout_vout above cout to vout;
   quantity Vvout_vp above vout to vp;
   !
quantity Vvout_vn above vout to vn;
--
-- These non-branch quantities are used as temporaries in building up
-- the equation for Icout
--
   quantity gain_current, dp_current, outstage_current: current;
begin
--
-- local terminal Cout
   Icout == gain_current + dp_current - outstage_current;
   Icout_vout == Vcout_vout/r0;
--
-- input stage.
   Iinput == Vinput/rin;
   Ivin == iin;
   Ivinb == iin;
--
-- gain stage.
-- The following relationship must hold between gmnon,dvmax and imax
-- to prevent a discontinuity in gain_current
   assert abs(gmnon*dvmax) = abs(imax);
   if Vinput > dvmax use
      gain_current == imax;
   elsif Vinput < -dvmax use
      gain_current == -imax;
   else
      gain_current == gmnom * Vinput;
   end use;
--
-- dominant pole.
   dp_current == c1 * Vcout'dot + Vcout/r1;
--
-- output stage limiting.
-- outstage_current is zero when Vvout_vp=-so!
ft or Vvout_vn=+soft
-- so there is no discontinuity at the transit

ion points
   if Vvout_vp > -soft use
      outstage_current == gmnom * (Vvout_vp+soft);
   elsif Vvout_vn < soft use
      outstage_current == gmnom * (Vvout_vn-soft);
   else
      outstage_current == 0.0;
   end use;
end basic;
```

FIG. 9-2. An amplifier, as described in VHDL-AMS. This code example was excerpted from a conference paper entitled "Mixed-Mode Simulation," by Kenneth Bakalar from Compass Design Automation, now Synopsys.

```
entity nonlinear_amp is
generic (k1, k2, k3: real); port (quantity a_in, a_out : real);
end entity amp;
architecture eq of amp is
begin
a_in*k1+a_in**2*k2+a_in**3*k3==a_out; - nonlinear effects included
end architecture;
...
The example shows that the entity "nonlinear_amp" contains the parameters k1, k2, k3
The test bench for the amplifier is the following:
# amplifier jobs: - comment line
res1=HBparametric([3.e6,5,0.01,1000])# - performs HB
print "================= results ================"
print res1; # - print res1 (complex structure)
print "================= tout.H1 =================".
print res1.getX("tout'reference",[1]) # of "tout'reference" variable
# - gets and prints only spectrum
# of "tout'reference" variable at first harmonic
```

FIG. 9-3. Code for defining an amplifier using VHDL-FD. This code example was excerpted from "VHDL-AMS Extensions Enable RF Harmonic Balance Simulation," originally published in High Frequency Electronics magazine in March 2004 and written by Mark Rencher, Ridgetop Group.

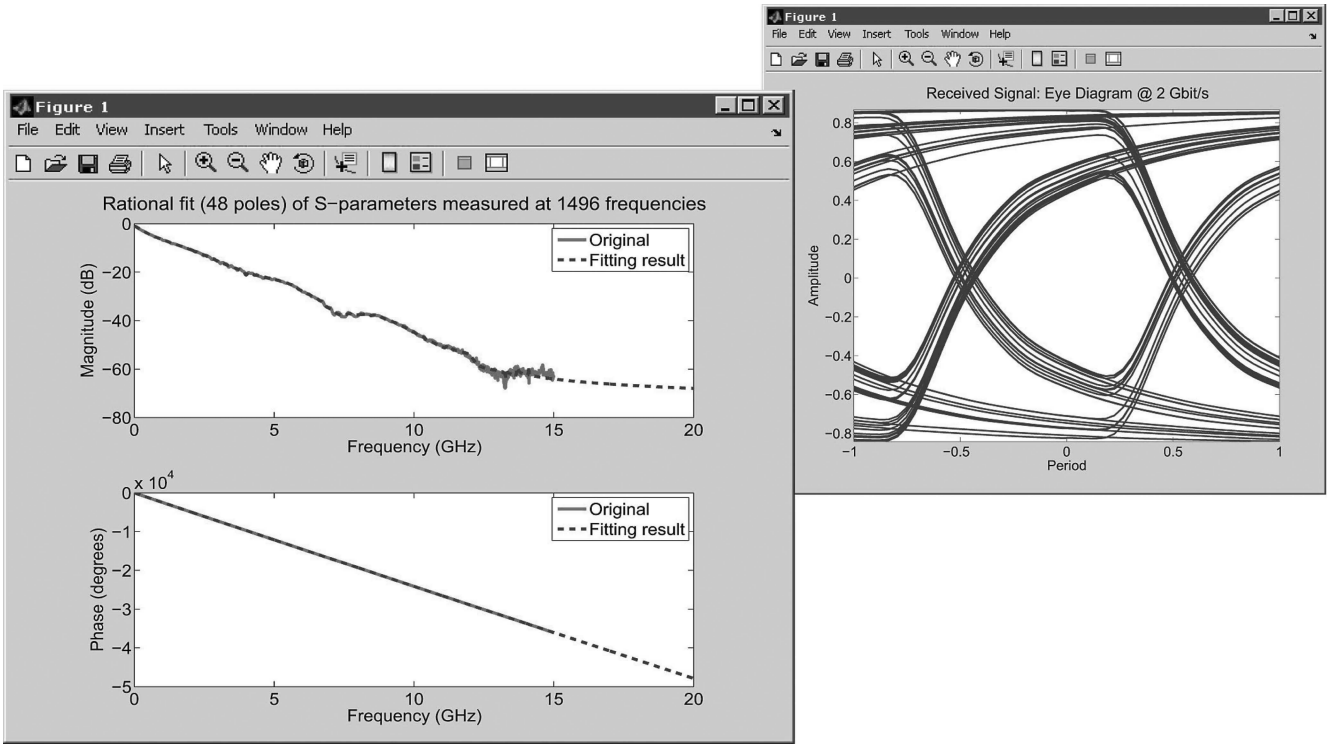


FIG. 9-4. RF Toolbox from The MathWorks enables engineers to design, model, analyze, and visualize networks of radio frequency components.

traditional programming languages such as C and C++. It is used for numerical computations, symbolic computations and scientific visualizations and runs in interpreted, as opposed to compiled, mode.

RF toolbox is a MATLAB function and class library that RF engineers use for designing, modeling, analyzing, and visualizing networks of RF components (Fig. 9-4). It works within the Simulink environment and offers users a library of blocks to model the behavior of RF amplifiers, mixers, filters, and transmission lines.

Simulink is a block library tool for modeling, simulating and analyzing dynamic systems. It works with MATLAB.

SPICE

SPICE is a powerful general-purpose analog circuit simulator, developed at the University of California, Berkeley. It is used

to verify circuit designs and to predict the circuit behavior. It falls into the category of an equivalent-circuit model software package. The SPICE simulator was long considered the de-facto industrial standard for computer-aided circuit analysis. However, the limitations of SPICE have forced the creation of new languages like Verilog and VHDL-AMS.

RFIC DESIGN FLOW

The process of chip design from concept to production is called a design flow. A design methodology is a set of procedures that accompany a design flow to achieve a particular outcome. For example, a design tool vendor specializing in locating and identifying signal integrity issues may develop a methodology that helps ensure any potential signal integrity issues are eliminated early in the design flow when they are easier to find and cheaper to fix. Note that most design tool vendors, especially those known for offering solutions that span the entire product lifecycle (e.g., from R&D to production), typically offer their own individual flows and methodologies.

A typical RFIC design flow is pictured in Fig. 9-5. Recall that while digital designers have grown accustomed to unified design flows, RF designers must often piece together tools from various vendors to develop a unified flow for their designs. EDA vendors like Agilent Technologies, Ansoft, Applied Wave Research (AWR), Cadence Design Systems, and Mentor Graphics are now working to rectify this situation via the development of RFIC flows developed around their key RF design tools and platforms. Therefore, while the flow depicted in Fig. 9-5 may be typical, it is by no means the only possible RFIC flow available today.

A functional representation of this flow is provided in Fig. 9-6. The key steps that comprise this flow are system design, circuit design (e.g., design capture and simulation), circuit layout, parasitic extraction, and full-chip verification.

System Design

During the system-level design phase, an RF specification in the form of a behavioral model is created using either Matlab, C or some other system-level design language. A behavioral model is essentially a model that describes what the system does. To create the model, the engineer begins with perfect RF components and then adjusts the performance specifications of the components until performance degradation occurs. The goal of this process is to determine what level of noise, nonlinearity, and frequency domain distortions the system can tolerate.

A behavioral testbench is also developed during this phase. It will serve as the framework for more complex mixed-level (e.g., analog, digital and RF) simulations, where blocks can be inserted at the transistor level and verified in a system context. The behavioral model and testbench are often referred to as an executable specification and are used together to validate the specification. Once the specification has passed verification (via any number of functional verification tools), it is passed to the circuit designer

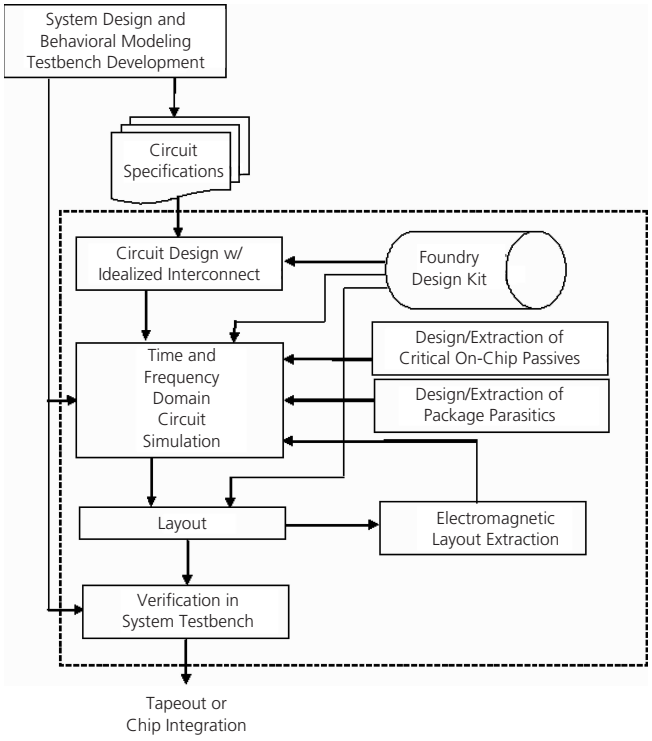


FIG. 9-5. This flow chart, courtesy of UMC, depicts a typical RFIC design and verification flow.

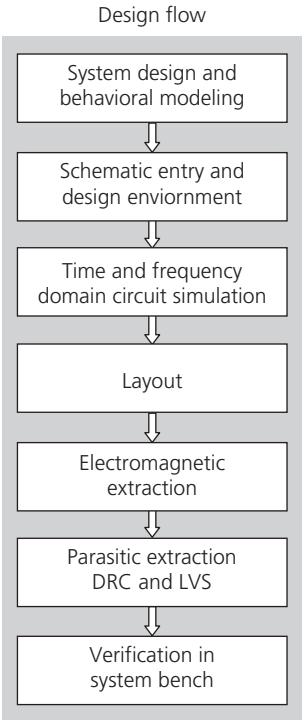


FIG. 9-6. Functional design flow representation.

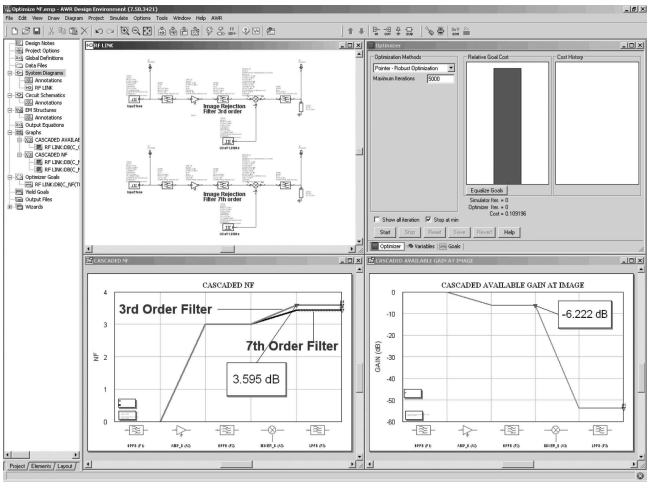


FIG. 9-7. The Visual System Simulator (VSS) from Applied Wave Research is a comprehensive software suite for the design of complete, end-to-end communications systems. The tool's RFA RF Budget Analysis module, shown here, enables designers to find potential pitfalls early in the design process, at the system-level design phase, thereby saving significant design cycle time and speeding products to market. Using it, designers can make traditional RF cascaded measurements such as gain, noise figure, and third-order intercept, inclusive of image noise, along a communication link.

where the detailed design, RTL or board-level schematics, are created and compared to the original behavioral description. When they match, the design is deemed "functionally" complete.

Note that the engineer might also utilize an advanced architectural planning tool to help determine the design's specification (Fig. 9-7). Such tools work by allowing the engineer to specify the correct system architecture. They then generate suitable specifications for each of the underlying components in complex communications designs. As a result, they help designers reduce time-to-market by eliminating iterations and rework, and cut system costs by ensuring that components are not over-specified and thus unnecessarily expensive.

Circuit Design

During the circuit design phase, the verified executable specification is further refined to incorporate results generated by the engineer while creating the actual circuit design. The process normally begins with the conversion of the specification into a block diagram of the various functions that the circuit must perform. The various blocks are considered in detail still—at an abstract stage—but with much more focus on the details of the electrical functions to be provided. The block diagram is then synthesized into a netlist for simulation. This will allow the engineer to find mistakes early in the design cycle before a physical device is fabricated.

Circuit simulation is performed in the time and frequency domains to characterize critical performance metrics. The choice of domain will depend on a number of factors including: the type of circuit and its size, the type of simulation and desired output, the designer's comfort level, and any personal preference the designer may have (Fig. 9-8). As circuits are completed at block level, they are verified within the top-level context with behavioral stimulus and descriptions for the surrounding chip. If the engineer is unable to achieve the specification for a certain component, the system-level model and executable specification can be modified accordingly.

There are two important things to note during the circuit-design phase. First, it is common during this stage to employ the use of modeling tools (e.g., for spiral inductors). This task will be discussed in detail later in the chapter. It is worth noting, however, that these tools allow you to determine what is and is not feasible to achieve with your circuit. The information that you obtain from this exercise is then fed back into earlier stages of the design process to verify system-level performance.

Secondly, prior to the start of the circuit design phase, the engineer will select a process technology (e.g., 0.18 or 0.25 micron design geometries). Once this selection is made, a process design kit (PDK) can be obtained from the appropriate foundry or IC fabrication house and installed within your design environment. The PDK helps jump start the design process by providing you with all of the foundry-specific process models and data that you will need for circuit design. The PDK is provided in a file format that is compatible with the design environment (e.g., IC design flow and tools) you are using.

Circuit Layout

The next step in the process is to perform circuit layout. Here, the individual circuit components which you have chosen to carry out each function in the system, are physically laid out and the electrical connections of each component are decided. Automated design-rule-driven and connectivity-driven layout tools may be used judiciously, especially to take advantage of direct ties to schematic and design-rule checking. Critical analog blocks are generally manually routed using a full custom approach to ensure that highly sensitive analog circuitry meets specifications.

Parasitic Extraction

The next step in the design process is to extract parasitics. This is a crucial step as high-speed requirements make RF circuits extremely sensitive to the effects of parasitics, including parasitic inductance, passive component modeling, as well as signal integrity issues. Once extracted, these effects are added to the circuit simulations.

For sensitive blocks like VCOs and critical radio blocks, full-wave three-dimensional electromagnetic (EM) simulation allows the extraction of the full layout at the block level. This rigorous method simulates all high-frequency layout effects including on-chip inductors, interconnect, coupling between on-chip passives and to other interconnect structures, and substrate coupling. No assumptions are made regarding parasitics or coupling. The resulting accurate models of these effects are used to replace the models that were created earlier in the design process.

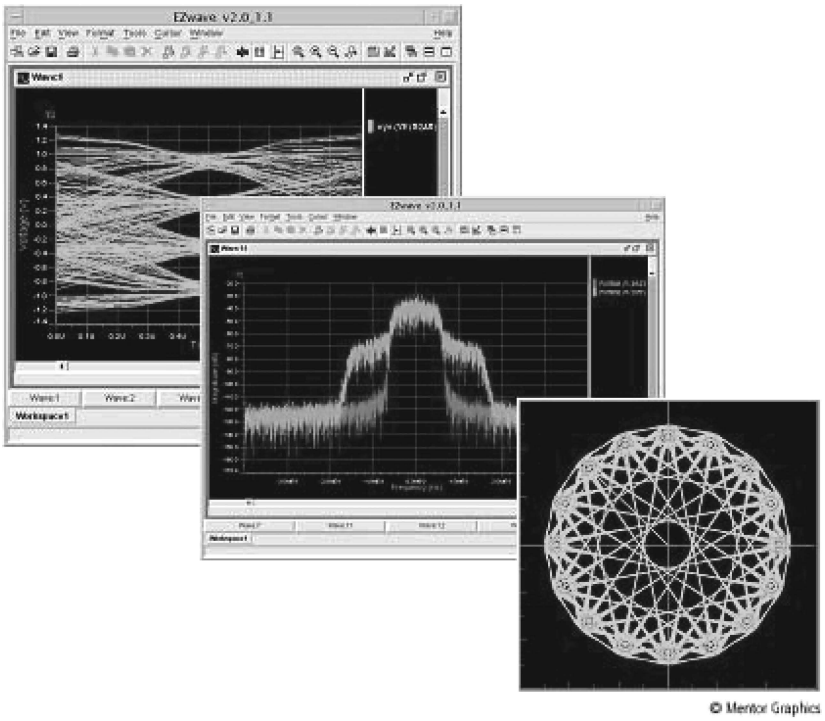


FIG. 9-8. The Eldo RF simulator from Mentor Graphics provides a set of dedicated algorithms to accurately and efficiently handle the multi-GHz signals in modern wireless communication applications. It features steady-state analysis of RFIC circuits excited with periodic (single-tone) or quasi-periodic (multi-tone) sources as well as a complete RF toolbox, including Smith Chart diagrams, gain and stability circles, and minimum noise figure. Eldo RF is part of the Mentor IC design flow, which includes Design Architect IC and IC Station for front-end and back-end analog and mixed-signal design.

Note that net-based RLC extractors also have their place in the RFIC flow, but they require designer input to manage which parasitic effects to include in the extraction. Unfortunately, it is not always clear which parasitic effects are most critical in the circuit context. Rigorous EM extraction of the entire block should be used to remove any doubt in the process.

Full-Chip Verification

The final step in the design process, prior to tape-out or additional chip integration, is to perform full-chip verification in a system (e.g., behavioral) testbench. The verification can include transistor-level circuits for multiple circuit blocks with incorporation of all extracted parasitics. The system should allow designers to select the particular level of abstraction for individual circuit blocks in order to make reasonable trade-offs between accuracy and simulation run time.

RFIC DESIGN FLOW EXAMPLE

What follows is an example of a wireless RFIC flow that is part of the Cadence Design Systems RF Design Methodology Kit which is based on the Advanced Custom Design Methodology (ACD).

Fig. 9-9 depicts the wireless RFIC flow. In this flow, the design collateral from the system design process is used as the first and highest abstraction level. System-level descriptions form the executable testbench for the top-level chip. Models of the surrounding system are combined with a high-level model of the chip, producing an executable specification. System requirements serve as the first specification to drive the chip-level requirements, and ultimately turn into repeatable testbenches and regression simulations.

Once a high-level executable specification is achieved, the design process continues by identifying areas of concern in the design. Plans are then developed for how each area of concern will be verified. The plans specify how the tests are preformed and which blocks are at the transistor level during the test. Resist the temptation to specify and write models that are more complicated than necessary. Start with simple models and only model additional effects as needed.

HDL Multi-Level Simulation

This process starts with HDL modeling (including all RF blocks and any analog content and/or digital blocks) for the entire RFIC being added to the system-level testbench. First a behavioral model of the full chip within a top-level testbench is created. Initially it is used to verify the partitioning, block functionality and ideal performance characteristics of the IC. Then, it serves as the basis to facilitate mixed-level simulations, where blocks can be inserted at the transistor level and verified in a top-level context. This behavioral model/testbench also serves as the regression template, allowing for continuous verification as blocks mature.

In the full simulation environment, several views of the same circuit will exist including, for example, a behavioral view, a

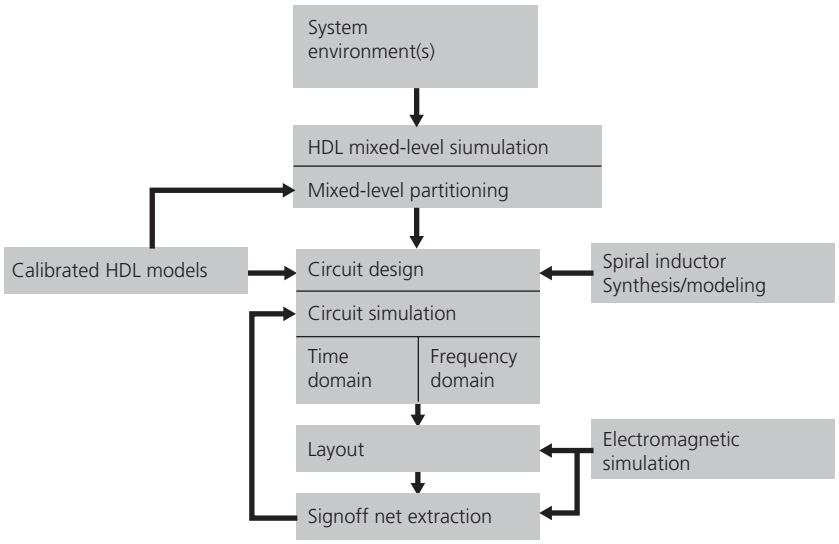


FIG. 9-9. Wireless RFIC flow utilized as part of the Cadence Design Systems RF Design Methodology Kit.

pre-layout transistor-level view, and several views of parasitic information. As blocks mature, it may be necessary to add more transistor-level information to test RF/analog and RF/digital interfaces. This will require the use of a mixed-signal simulator capable of handling analog, digital, and RF descriptions and mixed behavioral-level with transistor-level abstractions. You will need to pick the appropriate views of each block and manage the runtime versus accuracy tradeoffs through simulation options by—as an example—sending the transistors to a FastSPICE simulator or keeping the transistors in a full SPICE mode.

Block Circuit Design

During this stage a preliminary circuit design is created that allows for early circuit exploration and a first-cut look at performance specifications. This early exploration leads to a top-level floorplan, which for RFIC is sensitive to noise concerns and block-level interconnect. Passive components (e.g., spiral inductors) are then synthesized and an initial placement of them on the chip is completed.

Note that this approach enables two key activities: creating early models for spiral inductors that can be used in simulation before the block-level layouts are complete, and allowing for an initial analysis of mutual inductance between the spirals.

Next, simulation is performed. A single PDK and associated environment allows for a smooth determination and selection of the simulation algorithm desired. Results are displayed through an appropriate display for the simulation type selected. As circuits are completed at block level, they are verified within the top-level context with behavioral stimulus and descriptions for the surrounding chip.

Physical Implementation

Layout automation (e.g., automated routing, connectivity-driven layout, design-rule-driven layout, and placement) is used to implement the circuit. Analog-capable routers can help with differential pairs and shielding wires, and allow for manual constraints per line. Highly sensitive circuitry requires a manual approach.

Parasitic Extraction

Once the layout is complete, EM simulation is used to provide highly accurate models for passive components. Net-based parasitic extraction becomes a key element of the process as layouts emerge. Less sensitive interconnects may require RC only, whereas more sensitive lines may require RLC. For lines with spirals attached, these can be extracted fully with RLC plus the associated inductor component, even with substrate effects added for those lines that are the most sensitive. Lines that contain a "full" extraction can be mixed and matched with the component models for passive components that were created earlier.

Also during this stage, designers will check to see whether noisy circuits (such as digital logic and perhaps PLLs) are affecting the highly sensitive RF circuits. If so, they can either modify the floorplan accordingly or add guardbands around the noisy circuitry.

Note that it is often impractical to both simulate the entire design at transistor level and include all the parasitic information. One approach is to extract calibrated behavioral models using the extracted view of the design blocks. This will not capture the effects of the parasitics on interconnect between blocks; therefore, hierarchical extraction capabilities that extract only parasitics of interconnect between design blocks is required.

Calibrated Models

As blocks are completed, the initial behavioral models can be back annotated for key circuit performance parameters to provide more accurate HDL-level simulation. Verification of a block by mixed-level simulation, therefore, becomes a three-step process.

First, the proposed block functionality is verified by including an idealized model of the block in system-level simulations. Next, the functionality of the block as implemented is verified by replacing the idealized model with the netlist of the block. Finally, the netlist of the block is replaced by an extracted model. By comparing the results achieved from simulations that involve the netlist and extracted models, the functionality and accuracy of the extracted model can be verified. From then on, mixed-level simulations of other blocks are made more representative by using the verified extracted model of the block rather than the idealized model.

SIMULATION EXAMPLE 1

To better understand how to verify that the blocks in an RFIC (e.g., amplifiers, mixers, frequency dividers, and baseband chain) are wired together correctly, consider the example of a receiver with a wireless local area network (WLAN) 802.11b signal as shown in Fig. 9-10. This example utilizes the RF Design Environment (RFDE) from Agilent Technologies.

For this example, the direct-conversion receiver consists of an LNA, I and Q mixers and baseband receive chains, and a frequency divider to generate quadrature LO signals. In this example, the task is to supply a WLAN 802.11b input signal to the LNA, and verify that the I and Q baseband output signals still track the input modulation, at various points along the baseband chain (outputs from the mixers, outputs from the DC offset cancellation circuit, outputs from the VGAs and from the tunable filters).

The simulation uses time-domain baseband data sources that are the I and Q data of a WLAN 802.11b signal. These data sources modulate a sinusoid via an ideal modulator to generate the test signal at the input. If you have the I and Q time-domain data waveforms for some other modulation format, you may use this technique along with the Cadence Envelope Follower simulator or the Agilent Circuit Envelope simulator to generate a modulated test signal.

The simulation time required depends linearly on the desired stop time, with about 10 minutes being the minimum to get useful information. The circuit has 2547 devices, 1377 of which are nonlinear.

The data display in Fig. 9-11 shows comparisons between the input signal and the signals at various points in the receive chain. By changing the Vtest equation, you can select the test point to be the output of the mixer, the output of the DC offset cancellation circuit, the output of the first variable gain amplifier, or the output of the baseband filter. The plots compare the spectra, magnitudes, phases, real and imaginary parts, and the trajectories of the input and Vtest signals.

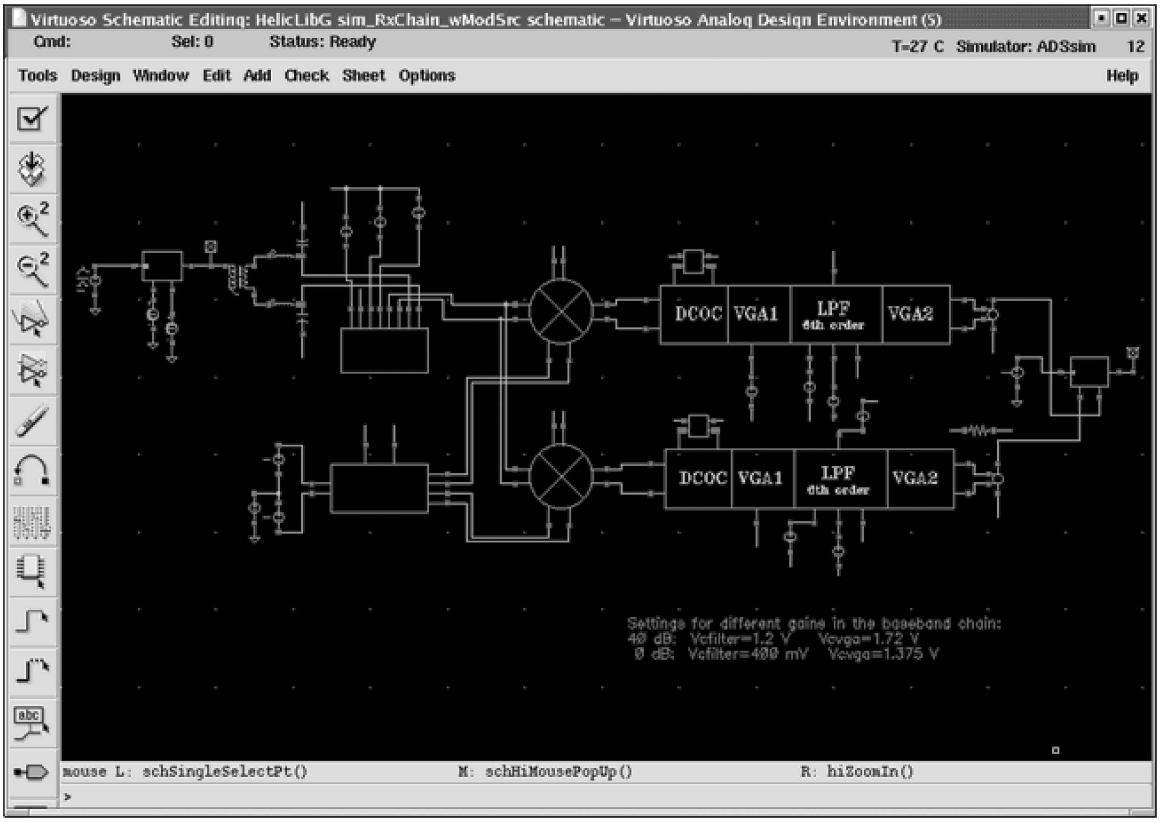


FIG. 9-10. This schematic, as viewed using Agilent Technologies' RFDE software design platform for large-scale RF/mixed-signal IC design, shows a direct-conversion receiver.

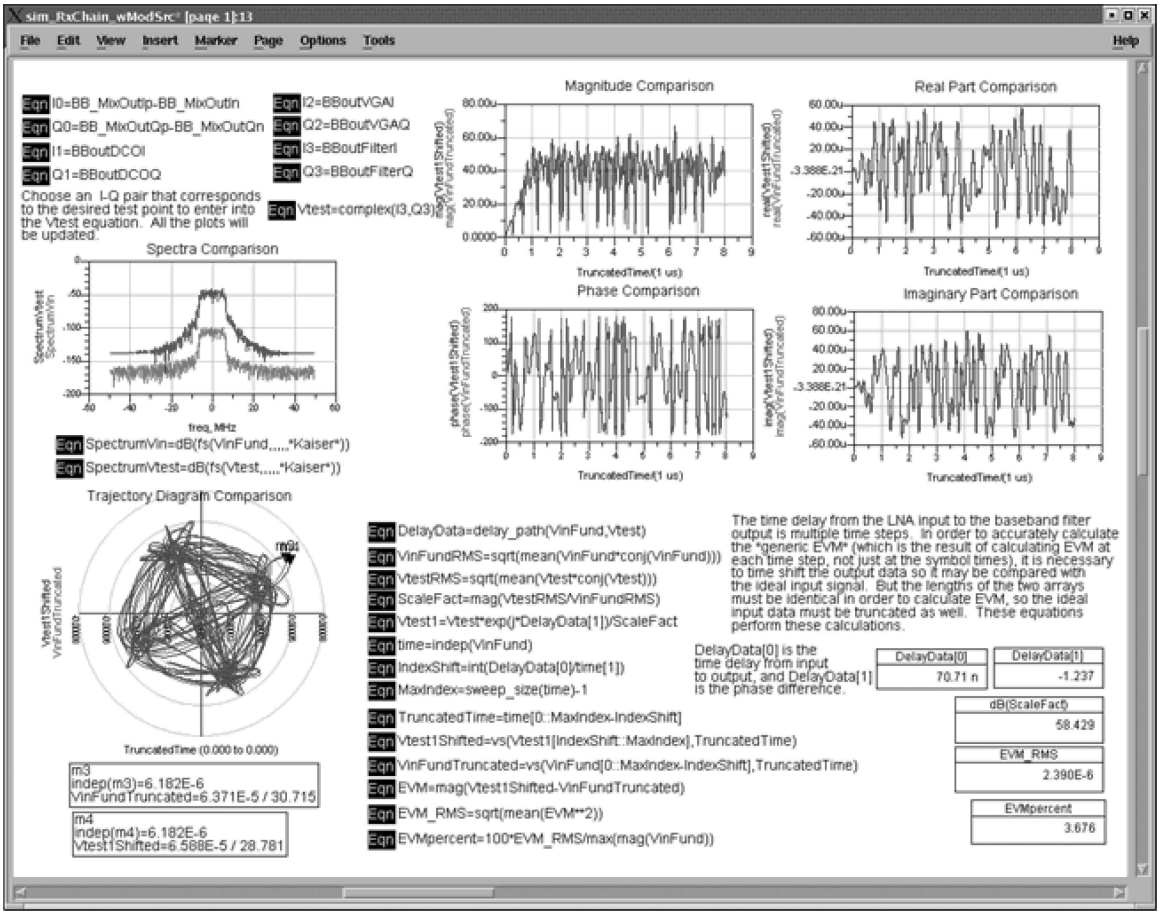


FIG. 9-11. Data display of a simulated WLAN 802.11b signal obtained using Agilent's RFDE software design platform. Note that the data display also calculates the "generic" EVM, which is the result of including every time point (not just the symbol samples) in the EVM calculation. This data may also be displayed as part of the Cadence simulation environment.

By changing the Vtest equation, you can see if there is a block in the receiver chain that significantly degrades the EVM. Also, you can make changes to various blocks to see if the EVM can be improved.

A good strategy for minimizing BER is to first run relatively quick simulations like this to test the EVM. When the EVM is minimized, the BER should be minimized as well, and these EVM simulations are much faster than simulating BER. Using this type of simulation, you can easily see whether your receiver or transmitter is wired up correctly and which block is causing degradation.

SIMULATION EXAMPLE 2

The high-frequency nature of RFIC circuits makes them especially vulnerable to circuit impairments like compression, noise, distortion and phase noise, as well as physical parasitics that include interconnect impedance and coupling. In order to handle these issues properly, you will require a simulator with the ability to simulate large circuits with extracted parasitics and high nonlinear designs. It should be capable of simulating signals at whatever frequency you want and running EM simulations on arbitrary layout structures, which is more accurate than using analytical models. It should also be able to run simulations that include post-layout extractions. Any limitation on the type, range, or capacity of simulation tools can impose major limits on accuracy, flexibility and design exploration.

As an example of the flexibility required for RFIC simulation, consider Fig. 9-12. The graphic depicts the spectrum for a direct downconversion of either a direct-conversion receiver or transmitter. Two input signals are shown near the local oscillator (LO) frequency, with downconversion to baseband terms at the output. The simulation can be run using either the Agilent Harmonic Balance or Circuit Envelope simulators available from either Cadence or Agilent via the Cadence SpectreRF or Agilent Advanced Design System (ADS) solutions. Both Cadence and Agilent provide these solutions.

Using the Harmonic Balance solution, this simulation would require three large-signal tones, one for the LO and one for each RF signal. A two-tone simulation can also be used if the frequency difference between Flo and Frf1 is an integer multiple of the frequency difference between Frf1 and Frf2.

With Circuit Envelope, only one large signal tone, for the LO, is required. Circuit Envelope simulation is a hybrid time- and frequency-domain simulation technique. Signals that are within the envelope bandwidth that is centered on each large-signal analysis tone are generated without requiring any additional large-signal analysis tones. The envelope bandwidth is equal to 1/(simulation time step). Tones that are within 0.5/(simulation time step) above 0 Hz are also generated.

The large-signal analysis frequencies in this Circuit Envelope simulation are the LO frequency and its harmonics. The LO signal can be generated by an oscillator and/or a frequency divider, in which case its frequency is determined automatically by the simulator while solving the oscillator and frequency divider.

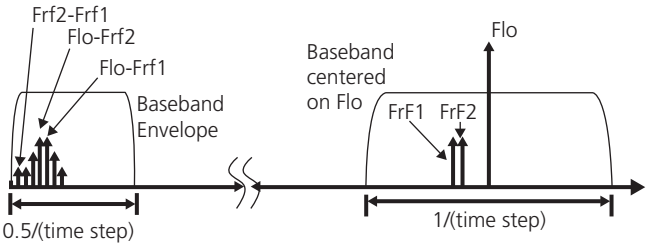


FIG. 9-12. Spectrum for a direct downconversion.

Fig. 9-13 displays the Circuit Envelope simulation results. The equations show how a particular spectral tone can be plotted as a function of the input signal amplitude. The design variables Flo, Frf1, and Frf2 specify the spectral tone. The input signal amplitude is VRFamp_dB.

Note that, in this example, the LO and its harmonics had to be simulated along with baseband signals near DC due to the large difference in signal frequencies. While a time-domain simulator cannot simulate such circuits efficiently, frequency-domain simulators like Harmonic Balance and Circuit Envelope are well suited for circuits of this type.

MODELING

As opposed to system-level behavioral modeling, which describes what the system does, modeling of active and passive on-chip elements during circuit design is done to improve the

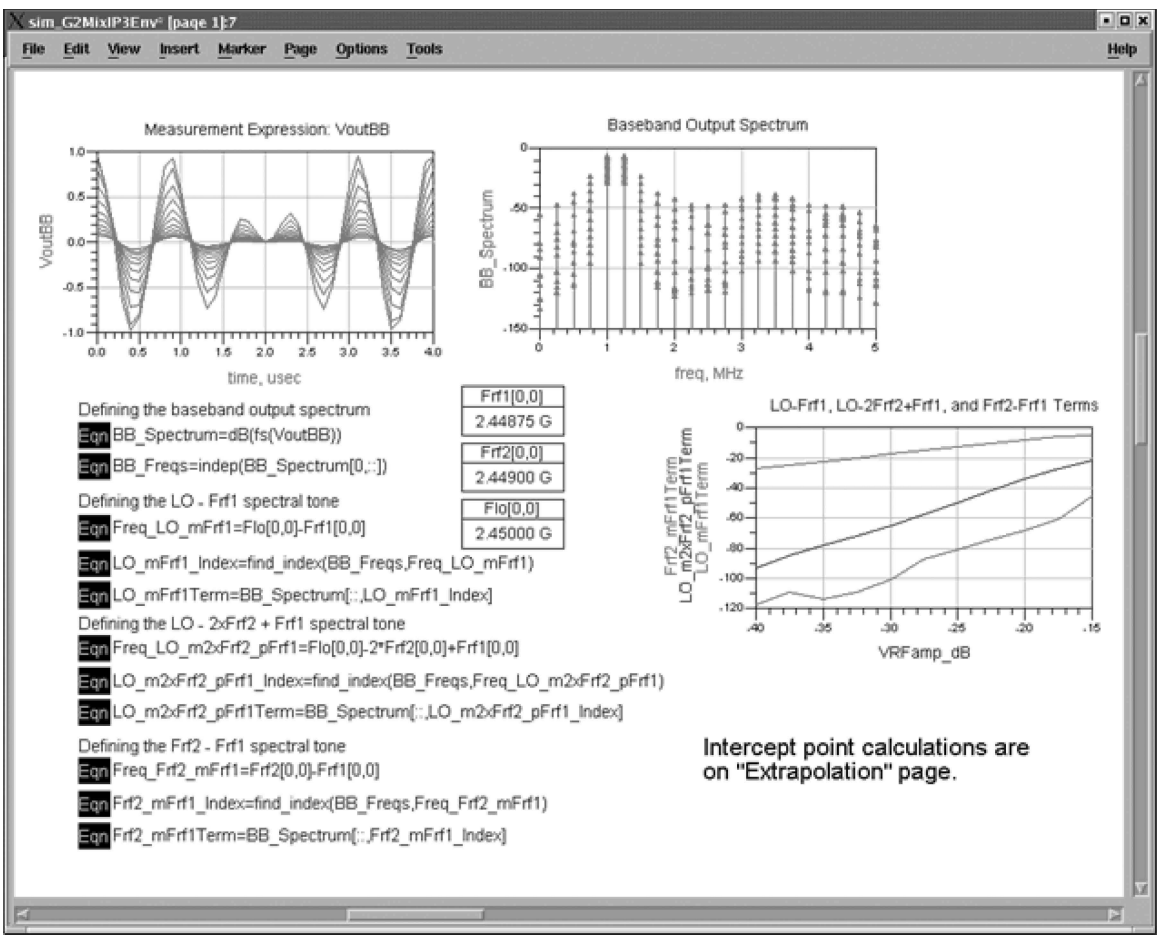


FIG. 9-13. Shown here are the results of simulating a downconverting mixer using the Circuit Envelope simulator as part of Agilent's Advanced Design System (ADS)/RFDE. The RF tones are at 2.44875 and 2.44900 GHz, while the LO is at 2.45 GHz. The simulation required about 4.5 minutes.

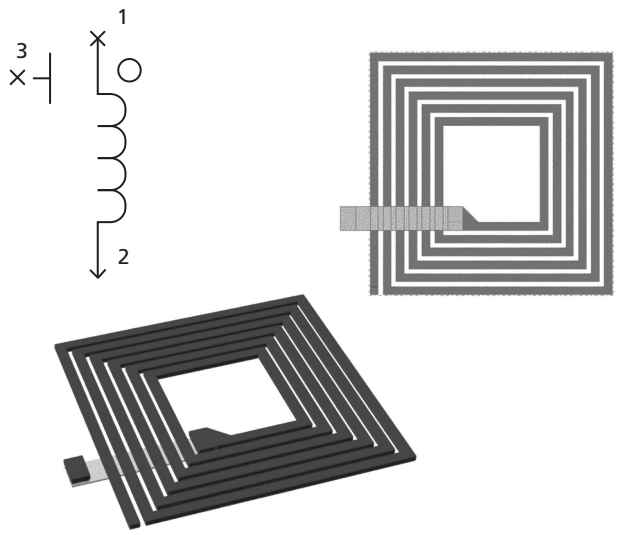


FIG. 9-14. Spiral filters are complicated structures, with non-zero thickness lines, underpasses, and more than one spiral metal layer possible. Graphic courtesy of Applied Wave Research.

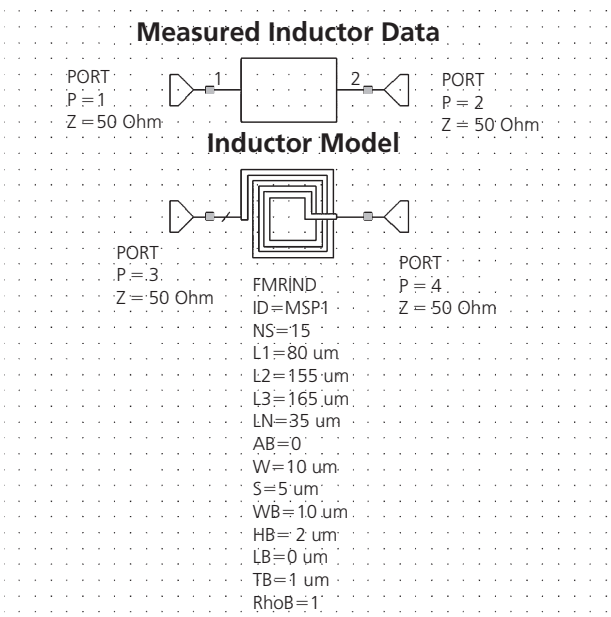


FIG. 9-15. Inductor models can be quite sophisticated. This model for a single layer spiral has thirteen physical parameters in order to match the experimental data accurately. Graphic courtesy of Applied Wave Research.

fidelity of the simulation. Especially problematic is the process of modeling spiral inductors on RFIC circuits, which are critical for filtering and tuning purposes (Fig. 9-14). When modeling these elements it is vital that the Q of the inductor and the resonance be predicted accurately. Because of their complex structures, this is often an extremely difficult, nontrivial task (Fig. 9-15). Note that the Cadence SpectreRF Simulator includes a Virtuoso Passive Component Modeler (VPCM) capability which enables synthesis, verification and modeling of spiral inductors and transformers. It is a complete flow, integrated in the Virtuoso Schematic editor and layout suite, and is compatible with Assura. It produces a complete PDK component with schematic, symbol, pcell layout, S-parameter file and Spectre models.

Note that a Foundry Design Kit (FDK) will provide many of the passive models (e.g., metal-insulator-metal or metal-oxide-metal capacitors) needed for designing RFICs. The RF Design Methodology Kit from Cadence can be used to help generate a design's layout and an electrical model. As long as design rules are followed and parameter ranges are not exceeded, these models are highly accurate. Because most foundries provide only corner-case device models for device modeling, these models may not be well suited for RF designs. Ideally, a model library should include the following: corner cases, statistical models, digital/analog mismatch models, pad models with RF electrostatic discharge, flicker-noise models, and substrate-resistance models. The library also should include well-proximity and shallow trench isolation stress effects.

Modeling Issues

If a suitable FDK is not available, the task of creating the model falls to the engineer. Some of the issues that will need to be considered when developing a spiral inductor model are:

- The model must have some form of frequency-dependent resistance to compensate for conductor losses.
  Normally, spiral inductors are placed on the top layer metal. The metal thickness and width are typically on the order of a few microns in each dimension. At GHz frequencies, these dimensions are on the order of a skin depth for aluminum or copper, forcing the current'distribution in the conductors to change (e.g., crowding toward the surface of the line). This current crowding increases the resistance in the inductor and decreases the Q.
- Inductance effects must be considered.
  Current crowding affects the inductance as a function of frequency. As it occurs at higher frequencies, the internal inductance will go down, as the current and flux are being excluded from the interior of the lines. Therefore, there is a frequency dependent inductance effect, which also affects the Q.
- Shielding can be used to isolate the inductor from the substrate and minimize substrate effects.
  The lossy substrate can contribute substantially to reduction of the Q by providing a resistive loss mechanism. Shielding placed below the inductor to isolate it from the substrate can mitigate the substrate's effect on Q.

- Make an assumption regarding location of the ground return.
  Inductance is only uniquely defined when there is a loop of current. Unfortunately, the modeler rarely has the luxury of knowing exactly where the return current is going to be. Therefore, certain assumptions will have to be made regarding where the return current is, e.g., on the substrate or on an intervening power plane.
- Capacitance can be difficult to predict.
  A multi-turn spiral has a large amount of capacitive coupling to itself, which does not change much with frequency. Charge will stay on the surface of a "good" conductor. Aluminum, for example, is a good conductor well past 100 GHz. However, it is often difficult to predict the capacitance, as the lines are three-dimensional structures. This is where experience and intuition can be important.
- Be careful when using the model in SPICE.
  Resistance and inductance will change with frequency. As a result, it may be tempting to make a model with frequency-dependent resistance and inductance. When the model is exported to SPICE, however, undesirable behavior can occur. The response can become completely nonphysical.
- Include underpass effect in the model.
  The signal must be brought out from the center port of the spiral inductor. Typically this is accomplished by connecting to a line on a lower metal layer, creating an underpass. It is important that the effect of the underpass be included in the model.

Electromagnetic simulation software can be used to obtain simulated data for inductor modeling and design. Several products are commercially available that use a variety of methods. They range from general-purpose simulators to those that are specifically customized for spiral inductor modeling in RFICs.

PCB DESIGN

The printed circuit board (PCB) is used to mechanically support and electrically connect electronic components using conductive pathways, or traces, that are etched from copper sheets laminated onto a nonconductive substrate. In high-volume production, the PCB is inexpensive, rugged and reliable.

The Flow

A typical PCB functional design flow is illustrated in Fig. 9-16. In the system specification phase all design-specific requirements are gathered and a specification is created. Next, a schematic representation of the specification is created or captured, using a range of schematic capture tools. Components for the schematic are chosen by the engineer based on a wide range of criteria, which might include such things as price, support and availability. A netlist is then automatically generated, which will serve as input to the PCB Layout tool.

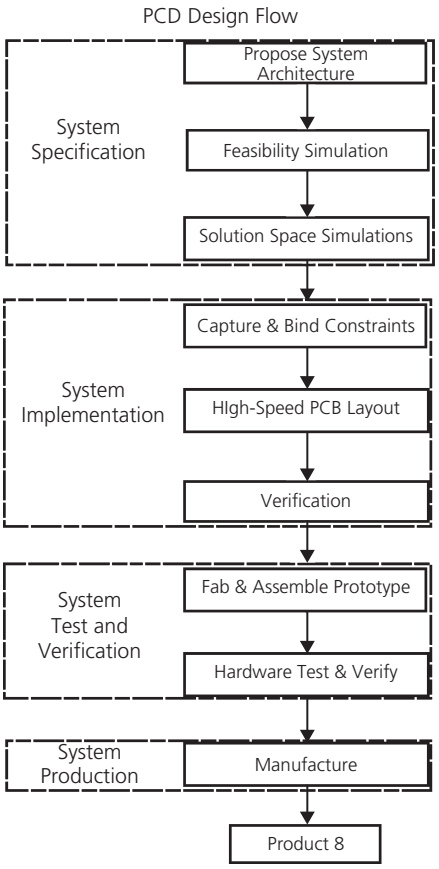


FIG. 9-16. Typical PCB design flow.

Schematic capture tools allow you to take components from a component library and place them on the schematic. You will need to edit them as required to ensure that the component package details and silkscreen legend to be printed on the PCB are as per requirements. If a component is not in the library, the component is defined and its definition and details are added to the component library. Existing library components may also need to be further edited to reflect the actual circuit diagram, a step which requires that the pins of a particular component have to be moved around as per desired circuit schematics. Once all the parts have been placed on the schematics, the required connections between pins of various components are drawn.

Next comes the PCB layout phase in which the schematic representation is laid out via the use of a PCB layout design tool. Considerations during this design phase include such things as design for manufacturability (DFM), signal integrity and EMI/EMC. These considerations will generally impose certain restrictions on the PCB layout design and should be included with your basic design rules (e.g., the rules that you follow for layout pertaining to things like track width, track spacing, pad sizes, via sizes, and routing types). Verification that the layout will provide the performance and function documented in the specification is then necessary.

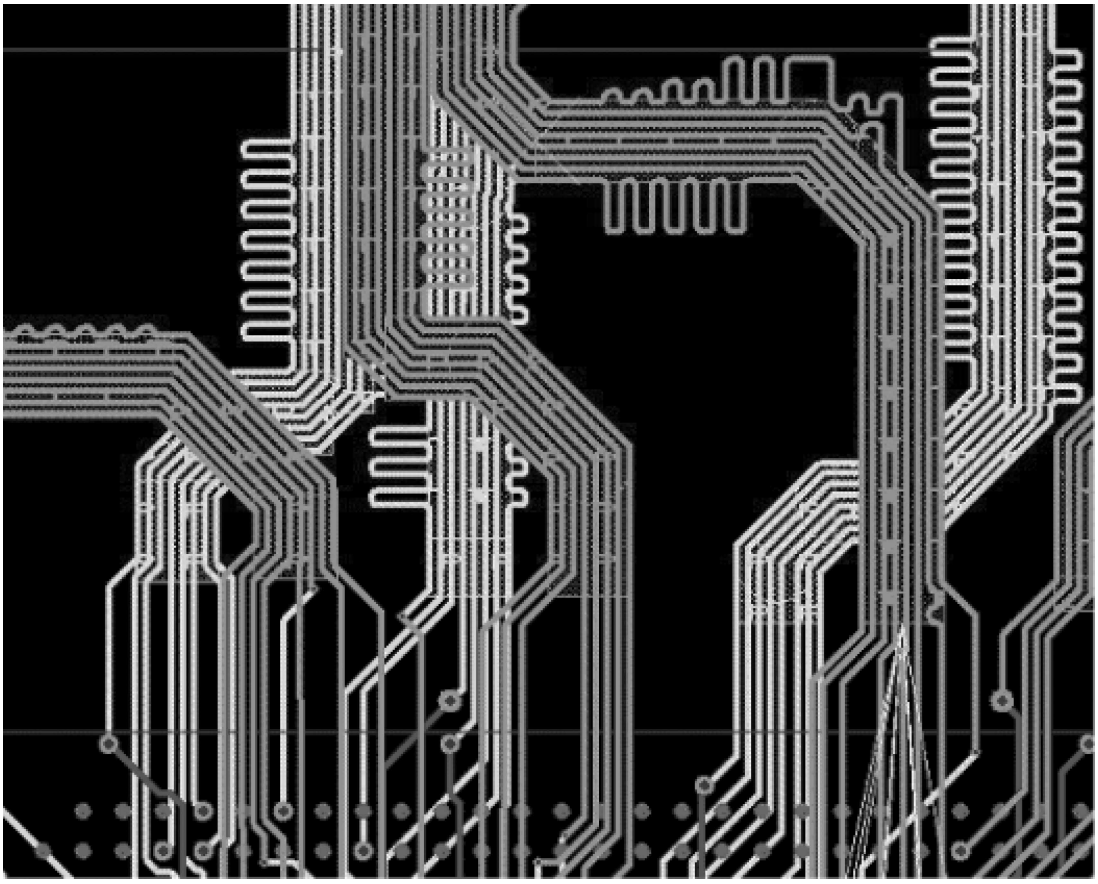


FIG. 9-17. Modern auto routers, such as the RE AutoRouter option to Mentor Graphics' Board Station RE PCB placement-and-routing environment, mimic the manual strategies used by a designer.

During layout, components that form the schematic are taken from the parts library or libraries and placed on the PCB layer. This process is aided by the use of PCB placement and routing (P&R) tools. All component placement constraints are taken into account. For example, heat dissipation considerations may dictate that some parts should be a certain minimum distance away from others, while signals that need to be limited in length may dictate close placement of certain parts that involve those signals.

Following placement, routing takes place. Traditionally, this process would have involved a combination of both manual and automatic routing techniques. Today, though, with 80–90% or more of signals considered critical it is no longer feasible to manually route critical nets. Instead, today's auto routers are able to route "correct by design" signals, following all the signal constraints for each signal. Modern auto routers also route signals in a manufacturing-optimized manner so that manual cleanup is not necessary (Fig. 9-17). While some engineers may choose to spend time manually cleaning up their designs, this process is just not "do-able" for very large designs, as it would take several weeks.

After routing, the engineer must perform a DRC check, whereby the PCB layout is tested against the netlist and the set of previously specified design rules. Any violations reported are corrected. After DRC, a final layout cleanup is done.

Note that radio transmitters and radio receivers are especially difficult to design. PCB designers must therefore minimize parasitic effects due to layout of components, or take them into account with a general model and use simulation software such as SPICE. Fortunately, many practical circuits can be laid out using a much simpler lumped element model.

At this stage, a final verification is done to ensure that the design will perform as expected. If any error is found then the PCB design process must be iterated (e.g., from design specification, layout, etc.) until such a time as the prototype passes verification. Once the design has been fully verified, it can be produced on a volume scale for delivery to the customer.

If desired, the engineer can create a prototype for the purpose of final verification, using the design files generated during PCB layout (e.g., ODB++ formatted manufacturing data). More often than not, though, it has become common to skip the prototype altogether. This process can be very time-consuming. More importantly, today's verification tools now make it possible to get the design right the first time, without having to build a prototype.

PCB Design Tools

At every stage throughout this process, design tools can be used to assist the RF engineer. These tools play a special role in ensuring design success, especially given the high-frequency circuit impairments in today's complex analog and RFICs, such as compression, noise, distortion and phase noise, as well as the physical parasitics like interconnect impedance and coupling, which make achieving design closure between RFIC's system and circuit, electrical and physical, and design and test activities difficult at best.

To aid RF designers, design tool vendors now offer not just point tools, but PCB design flows that are unified with both the IC and package domains. This allows the RF engineer to perform early analysis, I/O buffer planning, optimization and implementation across all three implementation domains.

PACKAGING

A package is used to house the component, system or subsystem being designed. RF packages are substantially different from digital IC packages. They operate at much higher frequencies and usually have far fewer I/O's. Consequently, greater care must be taken when doing the layout of individual signal interconnects and transmission line junctions.

Additionally, in the GHz range, the effects of the physical properties of interconnect on electrical signals (e.g., dispersion, radiation losses, resonance and skin effects and even the surface roughness of a conductor) can have a significant impact on RF performance. In fact, the electrical parasitics arising from the packaged product have become such a key problem for chip designers that it has now brought packaging into the spotlight as a system-level concern. These effects, therefore, must be carefully considered when designing RF packaging. Electromagnetic modeling and three-dimensional EM models often play a critical role in the RF package design process.

Options

There are a variety of packaging options and materials to choose from. Which one you choose will depend in part on your end application. For example, if the RF system in question is for a wireless mobile product, then you will need to trade off your packaging choices against customer-driven factors like cost, performance and size.

Packaging options range from standard single-die packages to System-in-Package (SiP). SiP technology builds on the innovative array interconnect of ball grid array (BGA). It allows multiple die with complementary device technologies to be combined in a single package. Passive devices may also be included in the package to deliver highly functional integration for digital and radio frequency applications.

Another packaging option, RF Stacked Die System-in-Package, is typically used in designs where the X-Y size constraint is the critical requirement, such as for wireless communication applications. It allows for the combination of die (e.g., 2 to 4 separate die) from different fabrication processes into a single package. Board area savings are realized by stacking the die vertically versus a side-by-side approach.

RF SiP is supported by design tools like the RF SiP Methodology Kit from Cadence Design Systems. The Kit accelerates the application of advanced EDA technologies to SiP designs for RF/wireless applications. It provides methodologies that maximize design productivity and predictability for customers leveraging the advantages of SiP implementation. An integrated set of SiP design products, built around proven methodologies, enables complete front-to-back SiP design and implementation, which you can leverage for techniques and methods to apply to your own design. The Kit is comprised of a complete documented step-by-step methodology and flow for designing a real, multi chip IC package. It includes all the design data, flow guides, and tutorials to jumpstart customers on a multi-chip RF SiP Module design methodology, including package layout. This methodology is also integrated with Cadence's Virtuoso environment.

RF packages can be designed in a wide range of materials (e.g., ceramics or advanced low dielectric materials). Two of the more common options in use are ceramics and laminates.

Ceramics

Ceramics are inherently robust and allow for hermetic packaging—a feature that makes them especially useful in applications requiring high reliability over a long period of time or for those that will be used in hostile environments such as in space or on the battlefield. Because there is such a wide range of ceramics technologies to choose from, the use of this material in RF packaging is quite common today.

Fabrication process options include thin film or thick film. Thin film technology offers the smallest feature size and the highest resolution. In contrast, thick film technology allows the printing of resistors and inductors directly onto the circuit. This technology is particularly useful in RF packaging where normal lumped components can be difficult to apply due to their inherent parasitic properties.

Low temperature co-fired ceramics (LTCC) is a multilayer ceramic substrate technology. It is especially well suited for RF applications and products where a high integration level and/or high reliability is needed.

Ceramic material options range from a standard material like high temperature co-fired ceramic (HTCC) alumina to more specialized options, like aluminum nitride with its excellent thermal properties or HiTCE ceramics, which have thermal expansion coefficients that are better matched to the expansion coefficient of PCB materials. Ceramic materials are also available with a variety of electrical properties. Lower dielectric constant ceramics, for example, allow for larger feature sizes and reduced propagation delays.

Laminates

A laminate is a material that is made up of bonded layers. A hybrid laminate is a PCB like structure, normally with a FR4

material core, that uses advanced materials on selected layers in order to meet rigorous RF performance requirements. These types of package substrates are particularly useful for RF systems where signals in the GHz frequency range may be routed on the substrate surface between various active circuit elements. They have become increasingly important in RF packaging due to their low cost.

Design Solutions

Given the possible trade-offs and packaging options RF engineers now have at their disposal, it is clear to see that designing an RF package requires a broad knowledge of material properties and assembly processes, as well as a thorough understanding of high-frequency signal behavior. Making the right choice can either make or break the success of your product. Therefore, the choice of packaging should never be an afterthought. You must implement a packaging strategy during the design phase in order to ensure that you will meet all your requirements for reliability, manufacturability, RF performance, size and cost.

Various EDA tools are now available to help you design an RF package. In addition to the EM modeling tools previously mentioned, there are also advanced packaging kits, such as the RF SiP Methodology Kit from Cadence Design Systems, which offer guidance on different design techniques (e.g., for SiP). They cover RF components, together with packaging strategies, to help you meet requirements for reliability, manufacturability, RF performance, size and cost. Kits also explain design rules for different types of packaging such as laminate and LTCC modules, as well as the trade-offs between them, and present working guidelines for IC partitioning decisions early in the design phase.

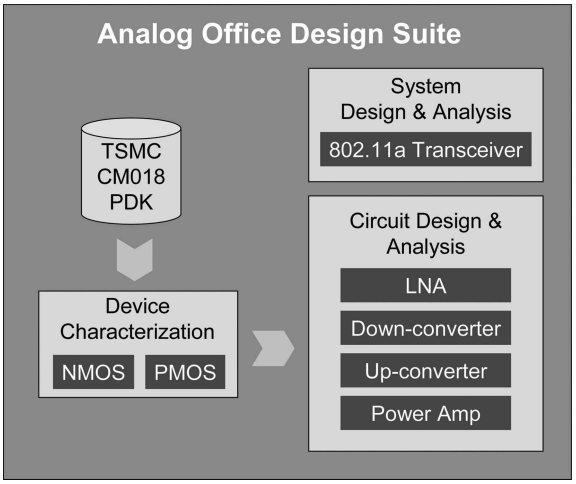


FIG. 9-18. The 802.11a design flow within the Analog Office design environment from Applied Ware Research.

CASE STUDY

Now that you have a better understanding of EDA tools, it is time to take a closer look at how they can be utilized to streamline the design process in the real world. The following case study examines the design and analysis of an IEEE 802.11a (wireless local area network) RF complementary metal-oxide semiconductor (CMOS) transceiver using the Analog Office software solution and design flow illustrated in Fig. 9-18. The design employs a conventional configuration widely used in the silicon RFIC community and 0.18-μm CMOS technology from Taiwan Semiconductor Manufacturing Company. Note that other design tools—such as the Genesys software environment from Agilent Technologies—may also be used for this design process.

System-Level Transceiver Design

The system-level diagram of the 802.11a transceiver, shown in Fig. 9-19, is captured using pre-built behavioral blocks in the Visual System Simulator (VSS) software. The design employs a

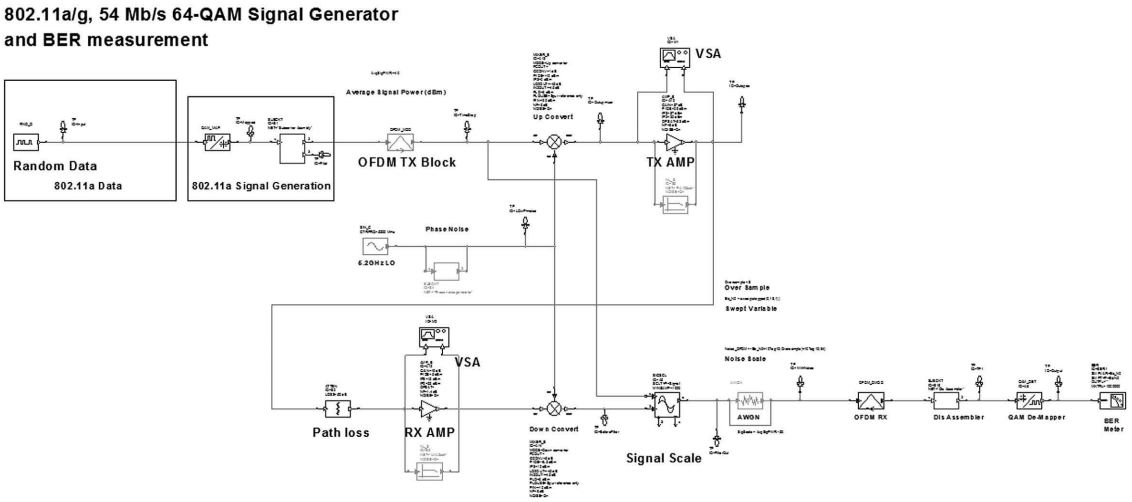


FIG. 9-19. The 802.11a transceiver system-level diagram with behavioral blocks representing major functions of the entire RF transceiver chain, including the channel characteristics.

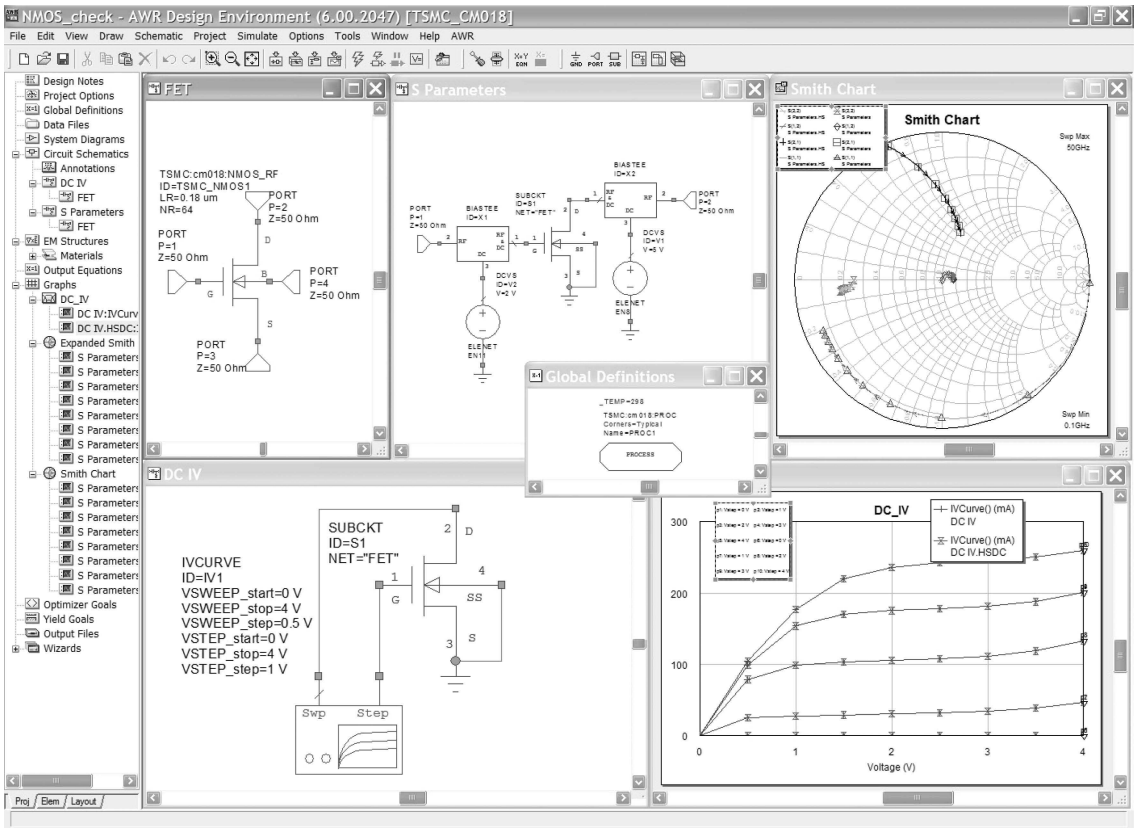


FIG. 9-20. The NMOS transistor is characterized before use in the circuit to ensure accurate and adequate performance.

homodyne architecture, i.e., direct conversion of the RF signal in 5.15 to 5.825 GHz band to baseband. The transceiver consists of a baseband and RF circuits. The transmitter baseband circuits include a 64 quadrature amplitude modulation (QAM) modulator and orthogonal frequency division multiplexing (OFDM) modulator. Receiver baseband circuits include a 64 QAM demodulator and OFDM demodulator. The transmitter RF circuits are comprised of up-converter mixers and power amplifiers, while the receiver RF circuits include up-converter mixers and low-noise receiver amplifiers. The local oscillator frequency is set to RF frequency as per the homodyne architecture.

The goal of the system-level design is to determine RFIC specifications from system-level specifications. During system simulation, the noise figure and nonlinearity of the power amplifier (PA), LNA, and up- and down-converter mixers and phase noise of the local oscillator (LO), are varied and system packet error rate (PER) is monitored against specifications.

Prior to the start of the circuit design phase, the TSMC 0.18-μm RF CMOS process is selected. A process design kit (PDK) is then developed and installed within the Analog Office design environment.

Circuit-Level Receiver Design

The homodyne receiver portion of the IEEE 802.11a transceiver consists of two main RF circuits: an LNA and a down-converter. A modulated RF signal in the 5.150 to 5.825 GHz band is the input to the LNA. The LNA must provide a sufficient gain to minimize the impact of mixer noise on the overall noise figure of the receiver. A single-ended, 50Ω input source is assumed for the LNA, which amplifies the input signal by 20 dB, with a noise figure of 1.2 dB.

The down-converter directly converts the amplified RF signal to DC. It consists of two mixers with LO signals offset by 90° between the I and Q mixers. The mixer has a conversion gain of 6 dB, and noise figure of 5 dB.

LNA Design

The LNA selected is a differential cascode common source amplifier. Note that in this example, noise figure, power consumption, input impedance, gain, and linearity are considered together in designing the LNA.

Device Characterization

The design optimizes the noise figure of the LNA, while taking power consumption into account. The optimum device width is about 160 μm (NR = 64). DC and AC characteristics of a device with a width of 160 μm are simulated, as shown in Fig. 9-20. The DC simulation displays the drain current (I_{ds}) with various bias settings of drain to source voltage (V_{ds}), and gate to source voltage (V_{gs}) of the negative-channel metal-oxide semiconductor

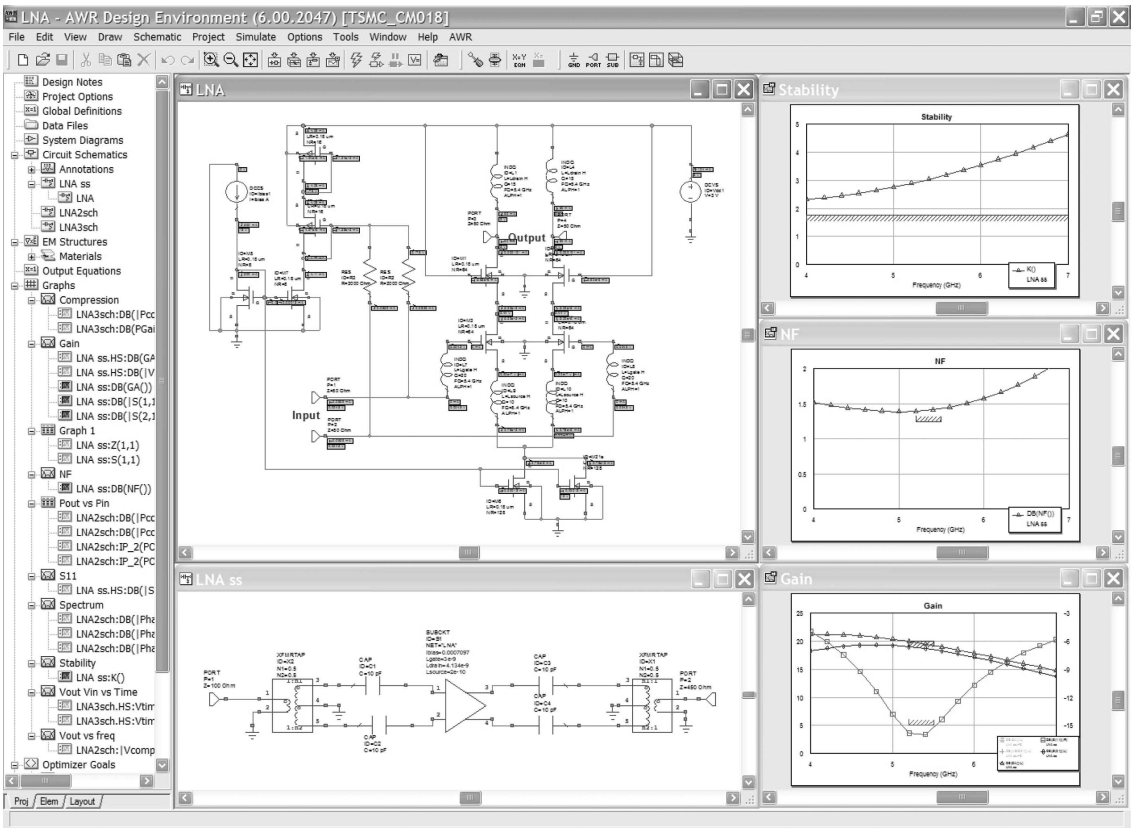


FIG. 9-21. LNA circuit design and analysis results including gain, noise figure, and stability.

(NMOS) device. A set-up for measuring I_{ds} vs. V_{ds} for various settings of V_{gs} is also shown. A set-up for measuring I_{ds} vs. V_{gs} can be used to extract g_m.

AC simulation displays f_T and Y- and S-parameters. A set-up for measuring a gain of a transistor versus frequency can be used to observe f_T of the transistor at a particular bias setting as the gain goes to 1. For the bias setting of interest, the f_T of the device is about 44 GHz. It is also used to monitor the Y-parameters. A set-up for measuring S-parameters of the transistor at a particular bias setting is shown.

Circuit Design

Once the device characteristics are simulated, a bias current of 14 mA is selected for the differential pair, as shown in Fig. 9-21. An estimated noise figure within the power constraint is about 1.5 to 1.9 dB. A noise simulation shows a noise figure of 1.2 dB in the frequency range of 4 to 6 GHz and 1.1 dB at 5.2 GHz.

Next, input impedance is matched to 100Ω differential impedance. Since input impedance of an NMOS device is mostly capacitive, a real term of the impedance must be provided. In order to minimize the noise, an inductive source degeneration method is used to present a real term in the frequency of interest. The capacitive term of input impedance is "resonated out" by a series inductor at the input. A high Q of the series inductor is desirable to minimize the noise figure. S11 is −20 to −30 dB in the frequency of interest.

Note that output impedance is not matched to 100Ω differential because the LNA is intended to be integrated with an I/Q demodulator. An inductor is used as a load at the drain to provide more headroom, to resonate out the capacitance seen at the output load, and to provide sufficient gain. The Q of the load inductors at the drain must be low enough to cover the entire band of 5.15 to 5.85 GHz. The gain of the LNA is 20 dB over the band of interest.

A stability factor, K, of the LNA is greater than 1 to ensure the stability over the band of interest and other frequencies. To prevent a negative resistance and thus avoid instability, it is important that the differential pair not see a capacitance at the source. In order to take advantage of automatic measurements of impedance, S-parameter, gain, noise, 1-dB compression point (and IP3), and intermodulation components, a balun is used to convert the differential inputs of the LNA input single-ended.

The LNA circuit is duplicated with differential input signal sources for different simulations. One circuit is used to simulate gain, noise, S11, and Z_{in}, and stability factor K. Another circuit with two tone frequencies and a power swept at the input port is used to simulate input IP3 and 1-dB compression point. The 1-dB compression point is −10 dBm and IP3 is >10 dBm.

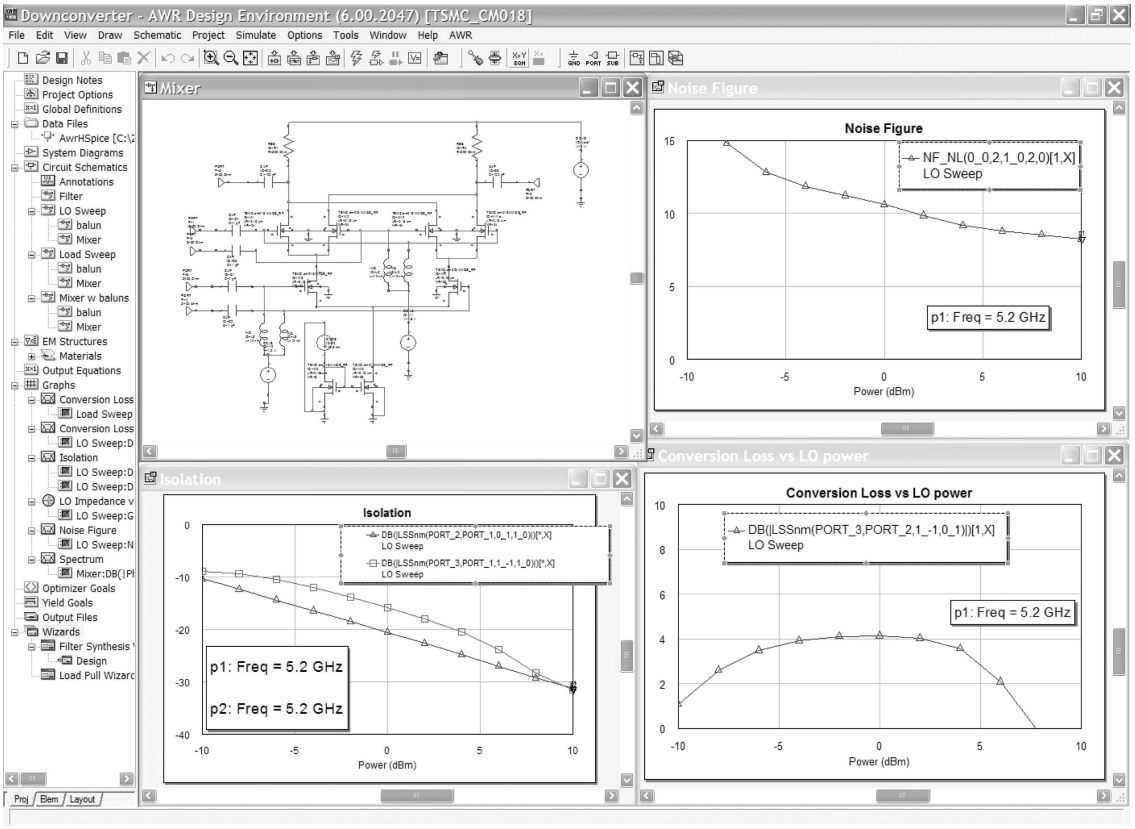


FIG. 9-22. The down-converter mixer design circuit schematic with bias voltages annotated and several simulation results including isolation, conversion loss vs. LO power and noise figure.

Another circuit with a single tone frequency and a power at the input port is used to simulate the input and output voltages in time domain.

Down-Converter Circuit Design

A down-converter consists of two mixers, converting from RF to DC. Fig. 9-22 shows an example of an RF mixer design using a Gilbert cell configuration. One mixer circuit is used for the simulation. The simulation shows a noise figure of only 5 dB, so it may not be including all the noise sources. The bias and the load resistor determine the conversion gain. The conversion gain is about 6 dB as shown in S21 of a large signal S-parameter simulation and noise gain simulation.

The source degeneration inductors at the source of the RF differential pair improve the linearity of the mixer. The 1-dB compression point is about −10 dBm of input power at the mixer RF port.

The isolation of LO-IF ports and LO-RF ports is simulated by mismatching the RF driver differential pair and LO switching pairs. The LO signal is monitored at the IF and RF ports separately, using a large signal S-parameter (LSS) NM measurement.

Another identical mixer circuit with a single RF input power level is used to simulate the intermodulation components of the mixer.

Transmitter Circuit Design

The direct conversion transmitter consists of two main RF circuits: an up-converter and a PA. The PA in the design has an output power of 40 mW and outputs a modulated RF signal in the 5.18 to 5.26 GHz band.

Up-Converter Design

The up-converter consists of two double-balanced active mixers for I and Q modulation. The lower differential pair takes baseband signals and the upper switching pairs take differential local oscillator signals tuned to the desired RF output signal. The local oscillator frequency should be varied from 5.18 to 5.26 GHz. For simulation purposes, the input center frequency is 50 MHz. If the outputs of I and Q paths are summed, the lower sideband will be selected as the output. If the outputs of I and Q paths are subtracted, the upper sideband will be selected as the output. The linearity requirement is set high to ensure no degradation of linearity in the overall transmitter for OFDM and 64 QAM modulation. The 1-dB compression point is set 10 dB higher than

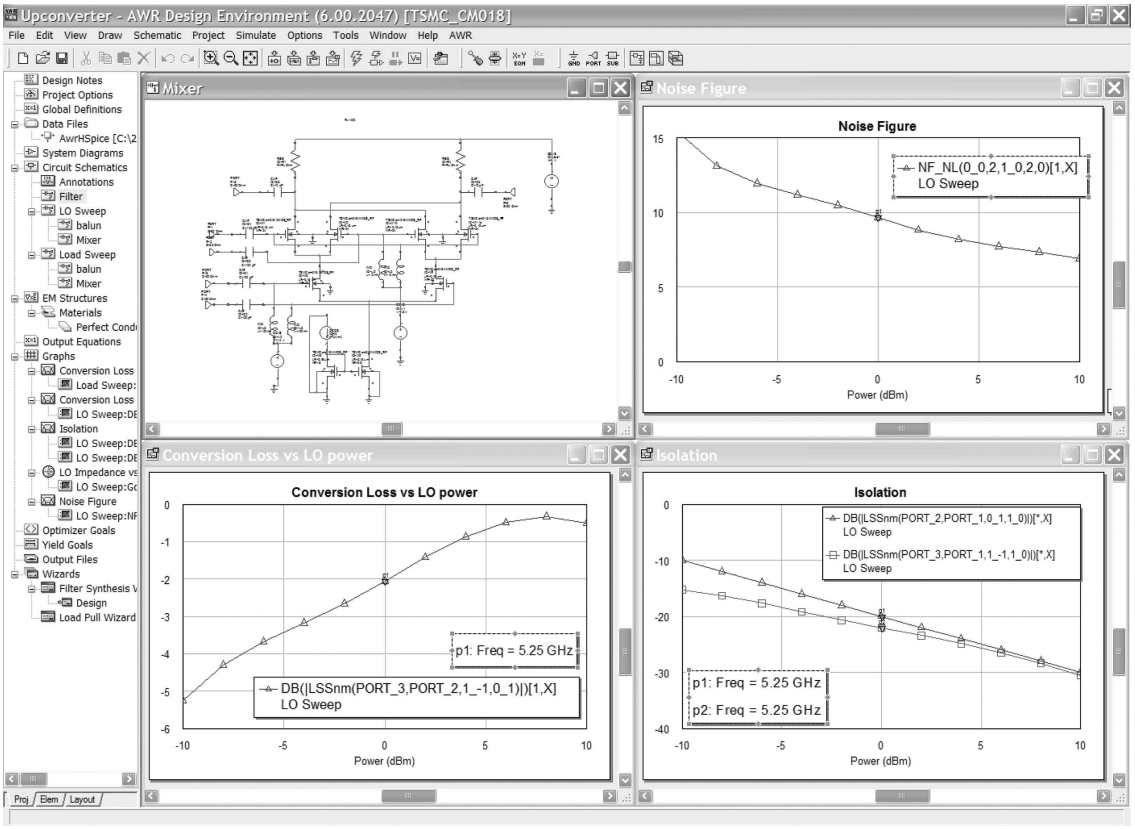


FIG. 9-23. Up-converter mixer design and analysis results, including noise figure, conversion loss vs. LO power, and isolation.

the operating power level. For instance, the input 1-dB compression point is 8 dBm and the input power level in use is −2 dBm or less. An RF system simulation using VSS software with a behavioral model for up-converter showed 8 dB back-off from the 1-dB compression point to guarantee no BER degradation, assuming 4 to 5 dB back-off from the 1-dB compression point for PA operation.

Mixer Design

The bias and size of the transistors are selected to provide the necessary linearity and gain. The mixer has a conversion loss of about 5 dB, as shown by the LSS_{21} measurement if the differential LO power level is 10 dBm. The noise figure is about 1.5 dB, which means it is not simulating correctly. The large signal impedance is monitored at the LO port for matching purposes.

It is important to provide a proper isolation between the LO and the output port. The LSS measurement monitors the LO signal level at the output port. The 1-dB compression point at the input is 8 dBm as shown by sweeping the baseband input power as the RF output is monitored. 8 dBm of a differential input power is referenced to a 100Ω input impedance load. Since the input of the up-converter is at baseband frequency, the input signal should have an option to sweep voltages rather than power. Two-tone simulations should show the intermodulation and harmonic components. A two-tone harmonic balance source with a fixed or swept power is applied to the input port while the intermodulation components are monitored at the output port. Small signal impedance and S-parameters as well as large signal impedance are monitored for the matching. Fig. 9-23 shows the various measurements mentioned above.

PA Design

The PA has a single-ended, common-source configuration with two gain stages. The target frequency band is 5.18–5.26 GHz. Although the average output power in this frequency band is 40 mW (16 dBm), OFDM and 64 QAM modulation signaling raises the required output peak power to 21 dBm. An RF system simulation using VSS software with a behavioral model for the PA also shows that 21 dBm is the required 1-dB compression point at the output. It is challenging to design a PA to deliver the necessary output power with a required linearity in CMOS technology at a low power supply voltage. Therefore, the bulk of the effort in this design is focused on exploring the design space to achieve optimum linearity.

PA Device Characterization

The output device size is selected by sweeping the current vs. bias and device size, given the trade-off between power gain, linearity and efficiency. Allowing for loss, the output power is designed for 22 to 23 dBm. Although the output resistance of

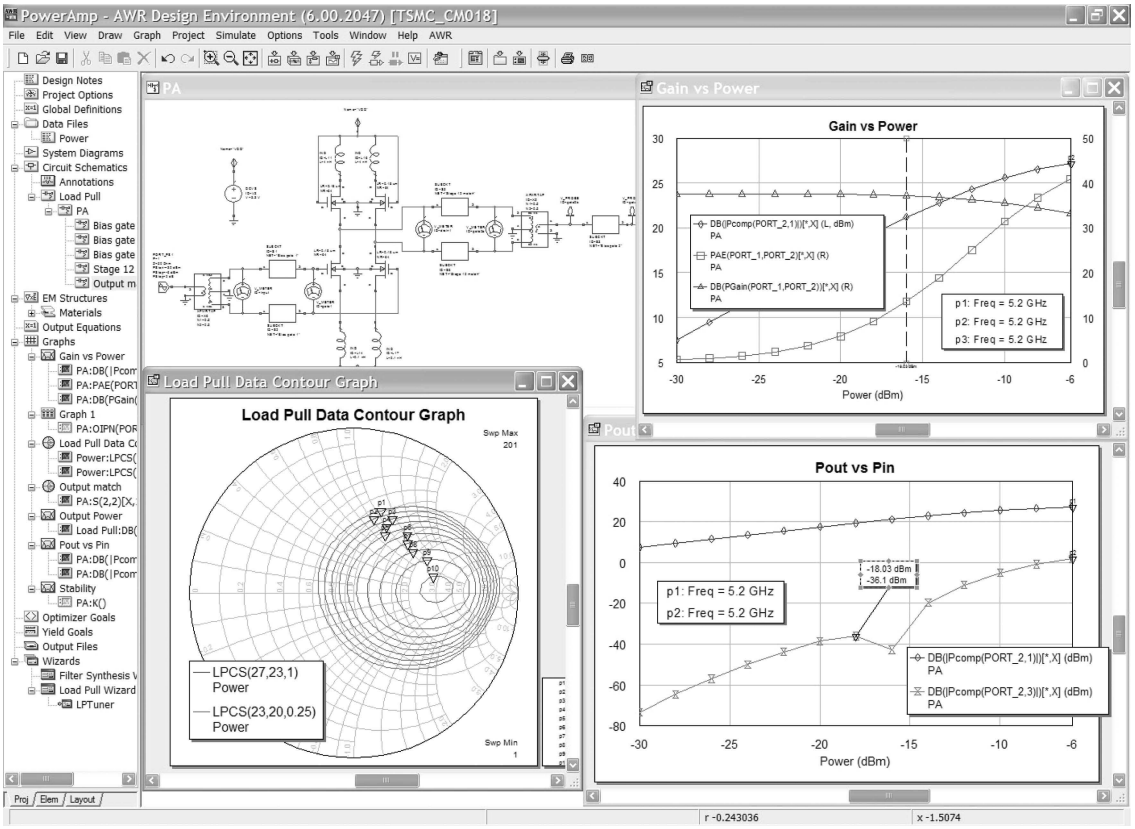


FIG. 9-24. The PA design and analysis results, including gain vs. power, pout vs. pin, and load pull data contour.

22Ω is required to deliver 23 dBm with a 3V supply voltage, a larger device of 960 μm/0.18 μm (thus smaller output resistance) is selected to achieve a better linearity and power gain. The gate is biased with a trade-off between the linearity and efficiency. The emphasis, however, is placed on the linearity, and therefore, it is biased as a class A amplifier (with $V_{gs} - V_{th} > 0.1$V). The device size of the first stage is 480 μm/0.18 μm.

PA Circuit Design

Given the output power and linearity requirement, the transistors are sized to provide the necessary load and gain. Each stage is first designed separately and then combined to resimulate the overall performance. Since a PA operates at large input and output levels, it is important to simulate it in a large signal domain. For instance, gain contours with an optimal load can be displayed in a Smith Chart. In Fig. 9-24, four gain contours with maximum incremental power gains of 20.8 dB and 0.1 dB are displayed along with S11 and S22. The input and output matching should maximize the gain. The contour plot shows that they are optimally matched. S11 and S22 are better than −25 dB over the frequency band of interest.

Prior to matching to 50Ω input and output ports, the circuit must be stabilized. Since the input impedance of the circuit is negative, a small inductor at the emitter to ground (simulating the bond wire connection) should help without degrading the noise figure. Additionally, a small resistor at the gate can provide stability over a wider band. The K value is greater than 2 at the frequency band of interest.

A proper matching of input and output delivers an optimal power gain. A matching network consists of a DC blocking capacitor to prevent the loading of circuits and an impedance transformation. A high pass L-match is used. The impedance at the input and output is transformed up to 50Ω. Since the input and output impedance changes with the power level, the matching should be monitored at a varying power level. The software displays large-signal impedance (Z-comp) as the input power level is swept.

The power-added efficiency is displayed by selecting "PAE". It is about 27% at the 1-dB compression point, as the PA is designed for high linearity.

Since the linearity is critical, it is important to simulate a 1-dB compression point and IM3 and IP3. The software simulates them by applying a single-tone or two-tone harmonic balance source. Two-tone harmonic balance simulation shows a large signal S21 (LSS21) of 21 dB with an input 1-dB compression point at 3 dBm (output at 22.6 dBm). The output IP3 is about 40 dBm and IM3 is about 33 dB down from the output power level of 18 dBm. The 1-dB compression point at the output of

22.6 dBm allows for better than 6 dB of peak-to-average ratio. The design target was 5 dB.

The resulting 802.11a RF CMOS transceiver is an optimum design. The LNA achieves adequate gain, low noise, and IP3, and input match and stability are also adequate. The mixer achieves adequate conversion gain and isolation, low noise figure, and IP3.

SUMMARY

EDA solutions for the RF industry are certainly far from being as mature and complete as those available to digital designers. Nonetheless, they can be used to more accurately and quickly design RFICs and PCBs. A number of design flows and tools now support these efforts, regardless of how simple or complex the design may be.

RF and Antennas

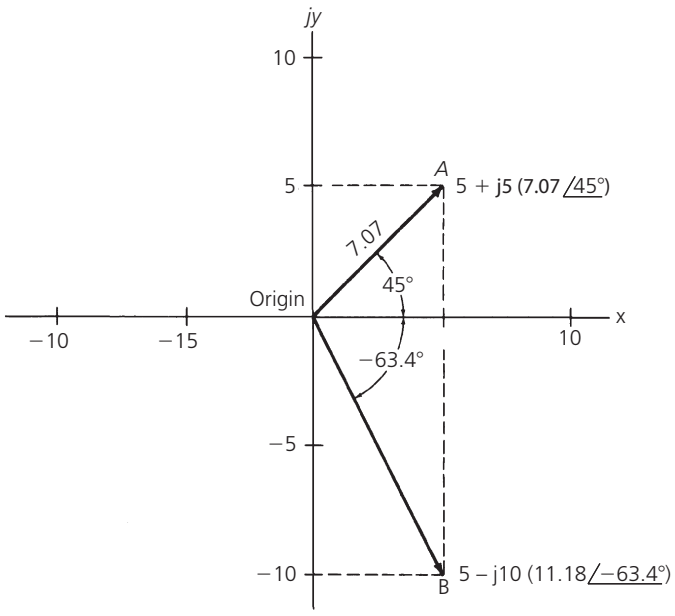
WHAT IS RF?

RF design is affected by residual circuit values, such as additional capacitances and inductances, exhibited even by transmission lines and circuit traces at high-enough frequencies. These residual or "parasitic" circuit elements can even be found in the structures of transistors used for RF and microwave frequencies.

At lower frequencies, the physical wavelengths of signals are very large so that the size of electrical components has little impact on these signals. For example, the very-low-frequency (VLF) range, which encompasses much of the human hearing range of 20 Hz to 20 kHz, features signals with wavelengths as long as 100 km (3 kHz). The high-frequency (HF) band, with frequencies from 3 to 30 MHz and wavelengths from 100 to 10 m, has long been used for military tactical radios and for amateur ("ham") radio operators (the 30-m or 10-MHz range) for the long-distance propagation properties of the waves.

Signals in the very-high-frequency (VHF) range of 30 to 300 MHz and the ultra-high-frequency (UHF) range of 300 MHz to 3 GHz have long been used for television broadcast, although in recent decades the 800- and 900-MHz bands are better known for use by cellular and cordless telephones. As cellular telephones have migrated to higher-frequency use in the bands around 1900 through 2100 MHz, additional wireless applications have grown in spot frequencies such as 2450 MHz, such as wireless local area networks (WLANs) and Bluetooth®. The over-used frequency is also the designated frequency for industrial heating and microwave ovens, since 2450 MHz is the water-absorption frequency that results in heating of water-containing materials.

Super-high frequencies (SHFs) from 3 to 30 GHz are generally known as microwave frequencies, and extremely high frequencies (EHFs) from 30 to 300 GHz are known as millimeter-wave frequencies because of their small signal wavelengths. Microwave frequencies at 12 and 18 GHz are commonly associated with direct-broadcast-satellite (DBS) television applications and line-of-sight digital radios. The smaller wavelengths of millimeter-wave signals make them well suited for radar systems, especially within specific narrow operating-frequency bands, such as 77 GHz, which is commonly used in automotive collision-avoidance systems.



Frequency Band	Range	Wavelengths
VLF	3 to 30 kHz	100 to 10 km
LF	30 to 300 kHz	10 to 1 km
MF	0.3 to 3 MHz	1 to 0.1 km
HF	3 to 30 MHz	100 to 10 m
VHF	30 to 300 MHz	10 to 1 m
UHF	300 to 3000 MHz	100 to 10 cm
SHF	3 to 30 GHz	10 to 1 cm
EHF	30 to 300 GHz	10 to 1 mm

TABLE A1. Frequencies and wavelengths

A Word about Antennas

Antennas make the work of the RF front end possible. Because they are relatively simple structures with an entirely electromagnetic (EM) behavior, antennas are among the most misunderstood component in the radio front end. In specifying antennas for a radio front end, it is enough to know that the antenna must

be properly connected to the remainder of the RF front end by means of an antenna coupling network. Also, the function of that network differs depending on whether it is used with the front end as a receiver or as a transmitter. When used with the front-end receiver, the task of the antenna-coupling network is to provide the highest sensitivity possible. To do this, the network should achieve the lowest loss possible, since loss translates into degraded noise figure. In some cases, the losses of interconnecting cable must also be considered when judging the effects of antenna matching to system receive sensitivity. When used with the transmitter, the antenna coupling network must produce the maximum radiated power from the associated transmit electronics.

VECTOR Algebra

Many of the design equations contained in earlier chapters require that the user be familiar with vector algebra. It is the intent of this appendix to provide, for those who are unfamiliar with this subject, a working knowledge of vector addition, subtraction, multiplication, and division.

As illustrated in Fig. B-1, a vector may be expressed in either rectangular or polar form. In *rectangular* form, the vector quantity is expressed as a sum of its coordinate parts. Thus, the vector A shown in Fig. B-1 can be expressed as the sum of 5 units in the x direction and 5 units in the y direction, or $A = 5 + j5$. That same vector may be expressed in *polar* notation as a distance (R) from the point of origin at an angle (θ) from the x axis. If vector A were measured, its length would be found to be 7.07 units at an angle of 45° from the x axis. Thus,

$$A = 5 + j5 \quad \text{or} \quad A = 7.07\angle 45^\circ$$

Similarly, vector B can be expressed in rectangular form as $5 - j10$ or in polar form as $11.18\angle -63.4^\circ$. Note that negative angles are measured clockwise from the x axis while positive angles are measured counterclockwise.

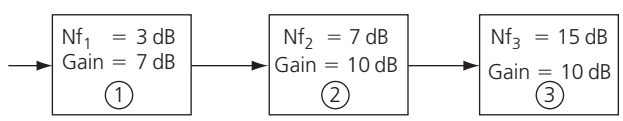
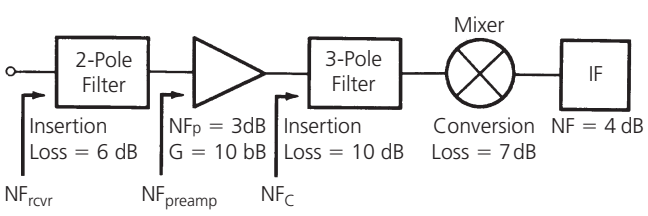


FIG. B-1. Vector coordinates in rectangular and polar form.



RECTANGULAR/POLAR AND POLAR/RECTANGULAR CONVERSION

Rather than plotting a vector to graphically determine its component parts, it is more convenient to perform a few simple mathematical calculations. Any vector expressed in rectangular form may be converted to polar form (Example B-1) using the following formulas:

$$R = \sqrt{x^2 + y^2} \quad \text{and} \quad \theta = \arctan\frac{y}{x}$$

The conversion from polar to rectangular notation (Example B-2) can be made by using the following formulas:

$$x = R\cos\theta \quad \text{and} \quad y = R\sin\theta$$

EXAMPLE B-1

The input impedance of a transistor is found to be $Z = 25 - j10$ ohms. Express this impedance in polar notation.

Solution

The magnitude of the resulting vector (R) is found as:

$$\begin{aligned} R &= \sqrt{x^2 + y^2} \\ &= \sqrt{625 + 100} \\ &= 26.9 \end{aligned}$$

The resulting angle from the x axis is found to be:

$$\begin{aligned} \theta &= \arctan\frac{y}{x} \\ &= \arctan\frac{-10}{25} \\ &= -21.8^\circ \end{aligned}$$

Thus, $Z = 25 - j10$ ohms can also be expressed as $Z = 26.9\angle -21.8^\circ$ ohms.

VECTOR ADDITION

Two vector quantities can be added by performing two separate additions—one for the respective x components and one for the respective y components (Example B-3). Of course, the resultant may be expressed in either rectangular or polar form.

EXAMPLE B-2

The input impedance of a transistor is found to be $Z = 26.9\,\underline{/-21.8°}$. Express this impedance in rectangular form.

Solution

First:

$$\begin{aligned} x &= R\cos\theta \\ &= 26.9\cos(-21.8°) \\ &= 26.9(0.9285) \\ &= 25 \end{aligned}$$

and then:

$$\begin{aligned} y &= R\sin\theta \\ &= 26.9\sin(-21.8°) \\ &= 26.9(-0.3714) \\ &= -10 \end{aligned}$$

Thus, $Z = 25 - j10$ ohms.

EXAMPLE B-3

An impedance of $Z_1 = 11.18\,\underline{/63.40°}$ ohms is added in series with an impedance of $Z_2 = 18.03\,\underline{/-56.3°}$ ohms. What is the resulting series impedance (Z_T) expressed in rectangular form?

Solution

Before the addition can be performed, the polar quantities of the problem must be transformed to rectangular notation. For Z_1:

$$\begin{aligned} x_1 &= R_1\cos\theta_1 \\ &= 11.18\cos(63.4°) \\ &= 5 \\ y_1 &= R_1\sin\theta_1 \\ &= 11.18\sin(63.4°) \\ &= 10 \end{aligned}$$

Thus,

$$Z_1 = 5 + j10 \text{ ohms}$$

For Z_2:

$$\begin{aligned} x_2 &= R_2\cos\theta_2 \\ &= 18.03\cos(-56.3°) \\ &= 10 \end{aligned}$$

EXAMPLE B-3 (Continued)

$$\begin{aligned} y_2 &= R_2\sin\theta_2 \\ &= 18.03\sin(-56.3°) \\ &= -15 \end{aligned}$$

Thus, $Z_2 = 10 - j15$ ohms.

To perform the addition, add the respective x components and the respective y components.

$$\begin{aligned} x_T &= x_1 + x_2 \\ &= 5 + 10 \\ &= 15 \\ y_T &= y_1 + y_2 \\ &= 10 - 15 \\ &= -5 \end{aligned}$$

Thus, $Z_T = 15 - j5$ ohms.

VECTOR SUBTRACTION

Vector subtraction is performed in a similar manner to that of addition (Example B-4). The two vector quantities must first

EXAMPLE B-4

Using the following values:

$$V_1 = 11.18\,\underline{/63.40°}$$

$$V_2 = 18.03\,\underline{/-56.3°}$$

Perform the calculation, $V_T = V_1 - V_2$

Solution

Both quantities must first be expressed in rectangular form. For V_1:

$$\begin{aligned} x_1 &= R_1\cos\theta_1 \\ &= 11.18\cos(63.4°) \\ &= 5 \\ y_1 &= R_1\sin\theta_1 \\ &= 11.18\sin(63.4°) \\ &= 10 \end{aligned}$$

and, then, for V_2:

$$\begin{aligned} x_2 &= R_2\cos\theta_2 \\ &= 18.03\cos(-56.3°) \\ &= 10 \end{aligned}$$

Continued on next page

EXAMPLE B-4—Cont

$$y_2 = R_2 \sin\theta_2$$
$$= 18.03 \sin(-56.3°)$$
$$= -15$$

Subtracting the x and y components, we get:

$$x_T = x_1 - x_2$$
$$= 5 - 10$$
$$= -5$$
$$y_T = y_1 - y_2$$
$$= 10 - (-15)$$
$$= 25$$

Therefore, $V_T = -5 + j25$.

be expressed in rectangular form, and their respective x and y components may then be subtracted.

VECTOR MULTPLICATION

Multiplication of two vectors is accomplished by first converting both vectors to polar form. The magnitudes (R) of the vectors are then multiplied and their angles are added (Example B-5). Thus,

$$R_T = R_1 R_2 \quad \text{and} \quad \theta_T = \theta_1 + \theta_2$$

EXAMPLE B-5

For a transistor, $S_{21} = 5.6\angle 60°$ and $S_{12} = 0.1\angle 30°$. Find the product $S_{21}S_{12}$.

Solution

Both S parameters are already in polar form, therefore:

$$R_T = R_1 R_2$$
$$= (5.6)(0.1)$$
$$= 0.56$$

and,

$$\theta_T = \theta_1 + \theta_2$$
$$= 60° + 30°$$
$$= 90°$$

Thus, the product $S_{21}S_{12}$ is equal to $0.56\angle 90°$.

VECTOR DIVISION

Vector division is performed by first converting both vectors to polar form. The vector quotient is then found by *dividing* the magnitudes and subtracting the angles (Example B-6). Use the formulas:

$$R_T = \frac{R_1}{R_2} \quad \text{and} \quad \theta_T = \theta_1 - \theta_2$$

EXAMPLE B-6

Perform the following vector division:

$$V_T = \frac{V_1}{V_2}$$

where

$$V_1 = 40\angle 60°$$
$$V_2 = 5 + j5$$

Solution

V_1 is already in polar form. Convert V_2 to polar form.

$$V_2 = 7.071\angle 45°$$

Divide the magnitudes.

$$R_T = \frac{R_1}{R_2}$$
$$= \frac{40}{7.071}$$
$$= 5.66$$

Subtract the angles.

$$\theta_T = \theta_1 - \theta_2$$
$$= 60° - 45°$$
$$= 15°$$

Therefore, the quotient is equal to $5.66\angle 15°$.

REAL, IMAGINARY, AND MAGNITUDE COMPONENTS

Several references are made throughout the text to the "real part," the "imaginary part," and the "magnitude" of a complex vector (Example B-7). These components are described as follows:

When given the complex vector V, where

$$V = R\angle\theta$$
$$= x + jy$$

the *real part* of the vector V is given as:

$$\text{Re(V)} = x$$

the imaginary part of the vector V is given as:

$$\text{Im(V)} = jy$$

and the magnitude of the vector V is, then, given as:

$$|V| = R$$

EXAMPLE B-7

Given the complex vector $V = 10\angle 60°$, find Re(V), Im(V), and $|V|$.

Solution

First, express the vector in rectangular form.

$$x = R\cos\theta$$
$$= 10\cos(60°)$$
$$= 5$$

$$y = R\sin\theta$$
$$= 10\sin(60°)$$
$$= 8.66$$

Therefore, $V = 5 + j8.66$ and

$$\text{Re}(V) = 5$$
$$\text{Im}(V) = j8.66$$
$$|V| = 10$$

BIBLIOGRAPHY

"2-3 GHz Broadband TRL Match." *SmithMatch User Manual*. February 26, 2007, p 12.

Anderson, L. H., "How to Design Matching Networks," *Ham Radio*, April 1978.

"Auriga: Modeling & Verification." [Online] Available http://www.ftlsystems.com/eda/Auriga/, March 23, 2007.

"Basic Smith Chart Principles/Techniques." SmithMatch User Manual.. February 26, 2007. p 10.

Besser, L., "Take the Hassle Out of FET Amp Design," *MSN*, September 1977.

Besser, L., "Update Amplifier Design With Network Synthesis," *Microwaves*, October 1977.

Bindra, Ashok. "Key Semiconductor Technology Trends to Watch." RF Design. [Online] Available http://rfdesign.com/rfic/cmos_and_compound_ics/semiconductor-technology-trends-0107, January 1, 2007.

"Bipolar junction transistor." [Online] Available http://en.wikipedia.org/wiki/Bipolar_junction_transistor, August 1, 2007.

"Bipolar Transistor." [Online] Available http://searchsmb.techtarget.com/sDefinition/0,,sid44_gci211668,00.html, March 28, 2001.

Blyler, John, "RF Board Designers Confront Many Challenges," Embedded Intel magazine, Spring 2007 [http://www.embeddedintel.com/technology_applications.php?app=245]

Blyler, John E., "Wireless Networking: Overview and Roadmap", IEEE Oregon Section (IEEE 802.11 Wireless LAN Technology), 2001.

Blyler, John, "Understanding Low-Power Designs," Wireless Systems Design magazine, June 2001, Pp. 57–58.

Breed, Gary. "A Heritage of technology and Innovation: The History of Agilent EEsof EDA." [Online] Available http://eesof.tm.agilent.com/pdf/history_2001.pdf, 2001.

Brown, W. L., "Don't Guess at Bias Circuit Design," *Electronic Design*, May 9, 1968.

Browne, Jack, "ADCs Clear Way To Digital Receivers," Microwave & RF magazine (www.mwrf.com).

Burwasser, A., "Wideband Monofilar Autotransformers," *RF Design*, January/February 1981.

Burwasser, A., "How to Design Broadband JFET Amplifiers to Provide Top Performance From VLF to Over 100 MHz," *Ham Radio*, November 1979.

Byrne, Patrick J. "Innovation in Measurement: How the dynamics of the electronics industry are leading to measurement innovation." Agilent Measurement Journal. First quarter 2007, Pp. 9–10.

"C++." [Online] Available http://en.wikipedia.org/wiki/C%2B%2B, August 5, 2007.

"Cadence RF Design Methodology Kit." [Online] Available http://cadence.com/products/kits/RF_Design/index.aspx, March 14, 2006.

"Cadence RF SIP Methodology Kit." [Online] Available http://cadence.com/products/kits/rf_sip_methodology/index.aspx#, March 14, 2006.

Carlson, A., *Communications Systems*, McGraw-Hill, New York, NY, 1975.

Carson, R., *High Frequency Amplifiers*, John Wiley & Sons, New York, NY, 1975.

Chatterjee, Pallab. " That 80's Short or AMS IP Creation Flashback." Wireless Chip Designer, [Online] Available, July 21, 2006.

Chern, Jue-Hsien. "Transitioning To A New Mixed-Signal Design Methodology." Chip Design, [Online] Available, February 20, 2007.

Cohn, S. B., "Direct-Coupled-Resonator Filters," *Proceedings of the IRE*, February 1957.

Cohn, S. B., "Dissipation Loss in Multiple-Coupled Resonator Filters," *Proceedings of the IRE*, August 1959.

"dBw." RFID Glossary. [Online] Available http://www.rfid4u.com/glossary.asp, February 25, 2007.

"Decibel." Wikipedia. [Online] Available http://en.wikipedia.org/wiki/Decibel, August 2, 2007.

DeMaw, M. F., *Practical RF Communications Data for Engineers and Technicians*, Howard W. Sams & Co., Inc., Indianapolis, IN, 1978.

DeMaw, M. F., "The Practical Side of Toroids," *QST Magazine*, June 1979.

DeMaw, M. F., "Magnetic Cores in RF Circuits," *RF Design*, April 1980.

"Designing Impedance Matching Networks." [Online] Available http://www.mathworks.com/applications/dsp_comm/demos.html?file=/products/demos/shipping/rf/imped_match.html, May 3, 2007.

Dishal, M., "Exact Design and Analysis of Double and Triple Tuned Band-Pass Amplifiers," *Proceedings of the IRE*, June 1947.

Dunn, John. "Spiral Inductor Modeling On RFICs." March 2007, Pp. 1–7.
"Example: Efficiently Simulating a Direct-Conversion Receiver or Transmitter." [Online] Available http://eesof.tm.agilent.com/products/design_flows/rfic/examples_05.html, March 21, 2007.
"Example: Verification and Troubleshooting of a Receiver with a WLAN 802.11b Signal." [Online] Available http://eesof.tm.agilent.com/products/design_flows/rfic/examples_01.html, March 21, 2007.
"Field-Effect Transistor." [Online] Available http://searchsmb.techtarget.com/sDefinition/0,,sid44_gci213961,00.html, August 8, 2000.
"Field effect Transistor." [Online] Available http://en.wikipedia.org/wiki/Field_effect_transistor, August 3, 2007.
Fink, D. G., *Electronics Engineers' Handbook*, McGraw-Hill, New York, NY, 1975.
Fisk, J., "How to Use the Smith Chart," *Ham Radio*, March 1978.
Geffe, P., *Simplified Modern Filter Design*, John F. Rider, New York, NY, 1963.
"GENESYS." [Online] Available http://www.eagleware.com/products/genesys/genesys.html, 2006.
Grammer, G., "Simplified Design of Impedance Matching Networks," *QST Magazine*, March 1957.
Grossman, M., "Focus on Ferrite Materials: They Star as HF Magnetic Cores," *Electronic Design*, April 30, 1981.
Gupta, S., Bayless, J., and Peikari, B., *Circuit Analysis*, Intext Educational Pub., Scranton, PA, 1972.
Hall, C., "RF Chokes, Their Performance Above and Below Resonance," *Ham Radio*, June 1978.
Hardy, J., *High Frequency Circuit Design*, Reston Publishing Co., Inc., Reston, VA, 1979.
Harper, C. A., *Handbook of Components for Electronics*, McGraw-Hill, New York, NY.
Heimlich, Mike. "Challenges and Solutions: Module Design." PowerPoint Presentation. May 4, 2007, Pp. 1–65.
Heimlich, Mike and Scott Wedge. "Meeting the Multi-Gbps Design Challenge with Signal Integrity Design." PowerPoint Presentation. May 4, 2007, Pp. 1–28.
Hejhall, Roy, "Systemizing RF Power Amplifier Design," *Motorola AN-282A*, 1975.
Hejhall, Roy, "RF Small Signal Design Using Two-Port Parameters," *Motorola AN-215A*, February 1977.
Hejhall, Roy, "Understanding Transistor Response Parameters," *Motorola AN-139A*, 1974.
Hewlett-Packard Components, "Transistor Parameter Measurements," *Hewlett-Packard AN-77-1*, February 1967.
Hewlett-Packard Components, "A Low Noise 4 GHz Transistor Amplifier Using the HXTR-6101 Silicon Bipolar Transistor," *Hewlett-Packard AN-967*, May 1975.
Hewlett-Packard Components, "A 6 GHz Amplifier Using the HFET-1101 GaAs FET," *Hewlett-Packard AN-970*, February 1978.
Hewlett-Packard Components, "S-Parameter Techniques for Faster, More Accurate Network Design," *Hewlett-Packard AN-95-1*.
Hewlett-Packard Components, "S-Parameter Design," *Hewlett-Packard AN-154*, April 1972.
Hewlett-Packard Components, "Microwave Transistor Bias Considerations," *Hewlett-Packard AN-944-1*, October 1973.
"High-pass filter." Wikipedia [Online] Available http://en.wikipedia.org/wiki/High-pass_filter, July 22, 2007.
Hoff, I. M., "Pi-Network Design for High-Frequency Power Amplifiers," *Ham Radio*, June 1978.
Hooijmans, Pieter W. "RF Front End Application and Technology Trends." Proceedings of Design Automation Conference, June 2–6, 2003, Pp. 73–78.
Horvitz, Robert. "Frequency Conversion and New Technologies: Challenges and Opportunities in Spectrum Management," Presented at the Center for Strategic Research, Moscow, June 20, 2006.
Howe, H. Jr., "Basic Parameters Hold the Key to Microstrip-Amplifier Design," *EDN*, October 5, 1980.
"Impedance Matching and the Smith Chart: The Fundamentals." [Online] Available http://www.maxim-ic.com/appnotes.cfm/appnote_number/742, March 23, 2001.
"Introduction to the RF BS XE Flow." RF in BSXE Flow User's Manual, V8.13-1. Pp. 1–10.
Kaufman, M. and Seidman, A., Handbook of Electronics Calculations for Engineers and Technicians, McGraw-Hill, New York, NY, 1979.
Kraus, H. and Allen, C., "Designing Toroidal Transformers to Optimize Wideband Performance," *Electronics*, August 16, 1973.
Krauss, H., Bostian, C., and Raak, F., *Solid State Radio Engineering*, John Wiley & Sons, New York, NY, 1980.
Lenk, J., *Handbook of Simplified Solid State Circuit Design*, Prentice-Hall, Englewood Cliffs, NJ, 1971.
Litty, T. P., "Microstrip RF Amplifier Design," *RF Design*, January/February 1979.
Llamas, Ramon and Ryan Reith. "Converged Mobile Device Market Surges Ahead on 42% Growth in 2006, According to IDC." [Online] Available http://www.idc.com/getdoc.jsp?containerId=prUS20578607, February 26, 2007.
"Low Noise Amplifer Design with GENESYS." May 22, 2007, Pp. 1–24.
"Low-pass filter." Wikipedia. [Online] Available http://en.wikipedia.org/wiki/Low-pass_filter, August 2, 2007.
Miller, J., *Solid State Communications: Design of Communications Equipment Using Semiconductors*, McGraw-Hill, New York, NY, 1966.
Millman, J. and Halkias, C., *Integrated Electronics: Analog Digital Circuits & Systems*, McGraw-Hill, New York, NY, 1972.
Mini-Circuits, "Constant-Impedance IF Bandpass Filters Improve Circuits Performance," www.minicircuits.com.
Motorola Semiconductor Products Inc., "Semiconductor Noise Figure Considerations," *Motorola AN-421*, August 1976.
Myoung, Robert and Denis Soldo. "System in Package (SiP) and Stacked Package Solutions." 2006 Worldwide Application Workshop, Pp. 1–11.
Nash, Eamon, "Measurement and Control of RF Power," Analog Devices (www.analog.com).
Neilson, Dave and Tom Quan. "System/Circuit Design and Analysis of an IEEE 802.11a RF CMOS Transceiver." Application Note, March 2007, Pp. 1–10.
"ParkerVision: Our Story." [Online] Available http://www.parkervision.com/about_us/our_story.php, February 20, 2007.
"PCB Design Process." [Online] Available http://www.pcbdesignexpress.com/designprocesses.jsp, March 23, 2007.
Quan, Tom. "Design Suite Provides Advanced Technology for Analog and RFIC Design." High Frequency Electronics. May, 2004, Pp. 50–55.
"Radio – History of Radio." [Online] Available http://www.telecompie.com/radio/history.html, February 13, 2007.

Reference Data for Radio Engineers, Howard W. Sams & Co., Inc., Indianapolis, IN, 1975.
Rencher, Mark. "VHDL-AMS Extensions Enable RF Harmonic Balance Simulation." High Frequency Electronics. March 2004, Pp. 44–47.
"RF Design Environment 2006." [Online] Available http://eesof.tm.agilent.com/products/rfde2006_update1.html, March 16, 2007.
"RF Design Environment – Verilog-A Complier." [Online] Available http://eesof.tm.agilent.com/products/rfde2003c_verilog.html, March 16, 2007.
"RF Design Methodology and Flow." White Paper. 2006, Pp. 1–8.
"RF IC Flow." [Online] Available http://cadence.com/whitepapers/rfic_flow_wp.pdf, 2004, Pp. 1–7.
RF Packaging – Technical, [Online] Available http://www.kjas-interconnect.com/RFTechnical.html, 2002.
"RF Toolbox 2.1." [Online] Available http://www.mathworks.com/products/rftoolbox/index.html, March 25, 2007.
"RLC Circuit." Wikipedia [Online] Available http://en.wikipedia.org/wiki/RLC_circuit, June 22, 2007.
Rousset, Jean, "From W-CDMA RF Front-End Specifications to Requirements," consultant to Agilent Technologies (www.agilent.com).
Ruthroff, C., "Some Broadband Transformers," *Proceedings of the IRE*, August 1959.
Sando, S., "Very Low-Noise GaAs FET Preamp for 432 MHz," *Ham Radio*, April 1978.
Scherer, D., "Today's Lesson–Learn About Low-Noise Design," *Microwaves*, April 1979.
Sentz, R., *Voltage and Power Amplifiers*, Holt, Rinehart & Winston, Inc., New York, NY, 1968.
"SerDes Signal Integrity Design and Analysis." Seminar. December 2006, Pp. 9–22, 30–31.
Sevick, J., "Broadband Matching Transformers Can Handle Many Kilowatts," *Electronics*, November 25, 1976.
Shuch, Paul, "Solid State Microwave Amplifier Design," *Ham Radio*, October 1976.
Shulamn, J. M., "T-Network Impedance Matching to Coaxial Feedlines," *Ham Radio*, September 1978.
Smith, P. H., *Electronic Applications of the Smith Chart: In Waveguide, Circuit, & Component Analysis*, McGraw-Hill, New York, NY, 1969.
Stickley, John and Wade Stone. "Combining the power of MATLAB for signal generation and analysis with an RTL representation in a high performance emulator." Chip Design, [Online] Available http://www.chipdesignmag.com/print.php?articleId=258?issueId=0, April 10, 2007.
"SystemVerilog." [Online] Available http://en.wikipedia.org/wiki/SystemVerilog, July 21, 2007.
"SytemC: Definition and recommended Links." [Online] Available http://www.eeglossary.com/systemc.htm, July 25, 2007.
"Transistors." [Online] Available http://www.ic2forum.com/viewtopic.php?t=82, November 9, 2006.
Trout, B. "Small-Signal RF Design With Dual-Gate MOSFETs," *Motorola AN-478A*, 1971.
"Using GENESYS with VNA Data for Matching." Using GENESYS with a Vector Network Analyzer, May 15, 2007, Pp. 1–16.
"Verilog." [Online] Available http://en.wikipedia.org/wiki/Verilog, August 3, 2007.
"Verilog-AMS." [Online] Available http://en.wikipedia.org/wiki/Verilog-AMS, December 22, 2006.
"VHDL" [Online] Available http://en.wikipedia.org/wiki/VHDL, August 2, 2007.
Viklund, Per. Personal Interview. March 7, 2007.
Wetherhold, E. E., "Design 7-Element Low-Pass Filters Using Standard Value Capacitors," *EDN*, January 7, 1981.
Williams, A. B., *Electronic Filter Design Handbook*, McGraw-Hill, New York, NY, 1981.
Witzke, Edward, "An Approach to Wireless Communications at Sandia National," SAND2002-3312, October 2002 [http://www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/2002/023312.pdf].
Yen, Albert. "Foundry Support Is Critical to SoC Design and Implementation." Chip Design. [Online] Available http://www.chipdesignmag.com/display.php?articleId=964&issueId=20, December 2006/January 2007.
Ziemer, R. E. and Tranter, W., *Principles of Communications: Systems, Modulation & Noise*, Houghton Mifflin Co., Boston, MA, 1976.
Zverev, A., *Handbook of Filter Synthesis*, John Wiley & Sons, New York, NY, 1967.
Zverev, A., "Introduction to Filters," *Electro-Technology*, June 1964.

D

T