SECOND EDITION

Complete Wireless Design

COTTER W. SAYRE

Complete Wireless Design

This page intentionally left blank

Complete Wireless Design

Cotter W. Sayre

Second Edition

New York Chicago San Francisco
Lisbon London Madrid Mexico City
Milan New Delhi San Juan
Seoul Singapore Sydney Toronto

0-07-164272-2

The material in this eBook also appears in the print version of this title: 0-07-154452-6.

DOI: 10.1036/0071544526

To my lovely wife, Linda, without whom this book would not have been possible.
And to my wonderful mother, Jan, for all her love and support.

This page intentionally left blank

About the Author

Cotter W. Sayre of San Jose, California, has worked as a senior RF design engineer at Micro Linear and Radix corporations, as well as a wireless hardware design engineer for 3Com Corporation's Advanced Development Group. His specialty is design, simulation, layout, testing, and troubleshooting of wireless transmitters and receivers up to 6 GHz. Mr. Sayre is a member of the Institute of Electrical and Electronics Engineers and the IEEE Microwave Theory and Techniques Society.

This page intentionally left blank

Contents

Preface

This Second Edition of *Complete Wireless Design* not only supplies the reader with a solid practical grounding in the latest RF circuit and systems operation, but also explains, in step-by-step detail, how to plan, design, simulate, build, and test a complete RF/microwave voice and data wireless radio, all the way from the discrete circuit level to the finished wireless system.

Sections of all chapters within this new edition have been expanded, enhanced, and fully updated, with the addition of new explanatory figures that have been optimized for clarity. New chapters have been added that detail RF testing procedures, printed circuit board layout, RF software simulation techniques, and small antenna design. A major new feature in this book is the worked-out design examples (*Quick Examples*) with full RF simulation results. The enclosed, comprehensive computer CD has also been expanded and updated to include more free, cutting-edge RF programs, and will enable the reader to design and simulate virtually any lumped or distributed RF circuit, or system typically needed in low-power consumer design.

Unlike many wireless books, *Complete Wireless Design* does not simply present predesigned circuits and expect readers to modify them in some haphazard fashion for their own wireless applications, nor does the book present overly complex equations for the design of wireless circuits and systems which most readers, even engineers, would have difficulty understanding, much less applying. Instead, *Complete Wireless Design* will allow the reader to easily and quickly, by employing basic algebra and the enclosed software programs, design cutting-edge oscillators, amplifiers, mixers, filters, PLLs, frequency multipliers, switches, microstrip elements, AGC loops, power splitters, attenuators, diplexers, antennas, and PCBs, as well as plan the entire wireless communications system.

Any supplementary information, text or software updates, or errata for the Second Edition of *Complete Wireless Design* will be posted at *http://cotter.sayre.googlepages.com/home*.

I'm very excited about this new edition, and I hope you will be too.

Cotter W. Sayre

This page intentionally left blank

Acknowledgments

A book of the comprehensive and detailed scope of this one would have been impossible to write without extensively studying—and applying—the prior work of many others who came before. I owe a special thanks to the incredibly practical and informative books and articles of Randy Rhea, Dean Banerjee, Chris Bowick, Peter Vizmuller, and Les Besser; and to all the other excellent authors and engineers mentioned within this book's bibliography.

I also wish to thank the following individuals for their assistance and contributions in giving permission to reproduce the enclosed RF software, as well as the software documentation:

- Stefan Jahn, lead developer, for the RF simulator *Qucs*.
- Kirt Blattenberger, senior engineer and owner of RFCafé.com, for the systems simulator *RF Workbench*.
- Mike Engelhardt, senior Linear Technology engineer, for the Spice simulator *LTSpice*.
- Neil W. Heckt, senior engineer and owner of AADE.com, for the lumped filter synthesis software *AADE Filter Design and Analysis*.
- Arie Voors, developer and computer engineer, for the NEC 3D antenna simulator *4NEC2X*.
- Christopher A. Schell, senior engineer at National Semiconductor, for the PLL design software *EasyPLL*.
- Robert L. Myers, AppCad Program Manager, for Agilent's RF design software *AppCad*.
- Dr. James C. Rautio, president of Sonnet Software, for the EM simulator *Sonnet Lite em*.
- Applied Radio Labs for the PLL design software *ADIsimPLL*.

And thank you to the following individuals:

- Dale Henkes, President of Applied Microwave, for the use of Linc2Pro.
- Daren McClearnon, Marketing Department, Agilent Corp., for the use of *Genesys' Harbec*.

This page intentionally left blank

Complete Wireless Design

This page intentionally left blank

CHAPTER 1
Wireless Essentials

A firm understanding of how passive and active components function at high frequencies, as well as a strong grasp of the fundamental concepts of transmission lines and scattering parameters (*S*-parameters) are essential to successful circuit design.

1.1 Passive Components at Radio Frequency

1.1.1 Introduction

Lumped (physical) resistors, capacitors, and inductors are not the "pure" components they are assumed to be at lower frequencies. As shown in Fig. 1.1, their true nature at higher frequencies has undesirable resistances, capacitances, and inductances, which must be taken into account during design, simulation, and layout of any wireless circuit.

At microwave frequencies the lengths of all component leads must be minimized in order to decrease losses due to lead inductance. *Surface-mount devices (SMDs)* are perfect for decreasing this lead length, and thus the series inductance, of any component (Fig. 1.2). Even the board traces that connect the passive components, if the trace is longer than approximately 1/20th of a wavelength, must be converted to transmission-line structures. On printed circuit boards (PCBs), *microstrip* is ideal for this, since it maintains a constant 50-Ω impedance throughout its entire length without adding any undesired inductance or capacitance.

As the frequency of operation of any wireless circuit begins to increase, so does the requirement that the actual physical structure of all of the lumped components themselves be as small as possible, since the part's effective frequency of operation increases as it shrinks in size. The smaller package decreases the typically detrimental distributed reactances, and raises the frequency of the series and parallel resonances.

Maintaining a high *unloaded quality factor* (*Q*) in each individual component is vital to minimize circuit losses. And as shown in Fig. 1.3, the unloaded *Q* of a capacitor decreases with frequency, while the unloaded *Q* of an inductor will actually increase with frequency, after which the *Q* will drop rapidly.

1.1.2 Resistors

A resistor with a value of over a few hundred ohms will begin to decrease in resistance as the frequency of operation is increased (Fig. 1.4). This is caused by the distributed capacitance that is always effectively in parallel with the resistor, shunting the radio frequency (RF) signal around the component, and thus lowering its effective value of resistance. The distributed capacitance is especially problematic not only as the

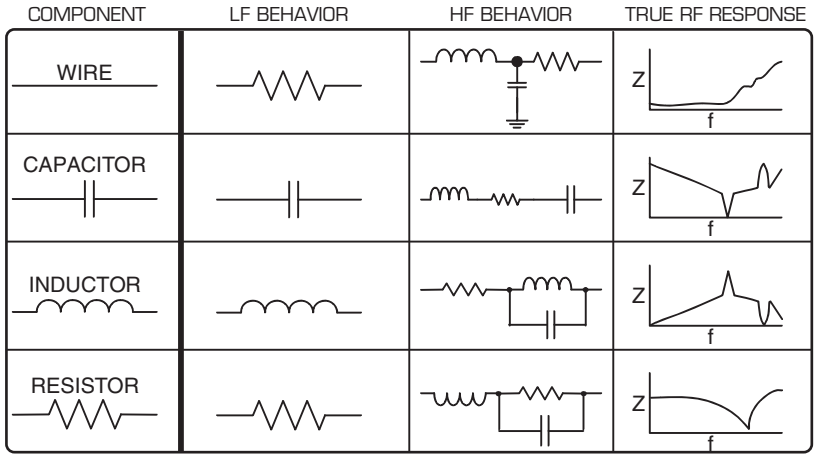


Figure 1.1 A component's real-life behavior at high frequencies.

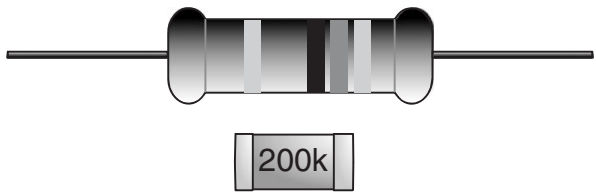


Figure 1.2 A surface-mount (SMD) resistor, along with the older through-hole type.

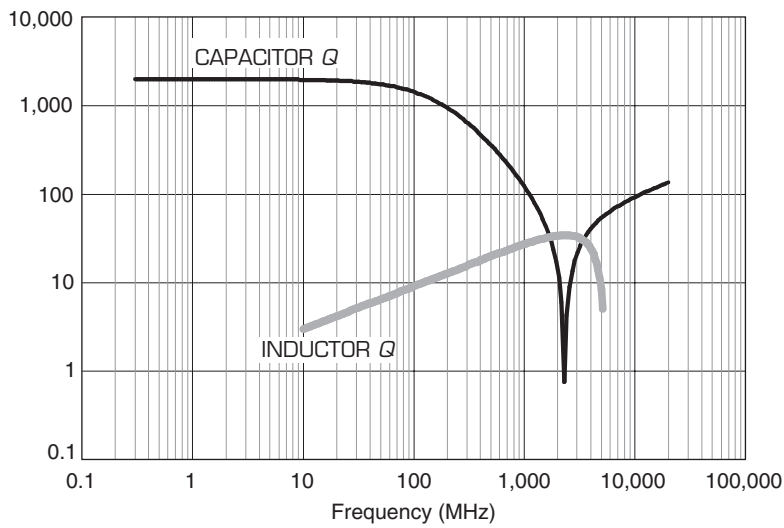


Figure 1.3 Capacitor and inductor *Q* over frequency, using surface-mount components.

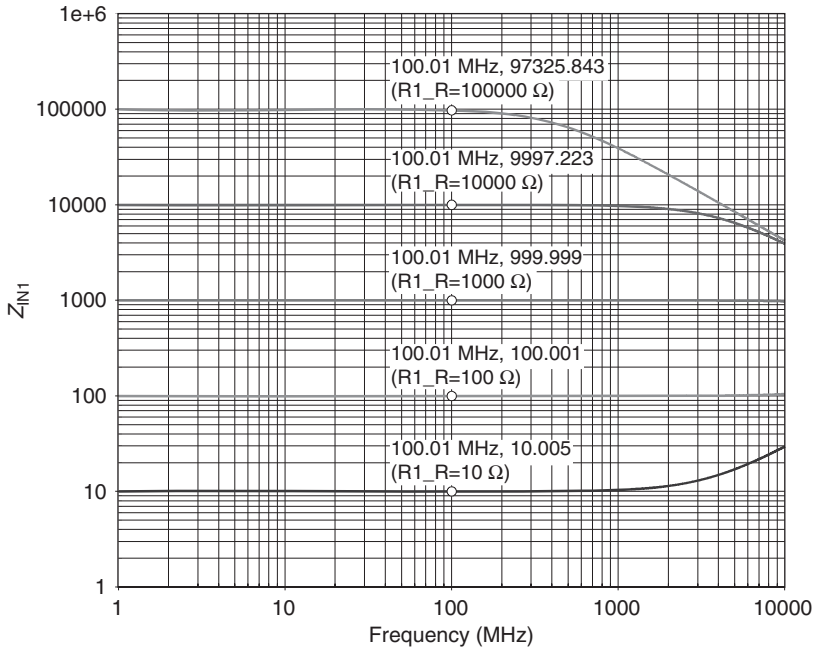


Figure 1.4 The change in a resistor's impedance over frequency.

frequency increases, but also as the resistance value increases. Indeed, if the resistor is not of the high-frequency, thin-film type, a high-value resistor may lose much of its marked resistance to this capacitive effect at relatively low microwave frequencies. Nonetheless, even thin-film resistors will begin to deviate from their rated resistance values when operated above a few hundred megahertz.

1.1.3 Capacitors

Capacitors at RF and microwave frequencies must be chosen not only for their cost and temperature stability, but also for their ability to properly function at high frequencies. As shown in Fig. 1.1, a capacitor will have an undesired lead inductance that begins to adversely change the capacitor's characteristics as the frequency is increased.

A capacitor's published rated value is normally taken at 1 MHz. Therefore, before selecting a capacitor for any critical high-frequency use, it is important to confirm that it is actually capable of meeting our performance criteria over our required band of interest. This can most easily be accomplished by employing the capacitor manufacturer's software, which will describe a particular microwave capacitor model's complete performance over frequency. With such software we can, within seconds, verify the capacitor's true impedance, *S*-parameters, unloaded *Q*, equivalent series resistance (ESR), and apparent capacitance over a very wide bandwidth.

The change in a capacitor's characteristics with frequency is most pronounced when the package's lead inductance resonates with the capacitance of the physical component itself, resulting in a series resonance with a total reactance of nearly 0 Ω. Indeed, resonating a capacitor can even be purposeful. A *j*0 type capacitor is a normal capacitor

that is operated within its *series resonant frequency (SRF)*, which creates a very low series impedance for the frequency of interest. This is perfect for coupling and decoupling at very high frequencies. Above the component's SRF, the capacitor will then become more inductive than capacitive, making it important to confirm that the circuit's design frequency will not reach too far over this SRF value. The capacitor will eventually reach a point called the *parallel resonant frequency (PRF)*, which is located at approximately twice the capacitor's SRF. The PRF places the capacitor in a very high impedance mode, but may not be an issue—and indeed may only be just barely noticeable—when viewed on an RF simulator's graph. This is due to resistive losses within the capacitor itself, which may remove any obvious S_{21} notch at the PRF. But for broadband RF applications, a minimized S_{21} PRF notch over a wide frequency range, and with low overall component insertion losses, is much more desirable than a capacitor that has an obvious high-loss, high-attenuation PRF S_{21} notch (or notches) within the band of interest.

Due to the parasitic effects that the PCB's substrate, traces, and pads have on a capacitor, its SRF will vary from the part's data sheet value. In other words, the SRF of a capacitor is highly dependent on the PCB layout itself. Further, the capacitor's SRF value, as quoted in the data sheet, will actually decrease by up to 40% when a capacitor is mounted in a shunt (trace to ground) configuration on the PCB. Consequently, because of the overall PCB board effects and capacitor tolerances, it is best to select a capacitor that has an SRF that is at least 10% higher than the highest frequency we want to couple or bypass. Clearly, for impedance-matching applications, if we plan on using anything close to the calculated component values, the capacitor's SRF must be significantly higher than the highest frequency we plan to match.

For RF decoupling purposes, we will normally be required to parallel two or more capacitors to effectively cover an extremely wide bandwidth. Occasionally, we may even need to parallel up to two capacitors in wideband series *coupling* applications (Fig. 1.5). Nonetheless, for either coupling or decoupling, these multiple capacitor arrangements should always be simulated with the proper software models to confirm that there will be no major resonant interactions between the two capacitors' unavoidable physical parasitics, since this paralleling can create unexpected performance degradation, especially of S_{21}. The result of this undesired parasitic interaction is called *antiresonance*, and its consequences are shown in Fig. 1.6. The lower frequency capacitor will directly contribute the most to these antiresonance effects, as it will be supplying more of the parasitic inductances than the high-frequency RF capacitor. The low-frequency capacitor will also be traversing, at some frequency, its series resonance point, and thus reaching into the inductive regions of the Smith chart.

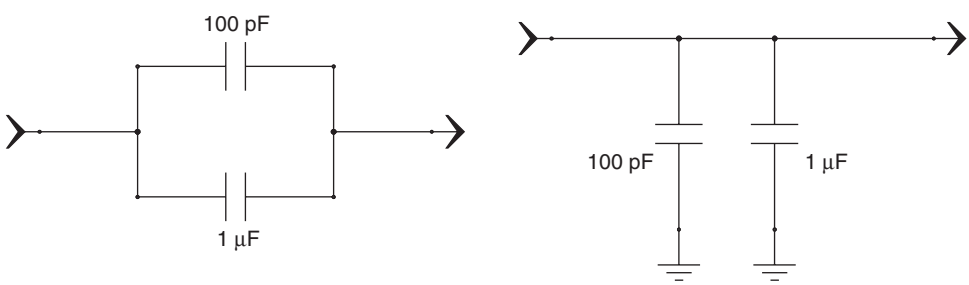


Figure 1.5 Two paralleled capacitors in series and shunt configurations, used for coupling and decoupling.

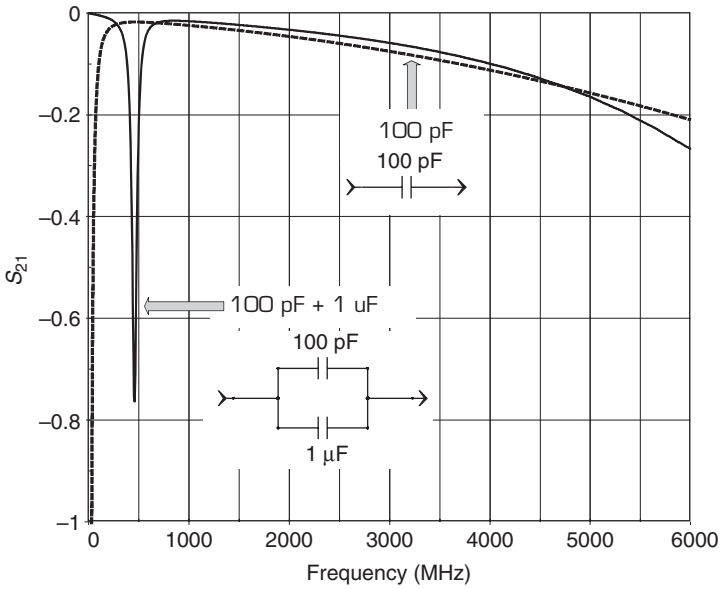


Figure 1.6 Resonances shown when paralleling two nonideal capacitors for wideband coupling, with better wideband S21 performance using a single 100-pF capacitor. (50-Ω ports. PCB mounting effects not modeled.)

We can lessen this effect of parasitic interaction between two or more decoupling shunt capacitors by using a very small resistor (or ferrite bead) in series with the largest decoupling capacitor, with this resistor's value being no higher than 10 Ω.

When using the typical single capacitor series coupling design for an ultra-broadband application, the capacitor may need to operate all the way from a very low frequency to within close proximity to its parallel resonant frequency. In this special case, it is perfectly acceptable, in non-matching applications, to actually use the capacitor at slightly *above* its series resonant frequency, as long as its inductive reactance does not become too great, and the magnitude of the capacitor's impedance at both its low and high frequencies remains small. In these very wideband applications, it may even be better to select an X7R rather than a negative-positive-zero (NPO) capacitor, due to the X7R's superior capacitance per package size. This is especially important if our lowest frequency extends into the lower megahertz regions.

It should be stressed again that the minimization of a capacitor's parasitics, and thus the higher its particular SRF will be, is absolutely critical for tuned circuits applications, where the capacitor should have an SRF far above the required design frequency. However, the higher the value of the capacitor, the lower its SRF, and thus the closer the capacitor will be to its inductive region. Consequently, a higher value capacitor will demonstrate a higher inductance, on average, than a smaller value capacitor of the same case size and model series. This makes it necessary to reach a compromise between the capacitive reactance of the selected capacitor, and the series resonance point of the component in the particular application. In other words, a coupling capacitor that is expected to have a capacitive reactance of 1 Ω at the frequency

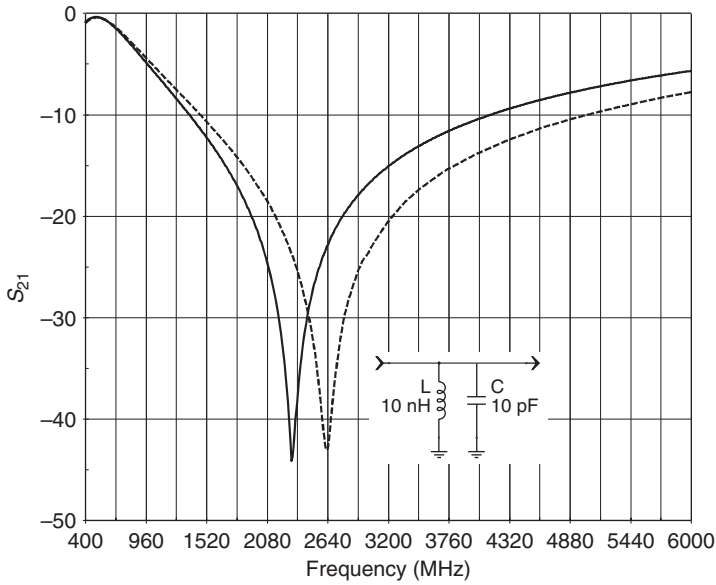


Figure 1.7 At RF, the frequency response of a circuit varies depending on which manufacturer's 10-pF capacitor is used.

of interest may actually be a much poorer choice than one that has a capacitive reactance of 5 Ω (unless the capacitor is chosen to operate as a *J*0 type).

Since component parasitics vary from manufacturer to manufacturer, and even from model to model, for the exact same capacitance value, if we were to substitute another company's capacitors of the exact same size and value, no matter how high or low the quality of the substitute part, we would obtain a slightly different response or impedance for any high-frequency filtering, matching, or coupling network (Fig. 1.7). This type of errant capacitor substitution could now also create lower output power, decreased efficiency, or increased instability in an RF power amplifier, or a different noise figure, return loss, or gain in low-noise amplifier (LNA).

Additionally, at elevated microwave frequencies, even small capacitor value changes can severely affect circuit performance. A tiny variation in a capacitor's value of but 0.2 pF (or less) may alter an optimized RF match or coupling circuit's impedance, even when the same make and model of capacitor is used. This would be classified as a *tolerance* issue, since an ordinary 1.2-pF capacitor's lack of a tight tolerance can permit its value to vary from 0.95 to 1.45 pF. As a result, all components used in critical, high-frequency (> 900 MHz) circuits should employ tight tolerance capacitors when their value is at or below 3.3 pF (for inductors, < 3.3 nH), due to the way low-value component tolerances are specified (this tolerance specification is normally quoted as ±0.25 pF, or at some other fixed capacitance value, for most 5-pF and below capacitors).

Capacitor Types

Different varieties of capacitors have completely different applications. The following paragraphs discuss the various common capacitor types, and their uses in today's wireless circuits.

Electrolytic capacitors, both aluminum and tantalum, are utilized for very low frequency coupling and decoupling tasks. They have poor ESR and high DC leakage and many are polarized. However, they possess a very large amount of capacitance per unit volume, with this value ranging from greater than 22,000 μF down to 1 μF for the aluminum types. The aluminum electrolytics have a limited life span of between 5 and 20 years while tantalums, with their dry internal electrolyte, have a much longer lifetime and less DC dielectric leakage. Unfortunately, tantalums have a lower range of values (between 0.047 μF and 330 μF), and a lower maximum working voltage rating.

Metallized film capacitors are commonly good up to about 6 MHz, and are also used for low-frequency decoupling. These capacitors are available in capacitance ranges from 10 pF to 10 μF, and include the polystyrene, metallized paper, polycarbonate, and Mylar (polyester) families. Metallized film capacitors can be constructed by thinly metallizing multiple dielectric layers.

Silver mica capacitors are an older, less used type of high-frequency capacitor. They have a low ESR and good temperature stability, with a capacitance range available between 2 to 1500 pF.

Non-surface-mount leaded *ceramic* capacitors are found in all RF circuits up to a maximum of 600 MHz. They are available as a single-layer type (*ceramic disk*), and as stacked ceramic (monolithic) structures. Capacitance values range from 1.5 pF to 0.047 μF, with the dielectric available in three different grades: *NPO* (*COG*) for critical temperature–stable applications with tight capacitance tolerances values of 5% or better (with a capacitance range of 10 to 10,000 pF); X7R types, with less temperature stability and a poorer tolerance (±10%) than NPO, with available values of 270 pF to 0.33 μF; and Z5U types, which are typically utilized for bypass and coupling due to their extremely poor capacitance tolerances (±20%) and inferior temperature stability, with a capacitance range of between 0.001 and 2.2 μF. Z5U (and Y5V), capacitors can drop almost all of their marked capacitance at low temperatures, so other less temperature-sensitive capacitors must be used if the product is meant to function properly at low temperatures.

However, by far, the dominant microwave frequency capacitor today is the surface-mount technology (SMT) *ceramic* and *porcelain chip* types, which are used in all parts of RF circuits up to about 15 GHz. The SMT passives are available in many different package sizes, but only the smaller sizes are of interest in microwave circuit design, such as 0201 (20 mils long × 10 mils wide), 0402, 0603, and 0805s. Nonetheless, even when employing these ultra-high-quality RF and microwave chip capacitors, maximum capacitance values as well as package sizes must still be small in order to function properly at elevated frequencies. Depending on frequency, a maximum value of 10 pF or less may be all that we can use in a particular circuit due to the increasing internal parasitic inductance as the part's capacitance value is increased.

Microwave chip capacitors are available in multilayer and single-layer configurations, with the multilayer types normally coming in the basic SMD package sizes described above. The single-layer capacitors are more specialized and expensive, and can be a bit more difficult to mount on a PCB due to their nonstandard mounting terminations, which may consist of a conductive "ribbon" for bonding the part to the PCB's microstrip. However, single-layer capacitors can operate at much higher frequencies than multilayer (up to 10s of gigahertz), but will possess much lower maximum capacitance ranges.

1.1.4 Inductors

A significant, high-frequency effect in any inductor is *undesired distributed parasitic capacitance*, which is a capacitance that is in parallel with the actual desired inductance of the coil (Fig. 1.1). This means that there must be some frequency that will force the coil's inductance to be in parallel resonance with this distributed capacitance, causing a high impedance peak to form at that particular frequency. In fact, the impedance created by this parallel resonance would be infinite if not for the small value of wire resistance found in series with the inductor's own structure. The point of resonance is referred to as the *self-resonant frequency* (SRF) of the inductor (Fig. 1.8), and if the inductor is expected to operate in a matching circuit, this SRF must be at a much higher value than the circuit's own frequency of operation. Indeed, RF inductors are built with small form factors in order to decrease this distributed capacitance effect, and thus increase the SRF frequency. However, this will also have the unavoidable side effect of lowering the maximum possible available inductance, as well as the part's current-carrying ability.

An inductor parameter that is especially important for any tuned circuit is the *unloaded Q*, or quality factor, of the inductor. The *Q* indicates the quality of the inductor at a certain test frequency, with the *Q* equaling the inductive reactance divided by the combined DC series resistance, its core losses, and the skin effect of the coil. As the frequency through the inductor increases the *Q* will also increase, but at a particular frequency the *Q* of an inductor will begin to decrease rapidly due to the skin effect raising its resistance. The inductor's *Q* will soon reach its lowest point at the component's SRF frequency (Fig. 1.9).

The coil's DC series resistance is the amount of physical resistance created by the inductor's own internal or external conductive elements, and influences not only the *Q* of a coil as mentioned above, but will also drop a significant amount of any DC bias voltage. This fact is important when choosing a coil for a circuit that demands that the

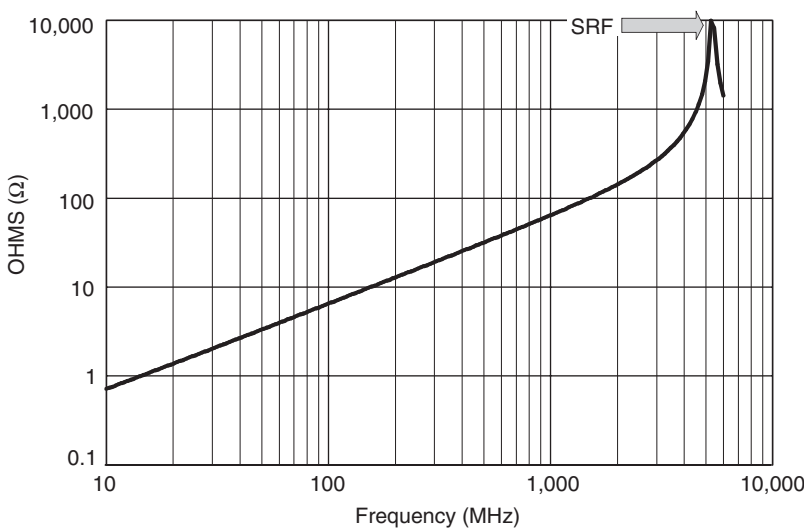


Figure 1.8 The impedance of a particular brand of SMD 10-nH inductor across frequency, showing its SRF.

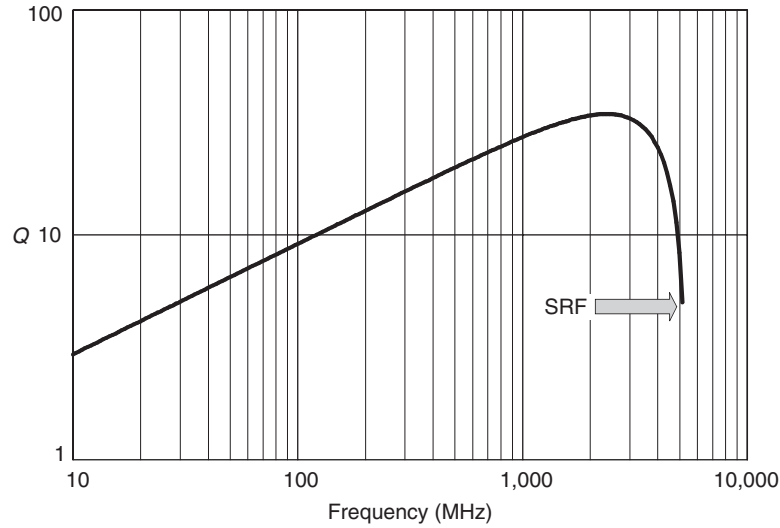


Figure 1.9 The *Q* of a particular brand of SMD 10-nH inductor across frequency, showing its SRF.

inductor itself must not cause an excessive DC voltage drop, as this can cause erratic circuit operation due to decreased voltages available to the biased active device.

The last major loss effect that creates problems in high-inductance coils at high frequencies is created by coil-form losses, which may become substantial due to hysteresis, eddy currents, and residual losses. As the frequencies continue to increase, the only acceptable type of inductor core material at microwave frequencies is typically that of the wire-wound air-core and the monolithic ceramic types.

Inductor Coil Design

There are times when the proper value or type of inductor is simply not available for a small project or prototype, and one must be calculated and constructed for non-microwave applications.

For a high-frequency, single-layer air-core coil (a *helix*), we can calculate the number of turns required to obtain a desired inductance with the formula shown below. However, the formula is only accurate for coils with a length that is at least half the coil's diameter or longer, while accuracy also suffers as the frequency is increased into the VHF region and above. This is due to excessive conductor thickness versus coil diameter.

$$n = \frac{\sqrt{L\left[(18d) + (40\ell)\right]}}{d}$$

where n = number of single-layer wire turns required to meet the desired inductance L [Only varnished (magnet) wire should be used in coil construction to prevent turn-to-turn shorts.]

L = desired inductance of the air coil, μH

d = diameter of the inside of the coil (the same diameter as the form used to wind the coil), in

ℓ = length of the coil, in (If this length is not met after winding the turns, then spread the individual coils outward until this value is reached.)

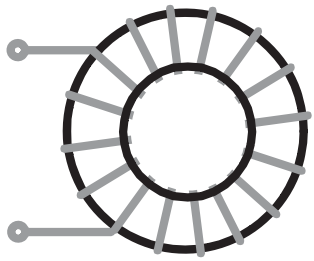


Figure 1.10 A toroidal core inductor.

Toroids

Inductors that are constructed from doughnut-shaped powdered iron or ferrite cores are called toroids (Fig. 1.10). Ferrite toroidal cores can function from as low as 1 kHz all the way up to 1 GHz, but the maximum frequency attainable with a particular toroid will depend on the kind of ferrite material employed in its construction. Toroids are found mainly in low to medium power, low RF frequency designs.

Since they will exhibit only small amounts of flux leakage, toroidal inductors are valuable components, and are thus far less sensitive to coupling effects between other coils and the toroid inductor itself. This circular toroidal construction prevents the toroid from radiating RF into surrounding circuits, unlike air-core inductors (and transformers) that may require some type of shielding and/or an alteration in their physical positioning on the PCB. Since almost every magnetic field line that is created by the primary makes it to the secondary, toroid transformers are also very efficient. Standard nontoroid air-core transformers do not share these abilities.

At low frequencies, toroids are used to remove hum from reaching a wireless receiver from the noisy main's power supply, as well as decreasing any wireless transmitter-generated interference from entering these same power lines. This is accomplished by inserting toroidal inductors in series with the main's supply, choking out most of the undesired "hash" emanating from these sources.

Toroids are identified by their outer diameter and their core material. For instance, an FT-23-61 core designation would indicate that the core is a ferrite toroid (FT), with an outer diameter of 0.23 in, and comprised of a 61-mix type of ferrite material. A simple "T" designation (instead of an FT) would indicate a powdered iron core, as opposed to a ferrite core.

Toroid Coil Design

As mentioned above, powdered iron toroidal inductor cores are available up to 1 GHz. To design and wind an iron toroidal inductor or choke, the A_L must be found on the core's data sheet. A_L symbolizes the value of the inductance in microhenrys when the core is wrapped with 100 turns of single-layer wire. All the inductor designer need do in order to design a powdered iron toroidal coil is to choose the core size that is *just* large enough to hold the required number of turns:

$$N = 100\sqrt{\frac{L}{A_L}}$$

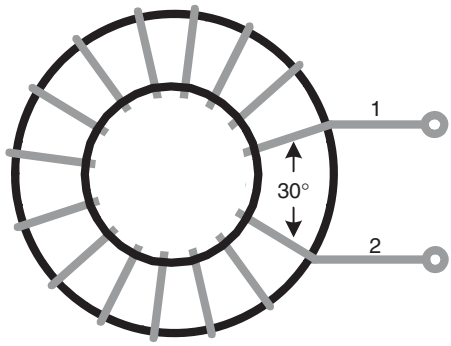


Figure 1.11 A toroid inductor, showing the proper winding around the core.

where N = number of single-layer turns for the desired value of L
L = desired inductance for the coil, mH
A_L = value, as read on the core's data sheet, of the chosen size and powdered iron mix of the core, mH per 100 turns

Or if designing for a *ferrite* toroidal core, then use the formula:

$$N = 1000\sqrt{\frac{L}{A_L}}$$

where N = number of single-layer turns
L = inductance desired, mH
A_L = value, as read on the core's data sheet, of the chosen core size and ferrite mix, mH per 1000 turns

Note: *A_L values have a tolerance of typically ± 20%. The core material must never become saturated by excess power levels, either DC or AC.*

Wind a single layer toroidal inductor or transformer with a 30° spacing between ends 1 and 2, as shown in Fig. 1.11, to minimize distributed capacitance, and thus to maximize inductor Q. The chosen mix for the core determines the core's maximum operating frequency.

1.1.5 Ferrite Beads

Above roughly 100.MHz, ferrite beads will dissipate undesired RF and high-frequency noise as heat ($R > X_L$), just as a resistor would, but have little affect on lower AC frequencies and DC. Thus, desired audio frequencies and DC power easily passes through these components with little loss.

Ferrite beads are rated for their frequency of operation (some models can function up to 1 GHz), as well as their RF impedance, DC resistance (DCR), and maximum DC current (in milliampere), and are readily available in both surface-mount and through-hole packages. Their equivalent circuit is shown in Fig. 1.12.

A typical RF bead's impedance characteristics over frequency are as shown in Fig. 1.13. Such beads are available with series impedances ranging from 10 to over 1000 Ω (at 100 MHz), depending on the particular model. Beads inherently do not

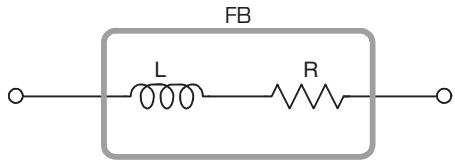


Figure 1.12 The equivalent circuit of a ferrite bead.

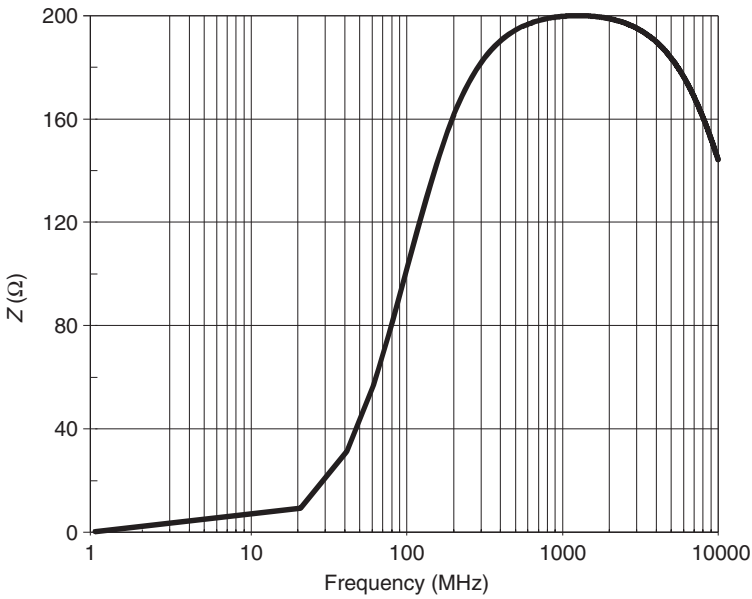


Figure 1.13 A particular ferrite bead's impedance over frequency.

support spurious or resonant modes, nor do they contribute to an active device's oscillatory behavior (as an inductor readily can). This positive attribute is due to their low inductance and high resistance characteristics at high frequencies, as inferred above.

To employ ferrite beads in a wireless design, simply select a part with the required impedance at the desired frequency, below a preferred DC resistance (DCR), and within a maximum mandatory DC bias current rating. Small ferrite beads should be used only in lower DC current draw applications in order to prevent ferrite saturation, since saturation would decrease the bead's rated impedance at a particular frequency.

1.1.6 Transformers

RF transformers are typically purchased as a complete component, but can also be constructed in toroidal form (Fig. 1.14). Toroids have replaced most aircores as interstage transformers in the majority of low-frequency radio designs (Fig. 1.15).

With the proper core material, toroidal transformers are effective up to 1 GHz. As these transformers are subjected to an increase in frequency, however, the capacitance between the transformer's windings begins to become more of a limiting factor. This

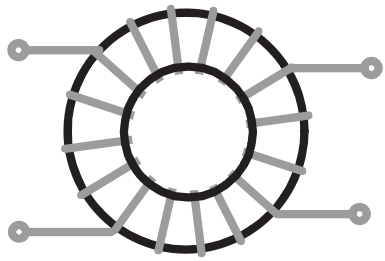


Figure 1.14 A toroid used as a transformer.

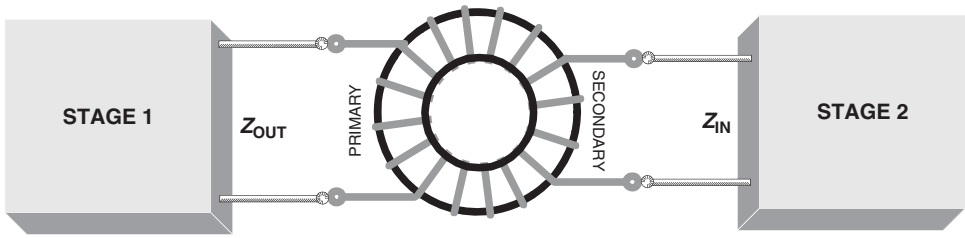


Figure 1.15 Impedance matching with a toroidal transformer.

internal capacitance will decrease the transformer's maximum operating frequency, since the signal to be transformed will now simply pass right through the transformer. This effect can be minimized by choosing a high-permeability transformer core, which permits fewer turns for the very same desired reactance, thus forming less distributed capacitance between the transformer's input and output.

Toroidal Transformer Design

For proper toroidal transformer operation the reactances of the primary and secondary windings must be four or more times greater than the source and loads of the transformer at the lowest frequency of operation. As an example, if a 1:1 transformer's primary had a 50-Ω amplifier attached to its input, and the secondary had a 50-Ω antenna at its output, then the primary winding's reactance (X_P) should be at least 200 Ω, while the secondary winding's reactance (X_S) of the transformer should also be 200 Ω at its lowest frequency of operation.

To design a toroidal transformer, first calculate the required reactances of both the primary and the secondary of the transformer at its lowest frequency:

1.

$$X_P = 4 \times Z_{OUT} \qquad \text{and} \qquad X_S = 4 \times Z_{IN}$$

where X_P = required primary reactance at the lowest frequency of transformer operation, Ω
Z_{OUT} = output impedance of the prior stage, Ω
X_S = required secondary reactance at its lowest frequency, Ω
Z_{IN} = input impedance of the next stage, Ω

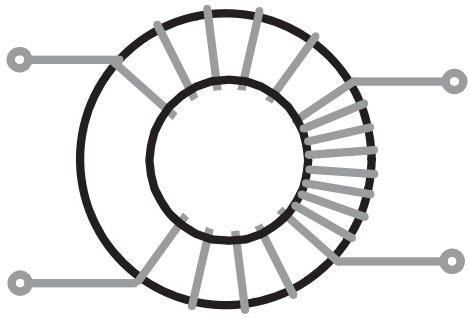


Figure 1.16 Proper winding for a toroidal transformer.

2. Now, calculate the inductance of the primary and secondary windings:

$$L_P = \frac{X_P}{2\pi f_{\text{LOW}}} \qquad \text{and} \qquad L_S = \frac{X_S}{2\pi f_{\text{LOW}}}$$

3. Choose a core that can operate at the desired frequency, with a high permeability, and as small a size as practical. Then, calculate the number of primary and secondary turns required:

$$N_S = 100\sqrt{\frac{L_S}{A_L}}^{*} \qquad \text{or} \qquad N_S = 1000\sqrt{\frac{L_S}{A_L}}^{*}$$

and then

$$N_P = N_S\sqrt{\frac{L_P}{L_S}}$$

4. Now wind the primary as a single layer around the entire toroid, while the secondary should be wound over the top of the primary winding at one end (Fig. 1.16). Reverse the windings for a step-up transformer.

1.2 Semiconductors

1.2.1 Introduction

Semiconductors, as opposed to the vacuum tubes of the past, are small, dependable, rugged, and need only low bias voltages. These active devices are utilized to not only amplify signals, but to also mix and detect such signals, as well as generate these signals through oscillation. Indeed, integrated circuits, and thus most modern wireless devices, would not even be possible without semiconductors. The following is a quick overview of the most common semiconductor components used in electronics.

*Formula for N_S will depend on how A_L is given in data sheet: 100 for microhenry; 1000 for millihenry.

1.2.2 Diodes

PN Junction Diode

A positive-negative (PN) junction diode (Fig. 1.17) is composed of both N- and P-type semiconductor materials that have been fused together. The N-type material will contain a surplus of electrons, called the *majority carriers*, and only a small number of holes, the *minority carriers*. The reason for this overabundance of electrons and lack of holes is the insertion of impurities, called *doping*, to the pure (or *intrinsic*) semiconductor material. This is accomplished by adding atoms that have five outer shell, or *valence*, electrons, as compared to the four valence electrons of intrinsic silicon.

The P-type material will have a surplus of holes and a deficiency of electrons within its crystal lattice structure due to the doping of the intrinsic semiconductor material with atoms that contain three valence electrons, in contrast to the four valence electrons of pure silicon. Thus, P-type semiconductor current is considered to be by hole flow through the crystal lattice, while the N-type semiconductor's current is created by electron flow.

In a diode with no bias voltage (Fig. 1.18), electrons are drawn toward the P side, while the holes are attracted to the N side. At the fused PN junction a *depletion region* is created by the joining of these electrons and holes, generating neutral electron-hole pairs at the junction itself, while the depletion region on either side of the PN junction is composed of charged ions. If the semiconductor material is silicon, then the depletion region will have a barrier potential of 0.7 V. This depletion region will not increase above this 0.7 value, however, since any attempted rise in majority carriers will now be repulsed by this same barrier voltage.

When a voltage is applied to the PN junction of sufficient amplitude, and of the suitable polarity, the semiconductor diode junction will be forward biased (Fig. 1.19). This will cause the barrier voltage to be neutralized, and electrons will then be able to flow freely. The bias, consisting of the battery, has a positive terminal which repulses the holes but attracts the electrons, while the negative battery terminal repels the electrons into the positive terminal. This action produces a current through the diode.

If a reverse bias is applied to a diode's terminals, as shown in Fig. 1.20, the depletion region will begin to enlarge. This is caused by the holes being attracted to the battery's negative terminal, while the positive terminal draws in the electrons, forcing the diode to function as a very high resistance. Except for some small leakage current, very little current will now flow through the diode. The depletion region will continue to expand

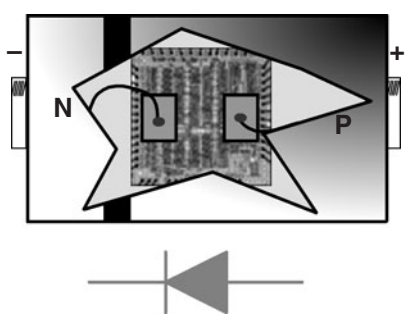


Figure 1.17 The semiconductor diode, showing internal die and bond wires.

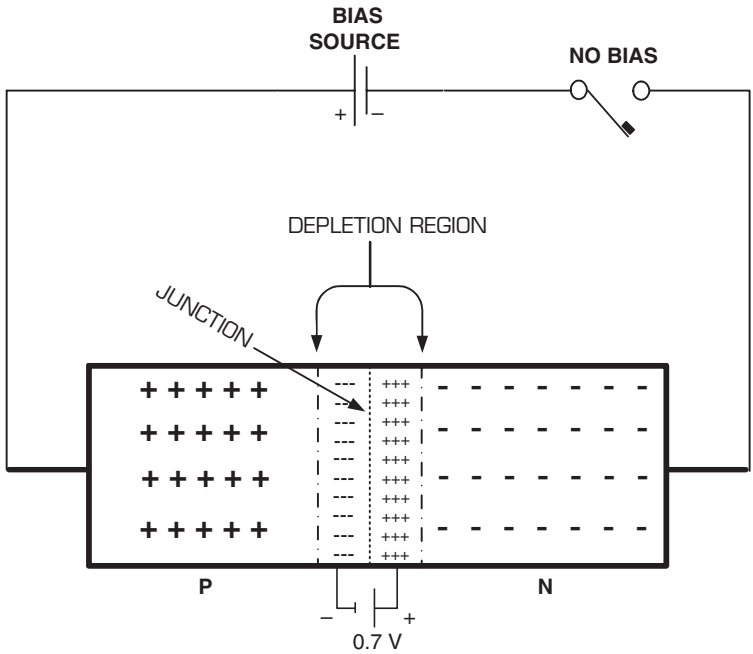


Figure 1.18 A diode shown with zero bias and its formed depletion region.

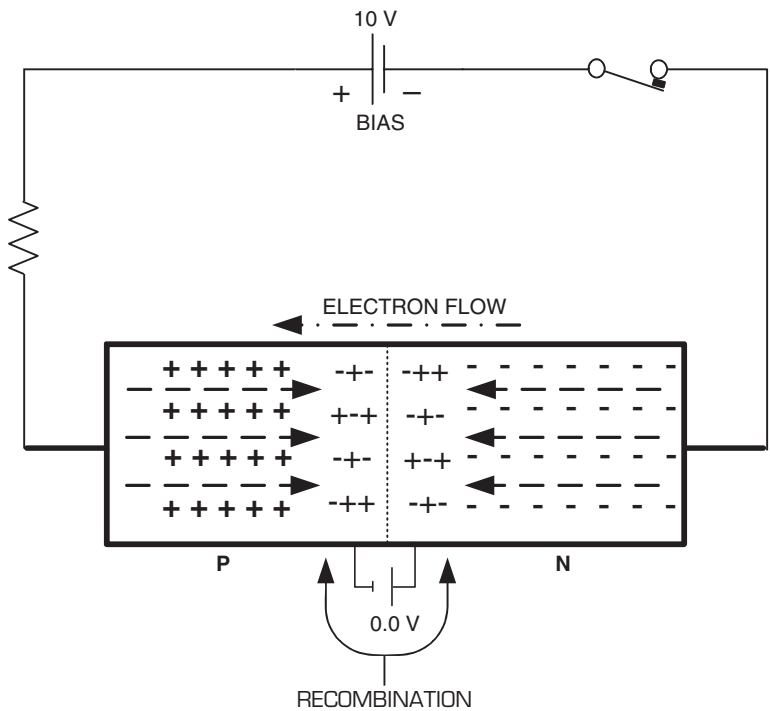


Figure 1.19 A diode with sufficient forward bias to conduct electrons.

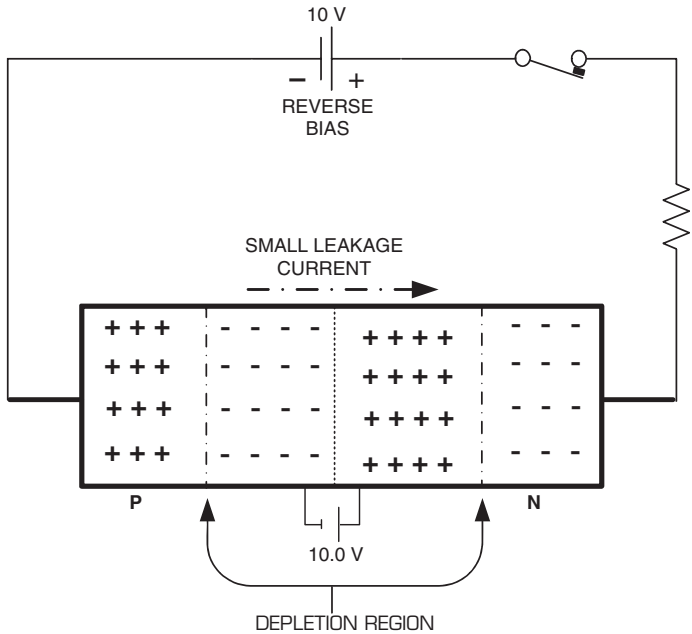


Figure 1.20 A diode with reverse bias applied and the resultant reverse leakage current flow.

until the barrier potential equals that of the bias potential, or until breakdown occurs, causing unchecked reverse current flow, and therefore damaging or destroying the diode itself.

As shown in the characteristic curves for a typical silicon diode (Fig. 1.21), roughly 0.7 V will invariably be dropped across a forward-biased silicon diode, no matter how

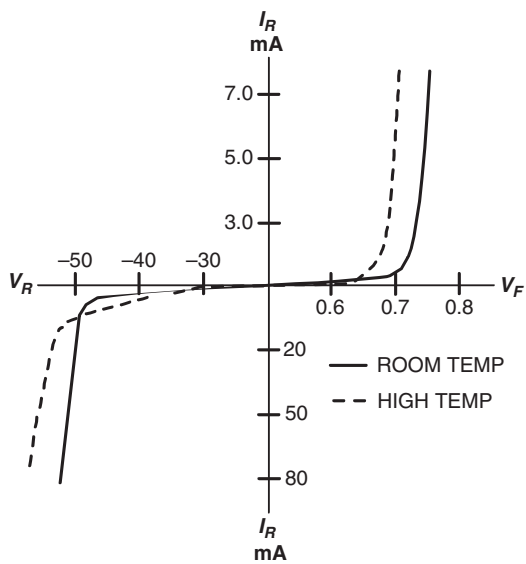


Figure 1.21 The characteristic curves of a silicon diode.

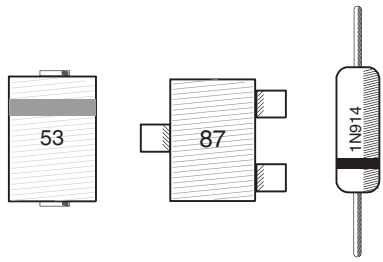


Figure 1.22 Three common small-signal diode packages.

much its forward current increases. This is because of the small value of dynamic internal resistance inherent in the diode's semiconductor materials.

Miniature glass and plastic diode packages (Fig. 1.22) are utilized for low-current circuits, while power diodes are used for high forward currents of up to 1500 A.

These are some of the more important rectifier diode specifications:

$I_{F(MAX)}$: Maximum forward current that can flow through the diode before its semiconductor material is damaged.

I_R: Diode's temperature-dependent reverse leakage current while in reverse bias.

PIV: Reverse-biased diode's *peak inverse voltage*, which is the maximum reverse voltage that should be placed across its terminals.

Zener Diode

The Zener diode (Fig. 1.23) uses a diode's capability to operate with reverse bias until *avalanche* (reverse breakdown) results, but without the Zener being destroyed in the process. This ability to safely operate in reverse breakdown is a huge advantage, since any changes in current through the Zener, no matter how large, will not affect the voltage dropped across the diode (Fig. 1.24), thus making the Zener an excellent choice for voltage regulation and voltage reference circuits.

At what voltage the Zener falls into avalanche is governed by its *Zener voltage* (V_Z). But each diode, even when rated at the same V_Z, will hit this point at a slightly different voltage, which is why there are different tolerances available for the Zener diode (20%, 10%, 5%, and 1% are common).

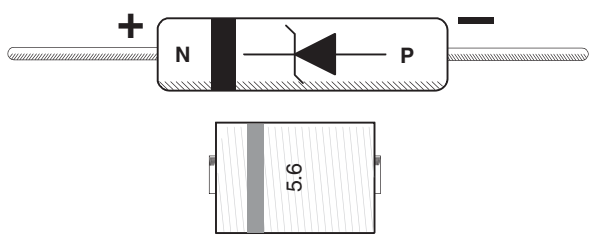


Figure 1.23 The Zener diode, showing SMD and through-hole packages.

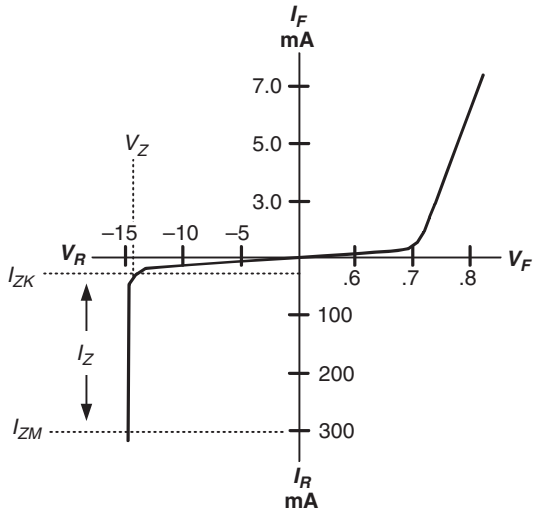


Figure 1.24 The characteristic curves of a Zener diode.

In some critical circuits it must be considered that a Zener's voltage ratings change over temperature. More temperature stable Zeners are available, such as *voltage reference diodes* and *temperature-compensated Zener diodes.*

The following are a few of the more important Zener diode specifications:

I_{ZM}: Maximum Zener current before the diode is destroyed.

V_Z: Reverse voltage across the Zener that changes very little with an increase in current.

I_Z: Zener current required to maintain the diode within its V_Z region.

P_D: Maximum approved power dissipation for the diode.

Varactor Diodes

As in the above Zener diodes, varactor diodes (Fig. 1.25) also operate under reverse bias. However, varactor diodes act as voltage-variable capacitors. And since we know that increasing the width of the dielectric in a capacitor will decrease its capacitance, and decreasing the width will increase its capacitance, we can use a similar effect to our advantage in varactor diodes. Increasing the reverse bias across the varactor increases the thickness of its depletion region, with this depletion region acting as *dielectric*, and thus decreasing the diode's capacitance (Fig. 1.26). Decreasing the reverse bias voltage will have the exact opposite effect, and will increase the diode's capacitance by decreasing the depletion region.

Figure 1.27 shows the capacitance variations versus the diode's reverse voltage for one kind of varactor, the *hyperabrupt* type. The other common type is called an *abrupt* varactor. Both are commonly used in *voltage-controlled oscillators* (VCO), and are readily available in many diverse capacitance values for almost any RF application.

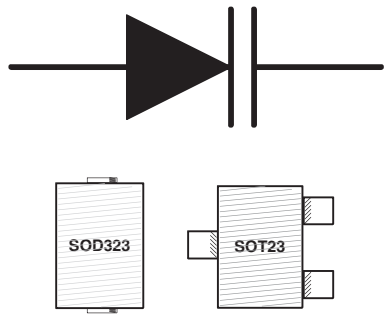


Figure 1.25 Schematic symbol, with common packages, for a varactor diode.

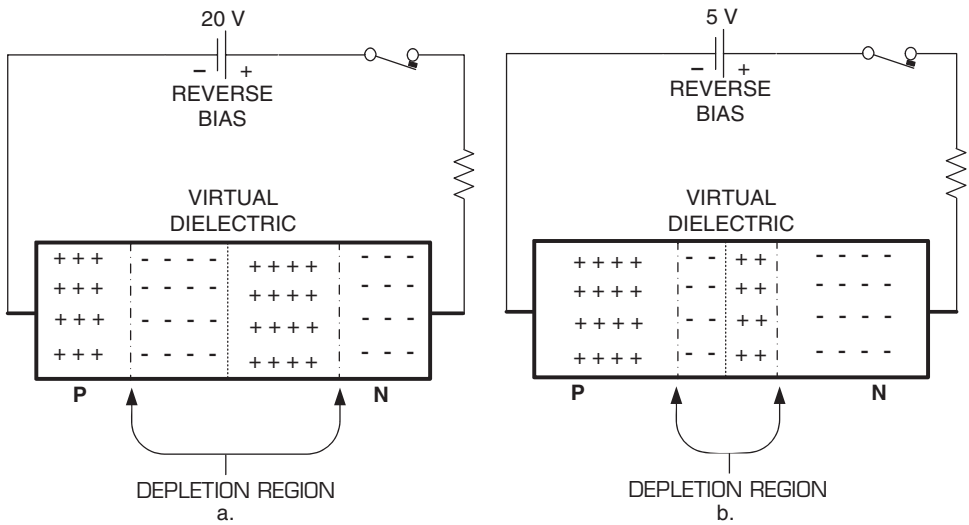


Figure 1.26 The formation of the virtual dielectric in a varactor diode with two different reverse bias voltages: (*a*) Low capacitance; (*b*) High capacitance.

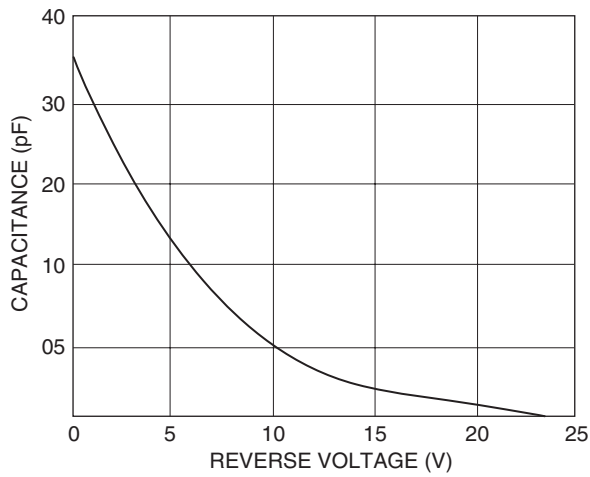


Figure 1.27 Capacitance versus the applied reverse voltage for a hyperabrupt varactor diode.

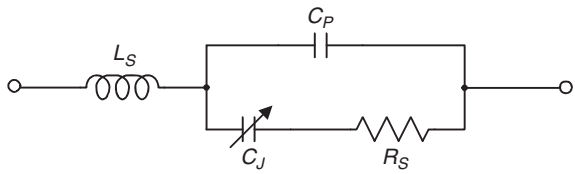


Figure 1.28 Equivalent circuit of a varactor diode.

The abrupt's main positive characteristic is a high unloaded *Q* (i.e., a low series resistance). This is a valuable characteristic, since the higher the *Q* of an oscillator tank circuit, of which the varactor is a very critical part, the lower the phase noise output of the VCO. Also, in voltage-variable filter applications, the higher the *Q* of the varactor, the tighter the filter's response characteristic, and the lower the filter's insertion loss will be.

However, an abrupt varactor has two major disadvantages that limits its use in many modern wideband VCOs, and that is their low tuning capacitance range versus tuning voltage, along with a high required DC tuning voltage. Indeed, the abrupt's tuning voltage must be as large as 1 to 30 V in order to obtain a small 5- to 18-pF capacitance change. Some abrupts may require a tuning voltage all the way up to 60 V.

The hyperabrupt diode is more popular in today's low-voltage, wideband environment. These varactors possess the advantage in that they can tune over a very wide capacitance range with a low tuning voltage (such as, from 10 to 200 pF with a 0.8 to 10 V tuning voltage), and they have superior linear voltage versus capacitance characteristics than abrupt types. Still, a hyperabrupt's main negative attribute is its relatively low *Q* in comparison to the abrupt type, along with the inferior VCO phase noise characteristic attendant with this lower *Q*.

A varactor's capacitance will vary somewhat over temperature, but this will normally not be of any concern for its most common application, which is within a VCO's tank circuit in a phase-locked loop (PLL) synthesizer. The PLL loop will automatically compensate for any such temperature drifts in the varactor's capacitance.

A varactor's equivalent circuit is shown in Fig. 1.28, and consists essentially of parasitic series inductance (L_S), parasitic parallel capacitance (C_P), voltage-variable capacitance (C_J), voltage-variable resistance (R_S).

Positive Intrinsic Negative Diodes

Positive intrinsic negative (PIN) diodes are constructed of a thin intrinsic layer sandwiched between positive and negative doped layers. Above certain frequencies (greater than 50 MHz) PIN diodes do not act as normal PN junction rectifier diodes, but as current-controlled resistors. Their *carrier lifetime* specification decides the diode's low frequency limit, under which the PIN begins to function simply as a normal small signal junction diode.

PIN diodes can be operated as RF switches and attenuators, and will have a much lower On resistance than PN junction rectifier diodes. PINs can be biased to supply a wide resistance range all the way down to 0.5 Ω and up to 10 kΩ with the application of a DC control current. When employed as an electronic switch, this control current is switched On or Off, thus forcing the PIN to a very low resistance (On), or to a very high

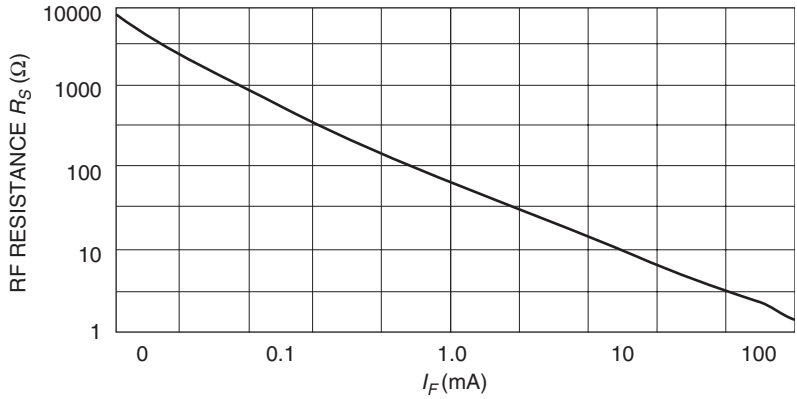


Figure 1.29 Forward bias current and RF resistance for one particular model of PIN diode.

resistance (Off), depending on the bias voltage. When a PIN is used in an attenuator circuit, this control current can be changed in a linear manner, in nondiscrete steps, forcing the PIN to alter its resistance anywhere from between its lowest to highest rated resistance values. Figure 1.29 displays a typical PIN diode's forward bias current and its resultant series RF resistance (R_S).

Schottky Diode

The Schottky diode is constructed of a metal that is deposited on a semiconductor material, creating an electrostatic boundary between the resulting Schottky barrier. These diodes can be found in microwave detectors, double-balanced modulators, harmonic generators, rectifiers, and mixers. Some Schottky diodes can function up to 100 GHz, have a low forward barrier voltage, and are mechanically sturdy.

Zero-bias Schottkys are a type of diode with a very low forward voltage. Figure 1.30 displays their *I-V* curves showing their forward voltage and the resultant forward current.

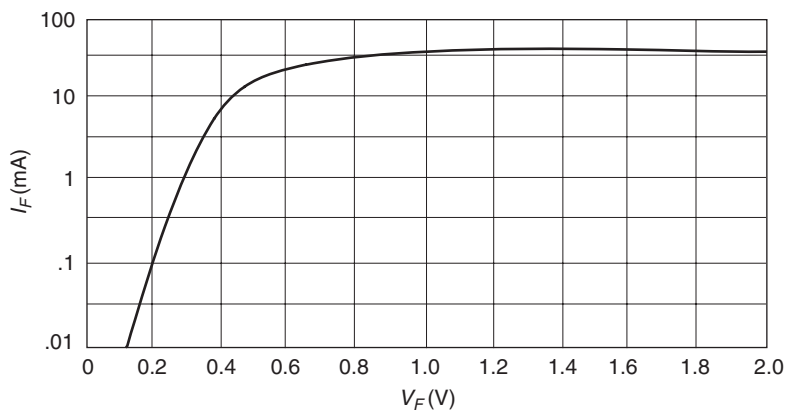


Figure 1.30 Zero-bias Schottky diode *I-V* curves showing forward voltage and the resultant forward current.

Gunn Diodes

Gunn diodes can be forced to function as an oscillator at microwave frequencies. The transit time of an electron through the Gunn diode determines the actual frequency of oscillation and, when the diode is inserted into a suitable resonant cavity, the Gunn device can oscillate at frequencies of up to 100 GHz. However, the higher the frequency of the Gunn, the thinner it must be, which lowers its power dissipation abilities.

Step-recovery diodes

A *step-recovery diode (SRD)* is a special diode employed in some microwave-frequency multiplication circuits. The SRD functions in this role by switching between two opposite impedance conditions, low and high. This change of state may occur in only 200 ps or less, thus discharging a very narrow pulse of energy. An SRD can best be visualized as a capacitor that stores a charge, then discharges it at a very rapid rate, creating a waveform that is plentiful in harmonics. Due to their high cost, SRDs are not as popular as they once were.

1.2.3 Transistors

Bipolar Junction Transistor

A *bipolar junction transistor (BJT)* is constructed of negative-postive-negative (NPN) or positive-negative-positive (PNP) doped regions, with the NPN type being by far the most common. The *emitter* of the transistor provides the charges, while the *base* controls these charges. The charges that have not entered the base are gathered by the *collector*.

Figure 1.31 reveals a silicon NPN transistor that has its emitter and base forward biased, with the collector reversed biased, to form a simple solid-state amplifier. The negative terminal of the emitter-base battery repels the emitter's electrons, forcing them into the ultra-thin base. But this narrow-base structure, due to the small amount of holes available for recombination, cannot possibly support the large number of electrons coming from the emitter. This is why base current is always of such a small value, since the majority of the electrons, over 99%, are attracted by the positive potential on the much larger collector, where they continue to flood into the collector's positive bias supply. This action is what forms the transistor's output current.

From the forgoing explanation, we see that $I_E = I_B + I_C$ and $I_B = I_E - I_C$; meaning that the currents through a transistor are completely proportional. Thus, if the emitter current doubles, then so will the currents in the base and the collector. But more important, this also means that if a small external bias or signal should increase this small base current then a proportional, but far greater, emitter and collector current will flow through the transistor. If this collector current is sent through a high output resistance, this action will produce voltage amplification.

The input port of a common-emitter transistor has a low resistance due to its forward bias, so any signal inserted into the base-emitter junction will be across this low input resistance, thus causing the bipolar transistor to be current controlled by both the DC bias and any external signal voltages. This is shown in the BJT's characteristic curves of

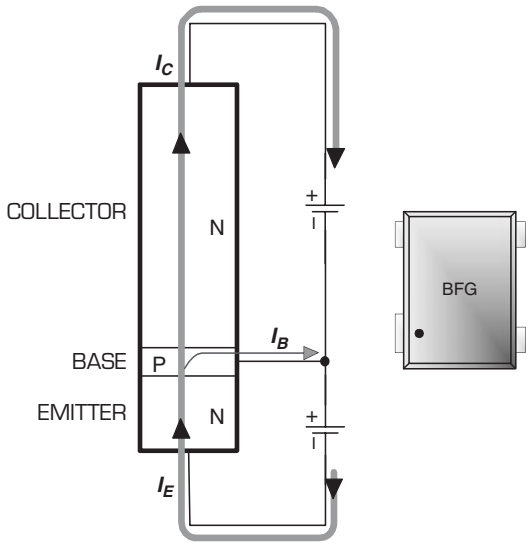


Figure 1.31 The current flow through the emitter, base, and collector of a bipolar NPN transistor, with its typical SMD package.

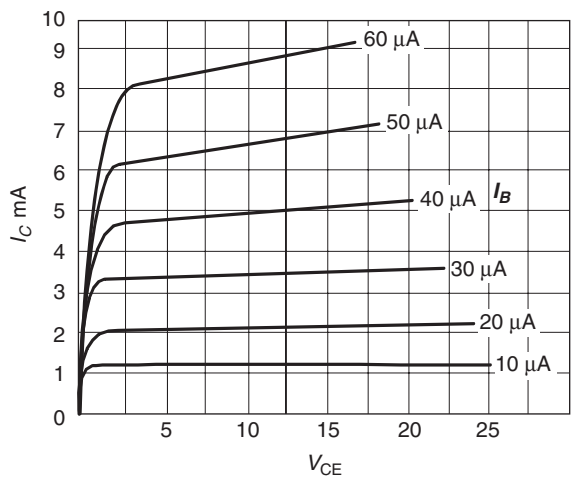


Figure 1.32 The characteristic curves for a bipolar transistor.

Fig. 1.32. The input signal, such as an RF or audio signal, will add to or subtract from this DC bias voltage across the transistor.

Before significant collector current can flow, the transistor's emitter-base barrier voltage of approximately 0.6 V_{BE} (for silicon) must be overcome. This particular task is performed by the base bias circuit. In a linear amplifier, the transistor's initial operating point is set by the bias circuits to be around 0.7 V, so as to permit any incoming signal to swing both above and below this value. The region of active amplification of a BJT is

only about 0.2 V wide, so any voltage between *saturation* (0.8 V) and *cutoff* (0.6 V) is the only area that a semiconductor is capable of amplifying in a linear manner. Between these two V_{BE} values of 0.6 V and 0.8 V the I_B, and thus the I_C, is controlled. In other words, the saturation region is reached when the bias input RF signal is high enough to force the amplifier to conduct its maximum collector current. This is a highly nonlinear area of operation. Dropping down to the cutoff region biases the transistor so that it only conducts a small leakage current (an Off state). In this state, only a strong RF input signal can force the collector to conduct more current, but only in pulses. Again, a highly nonlinear area of operation.

Therefore, a BJT can be thought of as a current-controlled resistance, with a tiny base current controlling the transistor's resistance, which influences the much larger emitter-to-collector current. This collector current is then made to run through a high load resistance, generating an amplified output voltage.

A few of the more common transistor specifications found in a BJT's data sheet:

BV_{CBO}: C*ollector-to-base breakdown voltage* is the amplitude of collector voltage that will normally break down the collector junction.

$P_{D(MAX)}$: Maximum total power dissipation a transistor is capable of in an ambient air temperature of 25°C.

$T_{J(MAX)}$: Maximum internal junction temperature before the semiconductor material breaks down.

$I_{C(MAX)}$: Maximum collector current of the BJT.

f_T: C*urrent gain-bandwidth product* is the frequency that a common-emitter transistor will be at a *beta* of unity.

f_{ae}: *Beta cutoff frequency* is the frequency that the BJT's *beta* decreases to 70.7% of its low frequency value.

I_{CEO}: Temperature-dependant leakage current that occurs from the emitter to the collector with the base open.

f_{MAX}: *Maximum oscillation frequency* is the maximum frequency that the transistor is capable of any power amplification whatsoever, or a maximum power gain of unity (0 dB).

S_{21}: Magnitude of the transistor's forward gain in a 50-Ω system, typically quoted in dB (this S_{21} value may vary by up to ±20% from one transistor to another due to production lot tolerances).

Junction Field Effect Transistors

Since a junction field effect transistor's (JFET) input gates are always reverse biased, JFETs will have a very high low-frequency input impedance, so are thus voltage controlled. JFETs are also quite capable of receiving an input of up to several volts, as compared to the bipolar transistor's few tenths of a volt. In addition, JFETs create less internal noise than a BJT, but will display lower voltage gain and more signal distortion.

As shown in Fig. 1.33, the structure of a JFET is composed of a *gate*, a *source*, and a *drain*. The JFET's terminals are voltage biased in such a way that the drain-to-source voltage (V_{DS}) causes the source to be more negative than the drain. This lets the drain current (I_D) flow from the source to the drain through the N-channel.

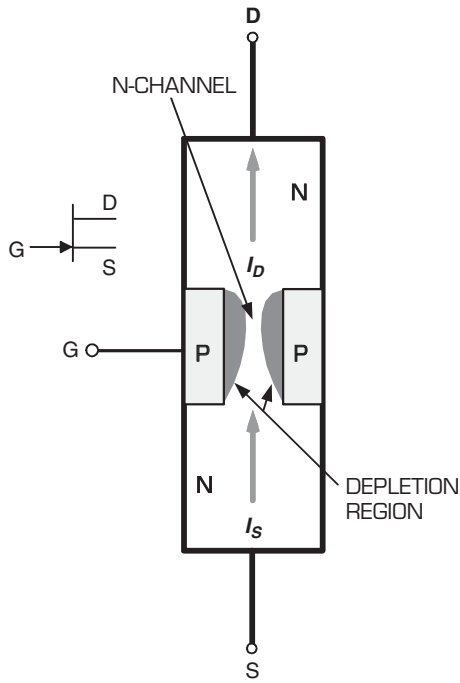


Figure 1.33 The internal structure of, and current flow through, a JFET.

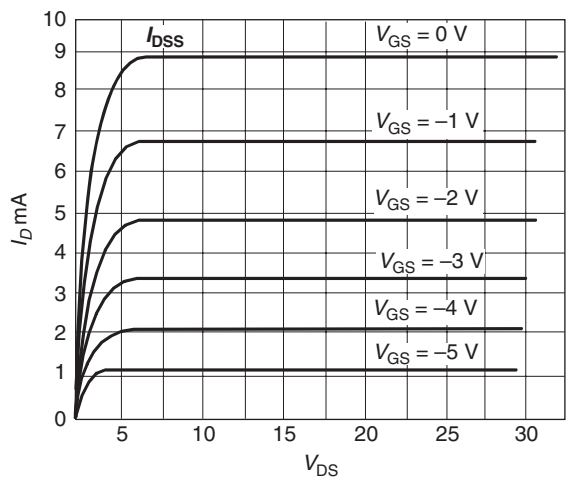


Figure 1.34 A JFET's characteristic curves.

The JFET characteristic curves of Fig. 1.34 readily indicate that a JFET is a normally On device when there is no bias voltage present at the gate. This permits the maximum current (I_{DSS}) to flow from the source to the drain. When the gate and source are presented with a negative voltage ($-V_{GS}$), an area lacking charge carriers (the *depletion region*) starts to form within the JFET's N-channel. This N-channel depletion region functions as an

insulator; so as the JFET becomes increasingly reverse biased and increasingly exhausted of any charge carriers, the N-channel continues to be narrowed by this developing depletion region. The channel's resistance rises, decreasing the JFET's current output into its load resistor, which lowers the device's output voltage across this resistor. As the negative gate voltage of $-V_{GS}$ is increased, the depletion region continues to widen, decreasing current flow even further. But a point is ultimately reached where the channel is totally depleted of all majority carriers, and no more decrease in current flow is possible. This lack of current flow is referred to as $V_{GS(OFF)}$. In short, the V_{GS} successfully controls the JFET's channel resistance, and thus its drain current. However, it is important that the drain-to-source voltage V_{DS} should be of a high enough amplitude to allow the JFET to operate within its linear region, or above pinch-off (V_P).

Pinch-off is simply an area where the drain current will stay constant, even if the drain-to-source voltage is increased; now only the gate-to-source voltage can affect the drain current. Pinch-off may also refer to the JFET's negative gate voltage value V_P that is needed to "pinch-off" all source-to-drain current flow to some particular low leakage value, and at a specific V_{DS}.

A few of the more common JFET parameters are

I_{DSS}: Maximum JFET drain current possible (with a V_{GS} at 0 V).

g_m *or* g_{fs}: T*ransconductance gain* (or $\frac{\Delta I_D}{\Delta V_{GS}}$), measured in siemens or mhos.

$V_{DS(MAX)}$: Maximum safe drain-to-source voltage.

V_P: Pinch-off voltage, which is the minimum V_{DS} required for the JFET's linear operation.

P_D: JFET's maximum power dissipation rating.

Metal-Oxide-Silicon Field-Effect Transistors

Metal-oxide-silicon field-effect transistors (MOSFETs) use a gate structure that is well insulated from the source, drain, and channel. This produces an active device with an almost infinite DC input resistance. However, this high input resistance is significantly decreased by its bias components, as well as during high-frequency operation. In fact, as the frequency of operation is increased, the MOSFET's input impedance approaches that of a BJT's.

MOSFETs are available in one of two modes: the *depletion-mode* type, a normally-On device, and the *enhancement-mode* type, a normally-Off device.

Drain current in a depletion-mode N-channel MOSFET (Fig. 1.35) is controlled through the application of negative and positive gate voltages (Fig. 1.36). By raising the negative voltage at the MOSFET's gate, we would soon reach a point where, due to the channel being depleted of all majority carriers, no significant drain current can flow. But as the gate-to-source voltage (V_{GS}) becomes less negative, more current will begin to flow. Even as we pass 0 V for V_{GS} the drain current will still continue to rise, since at zero V_{GS}, unlike the JFET, the maximum current for the device has not yet been reached. Nonetheless, the drain current is still quite substantial, since many majority carriers are present within the depletion MOSFET's N-channel. The V_{GS} increases until it reaches some maximum positive value. Now the maximum number of electrons have been drawn into the N-channel, and there is the maximum current flowing through the channel and into the drain.

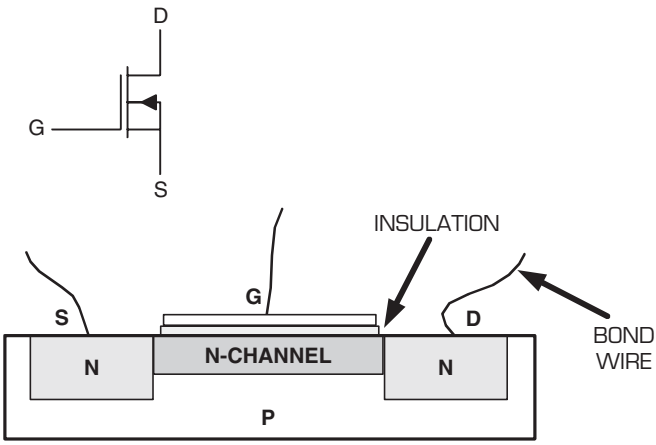


Figure 1.35 The internal structure of an N-channel depletion-mode MOSFET.

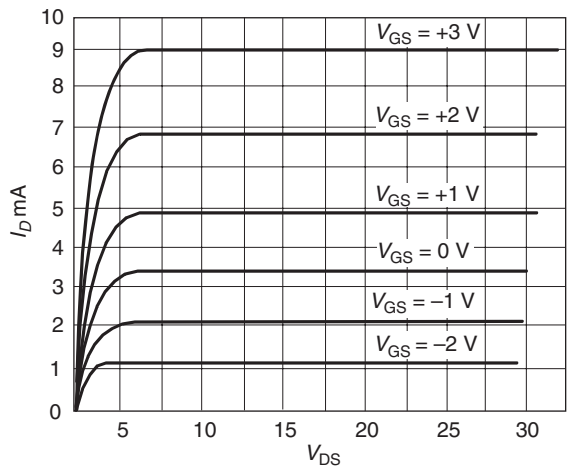


Figure 1.36 The characteristic curves of an N-channel depletion-mode MOSFET.

Depletion MOSFETs can be found operating extensively in wireless circuits due to their low noise characteristics. A similar structure, but employing two gates within a single device, is the dual-gate MOSFET (Fig. 1.37). Dual-gate MOSFETs are utilized in mixers and AGC-controlled amplifiers, with each of the MOSFET's gate inputs having an equal control over the drain current.

The other type of MOSFET, the enhancement-mode or E-MOSFET (Fig. 1.38) is, as mentioned above, a normally-Off transistor. So, almost no source-to-drain current flows when there is no bias across the E-MOSFET's gate, as shown in the characteristic curves of Fig. 1.39. However, almost any positive voltage that is placed across the gate will produce a channel between the device's source and drain (Fig. 1.40). Thus, as electrons are pulled to the gate, an N-channel is created within the P-type substrate. This action permits electrons to flow toward the positively charged drain, creating a continuous current flow.

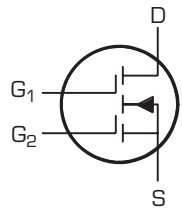


Figure 1.37 A dual-gate MOSFET's schematic symbol.

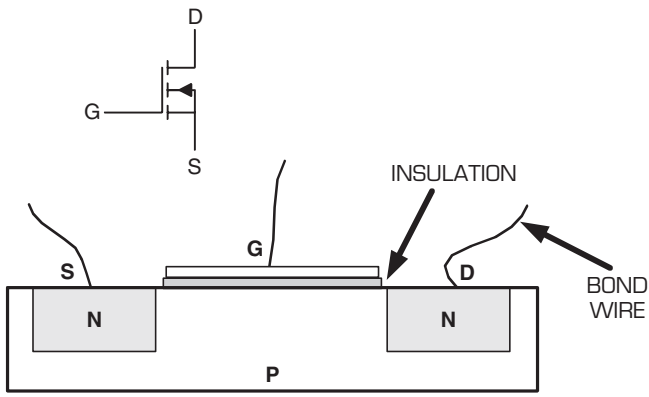


Figure 1.38 The internal structure of an enhancement-mode MOSFET.

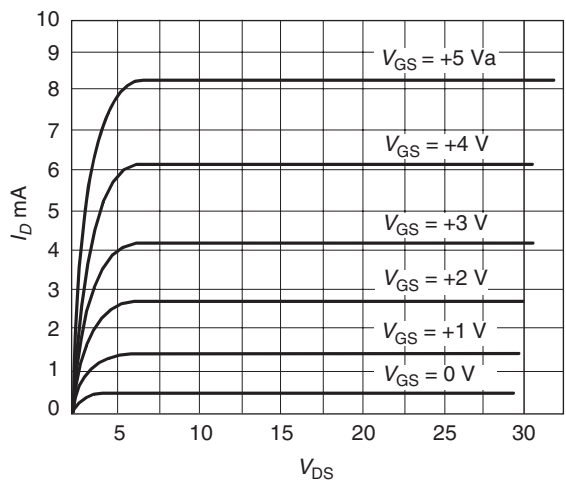


Figure 1.39 The characteristic curves of an enhancement-mode MOSFET.

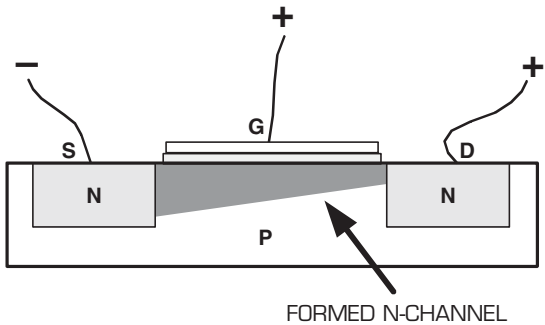


Figure 1.40 The formation of the N-channel in an E-MOSFET's substrate by a positive gate voltage.

Nonetheless, enhancement-mode MOSFETs will have a 1-V gate threshold voltage before any significant drain current will flow. E-MOSFET power devices must use a positive gate bias to overcome this initial gate threshold voltage in order to optimize gain and output power. This bias requirement means that, unlike a BJT, the input to an E-MOSFET cannot simply employ a zero gate bias to run in Class C power amplifier operation.

E-MOSFETs are popular in digital ICs as voltage-controlled switches, and are found as the active element in many HF, VHF, and UHF power amplifiers and drivers. The E-MOSFET's use in these particular applications is due to its superior parameters to that of a typical power BJT, such as higher input impedance and gain, increased thermal stability, lower noise, and a higher tolerance for load mismatches. Another advantage that any MOSFET enjoys over a BJT is its lack of any possibility of thermal runaway, as MOSFETs are designed to have a *positive temperature coefficient* at high drain currents. This means that, as the temperature increases, a MOSFET will actually decrease its source-to-drain current, instead of increasing its current output as a BJT will. This makes *thermal runaway* impossible, with temperature stabilization components being less necessary (except to stabilize the MOSFET's *Q*-point bias). As well, a MOSFET's input and output impedances will change much less for different input drive levels than a BJT's, along with the MOSFET's better single-stage stability and about 20% more gain. MOSFETs also have a very high VSWR survivability, second only to a BJT that employs an emitter ballast resistor. On the negative side, MOSFETs are very sensitive to destruction by static electricity, with almost any electrical spark causing possible damage to the gate insulation. N-channel enhancement MOSFETs, the most common in RF power applications, as well as depletion-mode power MOSFETs in Class C and B operation, can also begin to exhibit low-frequency oscillations if they are directly paralleled to increase their RF output level. MOSFETs also have inferior low-order IMD performance over that of a BJT.

1.3 Microstrip Design

1.3.1 Introduction

At microwave frequencies, microstrip (Fig. 1.41) is employed not only as a transmission line on PCBs, but also as equivalent passive components, tuned circuits, and high-*Q* microwave filters and resonators.

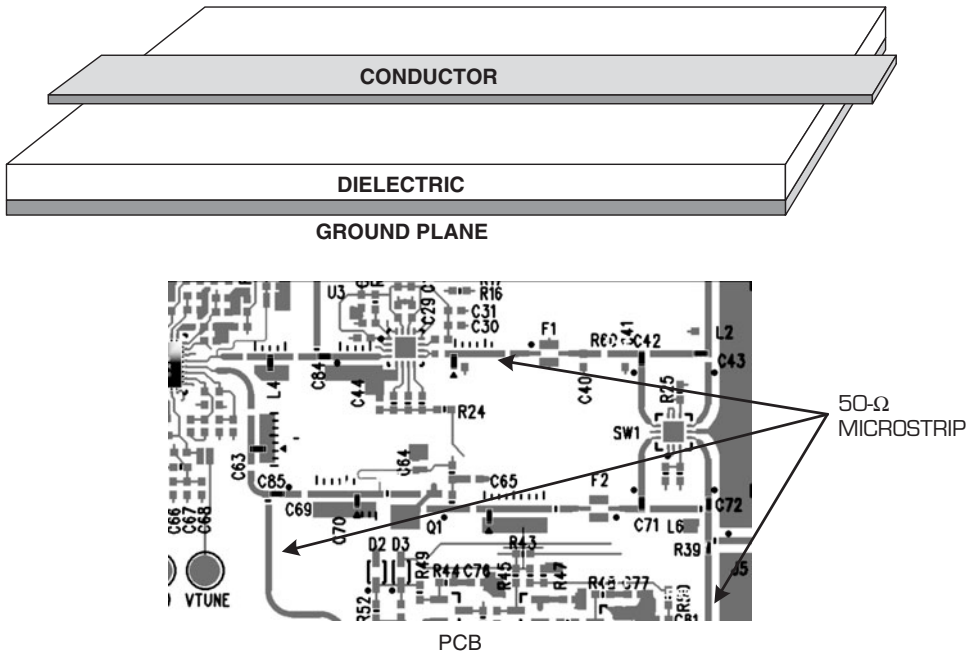


Figure 1.41 Microstrip, showing the dielectric and conductive layers.

Microstrip transmission line is used in microwave applications due to its low loss, low cost, small size, ease of implementation, and the ability to mount components, such as surface-mount capacitors, resistors, and transistors, directly onto the microstrip itself.

Most microstrips are unbalanced transmission lines and, because of their unshielded nature, can radiate some RF. However, radiations from properly terminated microstrips are fairly small. *Stripline* (Fig. 1.42) is similar to microstrip, but is placed between the metallization layers of a PCB and, due to these balanced twin groundplanes, does not radiate.

The characteristic impedance of microstrip transmission line is governed by the width of the conductor, the thickness of the dielectric, and the dielectric constant; with low impedance microstrip lines being wide, and high impedance microstrip lines being

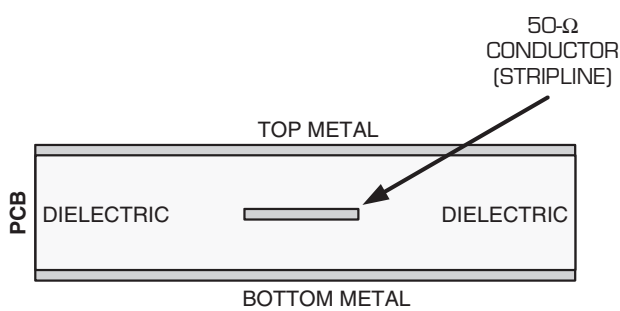


Figure 1.42 Stripline, showing the dielectric and conductive layers.

narrow. But the most important attribute of terminated microstrip as transmission line is that its impedance does not change with frequency, or with length. The normal characteristic impedances of microstrip and stripline are designed to be anywhere between 10 to 110 Ω, with 50 Ω being the universal norm for RF transmission line use. Microstrip is very common in frequencies of operation at 250 MHz and above.

1.3.2 Microstrip as Transmission Line

A 50-Ω microstrip is utilized in microwave circuits to prevent reflections and mismatch losses between physically separated components, with a calculated nominal width that will prevent the line from being either inductive or capacitive at any point along its length. In fact, with a source's output impedance matched to the microstrip, and the microstrip matched to the input impedance of the load, no standing or reflected waves will result. Consequently, there will be no power dissipated as heat, except in the actual resistance of the copper and dielectric as I^2R losses.

In microstrip, the dielectric constant (E_r) of the PCB's substrate material will not be the sole E_r that the microstrip transmission line itself "sees." This is due to the flux leakage into the air above the PC board, combined with the flux penetrating into the dielectric. So the actual *effective dielectric constant* (E_{EFF}), which is the true dielectric constant that the microstrip will now see, will be at some value between the surrounding air and the true dielectric constant of the PCB.

Due to the small RF field leakage that emanates from all microstrips, these types of transmission lines should be isolated by at least two or more line widths away from other traces and circuits in order to decrease any mutual coupling effects. To lower the chances of crosstalk even further an isolation ground trace may possibly be necessary between two such lines.

To decrease any impedance bumps at high microwave frequencies, microstrip should always be run as short and as straight as possible, with any angle using a mitered or slow round bend (Fig. 1.43).

Another issue to watch for when designing microwave circuits with microstrip transmission lines is the *waveguide effect* (see *RF Shielding Resonances*): Any metal enclosure that is used to shield the microstrip, or its source or load circuit, may act as a waveguide, and drastically alter the circuit's behavior. This effect can be eliminated by changing the width of the shield to cover a smaller area, or by inserting a special microwave foam attenuator material within the top of the enclosure.

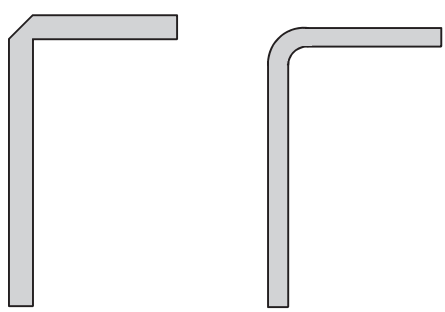


Figure 1.43 Proper way to work with bends in microstrip lines: (*a*) miter; (*b*) curve.

Microstrip Transmission Line Design

Use the following equation to plug in different microstrip widths to obtain the desired impedance:

$$Z_O = \frac{377}{\left(\frac{W}{h} + 1\right)\sqrt{E_r + \sqrt{E_r}}}$$

where Z_O = characteristic impedance of the microstrip, Ω
W = width of the microstrip conductor (use same units as h below)
h = thickness of the substrate between the groundplane and the microstrip conductor (use same units as W)
E_r = dielectric constant of the board material

1.3.3 Microstrip as Equivalent Components

Distributed components, such as inductors and capacitors, can be formed from microstrip transmission-line sections on PCBs at microwave frequencies. A series or shunt inductor can be formed from a thin trace (Fig. 1.44), a shunt capacitor can be formed from a wide trace (Fig. 1.45), and even a transformer can be formed by varying the width of the microstrip (Fig. 1.46).

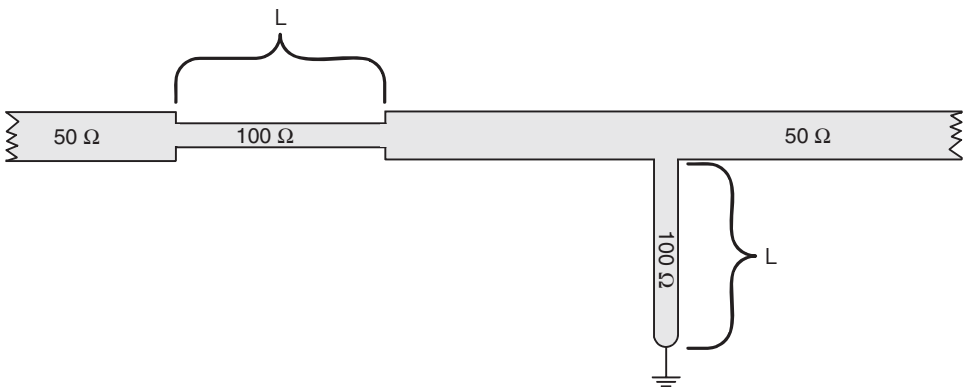


Figure 1.44 A distributed series and shunt inductor.

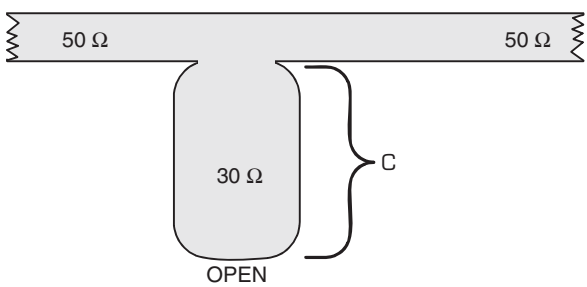


Figure 1.45 A distributed shunt capacitor.

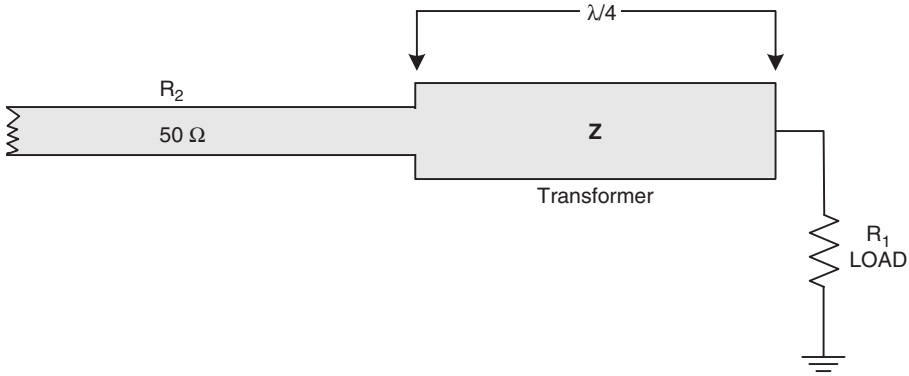


Figure 1.46 Using a distributed transformer for resistive matching.

Distributed equivalent microstrip works in replicating a lumped component because high impedance (thin) microstrip traces possess very little capacitance (due to the small surface area of the structure), but possessing a useful amount of inductance. This inductance is caused by the current flowing through the constricted microstrip element, which is both thin and long. In other words, we are taking a normal 50-Ω microstrip, with its built-in distributed inductance and capacitance that *makes* the microstrip line 50 Ω throughout its entire length, and simply minimizing the capacitance, while maximizing its inductance. The reverse is also true if we want to create a distributed capacitor, since the 50-Ω microstrip can now be formed as a wide strip, with the PCB's copper groundplane and its dielectric being located directly beneath it. Thus, the much larger surface area of this fatter microstrip (over its 50-Ω configuration) appears as a capacitor at RF frequencies. However, we cannot accurately, nor with any value above 1 pF or so, replicate *series* capacitors in distributed form, so we should select an appropriate lumped matching network that is devoid of any such series capacitors, and then convert it to a distributed structure.

The accuracy of the equivalency itself will only be exact for frequencies that are less than 30° long across the distributed equivalent shunt component. Equivalency is still possible with longer lengths, but keeping all equivalent component elements at less than 30° will supply the best performance. Maintaining this shorter length in the distributed component design will sometimes demand compensating, which we can do by narrowing the distributed inductor's trace width (increasing impedance), or widening the distributed capacitor trace width (decreasing impedance), to keep the total length under 30°.

Due to the increased losses and the lack of width repeatability during board fabrication, we cannot normally use equivalent inductors that are narrower than 6 mils.

Distributed Equivalent Component Design

As stated above, it is important to strive to make a distributed component shorter than 30° out of the 360° of an entire wavelength, or the *equivalent component effect* will begin to depart more and more from that of an ideal lumped component. To calculate how long 30° is out of 360°, simply divide 30 by 360, then multiply this value by the actual wavelength of the signal on the PCB, keeping in mind that the signal's wavelength in the substrate will not be the same as if it were traveling through a vacuum.

To find the actual wavelength of the signal, which is being slowed down by the PCB's substrate material, calculate the microstrip's *velocity of propagation* (V_p). First, find the *effective dielectric constant* (E_{EFF}) of the microstrip, since the signal will be partly in the dielectric and partly in the air above the microstrip, which affects the propagation velocity through this combination of the two dielectric mediums.

Step 1

$$E_{EFF} = \frac{E_r + 1}{2} + \left(\frac{E_r - 1}{2} \times \frac{1}{\sqrt{1 + \left(\frac{12h}{W} \right)}} \right)$$

where E_{EFF} = effective dielectric constant that the microstrip sees
E_r = actual dielectric constant of the PCB's substrate material
h = thickness of the substrate material between the top conductor and the bottom groundplane of the microstrip, in the same units as W
W = width of the top conductor of the microstrip, in the same units as h

Step 2

$$V_P = \frac{1}{\sqrt{E_{EFF}}}$$

where V_p = fraction of the speed of light as compared to light in a vacuum
E_{EFF} = effective dielectric constant of the PCB's substrate as calculated above

Step 3: Calculate the wavelength of the signal of interest in a perfect vacuum:

$$\lambda_{VAC} = 11800 \div f$$

where λ_{VAC} = wavelength of f in a true vacuum, mils
11800 = speed of light for f, mils
f = frequency of the signal of interest, GHz

Step 4: Multiply the velocity of propagation (V_p) times the wavelength (λ_{VAC}) of the signal as calculated above in order to arrive at the wavelength of the signal of interest (λ), in mils, when the signal is placed into the microstrip, or

$$\lambda = V_P \times \lambda_{VAC}$$

Distributed Parallel (Shunt) Capacitor

To design an equivalent distributed capacitor for microwave applications, follow these steps.

Step 1: Knowing the capacitance of the desired component for the circuit, calculate the reactance of the shunt capacitor required, at the frequency of interest, by the common formula:

$$X_C = 1 \div 2\pi fC$$

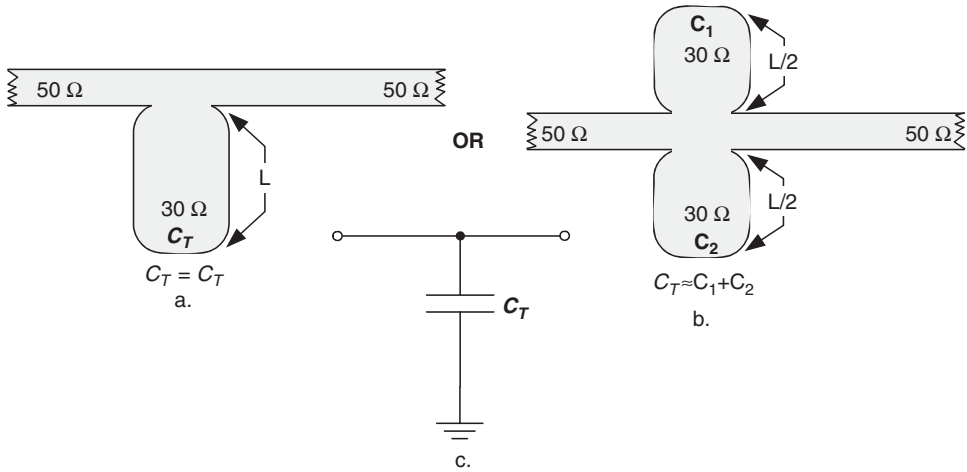


Figure 1.47 (*a*) Distributed shunt capacitor; (*b*) split into two shunts equaling the single capacitor; and (*c*) with an equivalent lumped shunt capacitor.

Step 2: Utilize 30-Ω microstrip ($Z_L = 30\ \Omega$) for the substrate's dielectric in use. Find the microstrip width required for this 30-Ω value by using either one of the many microstrip calculation programs available for free on the Web (such as HP's AppCad, AWR's TXLine, Daniel Swanson's MWTLC, and so on) or the formula above. As shown in Fig. 1.47, the microstrip of the equivalent shunt capacitor is open, and not grounded, at its end. The capacitor section is also attached to the 50-Ω microstrip transmission line by a small tapered section to improve the transition, with a further improvement possible by splitting the capacitor in two and placing it on both sides of the transmission line.

Step 3: Calculate the required microstrip's length to become a capacitor of value X_C, as calculated above, with this formula:

$$\frac{\left(\frac{30}{X_C}\ \text{ARCTAN}\right)}{360} \times \lambda = \text{Length}$$

where X_C = capacitive reactance required in the distributed circuit, Ω
Length = length of the microstrip required to imitate a lumped component of value X_C, mils (which should not be longer than 30°, or 12% of λ)
λ = wavelength of the frequency of interest *using the substrate of interest* (or $V_p \times \lambda$), mils

A Quick Example Design a Distributed Shunt Capacitor (Fig. 1.48)

Goal Create a shunt distributed capacitor to replace a lumped part. The frequency of operation is 5.8 GHz and the capacitance required is 0.5 pF. The substrate is FR-4 ($E_r = 4.6$), with a thickness of 20 mils.

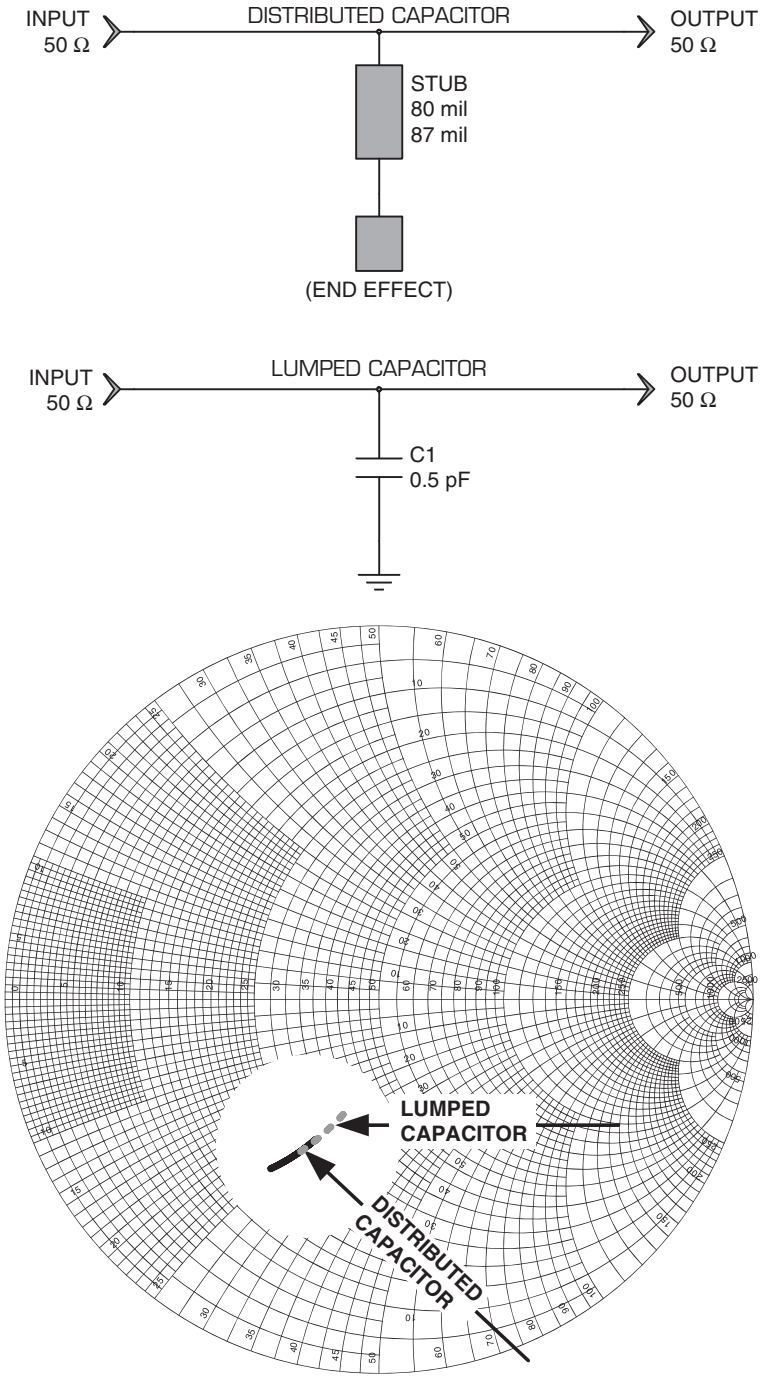


Figure 1.48 A shunt distributed capacitor vs. a shunt lumped capacitor over a narrow frequency range (5 to 6 GHz).

Solution

1. Capacitive reactance of a 0.5-pF capacitor is 55 Ω.
2. The width of the distributed capacitor's microstrip will be 80 mils.
3. The length will be 87 mils.

Series Inductor

As shown in Fig. 1.49, the equivalent series inductor is placed in series with the 50-Ω microstrip transmission line, or placed between other distributed or lumped components.

Step 1: Knowing the inductance required of the distributed inductor, calculate the reactance, at the frequency of interest, by the common formula:

$$X_L = 2\pi fL$$

Step 2: Utilize 100-Ω microstrip (Z_L = 100 Ω) for the substrate's dielectric in use. Find the microstrip width required for this 100-Ω value by either working with one of the many microstrip calculation programs available for free on the Web (such as HP's AppCad, or AWR's TXLine, or Daniel Swanson's MWTLC, and so on.), or by employing the microstrip formula above.

Step 3: Calculate the microstrip's required length to become an inductor of value X_L:

$$\frac{\left(\frac{X_L}{100}\ \text{ARCSIN}\right)}{360} \times \lambda = \text{Length}$$

where X_L = inductive reactance needed in the distributed circuit, Ω
Length = length of the microstrip required to imitate a lumped component of value X_L (should never be longer than 30°, or 12%, of λ), mils
λ = wavelength of the frequency of interest using the substrate of interest (or $V_P \times \lambda$), mils

A Quick Example Design a Distributed Series Inductor (Fig. 1.50)

Goal: Create a distributed series inductor to replace a lumped part. The frequency of operation is 5.8 GHz and the inductance required is 1 nH. The substrate is FR-4 (E_r = 4.6), with a thickness of 20 mils.

Solution:

1. Inductive reactance of a 1-nH capacitor is 36 Ω.
2. The width of the distributed inductor's microstrip will be 7 mils.
3. The length will be 69 mils (using ARCSIN).

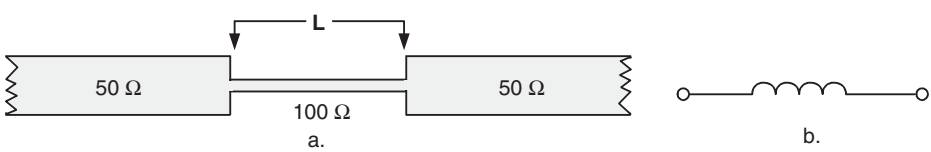


Figure 1.49 (*a*) Series distributed inductor with (*b*) equivalent lumped circuit.

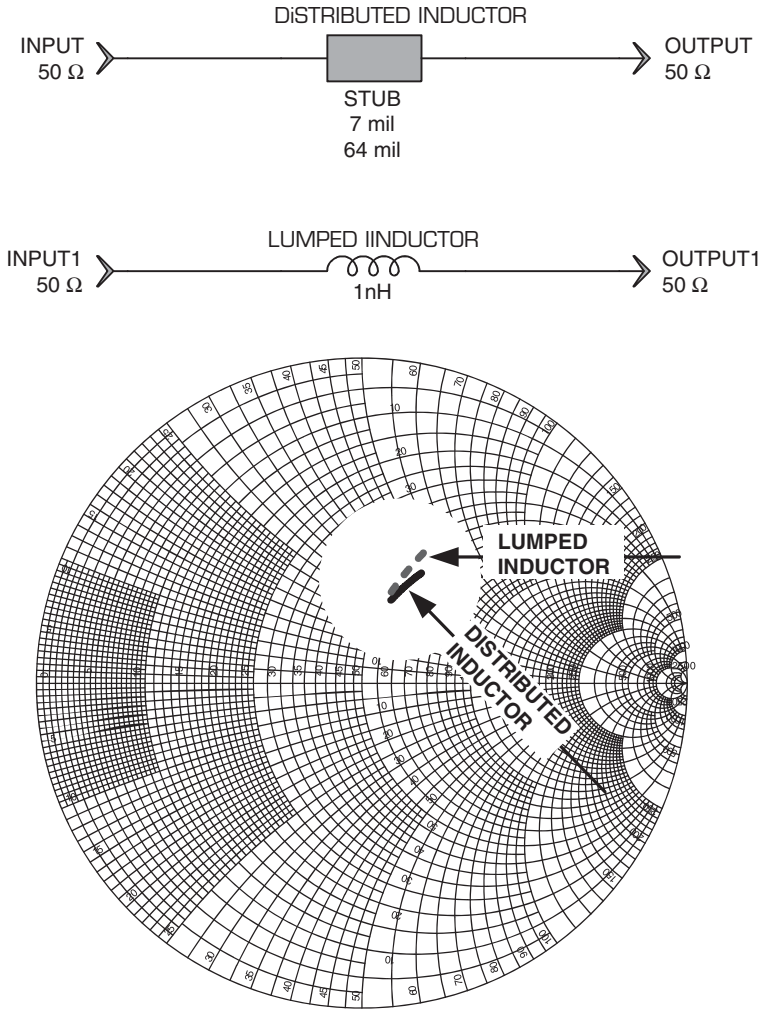


Figure 1.50 A series distributed inductor vs. a series lumped inductor over a narrow frequency range (5 to 6 GHz).

Parallel (shunt) inductor

As shown in Fig. 1.51, the equivalent shunt inductor is grounded at one end (a grounded stub) through a via to the groundplane of the PCB or, as will be shown, it may also be RF grounded through a distributed equivalent capacitor to ground.

Step 1: Knowing the inductance required within the circuit, calculate the reactance of the shunt inductor, at the frequency of interest, by the common formula:

$$X_L = 2\pi fL$$

Step 2: Use 100-Ω microstrip (Z_L = 100 Ω) for the substrate's dielectric in use. Find the microstrip width required for this 100-Ω value by either using one of the many

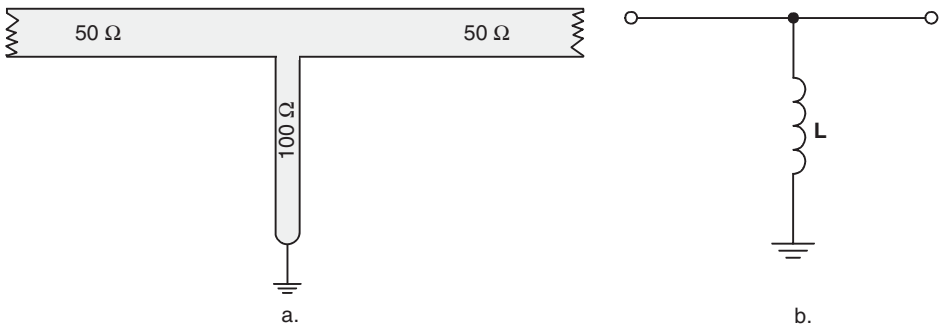


Figure 1.51 (*a*) Shunt distributed inductor with (*b*) equivalent lumped circuit.

microstrip calculation programs available for free on the Web (such as HP's AppCad, or AWR's TXLine, or Daniel Swanson's MWTLC, and so on), or calculate with the microstrip formula above.

Step 3: Calculate the microstrip's required length to become an inductor of value X_L:

$$\frac{\left(\frac{X_L}{100}\ \text{ARCTAN}\right)}{360} \times \lambda = \text{Length}$$

where X_L = inductive reactance needed in the distributed circuit, Ω
Length = length of the microstrip required to imitate a lumped component of value X_L (should never be longer than 30°, or 12%, of λ), mils
λ = wavelength of the frequency of interest using the substrate of interest (or $V_P \times \lambda$), mils

A Quick Example Design a Distributed Shunt Inductor (Fig. 1.52)

Goal: Create a distributed shunt inductor to replace a lumped part. The frequency of operation is 5.8 GHz and the inductance required is 1 nH. The substrate is FR-4 ($E_r = 4.6$), with a thickness of 20 mils.

Solution:

1. Inductive reactance of a 1-nH capacitor is 36 Ω.
2. The width of the distributed inductor's microstrip will be 7 mils
3. The length will be 64 mils (using ARCTAN)

Choke [Radio Frequency Choke (RFC)]

The distributed choke is RF grounded (a grounded stub) through a distributed, or lumped, capacitor (Fig. 1.53), or by being directly grounded through the groundplane [Fig. 1.54(*a*)]. The width of a distributed choke is that of 100-Ω microstrip for the substrate's dielectric in use (Z_L = 100 Ω: 100 Ω is the impedance of the microstrip only, and *not that of the equivalent choke*). Find the microstrip width required for this 100-Ω value by either using one of the many microstrip calculation programs available for free on the Web (such as HP's AppCad, AWR's TXLine, Daniel Swanson's MWTLC, or so on),

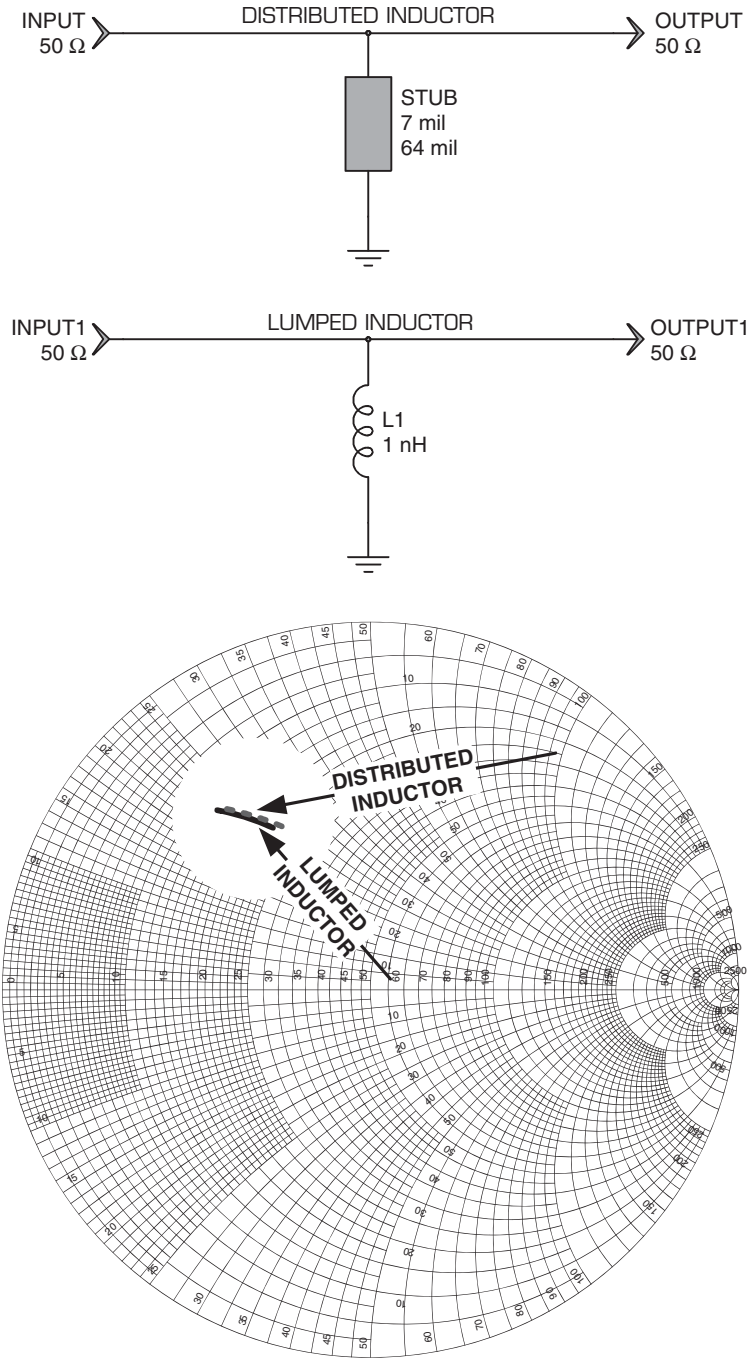


Figure 1.52 A shunt distributed inductor vs. a shunt lumped inductor over a narrow frequency range (5 to 6 GHz).

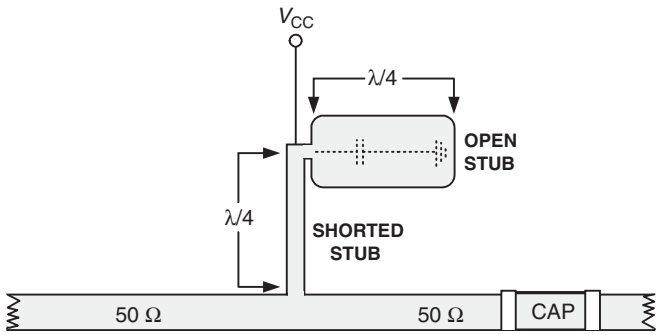


Figure 1.53 A distributed DC bias decoupling network.

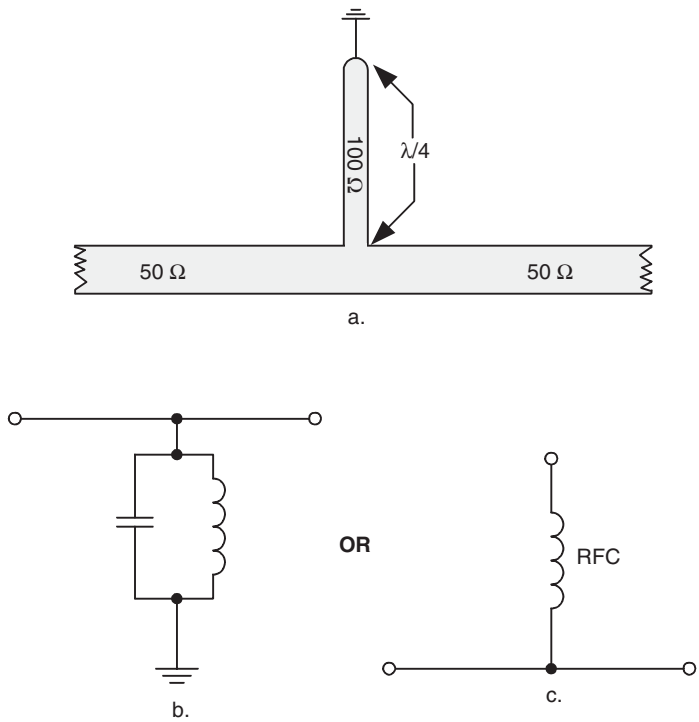


Figure 1.54 (*a*) Distributed choke with (*b*) equivalent lumped tank circuit, or as (*c*) lumped choke.

or calculate starting with the microstrip formulas above. The length of the choke will be exactly $V_p \times \lambda/4$, or 90° electrical. The distributed choke is theoretically now completely open due to the distributed circuit being at precisely $\lambda/4$.

The equivalent choke can be used in the bias decoupling circuit of Fig. 1.55(*a*) as such: *L* acts as a shorted quarter-wave stub due to the RF ground provided by *C*; *C* is behaving as an open stub to work as an RF short (by being exactly $\lambda/4$), while also being wide to lower its impedance even further; and R_{BIAS} and C_1 function as low-frequency

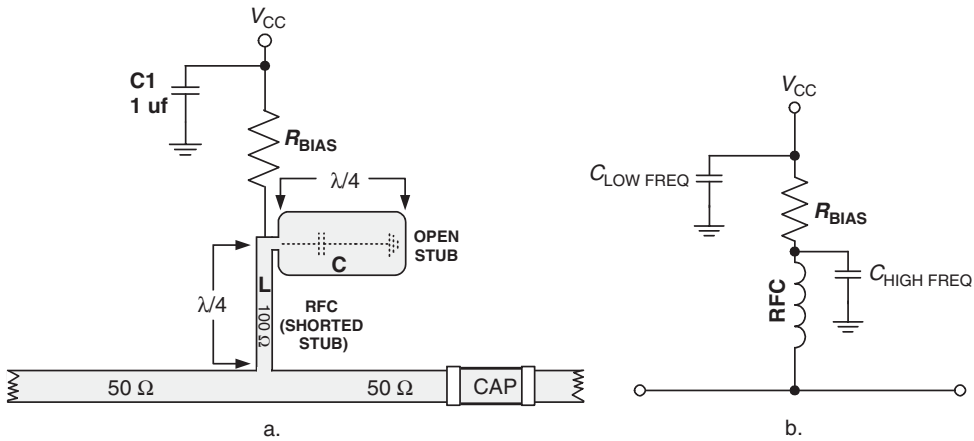


FIGURE 1.55 (*a*) Distributed DC bias decoupling with (*b*) equivalent lumped circuit.

decoupling [R_{BIAS} can also act as a bias resistor for a monolithic microwave integrated circuit (MMIC)].

A Quick Example Design a Distributed Bias Choke (Fig. 1.56)

GOAL: Create a distributed bias choke to replace a lumped part. The frequency of operation is 5.8 GHz. The substrate is FR-4 ($E_r = 4.6$), with a thickness of 20 mils.

Solution:

1. The width of the distributed choke's microstrip will be 7 mils.
2. The length will be 293 mils.

Transformer

The narrowband transformer of Fig. 1.57 is employed for resistive terminations only, such as between different values of microstrip, or between two resistive stages, or between two reactive stages with the reactances tuned out by a capacitor or an inductor. The transformer's length is exactly $V_p \times \lambda/4$, or 90° electrical, and its impedance can be calculated by:

$$Z = \sqrt{R_1 R_2}$$

After the impedance is found, calculate the "Z" section's required width by either employing one of the many microstrip calculation programs available for free on the Web (such as HP's AppCad, or AWR's TXLine, or Daniel Swanson's MWTLC, and so on.)

A Quick Example Design a Distributed Transformer (Fig. 1.58)

GOAL: Create a distributed transformer to match 40 to 80 Ω. The frequency of operation is 5.8 GHz. The substrate is FR-4 ($E_r = 4.6$), with a thickness of 20 mils.

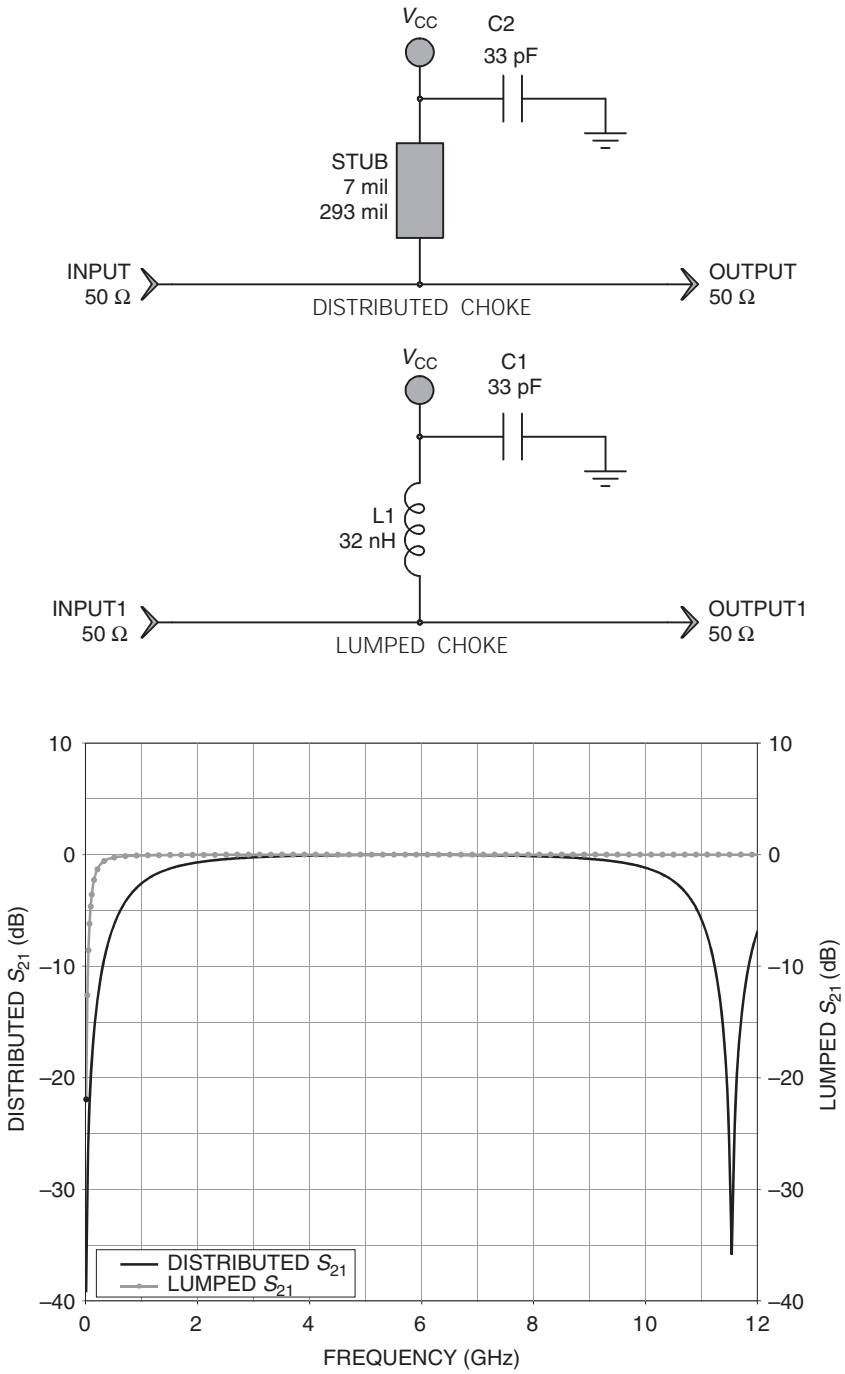


Figure 1.56 A 5.8-GHz distributed choke vs. a lumped (ideal) choke over a very wide frequency range (0 Hz to 12 GHz).

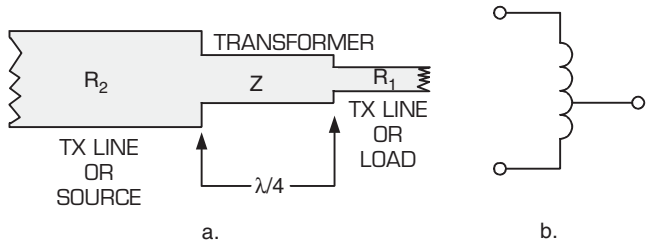
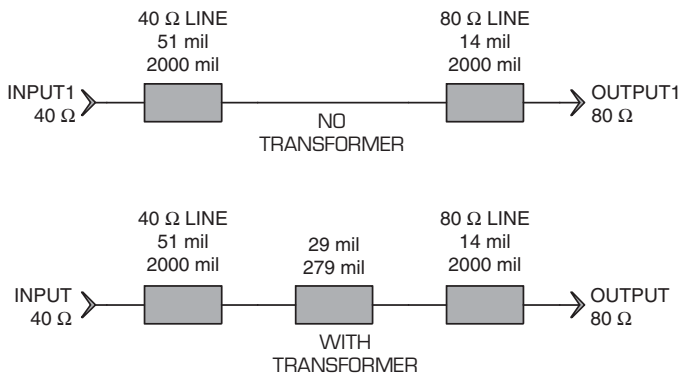


FIGURE 1.57 (*a*) Distributed transformer for resistive transformations, with (*b*) equivalent lumped circuit.



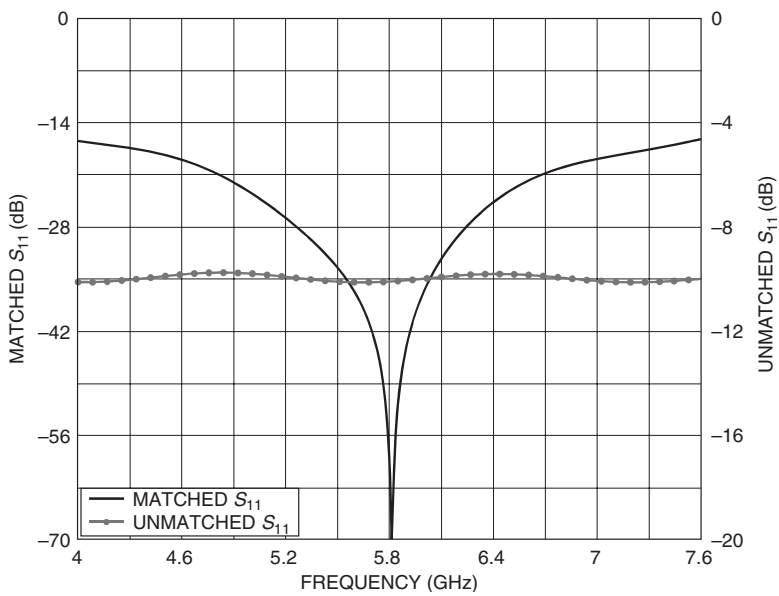


FIGURE 1.58 A microstrip circuit with and without a distributed 5.8-GHz impedance matching transformer (4 to 7.6 GHz).

Solution:

1. The impedance will be 56.5 Ω.
2. The width will be 29 mils.
3. The length will be 279 mils.

General microstrip component equivalency issues

Distributed inductors, transformers, capacitors, and series and parallel tank circuit's will only function for that particular dielectric constant, board thickness, and frequency used in the original equivalency calculations.

As stated above, the length of the equivalent inductor and capacitor elements should not be longer than 12% (30°) of a wavelength λ, or they will begin to lose their lumped component equivalency effect. When calculating the wavelength of the frequency of interest, the velocity factor of the substrate employed must be considered, since this changes the actual wavelength of the signal over that of free air. And inasmuch as the wavelength of the signal varies with the propagation velocity of the substrate, and the dielectric constant varies the V_p, and the dielectric constant changes somewhat over frequency, then all distributed components are frequency and dielectric constant dependent.

Due to microstrip's electromagnetic (EM) field leakage, when shielding microstrip-distributed equivalent capacitors and inductors (and microstrip's transmission lines), the RF shield should be kept at least five substrate thicknesses above the copper, or a disruption within the field, with resulting impedance variations, can occur.

The calculations for a frequency's velocity of propagation (V_p) will change slightly with the *width* of the microstrip conductor. This is due to the electric field that is created by the signal not being bounded by the dielectric and groundplane, but by air, on one side of the microstrip.

Figure 1.59 displays the proper and improper method to construct a distributed inductor and capacitor equivalence circuit, which in this case is being used as a three-pole lowpass filter. The proper way to position the microstrip-distributed inductors and capacitors is shown in Fig. 1.59(*b*) and Fig. 1.59(*c*), which have been calculated to be 246 mils long for the series inductors, and 200 mils long for the shunt capacitor. This layout will clearly permit the microstrip lengths and widths, as calculated, to function as desired. But in Fig. 1.59(*a*), the length of each of the distributed inductors is now far less than calculated. Figure 1.59(*c*) is the optimum approach, and minimizes any length and impedance discontinuity issues. Thus, each equivalent distributed component must be laid out properly, and with no length ambiguities, or improper circuit operation will result.

The capacitive *end effect* of open stubs should be taken into account in all distributed designs. This end effect creates an open stub that is approximately 5% longer electrically than the microstrip actually is physically on the PCB, causing the stub to resonate at a frequency lower than expected. This can be corrected by removing 5% from the calculated length of the open stub.

When shortening a distributed structure on a PCB, such as when using meandering line inductors, we should always check its reactance in an EM simulator to confirm that bending the line has not modified its performance unexpectedly.

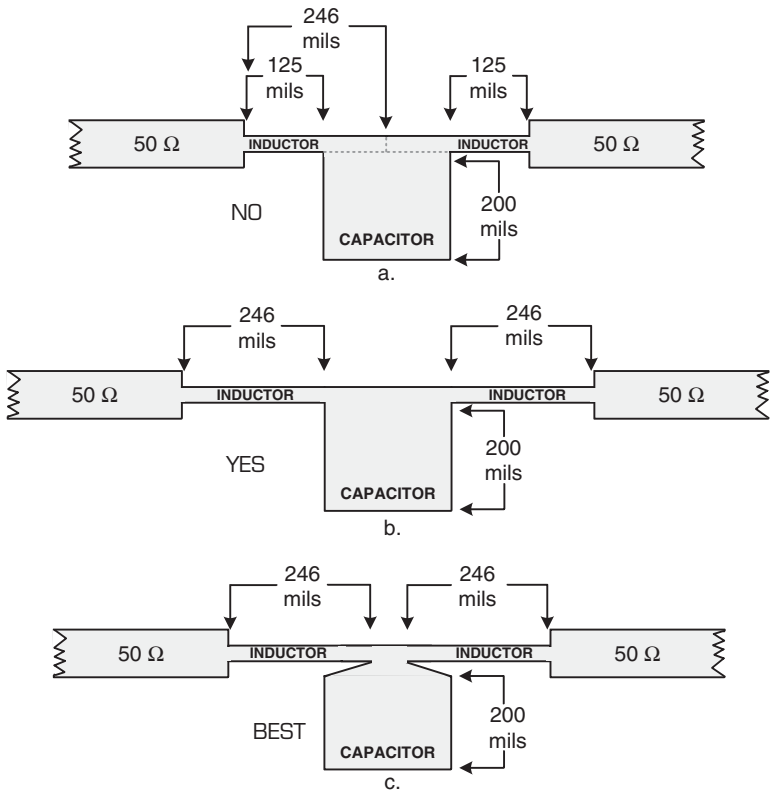


Figure 1.59 (*a*) Improper, (*b*) acceptable, and (*c*) preferred layout of equivalent microstrip components. (Desired length of each inductor section is 246 mils.)

1.4 Transmission Lines

1.4.1 Introduction

Transmission lines are conductors intended to move current from one location to another without radiating, and at selected impedance. There are two kinds of lumped RF transmission lines: *unbalanced*, normally in the form of coaxial cable, and *balanced*, such as twin-lead. By employing microstrip or stripline design techniques, both balanced and unbalanced transmission lines can be replicated on a printed circuit board.

Waveguide, a type of transmission line, can still be found in high-powered microwave transmitters, but is normally far more expensive than coaxial cable, is bulkier, and is also much harder to work with.

1.4.2 Transmission Line Types

Balanced lines are characteristically 300-Ω *twin-lead* (Fig. 1.60), and since there is no single conductor that is at ground potential, these types are distinctly different from

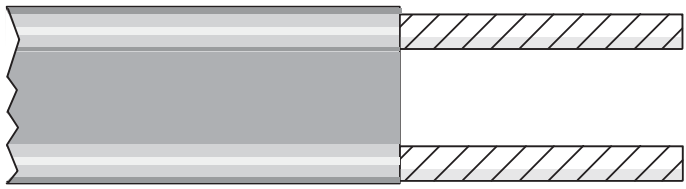


Figure 1.60 Twin-lead transmission line (for low frequencies).

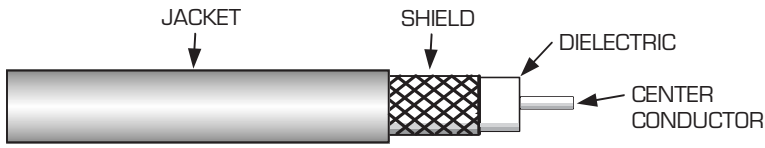


Figure 1.61 Flexible coaxial cable.

unbalanced coaxial line. In fact, balanced line has an equal-in-amplitude, but opposite-in-phase, signal present on each of its two conductors.

Commonly operated as a feedline to a television or FM receiver antenna or, more infrequently, as a balanced feed to a dipole transmitting/receiving antenna, twin-lead has very little intrinsic line loss, and is able to survive very high line voltages. However, twin-lead is not available in the impedance required for most modern transmitters and receivers (50 Ω).

By far, the most popular line today is the unbalanced type, which comes in the form of *coaxial* cable (Fig. 1.61). Coax is shielded with varying degrees of copper braid (or aluminum foil) to prevent it from receiving or radiating any signal (the inner conductor carries the RF current, while the outer shield is at ground potential).

Coax cable comes in many diameters, qualities, and losses. It is normally of the more common flexible type, covered with a protective rubber sleeve. Or it can be of the semirigid type, with its solid copper outer conductor. There is even flexible coax available that can reliably function up to frequencies as high as 50 GHz, but it is extremely expensive.

The dielectric constant of the coax's dielectric, and the spacing of its inner and outer conductors, as well as the diameter of the center conductor, govern the characteristic impedance of the transmission line.

Generally, the larger the diameter of the coaxial cable, the higher in frequency it can operate, and the smaller its losses. This is not true at the higher microwave frequencies, where the diameter of the cable can approach a certain fraction of the signal's wavelength, causing high transverse electric and magnetic (TEM) losses due to the coax transitioning to an undesired waveguide mode.

Now that coaxial cables can work in the high microwave regions, waveguide (Fig. 1.62) has become far less widespread. Whenever possible, modern microwave designs have removed waveguides in favor of low-loss, semi-rigid coax cables to transmit and receive signals. However, it is still favored as the transmission line of choice in certain demanding microwave high-power applications, and can be found in round or rectangular hollow metal shapes, which are optimized to transport microwave radiation from one point to another, with minimal signal losses, over very long distances. The actual size of the

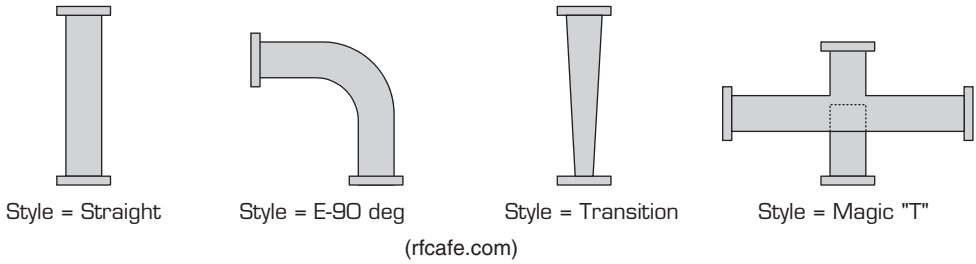


Figure 1.62 Waveguide microwave transmission line.

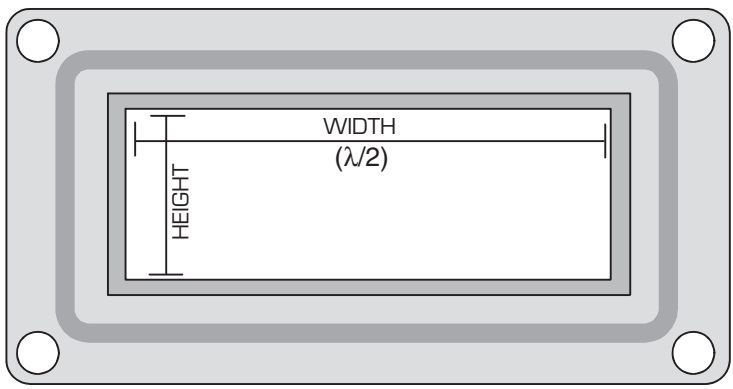


Figure 1-63 The width of waveguide should be half a wavelength at its frequency of operation.

waveguide itself will govern its working frequency (Fig. 1.63), with 1/4 wavelength straight or loop probes adopted to inject or remove the microwave energy from the waveguide structure.

Waveguides perform as a type of highpass filter, since they will propagate microwave radiation above their working frequency, but not below their cutoff frequency. However, mode shifts can arise within the waveguide structure, limiting the highest frequencies they are capable of propagating, and making a waveguide more of a very wide bandpass filter.

1.4.3 Transmission Line VSWR

Voltage standing wave ratio (VSWR) is the ratio between the reflected voltage and the forward voltage on a mismatched transmission line, or the ratio of a voltage maximum to the voltage minimum.

With a frequency source's output and its transmission line at the same impedance, and with the transmission line also equal to the load's input impedance, no standing or reflected waves can exist on the line. Thus, no excess power will be dissipated as heat, apart from the transmission line center conductor's natural resistance and skin effect losses. The line will appear infinitely long, with no power reflected back into the source, while sending the maximum RF power on to the load. The transmission line is thus

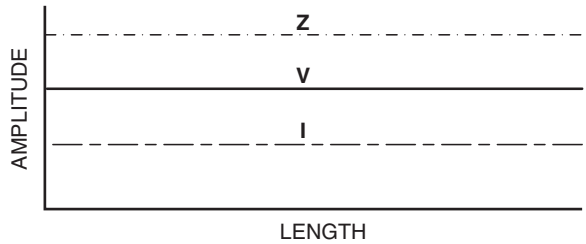


Figure 1.64 Voltage, current, and impedance distribution on a matched line, with no standing waves.

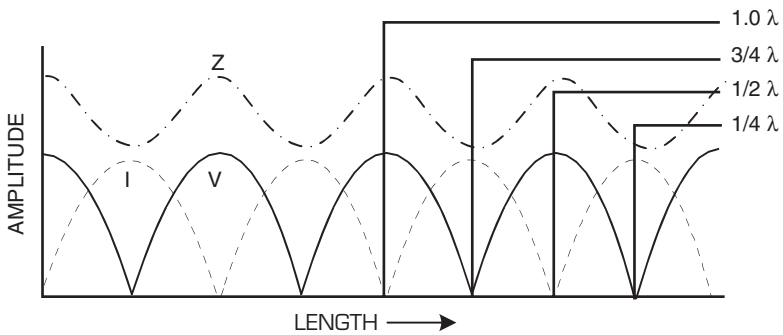


Figure 1.65 Voltage, current, and impedance distribution on a nonterminated transmission line, with high standing waves.

considered to be *flat line* (Fig. 1.64), with efficient power transfer and a low VSWR. However, if there were high standing waves (a high VSWR) existing on the transmission line (Fig. 1.65), the line's dielectric and/or the wireless transmitter's final amplifier could be damaged, and RF losses increased.

The highest transfer of RF energy will occur when the output of the transceiver and transmission line and antenna are all at the same impedance, which is a matched condition. However, when any component of this power amplifier-to-antenna system is not matched, then some of the RF energy is not passed on to the antenna, and a certain amount is actually reflected back toward the transmitter's output stage, producing peak voltage (and current) standing waves every 180° ($\lambda/2$) on the line. The magnitude of this reflection is dependent on the severity of the mismatch (i.e., the higher the VSWR the higher the reflected energy). As an example, when inserted into a perfectly matched environment (VSWR = 1:1) at 900 MHz, a certain brand of RG-58 has a loss of 0.65 dB over a 3-ft long section of cable. Thus, the efficiency of the power transfer through the coax is roughly 86%. On the other hand, when the RG-58's load is mismatched, creating a VSWR that is increased to 10:1, the total coax losses contributed by that exact same 3 ft will now increase to about 2.5 dB, or a power transfer efficiency of only 55%.

We can improve these numbers by improving our coax. For instance, moving up a bit to RG-8, with its 50-Ω 1:1 matched loss of only 0.35 dB over a 3-ft section, we can expect a mismatched loss of only 1.5 dB with that same 10:1 VSWR, or a power efficiency of 70%.

The ratio between the forward RF wave and the reflected RF wave is referred to as the *SWR*, while the *VSWR* (even though they are used interchangeably) is more accurately the ratio between the sum of the forward RF voltage and the reflected RF voltage.

1.5 S-Parameters

1.5.1 Introduction

S-parameters (*scattering-parameters*) characterize any RF device's complicated behavior at different bias points and frequencies, and are useful in allowing the circuit designer the ability to easily calculate a wireless device's gain, return loss, stability, reverse isolation, matching networks, and other vital parameters.

S-parameters are effective for any small signal RF linear circuit design application, passive or active. In amplifiers, *S*-parameters are only accurate for Class A devices running under 1 W, and are not considered useful in most power amplifier design.

Power is only one part of the equation as to what RF amplifier can be designed and described using *S*-parameters, as an amplifier must also be operated within its linear region. This will exclude any amplifier, even under nonpower conditions, that is biased at anything other than Class A, such as Class AB, B, or C. However, Class AB amplifiers are usually considered as performing acceptably when preliminarily designed using *S*-parameters, while even a few Class B designs have at least been started off employing *S*-parameters techniques.

What are *S*-Parameters?

For small signal transistors and linear passive networks meant to operate at frequencies above 10 to 50 MHz, *S*-parameters are utilized to design a circuit's input and output matching networks for maximum gain, as well as to define the input and output *reflection coefficients* of any linear network that is terminated at both its ports with $50 + j0$.

Reflection coefficients are the ratio of the reflected wave to the forward wave, they are a measure of the quality of the match between one impedance and another, or $V_{\text{REFLECTED}}/V_{\text{FORWARD}}$, with a perfect match equaling zero and the poorest match equaling one. Reflection coefficients can be expressed in rectangular ($\Gamma = R \pm jX$) or polar ($\Gamma = P \angle \pm 0$) forms.

When designing amplifiers, all *S*-parameters are collected and are valid for only one combination of V_{CE}, I_C, and f_r. However, this is not as limiting as it may seem, since multiple frequencies will always be given within the transistor's *S*-parameter file (called an *S2P file*; Fig. 1.66). Many microwave transistor manufacturers will also supply multiple *S*-parameter files with different selections and combinations of V_{CE} and I_C for each active device. This allows the engineer more flexibility in common-emitter amplifier bias design.

S-parameters can be collected for any linear device, whether active or passive, and not only to be used in calculating matching circuit elements, but also to simulate a complete circuit within a computer for gain, stability, and return loss. These *S*-parameters measurements are required in virtually all of high-frequency RF design, since at elevated frequencies many Spice models completely break down. This poor Spice performance at RF is due to the scarcity of proper modeling of a device's natural internal parasitic distributed capacitances and inductances, the inability to use microstrip, and the problems that occur as the wavelength of the RF signal becomes a significant portion of the total size of the physical components. These issues will have a significant effect on a circuit's true behavior at RF frequencies. However, many high frequency Spice packages will now permit not only accurate simulations up to almost 1 GHz, but also permit the use of linear *S*-parameter models.

```
! FILE NAME (21bfg425.S2P)
! MODEL OF TRANSISTOR (BFG425W)
! TRANSISTOR BIAS (V2=2.000 E+000 V, I2=1.000 E–003 A)
# GHz  S  MA  R 50
!          S11          S21          S12          S22
!Freq(GHz) Mag   Ang    Mag   Ang      Mag   Ang      Mag   Ang
   0.040  0.950 –1.927  3.575 177.729  0.003  83.537  0.996 –1.116
   0.100  0.954 –5.309  3.518 175.247  0.007  87.057  0.996 –3.082
   0.200  0.951 –10.517 3.504 170.441  0.014  82.341  0.991 –6.343
   0.300  0.947 –15.891 3.496 166.534  0.020  78.681  0.988 –9.405
   0.400  0.941 –20.987 3.493 161.221  0.027  75.109  0.982 –12.576
   0.500  0.935 –26.297 3.476 156.531  0.033  71.254  0.974 –15.593
   0.600  0.928 –31.508 3.433 151.954  0.040  67.636  0.965 –18.605
   0.700  0.919 –36.669 3.384 147.515  0.046  63.875  0.954 –21.674
   0.800  0.910 –41.871 3.350 143.152  0.051  60.357  0.943 –24.600
   0.900  0.898 –46.948 3.317 138.801  0.057  56.929  0.930 –27.559
   1.000  0.886 –52.161 3.272 134.309  0.062  53.488  0.916 –30.396
   1.100  0.874 –57.181 3.223 130.114  0.067  50.181  0.903 –33.098
   1.200  0.861 –62.218 3.171 125.837  0.071  46.955  0.888 –35.859
   1.300  0.849 –67.154 3.119 121.786  0.075  43.791  0.873 –38.531
   1.400  0.835 –72.157 3.072 117.682  0.079  40.631  0.857 –41.151
  11.500  0.845  –2.938 0.375 –130.163 0.134 –104.397 0.607  14.337
  12.000  0.848  –9.981 0.326 –139.789 0.124 –115.184 0.658   4.326

!BFG425W_noise.xls                                  !Philips part #:BFG425W
! Bias condition: Vce = 2 V, Ic = 1 mA              # MHz dB Ratio deg ohm

! Freq. Fmin Gmag Gangle Rn
900 0.7 0.67 19.1 0.40                              2000 1.3 0.56 57.5 0.36
```

Figure 1.66 A S2P S-parameter file at a set bias condition and at assorted frequencies. Text after "!" are comments for the user, and are ignored by the simulation programs.

S-parameters are described by S_{21}, S_{12}, S_{11}, and S_{22}. S_{21} is the forward-transmission coefficient, representing the stage gain, and is pronounced *"es two one"*. S_{12} is the reverse transmission coefficient, representing the reverse gain (isolation). S_{11} is the input reflection coefficient, representing the input return loss. S_{22} is the output reflection coefficient, representing the output return loss. Figure 1.67 combines all of these *S*-parameters into the gains and reflections of a single "black box" for a simplified graphical description.

S-parameters are voltage ratios referenced to 50 Ω, since a voltage comparison is all that is needed to find any *S*-parameter at the same impedance value. But a complete *S*-parameter can only be a vector quantity, thus taking into account not only the magnitude, but also the phase. This allows the analysis of stability and complex impedances while both ports of an active device are terminated into 50 Ω.

Considering that *S*-parameters are the voltage ratio between a port's input potential to the reflected potential, or from the input potential to the output potential, then it can be seen that any RF circuit can easily be accurately characterized using just *S*-parameters.

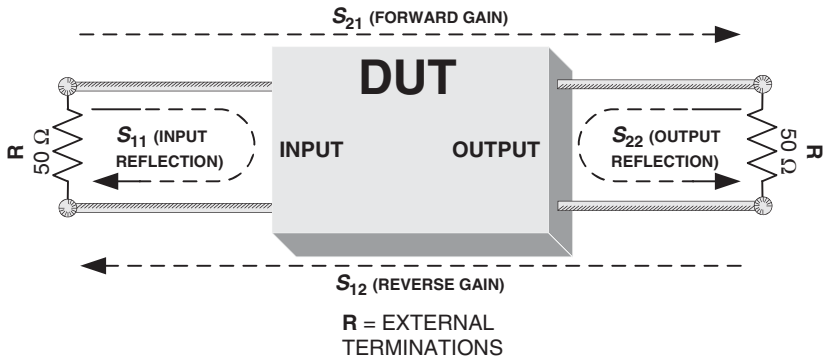


Figure 1.67 A two-port network showing transmission and reflection parameters.

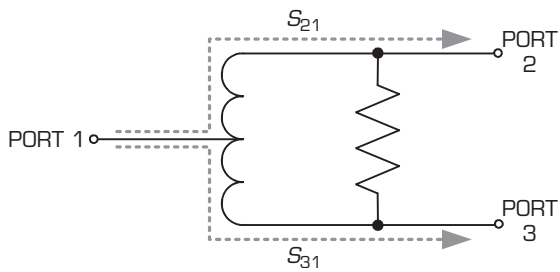


Figure 1.68 Three-port splitter demonstrating forward *S*-parameters.

It is far easier for a designer to deal with decibels than with voltage levels, so the majority of S_{21}, S_{11}, S_{12}, and S_{22} values will be in dB which, when compared to the measured voltages in a 50-Ω system is:

$$S_{XX}\,(\text{dB}) = 20\text{LOG}_{10}\left|S_{XX}\right|$$

where *XX* = either 21, 12, 11, or 22

Even though *S*-parameters are frequently associated with two-port devices, they work equally well with three, four, or more ports by the addition of the suitable subscripts, such as S_{31} for the forward gain through one branch of a splitter, as shown in Fig. 1.68.

1.5.2 *S*-Parameter Measurement

For increased clarity on just what *S*-parameters are, the measurement of an active device is shown in Fig. 1.69, and demonstrates a simple test setup for a method of taking the *S*-parameters in the forward direction.

For example, to obtain the *S*-parameters for a BJT, the bias voltage is injected at "Bias Voltage 1" (with *L* and *C* acting as decoupling components), which will control the transistor's base current, and thus the collector current, of the device under test (DUT). The emitter is grounded, while the V_{CE} for the BJT's collector is supplied by "Bias Voltage 2." The "Bias Voltage 2" furnishes the Class A static DC bias conditions for

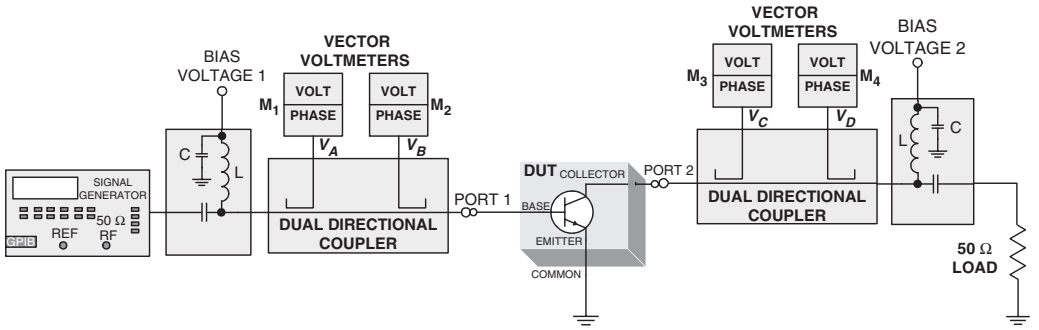


FIGURE 1.69 Test setup for forward *S*-parameter measurement.

the transistor. This hints as to why all *S*-parameter files are supplied only for a specific *fr*, I_{CE}, and V_{CE}, and why other $f_{r's}$, $I_{CE's}$, or $V_{CE's}$ will change the device's *S*-parameters (see *What are S-Parameters?* above). An AC signal is now injected into *Port 1*, through the dual directional coupler, by the 50-Ω RF signal generator (*S*-parameters are taken only when terminated at the input and output by 50 Ω, or 50 + *j*0). The *vector voltmeter*, M_1, which is a device that is able to measure not only the voltage of a signal but also its phase, reads the amplitude of V_A, and its phase, of the signal into Port 1 of the BJT, while the M_2 meter reads the amplitude of V_B and its phase of the signal that is reflected back from the input of the BJT due to the impedance mismatch as compared to the system impedance of 50 Ω. When the amplitude of the reflected signal V_B is divided by the amplitude of the V_A signal from the generator, the magnitude of S_{11}, or the *input reflection coefficient*, can be obtained by V_B/V_A. This value will be less than unity.

The phase angle difference between V_B and V_A depicts the phase angle of S_{11} by $\angle\theta_B - \theta_A$. So the *S*-parameter S_{11} will equal:

$$S_{11} = \frac{V_B}{V_A} \quad \angle\theta_B - \theta_A$$

S_{21}, or the *forward transducer gain*, equals the voltage measured at V_C by M_3, and will be at some value that is greater than unity, since it is the amplified value of V_A or V_C/V_A. And the phase difference between V_C and V_A is measured as $\angle\theta_C - \theta_A$. So S_{21} will equal:

$$S_{21} = \frac{V_C}{V_A} \quad \angle\theta_C - \theta_A$$

Figure 1.70 shows one technique for measuring the *reverse S*-parameters of an active device. Basically, the setup of Fig. 1.69 above is flipped, but the bias voltages and the DUT orientation remain the same, with the input now terminated with 50 Ω and the active device's output now fed by the 50-Ω signal generator. A signal is injected into Port 2, through the dual directional coupler, by the 50-Ω signal generator. The *vector voltmeter*, M_4, reads the amplitude, V_D, and its phase of the signal at Port 2 of the DUT, while the meter, M_3, reads the amplitude, V_C, and its phase of the signal reflected back from the output of the DUT (due to any impedance mismatch from the system's impedance of 50 Ω). When the amplitude of the reflected signal, V_C, is divided by the amplitude of the incident signal from the signal generator, V_D, then the magnitude of

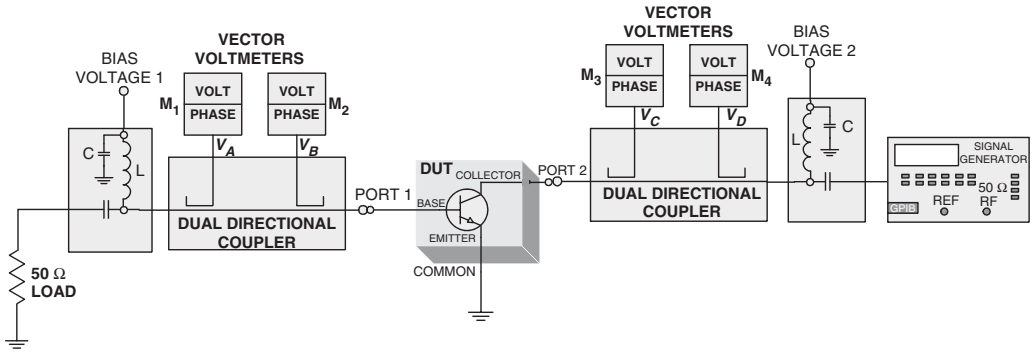


FIGURE 1.70 Test setup for reverse *S*-parameter measurement.

S_{22}, or the *output reflection coefficient*, can be obtained by V_C/V_D. This value will be less than unity. The phase angle variance between V_C and V_D stands for the phase angle of S_{22} by $\angle\theta_C - \theta_D$. So S_{22} will equal:

$$S_{22} = \frac{V_C}{V_D} \quad \angle\theta_C - \theta_D$$

S_{12}, or the *reverse transducer gain* equals the voltage measured at V_B by M2 divided by V_D or:

$$S_{12} = \frac{V_B}{V_D}$$

Which is also referred to as the amplifier's *isolation*, and is desired to be as low a value as possible. This is because we would like to have a gain in the forward direction only, with as little reverse gain, and thus as little adverse interaction, from reflections back into the amplifier's output port as possible.

The phase difference between V_B and V_D is measured as $\angle\theta_B - \theta_D$. So S_{12} will equal:

$$S_{12} = \frac{V_B}{V_D} \quad \angle\theta_B - \theta_D$$

To convert these values into decibels as desired:

$$|S_{11}|\text{dB} = 20\text{LOG}|S_{11}|$$

$$|S_{22}|\text{dB} = 20\text{LOG}|S_{22}|$$

$$|S_{21}|\text{dB} = 20\text{LOG}|S_{21}|$$

$$|S_{12}|\text{dB} = 20\text{LOG}|S_{12}|$$

The DUT, a BJT transistor, has now been characterized at a single frequency and a single bias point by the use of *S*-parameters. These gathered *S*-parameters could now

be inserted into an RF circuit simulation program, or they could be used to design an input and output matching circuit for the BJT. *S*-parameter matching and simulation will be discussed in further detail in the Chap. 3.

1.6 Noise in Circuits and Systems

1.6.1 Introduction

Noise is of crucial concern in radio, since the higher a signal is above the noise, the higher will be its signal-to-noise ratio (SNR), and the farther away it can be detected with the desired BER. There are two primary classifications of noise: *circuit generated* and *externally generated*. Both are unavoidable and limit the possible gain of any receiver's amplifiers. However, noise can be minimized by careful and cautious circuit and systems design. For instance, if we employ LNAs and tight (low loss) filtering in the front-end of receivers, decrease the noise contributions from the receiver's own local oscillators, eliminate the image frequency and image noise, and employ proper shielding and layout techniques, we can significantly improve on the receiver's noise floor amplitude.

Noise created within circuits and systems produce a chaotic, fluctuating, and very wide frequency ranging voltage. *White noise*, also referred to as *Johnson* or *thermal noise*, is created by a component's electrons randomly moving around due to thermal energy. *Shot noise*, because of its characteristics, can also be considered as another type of white noise, but is caused by electrons entering the collector or drain of a transistor, and by the indiscriminate movement of electrons across any semiconductor junction.

Due to shot noise effects, Zener diodes are especially problematic as noise contributors. In fact, all Zeners will add shot noise, but the designers of such diode's will have minimized other, more complex, noise contributors in any low-noise Zener type.

External noise is caused by not only man-made sources of electromagnetic interference, such as dimmer switches, car ignitions, electric engines, and the like, but also from natural sources, such as the static created by atmospheric lightning, as well as *space noise* produced by solar flares and sunspots, along with *cosmic noise* induced by the stars radiating interfering signals in all directions.

Circuit-generated noise power, in Watts, can be calculated by the formula shown below. This simple formula states that the two contributors to noise of a circuit are its temperature and bandwidth, and lower temperatures and lower bandwidths result in lower noise contributions. The actual carrier frequency of the signal itself has absolutely no effect on the production of this noise:

$$P_N = KTB$$

where P_N = noise power, W
K = Boltzmann's constant, or 1.38×10^{-23}
T = circuit's temperature, K
B = circuit's bandwidth, Hz

A related noise contributor is from outside origins, such as a signal source, and is created by the same mechanism as above. It is referred to as *source noise*, and can be calculated by:

$$NF = 10 \ \ LOG_{10}\left(\frac{P_{NO}}{P_{NI}}\right) \qquad \text{at} \ \ 290K$$

where NF = noise figure, dB
P_{NO} = output noise power, W
P_{NI} = input noise power, W
290K = reference temperature used in most measurements, K

This page intentionally left blank

CHAPTER 2
Modulation

Different modulation schemes have been adopted for the radio services, such as broadband amplitude modulation (AM) and frequency modulation (FM) for broadcast, narrowband FM for line-of-site two-way voice communications, single sideband (SSB) for long-distance voice communications via the ionosphere, and digital modulation for high-speed point-to-point or multipoint microwave radio communications links.

2.1 Amplitude Modulation

2.1.1 Introduction

Amplitude modulation is the earliest modulation method for wireless voice communications. Not only is it very simple and cheap to work with from a hardware standpoint, but is still extensively used today for commercial and shortwave broadcast, as well as in certain citizen's band radios and a few ham radio systems.

2.1.2 AM Fundamentals

Modulation is the way we insert baseband information on an RF carrier wave. The baseband information can be voice, digital data, analog video, and so on. Demodulation is the procedure of extracting this baseband information, which is then sent to a speaker for voice and music, or on to digital circuits for processing or storage.

The most basic way we have of imprinting voice, data, or music on an RF carrier is by modulating the amplitude of this carrier (Fig. 2.1). The unmodulated RF, which is produced by an oscillator, functions as the carrier that will transport the baseband modulation through space to a receiver. The baseband is the intelligence—always at a much lower frequency than that of the RF carrier—that is inserted onto this carrier through nonlinear mixing of the two signals. As seen in the time domain, the amplitude of the RF carrier is modified at the rate of the baseband's own amplitude and frequency variations. In fact, if the amplitude of the baseband signal increases, then so will the amplitude of the RF carrier (Fig. 2.2); while decreasing the baseband's amplitude decreases the amplitude of the carrier (Fig. 2.3).

The baseband modulation travels with the RF carrier to the receiver. The receiver then takes these amplitude variations that are riding on the carrier and removes them, thus converting them back into the original audio amplitude variations that were inserted at the transmitter. The recovered baseband is then amplified and fed into a speaker, or some other appropriate transducer. The actual percent of modulation is what controls the final amplitude of the detected signal, and the higher the amplitude

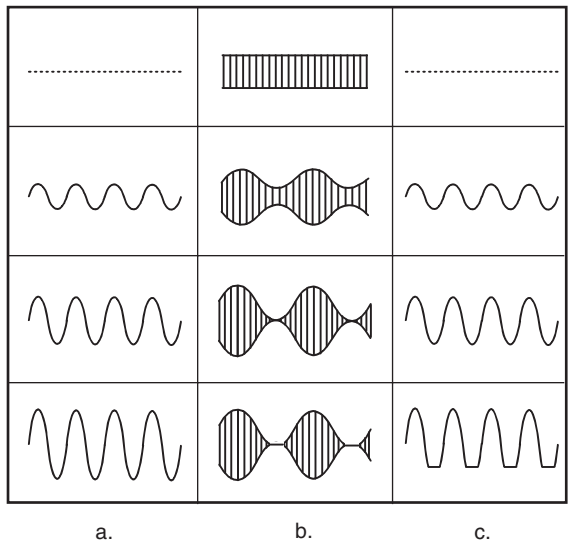


Figure 2.1 (*a*) Baseband modulation at various amplitudes; (*b*) a carrier unmodulated and at various amplitude modulation percentages; (*c*) the demodulated waveform at baseband.

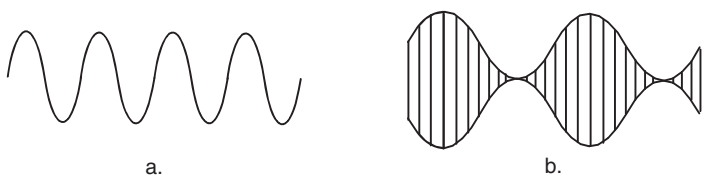


Figure 2.2 (*a*) The baseband audio modulation; (*b*) the 100% AM-modulated RF waveform.

of the baseband signal the higher the volume at the receiver's speaker, as shown in Fig. 2.1.

When the baseband signal is modulated at the transmitter, both the positive and negative alternations of the RF carrier will be influenced symmetrically. This means that the missing negative alternation lost by the Class C collector modulation circuit will be recreated again by the tuned output tank of the transmitter's final amplifier, forming a mirror image of the positive alternation (Fig. 2.4).

Sidebands are formed by this modulation between a carrier and its baseband signal, which are viewable in the frequency domain as shown in Fig. 2.5. These sidebands are created by the modulator's nonlinear mixing circuit producing sum (the upper sideband) and difference (the lower sideband) frequencies. But it is the phase relationships between the RF carrier and the upper and lower sidebands that actually creates a new waveform that will deviate in amplitude in the time domain (Fig. 2.6). This effect is produced when the two sidebands and the carrier are in phase, causing the amplitude of the modulated carrier waveform to be double that of the unmodulated carrier; and when the carrier and the two sidebands are completely out of phase, the amplitude of this new carrier waveform will be virtually zero. The new waveform will therefore have high peaks and low valleys (Fig. 2.7).

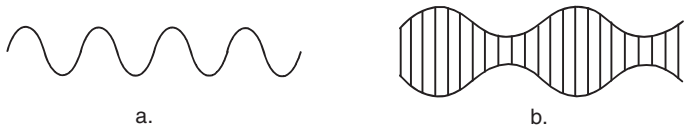


Figure 2.3 (*a*) The baseband audio modulation; (*b*) the 50% AM-modulated RF waveform.

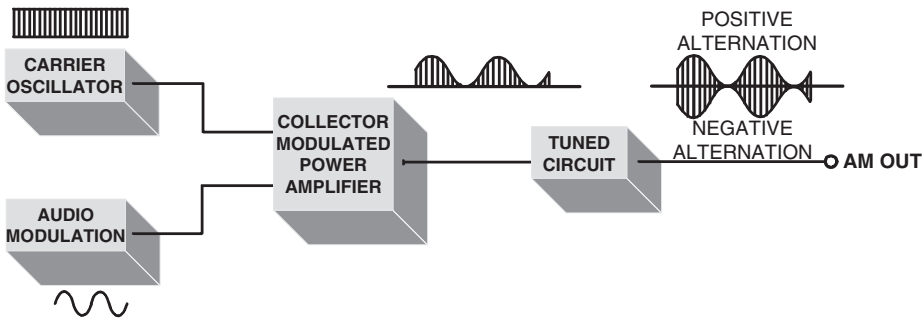


Figure 2.4 A Class C collector modulator for AM.

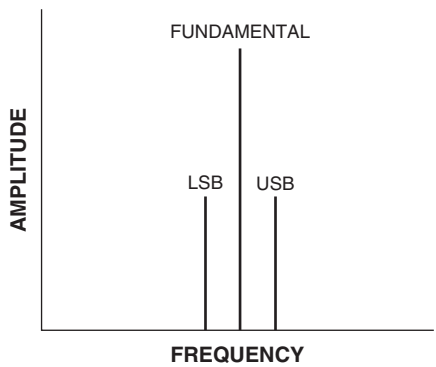


Figure 2.5 A carrier with its sidebands when AM modulated by a single baseband tone.

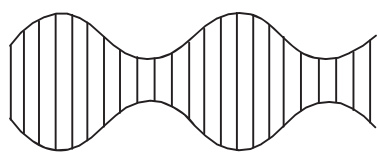


Figure 2.6 A single-tone amplitude-modulated RF carrier in the time domain.

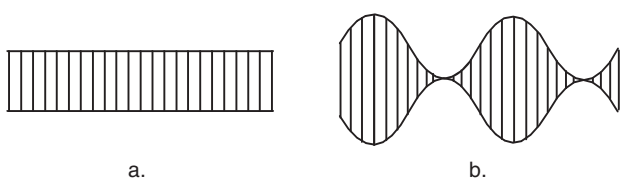


Figure 2.7 (*a*) An unmodulated carrier and (*b*) a 100% amplitude-modulated carrier in the time domain.

In the time domain, the percent of modulation of an AM signal can be found on an oscilloscope displayed by the following formula:

$$\%\ \text{MOD} = \frac{V_{\text{PEAK}} - V_{\text{MIN}}}{V_{\text{PEAK}} + V_{\text{MIN}}} \times 100$$

However, when such a modulated signal is observed in the frequency domain, the RF carrier's frequency and amplitude will not actually change, whether it is modulated or not (Fig. 2.8). This confirms that the carrier itself holds no information that can be demodulated, but that the information is in fact embodied within the two sidebands only. Indeed, when an AM signal is inspected in the frequency domain, we clearly see that when the transmitter's baseband modulation is varied both in frequency and amplitude, the carrier will stay at its original frequency and amplitude; while only the sidebands themselves will change in frequency and amplitude (Fig. 2.9). This distinctly verifies that there is no actual information contained within the RF carrier, but only

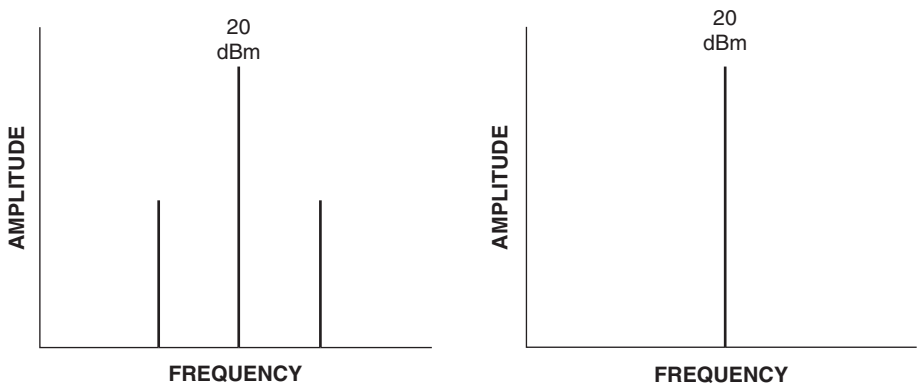


Figure 2.8 In the frequency domain the amplitude of an AM carrier remains the same whether (*a*) modulated or (*b*) not.

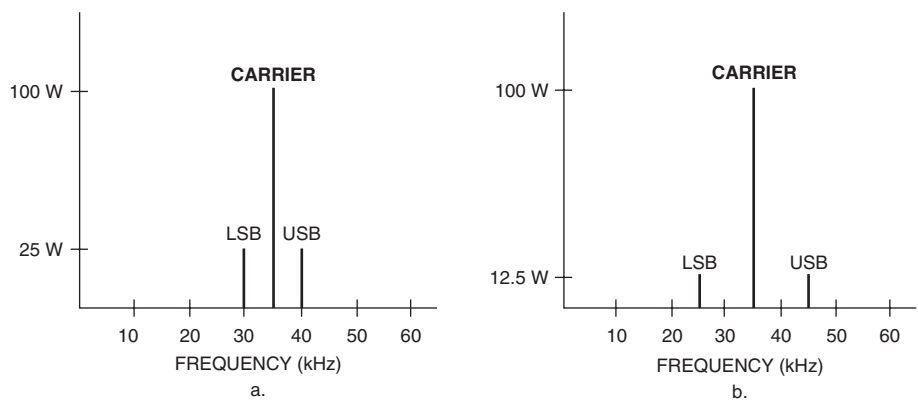


Figure 2.9 When the amplitude and/or frequency of the baseband modulation changes, only the sidebands are affected: (*a*) 5 kHz, high amplitude; (*b*) 10 kHz, low amplitude modulation.

within its sidebands; with each sideband holding the same information and power as the other.

These sidebands, both the upper (USB) and the lower (LSB), can be found at the sum and difference frequencies of the carrier and modulating frequencies: $f_{CARRIER} - f_{AUDIO}$ = LSB and $f_{CARRIER} + f_{AUDIO}$ = USB. As an example, if a 1000-kHz carrier is modulated with a 10-kHz baseband audio tone, then the sideband frequencies can be located at 1000 kHz + 10 kHz = 1010 kHz, and at 1000 kHz − 10 kHz = 990 kHz. The bandwidth of this signal will thus be twice the baseband frequency, or the upper sideband minus the lower sideband. In the above example, the bandwidth is 2×10 kHz, or 20 kHz.

Only the *total* power changes in an AM signal, or $P_T = P_C + P_{LSB} + P_{USB}$. This is because, as discussed above, the carrier power remains unchanged no matter what the baseband modulation amplitude we use, and only the sideband amplitudes will vary. Thus, the total power of an AM signal will equal the sum of the carrier and sideband powers. In fact, when the carrier is amplitude modulated by the baseband, the antenna current will rise due to the power added by the increasing sidebands.

An enormous amount of sidebands are created during normal voice modulation, and are located at many different frequencies and amplitudes. As shown in Fig. 2.10, a spectrum analyzer display of a voice AM signal can become quite complex. This is why we must employ single- or dual-tone baseband input signals for testing purposes.

The majority of AM voice transmitters will confine their modulation frequencies to between 300 and 3000 Hz to limit transmitted bandwidth. Limiting the baseband frequencies is easily accomplished by the use of a bandpass filter located just after the first audio (microphone) amplifier. To prevent AM overmodulation, an amplitude limiter circuit can also be employed to limit the maximum audio baseband amplitude, since overmodulation causes an unwelcome increase in transmitted bandwidth due to spectral *splatter*, as well as distortion. Splatter is harmonic generation within the original baseband frequencies created by clipping of the signal's modulation envelope; this action further modulates the RF carrier, producing adjacent channel interference

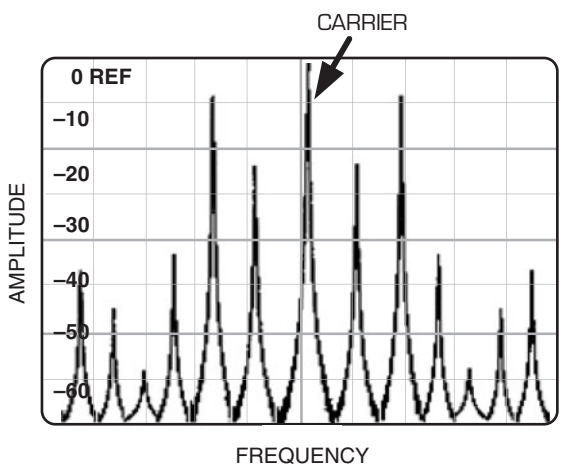


Figure 2.10 A voice signal, as viewed in the frequency domain, is comprised of many sidebands at various frequencies and amplitudes.

(ACI). The distortion level is now increased because part of the AM signal is not actually present at the demodulator (Fig. 2.1), so intelligibility of the received signal is degraded.

2.1.3 AM Power Measurement

The power of an AM signal can be measured as the *peak envelope power* (PEP), which is utilized to gauge the average peak power, with 100% modulation applied, of the transmitted signal, or:

$$\text{PEP} = V^2_{\text{RMS}}/R \qquad \text{or} \qquad V_{\text{RMS}} \times I_{\text{RMS}} \qquad \text{or} \qquad I^2_{\text{RMS}} \times R$$

The carrier power can also be calculated with these same formulas, but with zero transmitter modulation applied.

NOTE: *As an important aside to AM power measurement, increasing the transmitter's range by increasing its power is not linear. To extend a transmitter's distance by two, multiply the transmitter's power, in watts, by 2^2 (or 4 × power). To increase the distance by three, multiply the power by 3^2, (or 9 × power).*

2.1.4 AM Disadvantages

The disadvantages of AM are many: The bandwidth of an AM signal is twice what is required for the reception of the intelligence being sent, since only one sideband is absolutely necessary to convey the baseb and information; a significant amount of power is in the carrier, which is not even required to furnish the intelligence; the phase relationship between the carrier and the sidebands must be precise, or severe fading will result within the demodulated signal (which is quite difficult to maintain under most atmospheric and multipath conditions).

2.2 Frequency Modulation

2.2.1 Introduction

Frequency modulation was originally invented as an answer to the many deficiencies inherent in AM, which is fundamentally that of excessive noise sensitivity. Since noise is normally produced by undesired amplitude variations in a signal, this can easily be removed in frequency-modulated receivers by amplitude limiters.

Two techniques can be employed to generate an FM signal. The first, which directly alters the frequency of the carrier in step with the baseband's amplitude variations, is called *direct FM*; and the second method, *indirect FM*, changes the phase of the carrier, which creates phase modulation. However, both of these techniques produce the end effect of frequency modulation of the RF carrier. Both methods are classified under the designation of *angle modulation*.

2.2.2 FM Fundamentals

Modulation is the method we use to insert baseband information on an RF carrier wave. The baseband information can be voice, digital data, analog video, and so on. Demodulation is the procedure of extracting this baseband information, which is then

sent to a speaker to reproduce the original voice and music, or on to digital circuits for processing or storage.

FM accomplishes this modulation process by altering the carrier's frequency in step with the baseband signal's changes in amplitude. When this frequency-modulated RF carrier arrives at the receiver, the frequency variations as created by the original baseband modulations are changed back into amplitude variations. This baseband is then amplified and inserted into an appropriate transducer. As stated, in FM the baseband's amplitude alters the frequency of the RF carrier, and not the amplitude as it does with AM, while the amount of the actual *frequency deviation* of the FM carrier is dependant on the increase or decrease in this baseband's amplitude. Frequency deviation is considered to be the amount the RF carrier deviates from its center frequency in *one* direction during modulation. Without any baseband modulation present, however, the frequency of the RF carrier will stay at the transmitter's predetermined *center frequency*, which is the frequency of the master oscillator after any multiplication. Thus, as the baseband modulation occurs, the carrier will increase and decrease in frequency; as the baseband swings positive in amplitude, the carrier will increase in frequency, but as the baseband modulation swings negative in amplitude, the frequency of the carrier will fall below its rest frequency (Fig. 2.11).

The *frequency* of the baseband signal will change the rate that the frequency-modulated RF carrier intersects its own rest frequency, and will vary at this same baseband rate. As an example, if a baseband audio tone is inserted at 2 kHz, the FM carrier will actually swing past its own rest frequency 2000 times in 1 second.

Unlike AM modulation, the percent of modulation for FM is directed by government rules and regulations, and not by any natural limitations. For instance, narrowband voice communications is considered to be 5-kHz deviation for 100% frequency modulation, while with wideband FM broadcast, a maximum allowed deviation is 75 kHz. But if the baseband signal's amplitude should induce the FM deviation to go above this 100% limit, then more frequency sidebands will be created, which broadens the bandwidth conceivably causing interference to any adjacent channels.

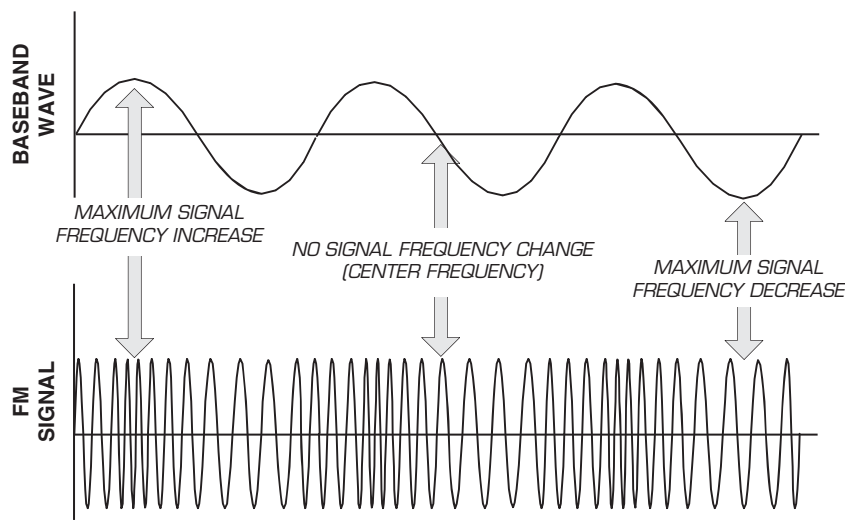


Figure 2.11 A baseband modulating signal's effect on a carrier during frequency modulation.

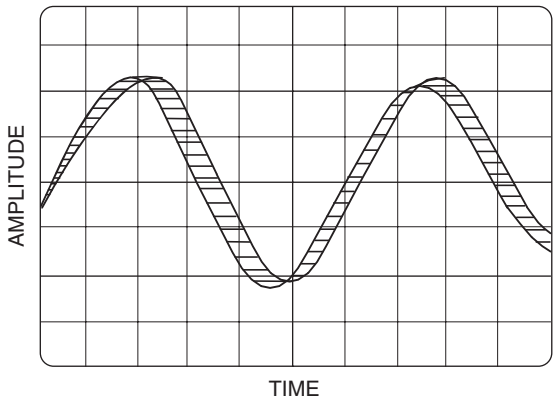


Figure 2.12 A time domain view of frequency modulation.

As shown in Fig. 2.12, when FM is observed on an oscilloscope in the time domain, the modulated RF carrier will not change in amplitude, but only in frequency (exaggerated here for clarity). These rapid frequency fluctuations are evidenced by the shortening and lengthening of the carrier's wavelength on the scope's cathode ray tube display, creating a blurring of the signal. And since wavelength equals the speed of light divided by the frequency, we can readily see that any shift in wavelength corresponds to a change in frequency.

The total FM transmitter power will always stay constant during baseband modulation, so the combined power or voltage in an FM signal will not vary whether it is modulated or unmodulated. However, any sidebands formed by the modulation must gain their power from the carrier itself. This carrier must then sacrifice some of its own power in the creation of these FM sidebands. For instance, let us assume that an FM transmitter is sending out an unmodulated carrier at 100 W. When the RF carrier is modulated by the baseband signal it must give some—or even all—of its power to these sidebands. Thus, the carrier and its *significant sidebands* must all total-up to the original 100 W that was present in the unmodulated carrier. At certain *modulation indexes* (see later in the chapter), the carrier itself will actually vanish, while the sidebands will now contain all of the power.

An infinite number of sidebands will be created during the modulation process, since the carrier is sent through an infinite number of various frequency or phase values by the continually changing baseband frequencies. This action produces an infinite amount of sideband frequencies, with even the amplitude of a single test-tone (which is changing only in a sinusoidal manner) having an infinite number of discrete amplitudes within a single cycle.

Due to the difficulties inherent in *infinite*, the concept of the *significant sideband* was created. Significant sidebands are any sidebands with an amplitude that is 1% or more of the amplitude of the unmodulated carrier. When a sideband is below this level it may be ignored, while the higher the amplitude of the baseband modulation the higher the number of these significant sideband frequencies produced.

However, unlike amplitude modulation, more than one pair of sidebands will be created for each single-tone modulation (Fig. 2.13). The sidebands are also separated on each side of the carrier and from each other by an amount that is equal to the frequency of the single-tone baseband signal.

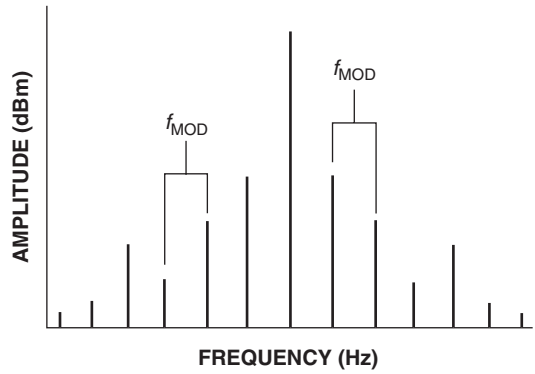


Figure 2.13 A single tone baseband signal creating multiple sidebands in FM.

The ratio between the FM carrier's instantaneous frequency deviation (f_{DEV}) divided by the instantaneous frequency of the modulation (f_{MOD}) is an important FM specification, and is referred to as the *modulation index*. The modulation index can tell us the number and amplitudes of all significant sidebands generated during FM modulation by simply reading the chart of Table 2.1. To use this table, first calculate the FM signal's modulation index by f_{DEV}/f_{MOD}; take this number and find its value under the "Modulation Index" column; now read across. The relative amplitude of the carrier, and each sideband with its number of significant sidebands, will be shown.

As the frequency of the modulation waveform increases, the amount of significant sidebands decreases. However, the actual bandwidth that such a signal consumes increases, since the sidebands spread out from the carrier and each other because they are equal to the frequency of the modulation waveform, which has just increased in frequency.

Modulation Index	Carrier	Pairs of Sidebands to the Last Significant Sideband							
		1st	2nd	3rd	4th	5th	6th	7th	8th
0.00	1.00	—	—	—	—	—	—	—	—
0.25	0.98	0.12	—	—	—	—	—	—	—
0.5	0.94	0.24	0.03	—	—	—	—	—	—
1.0	0.77	0.44	0.11	0.02	—	—	—	—	—
1.5	0.51	0.56	0.23	0.06	0.01	—	—	—	—
2.0	0.22	0.58	0.35	0.13	0.03	—	—	—	—
2.5	–0.05	0.50	0.45	0.22	0.07	0.02	—	—	—
3.0	–0.26	0.34	0.49	0.31	0.13	0.01	0.01	—	—
4.0	–0.40	–0.07	0.36	0.43	0.28	0.13	0.05	0.02	—
5.0	–0.18	–0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02

Table 2.1 Numbers and Amplitudes of FM Sidebands

We can also find the bandwidth of the modulated RF signal by multiplying the number of significant sidebands by two, then multiplying by the maximum modulating frequency, or $BW = 2N \times f_{MOD(max)}$.

The following is a basic example of the modulation index and its effect on what we might see in the frequency domain. With a modulation index of zero, we would be generating no sidebands at all (Fig. 2.14), since this would be just a simple continuous wave (CW) carrier with no baseband modulation. But as the modulation index increased to 1.5, we see in Fig. 2.15 that the sidebands will start to consume more bandwidth. This is a good example of why the frequency of the baseband modulation, and its amplitude, must be controlled in order to lower FM bandwidth demands and adjacent channel interference (ACI).

The two-way, narrowband FM radio modulation index is normally maintained to under a maximum of two, since the maximum allowed frequency deviation would be approximately 5 kHz, with a maximum baseband audio frequency of about 2.5 kHz. Thus, a bandwidth of between 12 and 20 kHz is customarily considered to be adequate to pass enough sideband power for ample intelligibility of voice signals. However, for FM broadcast use, the maximum deviation is set at 75, with the maximum audio frequency being 15 kHz and the deviation ratio at 5, or 75/15 kHz, as mentioned above. This allows for high fidelity music to pass without significant loss of quality.

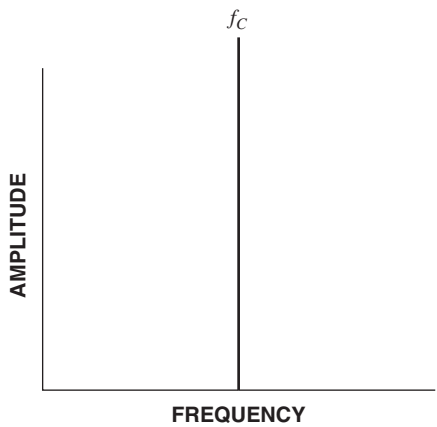


Figure 2.14 An FM signal with a modulation index of 0, or an unmodulated FM carrier.

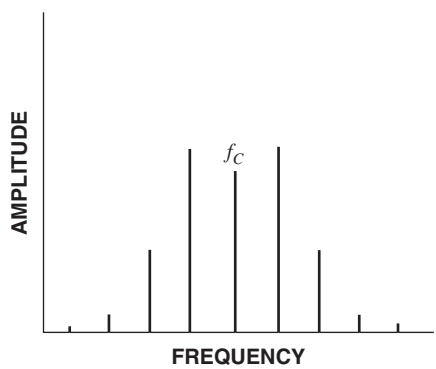


Figure 2.15 An FM signal with a modulation index of 1.5, showing multiple sidebands.

2.2.3 FM and AM Comparisons

Frequency modulation holds many benefits over amplitude modulation: Superior noise immunity helped by amplitude limiting to eliminate AM noise; decreased high frequency noise constituents due to preemphasis circuits, which boost the higher frequencies at the transmitter, along with deemphasis, which attenuates the now overemphasized frequencies at the receiver; FM's *capture effect*, which forces any undesired signal that is near, or at the same, frequency as the desired signal to be rejected. And, since FM does not have a delicate modulation envelope, as does AM, FM does not require Class A linear transmitter amplifiers, but instead can utilize the far more efficient saturated types in both its RF and intermediate frequency (IF) sections. Also, transmitter efficiency in FM is quite high, considering the transmitter itself can be modulated by low-level techniques, needing little baseband modulation power.

Frequency modulation does have its drawbacks. Increased bandwidth is necessary because of the additional sideband production over AM; the broadcast FM transmitter and receiver are more expensive to design and construct due to their higher frequencies of operation, along with higher stability requirements; bouncing the FM signal off the atmosphere's ionosphere creates distortion of the FM wave, so it is normally (unless repeaters are used) strictly a line-of-sight communications medium.

A reprise of some of the more important FM terms:

Center frequency, sometimes referred to as the *rest frequency*, is the FM transmitter's carrier frequency with 0% modulation.

Frequency deviation is the amount the RF carrier shifts from its center frequency in a single direction when modulated.

Frequency swing is the movement of the modulated carrier on *both* sides of the center frequency, or twice the *frequency deviation*.

Modulation index, which is employed when one tone, at a steady deviation, is transmitted, and is the ratio between the carrier's instantaneous frequency deviation divided by the instantaneous frequency of the modulation.

Deviation ratio is the ratio between the *maximum* frequency deviation, with 100% modulation, divided by the *maximum* audio modulation frequency.

2.3 Single-Sideband Modulation

2.3.1 Introduction

Due to the highly stable oscillators and phase-locked loops available today, along with advances in mass production and RF-integrated circuits, low-cost transmitters and receivers for SSB communications have become quite popular, and have almost completely replaced the older AM long-range voice communication systems.

2.3.2 SSB Fundamentals

Single-sideband suppressed carrier is a form of AM, but transmits only a single sideband, rather than the two sidebands and the complete carrier of amplitude modulation. In SSB the carrier, which holds no information, and the other sideband, which duplicates

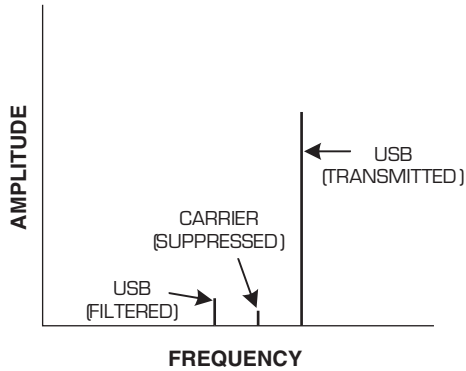


Figure 2.16 A single-sideband signal in the frequency domain.

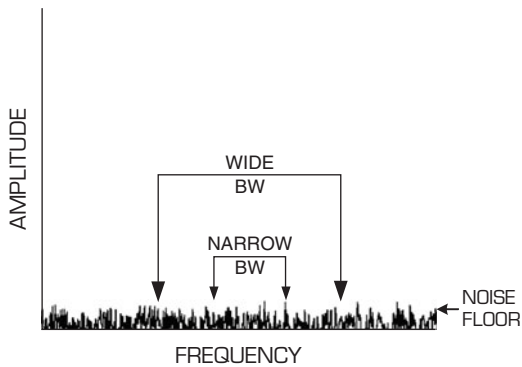


Figure 2.17 A narrower received bandwidth means less noise.

the information present in the transmitted sideband, are strongly attenuated (Fig. 2.16). Since one of the sidebands is attenuated, SSB requires only half the bandwidth that AM consumes for its transmissions, which also translates into less noise received (Fig. 2.17). Although amplitude modulation's fading characteristics are quite poor, but in SSB fading is much less of a problem. This is because the multiple phase dependencies between all of AM's transmitted elements—both of the sidebands and the carrier—need not be sustained in SSB, inasmuch as only a single sideband is actually being transmitted. Power efficiency is also much higher in SSB than in AM due to the power savings in transmitting only a single sideband, while further power is conserved since a transmitted signal is only produced when the baseband is actually present at the modulator.

2.3.3 SSB Modulation

The SSB transmitter (Fig. 2.18) creates a single-sideband signal by inserting both the oscillator-generated carrier (OSC) and a modulating audio signal from the AUDIO AMP into the *balanced modulator*. The balanced modulator nonlinearly combines, or mixes, the carrier and baseband inputs, producing both lower and upper sidebands. The modulator will also severely attenuate the carrier from the OSC stage by phase cancellation or common-mode rejection methods. The ensuing *double-sideband suppressed carrier* (*DSB*) signal is then injected into the next stage, which will be a USB/LSB filter.

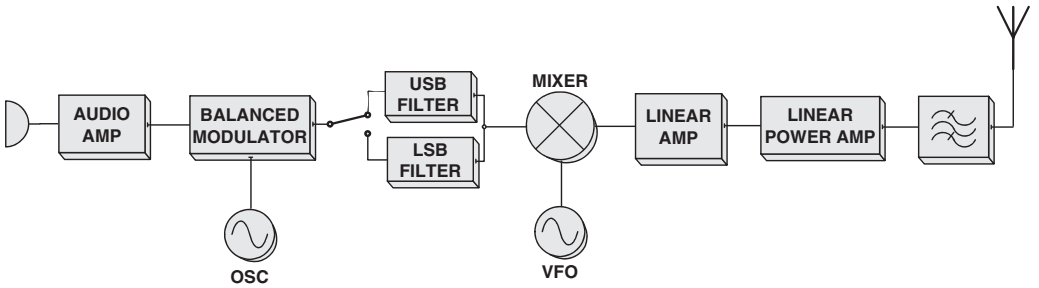


Figure 2.18 A typical filter-type single-sideband transmitter.

These filter stages of the SSB transmitter will be comprised of very selective bandpass filters that have a center frequency to pass either the upper or lower sidebands. There are nonfilter phase-cancellation methods that can be utilized to reject the undesired sideband by twin balanced modulators and phase-shifter circuits. Either way, the SSB signal will then be upconverted, amplified, and sent out through the antenna. However, since the modulated signal will contain a nonconstant amplitude modulation envelope that can easily become distorted, linear amplifiers must be utilized throughout a SSB system's signal path.

The RF signal is then picked up at the SSB receiver's antenna, filtered, amplified, and downconverted (Fig. 2.19). The signal is inserted into a type of nonlinear mixer called a *product detector*, along with the *carrier oscillator* (or beat frequency oscillator [BFO]) frequency to supply the missing carrier, with the output being a baseband signal. This baseband is amplified and sent on to headphones or to speakers.

Depending on the number of modulating tones and their amplitudes, different time-domain outputs can be viewed on an oscilloscope. If a single baseband tone is injected into a SSB modulator, a steady RF signal in both amplitude and frequency will be created, as in Fig. 2.20. This is simply a CW signal. However, a two-tone baseband

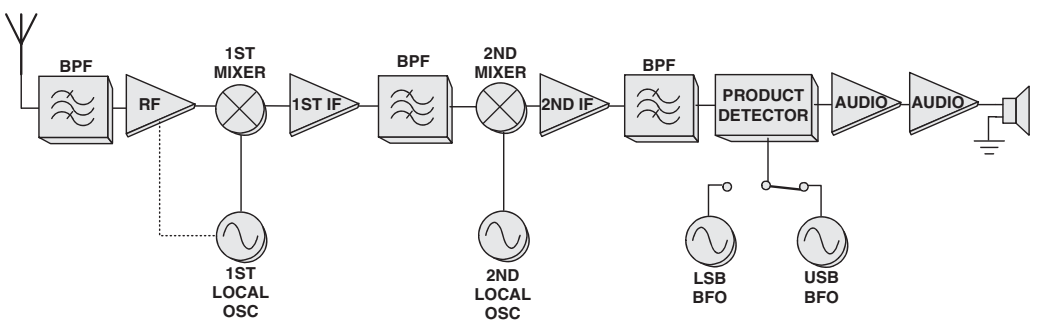


Figure 2.19 A typical dual-conversion single-sideband receiver.

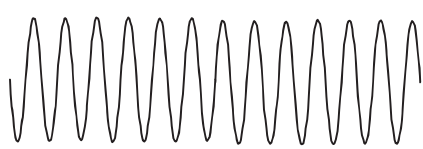


Figure 2.20 A single-tone SSB signal in the time domain.

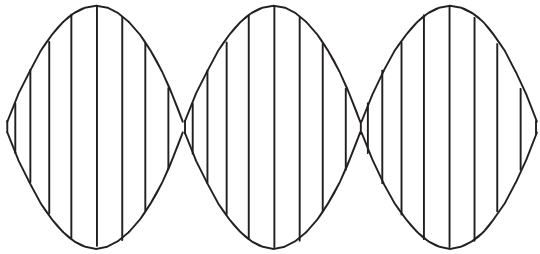


Figure 2.21 A two-tone SSB signal showing its modulation envelope.

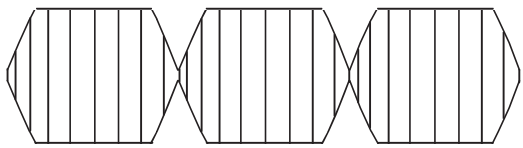


Figure 2.22 An overmodulated two-tone SSB signal.

signal will generate the consummate SSB modulation envelope display of Fig. 2.21, with the amplitude of the modulation envelope dependent on the baseband modulation level. The two-tone RF signal will start to flattop (Fig. 2.22) if overmodulation occurs, causing extreme distortion and spurious outputs.

2.3.4 SSB Output Power

The measurement of output power in SSB is the same as in AM. The *peak envelope power* being the measurement of the average peak power of the transmitted signal with 100% modulation. PEP can be calculated by V^2_{RMS}/R, or $V_{RMS} \times I_{RMS}$, or $I^2_{RMS} \times R$ of the maximum modulated peak.

2.4 Phase Modulation

2.4.1 Introduction

Phase modulation (PM) is used as the basis for almost all digital modulation in use today. PM varies the instantaneous phase of an unmodulated RF carrier in order to modulate it with a lower frequency analog (Fig. 2.23) or digital baseband signal (Fig. 2.24).

2.4.2 PM Fundamentals

When using analog baseband waveforms to phase modulate a carrier, the carrier is phase shifted continuously through an infinite number of phase states. As soon as the analog-modulated waveform moves positive, the RF or IF carrier will change phase in one

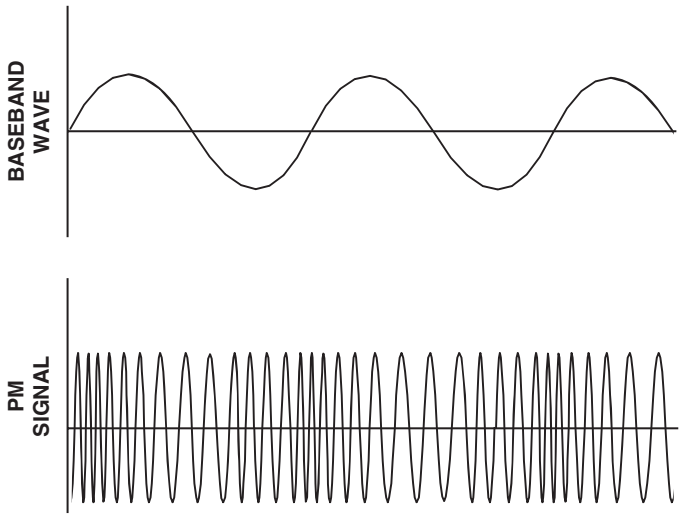


Figure 2.23 Analog PM modulation in the time domain.

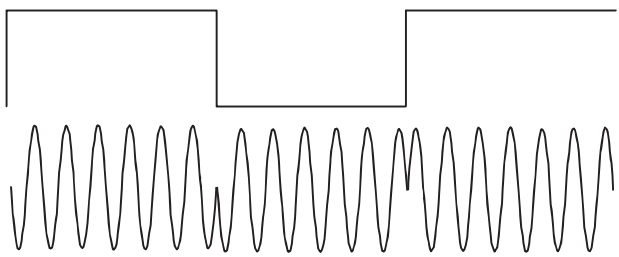


Figure 2.24 Phase modulation by a digital baseband signal (phase shift keying).

direction by lagging the phase of the unmodulated carrier by a phase angle magnitude that is dependent on the modulation's amplitude. And as soon as the modulation waveform moves negative, the carrier will change phase in the reverse direction by leading the phase of the unmodulated carrier by a phase angle magnitude that is dependent on the modulation's amplitude. Thus, the exact phase angle of the carrier at any given time is dependent on the amplitude of the baseband modulation input.

The smooth analog-modulated phase transitions, when viewed in the time domain, are virtually identical to regular analog frequency modulation. This is because, as with both FM and PM modulation, the following peak of the modulated carrier will occur at a different point in time than if the carrier were not phase or frequency modulated at all (i.e., just a simple CW signal). So, *FM* modulation will cause a *phase shift* of the RF carrier, while *PM* modulation will cause a *frequency shift* of the RF carrier. Both of these effects are incidental, however, and are but an unavoidable result of the modulation itself.

The carrier's phase is forced by the baseband modulation to change at the pace of the lower baseband modulating signal's frequency, and the extent of this phase shift, in any direction, is equal to the amplitude of the modulating waveform. Thus, if we have a 10-kHz modulating signal that has a 1 V amplitude, and this modulating signal causes

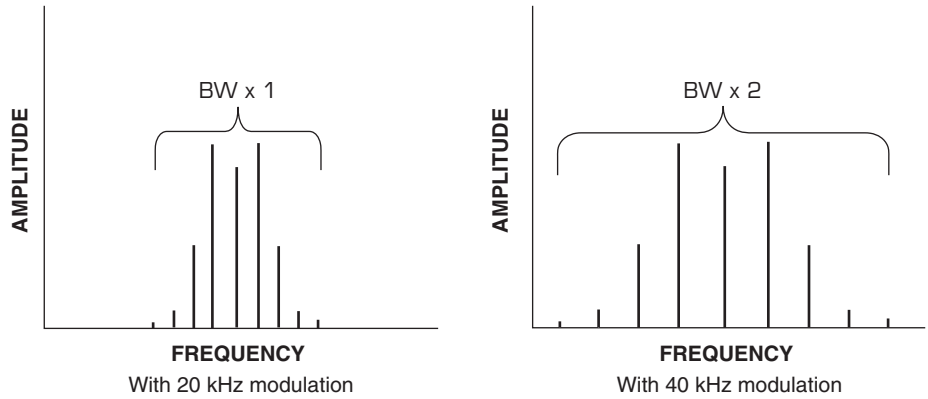


Figure 2.25 Phase modulation in the frequency domain.

the carrier to shift in phase by 10°, then the phase of the carrier will actually pass through 0° phase on its way to leading and lagging the phase angle at a rate of 10,000 times in 1 second, and the maximum magnitude of the leading and lagging phase-shift will take place at the maximum amplitude point of the modulating signal.

Phase modulation, as in frequency modulation, has an infinite number of sidebands in the frequency domain, and these sidebands are spaced evenly from each other at a frequency that is the same as that of the modulating signal itself. In the frequency domain, the only visible difference between PM and FM is that the power of each PM sideband is dependent on the modulating waveform's amplitude. Consequently, the significant number of PM sidebands and their amplitude ratios are the same, no matter what the modulating frequency (Fig. 2.25). This would also imply that as the frequency of the PM modulation is increased, the bandwidth of the transmitted signal increases. Therefore, the frequency deviation of the carrier during phase modulation is dependent not only on the amplitude of the modulating signal, as in FM, but also on its frequency.

2.4.3 PM Disadvantages

PM, when subjected to the rapid On/Off switching of the baseband digital modulation, causes rapid phase changes that widen the signal's bandwidth through the splattering of RF energy over a wide spectrum. In digital radios, much attention and effort is directed toward the filtering of the baseband signals in order to substantially decrease this bandwidth.

2.5 Digital Modulation

2.5.1 Introduction

With the advent of digital modulation techniques, far higher data rates are possible within constrained bandwidths, and at higher reliability and noise immunity levels, than the older modulation methods of FM, AM, *FSK* (*frequency shift keying*), *OOK* (*on-off keying*), *PWM* (*pulse width modulation*), *PPM* (*pulse position modulation*), *PAM* (*pulse amplitude modulation*), and the like.

Contemporary digital modulation have much in common with some of these older discrete modulation techniques, such as OOK and FSK. Both old and new modulation methods possess discrete states at discrete times, whether these particular states are amplitude, phase, frequency, or amplitude/phase driven, and will define the information being transmitted. Indeed, it is the number of these possible states that governs the amount of data that can be transmitted across an RF link.

2.5.2 Digital Modulation Types

Modulation methods, in general, can most easily be viewed with a *phasor diagram* (Fig. 2.26). *I* is the in-phase (0°) reference plane, while *Q* is the quadrature (90°) reference plane, and in-between these two *I* and *Q* states is the signal (*S*), which can vary in phase (θ) and amplitude (*A*). Since any digital modulation will alter either the phase or amplitude (or both) of a carrier, this is a very effective way to visualize various modulation methods.

As an example of analog modulation versus digital modulation, Fig. 2.27 displays the phasor diagram of analog phase modulation which, since it possesses no information in the carrier's amplitude, presents a full circle as the carrier rotates from 0° to 360°. On the other hand, since analog AM modulation contains no phase information within its carrier, its phasor will only vary in amplitude (Fig. 2.28).

More efficient digital-like methods of analog modulation are possible by using only discrete states in the phasor diagrams. The simplest, as stated above, is OOK modulation (Fig. 2.29), a type of *ASK*, which is commonly employed to send human-readable Morse code. OOK can also be used to transmit machine-readable 1s and 0s, permitting 1 bit of data to be sent between each discrete amplitude transition.

Another basic digital modulation is *noncoherent FSK*, which involves shifting the frequency of, typically, a voltage-controlled oscillator to two (or more) different discrete frequencies (Fig. 2.30). FSK is not only extremely simple to implement at both the receiver's and the transmitter's baseband sections, but can also take advantage of fully saturated, and thus highly efficient, power amplifiers at the transmitter's output stage.

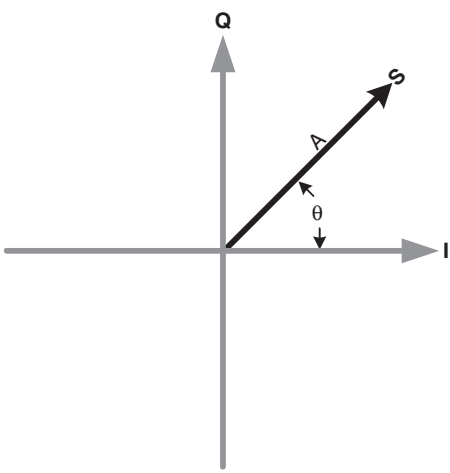


Figure 2.26 A phasor diagram.

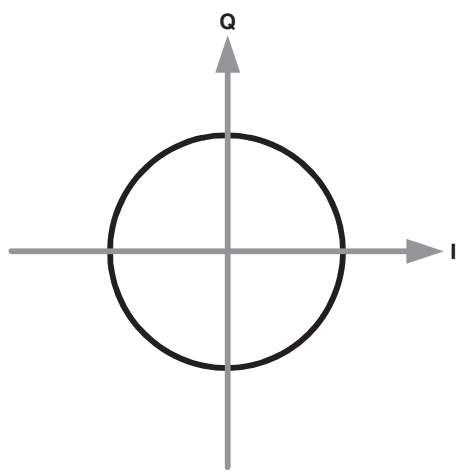


Figure 2.27 A phasor diagram of phase modulation.

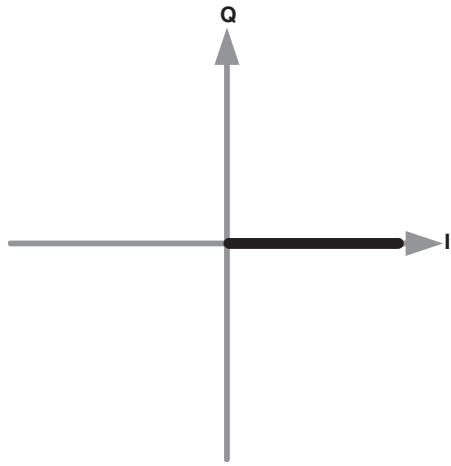


Figure 2.28 A phasor diagram of amplitude modulation.

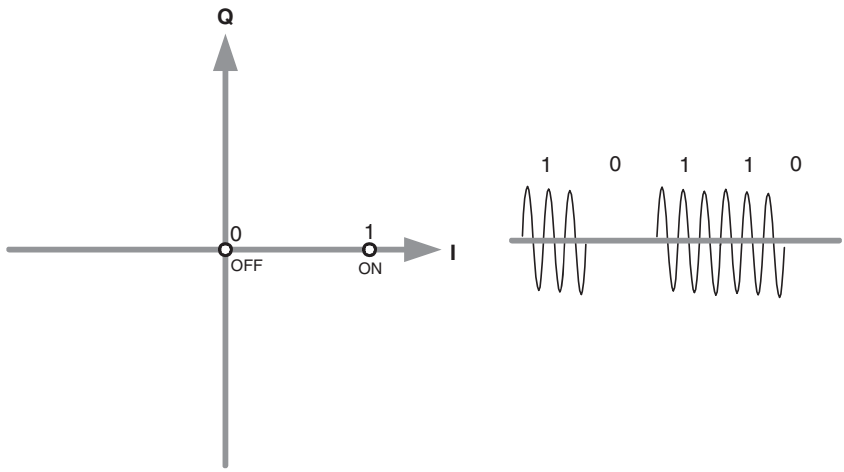


Figure 2.29 A phasor diagram of OOK modulation with accompanying time domain sine

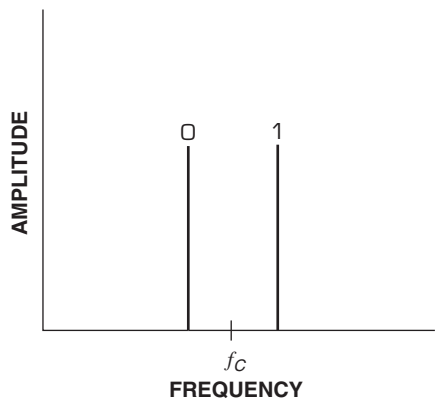


Figure 2.30 Frequency shift keying (FSK) in the frequency domain.

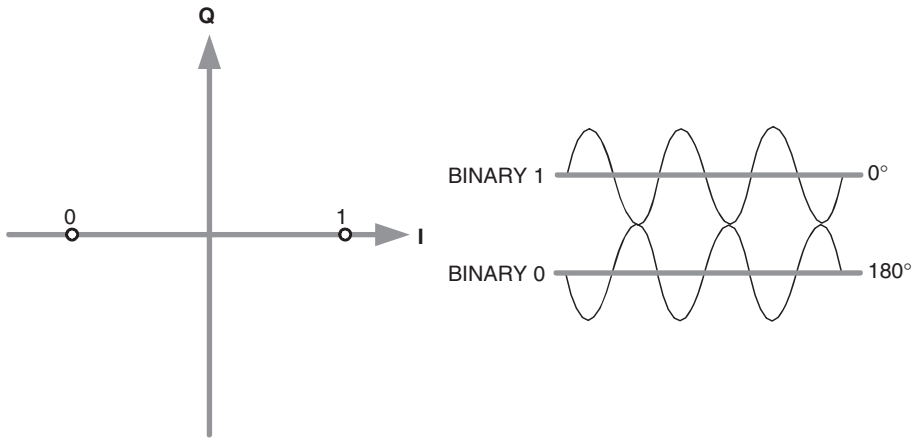


Figure 2.31 A phasor diagram of BPSK modulation with accompanying time domain waveforms.

However, it suffers from a demanding bit error ratio (BER) requirement and low spectral efficiency. FSK can be found in many low-cost, low-data-rate wireless systems, such as cordless phones, wireless audio speaker systems, and RF consumer telemetry systems.

Instead of varying the RF carrier's amplitude in discrete states to transmit information while leaving the phase unaffected, as in OOK, we can reverse this concept and ignore the amplitude of the carrier, while changing the phase of the signal in discrete states. In order to decrease the effects of phase jitter, these phase states should be as far removed from each other as possible, which is why 0° and 180° are commonly used (Fig. 2.31). This type of modulation is the most basic of what is considered to be "true" digital modulation, and is referred to as *binary phase shift keying* (*BPSK*), with a 0° reference phase indicating a 1, and a 180° discrete state indicating a binary 0.

When we modulate the phase of a carrier into four discrete states, we have *quadrature phase shift keying* (*QPSK*). As shown in Fig. 2.32, four discrete phase states

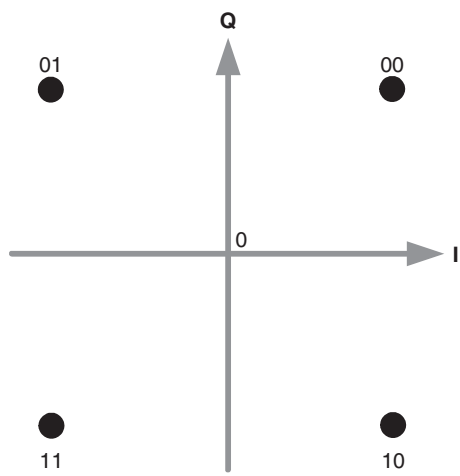


Figure 2.32 A constellation diagram of a QPSK signal.

have been selected to convey information, unlike analog phase modulation, which has infinite phase points as it rotates from 0° to 360°. These four discrete states for QPSK can be 0°, 90°, 180°, and 270°, and are located on a constant amplitude carrier. The four states supply 2 bits for each shift of phase (00, 01, 10, 11), instead of 1 bit (1, 0) as in the above BPSK system. This technique would clearly contribute double the information within the identical bandwidth and time period.

However, when we say that the carrier of a QPSK signal is of a "constant amplitude" during modulation, this is not quite true. Amplitude variations may play no role in actually transferring information across a QPSK-modulated wireless link, but amplitude variations of the carrier do occur, since QPSK includes AM components within its modulation envelope. We will go into this important issue in further detail below.

Quadrature amplitude modulation (*QAM*) is the most widespread digital modulation method in use today for sending data at very high bit rates across terrestrial microwave links, and employs a combination of amplitude and phase modulation. QAM utilizes various phase shifts to the carrier, with each of these phase shifts having the ability to possess two or more discrete amplitudes. In this way, every amplitude/phase combination can symbolize a different and distinct binary value. As an example, if working with QAM-8, a digital value of 111 could be represented by a carrier that displays a phase shift of 180° and an amplitude of +2; or 010 can be symbolized if the phase is shifted to 90° with a amplitude of −1. QAM-8 exploits four phase shifts and two carrier amplitudes for a total of eight possible states, or 3 bits of 000, 001, 010, 011, 100, 101, 110, and 111 that can be transmitted. Another example of quadrature amplitude modulation, QAM-16, is shown in Fig. 2.33, and provides 4 bits per AM/PM change.

More data can be transmitted within an allocated bandwidth or time period as the number of AM/PM states are increased, since more bits per change can now be encoded. But as the number of the AM/PM states are increased, each state is forced closer together, and thus noise will become more of a problem for the signal's BER. This means

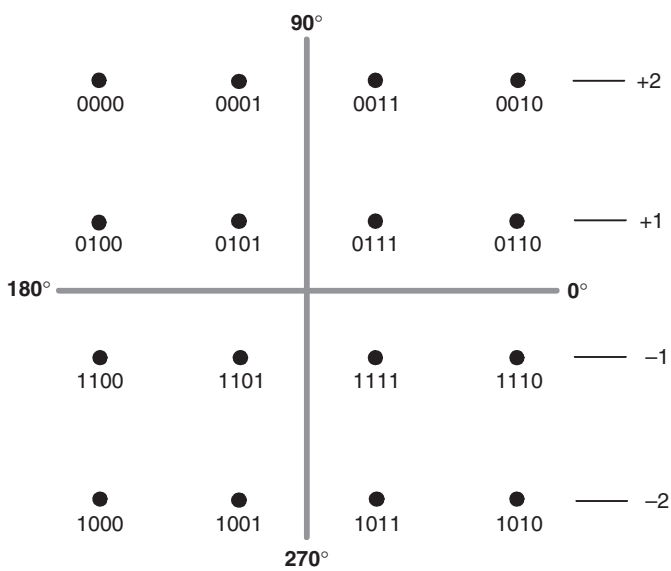


Figure 2.33 A constellation diagram for 16-QAM.

that the higher the QAM state, the more it can be subjected to damaging interference under comparatively low-noise conditions, since the signal's signal-to-noise ratio (SNR) must be increased for each increase in the number of states so as to maintain the BER. For this reason long-range satellite communications, which is a high-noise medium (due to low output powers), will normally use the simpler modulation schemes, such as BPSK or QPSK. Low-noise terrestrial radio links can use 32, 64, 128, and 256 QAM (or above) to send data at much higher rates than possible with the older digital systems, and within the confined bandwidth restraints of most wireless information channels.

Due to the noiselike nature of all true digitally modulated signals, viewing such signals in the frequency domain will usually not tell us much about the complexity of the modulation, such as whether it is QPSK, QAM-16, or QAM-256, but only the signal's amplitude, frequency, flatness, spectral regrowth, and the like. In fact, a digital signal will normally appear the same whether or not it is transporting any data at all. This is due to the coding and encryption added to the RF carrier signal.

2.5.3 Digital Modulation Power

In describing the digital modulation formats as shown above, we are displaying either a static constellation diagram or a phasor diagram at a discrete point in time, and completely ignoring how these transitions from symbol to symbol occur. These transitions are especially important in the common four-phase state QPSK modulation format and its variants, because this governs whether this particular modulation will have a constant modulation envelope, or one that varies in amplitude. Indeed, a constant amplitude modulation envelope will allow the use of an efficient, near-saturated power amplifier that does not need to be backed off in its power output, while a QPSK modulation with a nonconstant amplitude requires a highly inefficient, linear amplifier that must be heavily backed off from its maximum power output in order to avoid massive spectral regrowth. Excessive spectral regrowth, a form of intermodulation distortion (IMD), will cause an otherwise legal signal to create interference to channels on either side of its own channel, and with a bandpass that no longer fits within the mandated federal communications commission (FCC) spectral limits.

The problem of a nonconstant amplitude RF carrier during QPSK modulation occurs because the carrier of regular QPSK will sometimes pass through zero amplitude on its way from one phase state to another, as displayed in Fig. 2.34, which causes the QPSK modulation envelope to vary in amplitude, thus disallowing the use of a nonlinear amplifier due to the excessive IMD production. This is not the same thing as in QAM, in which much of the information to be conveyed is actually *contained* within the amplitude variations of the signal. In QPSK, amplitude variations are but an annoying side effect of this method of digital phase modulation.

Using amplifiers that do not have to be as extensively backed off in power can be accomplished by employing a less bit rate–efficient modulation, such as *offset QPSK* (*O-QPSK*). O-QPSK modulation allows only changes in phase states that do *not* pass through the zero amplitude origin. Referring back to Fig. 2.32, this would mandate symbol changes between 00 and 01, or between 10 and 11, but not between 11 and 00, or 01 and 10. Therefore, O-QPSK does not allow the carrier to be completely snuffed out, and will undergo carrier amplitude variations of at most 3 dB. As stated above, this permits the use of more efficient nonlinear amplifiers, with higher output powers possible, and with less spectral regrowth.

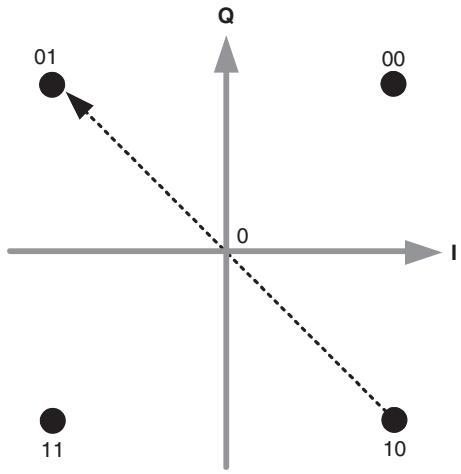


Figure 2.34 A QPSK signal passing through zero amplitude, quenching the carrier.

Another similar QPSK modulation format, referred to as $\pi/4DQPSK$, will also not permit the carrier to be completely suppressed, and in addition, allows for easier clock recovery during demodulation. Other modulation schemes have a modulation envelope that is completely constant envelope, and can thus employ very efficient nonlinear Class C amplifiers without producing any spectral regrowth. Two of these modulations are *minimum shift keying* (*MSK*) and its derivative, *Gaussian minimum shift keying* (*GMSK*).

Digital Power Measurement

The average power output of an analog amplitude–modulated transmitter will vary depending on the baseband waveform. Its *peak envelope power* need not be averaged over time, but can easily be measured over a single cycle of the signal's waveform using commonly available, low-cost test equipment. This is not quite the same situation with digitally modulated signals. The peak amplitude of a digital signal is completely unpredictable, and will vary dramatically due to its noiselike nature. To find the peak amplitude of these erratic signals, the power measurement must be taken over time to obtain a *statistical* peak amplitude, since there is only a statistical chance that a much higher peak will come along at any instance in time. This peak reading is then compared to the digital signal's *average* power. The average power is the same power that a DC signal would require to heat up a resistance element to the same temperature as the RF signal. With this type of average power test, any variation in wave shape can be measured.

If an RF signal did not vary its power over time, as it is with DC signals in which $P_{DC(PEAK)} = P_{DC(AVG)}$, then the average power would be delivered constantly to the load, with its peak power equaling its average power. But since RF signals do vary (as any AC signal will), their peak power will be quite different from the average power. This ratio between the peak and the average values of a modulated signal is referred to as

the *peak-to-average* ratio. The less this ratio the closer to the amplifier's *P*1dB an amplifier can be driven without producing excessive IMD products, since the occasional power peaks will be lower in amplitude with a modulation format that has a lower peak-to-average ratio. Thus the amplifier will not have to possess as much of a power margin to gracefully accept these higher amplitude power spikes, and so as not to create inordinate levels of IMDs. A signal with very high absolute peaks, but a low average power, will have a very poor peak-to-average ratio, forcing any amplifier that is working with this type of signal to have a large amount of reserve power to amplify, without excessive distortion, these occasional peaks. However, because these peaks are caused by the modulation shifting from one constellation point to another, and therefore will be erratic over time, the average power in a digital signal will be constant due to the digital signal's own encoding.

As stated above this peak-to-average power ratio, if high, can be excessively inefficient, since it forces the amplifier to be designed for, and to run at, higher output powers just to take care of these occasional peaks. The peak-to-average ratio, however, will vary with the symbol patterns and clock speeds, as well as the channel filter and bandwidth. General peak-to-average ratios, nevertheless, can be approximated to be about 5 dB for QPSK, 8 dB for QAM-64 and *orthogonal frequency division multiplexing* (*OFDM*), and up to 15 dB for *code division multiple access* (*CDMA*). This means that for some QPSK modulations, the intermittent peaks will rise above the root mean square (RMS) power by 5 dB.

To measure this peak-to-average ratio, first measure the power peaks with a fast-acting, digital-modulation capable, peak-reading power meter (such as the *Boonton 4400*), over a time of about 10 seconds. This will give a reasonably accurate indication of the signal's peak power. To measure the average power, see "Digital Signal Power Measurement" in Chap. 12, or use a special digital-modulation compliant *average power meter*. Now subtract the average from the peak, in dBm. This equals the peak-to-average ratio in dB.

Therefore, digital signals, due to their nonrepetitive, random nature, as well as the fact that all of their power is being spread out over frequency, rather than condensed in sidebands on both sides of the carrier as they are in analog signals, have no predictable, recurrent peak power points to measure. Since it is difficult to measure these peaks in a nonstatistical manner, we are forced to take an average measurement of the digital signal's power over its entire bandwidth. But how much bandwidth does the normal digital signal consume? It is considered to be the −30-dB bandwidth where most of the power of the digital signal resides, rather than just within the −3-dB bandwidth used in most analog signal measurements.

Any digital signal we work with will also have been filtered, since any unfiltered (i.e., ideal) digital signal would theoretically take up infinite bandwidth. Filtering a digital signal will force the signal to go from a square wave to a more rounded waveform, which allows the signal to be inserted into a much narrower bandwidth. But this also increases the required power that the power amplifier (PA) of the transmitter must, occasionally, transmit. In fact, this filtering is mainly responsible for the peak-to-average problem we just discussed, along with the signal passing through the origin, so the more filtering of the digital square wave we employ to decrease the transmitted bandwidth, the more we create a higher peak-to-average ratio (Fig. 2.35). More details on the filtering of digital signals is explained in "Digital Modulation Issues."

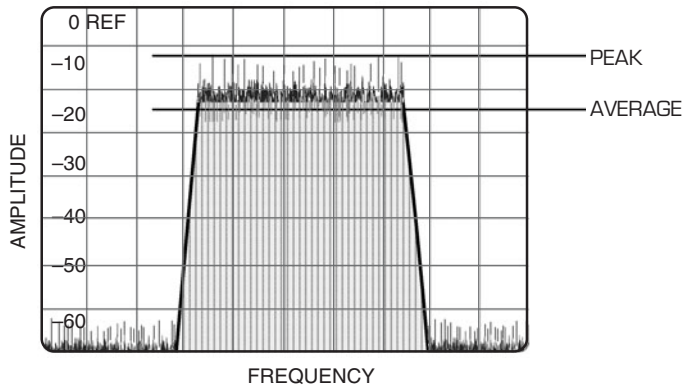


Figure 2.35 Peak and average amplitudes of a digital signal.

2.5.4 Digital Modulation Issues

In portable digital communications, everything is geared to not only supplying reliable communications at the lowest transmitted power and highest bandwidth efficiency possible, but also by maximizing the data rate. In fact, bandwidth, power, noise, and information capacity are all interrelated by *Shannon's information theorem*, which states that the speed of information transfer is limited by the bandwidth and the signal-to-noise ratio of a communication's channel:

$$C = W \ \mathrm{LOG}_2 \left(1 + \frac{S}{N}\right)$$

where C = capacity of the data link, *bps* (bits-per-second)
W = bandwidth of the channel, *Hz*
S/N = signal to noise *ratio*, not in dB

Indeed, Shannon's information theorem makes clear the following:

A. As we increase the information rate while under the presence of noise, we can avoid raising the BER by increasing the SNR.

B. We can increase the information rate by widening the bandwidth.

C. We can obtain an infinite information rate in a completely noiseless channel of *any* bandwidth, since the SNR would now be infinite as well. This does not infer that theoretically increasing our bandwidth to infinity would also permit our data rate to become infinite, since the noise level would increase as well.

The Shannon's theorem thus has a very practical application for any RF design engineer, since it clearly demonstrates that we can juggle bandwidth for RF power, and vice versa, in order to maintain a desired performance. It also clearly supplies us with an answer to "what is the highest theoretical data bit rate that can be sent over a communications channel of a specific bandwidth, under the influence of white additive noise, with a small BER, using a limited average signal power, per second."

Depending on error correction techniques, Table 2.2 displays different E_b/N_0's required of the various modulation formats to sustain a desired BER.

BER	E_b/N_0 (SNR in Parenthesis, dB)					
	QPSK	16 QAM	32 QAM	64 QAM	128 QAM	256 QAM
10^{-4}	8 (12)	13 (18.8)	15	17 (25)	19	21.5
10^{-5}	10	14	16	18	20.5	23
10^{-6}	11 (14)	15 (21)	17	19 (27)	21.5	24
10^{-7}	12	15.7	18	20	22.5	25
10^{-8}	12.5 (15.7)	16.2 (22.3)	18.5	21 (28.5)	23.5	25.7
10^{-9}	13	16.5	19	21.5	24	26.2
10^{-10}	13.2	16.7	19.2	21.2	24.2	26.5

Approximate and calculated with no channel coding or channel fading.

TABLE 2.2 Graph of E_b/N_0 vs. BER for QPSK and QAM

Since the *symbol rate* equals the *bit rate* divided by number of bits represented by each symbol, then a modulation format such as BPSK would transmit at a symbol rate that is equal to its bit rate, or bit rate = symbol rate. However, in modulation schemes that encode more than 1 bit per symbol, such as QPSK (2 bit/baud), the baud rate will be less than the bit rate (in this case, half). This, as discussed above, allows more data to be transmitted within a narrower bandwidth.

The *modulation index* (h or bits/symbol), also referred to as *bandwidth efficiency*, is measured in bits/s/Hz. The higher this modulation efficiency, the higher the data rate that can be sent through a certain fixed bandwidth. For instance, BPSK has an h of one, while QAM-64 has an h of six. However, a higher h comes at the expense of higher equipment cost, complexity, linearity, and an increased SNR required to maintain the same BER as the lower h systems. Table 2.3 displays the various common modulation schemes and their h values, number of states, amplitudes, and phases.

Adaptive equalization will correct certain signal impairments in real time, such as *group delay variations* (*GDV*), amplitude tilt, ripple, and notches. However, it will not improve impairments created by a nonlinear amplifier, noise, or interference, but will mitigate the sometimes massive multipath effects that can render a digitally modulated signal unreadable. Indeed, adaptive equalization is basically a dynamically varying adaptive filter that corrects the received signal in amplitude, phase, and delay, making

Type	Bits/Sym(h)	States	Amplitudes	Phases
BPSK	1	2	1	2
QPSK	2	4	1	4
8 PSK	3	8	1	8
16 QAM	4	16	3	12
32 QAM	5	32	5	28
64 QAM	6	64	9	52

TABLE 2.3 Common Modulation Schemes and Their Properties

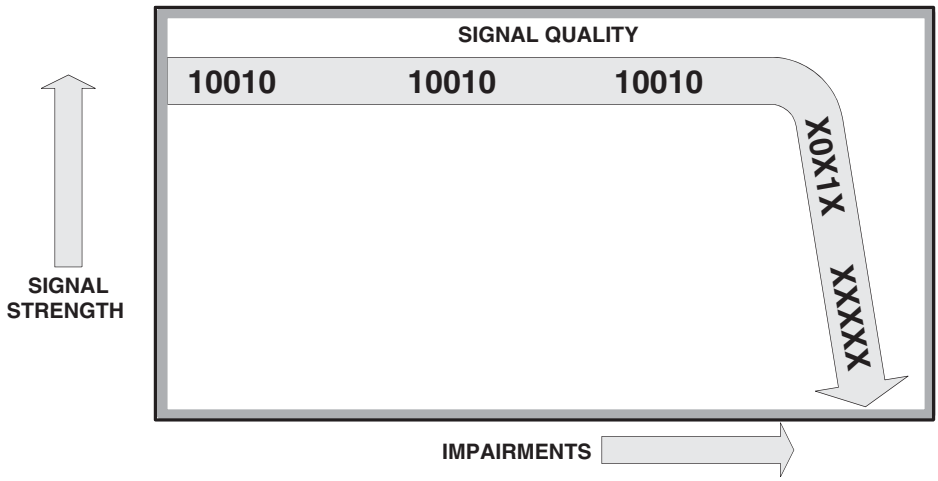


Figure 2.36 A digital signal and the cliff effect.

high-density modulations possible over extended ranges. Virtually, all terrestrial microwave communication systems employ some form of adaptive equalization located right after the receiver's demodulator.

Due to their nature, digital signals can maintain a relatively high quality at the receiver, even when very close to becoming unreadable due to various impairments. This makes the simple testing of a single digital signal at the receiver of little use, since the signal may actually be only a few dB in signal strength from crashing the entire link. This is referred to as the *cliff* (or *waterfall*) *effect*, and demonstrates why the BER of a digital signal will lessen to unacceptably high levels quite rapidly (Fig. 2.36).

However, digital communication systems can be examined for proper operation by sending and receiving certain digital test patterns that incorporate a recurring succession of logical 1s and 0s. The test then compares the impaired received pattern to the perfect transmitted pattern over time. The BER can then be established by contrasting the bits received that were incorrect with the total number of actual bits received.

Degradation in digital signal quality can be caused by many things. Reflections off of metallic surfaces (multipath), which produces amplitude ripple within the signal's passband; inadequate signal strength at the receiver, creating a decreased SNR, caused by the transmitter power levels being too low, or a high receiver noise figure (NF), or signal path attenuation caused by trees, weather, or Fresnel zone clearance problems; GDV and amplitude ripple, which can be produced by improperly designed or implemented analog filtering; increased noise floor under the desired signal caused by strong phase noise components in the frequency synthesizers of the conversion stages (reciprocal mixing); or noise and cochannel interference levels induced by interferers of all types.

Since many communications systems live or die by their BER figures, it is worthwhile to not only summarize what the dominant causes of BER degradation are in a digital communications system, but also to dig a little deeper into the reasons behind this increase in bit error rate. Decreased signal-to-noise ratio is the main mechanism for poor BER, since noise will smudge the symbol points, making their exact location hard to distinguish by the receiver's demodulator. Phase noise, another important contributor,

will also cause an input signal into a radio's frequency converter stage to be slightly changed at its output, as well as the reciprocal mixing issue mentioned above.

Phase noise is introduced by real-world local oscillators, which are not perfect single CW frequency sources. Since digital signals carry their information in the phase of the signal, this insertion of phase variances will create increased BER, with the density of the modulation affecting the severity of the bit error rate degradation. In other words, the higher the order of the QAM constellation (such as QAM-256) the more sensitive it is to relatively small levels of phase noise, since the constellation points are so densely packed, causing the phase/amplitude points to bisect digital decision boundaries. Another major impairment, IMD, will induce noiselike sidebands in a digital system, increasing distortion and decreasing the SNR, which degrades the BER, as well as creating *adjacent channel interference* (Fig. 2.37). *Group delay variations*, especially at the band edges of the system's analog filters, can add significant BER problems, since GDV forces the digital signal to arrive at different times at the system's output. *Ripple*, which are amplitude variations in the passband of an analog filter caused by poorly designed or implemented filter structures, can produce a high BER in many digital systems. Multipath itself will also cause both amplitude impediments (ripple and notches) and phase distortions of the signal through the process of phase cancellation, and decreases the received signal strength, lessening SNR and therefore degrading the BER.

Consequently, so as not to adversely influence the BER of phase/amplitude modulated signals, high density digital communications systems must be designed for

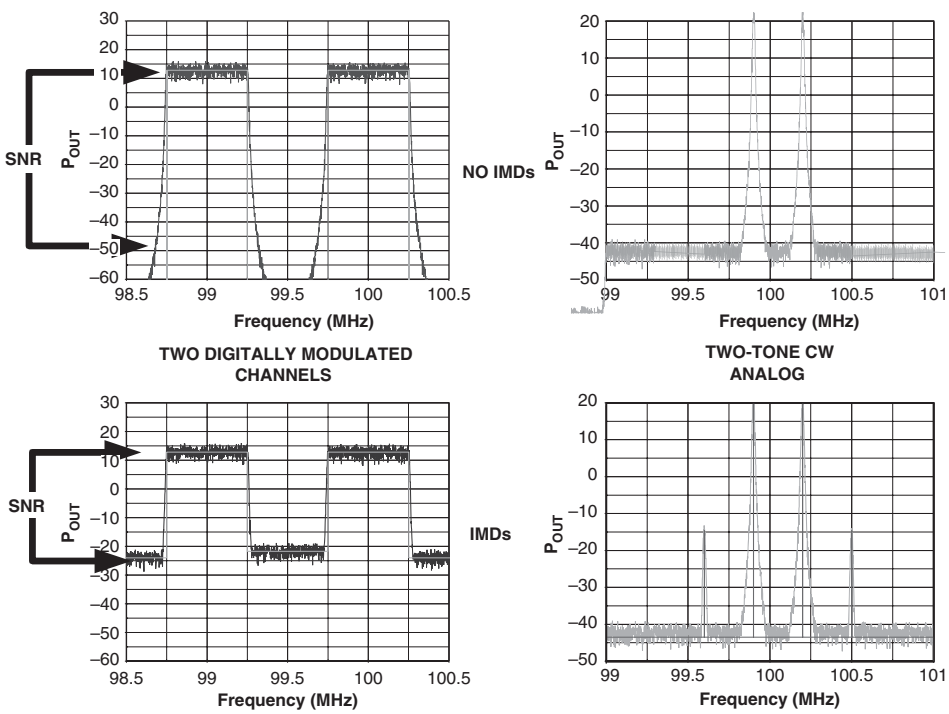


Figure 2.37 Digital signals as affected by IMDs, along with a standard CW two-tone display for comparison.

low levels of phase-noise, GDV, IMD level, amplitude ripple and shape, frequency deviation, multipath, and for the highest possible signal-to-noise ratio.

Baseband Filtering

Another very important issue in digital modulation that has, as yet, only been touched upon, is the effect that the baseband filtering has on the digital signals. Again, this filtering is employed to limit transmitted bandwidth to reasonable or legal levels. Our example for the following discussion will be with *filtered QPSK*.

As shown in Fig. 2.38, a quadrature modulator for QPSK transmitters receives a data bit stream, which is then inserted into the *bit-splitter*. The bit-splitter sends the *odd* bits to the I input of the quadrature modulator chip, and the *even* bits to the Q input. However, before exiting the modulator, these bits must first pass through a lowpass filter, which rounds off the bits' sharp rise and fall times. This shaping of the digital signal before it enters the I/Q modulator chip helps to avoid interference to the important central lobe of the RF or IF digital signals, and to specifically reduce the bandwidth that will exit the modulator chip. Notwithstanding, band limiting can also be added through a bandpass filter at the modulator's output, after the I and Q are linearly added in the combiner, along with the main lowpass filters in the I and Q legs.

Even at the receiver's demodulator, filtering is taking place. In fact, the filtering and band shaping is typically shared among the transmitter and receiver. The transmit filter reduces the *adjacent channel power* (*ACP*) into nearby channels, while the receive filter reduces the effect of ACP and noise on the received signal itself. This scheme is used in order to obtain a low ISI and BER by allowing an almost zero group delay variation from the input of the transmitter to the output of the receiver.

At the receive end, one method for demodulating an incoming received QPSK input is displayed in Fig. 2.39. The IF or RF enters the demodulator's input, where the signal is split into two paths and enters their respective mixers. Each mixer's local oscillator frequency (LO) input is fed by the *carrier recovery circuit*, which strips the carrier from the incoming signal at the exact frequency that the transmitted signal would be after going

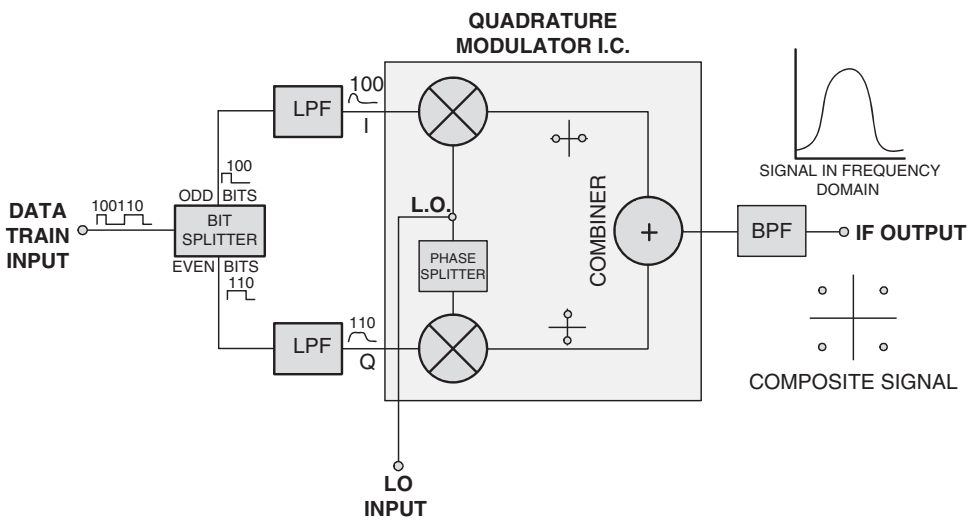


Figure 2.38 A QPSK modulator.

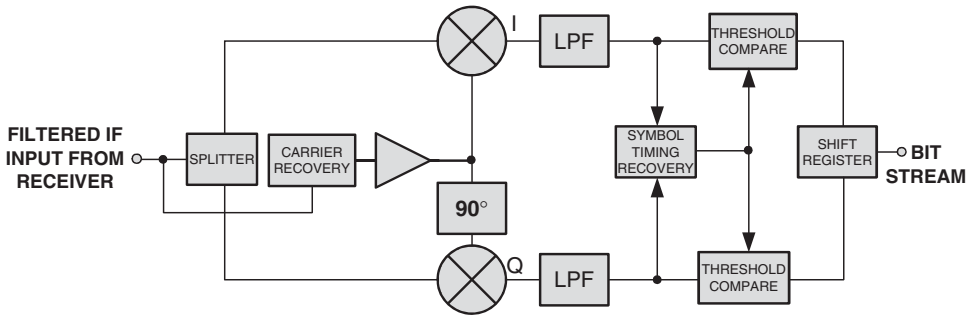


Figure 2.39 A QPSK demodulator.

through the receiver's conversion stages (if conversion stages are present). The output of these mixers are fed into the lowpass filters, which eliminate the now undesired IF signals. Some of this output from the LPFs is tapped and then placed into the *symbol timing recovery* and the *threshold comparison loop* to judge whether a 1 or a 0 is present. This also reshapes the digital data into a recognizable bit stream. If the SNR is high enough to assure a low BER, the bits from both mixers are then combined in the shift register as an almost exact replica of the originally transmitted binary signal.

However, the LPFs in the modulator and demodulator sections just discussed are not just any breed of filter. The lowpass filters must be of a very special type that will limit excessive intersymbol interference, since the demodulator would have great difficulty in deciding whether an input signal was a 1 or a 0 if high ISI were present. A *raised cosine filter* (a type of *Nyquist filter*) is commonly employed for this purpose. To decrease the required bandwidth needed to transmit the desired information, raised cosine filters slow the transitions of a digitally modulated signal from high to low and from low to high, and without degrading the ISI and the BER at the symbol decision times (as discussed above). These filters are usually matched, with one placed between the incoming data and the DAC in the transmitter, and the other half placed in the demodulator of the receiver. This replicates the response of a full Nyquist filter.

To compute the required bandwidth needed for a wide cosine filtered symbol rate, the formula is BW = symbol rate $\times (1 + \alpha)$, with an α (alpha) equal to a value between zero and one. It would be very bandwidth efficient if the BW could be exactly equal to the symbol rate, which is the same as an alpha of zero for a raised cosine filter. However, this is not quite practical. Anything over this value of zero for alpha is referred to as the *excess bandwidth factor*, because it is this bandwidth that is necessary beyond the symbol rate = BW value. We will always require an excess bandwidth greater than the symbol rate, or an alpha at some value that is over zero. Thus, if the alpha equaled one, the bandwidth necessary to transmit a signal would be twice the symbol rate. A contemporary digitally modulated radio, though, will usually filter the baseband signal to a value between an alpha of 0.2 and 0.5, with the corresponding decrease in bandwidth and the increase in the required output power headroom as compared to an alpha of 1. Figure 2.40 demonstrates the effect on the digital input signal's rise and fall times, the channel's bandwidth, and the received constellations, as the alpha is varied.

Due to the increased cost and complexity of building sustainably accurate filters, with high clock precision, in mass production environments, alphas lower than 0.2 are very uncommon. Any attempts at using very low alphas will also increase ISI to unacceptable levels, along with the added expense of producing amplifiers that must

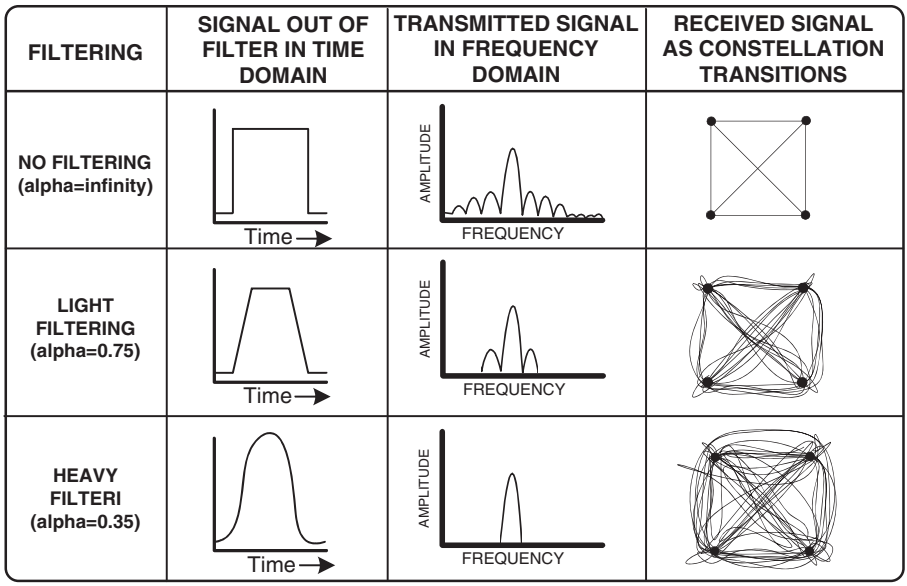


Figure 2.40 Baseband filtering effects on a digital QPSK signal.

be capable of greater peak output powers. Power back-off is required of these amplifiers due to the elevated power over-shoots (see Fig. 2.40) created by the increase in Nyquist filtering that limits the transmitted bandwidth of the digital signals. For heavily filtered QPSK, the excess peak power requires the solid-state power amplifier (SSPA) to have a *P*1 dB that is at least 5 dB over what would normally be required for an unfiltered signal to permit the power overshoots of the signal enough headroom so as not to place the SSPA into limiting, creating spectral splatter into adjacent channels. All signals that have a modulation envelope, whether used to carry information or not, will be affected by this Nyquist filtering, such as QPSK, DQPSK, and QAM modulations.

Gaussian filters are another method of slowing the transitions of the signal in order to decrease occupied bandwidth in the modulation scheme GMSK. Unlike raised cosine filters, there is a certain amount of unavoidable ISI created by this type of filtering. However, there are no power overshoots, which allows the use of more efficient amplifiers, with less power back-off required, than the raised-cosine filters discussed above.

It is important to note the difference between the filtering that takes place in each modulator/demodulator leg, as opposed to the rest of the analog radio sections. Since the I and Q signals at the input to the quadrature modulator are filtered separately in each I and Q input leg, then each of the I and Q legs of the modulator will lowpass filter at BW = symbol rate × (0.5 + α). But modulating the I/Q stream onto the IF, creating a double-sideband signal, will cause BW = symbol rate × (1 + α). So the actual shaping of the digital QPSK (or QAM) signal will take place in these modulator sections, comprised usually of a modem, with the rest of the radio design merely used to maintain this modulator-generated spectral shape while adding as little distortion and noise as possible to the already predefined modem signal. Thus, the actual analog filtration that occurs within the IF and RF sections of the analog transmitter and receiver units will be wider than BW = symbol rate × (1 + α), especially in block-up and -down converter designs.

2.6 Designing with Modulator/Demodulator ICs

2.6.1 Introduction

Quadrature (I/Q) modulators and demodulators are the most popular method today to perform modulation and demodulation of digital, as well as analog signals. Since it became possible to integrate them onto a single, low-cost chip, quadrature modulators have become hugely popular. These devices solve the complicated problem of imparting complex amplitude/phase information onto the RF or IF.

Any part of a signal's parameters can be modified by the quadrature modulator, whether it be phase, frequency, and/or amplitude, thus adding information to an unmodulated CW carrier. Simply employing a single mixer for this role would be unacceptable, since only one parameter (such as phase, for a BPSK signal) could be modified at a time, making an efficient digital modulation scheme infeasible.

Figure 2.41 shows a quadrature modulator for digital signals that is capable of varying two of the three modulation parameters. Typically, phase and/or amplitude are chosen, such as when generating BPSK, QPSK, or QAM. Many quadrature modulators are also proficient at generating AM, FM, CDMA, and SSB. The I/Q modulator shown will accept data at its I/Q inputs, modulate it, and then upconvert the baseband to hundreds of megahertzs. There are some specialized I/Q modulators that are actually capable of functioning into the gigahertz range.

Many quadrature modulators will also be fed by *digital-to-analog converters* (*DACs*) into their I/Q inputs (Fig. 2.42). The digital data is placed at the input to the DAC, which outputs in-phase (I) and quadrature-phase (Q) baseband signals into the I/Q modulator inputs. The I modulating signal enters the I input, where it is mixed with the LO, which converts it to RF or IF. The Q modulating signal enters the Q input, where it is mixed with the 90° phase shifted LO signal, which converts it to RF or IF. Both of these signals are then linearly added in the *combiner*, with each mixer outputting a two-phase state BPSK which, depending on the bits entering the modulator, will be in any one of four phase states. This combining of the two BPSK

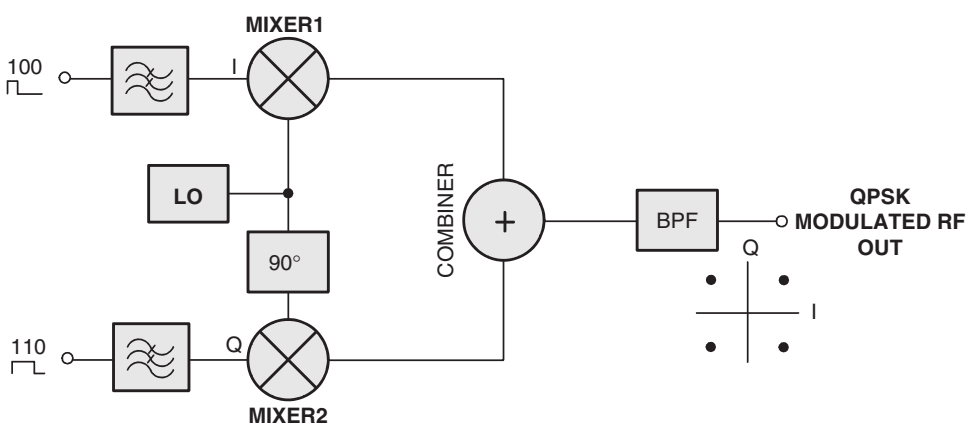


Figure 2.41 Simplified internal structure of a QPSK quadrature modulator.

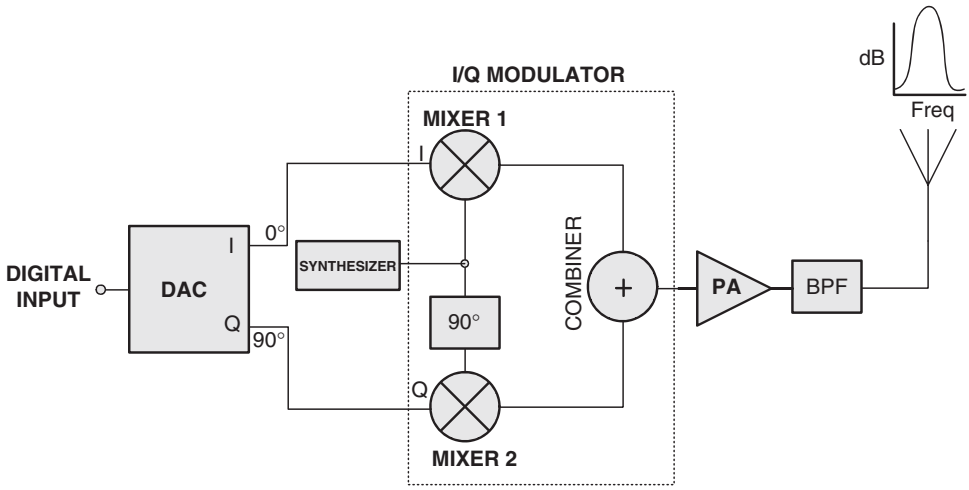


Figure 2.42 Simplified structure of a quadrature modulator using a DAC.

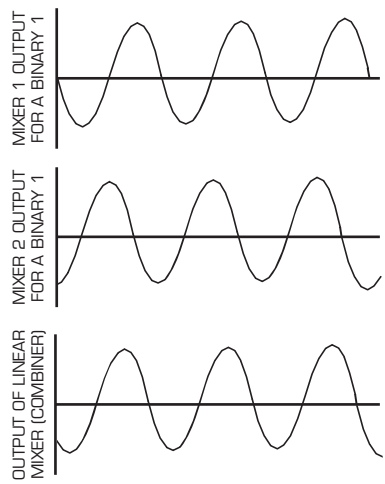


Figure 2.43 A signal's phase states through a quadrature modulator in the time domain.

signals produces QPSK, which is shown in the time domain in Fig. 2.43. Since each mixer's output is 90° phase shifted from the other, the algebraic summing of the combiner creates a single phase state out of the four possible. In other words, the incoming baseband signals to be modulated are mixed with orthogonal carriers (i.e., at 90°), and consequently will not interfere with each other. When the I and Q signals are summed in the combiner, they become a complex signal, with both signals independent and distinct from each other.

Now, the quadrature *demodulator* will take this incoming RF or IF signal, demodulate it, and then downconvert the signal to I/Q outputs into baseband for further processing

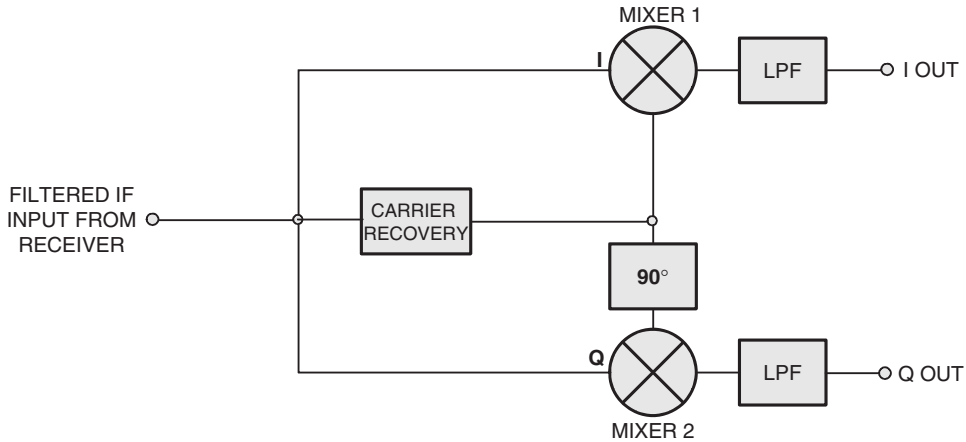


Figure 2.44 An I/Q demodulator for QPSK.

by digital logic circuits. An I/Q demodulator (Fig. 2.44) simply performs the reverse operation as the I/Q modulator above. It accepts the amplified and filtered RF or IF modulated signal, in this case QPSK, from the receiver's front end or IF section. The demodulator then recovers the signal's carrier in order to maintain the original phase information from the transmitter, and then splits it and inserted it in-phase into MIXER 1, and out-of-phase into MIXER 2. A baseband signal in I/Q format is then output at I_{OUT} and Q_{OUT} of the demodulator for processing.

Virtually, all high-speed digital communication systems will employ a predesigned modem, or MODulator-DEModulator (MODEM), for the task of modulating and demodulating the digital signal stream through the wireless system. Still, a methodology for the design of a modulator/demodulator is included in this section to assist in the construction of lower speed systems that may not have or need separate modems.

Another common digital demodulation technique is referred to as *sampled IF* (Fig. 2.45), and involves using a very high speed, wideband *analog-digital converter* (*ADC*) to seize the entire IF analog signal directly, rather than at baseband as the other prior digital radio designs do, and changes it into a sampled digital representation of

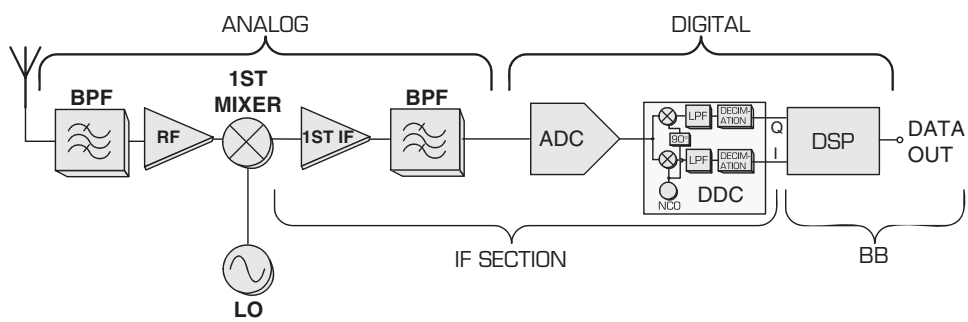


Figure 2.45 Receiver using IF sampling for I/Q demodulation into a DSP.

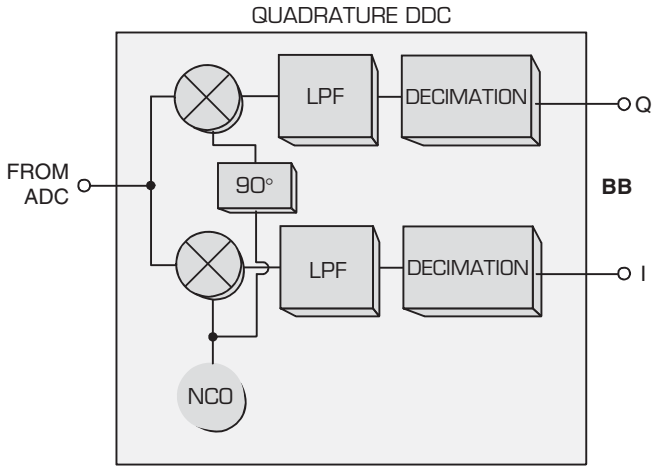


Figure 2.46 A quadrature DDC for complex modulations.

the IF. This is then fed into a *digital down converter* (*DDC*, Fig. 2.46), which digitally selects and mixes the sampled IF down to separate low frequency I and Q baseband components, filters them to narrow the bandwidth to accommodate only our desired baseband signal, and removes the sum frequency and any generated aliases. The circuit then performs *decimation* by discarding every *n* samples out of every 10 in order to reduce the huge ADC sampling rate. This decimation function has the positive result of reducing the final data rate. Thus, after the ADC's IF signal passes through the DDC and has undergone down conversion, I/Q translation, filtering, and data reduction of the desired and selected baseband portion of the IF, the now much lower baseband data rate output can be sent into a low cost DSP chip which, with the appropriate programming, can further process and demodulate the receiver's baseband signal, and output digital data to other sections of the receiver.

The transmitter section of this sampled IF system is shown in Fig. 2.47, and is essentially the reverse of the receiver, but uses a *digital up converter* (*DUC*).

The digital IF sampling technique is superior to that of the older analog baseband sampled architectures in that it avoids the analog circuit's I/Q mismatch and orthogonality impairments, as well as any DC offset and certain noise issues. However,

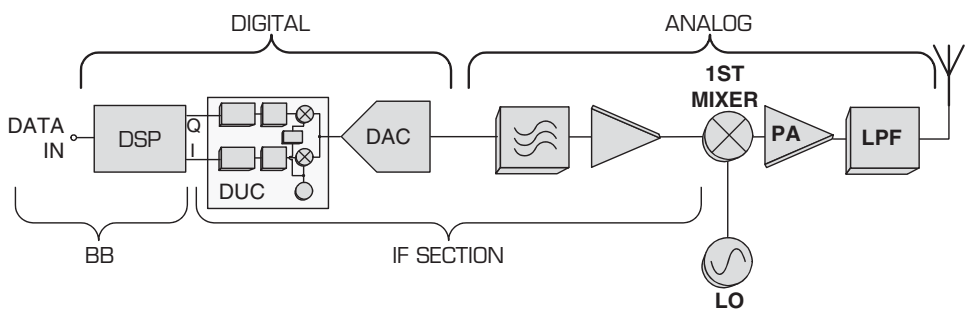


Figure 2.47 Transmitter using IF sampling for I/Q modulation.

digital IF sampling does require a very high-quality, high-speed, wide-bandwidth ADC to operate at the higher frequency and bandwidth levels of the IF chain, rather than at the much lower frequency baseband stages.

2.6.2 Designing with the RFMD RF2713

A popular chip that can perform both modulation and demodulation is the RFMD RF2713. Depending on a few component changes, it can be adopted as a modulator (Fig. 2.48), or as a demodulator (Fig. 2.49). The RF2713 is a monolithic IC that can operate with an IF from 100 kHz to 250 MHz, at a baseband frequency of anywhere from DC to 50 MHz, and with a V_{CC} of 3 to 6 V.

In the *modulator* configuration, pins 1 and 3 are the single-ended I and Q inputs, (which can also be driven differentially), while pins 2 and 4 are at RF ground. Since pins 1 and 3 are at a high input impedance (1200 Ω), 51-Ω resistors are added for 50-Ω matching, while the capacitors, at less than 1-Ω X_C, supply DC blocking. Pin 5 is at RF ground through a 0.1-μF capacitor. In normal modulator operation, pins 8 and 9 are left floating, while pins 10, 11, and 12 are direct-to-groundplane connected. Pin 13 is a high impedance (500 Ω) voltage-driven input, so a 51-Ω resistor can be placed in shunt to match to a 50-Ω LO signal input, if desired.

A low power LO with medium voltage outputs (0.1 to 0.8 V_{PP}) is required, as is a LO that is at twice the desired frequency of the carrier due to the internal divided-by-two frequency divider for the chip's 90° splitter, or LO = 2 × IF. Pin 14 supplies DC power to the chip, and must be adequately bypassed all the way from the lowest baseband

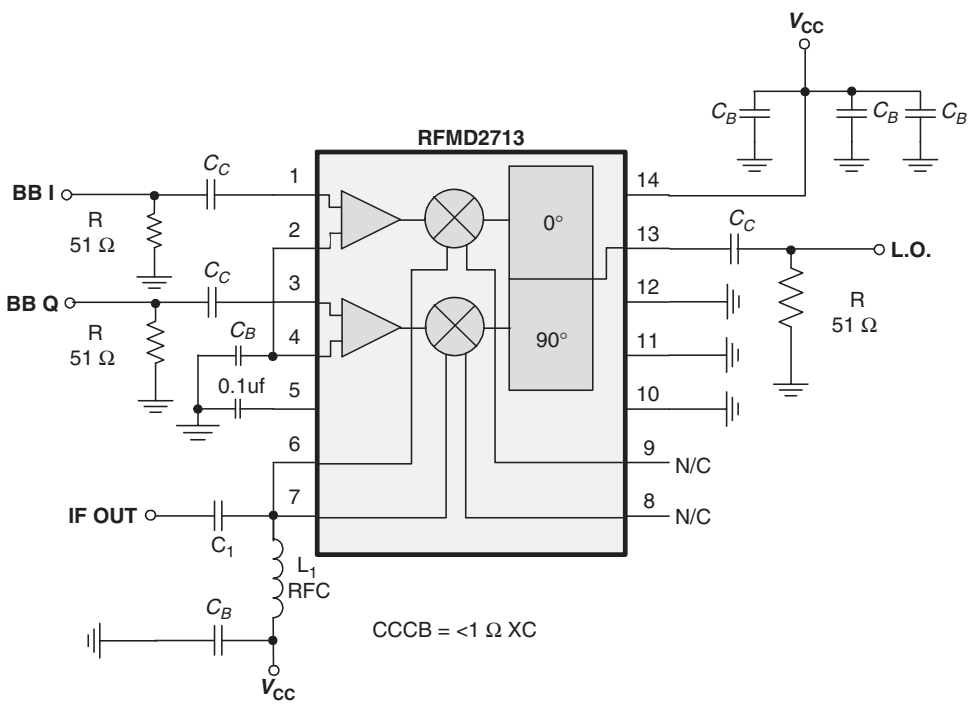


Figure 2.48 The RF2713 IC in its *modulator* configuration with support components.

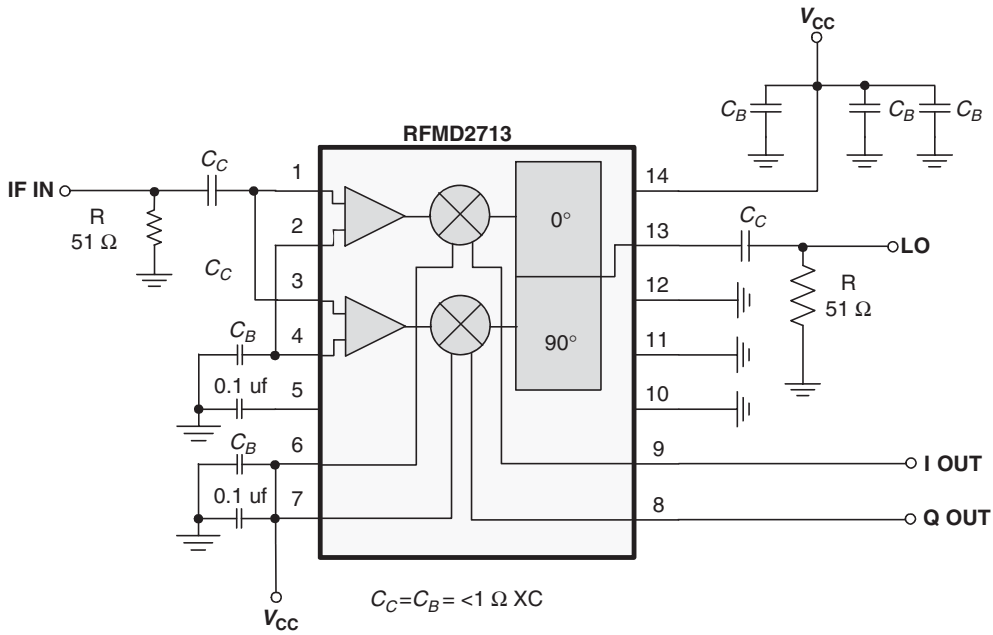


Figure 2.49 The RF2713 I.C. in its *demodulator* configuration with support components.

frequencies to the highest LO frequency. Pin 6 should be tied to pin 7, while V_{CC} is connected through an RFC to bias the internal open collector active mixers. However, since this signal output at pins 6 and 7 needs to see 1200 Ω to supply maximum P_{OUT} and gain, either a high impedance load or an LC-matching network that is tuned to the IF should be used to obtain a good impedance match into a lower impedance load.

When the RF2713 is in *demodulator* configuration pin 1, with pin 3 tied to it, has the IF injected into its high input impedance of 630 Ω, which is shunted by 51 Ω for matching. Pins 2 and 4 are connected together and placed at RF ground. Pin 5 is at RF ground through the 0.1 μF capacitor, while pins 6 and 7 are tied to VCC to bias the internal active mixers. Pins 8 and 9 are Q and I out, with 50-Ω outputs, but can only drive a high impedance load (>> 2000-Ω DC impedance), and are not internally DC blocked. Pins 10, 11, and 12 are sent directly to ground. Pin 13 is the high impedance voltage driven LO input, with 51 Ω shunted to ground for a 50-Ω oscillator, as desired. Pin 14 is V_{CC}, which must be adequately bypassed at all frequencies into, and out of, the demodulator.

CHAPTER 3
Amplifier Design

An amplifier is an active device that has the ability to amplify voltage, current, or both, at zero frequency (a DC amplifier), low frequencies (an audio amplifier), or at high frequencies (an RF amplifier). Considering that power is $P = VI$, then power amplification is only a normal outcome of this capability, since raising the current and/or the voltage will create power amplification.

AC amplifiers, whether for high or low frequencies, operate by allowing a small fluctuating external input signal to control a much greater DC output bias current. This small input signal changes the amplitude of the larger bias current, with the varying bias current then sent through a high-value output impedance or resistance component, which creates an AC output voltage due to $V = IR$. Depending on the amplifier's designated purpose, these output components may be composed of a resistor, an inductor, or a tuned circuit.

There are assorted circuit configurations to allow an amplifier to achieve different frequency responses, input and output impedances, gains, and phase shifts. Various bias circuits can be adopted to produce amplification at different efficiency and thermal stability levels, while special matching networks and coupling methods can be applied to match impedances and filter out undesired frequencies.

The most desirable specifications when designing an amplifier, such as a high P1dB, low noise, maximum efficiency, high gain, and good return loss, can frequently be in opposition with each other due to real-life internal transistor design limitations.

3.1 Amplifier Circuit Configurations

3.1.1 Introduction

Amplifiers come in three different basic flavors, each with its own distinct application and capability. They are referred to as *common-base*, *common-collector*, and *common-emitter* amplifiers, depending on whether the base, collector, or emitter is common to both the input and output of the amplifier circuit.

3.1.2 Common-Base Amplifier

With an input signal inserted at the emitter, and with the output taken from the collector circuit, we have our first configuration, the common-base amplifier (Fig. 3.1). The common-base (CB) can be found operating as a voltage amplifier for low input impedance circuits. It also possesses a high output impedance and a power amplification due to $P = V^2/R$, but current gain will always be a little less than unity. However, even though the CB amplifier has superior temperature stability and linearity, and can easily operate

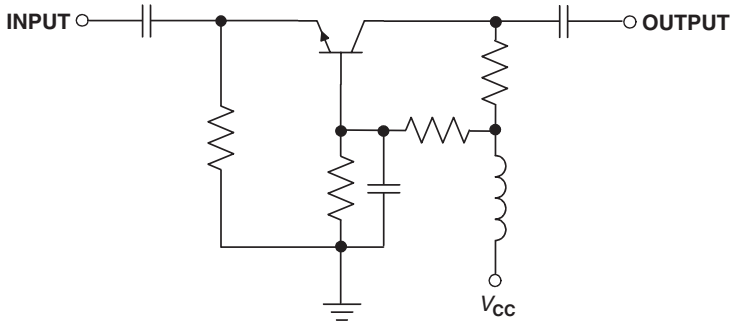


Figure 3.1 A common-base amplifier circuit.

at very high frequencies, it is not nearly as common as the next two configurations—the common-emitter and the common-collector amplifiers. This is due partly to the common-base's low input impedance (50 to 75 Ω). Nonetheless, CB amplifiers can occasionally be found at the 50-Ω antenna input of a receiver, or as Class C high-frequency amplifiers.

The JFET version of the BJT's common-base, a common-gate amplifier, can be seen in the IF of some receivers, with one such circuit shown in Fig. 3.2. C_2, C_3, R_2, and the radio frequency choke (RFC) are for decoupling; C_4 and C_6 are for RF coupling; C_5 can be tweaked to obtain a flatter frequency response throughout its passband; T_1 is for matching of its low input impedance, as required.

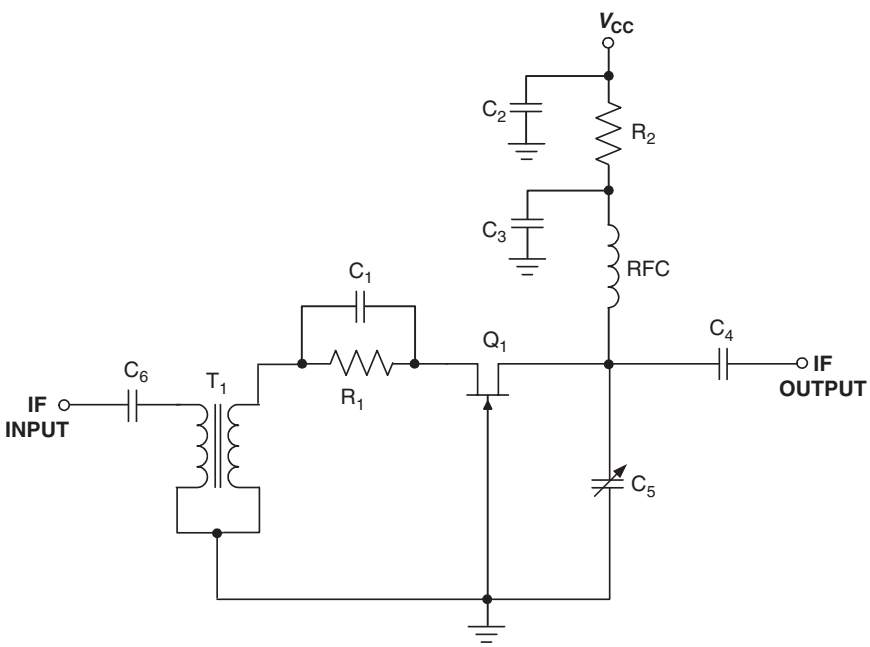


Figure 3.2 A common-gate JFET amplifier.

3.1.3 Common-Emitter Amplifier

The most popular amplifier circuit arrangement in all of electronics is the common-emitter (CE). A low frequency example is shown in Fig. 3.3. The bias circuit displayed in the figure is only one of the many ways to bias common-emitter amplifiers (see Sec. 3.6). A CE amplifier has a greater current and voltage gain combination than any other type. In fact, common-emitter amplifier configurations are capable of increasing not only voltage and current, but also make excellent power amplifiers.

A common-emitter amplifier functions when a signal is placed at the base of the transistor, an amplified output is extracted from the collector's output circuit. This output voltage will have been shifted by approximately 180° in phase when compared to the signal present at the amplifier's own input, due to the following action: As the signal at the transistor's base turns more positive, an increase in current will flow through the transistor. This decreases the transistor's resistance, and thus the voltage that is dropped across its collector-emitter junction, or from the collector to ground. Considering that the output signal will be taken from the voltage that is dropped across the transistor's collector—and the load resistor (R_C) will now be dropping the voltage that was formerly available to the collector—a shift in the phase at the amplifier's output is created that is precisely the reverse to that of the input signal.

At RF frequencies a large difficulty in CE amplifiers is an effect called *positive feedback*, which creates amplifier instability and oscillations due to the internal feedback capacitance between the transistor's collector and its base. This collector-to-base capacitance can be as high as 25 pF or more in certain types of bipolar transistors. At a specific frequency, this capacitance will send an in-phase signal back into the base input from the collector's output, which will create, for all intensive purposes, an oscillator. In other words, to give birth to the undesired CE oscillations, the internal capacitance and resistance of the transistor, along with other phase delays, must yield a powerful phase

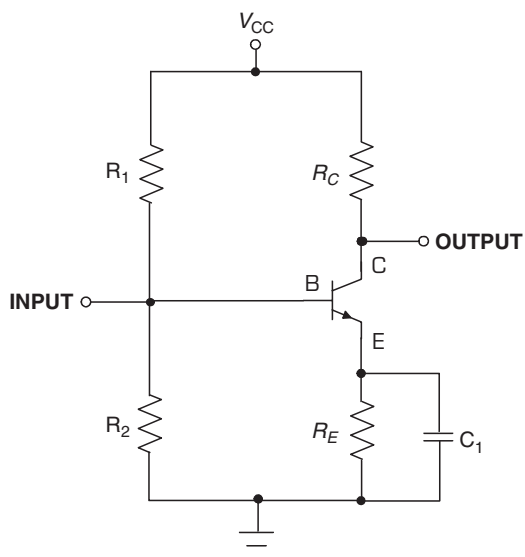


Figure 3.3 A low-frequency type of common-emitter amplifier.

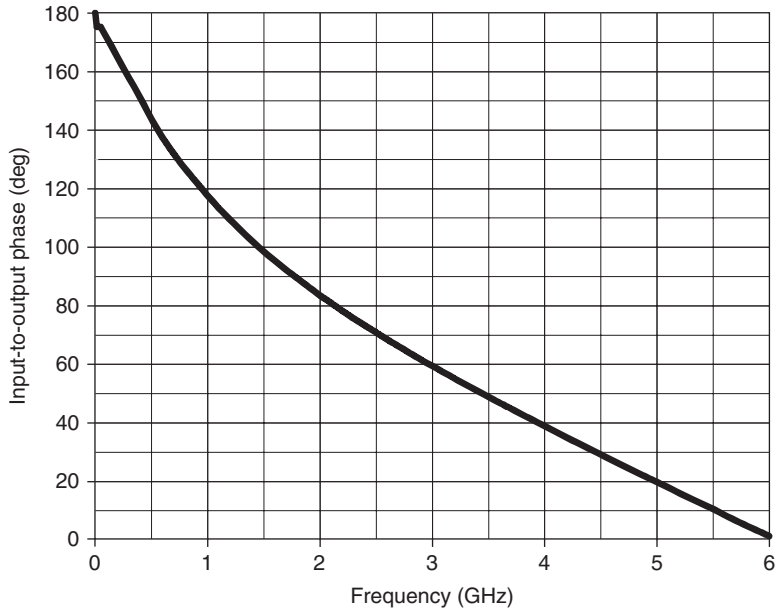


Figure 3.4 A phase-versus-frequency graph for a typical RF common-emitter unmatched amplifier circuit.

shift to the normally out-of-phase 180° feedback signal. But only those phase delays that are at a total of 360° (or 0°), furnishing the common-emitter with positive feedback to its base, will bring about amplifier instability and oscillations. Figure 3.4 illustrates a typical phase-versus-frequency response of a certain unmatched RF transistor in a common-emitter configuration (180° phase shift of a common-emitter amplifier will only have such a perfect 180° phase shift at 0 Hz (DC), and will rapidly decrease toward 0° phase shift as the amplifier's operational frequency is increased).

3.1.4 Common-Collector Amplifier

Figure 3.5 demonstrates the basic bias circuit configuration of a common-collector (CC) amplifier (also called an *emitter-follower*). The CC amplifier has the input signal inserted into its base, and the output signal removed from its emitter, which gives a current and power gain, but has a voltage gain of less than 1. This amplifier is used because of its high input impedance and low output impedance, making it beneficial as a buffer amplifier or as an active impedance matching circuit.

Unlike the common-emitter amplifier, there is no phase inversion between the CC's input and output, since as the input signal to the transistor's base rises in amplitude, the current through the active device will increase. This action forces a rise in the current through the emitter resistor, which increases the voltage drop across R_E, resulting in a 0° phase shift.

Most common-collector amplifiers do not possess any voltage robbing collector resistor, nor do they use a R_E bypass capacitor, which would also lower the output voltage at V_{OUT}.

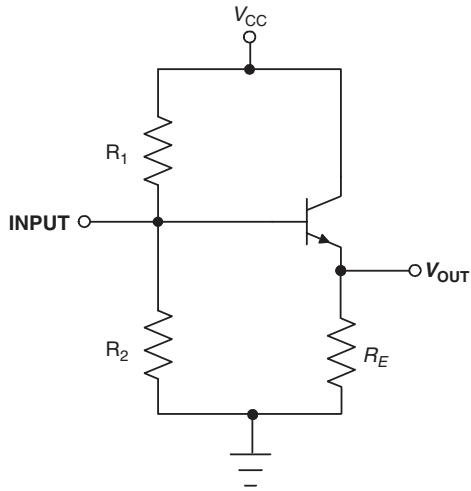


Figure 3.5 A typical common-collector amplifier circuit.

3.2 Amplifier Matching Basics

3.2.1 Introduction

When an amplifier's output impedance matches the load impedance, maximum power is transferred to the load (Fig. 3.6), and all reflections are eliminated. Thus, when an amplifier's output impedance does *not* equal its load impedance, such as the extreme

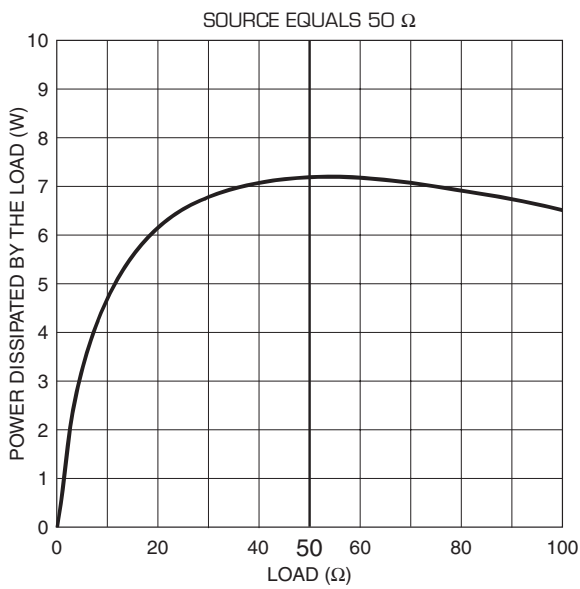


Figure 3.6 Maximum power theorem example, showing a 50-Ω source, with the load swept from 0 to 100-Ω.

example when the load (R_L) equals 0 Ω, then no voltage, and thus no power ($P = I \cdot V$, or $P = I \cdot 0$, or 0 W), can be developed across it. (However, the current drawn from the amplifier's power supply would be at a maximum possible value, or V_{CC}/R_S.) On the other hand, when R_L equals *infinite* ohms, then no current can possibly flow through such a load ($I = V/R$, or $I = V/\infty$, or 0 A), and thus no power can be dissipated by R_L ($P = I \cdot V$, or $P = 0 \cdot V$, or 0 W). However, the voltage drop across this R_L load resistor would be the entire value of V_{CC}. Therefore, between these two extremes there is a perfect, happy medium where R_S and R_L will create the highest possible P_{OUT}, and is shown below as P_{max}, which is exactly the point where R_S equals R_L. (The voltage drop across the matched load will only be *half* the total V_{CC}, since this entire circuit now forms a simple, evenly split voltage divider).

In spite of this perfect match, the point of maximum power transfer is not the same point as the maximum amplifier efficiency since, as just stated, in order to develop maximum power the Z_{OUT} of the amplifier must be the complex conjugate of the Z_{IN} of the load. But when Z_{IN} does equal Z_{OUT}, then exactly half of the power will be dissipated and wasted in the amplifier, and the other half used in the load, or an efficiency of 50%. This low efficiency level can be increased if the load has a higher input resistance, thus dropping more power across the load than across the output of the amplifier. Nonetheless, the total output power across the load will be less in this mismatched condition than if Z_{IN} actually equaled Z_{OUT}. In other words, the transfer of maximum power from a source to a load will not maximize efficiency, since maximum power transfer only occurs when the source impedance equals the load impedance. Maximum efficiency only occurs when the source impedance is zero and the load impedance is infinite. We can easily calculate an amplifier's efficiency in its DC case by:

$$\text{EFF}\ (\%) = R_{load}/(R_{load} + R_{source}) \times 100$$

To calculate the efficiency of a perfectly matched source to its load, we plug in the two 50-Ω values, or:

$$\text{EFF}\ (\%) = 50/(50 + 50) \times 100 = 50\%$$

As discussed above, and as the formula clearly shows, we obtained a mediocre efficiency of only 50% with a perfect conjugate match, yet this also supplies us with the maximum possible power transfer to the load. Thus, using Ohm's law for the 50-Ω R_{load} circuit in Fig. 3.7, we find that if 10 mA is flowing through the series source and load resistors ($I \cdot R$), then 0.5 V will be dropped across each resistor, and a power consumed in R_{load} of $P = I \cdot V$, or P = 10 mA × 0.5 V, or 5 mW, which is half the power of the complete circuit comprising R_{source} + R_{load}, or a total power dissipation of 10 mW. However, the figure also shows that using any other load resistance, whether higher or lower in value, will supply us with less output power.

Now, if we make R_{source} = 1 Ω and R_{load} = 1000 Ω, which is an extremely poor match for a good power transfer, we actually obtain this efficiency:

$$\text{EFF} = 1000/(1000 + 1) \times 100 = 99.9\%$$

This is, of course, an extremely high efficiency; almost perfect, with little power being wasted as heat in the driving source, yet the actual amount of power transferred to the load is $I_{R(load)}$ = 0.999 mA and $V_{R(load)}$ = 0.999 V, with the power consumed in R_{load} of P_{load} = 0.998 mW, or almost 100% of the total circuit power. But the total power of the entire circuit is only 0.999 mW, as opposed to the perfectly matched condition above,

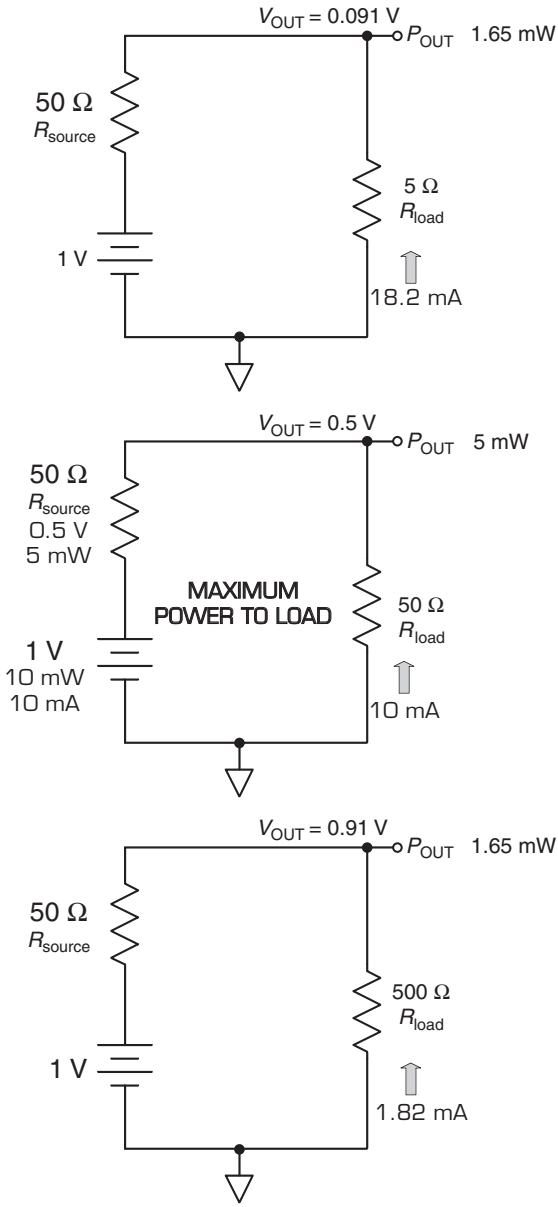


FIGURE 3.7 Three circuits with different load resistances, and the voltage, current, and power within each circuit.

which had a larger total available circuit power initially of 10 mW, and with a much greater resulting P_{load} of 5 mW, yet with an efficiency of only 50%.

Reversing this circuit and making R_{source} = 1000 Ω and R_{load} = 1 Ω will degrade all parameters. We will now not only have very poor efficiency (0.099%), but also very poor transfer of power to the load (0.998 μW).

Therefore, we can quickly see that efficiency increases as we increase the load's resistance, and decreases with a decrease in a load's resistance.

The above is used in audio amplifier design to maximize efficiency (and supply higher fidelity), as the driving amplifier (the source) is designed to have as low a Z_{OUT} as possible, while the input to the load will have a much higher Z_{IN}. This will guarantee that a significant amount of the audio signal is dropped across the load, and not across the source.

Nonetheless, the importance of a great impedance match from one RF amplifier stage to the next can readily be viewed by inspecting the formula below. Any impedance mismatches will end in a loss of power, referred to as *mismatch loss* (ML), and can be calculated by:

$$\mathrm{ML} = 10 \log_{10}\left[1 - \left(\frac{\mathrm{VSWR} - 1}{\mathrm{VSWR} + 1}\right)\right]^2$$

where ML = mismatch loss, dB, VSWR = voltage standing wave ratio in dimensionless units

3.2.2 Amplifier Matching Network Types

Matching allows the maximum power transfer and the attenuation of harmonics to be achieved between stages. There are numerous RF matching networks that can be employed to facilitate impedance matching and coupling between RF stages, as well as also supplying a certain amount of filtering for the signal. Using one of the various topologies of LC circuits within a matching network is far less expensive, and can reach far higher frequencies, than lumped transformer matching, which was so popular in the past.

L Network

One of the most common LC matching topologies, especially for narrowband impedance matching, is the simple L network, so named due to its L shape. With the appropriate topology, the L network can also furnish lowpass filtering to decrease harmonic output. Figure 3.8 shows a two-stage circuit without any matching network between the source and load, while Fig. 3.9 demonstrates the same two stages with a simple lowpass L network inserted, which is capable of matching a higher output impedance source to a lower input impedance load. Conversely, the lowpass L network of Fig. 3.10 matches a lower output impedance source to a higher input impedance load.

The loaded *Q*, which must be low for broadband circuits to obtain a wide bandwidth, is not selectable in L networks. But we can typically cascade multiple L networks to widen the bandwidth for use over a broadband, with a realistic number of cascaded L

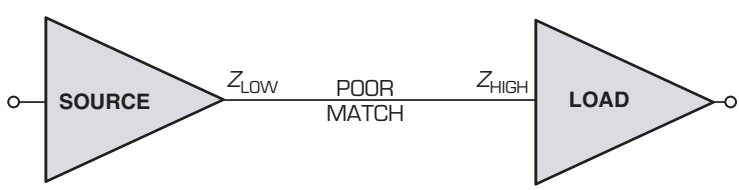


Figure 3.8 No impedance matching used between two amplifiers.

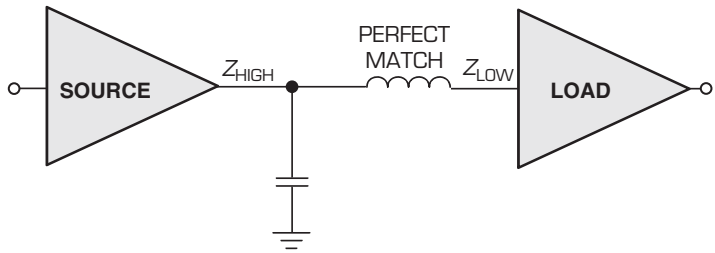


Figure 3.9 A high-to-low impedance matching L network between two amplifiers.

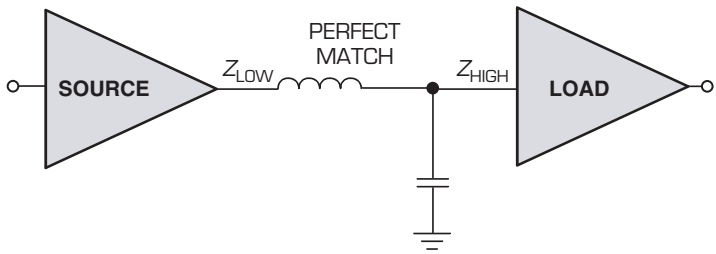


Figure 3.10 A low-to-high impedance matching L network between two amplifiers.

networks being a maximum of five. This practical limit is imposed because each new section we add will have less and less effect on improving the wideband match.

T Network

The T network (Fig. 3.11) is another popular impedance matching circuit, and can be designed to furnish almost any impedance matching level between two stages, and at a selectable loaded Q.

PI Network

The PI network is chosen for the same reason as the above T network, and can be found extensively in matching applications of all types. As shown in Fig. 3.12, we can alter the ratio between capacitors C_1 and C_2 so that the output impedance of the load can be matched to the source, as well as decreasing the harmonic output.

While the PI network's topology is as a lowpass filter, it can have a small resonant bandpass-like S_{21} gain peak, with an associated and excellent return loss (S_{11}), at a particular frequency. This resonant peak will only be present if the unloaded Q of the inductor and the loaded Q of the entire circuit is sufficiently high (Fig. 3.13).

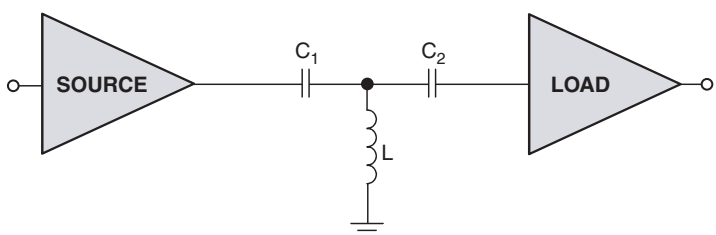


Figure 3.11 The impedance matching T network.

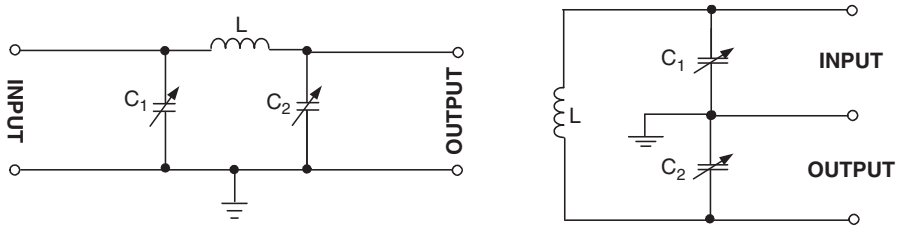


Figure 3.12 (*a*) The PI impedance matching network and its (*b*) equivalent circuit.

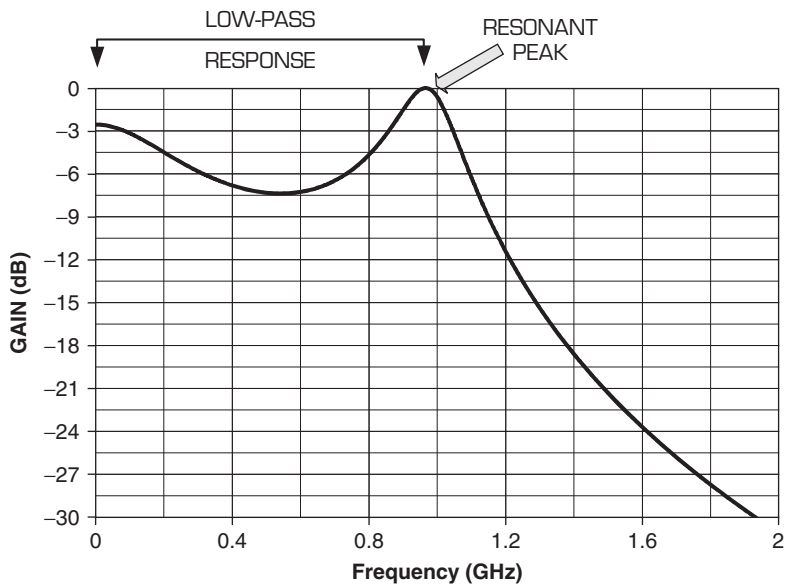


Figure 3.13 The frequency response of the PI matching network with resonant peak.

3.2.3 Amplifier Matching Network Issues

Two of the most important issues to keep in mind when designing microwave matching networks is the ever-present device parasitics that are a part of any active or passive component, and the effect that length has on RF circuit matching.

Designing RF matching networks in a spreadsheet program (or on paper) will supply results that will not take into account the vital component's parasitics, and will completely disregard printed-circuit board (PCB) trace and pad reactances, and the strong influence of circuit length. These issues will modify our expected RF circuit performance. However, at lower frequencies, much of this can be safely ignored, and the matching network design completed using the calculated ideal component values, without even considering PCB parasitic effects or the distance between each matching component and the source/load. At higher RF frequencies, where the distances between components and circuits can be a significant portion of a wavelength, this simplistic design methodology will adversely affect the expected performance of the matching network (or, indeed, of any other RF circuit), with the exact results dependent on where

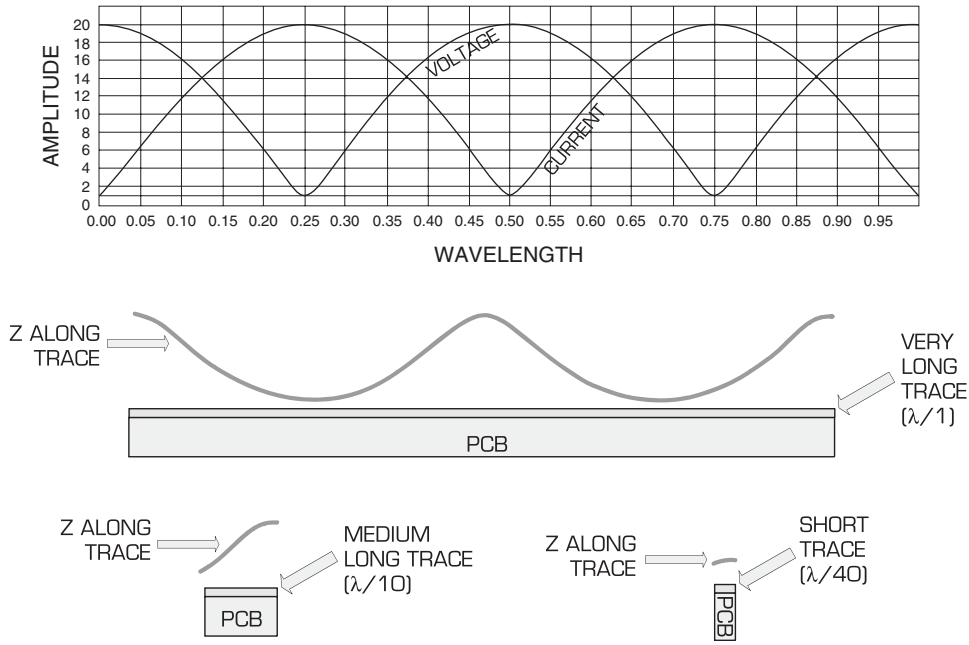


Figure 3.14 Voltage and current standing waves on a mismatched transmission line (one wavelength), and impedance variations along different lengths of microstrip with an unmatched load.

the circuit and individual components are located on the PCB. This is because on any mismatched line that is a significant portion of a wavelength, the impedance will vary along that line (Fig. 3.14). These variations on a mismatched line are due to the standing waves that are formed by the reflected RF waves bouncing off of the mismatch load, and then interacting with the forward wave, creating fixed peaks and valleys of voltage and current at every half wavelength along the trace. Thus, while the discrete matching component values are the exact same in Fig. 3.15 as they are in Fig. 3.16, and only the distance between each circuit element has been varied, this distance will completely destroy any predicted RF match.

We can use a rule of thumb, and say that when a trace or component reaches more than $\frac{1}{20}$ th of a wavelength in separation, then we must be extremely careful of all RF-related issues, or our design will not turn out as expected. Indeed, as can be seen in Fig. 3.14, even a minor distance of $\frac{1}{40}$ th of a wavelength along a transmission line will possess a slightly different impedance. This is why if we desire realistic, accurate simulation results, we must always have the close equivalent of the final PCB layout accurately duplicated within our final RF circuit simulation.

Even with accurate matching between two different circuits, certain practical problems can be created when the two matched stages are cascaded together. This can be the case, for example, when a 50-Ω filter is placed at the output of a 50-Ω amplifier, since only within each other's limited passbands will both stages actually "see" a pure 50-Ω match. Away from the center of their passbands, especially within the stopbands, these stages will start to become highly reflective, and are thus not at 50 Ω across a wide band. Further, even the passbands themselves will not be exactly 50 Ω, especially toward

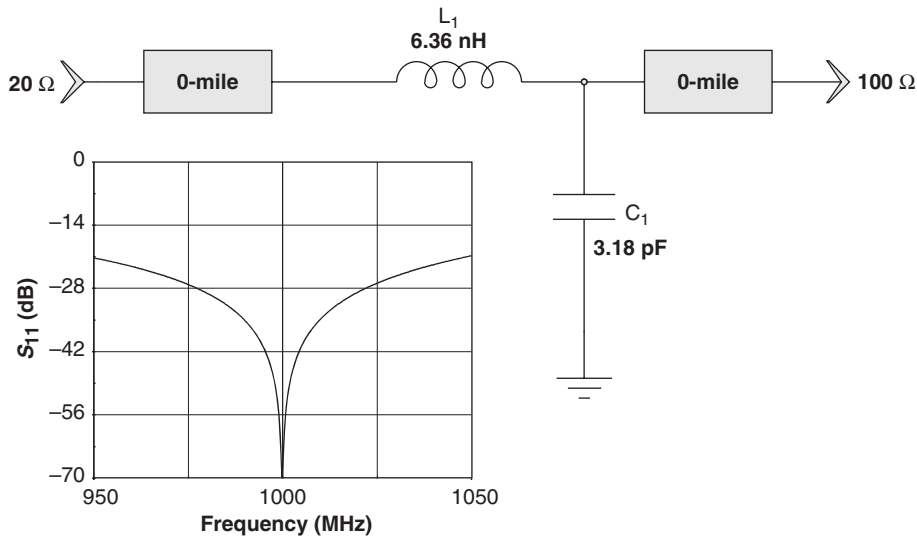


Figure 3.15 The match will be as calculated with zero microstrip length.

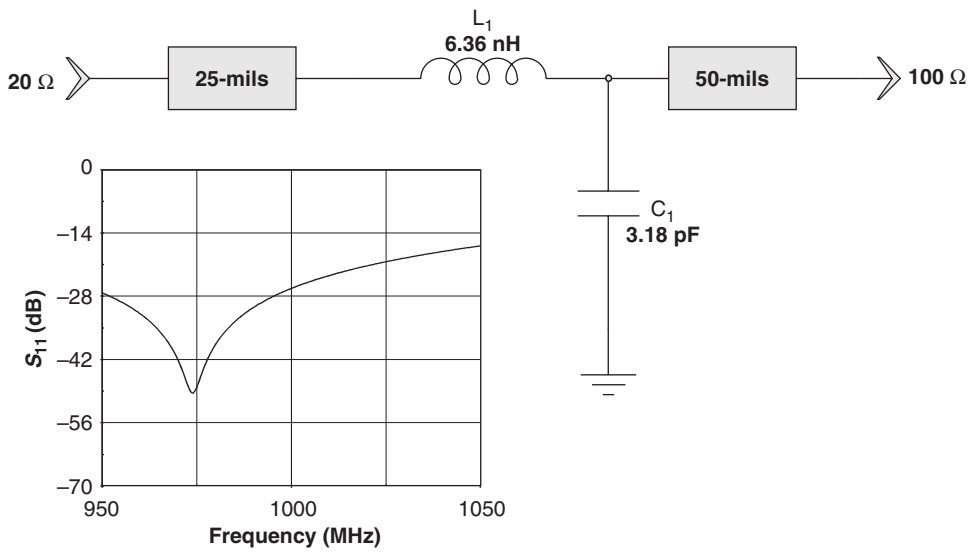


Figure 3.16 The calculated match will degrade with microstrip length.

the stage's band edges. This means that neither the filter nor the amplifier of these two stages will see 50 Ω at both of their ports over an infinitely wide bandwidth, as was assumed in the original filter or amplifier design. When separated by a certain length of 50-Ω microstrip, this can strongly influence a stage's expected response characteristic, as there will now be some undesired interaction between each stage, caused by the slight impedance mismatch in their passbands and the strong mismatch in their stopbands. This non-50-Ω nature of filters and amplifiers is created by their capacity to

frequency discriminate, which they accomplish by reflecting the unwanted frequencies, thus attenuating certain frequencies by a purposefully poor input/output impedance match, and through the easy passing of the desired frequencies through a good match. Depending on the length of the particular 50-Ω transmission line connecting these two stages, one stage's output impedance will be at a complex conjugate to the other output impedance, causing an almost perfect match at some specific frequency or frequencies. And as we lengthen or shorten this intervening 50-Ω microstrip, this matched condition will shift in frequency, which can cause excessive VSWR changes, gain ripple, instability, and/or loss of stopband performance at specific frequencies. So, unless we exactly replicate the specific length and width of all PCB microstrips between all stages during the RF simulation phase, we cannot tell what our true circuit performance will be like.

The above applies to any stage that will have frequency discrimination, such as: A matched (i.e., narrowband) amplifier connected to a filter network; a matched amplifier connected to another matched amplifier; an low-noise amplifier (LNA), in which we cannot even attempt to always conjugately match its input; or in a PA design, in which we do not conjugately match its output.

Even when we cascade two passive filters together for improved selectivity, it may be prudent to place a wideband 50-Ω matched amplifier, or a resistive attenuator pad, between the two filter stages. Both the wideband amplifier and the attenuator pad will supply strong isolation to each individual filter stage, as well as a wideband 50-Ω match.

3.3 Distortion and Noise Products

3.3.1 Introduction

Two unavoidable, but distinctly undesirable, elements of any electronic circuit are *distortion* and *noise*. Distortion can deform the carrier and its sidebands at the transmitter or receiver, causing spectral regrowth and adjacent channel interference, as well as a faulty, distorted replica of the original baseband signal, increasing the bit error rate (BER). Any added noise products, which can be contributed from almost any internal or external source, will also degrade the all important BER of the entire system.

3.3.2 Distortion

Distortion can form frequency intermodulation products by the internal nonlinear mixing of any signal with one or more other signals, or create harmonic distortion products when only one frequency is present. Distortion may have different causes, but comparable results: a modulated or unmodulated waveform that is altered in shape or amplitude from the original signal due to improper circuit response.

Frequency Distortion

Frequency distortion is produced when any passive or active circuit increases or decreases the amplitude of particular frequencies differently than the other frequencies. This is normally only a common problem in wideband IF or RF amplifiers if such stages are pushed to their extreme frequency limits. The frequency limitations of amplifiers have many causes, and mainly involve the active device's *transit time*, the negative effects of junction capacitance, and the reactive nature of the transistor's matching, filtering, coupling, and decoupling networks (which are all highly frequency-dependent, and will naturally act as bandpass, bandstop, highpass, or lowpass filters).

Amplitude Distortion

Amplitude distortion is a form of *nonlinear distortion*, and can be produced by the incorrect biasing of an amplifier, causing either saturation or cutoff of the transistor. This is extremely nonlinear behavior, generating harmonics and *intermodulation distortion products* (IMDs). Overdriving the input of an amplifier (*overload distortion*) will create this same effect, called *flat topping*, whether the bias is correct or not, causing both saturation and cutoff conditions. The harmonics and IMDs generated will produce interference to other services and/or to adjacent channels, as well as increase the system BER in a digital data radio, while a voice-band device will have an output signal with a harsh, coarse output.

Intermodulation Distortion

Intermodulation distortion (Fig. 3.17), quite similar to the *amplitude distortion* above, is produced when frequencies not harmonically related to the fundamental are created through nonlinearities in a linear Class A or a nonlinear Class C amplifier, or in a nonlinear mixer's output. IMD products can be formed by mixing together of the carrier with interferers, harmonics, IMD products from other stages, other channels, or sidebands, producing various spurious responses. And since these IMD products can fall in-band, or cause other signals to fall in-band, they can swamp the desired signal, creating severe interference. This also creates additional noise issues, which will degrade system performance and BER.

Also, when another neighboring transmitted signal (and/or its harmonics) arrives at a PA's stage, it too can mix together with the transmitter's carrier, causing IMDs to be created. This can be particularly problematic in dense urban environments, as there are many signals present that will modulate each other within the nonlinearities of a normal power amplifier, producing a multitude of sum-and-difference frequencies.

In these transmitter-to-transmitter cases, the IMDs can be attenuated by employing a wavetrap that is tuned to the interfering transmitter's frequency, and/or by shielding and proper grounding to prevent mixing within the other internal stages of the transmitter.

In a communications receiver this nearby-transmitter-driven IMD effect can be much worse, since the desired signal, along with a close-by transmitter's undesired signal, and/or its harmonics, may be allowed into the receiver's front end, creating the reception of unwanted signals and the obliteration of the desired frequency itself by these self-generated IMDs. This can be somewhat mitigated at the receiver by using an input notch filter, tighter bandpass filtering, amplifiers that are biased for maximum linearity, and confirming that the RF amplifiers are not functioning in a nonlinear region due to being overdriven by any input signal.

Because of its vital importance in the design of any amplifier, a more in-depth explanation of "intermod," is warranted. Intermodulation distortion is produced when two or more frequencies mix in any nonlinear device. It causes not only numerous sum and difference combinations of the original fundamental frequencies (*second-order products*; $f_1 + f_2$ and $f_1 - f_2$), but also intermodulation products of $mf_1 + n$ and $mf_2 f_1 - nf_2$, in which m and n are whole numbers. In fact, third-order intermodulation distortion products, which would be $2f_1 + f_2$, $2f_1 - f_2$, $2f_2 + f_1$, and $2f_2 - f_1$, can be the most damaging of the higher or lower IMDs. This is because the second-order IMD products would usually be too far from the receiver's or transmitter's bandpass to create many problems (see "Second-Order Intermodulation Distortion" below), and would be strongly

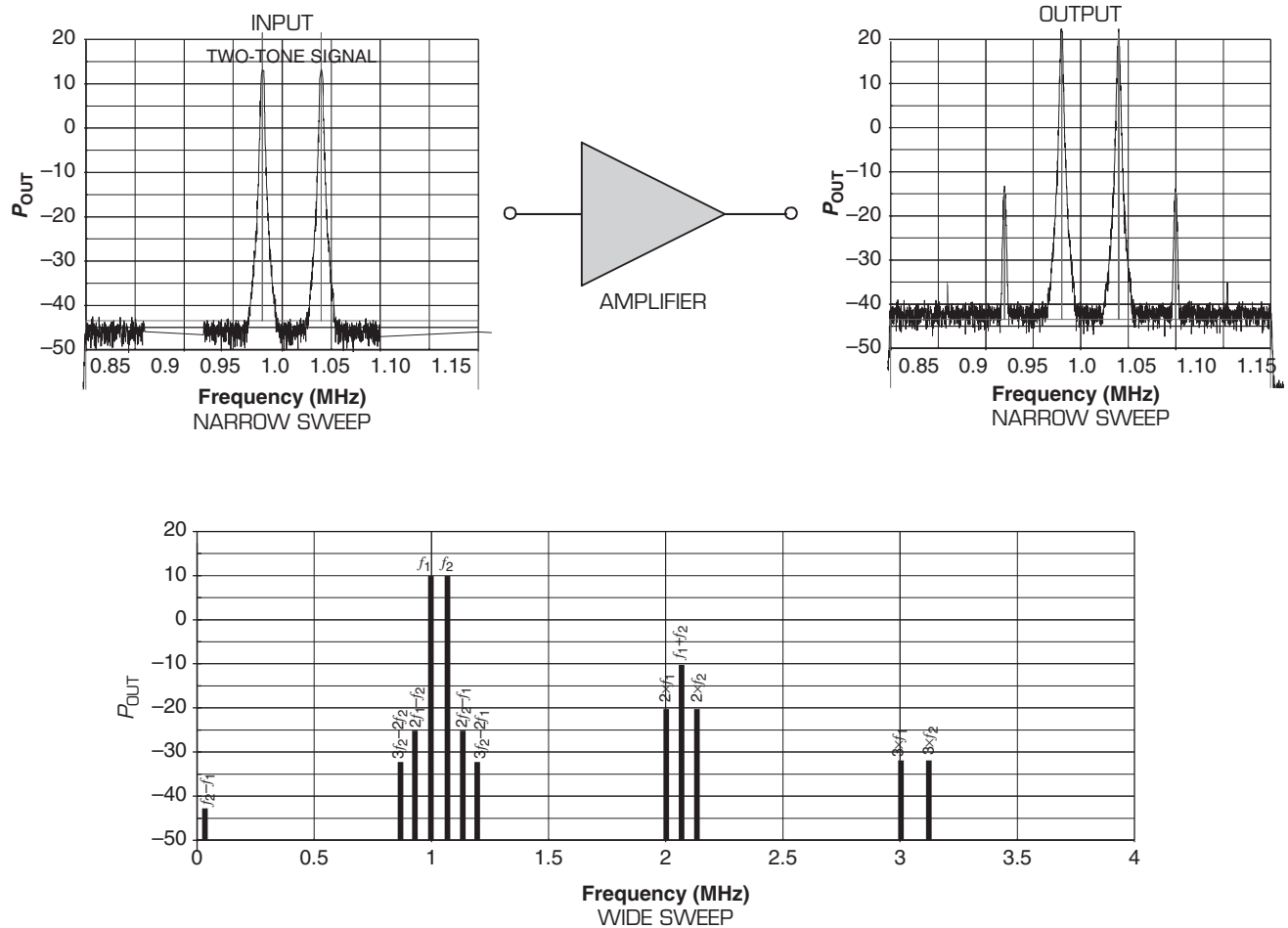


Figure 3.17 IMD generation through an overdriven or nonlinear amplifier, with wideband output frequency sweep.

attenuated by a narrowband amplifier's tuned circuits, the system's filters, and the selectivity of the antenna. As an example: Two desired input signals to a receiver, one at 10.7 MHz and the other at 10.9 MHz, would produce sum and difference second-order frequencies at both 21.6 and 0.2 MHz. These frequencies would be far from the actual passband of the receiver, and will be rejected by the receiver's selectivity. But the third-order IMDs formed from these same two signals would be at 10.5, 11.1, 32.3, and 32.5 MHz, with the most destructive frequencies being, of course, at 10.5 and 11.1 MHz. This is well within the passband of this particular receiver. Much higher order IMDs are created in receivers and amplifiers, so all IMDs up to the seventh order should be accounted for and, if they do fall within the band, must be at such a low amplitude that they cannot cause problems.

Third-Order Intercept Point

The third-order intercept point (TOIP or IP3) illustrates the entirety of third-order spurious products that will be created within the nonlinearities of a particular device—such as a linear amplifier, an active filter, or a mixer—when two different tones are injected into the input, causing multiple mixing frequency products at the stage's output that were not placed at its input.

The IP3 point (Fig. 3.18) is the area that undesired third-order output products will be at the same amplitude as the desired two-tone fundamental injected signals. However, this output IP3 point can never actually be reached, since the amplifier will go into saturation before this amplitude is ever truly attained. In fact, even though Fig. 3.18 does not show it, the third-order product's output power will gain-limit, just as the fundamental signal must, when the amplifier goes into saturation.

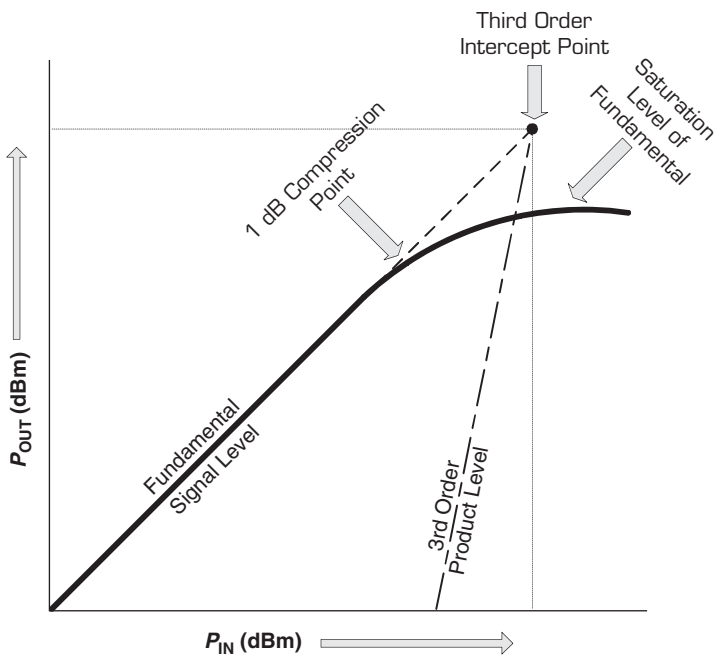


Figure 3.18 The third-order intercept and 1 dB compression points.

IP3 is the best indicator of how well a receiver will operate in the presence of powerful, close-in signals, and is a figure of merit that measures an electronic device's linearity. The linearity is tested by injecting two tones through the stage under test, with the tones thus being subjected to a certain amount of nonlinear distortion inherent within the DUT. This nonlinearity forces each tone to depart from a pure sine wave, producing harmonics and intermodulation products of various orders, permitting the IP3 to be measured (see *TOIP* (*IP3*) *Test* in Chap. 12). However, the value of the IP3 must be measured only when it is in its linear operating range; that is, when the DUT is not in compression.

Typically, an amplifier's third-order intercept point is located approximately 10 to 15 dB (OIP3 = P1dB + 10) above its P1dB compression point, but the output IP3 for a BJT amplifier can generally be easily approximated by:

$$\text{OIP3} = 10 \log(V_{CE} \cdot I_C \cdot 5)$$

where OIP3 = amplifier's output IP3, dBm
I_C = transistor's collector current, mA
V_{CE} = transistor's collector-to-emitter voltage, V

This clearly shows that the higher the bias level, the higher will be the IP3 of the amplifier. The I_C will normally be the easiest bias parameter to increase when we require a higher IP3.

1-dB Compression Point

The near maximum output power possible in an amplifier is the 1-dB compression point (P1dB), measured in dBm. This is the area where an amplifier begins to run out of room for its maximum output voltage swing, and thus becomes compressed. P1dB can be specified at the input or the output of a device.

Until this 1-dB compression point, which is reached when a high enough input signal is injected into the amplifier's input, an amplifier will have what is generally considered as a linear power output. At the amplifier's P1dB point the gain of the amplifier stage will depart from its linear value, causing the gain to actually decrease by 1 dB when compared to its small signal, low amplitude, value. So, for every decibel placed at the amplifier's input, no longer will there be a linear amplification of the signal; the output gain slope flattens, and soon no significant increase in output power is possible. Thus, a higher 1-dB compression point specification for a particular amplifier means it can output higher output power: If a particular amplifier has a P1dB of +30 dBm, then it is capable of a maximum output power of roughly +30 dBm (the maximum output power of an amplifier occurs a few decibels above this P1dB point, and is called the *saturated output power*).

Further, the lower the amplifier's RF power output is below its P1dB point, the lower the signal distortion will be. So for every 1-dB decrease in the fundamental input power, the second and third orders decrease in power by 2 and 3 dB, respectively. And the reverse is also true: for every decibel *increase* in fundamental power into the amplifier, the output second-order products will increase by 2 dB, while the output third-order products will increase by 3 dB. However, by increasing the desired input signals, there will be some point reached where the third-order products must (theoretically) be equal to the fundamental outputs. This, as discussed, is the *third-order intercept point*.

Second-Order Intermodulation Distortion

Second-order intermodulation distortion (IM2) has become an issue with the relatively new proliferation of wideband communications systems, especially direct conversion and low IF receivers, and includes the generation of spurious frequencies at $f_1 + f_2$, $2 \cdot f_1$, $2 \cdot f_2$, or $f_1 - f_2$ when two signals, f_1 and f_2, are present within the receiver. These spurious frequencies can be created by either the intermodulation distortion between an interferer (f_2) and the desired signal (f_1), or by two interferers (f_1 and f_2), falling in-band. For instance, if we had two signals coming into a wideband receiver at 400 MHz (f_1) and 401 MHz (f_2), then one of the IM2 products generated by the combination of these two signals would be located at 801 MHz ($f_1 + f_2$). Due to their extreme distance from the typical receiver's bandpass, these second-order signals would be filtered out of any narrowband receiver. However, in a modern wideband radio that possesses a very wide RF preselector (or no preselector at all) one of these f_1, f_2 frequency combinations may indeed create interference that falls in the receiver's IF bandpass. Such IM2 interference can be virtually eliminated if the receiver's RF front end can be sub-octave bandpass filtered.

These IM2 frequencies increase in amplitude versus the fundamental by a ratio of 2:1. Therefore, if we increase the desired signal level by 10 dB, then the IM2 products will increase by 20 dB.

The measurement of the second-order intercept point itself is considered to be the point where the two original signals and the spuriously generated second-order products would theoretically meet on the device's P_{IN} *versus* P_{OUT} curves (see "Third-Order Intermodulation" above).

Harmonic Distortion

Harmonic distortion occurs when an RF fundamental sine wave (f_r) is distorted due to nonlinearities within a circuit, generating undesired harmonically related frequencies ($2 \cdot f_r$, $3 \cdot f_r$, and so on). Interference to receivers tuned to many megahertz, or even gigahertz, away from the transmitter's output frequency is possible when these harmonics are broadcast into space (Fig. 3.19). The dominant cause of transmitted harmonics is overdriving a poorly filtered power amplifier, with an extreme case of distortion resulting in the sine wave carrier actually changing into a rough square wave. These nonperfect square waves contain not only the fundamental frequency, but numerous odd harmonics, as well as a certain amount of even harmonics.

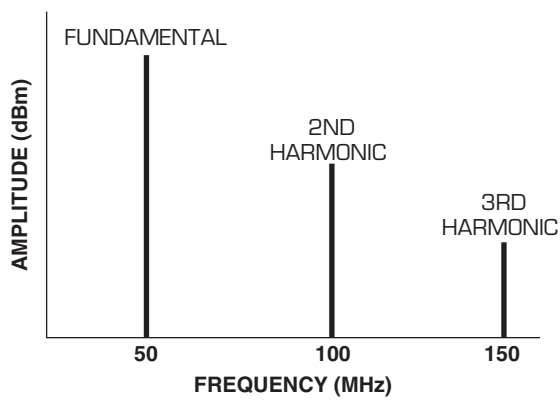


Figure 3.19 Harmonics in the frequency domain.

No active stage can be completely linear, with an inevitable number of harmonics being produced within all amplifiers. In a transmitter, these harmonics must be attenuated below the legal or system limits.

3.3.3 Noise

There are two principal classifications of noise: *circuit generated* and *externally generated*. Both limit the possible sensitivity and gain of a receiver, and are unavoidable, but can be minimized.

Circuit noise creates a randomly changing and wide frequency ranging voltage. There are two main causes: *white noise*, created by a component's electrons randomly moving around due to thermal energy, and *shot noise*, caused by electrons randomly moving across a semiconductor junction and into the collector or drain of a transistor.

External noise, produced by atmospheric upheavals like lightning, as well as *space noise* caused by sunspots and solar flares, and *cosmic noise* created by interfering signals from the stars, is exacerbated by man-made electromagnetic noise sources such as dimmer switches, neon lights, car ignitions, and electric motors.

3.4 Small-Signal Amplifier Design

3.4.1 Introduction

Small-signal amplifiers, which are always biased in their linear region, are needed to increase the tiny signal levels found at the input of a receiver into usable levels for the detector, or into the proper levels required for a transmitter's final power amplifier.

A microwave receiver's first RF amplifier will be of the Class A small signal, high gain type. It must not produce excessive noise, since any noise generated within this first stage will be highly amplified by later stages, decreasing the signal-to-noise ratio (SNR).

There are four vital considerations in any discrete RF amplifier design: the choice of the active device, the input and output impedance matching network, the bias circuit, and the physical layout. Each of these will be discussed in detail.

3.4.2 Small-Signal Amplifier Design with *S*-Parameters

Scattering parameters (*S*-parameters) characterize any RF device's behavior at different frequencies and bias points. With the information that *S*-parameters supply, the designer can calculate a device's gain, return loss, stability, reverse isolation, and its input and output impedances. Indeed, by using a device's *S*-parameters we can rapidly design an RF transistor's vital matching circuits.

Impedance matching of active devices is essential not only because the typical transistor will not naturally have a 50-Ω resistive Z_{IN} and Z_{OUT}, but its reactances will also vary over frequency. This means that for maximum power transfer into the system's impedance, which is normally 50 Ω, a matching network must be used to match the active device, and sometimes over a very wide band of frequencies. Since utilizing reactive LC components is the dominant matching technique, the match will only be perfect over a very narrowband of frequencies. There are, nevertheless, techniques for impedance matching that work quite well over a wider band of frequencies, and these will be discussed. For an even more in-depth discussion on matching, consult the influential work, *RF Circuit Design*, by Christopher Bowick.

As mentioned above, an unmatched active device will usually not possess 50-Ω resistive port terminations. Further, these ports will also be inductive or capacitive. This combination of a transistor's Z_{IN} and Z_{OUT} having not only a resistive, but also a reactive, element is referred to as a *complex impedance*. Thus, the matching network's job is not only to match the active device to the system's resistive impedance, but also to cancel the innate reactive elements within the transistor so as to permit a perfect 50 + *j*0 match—or 50-Ω resistive, with no capacitive or inductive reactances. This is called *conjugate matching*, and supplies an ideal impedance match.

However, in order to decrease the gain of a linear-biased transistor at various frequencies (called *gain flattening*), or to purposely design an amplifier with less gain, as well as for optimal noise figure (NF), a perfect match may not be desired for certain specialized applications.

When designing RF matching networks, we can take an *S*-parameter two-port file representation of the transistor, and initially ignore any effects the added DC biasing network may have on the active device in the final, physical design. This assumption is valid if only small amounts of RF feedback are produced by high values of R_f (Fig. 3.20; the feedback resistor) in an amplifier's bias network. In this way the *S*-parameters will be quite satisfactory for synthesizing not only the matching networks, but also the software simulations of the active circuit's responses. However, if an amplifier utilizes a low-value resistor for R_f in order to employ heavy RF feedback, then the device's *S*-parameter file calculation for the matching networks may no longer be completely valid for the transistor. In other words, the amplifier circuit's attributes, as defined by the device's *S*-parameter file, are accurate only when the bias network employs normal, high resistance values within its bias network.

S-parameter files (or ∗.S2P; Fig. 3.21) contain only the RF parameters for a few frequencies (usually not more than 20). So, when reading *S*-parameters on data sheets, or in the ∗.S2P text files themselves, we may find that our frequency of interest falls between two published values. For accuracy at our desired frequency when designing

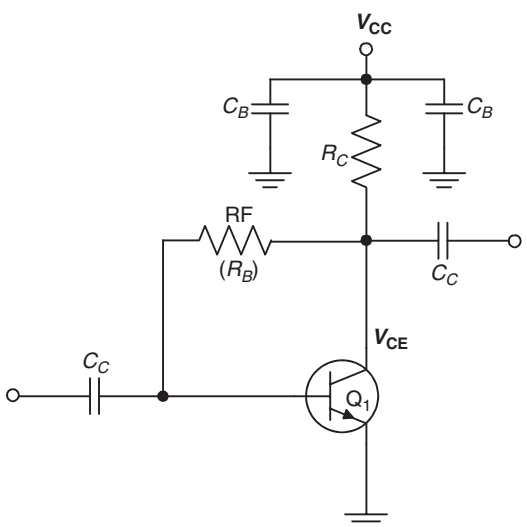


Figure 3.20 Collector feedback bias for a BJT.

```
! FILENAME (21bfg425.S2P)
! MODEL OF TRANSISTOR (BFG425 W)
! TRANSISTOR BIAS (V2=2.000 E+000 V,  I2=1.000 E–003 A)
# GHz  S  MA  R 50
!          S11          S21          S12          S22
!Freq(GHz) Mag   Ang    Mag   Ang    Mag   Ang    Mag   Ang
   0.040  0.950 –1.927  3.575 177.729 0.003  83.537 0.996 –1.116
   0.100  0.954 –5.309  3.518 175.247 0.007  87.057 0.996 –3.082
   0.200  0.951 –10.517 3.504 170.441 0.014  82.341 0.991 –6.343
   0.300  0.947 –15.891 3.496 166.534 0.020  78.681 0.988 –9.405
   0.400  0.941 –20.987 3.493 161.221 0.027  75.109 0.982 –12.576
   0.500  0.935 –26.297 3.476 156.531 0.033  71.254 0.974 –15.593
   0.600  0.928 –31.508 3.433 151.954 0.040  67.636 0.965 –18.605
   0.700  0.919 –36.669 3.384 147.515 0.046  63.875 0.954 –21.674
   0.800  0.910 –41.871 3.350 143.152 0.051  60.357 0.943 –24.600
   0.900  0.898 –46.948 3.317 138.801 0.057  56.929 0.930 –27.559
   1.000  0.886 –52.161 3.272 134.309 0.062  53.488 0.916 –30.396
   1.100  0.874 –57.181 3.223 130.114 0.067  50.181 0.903 –33.098
   1.200  0.861 –62.218 3.171 125.837 0.071  46.955 0.888 –35.859
   1.300  0.849 –67.154 3.119 121.786 0.075  43.791 0.873 –38.531
   1.400  0.835 –72.157 3.072 117.682 0.079  40.631 0.857 –41.151
  11.500  0.845 –2.938  0.375 –130.163 0.134 –104.397 0.607 14.337
  12.000  0.848 –9.981  0.326 –139.789 0.124 –115.184 0.658 4.326

!BFG425W_noise.xls                                  !Philips part #:BFG425W
! Bias condition: VCE = 2 V, IC = 1 mA               # MHz dB Ratio deg Ω

! Freq. Fmin Gmag Gangle Rn
900 0.7 0.67 19.1 0.40                              2000 1.3 0.56 57.5 0.36
```

Figure 3.21 ∗.S2P *S*-parameter file for set bias conditions and set frequencies. Text after "!" are comments for the user, and are ignored by simulation programs.

the matching network by hand calculation, we can take the *mean* value between the two closest frequencies within the file. As an example: *S*-parameters are given in a certain ∗.S2P file for 3 GHz and 4 GHz, but our design requires a center frequency of 3.5 GHz. To achieve an accurate design, take the mean value of each *S*-parameter at 3 GHz and 4 GHz. To compute S_{12} at 3.5 GHz:

$$\frac{S_{12}\text{MAG}(@3\text{ GHz}) + S_{12}\text{MAG}(@4\text{ GHz})}{2} = S_{12}\text{MAG}(@3.5\text{ GHz})$$

and:

$$\frac{S_{12}\theta(@3\text{ GHz}) + S_{12}\theta(@4\text{ GHz})}{2} = S_{12}\,\theta(@3.5\text{ GHz})$$

The analysis of a device's *S*-parameters will furnish the designer with a lot of data on the transistor, such as three different ranges of possible gains. These gains are: *MAG*, or the *maximum available gain* that the transistor can attain when perfectly matched (MAG is considered a figure of merit only); *transducer gain*, or the true gain of an amplifier stage, together with the effects of impedance matching and device gain, but not including power lost within the passive components; and *transducer unilateral gain*, or the measurement, in decibels, of an amplifier's power gain into an *unmatched* 50-Ω load (a worst case evaluation).

Another very meaningful piece of information that *S*-parameters are easily able to reveal is whether a particular active device will remain stable when presented with any impedance at its input and output port, or whether the transistor may begin to oscillate at some specific impedance combination. These stability calculations, using *S*-parameters, are referred to as the *Rollet stability factor*, or simply *K*. A transistor with a *K* of over *1* will be *unconditionally stable* at that particular frequency and DC bias point as chosen for the transistor at *any* input and output impedance it may be presented with. In other words, it will never begin to oscillate under (almost) any circumstance. However, if the value of *K* is under 1, then there will be some value of input and output impedance that will cause the amplifier to become *potentially unstable*. Thus, the amplifier may begin to oscillate. What value of impedances that will cause this instability will not be disclosed by the formula, and real-life feedback paths in the amplifier's final board layout can also cause unexpected stability problems with this simplistic, but very helpful, initial stability approach.

Linear Amplifier Design

In order to begin the design of any amplifier, we should first discover whether the active device we have chosen will remain stable at our frequency and bias of interest over all impedance variations. This stability should be maintained over a very wide region of frequencies, both low and high, for wide ranging stability.

Using the *K* stability formula:

$$K^* = \frac{1 + \left(|D_S|^2 - |S_{11}|^2 - |S_{22}|^2 \right)}{2 \cdot |S_{21}| \cdot |S_{12}|}$$

where

$$D_S{}^{\dagger} = S_{11}S_{22} - S_{12}S_{21}$$

Thus, if $K > 1$, then the active device will be unconditionally stable for all $Z_{IN'S}$ and $Z_{OUT'S}$ presented at its ports. This is by far the easiest transistor to design an amplifier with. But if $K < 1$, then the device is potentially unstable. If this is so, Z_{IN} and Z_{OUT} must be very cautiously selected; or pick a different active device with a $K > 1$; or opt for another transistor bias point that will give a $K > 1$; or use a neutralizing circuit; or place a low-value resistor at the amplifier's input or output (to decrease gain).

The following is an example of how to rapidly calculate whether our chosen transistor will be stable at 1.5 GHz, with a $V_{CE} = 10$ V and an $I_C = 6$ mA.

1. The *S*-parameters at that particular frequency and bias point are found to be (by looking at the transistor's *S*-parameter text file):

$$S_{11} = 0.195 \angle 167.6°$$

$$S_{22} = 0.508 \angle -32°$$

$$S_{12} = 0.139 \angle 61.2°$$

$$S_{21} = 2.5 \angle 62.4°$$

---

*Do *not* use full vector algebra; employ only the *S*-parameter's scalar magnitudes (e.g., $S_{11} = 0.35$). $|S_{xx}|$ means to ignore the *sign* of the magnitude, and always consider it positive.

†Use *full* vector algebra ($Z \angle \pm 0°$) in *S*-parameter calculations (e.g., $S_{11} = 0.35 \angle -45°$). How to multiply, subtract, divide, and add vectors is explained below.

2. First calculate D_S:

$$D_S{}^* = (0.195 \angle 167.6° \times 0.508 \angle -32°) - (0.139 \angle 61.2° \times 2.5 \angle 62.4°)$$

$$= 0.25 \angle -61.4°$$

3. Then calculate K:

$$K^\dagger = \frac{1 + |0.25|^2 - |0.195|^2 - |0.508|^2}{2|2.5| \cdot |0.139|} = 1.1$$

Since K is greater than 1, we see that we have a stable transistor at 1.5 GHz with the transistor's bias conditions as stated in the S-parameter file.

For the next calculation to be valid, K must be greater than 1, or unconditionally stable. This calculation is for MAG, or maximum available gain. Thus, if K is over 1 for a transistor, we can proceed to the MAG calculation to check if the transistor will give us the gain value we desire. MAG is never attained in practice, so only when the MAG is 20% or more above our required gain would we want to design with that particular transistor.

To calculate the MAG of the transistor, or the maximum gain that the transistor can attain when perfectly matched:

1. Calculate $B_1{}^\dagger = 1 + |S_{11}|^2 - |S_{22}|^2 - |D_S|^2$
2. B_1 judges if + or – will be adopted in the MAG equation in step 3 below. If B_1 returns a *negative* answer, use the + (positive) sign after K; if B_1 is *positive*, utilize the – (negative) sign after K.
3. $\text{MAG}^\dagger = 10\log\frac{|S_{21}|}{|S_{12}|} + 10\log\left(\left|K \pm \sqrt{K^2 - 1}\right|\right)$

As an example of a MAG calculation:

1. $B_1{}^\dagger = 1 + |0.195|^2 - |0.508|^2 + |0.25|^2 = +0.717$
2. Since B_1 has returned a positive number, the sign after K (1.1) in the equation below will be negative.
3. Complete for MAG:

$$\text{MAG}^\dagger = 10\log\frac{|2.5|}{|0.139|} + 10\log\left(\left|1.1 - \sqrt{1.1^2 - 1}\right|\right) = 12.56 + (-1.92) = 10.63 \text{ dB}$$

Thus, the amplifier will supply a maximum available gain of 10.63 dB.

After finding the transistor with a K greater than 1 at our desired frequency, and with a MAG greater than 20% of that required for our application, the actual Z_{IN} and Z_{OUT} of the transistor can then be calculated. These calculations will take into account the *reflected impedances* caused by S_{12}, which is the transistor's value of isolation in the

---

*Do *not* use full vector algebra; employ only the S-parameter's scalar magnitudes (e.g., $S_{11} = 0.35$). $|S_{xx}|$ means to ignore the *sign* of the magnitude, and always consider it positive.

†Use *full* vector algebra ($Z \angle \pm 0°$) in S-parameter calculations (e.g., $S_{11} = 0.35 \angle -45°$).

reverse direction (only if S_{12} had a value of zero would it have no effect on the transistor's Z_{IN} and Z_{OUT}. However, this is never the case). So, we will want to perform a *simultaneous conjugate match* to prevent the matching of the input port from changing the matching of the output port, and vice versa.

To begin the calculation of the transistor's input and output impedances:

1. Determine the value of C_2, which is used in one of the following equations, by:

$$C_2^{\dagger} = S_{22} - (D_S\, S_{11}^{\Psi})$$

where

$$D_S^{\dagger} = S_{11}\, S_{22} - S_{12}\, S_{21}$$

NOTE: S_{11}^{Ψ} *equals the complex conjugate of* S_{11}. *In other words, just change the sign of the angle, but not the* magnitude's *sign, of the* S_{11} *value (e.g.,* $S_{11} = +12 \angle +18°$ *, so* $S_{11}^{\Psi} = +12 \angle -18°$*).*

2. Calculate B_2, which is also used for a following calculation:

$$B_2{}^{*} = 1 + |S_{22}|^2 - |S_{11}|^2 - |D_S|^2$$

3. Then calculate the magnitude of the *load reflection coefficient* (Γ_L), which is the value of the impedance that the transistor must see at its output to be perfectly matched:

$$|\Gamma_L|^{*} = \frac{B_2 \pm \sqrt{|B_2|^2 - 4|C_2|^2}}{2|C_2|}$$

The sign used for $B_2 \pm$ is the opposite as that obtained in the B_2 calculation of step 2.

4. The angle is the same as that calculated in step 1 for the C_2 angle, but simply reverse this answer's sign.

Now, to calculate the *output impedance* (Z_{OUT}) of the transistor:

1. Follow the formula below [all signs (±) must be strictly maintained for all calculated numbers. $1 - \Gamma_L$ will subtract 1 from the *real* term of Γ_L, and will simply change the sign of the imaginary term, while $1 + \Gamma_L$ will add 1 to the real term of Γ_L and ignores the imaginary number completely]:

$$Z_{OUT}{}^{\dagger} = Z_{LOAD}\left(\frac{1 + (\Gamma_L^{\Psi})}{1 - (\Gamma_L^{\Psi})}\right)$$

---

*Do *not* use full vector algebra; employ only the *S*-parameter's scalar magnitudes (e.g., $S_{11} = 0.35$). $|S_{xx}|$ means to ignore the *sign* of the magnitude, and always consider it positive.

†Use *full* vector algebra ($Z \angle \pm 0°$) in *S*-parameter calculations (e.g., $S_{11} = 0.35 \angle -45°$).

NOTE: *Z_{LOAD} equals the transistor's load placed at its output (typically 50 Ω; written as 50 + j0).* $\Gamma_L{}^{\Psi}$ *equals the complex conjugate of* Γ_L. *In other words, just change the sign of the* angle, *but not the* magnitude's *sign, of the* Γ_L *value.*

To calculate the value of the transistor's *input impedance* (Z_{IN}) for the transistor's particular output impedance as calculated above:

1. $\Gamma_S{}^{\dagger} = \left[S_{11} + \dfrac{S_{12} \cdot S_{21} \cdot \Gamma_L}{1 - (\Gamma_L \cdot S_{22})} \right]^{\Psi}$

2. $Z_{IN}{}^{\dagger} = Z_{SOURCE} \left(\dfrac{1 + \Gamma_S{}^{\Psi}}{1 - \Gamma_S{}^{\Psi}} \right)$

NOTE: *Z_{SOURCE} equals the transistor's source (the prior stage) placed at its input. Typically 50 Ω; written as 50 + j0 for this formula.*

$\Gamma_L{}^{\Psi}$ *or* $\Gamma_S{}^{\Psi}$ *equals the complex conjugate of* Γ_L *or* Γ_S *respectively. If* Ψ *is outside a bracket, then the answer to everything within the bracket must be converted into this complex conjugate.*

Now that we have discovered the input and output impedances of our chosen active device, we can begin to impedance match its ports to obtain a simultaneous conjugate match for the transistor's required source and load. So the next step is to design the matching networks for our circuit.

As an example, the active device, a transistor, has these *S*-parameters at 1.5 GHz with a $V_{CE} = 10$ V and an $I_C = 6$ mA:

$$S_{11} = 0.195 \angle 167.6°$$

$$S_{22} = 0.508 \angle -32°$$

$$S_{12} = 0.139 \angle 61.2°$$

$$S_{21} = 2.5 \angle 62.4°$$

We will want the amplifier to run between two 50-Ω terminations. To design an input and output matching network to maintain maximum amplifier gain:

1. Calculate for *K*. Confirm unconditional stability ($K > 1$) by calculation or look-up table, if supplied. (IMPORTANT: All negative real number results *must* be used in all calculations *as* negative real numbers):

---

†Use *full* vector algebra ($Z \angle \pm 0°$) in *S*-parameter calculations (e.g., $S_{11} = 0.35 \angle -45°$).
All signs (±) must be strictly maintained for all calculated numbers. $1 - \Gamma_L$ will subtract 1 from the *real* term of Γ_L, and will simply change the sign of the imaginary term.

a. $D_S{}^{*} = S_{11}\,S_{22} - S_{12}\,S_{21} = 0.25\angle{-61.4°}$

b. $K^{\dagger} = \dfrac{1+|D_S|^2-|S_{11}|^2-|S_{22}|^2}{2\,|S_{21}|\cdot|S_{12}|}$

c. $K^{\dagger} = \dfrac{1+(0.25)^2-(0.195)^2-(0.508)^2}{2(2.5)(0.139)} = +1.1$

2. Calculate for MAG:

a. $\text{MAG}^{\dagger} = 10\log_{10}\left|\dfrac{S_{21}}{S_{12}}\right| + 10\log_{10}\left|K \pm \sqrt{K^2-1}\right|$

b. $B_1{}^{\dagger} = 1 + |S_{11}|^2 - |S_{22}|^2 - |D_S|^2 = +0.717$

c. $\text{MAG}^{\dagger} = 10\log\left(\dfrac{2.5}{0.139}\right) + 10\log\left(1.1 - \sqrt{(1.1)^2-1}\right) = 10.63\text{ dB}$

Since $B_1 = +0.717$, then the sign between 1.1 and the square root in step C is negative.

10.63 dB MAG *is alright for our needs, so*:

3. Calculate Γ_L required for a conjugate match for the transistor. As stated above, the Γ_S and Γ_L are the values that the transistor must have at its input and its output for a perfect match:

a. $C_2{}^{*} = S_{22} - (D_S\,S_{11}{}^{\Psi}) = (0.508\angle{-32°}) - [(0.25\angle{-61.4°})\,(0.195\angle{167.6°})]$

b. $C_2{}^{*} = 0.555\angle{-33.5°}$

c. $B_2{}^{\dagger} = 1 + |S_{22}|^2 - |S_{11}|^2 - |D_S|^2 = 1 + (0.508)^2 - (0.195)^2 - (0.25)^2$

d. $B_2{}^{\dagger} = +1.157$

e. Therefore, since B_2 equals "+," the sign equals "−" in the Γ_L equation below:

$$\Gamma_L{}^{\dagger} = \frac{B_2 \pm \sqrt{(B_2)^2 - 4|C_2|^2}}{2|C_2|} = \frac{1.157 - \sqrt{(1.157)^2 - 4(0.555)^2}}{2(0.555)} = 0.748$$

So the answer for the magnitude of Γ_L is:

$$\Gamma_L = 0.748$$

f. Now find the angle of Γ_L:

The angle equals the same value as $C_2 = (0.555\angle{-33.5°})$, but opposite in sign. Therefore, the angle of $\Gamma_L = \angle{+33.5°}$.

g. Our complete answer is $\Gamma_L = 0.748\angle{+33.5°}$

∗Use *full* vector algebra ($Z \angle \pm 0°$) in *S*-parameter calculations (e.g., $S_{11} = 0.35\angle{-45°}$).

†Do *not* use full vector algebra; employ only the *S*-parameter's scalar magnitudes (e.g., $S_{11} = 0.35$). $|S_{xx}|$ means to ignore the *sign* of the magnitude, and always consider it positive.

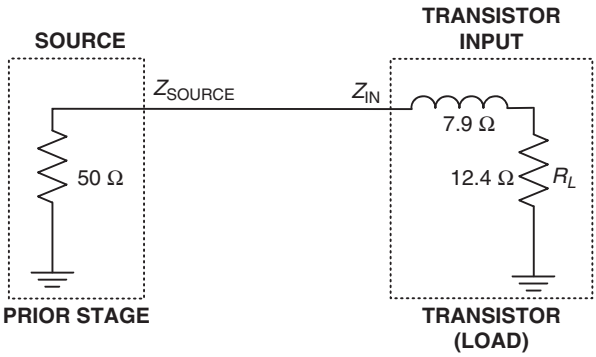


FIGURE 3.22 A transistor's input at RF.

4. Calculate the source reflection coefficient (Γ_S):

   a. $\Gamma_S{*} = \left[S_{11} + \dfrac{S_{12}\,S_{21}\,\Gamma_L}{1-(\Gamma_L \cdot S_{22})}\right]^{\Psi}$

   b. $\Gamma_S{*} = \left[0.195\angle 167.6° + \dfrac{(0.139\angle 61.2°)(2.5\angle 62.4°)(0.748\angle 33.5°)}{1-(0.748\angle 33.5°)(0.508\angle -32°)}\right]^{\Psi}$

   $= [0.61\angle 160.8°]^{\Psi}$

   c. $\Gamma_S = 0.61\ \angle -160.8°$

5. Calculate the input impedance of the transistor (Fig. 3.22):

$$Z_{IN}{*} = Z_{SOURCE}\frac{1+(\Gamma_S{}^{\Psi})}{1-(\Gamma_S{}^{\Psi})} = 50 + j0\left(\frac{1+(-0.576+j0.2)}{1-(-0.576+j0.2)}\right)$$

$$= 50 + j0\left(\frac{0.424+j0.2}{1.57-j0.2}\right) = 12.4 + j7.9$$

   where Z_{SOURCE} = transistor's source (the prior stage) placed at its input (Typically 50 Ω; written as 50 + *j*0 for this formula)

6. Now match Z_{SOURCE} to Z_{IN} with the matching procedures presented further in this chapter.

7. Calculate the transistor's output impedance (Fig. 3.23):

$$Z_{OUT}{*} = Z_{LOAD}\left(\frac{1+(\Gamma_L{}^{\Psi})}{1-(\Gamma_L{}^{\Psi})}\right) = 50 + j0\left(\frac{1+(0.624-j0.413)}{1-(0.624-j0.413)}\right) = 70.5 - j132$$

   where Z_{LOAD} = impedance of the transistor's load

8. Now match Z_{OUT} to Z_{LOAD} with the procedures presented later in this chapter.

9. At this time it is possible to calculate the transducer gain (G_T; the actual gain of an amplifier stage, which includes the effects of impedance matching and

∗Use *full* vector algebra ($Z \angle \pm 0°$) in *S*-parameter calculations (e.g., $S_{11} = 0.35\ \angle -45°$).

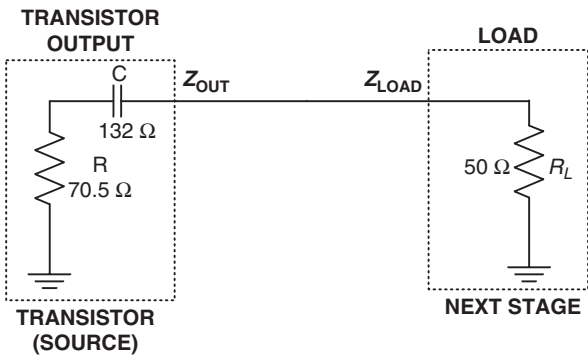


FIGURE 3.23 A transistor's output at RF.

device gain, but does not include power losses in real-world components. G_T will be quite close to the MAG value):

$$G_T = 10\left[\log_{10}\left(\frac{|S_{21}|^2\left(1-\left(|G_S|^2\right)\right)\left(1-\left(|\Gamma_L|^2\right)\right)}{\left|(1-S_{11}\Gamma_S)(1-S_{22}\Gamma_L)-S_{12}*S_{21}*\Gamma_L*\Gamma_S*\right|^2}\right)\right]$$

NOTE: *All signs (±) must be maintained for all calculated numbers. 1 − Γ_S will subtract 1 from the* real *term of Γ_S, and will change the sign of the* imaginary *term (or 1 − (0.2 + j0.1) = 0.8 − j0.1).*

$\Gamma_S{}^{\Psi}$ equals the complex conjugate of Γ_S. If Ψ is outside a bracket, then the answer to everything within the bracket must be converted into the complex conjugate.

3.4.3 Small-Signal Design and Vector Algebra

When required to utilize full complex numbers ($Z \angle \pm \theta°$) in our amplifier calculations, we can perform the required mathematical functions by the following methods:

To multiply polar quantities:

First, multiply the magnitudes; then add the phase angles.

To divide polar quantities:

First divide the magnitudes; then subtract the phase angles.

To subtract polar quantities:

First, convert to rectangular notation ($R + jX$; see *To Convert Polar Into Rectangular Form* below); then subtract $R_1 - R_2 = R_T$, and $jX_1 - jX_2 = jX_T$; then convert the rectangular answer back to polar.

∗Use *full* vector algebra in *S*-parameters that are so marked. Employ only the *S*-parameter's *magnitudes* if not so marked.

To add polar quantities:

Perform as in subtraction; but *add* the rectangular values ($R_T = R_1 + R_2$; $jX_T = jX_1 + jX_2$).

3.4.4 Small-Signal Amplifier Stability

A typical amplifier must be unconditionally stable across all frequencies and input/output impedances. An amplifier may oscillate at anywhere from low frequencies (where the port terminations are not matched to 50 Ω, and the transistor's gain is naturally high, until it is limited by the matching/coupling network) all the way up to the maximum frequency at which the amplifier still shows greater than unity gain. Also, when the transistor is not stable and begins to oscillate, it can shift the bias point of the stage, consuming more current and increasing internal device dissipation, and possibly causing its destruction.

What does instability in an amplifier look like? Figure 3.24 displays an oscillating amplifier as would be seen in the frequency domain on a spectrum analyzer, showing the possible areas of oscillations. Normally only one or two of these oscillations will "pop up," but they can easily be distinguished from other spurs by placing your finger in the area of the (low voltage and low current!) circuit, and then closely seeing if these spurs begin to shift in frequency. If so, then you are viewing instabilities in the amplifier, and these must be eliminated by stabilizing the circuit. (Even though most unstable amplifiers will oscillate at where the gain is highest, which is at the low frequencies, these low-frequency oscillations will mix with the in-channel frequencies, causing visible oscillations at and near our passband of interest.)

Some engineers erroneously believe that an amplifier need only remain stable when presented with 50 Ω, since that is what the input of the following stage will be—or at least is expected to be. However, unconditional stability across all frequencies and impedances is critical, since the next or prior stage, which would normally be a filter or another matched (narrowband) amplifier, will be presenting

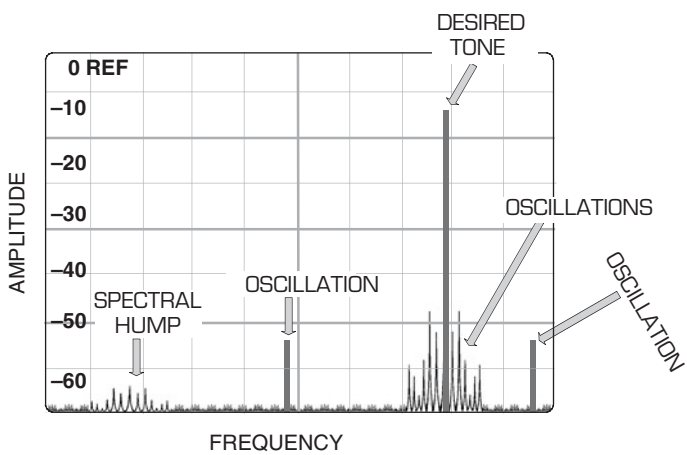


Figure 3.24 In the output of an unstable amplifier one or more of these spectra will be seen.

anything *but* 50 Ω when looking outside of their own passbands. This means that the amplifier may actually be seeing a VSWR of 18:1, or even higher, at certain out-of-band frequencies, and thus must be stable across the entire area that it possesses any appreciable gain.

Conditionally stable transistors may still be used in amplifier designs, but only if the amplifier stage is never presented with any impedance that would cause it to become unstable. Nevertheless, selecting a transistor that is unconditionally stable will give us one less thing to worry about. But since such a transistor may not be all that common, we may have to take a conditionally stable device and force it into unconditional stability. Such a conditionally stable transistor will have some frequencies that have conditional points of stability and, through inductive emitter degeneration, collector and base resistive loading, or negative feedback, we can make this conditionally stable transistor into a fully stable amplifier at all frequencies.

Stability of an amplifier stage is dependent on a transistor's temperature, bias, signal level, and H_{FE} spread, as well as the active device's positive internal feedback mechanism, excessively high gain outside of the desired bandwidth (usually at lower frequencies), and *external* positive feedback caused by support components, PCB layout, or an RF shield's *box modes* (see *RF Shielding Resonances*, Sec. 13.3.2).

Since amplifier instability will be the number-one major headache of any circuit design engineer, it will benefit us to delve back again into this important issue. As stated, virtually all active devices are quite stable when presented with a perfect 50-Ω source and load over the entire frequency range that the device exhibits gain. Indeed, most instability problems occur when the circuit designer does not take into account the elevated *low-frequency* gain of a normal amplifier (Fig. 3.25), and the transistor's possible instabilities when presented with anything other than a 50-Ω termination. This lack of a 50-Ω termination as the frequency is decreased can be due to two main causes: (1) The amplifier's matching circuits are good only for a narrow band of frequencies, so they

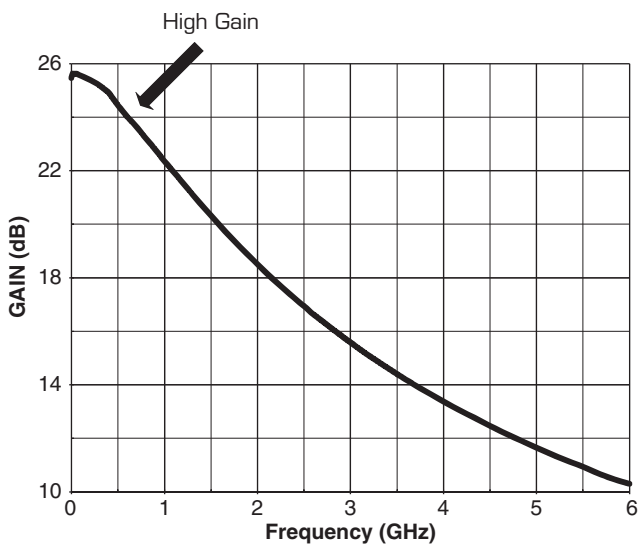


Figure 3.25 No gain compensation in an (unmatched) common-emitter amplifier.

will present the desired impedance to the transistor over a relatively restricted range; (2) The inductor adopted for decoupling of the low impedance power supply becomes closer to a short circuit as this frequency is decreased. These issues can create instability at low frequencies, as a low-value lumped or distributed inductor employed for RF decoupling will only present a true RF choke response over a higher band of limited frequencies. This means that as the frequency of the stage is decreased, the "open" circuit of the RF choke will begin to look more like just a piece of straight, low impedance, wire than a real choke, causing the amplifier to now possess a load that no longer appears as 50 Ω. This can create oscillations in a conditionally stable amplifier. One way to lessen this instability is to add a high-value low-frequency choke in series with the low-value (distributed or lumped) RF inductor. Such a high-value choke will be capable of eliminating these lower frequencies.

Another technique used to ensure that no low-frequency oscillations are created in an amplifier is to employ a 50-Ω resistor at the DC end of the bias circuit (Fig. 3.26). This allows the amplifier to operate into an excellent 50-Ω termination at frequencies so low that the distributed RFC would have little effect. The termination resistor works to stabilize a conditionally stable amplifier, and is only required if there is no other series-voltage-dropping resistor located between the transistor's collector and its V_{CC}. Looking at Fig. 3.26 again, we see that capacitor C_B also helps to shunt low frequency RF to ground, further decreasing the disruptive low-frequency RF gain.

Figure 3.27 is another configuration that is capable of maintaining strong decoupling from the power supply at lower frequencies, as discussed above. The circuit accomplishes this by using both a low- and a high-frequency choke to sustain a high impedance into

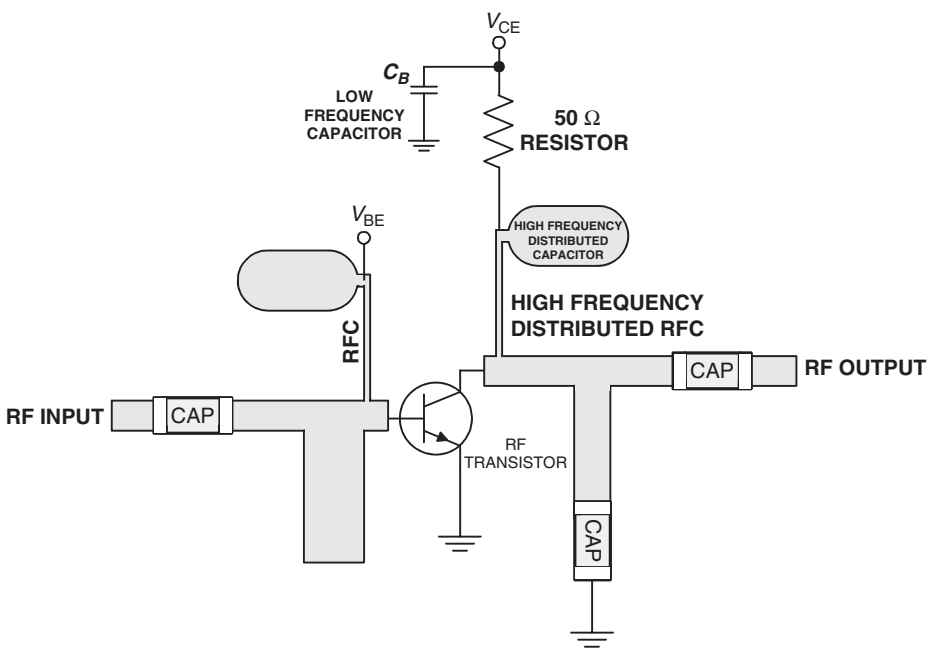


Figure 3.26 Terminating low frequencies into 50 Ω to prevent instability.

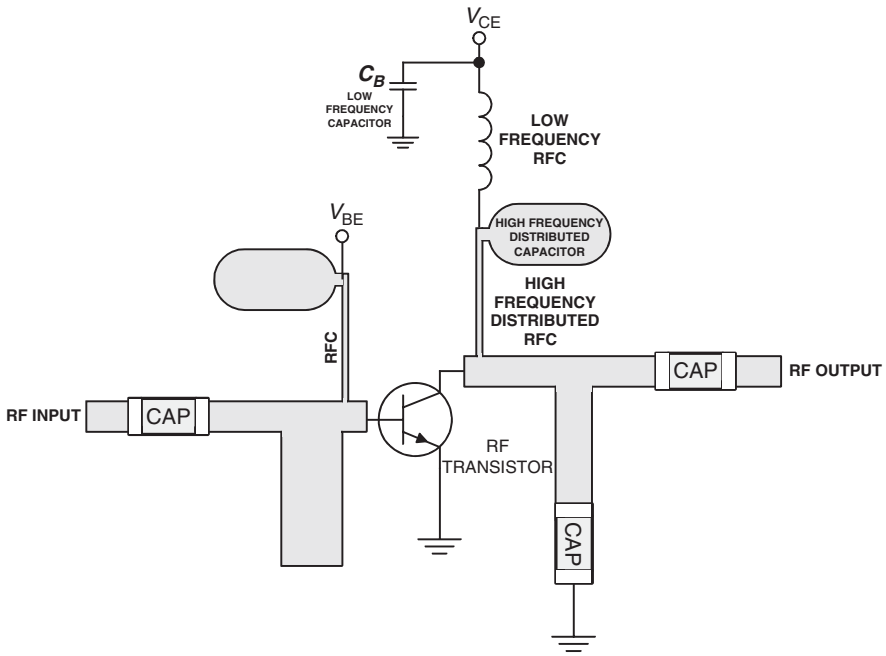


Figure 3.27 Low-frequency decoupling to prevent instability.

the power supply. A de-Q'ing resistor is sometimes required across the amplifier's base or collector bias choke in order to lower stage gain, and thus increase the amplifier's stability.

An amplifier can also be resistively loaded, either in series or in shunt, at its output (Fig. 3.28), thus increasing stage stability, but at the cost of some gain. When this method is used in an LNA circuit, the loading must only be employed at the amplifier's output so as not to degrade the noise figure, while a PA must only be loaded at its input to maintain efficiency, unless absolutely unavoidable. A low-value series resistor of about 2 to 20 Ω is common, while in shunt a 200 to 500-Ω resistor can be used. Most transistors should only need a single stabilization resistor, while others may require two for both high- and low-frequency stability.

Always carefully utilize a loading resistor that will balance the amplifier's increased stability versus the loss in stage gain, and if the instability is only at the lower frequencies, then we may also bypass a series stabilization resistor with a low-value capacitor, and thus gain will not be degraded at the higher, desired frequencies.

To completely maintain stability within an amplifier circuit it should always be considered that the PCB's circuit elements themselves may add a larger feedback path for oscillations than even the transistor itself. Since Barkhausen's criterion for oscillations, even though not always infallible, is a loop gain of unity or higher and an in-phase (regenerative) feedback from output to input, then we can see that at certain frequencies, and with a high enough feedback path somewhere on the PCB board, oscillation can become a problem if the layout is poor. This also means that the higher the gain of an amplifier stage, the more likely it is that oscillations will begin to break out. For predictable stability from a single stage, 25 dB is considered the maximum safe gain.

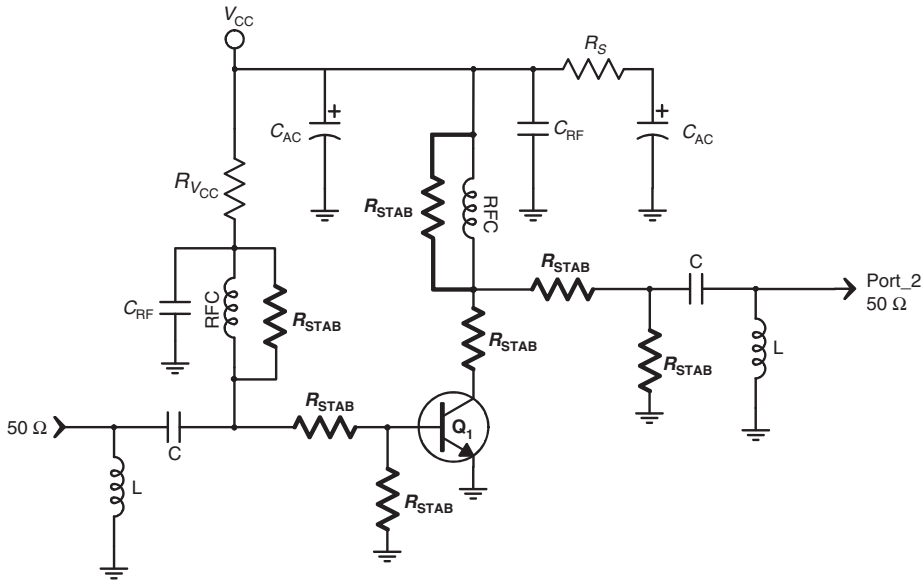


Figure 3.28 Possible locations for a shunt or series stabilization resistor to maintain a sufficient K and B_1. LNAs will demand only output stabilization resistors for an optimal NF, while PAs will demand only input resistors for maximum efficiency and P_{OUT}.

Amplifier feedback oscillations (comparable to "public address" speaker-to-microphone feedback) is an avoidable, and quite common, source of amplifier instability. It is strongly associated with common PCB layout issues, and can be suppressed by proper RF decoupling in the V_{CC} circuit; by not permitting RF traces from the amplifier's output port from getting near and coupling into its input port; and by making sure that the power supply line of some prior stage is not running under the RF or V_{CC} line of the current stage.

Further instability problems can be created if we permit a printed circuit board's groundplane to become excessively segmented, since amplifier oscillations may then occur, especially in PAs. This segmentation, or breaking-up of the groundplane (Fig. 3.29), can arise if any DC or RF board traces are placed on the dedicated groundplane layer of the PCB. When laying out a two-sided board, where real estate is at a premium, this is unfortunately relatively common. As with most instabilities, these oscillations may be at a single frequency or at multiple frequencies, and can be strongly affected in amplitude by touching almost any part of the top groundplane pour, as well as in amplitude *and* frequency by touching the input or output matching networks of the oscillating stage. With a segmented groundplane, amplifier operation will be unpredictable, and the amplifier may sometimes oscillate, while sometimes it may not. As with any unstable oscillator, this apparently intermittent oscillating issue can be caused by changes in temperature, frequency, RF input power, output impedance, or V_{CC}.

Even the small RF leakage that takes place in a time division duplex (TDD) system's receiver/transmitter RF switch can affect the stability of the transmitter section's PA, since the PA will still see some of the reactances within the receiver's (off) input port. Selecting a high isolation RF switch can decrease the importance of this issue.

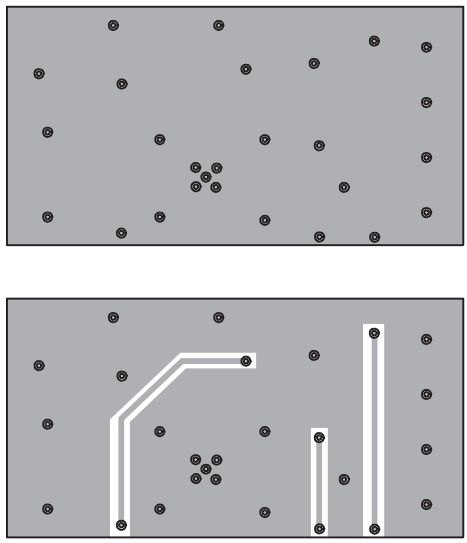


Figure 3.29 Bottom of PCB, showing the copper groundplane unsegmented, and heavily segmented.

When cascading multiple stages, we should always attempt to isolate the grounds of each amplifier so as to decrease the chances of any positive feedback loops, as well as strongly decouple each amplifier's V_{CC} line from other stages. This isolation of the grounds can simply involve each stage having separate vias to the PCB's groundplane, while the V_{CC} power supply that is providing for multiple amplifiers is an obvious source of massive feedback if the decoupling is not excellent.

When cascading amplifiers that are not matched to each other, stability can be highly affected if a transmission line of any significant length is added between these stages. This is because the trace itself can obviously alter the phase angles of the interstage impedances due to the varying voltage and current along any trace that is not terminated in 50 Ω. Nonetheless, unless poor layout causes RF feedback to be created, when all stages within a cascade are unconditionally stable, then the entire cascade should be stable as well.

Since the Rollet stability factor, *K,* is not valid for amplifiers with more than a single stage, we would have to employ what is called a *Nyquist stability test* in the form of an S-probe analysis. In fact, even using both K and B_1 is not a 100% guarantee of a stable circuit for even linear or nonlinear single stages, much less multistage amplifiers—and especially not for a circuit with a feedback loop. However, a Nyquist stability test can only be used when an explicit pair of source/load terminations are specified. Due to its limited utility, this type of stability test will not be outlined here.

Frequency sweeping an amplifier with a vector network analyzer to check for any excessive gain peaking (especially out of band) is a good way to test for instabilities, since the frequency point of highest gain has the strongest probability for instability and oscillations. However, a full mechanical load-pull test is the dominant way to vary the stage's VSWR and phase angles to check for unconditional stability at all supply voltages, device and ambient temperatures, and various load impedances. Indeed, we can perform a cursory check for stability by employing a simple *slide-stub tuner*, rather than invest in a fully automated load-pull setup.

While load-pulls are used routinely in power amplifier design, they should also be exploited to test small-signal amplifiers for unconditional instability. Such load-pull setups are extremely valuable, since a stability test between two ideal wideband 50-Ω terminations will not tell us if the amplifier will become unstable when a reactive circuit, such as a typical filter or antenna structure, is actually attached to it. This is because an amplifier, when placed in a real circuit cascade with other reactive circuits, will virtually never see a perfect 50-Ω resistance at all frequencies. The out-of-band impedance of the next stage, whether it is a filter, antenna, mixer, or amplifier, will be anything *but* 50 Ω, and will have an inductive or capacitive reactance. This can, and does, lead to unexpected instability in many realized amplifiers.

Neutralization (degenerative feedback) is sometimes used to stabilize a potentially unstable amplifier. Nonetheless, the amplifier neutralization procedure will only be successful if the positive feedback path that created the instability is internal to the transistor, and cannot be guaranteed if poor layout or the lack of input or output shielding creates the return feedback path. And due to the variations in input and output capacitance of a bipolar transistor with changes in frequency and bias currents, neutralization is also extremely problematic with wideband transistor amplifiers, as well as the neutralization retuning requirements faced with transistor-to-transistor lot variations. Nonetheless, using some out-of-phase feedback will help to improve stage stability, as well as IMD performance.

3.4.5 Small-Signal Design Approximations

Due to time constraints when searching for a small-signal transistor with a specific gain and stability, we may not want to be as formal with our calculations as we have been up to now. Ballpark figures will suffice. In fact, there is a much faster *S*-parameter method to obtain gain and stability figures called *scalar approximations*, and are utilized for amplifiers to obtain approximate design values.

For the all following formulas, only the magnitude of the S-parameters are employed, and not the phase angles:

1. G_{tu} (*transducer unilateral gain*) measurement, in decibels, of an amplifier's power gain into an unmatched 50-Ω load, a worst case gain value, can be roughly calculated by:

$$G_{tu} = 10 \cdot \log_{10} (|S_{21}|)^2$$

2. Mismatch losses (αp) at the transistor's input or output, in decibels, can be calculated by:

$$\alpha p_{IN} = -10 \cdot \log_{10} (1 - S_{11}{}^2)$$
$$\alpha p_{OUT} = -10 \cdot \log_{10} (1 - S_{22}{}^2)$$

With the total mismatch loss for the entire unmatched transistor as:

$$\alpha p_{TOTAL} = \alpha p_{IN} + \alpha p_{OUT}$$

3. MAG (maximum available gain) is calculated by:

$$MAG = G_{tu} + \alpha p_{TOTAL}$$

4. Since designing for MAG is not recommended due to an amplifier being potentially unstable at this high gain value, we would like to be able to compute the *MSG*, or the *maximum stable gain*:

$$\text{MSG} = 10 \cdot \log_{10}(|S_{21}| \div |S_{12}|)$$

Thus, if the MAG is smaller than the MSG, then the amplifier will be unconditionally stable (unless poor circuit layout produces an external feedback path).

Employing scalar approximations is a more rapid technique than the methods presented in the prior pages, since we are using only the *magnitude* of the *S*-parameters, and not the phase angle. As an example, we are given a transistor with the following *S*-parameters:

$$S_{11} = 0.195 \angle 167.6°$$

$$S_{22} = 0.508 \angle -32°$$

$$S_{12} = 0.139 \angle 61.2°$$

$$S_{21} = 2.5 \angle 62.4°$$

Therefore $S_{11} = 0.195$; $S_{22} = 0.508$; $S_{21} = 2.5$; $S_{12} = 0.139$

$$G_{tu} = 10 \cdot \log_{10}(|2.5|)^2 = 7.96 \text{ dB}$$

$$\alpha P_{IN} = -10 \cdot \log_{10}(1 - 0.195^2) = 0.168 \text{ dB}$$

$$\alpha P_{OUT} = -10 \cdot \log_{10}(1 - 0.508^2) = 1.29 \text{ dB}$$

$\alpha P_{TOTAL} = 0.168$ dB + 1.29 dB = 1.46 dB (Which demonstrates that about 1.46 dB will be gained by proper impedance matching.)

MAG = 7.96 dB + 1.46 = 9.42 dB (10.63 dB was calculated for this same transistor with the full MAG method described earlier.)

$$\text{MSG} = 10 \cdot \log_{10}(|2.5| \div |0.139|) = 12.55 \text{ dB}$$

With the MAG < MSG by over 3 dB, even with our approximation methods, we can see that this transistor will be very stable.

3.4.6 Small-Signal Matching Network Design

The most common amplifier matching networks are lumped LC type L, T, and PI circuits (Fig. 3.30).

Once all of the information on the parameters of an active linear device are assembled, we will need to design the stage's matching network. This network is required so the amplifier's impedances will exactly match the impedances of the circuit it will be inserted into (Fig. 3.31), so that we can obtain the maximum power transfer from one stage to another, and with no power reflections, or:

$$Z_{SOURCE} = R + jX \text{ must equal } Z_{LOAD} = R - jX$$

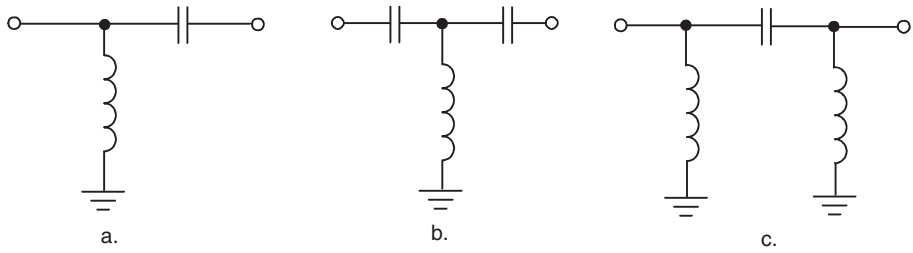


FIGURE 3.30 Three popular matching networks: (*a*) L; (*b*) T; and (*c*) PI.

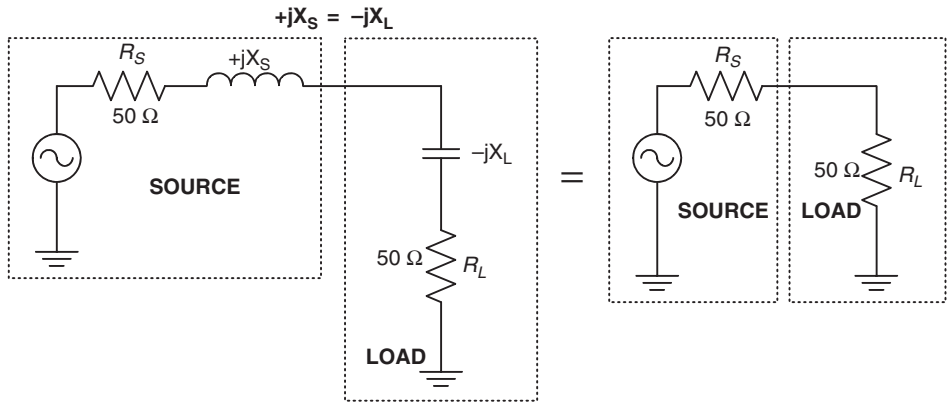


FIGURE 3.31 Canceling reactances along with equal resistances maximizes power transfer.

However, there is only one frequency which will be *perfectly* matched from source to load, since X_C and X_L are frequency dependent, or:

$$L = \frac{X_L}{2\pi f} \qquad \text{and} \qquad C = \frac{1}{2\pi f X_C}$$

Nonetheless, we may obtain a decent return loss over a very wide band of frequencies by proper matching techniques with the correct matching network.

When performing interstage matching, it is possible to choose between two methodologies: Match the output impedance of the first stage to the input impedance of the second stage. This uses the fewest amount of components. Or, match everything to 50 Ω for standardization, which also permits the testing of each stage of the final physical design with normal 50-Ω test gear.

One of the most critical considerations at these frequencies is that the higher the frequency, the physically closer we must keep our calculated matching network to the actual device to be matched. At *lower* frequencies, this is not nearly as much of a consideration. In other words, at the lower frequencies we can place our calculated matching network components almost anywhere on the PCB's microstrip and still obtain a very good match, but at RF frequencies this practice will dramatically alter the proper match. More on this important effect later.

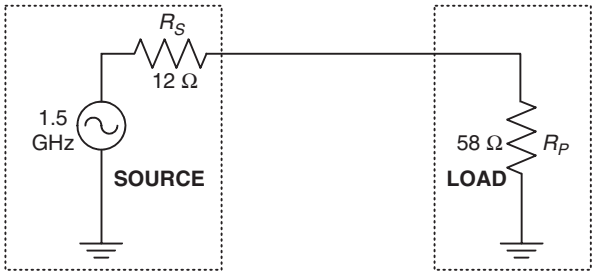


Figure 3.32 An unmatched source and load.

Lumped L Matching

The simple, but very popular, L matching network has the disadvantage that the loaded Q of the circuit cannot be chosen at the start of your calculations, as it can in the more complex networks shown below. In many applications, a low-loaded Q is desired to increase the bandwidth of a linear amplifier (as well as for nonlinear power amplifier design to decrease lossy circulating currents). Still, the value of Q is usually naturally low in an L network, and thus will suffice for most semi-wideband matching needs.

First, to design a basic *resistive- matching-only* L network, used for matching two different value resistances of R_S and R_p (Fig. 3.32), the network topology must be chosen. For a high-to-low impedance transformation, choose Fig. 3.9; for a low-to-high impedance transformation, choose Fig. 3.10:

1. Find the natural Q of the circuit by the formula:
   a. $Q_S* = Q_p* = \sqrt{\dfrac{R_P}{R_S} - 1}$
   b. $Q_S = Q_P = \sqrt{\dfrac{58}{12} - 1}$
   c. $Q_S = Q_p = 1.96*$
2. Find the reactance of element X_p of the L network (Fig. 3.33):
   a. $X_P = \dfrac{R_P}{Q_P}$
   b. $X_P = \dfrac{58}{1.96}$
   c. $X_p = 29.6\ \Omega$
3. Find the reactance of element X_S of the L network:
   a. $X_S = Q_S R_S$
   b. $X_S = 1.96 \times 12$
   c. $X_S = 23.5\ \Omega$

*Must be a positive number.

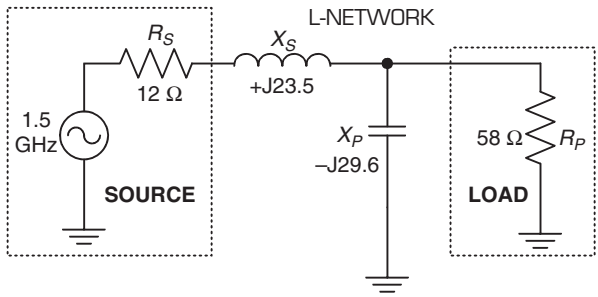


FIGURE 3.33 Matching two different source and load resistances with an L network.

4. To convert the calculated X_S reactance into an inductor value:
   a. $L = \dfrac{X_S}{2\pi f}$
   b. $L = \dfrac{23.5}{2\pi\ 1.5\ \text{GHz}}$
   c. $L = 2.5nH$
5. To convert the calculated X_p reactance into a capacitor value:
   a. $C = \dfrac{1}{2\pi f X_p}$
   b. $C = \dfrac{1}{2\pi(1.5\ \text{GHz})\ 29.6}$
   c. $C = 3.58pf$

   The completed matching network is shown in Fig. 3.34.

When we are in a high-to-low impedance matching situation, simply change the R_p designation to R_S, and R_S to R_p, and then use the same calculations as above. (This switch is needed because the L network's matching capacitor X_p is now in parallel with the source's resistance, instead of the load's resistance.)

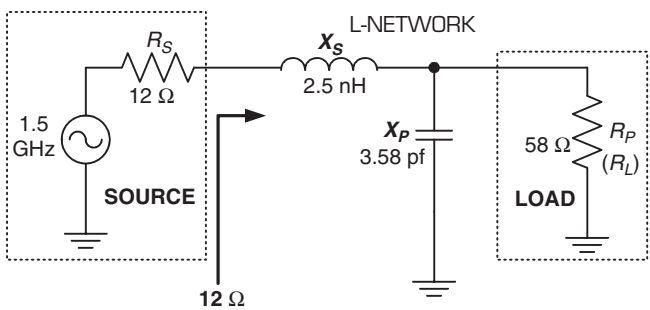


FIGURE 3.34 The final L network component values for a matched source and load.

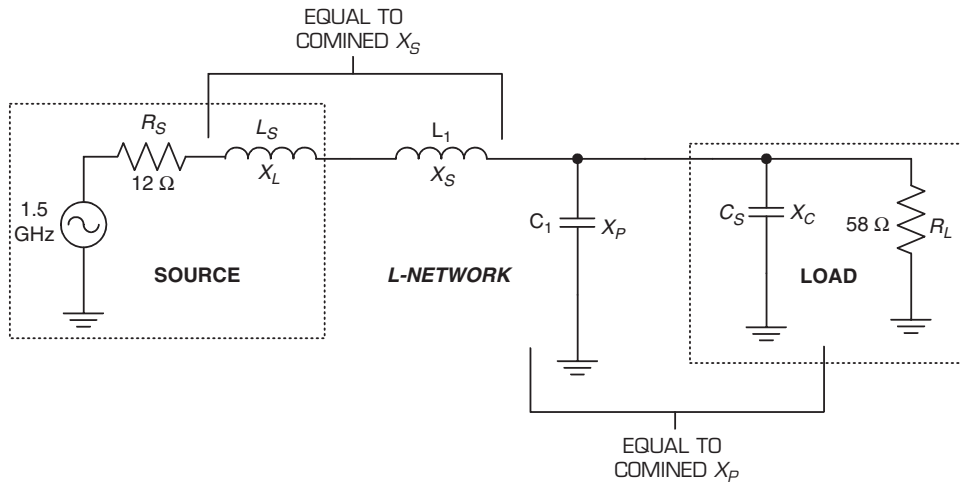


Figure 3.35 Circuit that requires the addition of components to absorb reactances.

When two different, but pure, resistances must be matched, the above technique is easily and rapidly applied to perform this task. However, if reactances must also be cancelled within one or both of these circuits—as well as each resistance matched—then one (or both) of the two following methods may be employed.

Absorption uses the reactances of the impedance matching network itself to absorb the undesired load and/or source reactances (Fig. 3.35). This is accomplished by positioning the matching inductor in series with any load or source inductive reactance. In this way, the load or source's X_L actually becomes *a part* of the matching inductor. The same outcome can be attained by positioning a matching capacitor in parallel with any load or source X_C, thus combining the two values into one larger value. This allows the internal stray reactances of both devices to actually contribute to the matching network; with these internal reactances now being subtracted from the calculated values of the LC matching components. In other words, the transistor's own stray reactances are now becoming an additive part of the matching network. This method is only useful if the stray internal reactances of the device are less than the calculated reactances required for a proper match, which is normally the case.

The other technique is *resonance*, which is utilized to resonate out the stray reactances of the device or circuit to be matched at our desired frequency, with a reactance that is equal in value, but opposite in sign; and then continuing on as if the matching problem were a completely resistive one ($R + j0$). This will make the internal stray reactances of the two devices or circuits disappear, thus allowing only the pure resistances to be easily dealt with.

The first approach, *absorption*, is demonstrated with the practical example of Fig. 3.36*a*:

1. Disregard *all* source and/or load internal reactances.
2. Place an L network in series with the internal stray X_L of the source, and the capacitance in parallel with the internal stray X_C of the load (Fig. 3.36*b*).

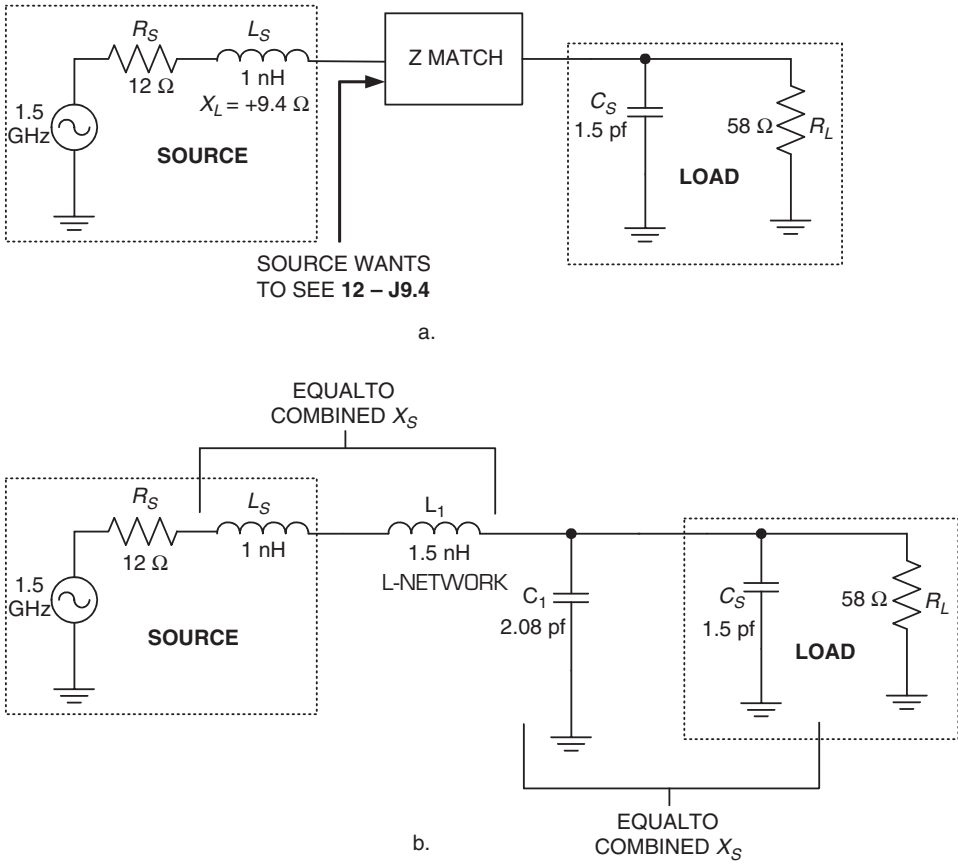


Figure 3.36 (*a*) Circuit that requires matching, and (*b*) the addition of components to absorb reactances.

3. While still neglecting all of the stray reactances, use the formulas and methods under the resistive lumped L matching explanation, as outlined above, to calculate and match R_S to R_L.
4. Subtract the internal stray reactance values from the L network's calculated values of L_1 (2.5 nH) and C_1 (3.58 pF); which in this case will be 2.5 – 1 nH = 1.5 nH = L_1; and 3.58 –1.5 pF = 2.08 pF = C_1.
5. The new L network component values are now the actual values required to obtain the proper 12 – *j*9.4 conjugate match for the 12 + *j*9.4 source (or $Z_L = 12 + j0$).

To design a matching network employing the second method, the *resonance* approach, view the example circuit of Fig. 3.37:

1. Resonate out the 1.5 pF of stray capacitance within the load by employing a shunt inductor with a value of $L = \dfrac{1}{|2\pi f|^2 C_{\text{STRAY}}}$, or L = 7.5 nH (Fig. 3.38). The

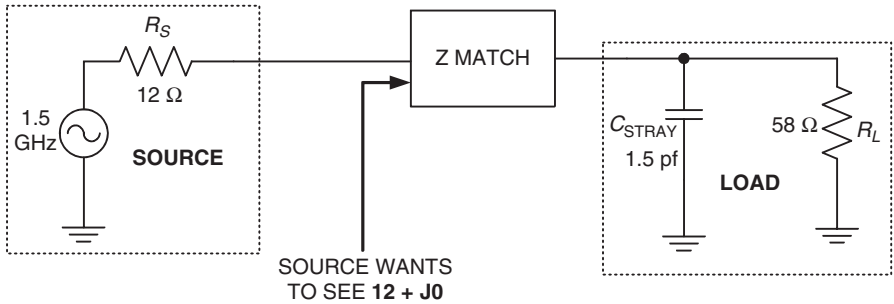


Figure 3.37 Example circuit for the resonance impedance matching design approach.

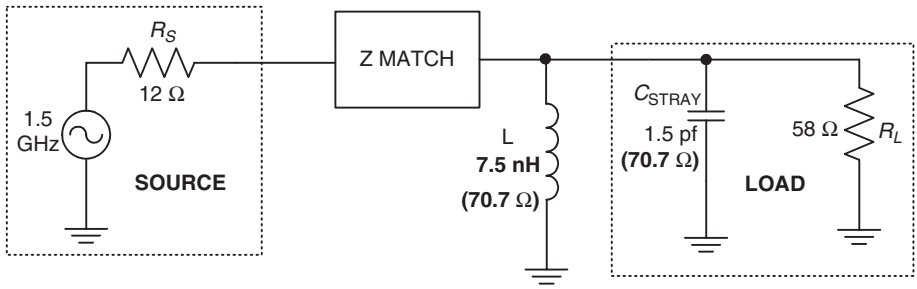


Figure 3.38 Canceling the load's stray reactance.

internal stray capacitance can now be considered as no longer existing within the load.

2. Since the source is purely resistive ($Z_S = R_S + j0$), and the load is now as well ($Z_L = R_L + j0$), we can utilize the formulas under the basic resistive lumped matching techniques outlined above to design an L network to match the source to the load.
3. Simplify by combining both of the inductors we now have (Fig. 3.39) with a single inductor by (Fig. 3.40):

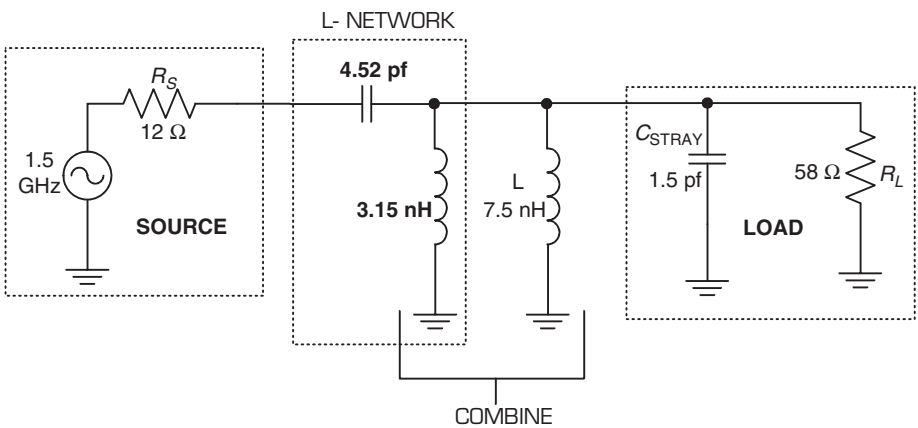


Figure 3.39 Adding the two inductors.

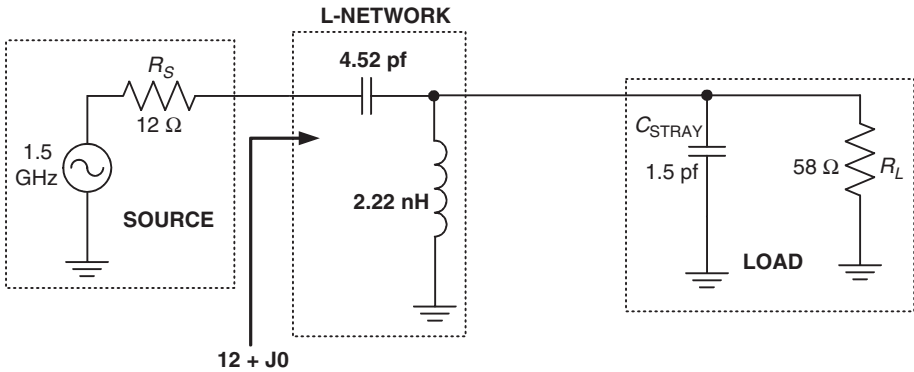


FIGURE 3.40 Combining the two inductors into one for an L network.

$$L_{NEW} = \frac{L_1 L_2}{L_1 + L}, \qquad \text{or} \qquad 2.22 \text{ nH} = L_{NEW}$$

The 50-Ω source is now perfectly matched to the complex load.

PI and T Network Matching

Three-element impedance-matching (PI or T) networks are popular in many narrowband applications. The narrowband popularity is due to the higher loaded Q over what the L network possesses, yet PI and T network's also permit almost any Q to be selected. Nonetheless, T and PI circuits can never be *lower* in Q than an L network. The Q desired for a particular application may be calculated with the following formula, assuming the utilization of high-Q inductors:

$$Q = f_C/(f_2 - f_1)$$

where Q = loaded quality factor of the circuit
f_C = center frequency of the circuit
f_2 = upper frequency that we will need to pass with little loss
f_1 = lower frequency that we will need to pass with little loss

Employ the guidelines below to design a PI network capable of matching two different pure *resistances* (Fig. 3.41). With the following design methodology, consider the PI network as two L networks attached back to back; with a *virtual* resistor ("R") in the center, which is used only as an aid in designing these networks and will not be in the final design.

1. Find "R," the virtual resistance, as shown in Fig. 3.42. In this example, the loaded Q of the PI network is chosen to be 10:

$$"R" = \frac{R_H}{Q^2 + 1} = \frac{58}{10^2 + 1} = 0.57\ \Omega$$

R_H is equal to whichever source or load resistance is larger: R_L or R_S.

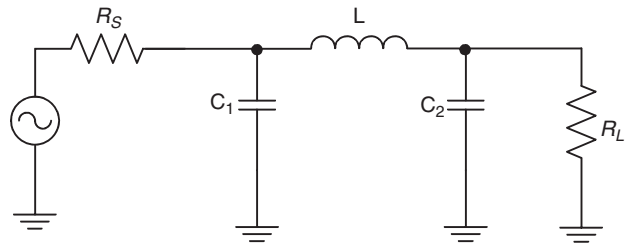


Figure 3.41 A PI matching network between a source and its load.

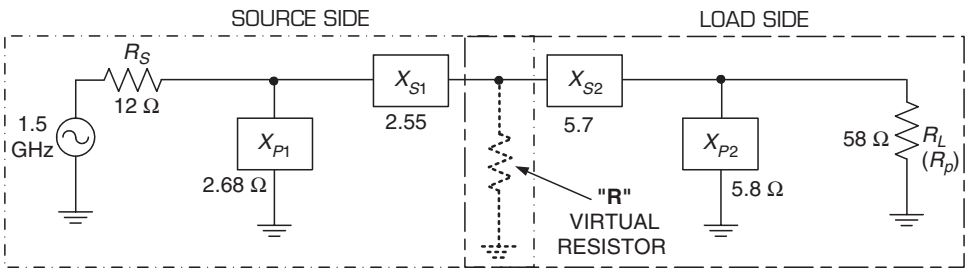


Figure 3.42 Using a "virtual resistor" to design a PI network.

2. Find X_{P2} and X_{S2} by:

$$X_{P2} = \frac{R_L}{Q} = \frac{58}{10} = 5.8\ \Omega \quad \text{and} \quad X_{S2} = Q \cdot "R" = 5.7\ \Omega \quad \text{for the } \textit{load} \text{ side values.}$$

3. Find the value of X_{P1} and X_{S1}:

$$X_{P1} = \frac{R_S}{Q_1} \quad \left| Q_1 = \sqrt{\frac{R_S}{"R"} - 1} = \frac{12}{4.48} \right| 4.48 = \sqrt{\frac{12}{0.57} - 1} = 2.68\ \Omega$$

$$X_{S1} = Q_1 \cdot "R" = 4.48 \cdot 0.57 = 2.55\Omega$$

4. Combine X_{S1} and X_{S2} ($X_{S1} + X_{S2}$) (Fig. 3.43).
5. One of four different PI matching configurations can be chosen, depending on whether we must get rid of stray reactances, pass or block DC, or filter excess harmonics (Fig. 3.44).
6. Convert the reactances calculated to *L* and *C* values by:

$$L = \frac{X_S}{2\pi f} \qquad \text{and} \qquad C = \frac{1}{2\pi f X_P}$$

To match two stages with a PI network, while canceling reactances *and* matching resistances (Fig. 3.45), observe the following procedures. Convert the load/source to/from parallel or series equivalences as required to make it easier to absorb any reactances. (These conversion equations are found later in this chapter).

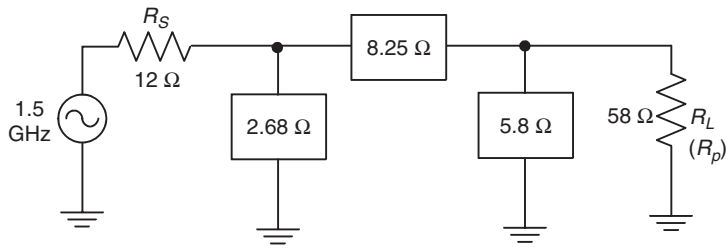


Figure 3.43 Reactance values as calculated for PI network.

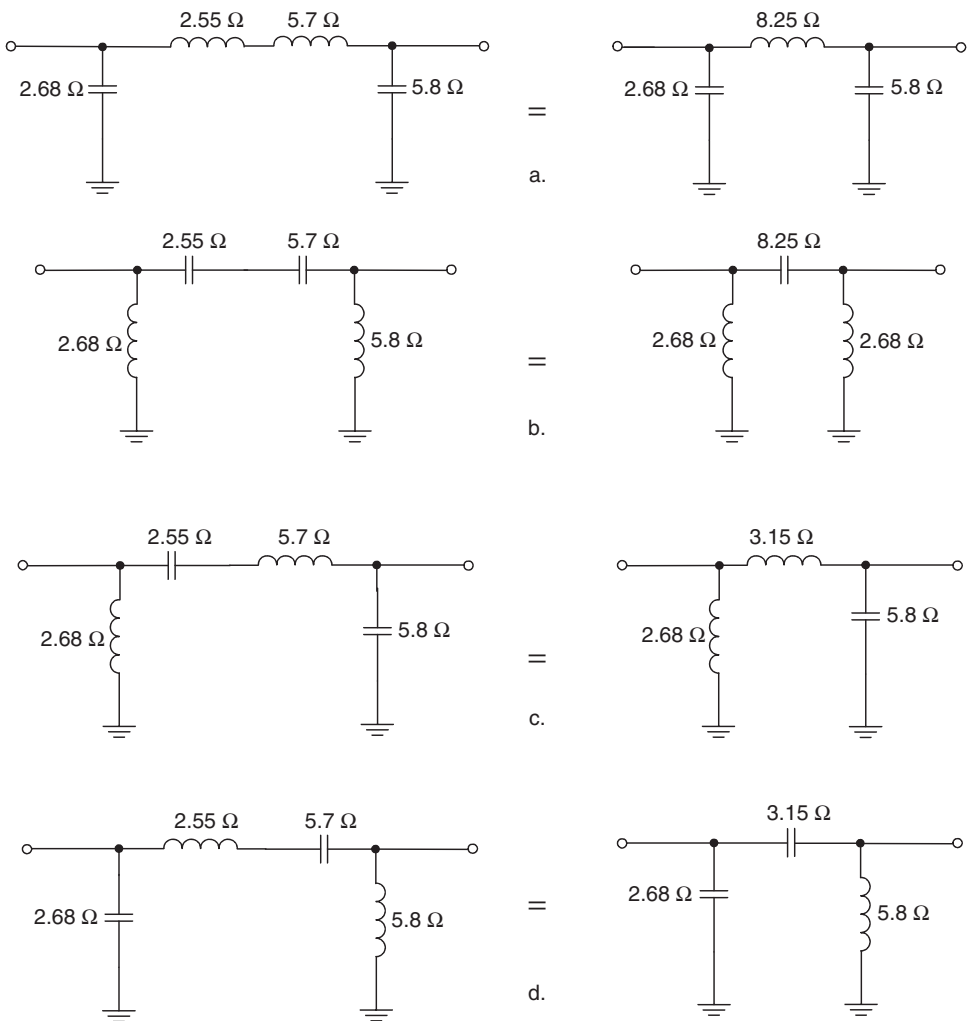


Figure 3.44 Different legitimate PI networks before and after combining components.

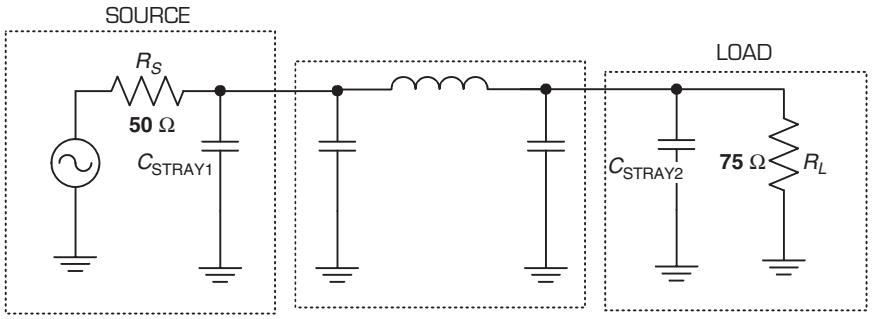


FIGURE 3.45 A PI network used in a resistive and reactive source and load.

1. Select a proper network topology that will *absorb* both stage's reactances. In this case, we would choose a PI network with two parallel capacitors (see above on *Absorption* under "Lumped L Matching").
2. Choose a desired loaded Q and frequency of operation.
3. Find "R," the virtual resistance:

$$"R" = \frac{R_H}{Q^2 + 1}$$

NOTE: *R_H is equal to whichever source or load resistance is larger: R_L or R_S.*

4. Find X_{C2} and X_{L2} by:

$$X_{C2} = \frac{R_L}{Q} \text{ and } X_{L2} = Q \cdot "R" \text{ for the } \textit{load} \text{ side values.}$$

5. Find X_{C1} and X_{L1} by:

$$X_{C1} = \frac{R_S}{Q_1} \quad \left| Q_1 = \sqrt{\frac{R_S}{"R"} - 1} \right. \qquad \text{and} \qquad X_{L1} = Q_1 \cdot "R"$$

6. As shown in Fig. 3.46, add X_{L1} and X_{L2} to form $X_{L\,(NEW)}$; combine $X_{C\,(STRAY1)}$ and X_{C1}; then combine $X_{C\,(STRAY2)}$ and X_{C2} (X_{C1} and X_{C2} must be smaller than $X_{C\,(STRAY1)}$ and $X_{C\,(STRAY2)}$ respectively, since adding two capacitor's reactances in parallel involves:

$$\frac{X_C \times X_{C(STRAY)}}{X_C + X_{C(STRAY)}} = X_C$$

Thus, if $X_{C(STRAY)} < X_{C1}$, then $X_{C(TOTAL)}$ will not be able to reach the proper X_C value. Also, increase X_{C1} until:

$$\frac{X_{C1} \cdot X_{C(STRAY1)}}{X_{C1} + X_{C(STRAY1)}} = X_{C(TOTAL)} = X_{C1} \qquad \text{or} \qquad \frac{X_{C1} \cdot X_{C(STRAY1)}}{X_{C1} - X_{C(STRAY1)}} = X_{C(NEW)}$$

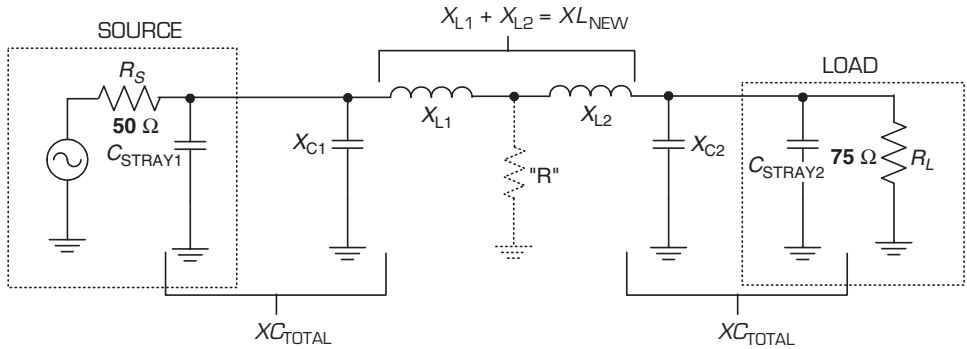


Figure 3.46 PI network before combining calculated components.

This is so X_{C1} and $X_{C(STRAY1)}$, in *parallel*, will still equal the computed value of X_{C1} ($X_{C(TOTAL)}$).

7. Convert the reactances calculated to L and C values by:

$$L = \frac{X}{2\pi f} \qquad \text{and} \qquad C = \frac{1}{2\pi f X}$$

The completed network is as shown in Fig. 3.47.

T networks are required when two low impedances need to be matched with a high loaded Q, and must be of a higher Q than that available with the L network type.

Follow this procedure to match two unequal and pure *resistances*, as shown in the example of Fig. 3.48.

1. Decide on the loaded Q (in this case 15), and the frequency (in this case 1.5 GHz).
2. Find the "R" value by "R" = $R_{S(MALL)}$ ($Q^2 + 1$); "R" = $12(15^2 + 1)$; "R" = 2712 Ω. $R_{S(MALL)}$ is the smaller value of the two resistances, whether it is R_S or R_L.

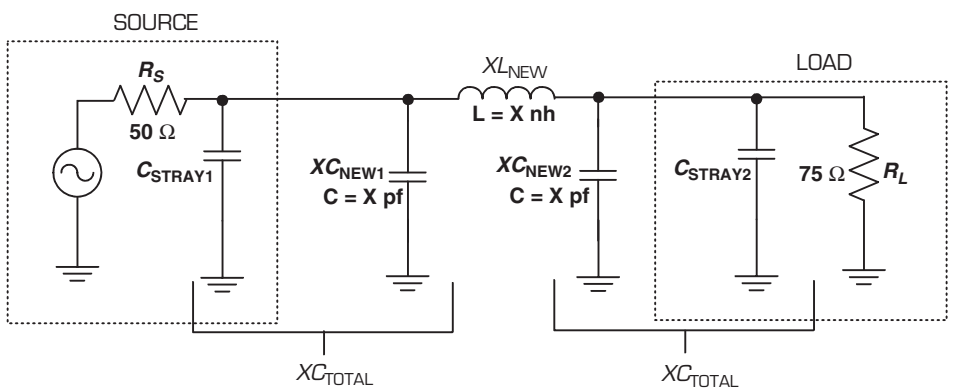


Figure 3.47 PI matching network after combining components.

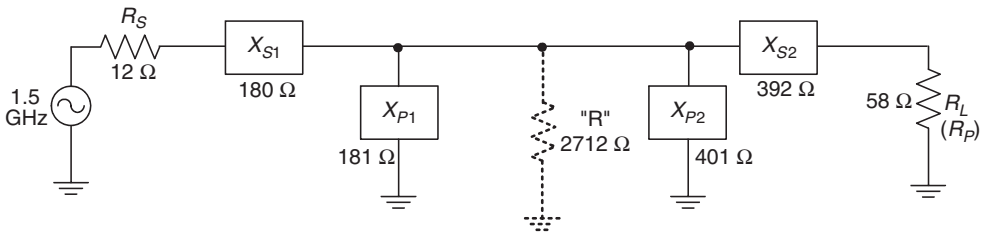


Figure 3.48 Designing a T network for use between a resistive source and load.

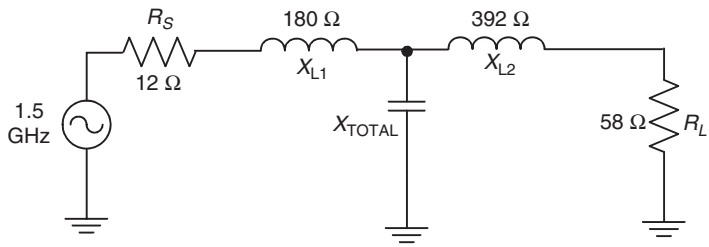


Figure 3.49 Values for a completed T network.

3. Find $X_{S1} = Q \cdot R_S = 15 \cdot 12 = 180\ \Omega$
4. Find $X_{P1} = \text{“}R\text{”}/Q = 2712/15 = 181\ \Omega$
5. Find

$$Q_2 = \sqrt{\frac{\text{"}R\text{"}}{R_L} - 1} = \sqrt{\frac{2712}{58} - 1} = 6.76$$

6. Find

$$X_{P2} = \frac{\text{"}R\text{"}}{Q_2} = \frac{2712}{6.76} = 401\ \Omega$$

7. Find

$$X_{S2} = Q_2 \cdot R_L = 6.76 \cdot 58 = 392\ \Omega$$

8. X_{P1} and X_{P2} are combined by:

$$X_{\text{TOTAL}} = \frac{X_{P1} \cdot X_{P2}}{X_{P1} + X_{P2}} = \frac{181 \times 401}{181 + 401} = 125\ \Omega$$

9. The circuit is shown completed in Fig. 3.49. Other possible circuit configurations can be used as required (Fig. 3.50). Figure 3.50*a*, *b*, and *c* are combined as in step 8 above, but the signs must be maintained for *b* and *c* due to the opposite reactance employed (“+” for inductors and “–” for capacitors).

Wideband Matching

Sometimes it may be necessary to design a low-loaded *Q*, and thus a very wideband, matching network. This can be done as follows, using Fig. 3.51*a* for a pure resistive load

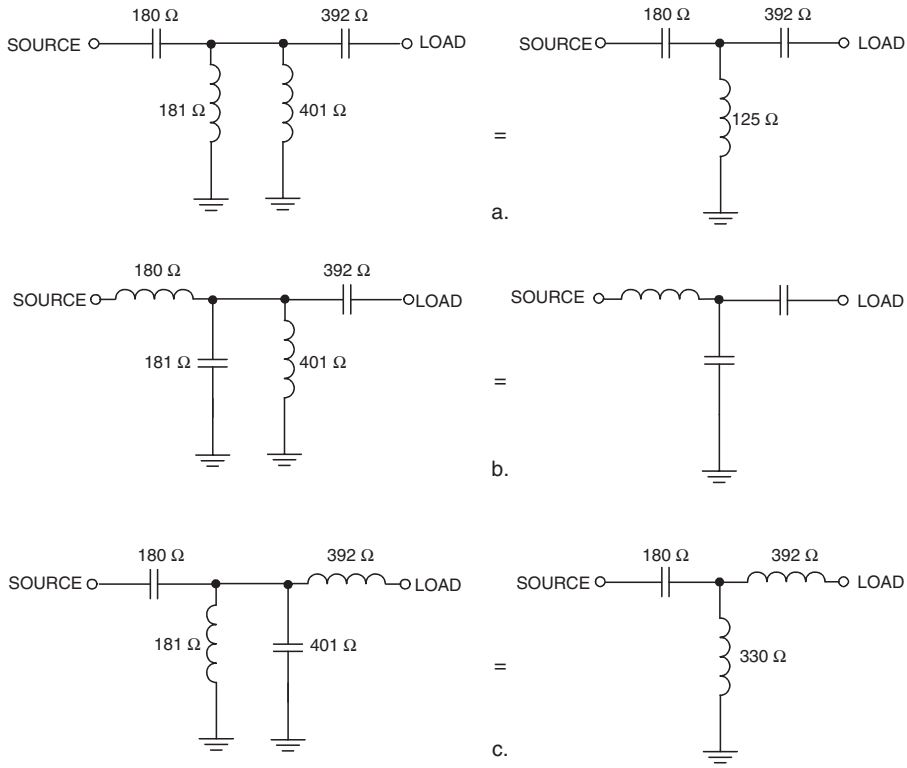


Figure 3.50 The T network and its various legal configurations before and after combination.

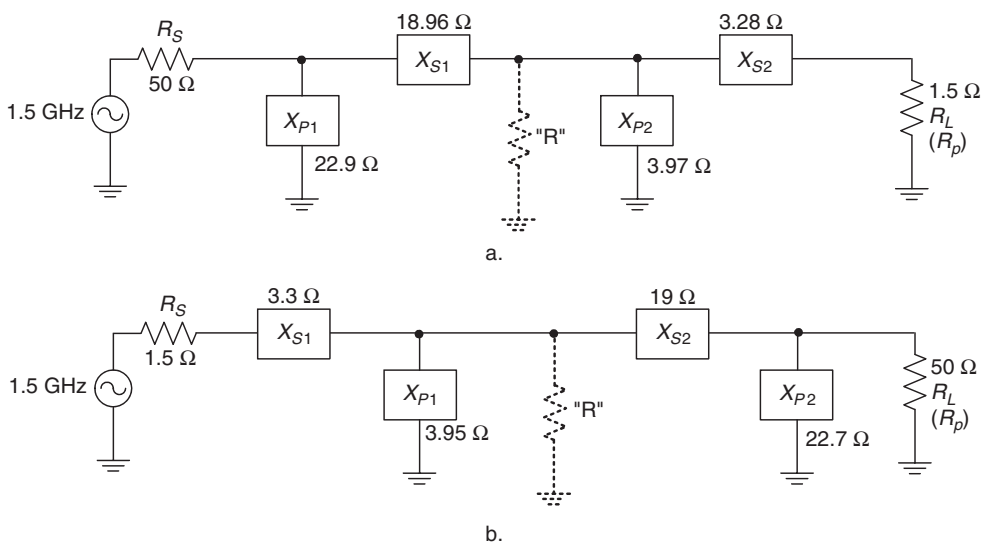


Figure 3.51 Two different configurations for a wideband matching network: (*a*) high to low; (*b*) low to high.

that is smaller than the pure resistive source; or by employing Fig. 3.51*b* for a pure resistive load that is larger than the pure resistive source. X_{S1} and X_{P1} can be considered as a separate L network from X_{S2} and X_{P2}, so each L may be oriented any way that is convenient. For instance, X_{S1} may be an inductor, so X_{P1} must then be a capacitor; however, X_{S2} may be the capacitor, with X_{P2} being the inductor.

1. Solve for "R":

$$"R" = \sqrt{R_S\, R_L} = 8.7\ \Omega$$

2. Solve for loaded Q:

$$Q = \sqrt{\frac{"R"}{R_{\text{SMALLER}}} - 1} = 2.2$$

3. Complete for Fig. 3.51*a* above:

$$X_{P2} = \frac{"R"}{Q_2} \quad \left| Q_2 = \sqrt{\frac{"R"}{R_L} - 1} = 3.97\ \Omega \right. \quad (Q_2 = 2.19) \text{ and } \quad X_{S2} = Q_2 \times R_L = 3.28\ \Omega$$

$$X_{P1} = \frac{R_S}{Q_1} \quad \left| Q_1 = \sqrt{\frac{R_S}{"R"} - 1} = 22.9\ \Omega \right. \quad (Q_1 = 2.18) \text{ and } \quad X_{S1} = Q_1 \times "R" = 18.96\ \Omega$$

4. Or complete for Fig. 3.51*b* above:

$$X_{P2} = \frac{R_P}{Q} = 22.7\ \Omega \qquad \text{and} \qquad X_{S2} = Q \times "R" = 19\ \Omega$$

$$X_{P1} = \frac{"R"}{Q} = 3.95\ \Omega \qquad \text{and} \qquad X_{S1} = Q \times R_S = 3.3\ \Omega$$

It is possible to match for increasingly wider bandwidths by adding sections as shown in Fig. 3.52:

1. Maximum bandwidth is always achieved if the ratios of each of the two ensuing resistances are equal, or:

$$\frac{"R"_1}{R_{\text{SMALLER}}} = \frac{"R"_2}{"R"_1} = \frac{"R"_3}{"R"_2} = \text{etc} = \frac{R_{\text{LARGER}}}{"Rn"}$$

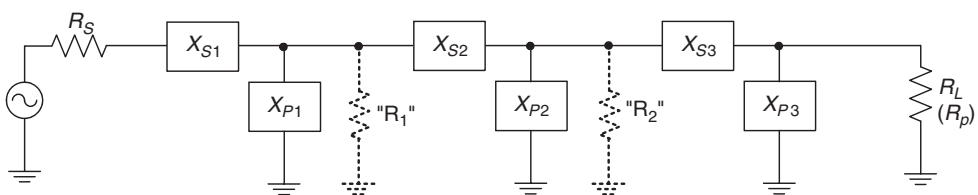


Figure 3.52 Matching for very wide bandwidths.

2. Design as in Fig. 3.51*b* above for this circuit if $R_L > R_S$; or adopt Fig. 3.51*a* design procedure and circuit elements if $R_L < R_S$.

Impedance Matching with Distributed Circuits

In certain high-frequency microwave applications, using distributed matching elements may be a lower cost and higher performance alternative to using lumped parts.

One limitation when using distributed matching elements is the inability to easily create series capacitors. Even though it is possible to design low-value distributed series capacitors into a microwave circuit, it is ordinarily too difficult and inaccurate a procedure to be truly practical. This means that, wherever possible, we should employ shunt distributed capacitors when matching impedances in our microwave designs. But what is the alternative if, for instance, we find that the series input impedance of a device is inductive, and we would like to tune it out? This would generally require a conjugate series capacitance to cancel the transistor's series input inductance. However, since we would like to get away from using a lumped series capacitor, we can convert the *series input impedance* (Fig. 3.53) of the device into an equivalent *parallel input impedance* (Fig. 3.54), which will now permit us to exploit a shunt distributed element to resonate out the input reactance of the device. The formulas to accomplish this conversion are:

$$R_P = R_S + \frac{X_S^2}{R_S} \qquad \text{and} \qquad X_P = \frac{R_P \cdot R_S}{X_S}$$

where R_P = equivalent parallel resistance, Ω
R_S = series resistance, Ω
X_S = series reactance, Ω
X_P = equivalent parallel reactance, Ω

Indeed, we could design the distributed matching circuit as we would the lumped type, and simply substitute the equivalent distributed components as explained under *Microstrip as Equivalent Components*, Sec. 1.3.1. owever, we may find that the calculated *L* and *C* values may be beyond the recommended maximum 30° per wavelength length

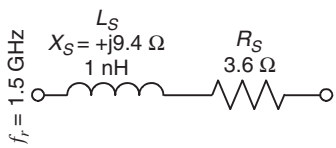


Figure 3.53 Series input impedance.

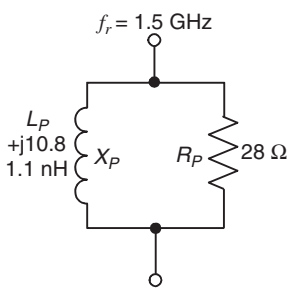


Figure 3.54 Parallel input impedance.

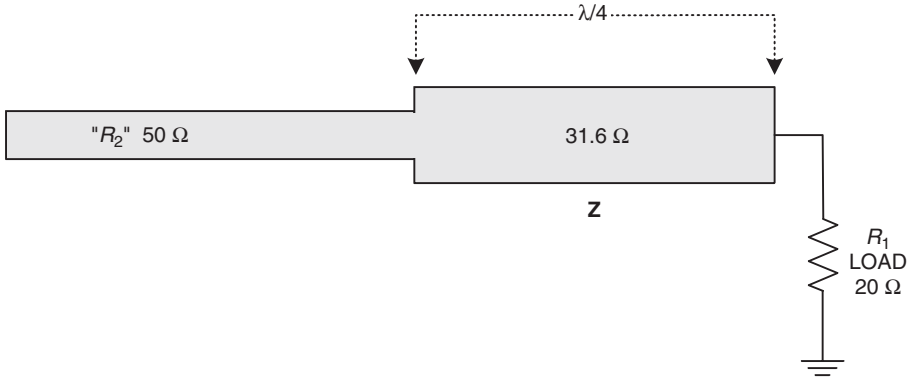


Figure 3.55 Using a distributed transformer to match a 50-Ω resistive source to an unequal resistive load.

limit imposed on accurate equivalent distributed components, thus being either unrealizable, or inferior to a lumped part. In that case, it may be far easier to utilize microwave *quarter-wave line matching*.

For small (and large) signal devices, quarter-wave line transformer (Fig. 3.55) matching can be accomplished as follows:

1. Calculate the input/output impedances of the device to be matched (a series impedance, or $R \pm jX$); or obtain these values from the data sheet.
2. Convert series $R \pm jX$ to parallel as required. Whether we elect to utilize parallel or series will depend on whether it would be easier, with microstrip, to resonate out the reactance in series, or in a parallel equivalence. (If a distributed part must be used for this purpose, a shunt capacitor is always desired).
3. Calculate the required microstrip width and length, at the frequency of interest, to simulate a lumped value that will cancel out the reactive component of the device being matched, making the input or output $R + j0$. Lumped microwave capacitors and inductors can also be utilized if the microstrip part is unrealizable due to being inordinately over 30° in length.
4. Then, match the now real (resistive) part of the transistor's input or output by employing the microstrip transformer. The microstrip transformer section is placed between the two mismatched impedances (in this case, 50 Ω for the system's transmission line impedance, and the 20 Ω for the transistor's input resistance). The transformer segment will be $\lambda/4 \cdot V_p$ long (V_p = *propagation velocity*; see *Microstrip as Transmission Line*, Sec.1.3.2), and as wide as a microstrip transmission line would be with an impedance of $Z = \sqrt{R_1 R_2}$ which, in this case, is 31.6 Ω).

Reflection Coefficients

The magnitude of the reflection coefficient (signified by ρ or Γ) of a circuit or transmission line is simply the ratio between the reflected wave and the forward wave of a signal, or:

$$\rho = \frac{V_{\text{REFL}}}{V_{\text{FWD}}} \qquad \text{and} \qquad \rho = \frac{\text{VSWR} - 1}{\text{VSWR} + 1}$$

The reflection coefficient will always be some value between 0 and 1, since the reflected wave's amplitude will never be higher in amplitude than the height of the forward wave. Most values of ρ, however, will contain both magnitude and phase, instead of simply magnitude as above. These reflection coefficients are an indicator of the quality of the match between one impedance and another, or V_{REFL}/V_{FWD}, with a perfect match equaling zero, and the worst match equaling 1. They can be expressed in rectangular ($\Gamma = R \pm jX$) or polar ($\Gamma = P\angle \pm 0$) forms.

Nevertheless, calculating just the magnitude ratio will allow the computation of the return loss and mismatch for any circuit.

Return loss: $\text{RL(in dB)} = 10\log_{10}(\rho^2)$

Mismatch loss: $\text{ML(in dB)} = 10\log_{10}(1-\rho^2)$

Conversions

There may be occasions when we must convert from the old *Y-parameters* (another way to characterize a transistor) into the newer *S*-parameters for our matching needs; this is rarely required today.

$$Y_{11}{}^{*} = \left[\frac{(1+S_{22})(1-S_{11})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-(S_{12}S_{21})}\right]\times\frac{1}{50}$$

$$Y_{12}{}^{*} = \left[\frac{-2\times(S_{12})}{(1+S_{11})(1+S_{22})-(S_{12}S_{21})}\right]\times\frac{1}{50}$$

$$Y_{21}{}^{*} = \left[\frac{-2\times(S_{21})}{(1+S_{11})(1+S_{22})-(S_{12}S_{21})}\right]\times\frac{1}{50}$$

$$Y_{22}{}^{*} = \left[\frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1+S_{22})(1+S_{11})-(S_{12}S_{21})}\right]\times\frac{1}{50}$$

However, there will be many instances when we must convert from rectangular ($Z = R \pm jX$) to polar ($Z = R\ \angle\theta$) notation, or vice versa, when designing matching networks. The manual technique outlined below, which is good only for *positive* real numbers, is but one method. A simple scientific calculator performs the job much faster and more accurately.

1. *To convert rectangular into polar form* ($R \pm jX$ to $Z\ \angle\theta$):
   a. $Z = \sqrt{R^2 + X^2}$
   b. $\theta = \tan^{-1}\dfrac{X}{R}$ ($\tan^{-1}$ = arc-tangent)

---

*Use *full* vector algebra ($Z\ \angle \pm 0°$) in calculations (e.g., $S_{11} = 0.35\ \angle -45°$).

2. *To convert polar into rectangular form* ($Z \angle\theta$ to $R \pm jX$):
   a. $R = Z(\cos\theta)$
   b. $X = Z(\sin\theta)$

There will also be many times when we must convert a *series* resistance and reactance into a *parallel* resistance and reactance in order to make a certain impedance matching problem easier to solve. Below is another technique for doing this.

To convert from series to parallel, using the examples of Figs. 3.53 and 3.54 ($f_C = 1.5$ GHz):

1. Find the Q of the series circuit:

$$Q = \frac{X_S}{R_S} \qquad \text{or} \qquad Q = 2.62$$

2. If $Q < 10$, use $R_P = (Q^2 + 1) \cdot R_S = 28\ \Omega$
3. If $Q > 10$, use $R_P = (Q^2) \cdot R_S$
4. $X_P = R_P/Q_P = 10.8\ \Omega \qquad (Q = Q_S = Q_P)$
5. $C_P = 1/(2\pi f X_P)$
6. $L_P = 2\pi f X_P$

To convert from parallel to series, also using Figs. 3.53 and 3.54:

1. Find $Q = R_P \div X_P$
2. $R_S = R_P \div (Q^2 + 1)$ [if $Q < 10$]
3. $R_S = R_P \div Q^2$ [if $Q > 10$]
4. $X_S = R_P\left(\dfrac{X_P R_P}{X_P^2 + R_P^2}\right)$
5. $C_S = 1 \div 2\pi f X_S$
6. $L_S = 2\pi f X_S$

Selective Mismatching

Designing an amplifier for a specific gain can be accomplished by selective mismatching at either its input or output port. This is an important technique, since we may not always require all of the gain that can be supplied by a particular transistor. A stage can thus be designed for an explicit gain by not matching the load to the source by some amount. This technique can be quite useful, but it is wise to attempt it only when we are using an unconditionally stable transistor circuit. Even so, fixed PI or T pad attenuators can be adopted safely for this purpose within the amplifier design, if NF is not a large concern.

To carry out selective output mismatching of a transistor to lower its gain by mismatch losses is demonstrated below (Fig. 3.56):

1. Choose the desired gain (G_{DESIRED}) for the amplifier.
2. Calculate $M_L = G_{\text{MAX}}$ (dB) $- G_{\text{DESIRED}}$ (dB)

where M_L = mismatch loss, dB
G_{MAX} = gain from the transistor's data sheet, or use MAG, dB
G_{DESIRED} = gain desired from the amplifier, dB

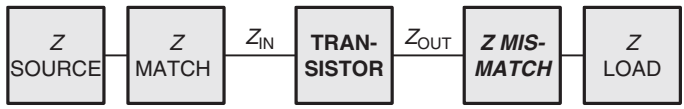


FIGURE 3.56 Selective output mismatching of an active device to lower stage gain.

3. Calculate "RATIO," which is the ratio between the transistor's *real* output impedance, $R_{Q(OUT)}$, and the matching network's input, $R_{IN(MATCH)}$ (to be calculated in the next step):

$$\text{RATIO} = \frac{1+\sqrt{1-\left(10^{\frac{-M_L}{10}}\right)}}{1-\sqrt{1-\left(10^{\frac{-M_L}{10}}\right)}}$$

4. Find the $R_{IN(MATCH)}$ ($R\upsilon$) of the matching network:

$$R\upsilon = R_{Q(OUT)} / \text{RATIO}$$

where $R_{Q(OUT)}$ = *real* part of the transistor's Z_{OUT}
$R\upsilon$ or $R_{IN(MATCH)}$ = *virtual* resistance at the matching network's input ($R\upsilon$ is used in calculations only, and is not a real circuit element)
RATIO = ratio between the transistor's real output impedance, $R_{Q(OUT)}$, and the matching network's input, $R_{IN(MATCH)}$

5. Cancel the reactance of the transistor's output by placing a reactance of the opposite value in series (Fig. 3.57; X_L). Now design the transistor's T matching network of Fig. 3.58 (L_1, L_2, C_1) to cancel all reactances in the load, but designed as if the transistor's true output impedance was now the new value of $R\upsilon$.
6. Remove $R\upsilon$ from the design (it is only used for the initial calculations). Combine all series reactances.

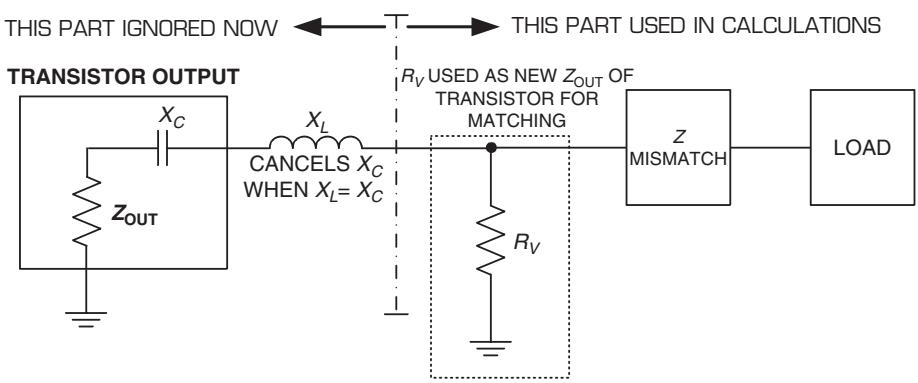


FIGURE 3.57 Using a virtual resistor (R_v) as the active device's temporary output resistance.

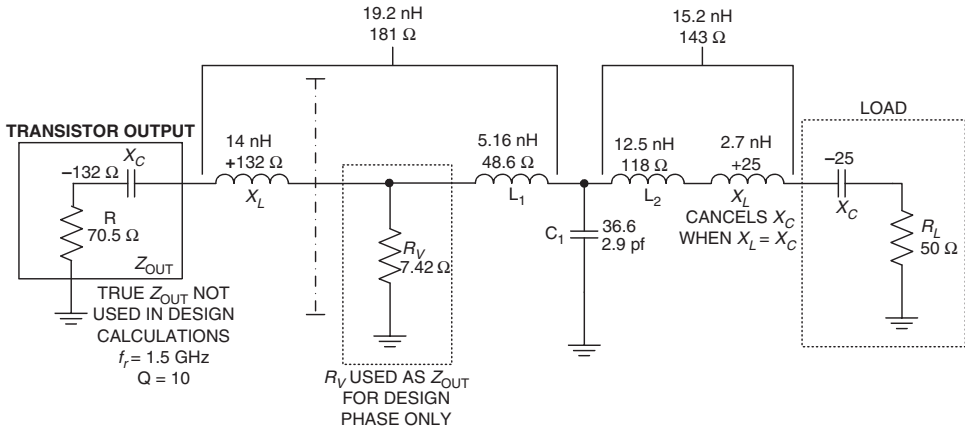


Figure 3.58 Combining the calculated values for a T matching network for selective mismatching.

7. An impedance mismatch is now formed, creating a drop in amplifier gain. This is due to mismatch losses caused by designing the transistor's output T matching network, as if the transistor had an output impedance of *R*υ instead of its true value. The completed mismatched output impedance amplifier is as shown in Fig. 3.59.
8. When using an active device that has high reverse isolation, design the input matching network for the transistor normally. (Use active devices that have a low S_{12} gain, and decrease the transistor's stage gain to a maximum of 25%. Some tuning will be involved due to the fact that no transistor has $S_{12} = 0$).

As an illustration, follow Figs. 3.56 to 3.59. Design a transistor amplifier with a gain of 6 dB at 1.5 GHz with the following device *S*-parameters:

V_{CE} = 10 V; I_C = 6 mA; $S_{11} = 0.195\angle 167.6°$; $S_{22} = 0.508\angle -32°$; $S_{12} = 0.139\angle 61.2°$; $S_{21} = 2.5\angle 62.4°$; MAG = 10.63 dB; MSG = 12.55 dB

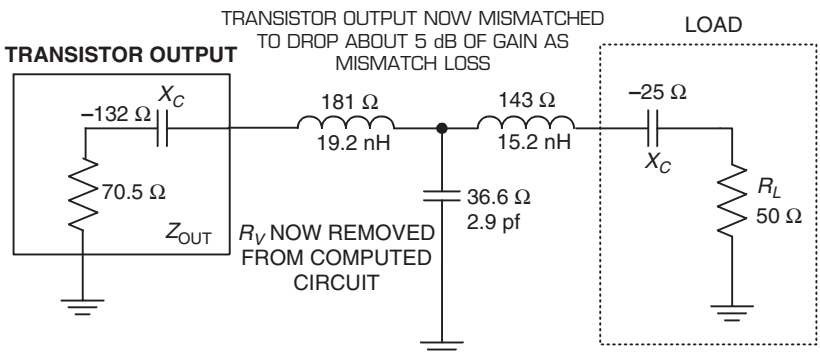


Figure 3.59 The completed mismatched amplifier with our desired gain.

1. Choose $G_{DESIRED}$ = 6 dB (or whatever value you would like the gain to be).
2. Calculate $M_L = G_{MAX}$ [dB] – $G_{DESIRED}$ [dB] = 10.63 – 6 ≈ 4.63 dB
3. Calculate

$$\text{RATIO} = \frac{1+\sqrt{1-\left(10^{\frac{-4.63}{10}}\right)}}{1-\sqrt{1-\left(10^{\frac{-4.63}{10}}\right)}} = 9.5$$

4. Calculate the virtual resistance ($R\upsilon$) required at the input to the impedance matching network:

$$R\upsilon = R_{Q(OUT)}/\text{RATIO} = 70.5/9.5 = 7.42\ \Omega$$

where $R_{Q(OUT)}$ = real part (resistance) of the transistor's output

5. Design matching network to cancel all reactances at the *transistor's* output and the *load's* input, while employing $R\upsilon$ as the new Z_{OUT} of the transistor for the impedance matching network. Remove $R\upsilon$, and combine all series reactances.
6. Design *input* matching network for a conjugate match to the transistor's input. To minimize the transistor's output reflected impedance from modifying its input impedance, and thus the calculated input matching circuit values, this last step is completely valid only if S_{12} is low.

LNA Matching Design

The LNA is the most vital stage of a typical microwave receiver. To safeguard a system's SNR the LNA must not contribute excessive noise, it must have high gain to minimize the noise contributions of all other succeeding stages, and it must have a high IP3 to minimize distortions (and thus minimize the received channel interference and improve the system's dynamic range).

Since modern system requirements demand that an LNA meet multiple, complex, and very challenging and interdependent specifications of low NF, high gain, good return loss, low current draw, high IP3, total stability, low voltage supply, small physical size, low cost, and low parts count, LNA design must be performed with care. Following all of the design information presented below will permit such an advanced LNA to be created.

Locate a transistor with a low NF at the desired frequency. The minimum NF for an LNA can be obtained from any transistor by carefully choosing its source load and bias point. The optimum source load and bias point can be found by looking at the transistor's *source resistance (R_S) versus collector current (I_C)* and *I_C versus NF* charts. Both charts may be available, at a limited number of frequencies, on the active device's data sheet. The optimal combination of *R_S versus I_C* for certain frequencies can also be found on a small Smith chart that may be printed on the transistor's sheet. The important LNA specification of $\Gamma_{S(opt)}$, the optimum source reflection coefficient for the lowest possible NF, can also be found on low noise transistor's data sheets.

Now, design a bias network to give the selected I_C, and match for the optimal source impedance of *Rs* at the transistor's input (instead of at the conjugate source impedance

value). In other words, if we match the low-noise active device to $\Gamma_{S(opt)}$ by forcing the LNA to "see," for instance, the antenna's 50-Ω output impedance as this particular desired $\Gamma_{S(opt)}$ value, we can make the LNA transistor supply its lowest possible NF (but at an increased VSWR over the perfect conjugate source match). We then terminate the transistor's output by:

$$\Gamma_L{}^* = \left[S_{22} + \frac{S_{12}\, S_{21}\, \Gamma_{S(opt)}}{1 - (S_{11}\Gamma_{S(opt)})} \right]^{\Psi}$$

where Γ_L = load reflection coefficient
$\Gamma_{S(opt)}$ = optimum source reflection coefficient for the lowest NF (as found on the transistor's data sheet)
Ψ = take the complex conjugate of the result

However, $\Gamma_{S(opt)}$ will normally be close to the complex conjugate of the device's input impedance and will, at times, be so close to $S_{11}{}^{\Psi}$ as to be almost identical. This would be the optimal situation, since both NF and VSWR will then be optimized, and will be more common in GaAs FETs that are operated above 2 GHz (GaAs FETs can be employed in the microwave regions above 3 GHz if BJTs cannot furnish the necessary gain and noise performance).

When the $S_{11}{}^{\Psi}$, the perfect input impedance match for maximum power transfer, and $\Gamma_{S(opt)}$, the perfect NF input match, diverge excessively from each other, system performance may suffer with an increased mismatch loss. (An LNA's input impedance should be as close as possible to 50 Ω so as to properly integrate with any input filter, without degrading the filter's performance characteristics, such as ripple.) A method we can use to compromise between a good VSWR and our desired NF is to employ a small amount of emitter or source lead inductance. If used under 2.5 GHz, where it is most needed in both BJTs and FETs, $S_{11}{}^{\Psi}$ and Γ_{opt} will draw closer together, thus maintaining NF and improving VSWR, as well as enhancing the third-order intercept and gain compression points, while slightly decreasing gain and with a minimal impact on stage stability. The increased inductance need only be added in the form of slightly longer emitter or source leads that are used between the transistor's package pins and the PCB's groundplane (Fig. 3.60). The length of these emitter leads must be found empirically, or approximated with RF software packages that are using transistor models with the appropriate noise data.

Stability under 2.5 GHz will not be dramatically affected by this slight increase in source inductance, but at higher frequencies this may not be the case, and instability of the LNA may result. Indeed, excessively high source inductance will create amplifier instability at both high and low frequencies, so the best inductance value should be chosen carefully. A compromise between stability, NF, S_{11}, and gain must be reached with both BJT and FET wideband low-noise amplifiers, since inductance at the emitter or source will, as stated, lower gain and increase stability at low frequencies, but can actually increase gain and decrease stability at high frequencies.

---

*Use *full* vector algebra ($Z \angle \pm 0°$) in *S*-parameter calculations (e.g., $S_{11} = 0.35 \angle -45°$).

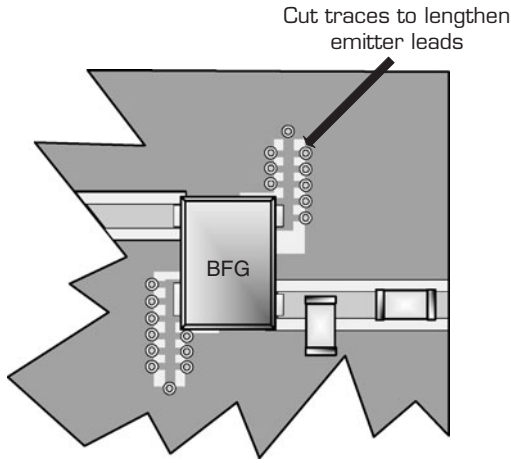


FIGURE 3.60 Adjustable trace lengths for an LNA's emitter leads.

If LNA stability is still a concern, adding a very low-value series resistor of 2 to 15 Ω between the transistor's collector and its output matching network will decrease stage gain, forcing LNA stability over a very wide band of frequencies. In fact, safety from oscillation in high gain amplifiers from VHF upward will normally demand some small value of LNA resistive collector loading, with higher values of the series resistor being more stable, but with smaller amplifier gain.

Thus, to recap a major LNA concept: Increasing LNA emitter lead length excessively, therefore adding too much emitter inductance, may severely degrade higher frequency LNA stability; there is, nonetheless, an ideal emitter lead length that will be a perfect compromise between the stability of the high frequencies, the in-band frequencies, and the lower frequencies. And if any of the lower frequencies are still shown to be unstable, we can try increasing the series collector stabilizing resistor, even though both gain and P1dB will be compromised whenever we improve stability resistively in this way. However, unconditional and wideband stability is by far the most important, and the most difficult to achieve, parameter in all of RF amplifier design.

As an example of BJT LNA design (also see Sec. 3.4.7), we find that the ideal bias point for the lowest NF for a certain low noise transistor is found to be $V_{CE} = 10$ V and $I_C = 6$ mA on the device's data sheet. Also found on the data sheet, we see that the optimal Γ_S for low-noise operation is $\Gamma_{S(opt)}$, and equals 0.65 $\angle 138°$ for this particular transistor for the frequency and bias of interest ($\Gamma_{S(opt)}$ may be referred to as Γ_o and Γ_{opt}). Along with $\Gamma_{S(opt)}$ we will find the G_a (the *associated gain* at the minimum NF, in dB) expected of the transistor at our desired operating frequency. The *S*-parameters of the transistor at the above bias conditions are also found to be $S_{11} = 0.35\angle 160°$; $S_{22} = 0.37 \angle -36°$; $S_{12} = 0.05 \angle 61°$; $S_{21} = 3.4 \angle 62°$. Therefore, design a stable LNA that can be placed within a 50-Ω receiver system at 500 MHz, and calculate the required impedance matching networks and gain.

1. Confirm that the transistor is unconditionally stable ($K > 1$). If not, stabilize with a small-value series resistor at its output port.

2. The $\Gamma_{S(opt)}$ for optimal NF, as stated on the data sheet, is $0.65\angle 138°$, which equals $-0.48 + j0.43$.
3. Find the input matching network's optimum NF match from the source to the transistor's input using this $\Gamma_{S(opt)}$ of step 2 (i.e., make the transistor's input "think" it is seeing $\Gamma_{S(opt)}$ as its source impedance).
4. Find

$$\Gamma_L^* = \left[S_{22} + \frac{S_{12}\, S_{21}\, \Gamma_{S(opt)}}{1 - S_{11}\, \Gamma_{S(opt)}} \right]^{\Psi}$$

where Γ_L = load reflection coefficient
$\Gamma_{S(opt)}$ = value as shown in step 2 above
Ψ = take the complex conjugate of the resultant

5. Match the transistor's output conjugately to the next stage, as presented earlier in this chapter.

NOTE: *An LNA can also be noise matched at its input port, and gain or IP3 matched at its output port. During simulation we can then tweak the stability, S_{11}, noise, and gain for the best compromise.*

When designing LNAs, we can also increase the collector current draw of the transistor to improve its vital IP3 (and gain) performance. This will obviously raise the stage's NF, as well as its current draw. Higher levels of collector-emitter voltage will also supply higher linearity, since there will now be more voltage swing obtainable at the LNA's output. This ability to improve the LNA's linearity is limited not only by the available supply voltage as set by the bias network, but also by the RF transistor's collector-emitter's limited breakdown voltage ($BV_{(BR)CEO}$). In fact, the $BV_{(BR)CEO}$ can be surprisingly low in some RF microwave transistors, and nearing or exceeding this point will start to increase the stage's NF, and may damage or destroy the transistor itself.

We can further improve the LNA's IP3 by up to 12 dB by attenuating the lower, or video, frequencies, which are the frequencies between DC and 40 MHz. This IP3 enhancement approach is sometimes referred to as the *envelope termination technique*. This will not improve a stage's P1dB, so the typical 10 dB difference between the IP3 and the P1dB specifications will no longer apply.

The video frequencies, if not removed, modulate the LNA's base and collector bias voltages, degrading the stage's IP3/linearity. The mixing products of the two injected test frequencies f_1 and f_2, which are used to perform the two-tone IP3 test, will be well within this video frequency range as $f_2 - f_1$, with the degradation of the stage's IP3/linearity caused by the fact that any decrease in the collector voltage of an amplifier forces it to saturate earlier, and since the bias *Q*-point of the stage will be shifted by a

---

∗Use *full* vector algebra ($Z \angle \pm 0°$) in *S*-parameter calculations (e.g., $S_{11} = 0.35 \angle -45°$).

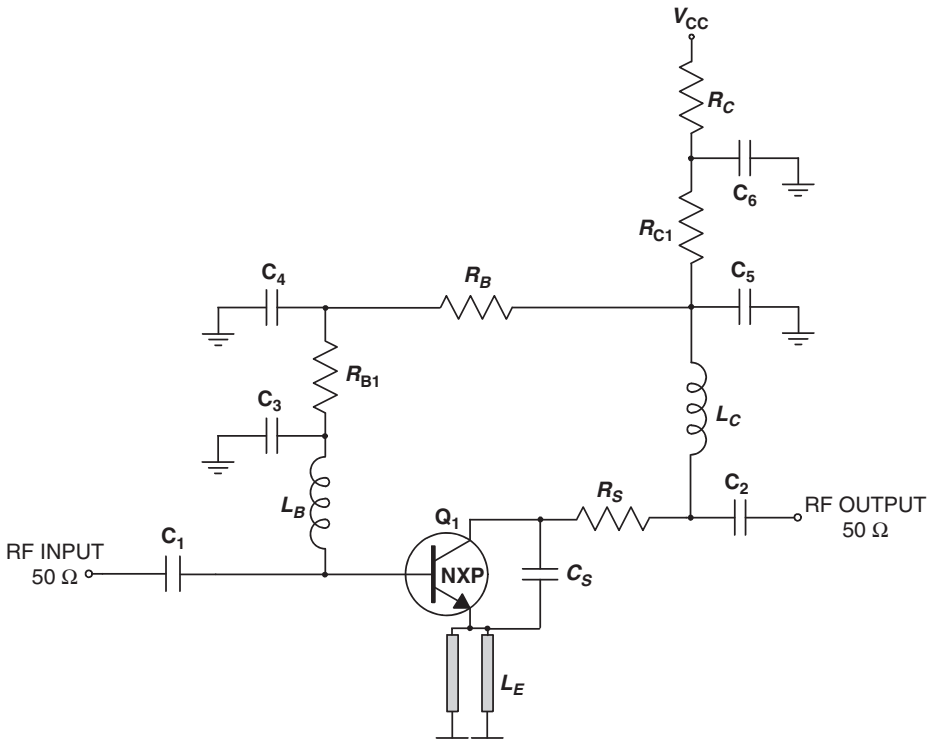


Figure 3.61 A complete LNA amplifier optimized for high IP3.

change in the base voltage, linearity is strongly affected by these low-frequency products of f_1 and f_2. Shunting $f_2 - f_1$ to ground by simply using high-value bypass capacitors (C_4 and C_6 of Fig. 3.61) will therefore significantly lessen any variations of the LNA's sensitive base and collector voltages.

Video frequency *base* bypassing has a large effect on improving IP3; *collector* bypassing a much smaller one. The shunting of the video frequencies becomes less successful with any amplifier that has a high collector current bias, which is why this technique is typically used only with low current amplifiers, such as LNAs. Also, there is a major caveat to this particular IP3 improving technique: When turning the *V*cc to the stage On/Off, we must always take into account the much longer time constant added to the switching time by these high-value capacitors, which obviously take time to charge up before the LNA will be able to operate properly.

There are further LNA design considerations that should be brought into play to optimize stage performance. As opposed to low-frequency amplifiers, an emitter bias resistor and its bypass capacitor are not used at these frequencies. This is because of the instability that can be created by the capacitor itself, and the increase in NF and the decrease in gain created by the emitter resistor if it is not fully and effectively bypassed across the entire passband. The capacitor and resistor also contain a certain amount of unavoidable series inductance (> 0.3 nH for SMD types), which may cause instability as well. For all of these reasons, there must be as little series emitter impedance as possible,

which demands that a direct and short emitter-to-ground connection take place between the transistor and the PCB.

So that we may insert a 50-Ω bandpass filter at the LNA's input port without degrading the BPF's ripple characteristics in receiver applications, it is important that an LNA's input impedance should be as close as possible to 50 Ω, as stated above. Therefore, we should design the LNA to possess an input return loss of approximately 10 dB, or better, even when matched for the amplifier's (almost) optimal NF. Normally, the amplifier's S_{22} will be significantly better than this, so adding an output BPF directly after an LNA will not be challenging.

If the LNA's collector-to-base DC bias resistor value of R_b is too low, the negative feedback contributed by this resistor can sometimes degrade the stage's entire NF, as well as stability. Thus, if the amplifier's bias circuit design demands a low value of R_b, then an RFC should be placed in series with the resistor to eliminate this problem.

When selecting a particular transistor to be used in an LNA design, we must keep in mind that the final implemented LNA will never supply quite the quoted NF or gain as promised in the device's data sheet. This is because of the unavoidable matching circuit and trace losses, the (usually) slight noise or gain mismatches, the use of consumer-grade PCB substrates, the bias current that is not exactly optimized for NF or gain, the use of stabilization loading resistors (either series or shunt), and the active and passive part-to-part tolerance variations of all circuit components. These real-life impediments will all negatively affect the quoted transistor's data sheet specifications.

When an LNA is biased for a minimum NF at 25ºC, a temperature-induced increase in collector current of 50% may increase the stage's NF by 1 dB. For instance, let's say we have just designed a 900-MHz LNA with a 25ºC NF, when biased at 10 mA, of 3 dB. If this collector current increases to 15 mA due to an ambient temperature rise, the LNA noise figure may then increase to 4 dB. If during this particular temperature change, however, we use a more complex and costly temperature-resistant bias network that is able to maintain the bias currents to much tighter tolerances, then both gain and NF of the LNA will be little affected.

The complete LNA stage, as shown in Fig. 3.61 above, employs all of the circuits and components as discussed in the above paragraphs: C_1 and C_2 are matching, blocking, and low-frequency gain rolloff capacitors; L_B is the input matching/blocking inductor, which can be used to assist in the input match, as well as block RF from being shunted to ground through C_4 (L_B has little affect on the low video frequencies); L_C is the output matching/blocking inductor used to perform a match of the output, as well as block RF from being shunted to ground through C_6 (L_C also has little effect on the low video frequencies); L_E is a printed microstrip inductance used to optimize the S_{11} and *gamma opt* points, as well as to force improved high-frequency stability and better IP3 (at the expense of some gain); R_S is a small-value resistor (< 15 Ω) used to stabilize the stage over a broad band, and helps to improve S_{22} (but will degrade *P1dB* and *IP3* somewhat; R_S is also less important for low-frequency stability when R_{C1} is used in circuit); C_3 and C_5 are the RF decoupling capacitors with an X_C of < 5 Ω; C_4 and C_6 are the actual IP3-improving video frequency decoupling capacitors used to shunt the low f_1 and f_2 mixing products to ground; R_B and R_C are the DC bias resistors and RF blocks; R_{B1} and R_{C1} are the 50-Ω resistors used to supply proper low frequency termination to significantly improve stage stability; and C_S is a low-value capacitor that is series resonant at some

undesired high frequency/high gain point, which is employed to improve stability at that specific frequency (C_S resonates with L_E, and should be utilized only if needed. It may also be exploited as part of the Z_{OUT} matching network).

3.4.7 Small-Signal Amplifier Class A Design Procedure

The complete step-by-step design method for a linear Class A amplifier is outlined below (Fig. 3.62).

To Design

1. Select the appropriate transistor for the desired frequency range, gain, NF, cost, package, and the availability of *S*-parameter files at different bias levels.
2. Select the appropriate bias for the transistor in your application by studying the transistor's data sheet. For instance, an LNA transistor may need a collector current of 5 mA, while a transistor for other uses may require a higher level of bias.
3. Insert the transistor's *S*-parameter model, which should be at the closest expected bias level of the device to be used, and check for stability all the way from the lowest to the highest frequencies available within the *S*-parameter file.
4. If there are any frequencies that have a K less than 1, or a B_1 below 0, then stabilize the transistor, while sacrificing gain and NF as little as possible. We may also want to check out other bias level *S*-parameter files for the same model transistor. (Figure 3.63 illustrates the exact same model transistor, but at three different bias levels of 5 mA, 10 mA, and 20 mA, and the resulting variations in K with collector current.) We may even exploit a transistor that has only conditional stability, but we would then need to guarantee that it would never see any impedance that

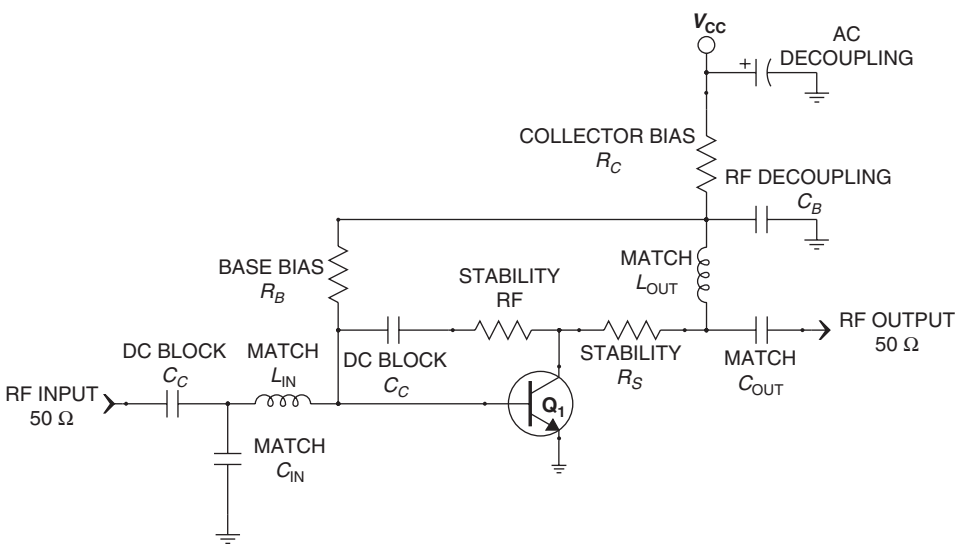


Figure 3.62 A complete small-signal Class A amplifier circuit, with RF matching, DC bias, and stability networks.

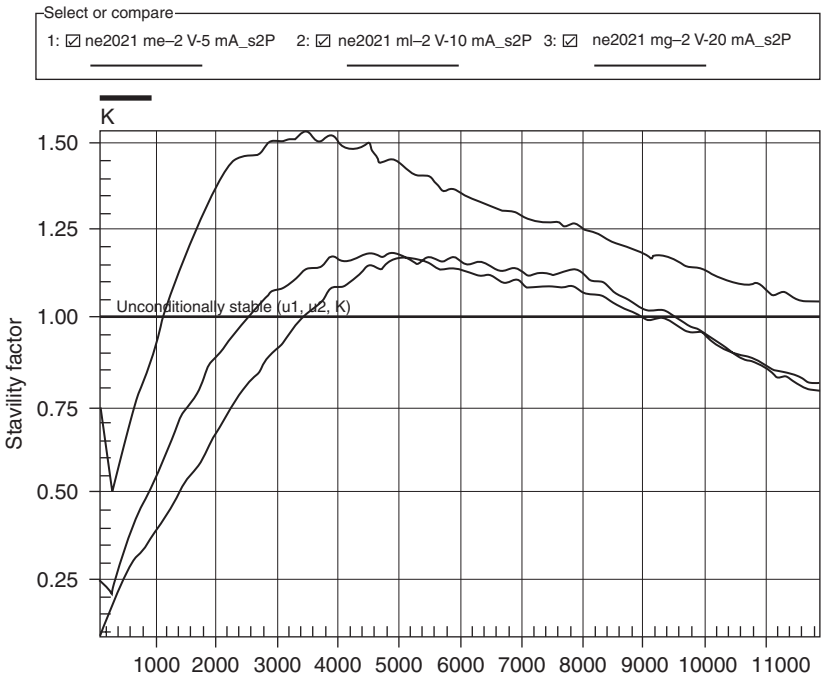


Figure 3.63 One example of the change in a transistor's stability with a change in bias level (stable above 1, unstable below 1).

was within its unstable region. (An amplifier can be considered as stable if $K > 1$, and both the input and output return losses are negative; however, also having a $B_1 > 0$ will then virtually guarantee complete stability.)

5. To stabilize the transistor, follow this procedure in the linear simulator: Insert and vary the value of a resistor that is first placed in series, and then in shunt, at the transistor's collector output port. Perform the same action at the device's input port (if NF does not matter), all the while checking for improvements in K across the widest frequency range available in the S-parameter file. This empirical manipulation of stabilizing components is the only method of improving the stability while minimizing any undesired gain reductions. To decrease these gain losses, the series resistor must be adjusted to the lowest possible value, while any shunt or feedback resistor must be optimized to be at the highest possible value. The resistor need only just barely force K and B_1, with about 10% stability margin, to indicate full wideband frequency amplifier stability. The shunt resistive components will typically range between 100 and 500 Ω, while the series resistors will extend from 2 to 15 Ω. Many active devices will be well stabilized with a single series output resistor, or a single output or input shunt resistor. In order to achieve complete wideband $K > 1$ and $B_1 > 0$ stability, other transistors may require a combination of methods. (As stated, for lowest noise, an LNA design will demand that there not be a series input resistor of any kind.)

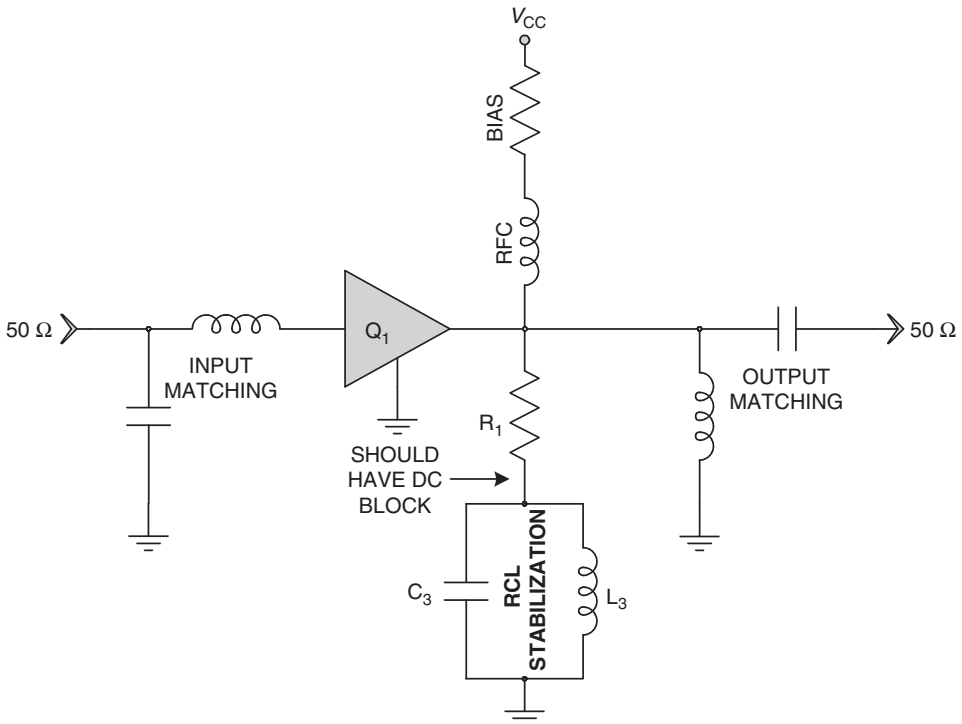


Figure 3.64 An RCL stabilization network.

There may be times when reactive stabilization circuits will meet our demanding requirements (Fig. 3.64), since they may be the only networks that can stabilize the transistor without unacceptable losses. The stabilization network's LC tank will be tuned to the amplifier's resonant frequency, with a high L/C ratio to maximize bandwidth, while the tank's series resistor will be tuned up and down between 30 and 100 Ω for optimum K/S_{21} performance.

Any shunt stabilization circuit, either resistive or reactive, must have a DC block inserted in that leg if it will short the amplifier's bias to ground to an unacceptable level.

6. We can now take the complete *S*-parameter file for the transistor, along with its stabilization components, and perform a simultaneous conjugate (or noise) match on the combination. As stated, we must include the actual stabilization components as part of the transistor during this matching, as these components would alter the match. Indeed, if these stabilization components must be modified in value to optimize K/S_{21} after the matching circuit is designed, then the input and output match itself must also be recalculated, or at least slightly retuned.
7. When the stabilization resistors are part of the amplifier circuit, the only way to quickly and easily perform such a simultaneous impedance match is by means of the included *Qucs* simulator program:

a. Open *Qucs*, locate and open the included project file "RF-TEMPLATE", then open the "RF-TEMPLATE.sch" file.
b. In the open RF-TEMPLATE.sch schematic, double click on the TRANSISTOR file block, and browse for and select your desired transistor *S*-parameter file.
c. Insert all required R or LC stabilization components.
d. In the schematic's workspace window, double click on the "S-Parameter Simulation" block, located just below and to the left of the circuit itself, and set its sweep frequency values to the minimum and maximum values that are available within your transistor's *S*-parameter file.
e. Run the simulation and go to the Smith chart that has S_{11} plotted. Place a Marker on the desired frequency. Now, select the Marker's display box with a single right mouse button click. In the menu that opens, select "*Two-Port Matching*". A "*Create Matching Circuit*" dialog will open. Select "*Create*". A matching network will immediately be positioned at the end of the mouse cursor. Press the left mouse button to drop this circuit into any open space of the schematic. Press your Esc key. Drag this new network into the stabilized *S*-parameter circuit.
f. Confirm that the stabilized match was created successfully by pressing F2 to simulate. Inspect all parameters to make sure that the amplifier's passband return loss and gain are as designed, and that there is no frequency within the wider frequency sweep that has a positive return loss, or a *K* of less than 1.

8. Design the DC bias network for a stable temperature operation of the transistor over its specified temperature range, and then confirm that the bias network has not modified the impedance match in any way. (If it has, investigate whether the collector bias resistor is of too low a value. If R_C is < 600 Ω, insert a series RFC in series to block any conducted RF.)
9. Add all the appropriate lumped and distributed component models, substrate, and microstrip to the circuit. Due to component and PCB parasitics, the amplifier may need to be extensively retuned for proper operation, especially when designed for elevated frequencies.

A Quick Example Design a Class-A RF Amplifier (Fig. 3.65)

Goal: Create a stable Class-A RF amplifier with collector feedback bias and 50-Ω matching. The complete specifications and parameters for the amplifier are:

$V_{CC} = 5$ V
$f_r = 2.4$ GHz
$NF = 2.8$ dB
$I_C = 10$ mA
$V_{CE} = 2$ V
$S_{21} = 19$ dB

Transistor = NEC NESG2021M05 ($H_{FE(LOW)} = 130$, $H_{FE(MID)} = 190$, $H_{FE(HIGH)} = 260$)

Solution:

1. For the DC bias circuit with collector feedback:
   a. $I_B = 53$ uA
   b. $R_b = 24.7$ kΩ
   c. $R_C = 300$ Ω

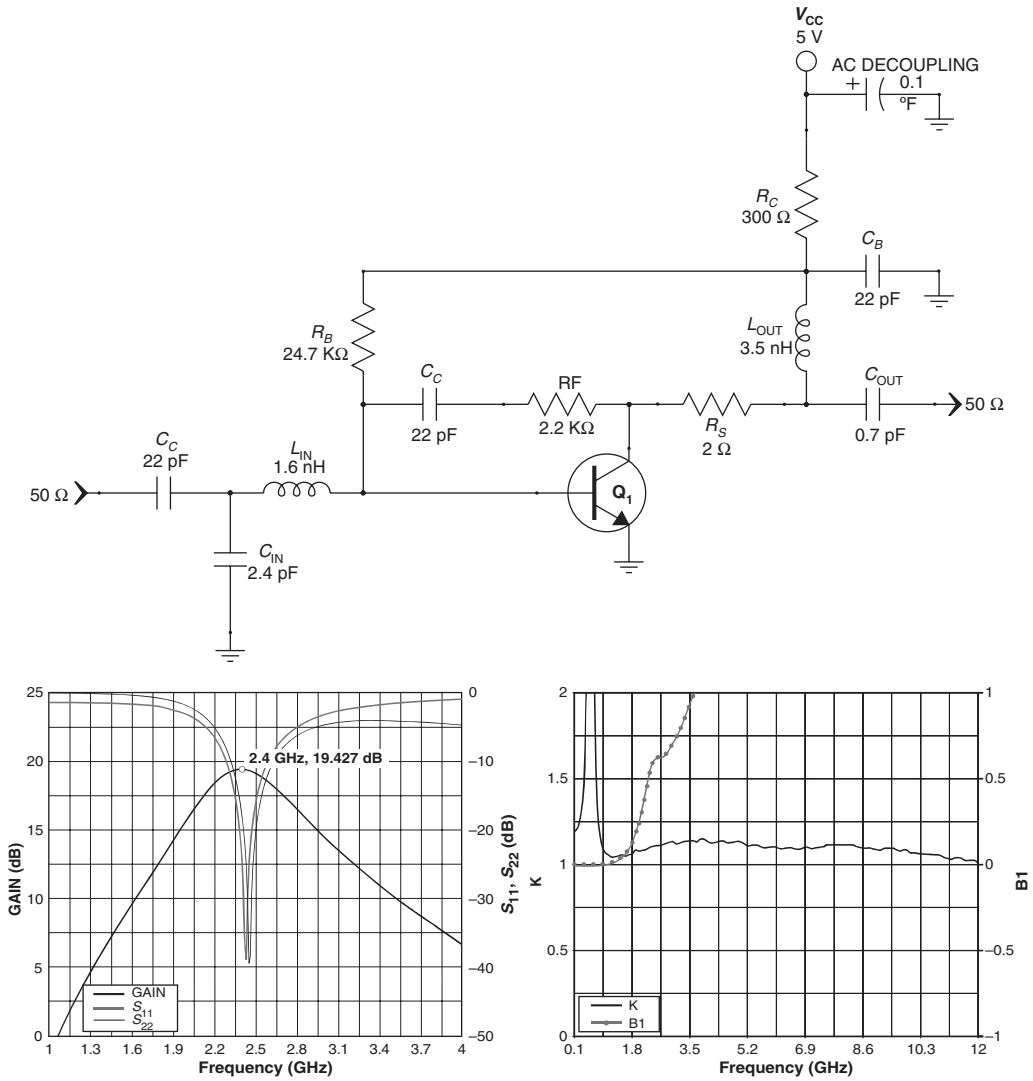


FIGURE 3.65 A complete worked example, with part's values and simulation results, of a Class A amplifier with bias, matching, and stabilization circuits.

2. For the amplifier's stabilization circuit (one solution):
   a. $R_S = 2\ \Omega$
   b. $R_f = 2.2\ \text{k}\Omega$
   c. $C_C = 22\ \text{pF}$
3. For the amplifier's RF matching circuit (one solution):
   a. C_C, $C_b = 22\ \text{pF}$
   b. $C_{in} = 2.4\ \text{pF}$
   c. $L_{in} = 1.6\ \text{nH}$
   d. $L_{out} = 3.5\ \text{nH}$
   e. $C_{out} = 0.7\ \text{pF}$

NOTE: *With this particular bias and a 5* V_{CC} *and a 2* V_{CE}*, the transistor's* I_C *will change by only +/–1 mA across a wide –25 to –65°C deviation, which will force only minimum changes in the amplifier's RF characteristics over temperature. Also, its* I_C *will be limited to no more than a further +/–1 mA change over the transistor's full* H_{FE} *production variations at 25°C. Due to the value of* R_C *being under 500* Ω*, an RFC must be added in series with the bias supply, unless the matching network will be using a collector matching inductor of sufficient value in series with* R_C.

3.5 Power Amplifier Design

3.5.1 Introduction

RF power amplifiers come in many common flavors, such as Class A, B, C, D, E, and F. Class A is the only amplifier type that is considered fully linear, and has the highest gain of any other class of large signal amplifier. Both the linearity and gain are due to the Class A amplifier's ability to fully amplify both the positive and negative alternations of an input signal. Class A can also effectively operate at higher frequencies than any other class (very near the f_{MAX} of the transistor itself) due to its lower level of output harmonics. Its overwhelmingly negative attribute is its very low efficiency, which can be critical in large-signal applications. Thus, Class A amplifiers are normally only used as PAs when the wireless system's modulation scheme will demand a very linear level of amplification, or if only low RF output powers are required. Nonetheless, in order to accommodate modern linear digital modulation schemes, many of the PAs designed today are Class A.

Class A power and small-signal amplifiers can be designed using *S*-parameters. Nonlinear, Class B and C power amplifiers cannot reliably exploit these parameters, but instead must depend mainly on *large signal input/output parameter* design. These large signal parameter values can be found in the power transistor's data sheet in rectangular notation (such as 1.1 – *j*3.2), and are available for both the transistor's input and output series impedances at a number of frequencies, V_{CC}'s, and P_{OUT}'s. The large signal impedances can also be made available as a Smith chart representation, or in the much easier to read tabular format, on the part's data sheet. If so desired, the series impedance values can be converted to a shunt impedance (as presented earlier in this chapter), if this information is not so provided. When provided, these numbers may be in the form of a separate graph of input and output parallel equivalent *resistance versus frequency*, and a graph of input and output parallel equivalent *capacitance versus frequency*, at a set V_{CC} and P_{OUT}.

Designing Class C power amplifiers with small-signal *S*-parameters will result in a circuit that is not optimized, and will not function as intended. This is because any transistor's input/output impedances, capacitances, and gain will be significantly different when run as a large-signal amplifier, as opposed to a small-signal, Class A stage. (Sometimes *S*-parameters were seen in power metal-oxide-silicon field-effect transistors (MOSFET) data sheets, but were only meant to be utilized in order to *approximate* a beginning design, which must then be tweaked in software, then hardware.)

Impedance matching, especially in RF power amplifiers, is required so that the transfer of energy to the next stage is accomplished with as little wasted power as possible. Matching (and biasing) permits any active device to perform as desired,

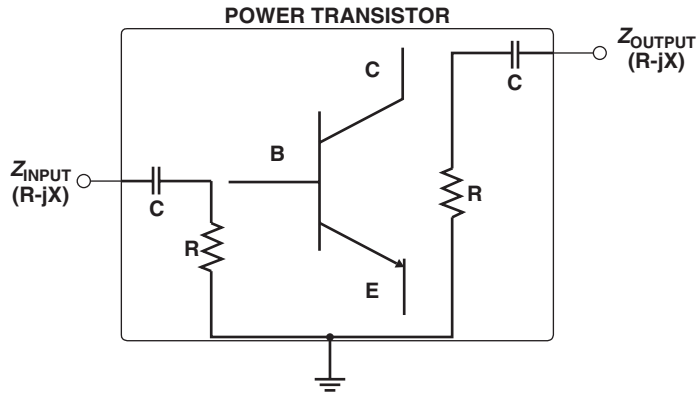


Figure 3.66 The complex input and output series impedance of a power transistor.

optimizing the output power, gain, return loss, NF, distortion, stability, and prevents ripples in the passbands of adjacent filters.

Since the input and output impedances of a power transistor are complex impedances (Fig. 3.66), and this impedance can be at very low values, which can necessitate an impedance transformation ratio of up to 20 times for a BJT, it can be seen that the efficient matching of a power amplifier to its source and load is anything but trivial. And the higher the desired output power of the large-signal transistor, the lower will be its output impedance. In fact, the design of power amplifiers of just a few years ago involved much trial and error, accomplished solely with physical bench tweaking of its matching and bias networks to obtain an efficient and workable amplifier, and one that possessed viable component values, and that did not self-destruct on start-up. To a lesser extent, this is still the case, but bench tuning is now used mainly to fine tune the PA for a low VSWR, high gain, maximum efficiency, and greatest output power after full nonlinear simulations have been performed.

The transistor's input and output impedances will also decrease with an increase in frequency, which further complicates the design of a PA's matching networks, especially since these impedances can be as low as 0.5 Ω. Thus, when matching a discrete driver stage to its PA with maximum efficiency, we would normally want to implement a direct match from the true output impedance of the driver to the true input impedance of the PA, instead of first forming a 50-Ω match at the output of the driver, and then another 50-Ω match for the input of the power amplifier, as this would needlessly transform the impedances from low to high, and then back from high to low.

By selecting a transistor with a high collector voltage requirement, we can increase its output impedance over a transistor that operates at lower values of collector voltage, or:

$$Z_{OUT} = \frac{V_C^2}{2P_{OUT}}$$

where Z_{OUT} = output impedance of the transistor, Ω
V_C = DC voltage at the collector, V
P_{OUT} = transistor's output power, W

There is, of course, a practical limit to this high collector voltage concept due to the available on-board (PCB) voltages, as well as internal transistor design issues at microwave frequencies.

When selecting the proper power transistor for a design, certain factors and specifications must be taken into consideration. The most important of these are maximum rated power output P_{OUT}, V_{CC} range, packaging types, cost per device, saturated gain, frequency of operation, maximum power input $P_{IN(MAX)}$, recommended class(es) of operation, device ruggedness, ease of matching, wideband stability, thermal issues, and so on.

The gain of the power transistor at the frequency of operation must, of course, fit the specified requirements, but selecting an active device with an excessive f_T will result in a much more delicate device. This is because one major method for the transistor designer to increase the operational frequency of the part is by making it physically smaller. However, a smaller device lowers its maximum safe power dissipation levels.

Most power transistors will also be specifically characterized for different bias *Q*-points (Class A, Class AB, Class C, and so on). If used at another *Q*-point, the transistor's parameters, such as gain, impedance, and even device lifetime, will change. Indeed, power gain is ordinarily at its peak with Class A amplifiers, and begins dropping as the forward bias is decreased, with Class C having the lowest large-signal gain of any common amplifier type. This change in bias will also affect the transistor's tolerance to impedance mismatches, which will be the greatest for Class C biased amplifiers, decreasing as the device gets closer to Class A.

As most modern low and medium output power PAs must run at small supply voltages, current draw can become quite high, which demands chokes and inductors capable of handling these increased currents. Another concern with power amplifiers over small-signal types is that any high-*Q* circuits located at the amplifier's output can result in high circulating tank currents, causing high dissipative losses and lower amplifier efficiency. Unfortunately, in many applications this is in direct conflict with the need to attenuate output harmonics with a high-*Q* output network.

Choosing the *Q* of the matching networks to be either high or low will depend on whether the amplifier itself will be operated in a broadband application. If it is, then the *Q* should be as low as possible to pass as wide a band of frequencies as possible, while also enhancing the amplifier's stability. Stability should not be compromised if we do not permit the matching network's loaded *Q* to exceed 5, even when designing for narrow bandwidths.

Indeed, power amplifier stability can become an almost impossible task if the transistor is operated significantly below its own power or frequency rating. This is caused by the increased gain over a safe, stable value when the transistor is not operated closer to its design specifications.

The physical PCB layout of power amplifiers must also be carefully watched. Excessively long emitter leads, which add undesired inductances, can cause severe degeneration, which will lower stage gain, or instability in higher-frequency applications. In Class C, common-base power amplifiers, the effects can be even more pronounced, and can rapidly lead to complete circuit instability.

Many power transistors today are protected against instant destruction, which can be caused by brief intervals of mismatch and instability, by modern fabrication techniques. This protection is important, since oscillations will create high peak voltages and collector currents, producing damage to an unprotected device.

3.5.2 Power Amplifier, Class C

Due to their use as highly efficient PAs and as frequency multipliers in nonlinear systems, we will delve a bit further into Class C amplifiers, and review some of the critical design issues not seen with most other types of transistor amplifier designs.

In order to survive Class C operation, the transistor should have a collector voltage breakdown that is at least three times the active device's own DC voltage supply. The reason: Class C amplifiers have low average output power (since the transistor conducts only for short, pulse-like periods), but demand very high input drive levels. Thus, the transistor's main Class C failure mode is the low value of the active device's own reverse breakdown voltage, which is unfortunately exacerbated by the RF input signal voltage going negative just as the transistor's collector voltage reaches its positive peak. This is especially problematic and dangerous if the load changes from design expectations, such as occurs if the system sustains a damaged or missing antenna or feedline during operation. In many cases this will cause a high voltage to form in the transistor collector's output tank, and any such device that is subjected to a higher than rated breakdown voltage may then start to conduct a very high current, destroying the transistor immediately, or damaging it so much that it suffers an immediate and permanent decrease in gain. Unfortunately, this particular BJT failure mode is very common when the transistor is in Class C bias.

One way we may lessen this voltage strain on a BJT is to decrease the chances of excessive reverse voltages from ever actually forming. We can do this by employing a low-*Q* inductor at the transistor's base and collector, or by using a ferrite bead in series with the base's normal or high-*Q* RFC, as opposed to using just a simple high-value resistor (Fig. 3.67). This low-*Q* circuit will suppress any high voltage ringing from occurring, which could possibly destroy the transistor. If the PA supplies under a few watts of RF output power, and its collector matching network starts with a shunt capacitor (such as in PI networks), none of this may be an issue.

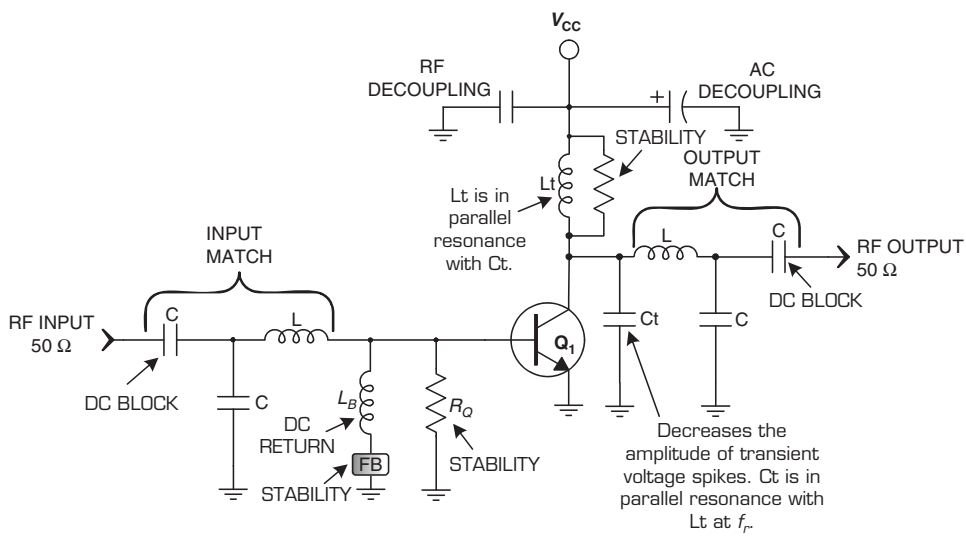


Figure 3.67 A Class C amplifier showing de-Qing ferrite bead at base of input inductor, as well as de-Qing resistors across all inductors (if needed).

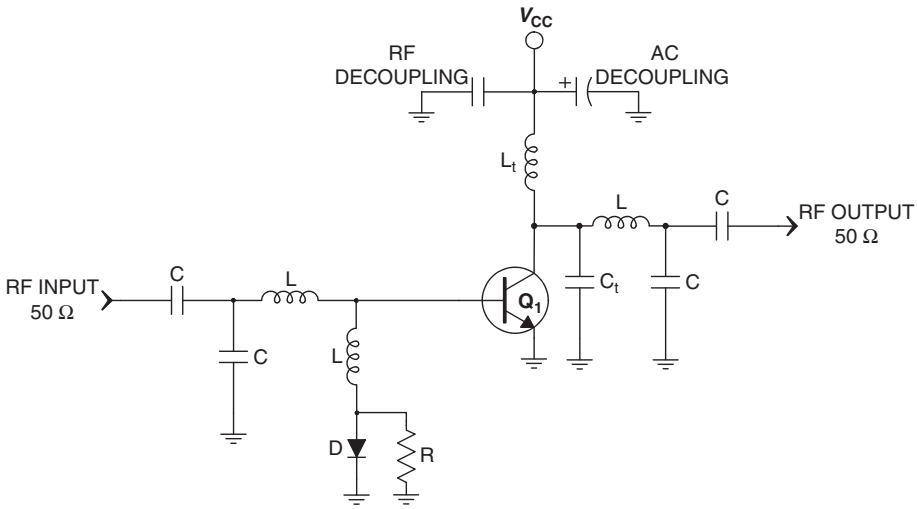


FIGURE 3.68 A Class C amplifier showing diode, which is used for stabilization and to suppress high spike voltages.

Another technique that can be used to safely employ a BJT in Class C operation is shunting the transistor's base with a diode, as in Fig. 3.68. This method restrains any spurious oscillations from forming across resistor *R*, while also eliminating any chance of transistor failure caused by excessively high voltages that are above the maximum permitted base voltage. The diode *D* prevents any voltage over and above that which the transistor can survive, and prevents excessive forward bias, and thus destruction, of the transistor. The inductor is used to shunt low frequencies to ground, acting as a filter.

R_{shunt} is normally used as the input component for many low-powered Class C amplifiers (Fig. 3.69), since it is very effective in suppressing oscillations caused by

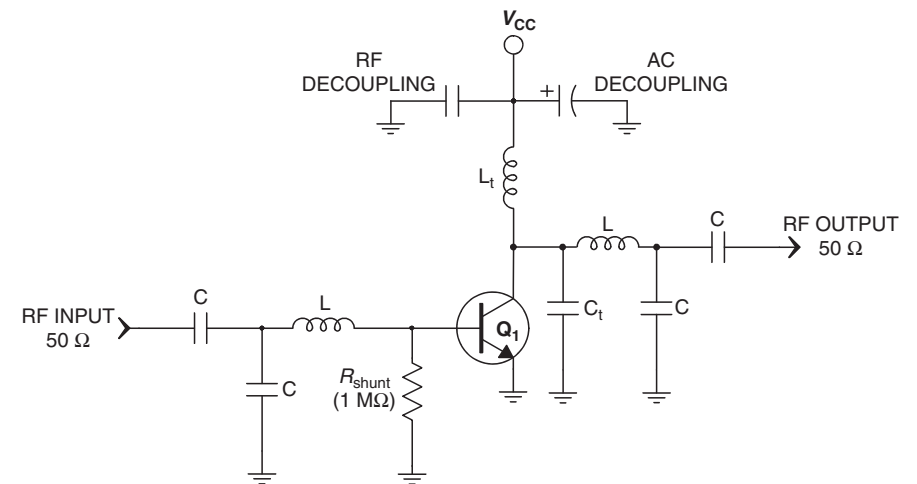


FIGURE 3.69 A Class C amplifier, showing the standard shunt input resistor topology used for stabilization.

variations in the amplifier's load, changes in input signal amplitude, or excessive frequency deviation. However, when the amplifier is being tuned, or if the load's output impedance varies, this simple Class C resistor shunt input can cause a breakdown of the transistor's collector-to-base diode, even when used in low power RF applications, which then strongly forward biases the transistor into destruction—or permanently reduces its performance.

3.5.3 Power Amplifier Design with Equivalent Impedances

The dominant procedure for high-power design of nonlinear amplifiers is the *large-signal series equivalent impedance* method, which characterizes a common-emitter Class AB or C power transistor's equivalent input and output impedances (Fig. 3.66 above). The large-signal series equivalent impedances can normally be found in the part's data sheet, and merely represents the active device's input impedance and its load's impedance at a specific frequency, RF power output, and bias condition in which the matched transistor will supply maximum stable gain. To clarify: These large-signal series equivalent impedances will specify the transistor's *own input impedance*, but *not* the device's own *output impedance*. Instead, these data sheet values indicate the load impedance that the amplifier *wants to see* for maximum gain and/or power.

In wideband power amplifier design, such a match would not guarantee that maximum efficiency would result, since the lower frequencies, where gain is naturally at its highest level, would be purposefully mismatched, while the higher frequencies will be matched to peak their gain.

Some power amplifier data sheets may also present the transistor's input and output impedances within separate graphs in parallel form, in which the device's *resistance versus frequency* and *capacitance versus frequency* are displayed at a specific V_{CC} and RF output power. This will simply indicate that the transistor should be visualized as having an internal resistance that is in parallel with its input capacitance, rather than in series as described above, at both its input and output ports.

Still other high-powered transistor data sheets may only supply the designer with an active device's parallel output capacitance. This will force the engineer to calculate the transistor's desired *optimum collector load resistance* that the transistor would like to see, at a particular RF output power and supply voltage, by using one of these basic formulas:

$$R_L = \frac{(V_{CC} - V_{SAT})^2}{2P} \qquad \text{or} \qquad R_L = \frac{(V_{CC})^2}{2P} \qquad \text{(with less accuracy)}$$

where R_L = required load resistance, Ω
V_{CC} = supply voltage, V
V_{SAT} = transistor's saturation voltage, V
P = RF output power level needed, W.

Armed with any of the above impedance information for the transistor, we can then conjugately match the transistor to its (normally 50 Ω) source and load, just as with the small-signal amplifiers earlier in this chapter. Due to real-life parasitic reactances inherent in all circuit components, pads, and traces, the PA will have to be bench tuned for optimal operation.

Even though it is transparent in the above power amplifier design information, it is still important to keep in mind that a large signal power transistor's output port is not truly *conjugately* matched to its load, but the port is instead presented with a specific load impedance, through the output matching network, that it *needs* to see in order to supply either maximum gain, maximum power, or maximum efficiency.

Matching for most high-powered amplifiers should normally consist of the T type, rather than the PI type, network. PI matching networks for high-powered amplifiers sometimes result in unrealistic component values when matching for the higher operating frequencies encountered into a 50-Ω load. Indeed, T networks are capable of much higher-frequency operation before this becomes a major problem. Both T and PI networks can be used, however, if the output impedance of the transistor is higher than its load, or the power output of the amplifier is under 15 W.

3.5.4 Power Amplifier Design Issues

When matching a discrete driver stage to its PA, we would normally want to implement a direct impedance match from the true output impedance of the driver to the true input impedance of the PA, instead of first forming a 50-Ω match at the output of the driver and then another 50-Ω matching network for the input of the PA. This would needlessly transform the impedances from low to high, and then back from high to low, decreasing efficiency through the already low port impedances involved at each of the nonlinear transistors.

When utilizing two identical power transistors as both our driver and PA, which can be a common situation in many sub-1-W RF amplifier designs, we can bias the PA stage at a higher collector current than the driver stage. This will permit the PA to handle the higher RF power levels, while we are then using only a single active device model number in our bill of materials.

Reflected power caused by a high VSWR condition between a PA and its load does not, in and of itself, cause a transistor's destruction or damage. Rather, a PA can be damaged or destroyed in a high VSWR environment simply because it is now looking at a completely *different* load impedance than it was designed for. This can cause damaging instability, excessive current draw, low efficiency, and decreased output power. High device power dissipation can then produce elevated heating of the transistor and/or excessively high voltages. These are the effects that can trigger transistor output stage damage, not the reflected power from the load itself destroying the output of the PA.

Another critical reason why we want low VSWR at a PA's output is to present a suitable input termination for the final output filter, which will preserve the filter's desired frequency response.

Heavy dissipation losses can occur in the finite Q of the inductors and capacitors at a PA's output, which decreases its RF output power and power added efficiency (PAE). This is because some of the RF energy produced by the transistor is given up as heat in the output matching circuit, with these resistive losses being as high as 1 dB for any PA in our power range of interest (under 36 dBm). Therefore, we should employ the lowest loss output network possible, using only high-Q components. High-Q components also naturally aid the inductors and capacitors in surviving the high DC and RF power levels that they may be subjected to, as there will be much less of an internal dissipative resistance to heat up the components themselves. Further, some of the DC bias voltage would be wasted if the resistance in the chokes were too high, since part of the power supply's voltage will be dropped across this undesired parasitic resistance, leaving less for the active device.

A similar issue, and still due to the high currents entailed in the operation of a PA, is to always utilize high-value, low-ESR electrolytic capacitors at the PA's power supply, as this type of capacitor can immediately supply the needed current to the amplifier stage, and without pulling down the entire voltage supply during this critical transient turn-on time. This will prevent all other circuits that are attached to the same power supply from being adversely affected by the PA's high-current turn-on. As well, this high-value capacitor must be supplied with the *V*cc voltage at all times, which means that the capacitor must be placed before any On/Off switch. If not, then during the transient turn-on period, it too will need to charge up and draw current, just as the PA will, which will only make matters worse.

Amplifier collector chokes not only demand a very low series DC resistance as mentioned above, but also must supply a very high impedance to the RF. If this impedance is not high enough, then some of the valuable RF output power generated by the PA will be wasted. Additionally, some of the RF signal will now feed into the bias voltage supply lines, causing possible oscillations of not only the power amplifier, but also to any other active stage with RF gain in the system. Therefore, the RF impedance from the PA stage to its DC power supply is kept high by employing a series inductor between these two points, and with enough inductive reactance to effectively choke off a very wide band of frequencies. A ceramic shunt capacitor is also placed between this inductor and the power supply to "short out" any RF that may have made it past the RFC. However, this simple arrangement may cause instability at low (video) frequencies, since this RF inductor will become, essentially, just a low impedance piece of wire, no longer acting as a choke at decreased frequencies. This cause of instability can easily be cured by adding a 50-Ω resistor in series with the inductor (Fig. 3.70), which forces the amplifier to always see a solid 50-Ω termination, even at the lowest of frequencies.

To go into more depth on a PA's input and output impedances, because of the confusion it can cause, a microwave power transistor's output impedance is actually

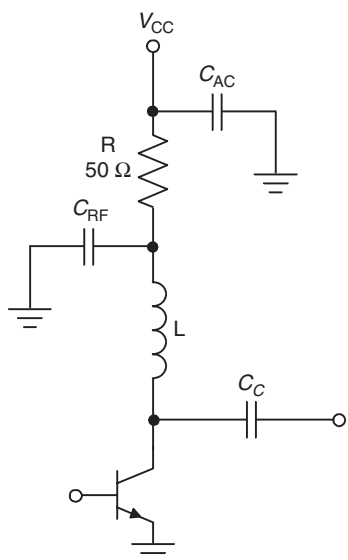


Figure 3.70 Forcing amplifier stability at low frequencies with a series stabilization resistor.

the conjugate of the load impedance that is necessary to realize the desired device performance within the PA circuit, and consequently the optimal load impedance for a PA is *not* the complex conjugate of its output, as it is in small-signal design. Indeed, the optimum load impedance (Z_{LOAD}) is the impedance of the load that the power amplifier would like to "see" to obtain optimum gain, output power, IMD, output match, or efficiency. Thus, we can readily see that the "output impedance" of a PA will rarely be a perfect conjugate match to the load, and that the output impedances as published in a power transistor's data sheet are normally the output impedances as specified for the active device, rather than the complex conjugate of the device's own impedances. In fact, these impedances are determined from the complex conjugate of the *load's* impedance that supplies the PA with the highest gain, while also maintaining complete PA stability, at a specific P_{OUT} and P_{IN}, frequency, collector current, nominal supply voltage, and class of operation (i.e., Class A, AB, or C). As a result, it is rarely advisable, or even possible, or even necessary, to measure the output impedance of a nonlinear PA directly. We may even damage the front end of our vector network analyzer during this particular process, since we must measure such nonlinear power amplifiers with the full input power applied in order to force the transistor into its true class of operation. The resultant high RF output power of the PA may instantly damage the VNA's front end during this "hot" measurement of return loss or impedance (the average VNA must not be hit with more than 1 W, or 30 dBm, of RF input power).

It should also be kept in mind that there is no way to maximize all of the important parameters of a PA design with a single "perfect" input/output matching network or ideal DC bias level. We can only perform the best compromise in our selection of which parameter we need to optimize, and efficiency, output power, power gain, return loss, or linearity must all be played against the other.

When in saturation, a nonlinear PA's gain, PAE, and linearity are most affected by the reflections of its own harmonics back into its output port. These reflections are caused by the next stage, which will normally be a bandpass or lowpass filter, as well as an antenna. Both the filter and the antenna will naturally be a poor match for the fundamental's harmonics, and therefore reflect them back to the PA. For this reason, designing a nonlinear PA that will be used with some form of output filter network, or a narrowband antenna, must be simulated and tuned together. The filter structure can drastically affect the initial tuning of the PA's input and output matching networks, as well as almost every other major amplifier parameter. This is because nonlinear PAs will always react somewhat differently when taken from the wideband 50-Ω terminated world of the RF simulator or load-pull station, and placed into an environment where the PA sees a relatively narrowband load, such as an output filter and/or antenna element. Indeed, a full RF nonlinear PA simulation must be performed using the exact filter design, with the exact same layout, as will be employed on the physical PCB of the completed wireless system.

After the PA/filter combination is actually built on the PCB, we are ready for bench testing to optimize the PA's gain, P_{OUT}, PAE, linearity, and stability. The more nonlinear the device, the more critical is this final step.

When we cannot afford to lose either PA efficiency or DC bias temperature stability, we may need to go to a discrete active bias method for low voltage, high current power amplifier applications, as described in Sec. 3.6.5 There are also fully integrated bias controllers available that will be able to operate at even lower voltages than the discrete active designs, and have less of a wasted internal voltage drop across them. For very low V_{CC} applications, they would be almost ideal, except for their increased cost.

However, the typical discrete passive collector feedback bias is normally sufficient for most low powered (less than 2 W) PA requirements.

3.5.5 Power Amplifier Load Pull

Load-pulling is a method that continually modifies the PA transistor's load impedance, during which we are continually measuring its performance characteristics. Source pulling, on the other hand, is used to continually vary the transistor's source impedance. In tuning of a design, the source input of a PA is usually presented with the impedance that will provide the amplifier with maximum gain, while the PA's output load impedance is varied around the Smith chart for optimal output power.

As stated, since the optimal load impedance for a PA is *not* the complex conjugate of its output, little good would come from directly measuring its Z_{OUT} with a VNA. What we need to do is discover the exact load that the PA would like to see at a specific input power level in order to supply maximum gain, output power, and efficiency. This can be accomplished on the bench by employing either manual or automatic load pull methods. Critically, a load pull setup will also assist us in confirming that a PA is capable of maintaining full stage stability over all input/output impedances, frequencies, and bias voltages.

Automatic load/source pulling equipment, referred to as *ATS* (*automated tuner system*), requires a full automated load-pull instrument, such as that sold by Maury Microwave. This complex device will automatically step through a huge number of input/output impedances across most of the Smith chart.

Manual load pulling involves a simple and low-cost *slide-stub tuner*, a signal generator, power supply, spectrum analyzer, VNA, and a bias-T. Performing a manual load-pull involves biasing the amplifier under test through the bias-Ts that are placed at both the transistor's input and output ports (Fig. 3.71). These slide-stub tuners are then connected to the output and input of the bias-Ts. Next, a 30-dB 50-Ω pad attenuator is placed on the stub tuner's output, which is then connected to the signal generator and spectrum analyzer. (The 30-dB pad is used to protect the front end of the spectrum analyzer against damage or overload, as well as to supply a precise broadband resistive 50-Ω termination for the tuner's output). Now, while injecting the RF signal at the exact required input power (this exact injected value of P_{IN} must be known in order to not only tune the PA for the proper power level match, but also to accurately calculate the power gain of the device), adjust the slide-stub tuners at the device's input and output ports while looking for the highest stable RF power output from the transistor, as well as its lowest current draw (i.e., highest efficiency). The input of the amplifier is normally presented with the impedance that will provide the PA with maximum gain, while the PA's output load impedance is varied around the Smith chart for a maximum performance characteristic, such as maximum output power. Remove the active device and measure the input and output impedance of the bias-Ts and stub tuners. Now, these exact impedance values must be duplicated with a discrete or distributed matching and bias network when designing the final PA.

A point to keep in mind when measuring a tuner's impedance after it has tuned the input/output of the PA, is that there can be a certain amount of inaccuracy introduced when the VNA must measure very low impedances, which can be typical of many high powered active devices. The VNA inaccuracy is due to the measurement uncertainty inherent in all such test and measurement equipment, and can be an issue if the active device's real impedances are under 2 Ω. Extreme inaccuracies will also occur as the wavelength shrinks at high frequencies versus circuit length.

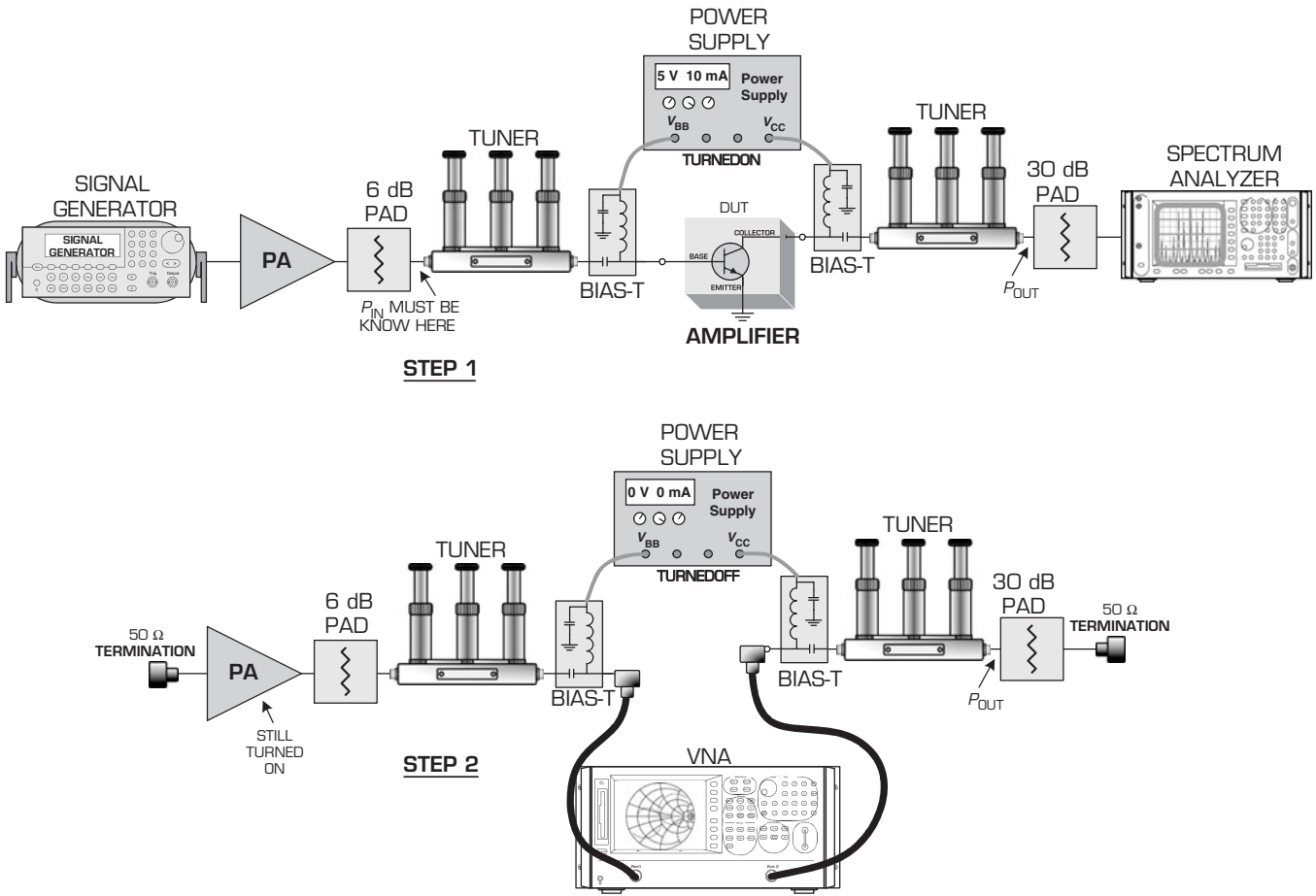


Figure 3.71 Manual load-pulling setup for an amplifier.

3.5.6 Power Amplifier Design Steps

We can design and tune an RF power amplifier (Fig. 3.72) with either of three methods: (1) Find the necessary PA circuit values by using a nonlinear RF simulator; (2) By empirically determining the PA's optimal design values through physical, on the bench, load-pull methods; or (3) By using a combination of software simulation and physical bench optimization.

We can further break down software simulation by whether we design the PA's matching circuits by making use of any of the transistor's specified input and output impedances on the part's data sheet (if available), or by employing software load-pull (if available in your nonlinear software package). Bench optimization/design may also follow two paths: by either employing physical manual/automatic load-pull equipment, or via empirical circuit component substitutions.

To Design

1. Select the proper active device that fits the requirements of frequency, gain, P1dB, V_{CC}, cost, package, availability, and so on. It should also have (accurate) nonlinear models available.
2. At the desired frequency of operation, look through the transistor's data sheet for the *output power versus Input Power* graph (or an *output power versus frequency* graph) to find the level of RF input power needed to drive the transistor to a specific RF output power. Or only the *power gain* (G_p) at a specific output power, V_{CE}, and frequency may be given, in which case subtract the G_P from the P_{OUT} value to obtain P_{IN}.

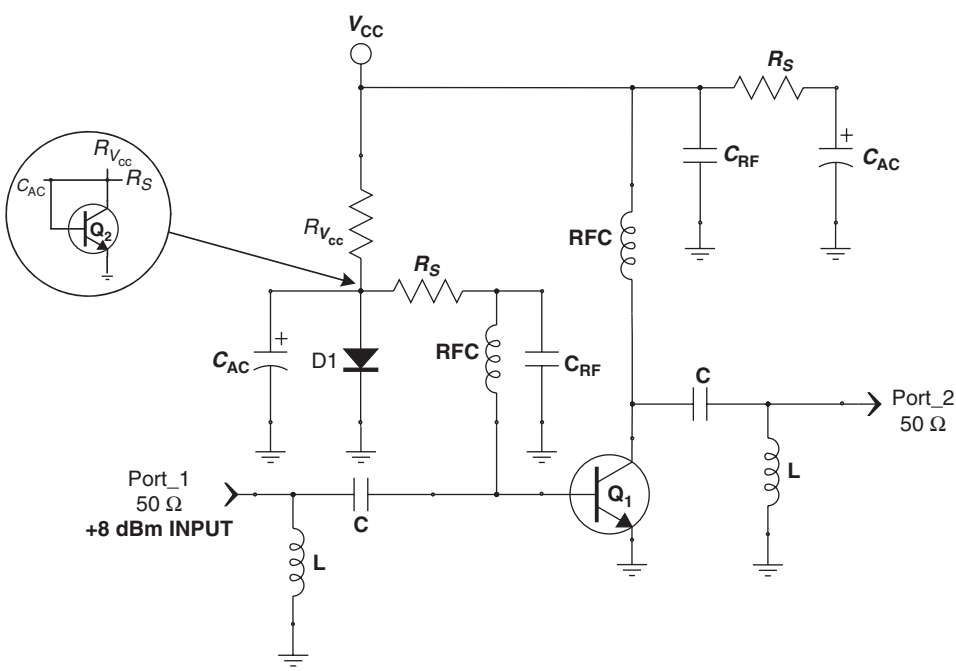


Figure 3.72 A complete Class AB power amplifier, with matching, active bias, and decoupling.

3. In the transistor's data sheet look for its *Series Equivalent Impedance*, either found on a Smith chart, graph, or table, to obtain the transistor's series input impedance Z_{IN} and its desired load impedance Z_{OUT} (Z_{OL}) at the desired frequency, P_{OUT}, and V_{CE}. This may also be specified simply as the transistor's *input impedance* and *output impedance* across a specific frequency range. Either way, this is normally, but not always, specifying the transistor's *own* input impedance (Z_{IN} or Z_i), but not its own output impedance. Instead, Z_{OL} (or Z_L) is indicating the *load impedance* that the PA's transistor *wants to see* for maximum gain or power.
4. Now, as for small-signal amplifiers, follow the same RF matching network design procedures while using the above Z_{IN} and Z_L impedance data to obtain a separate match at each port. Depending on the particular matching network selected, any matching inductors may be able to be used in place of the RFC bias chokes.
5. Design the bias network for the particular class of amplifier needed, as presented under Sec. 3.6.5. Normally this will be the diode or transistor Class AB bias topology. To instantly supply the high collector current required by the transistor as it pulses on, use a high-value electrolytic "storage" capacitor placed above the collector's RFC.
6. Use nonlinear simulation as recommended under the Sec. 11.1, *RF Simulator Design Software,* tuning the circuit for optimum power, gain, stability, efficiency, and input return loss. Now, replace all ideal components, and insert accurate passive models and PCB microstrip into the simulation, and retune.
7. Layout the printed circuit board as recommended under the *EMI Control and PCB Layout* in Chap. 13, and have the board fabricated and assembled.
8. On the bench, further tune the physical PA's matching and bias networks for optimal P_{OUT}, gain, PAE, linearity, and stability.
9. To check for proper PA stage stability over temperature and impedance, use a manual or automatic load-pull procedure to vary the PA's load impedances, as well as bias supply voltages, while the circuit is heated/cooled (see Sec.3.5.7).
10. Perform a final test of all appropriate PA parameters at low, mid, and high frequencies, as described under the *RF Testing* chapter. Record the amplifier's current draw, fundamental output power, harmonic power levels, saturated gain, stage stability, and all other desired parameters.

NOTE: *When the transmitter's PA stage is connected to a filter network and/or narrowband antenna, then the stability, IP3, PAE, as well as the level of the RF power output itself, may change drastically. This is caused by the nonlinear transistor's generated harmonics being reflected off the filter's and/or antenna's stopbands, and being returned, at various amplitudes and phases, back into the amplifier's output port. The more nonlinear the device, the stronger the harmonics, and the more the PA will be affected by this. Harmonics at the PA's input port will also have an influence. Thus, after an initial tune of the PA alone as a single stage, it is always wise to perform a complete PA tune with any filter inserted at the amplifier's output, in the exact PCB layout of the final and completed circuit. Further, due to a PA's output rarely being at a perfect 50* Ω*, any filter at its output will be strongly influenced by this poor termination, shifting the filter's amplitude and frequency response away from the as-expected design values.*

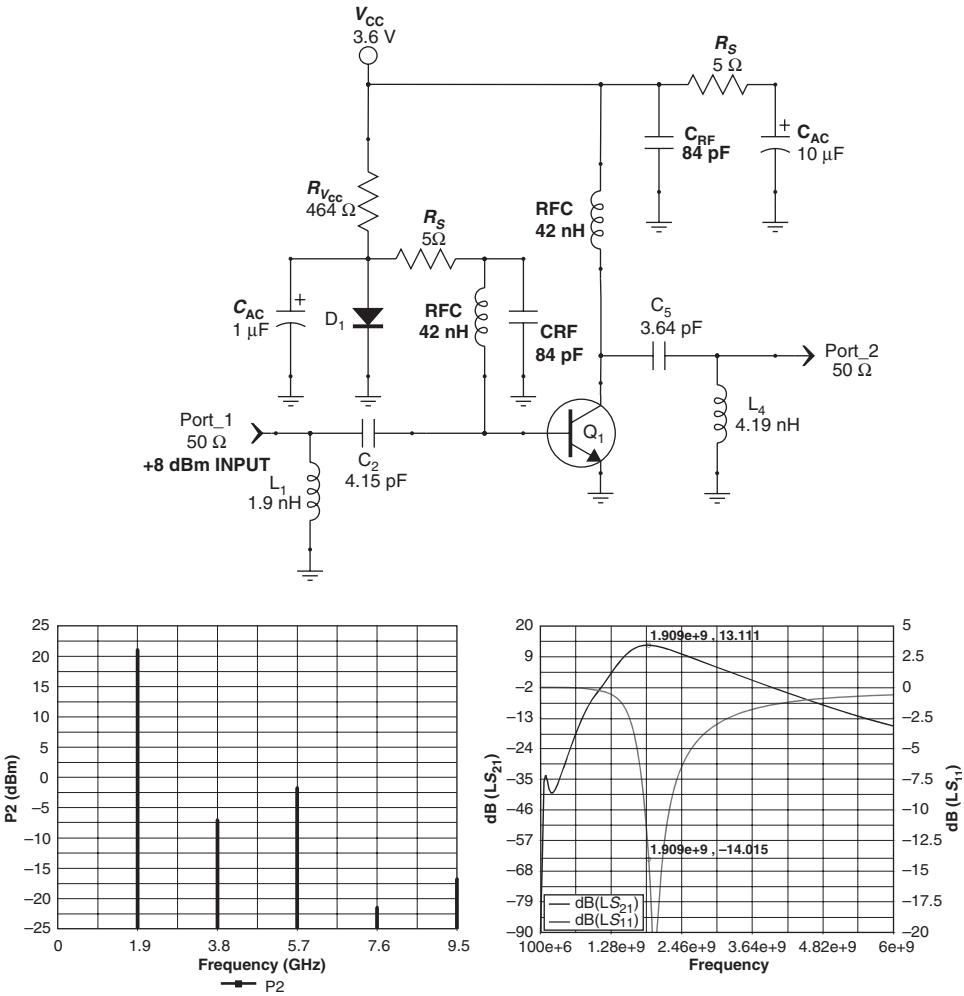


FIGURE 3.73 An example design of a complete Class AB power amplifier with active bias and matching, showing the Agilent Genesys nonlinear simulation of its RF output power and large signal S-parameters.

A Quick Example Design an RF Power Amplifier (Fig. 3.73)

Goal: Create a stable Class-AB RF power amplifier with diode bias and 50-Ω matching. The specifications and parameters for the circuit are:

$V_{CC} = 3.6$ V

$f_r = 1.9$ GHz

$Ga = 13$ dB

$P_{IN} = 8$ dBm

$P_{OUT} = > 21$ dBm

Transistor = NXP BFG480W ($I_{C(MAX)} = 250$ mA, $h_{FE(MIN)} = 40$)

One possible solution:

1. $R_{V(cc)} = 464\ \Omega$
2. D_1 = any small-signal diode capable of dissipating the maximum power required
3. $L_1 = 1.9$ nH, $C_2 = 4.15$ pF, $C_5 = 3.64$ pF, L4 = 4.19 nH (BFG-480W's data sheet states that it has an input impedance of $8.5 + j1.4$, and wants to see a load of $25 + j2$ when at 3.6 VCE, 1.9 GHz, 100 mW. The L_1/C_2 and C_5/L_4 input and output L networks are designed to satisfy these requirements.)
4. RFC = 42 nH, $C_{RF} = 84$ pF, $R_S = 5\ \Omega$

NOTE: *When PA is modeled and/or built with real, nonideal passive parts, all of the above 84 pF RF decoupling capacitors would be replaced by approximately 18 pF (for 0402 case size) values to take advantage of their* series resonance frequency *(SRF). This applies equally to the 42 nH RFC inductors for their parallel* self resonant frequency *(SRF).* R_S *is used to minimize undesirable parallel resonant interactions of the large value* C_{AC}*'s with the RF decoupling capacitors* C_{RF}*'s. Use a ferrite bead for improved performance.*

3.5.7 Power Amplifier Stability, Tests, and Cures

Instability in RF amplifiers can take the form of oscillations at almost any frequency, and may even damage or destroy the transistor. These spurious oscillations will arise at specific, or very wide ranging, frequency or frequencies, and over a particular bias, drive level, temperature, or output load impedance. Eliminating this instability will normally require a reduction of stage gain or output power through the complete retuning of the PA, or by the application of negative feedback—all of which will reduce the PAE. Indeed, a particularly poor RF layout may, in rare cases, demand a completely new PCB design and board fabrication due to electromagnetic, magnetic, or capacitive coupling from the PA's output port to its input port. This can cause positive feedback, resulting in oscillations (as can a transistor's own interelectrode capacitance feeding an in-phase signal back from its collector to its base).

A power amplifier, when properly biased and matched, must be stable across all output impedances that it expects to see, especially if connected to an antenna that is directly interfacing with the outside world. Testing this type of stability is fairly rapid with the use of an automated load-pull station. However, if an automated load-pull station is not available, a quick and dirty method of impedance-based stability analysis of a low power (< 3 W) PA is by attaching the appropriate antenna to the amplifier's output port and, over various supply voltages and temperatures, place your hand near, around, and at various distances from the side and top of the antenna. If the amplifier does not oscillate during this procedure, as checked with a sniffer probe or a 20-dB coupler attached to a spectrum analyzer set for a wide sweep, then the stage can be considered sufficiently stable, unless component-to-component tolerances are excessively poor with succeeding PA builds.

Stability testing over temperature is critical, as the stage may only oscillate when thermally hot due to the typical changes in transistor parameters with temperature. This is referred to as *thermal feedback*.

Class C power amplifiers, too, must remain stable under any load or V_{CC}, since instability can destroy the transistor because of the increased collector currents and high voltages created while oscillating. One way to rapidly test Class C amplifier stability is to place its output port into a slide-stub tuner, and drive the stage's input at its design frequency. If, while varying the amplifier's RF input drive power level, the

slide-stub tuner's impedance values, and the V_{CC}'s amplitude, the amplifier remains stable under these diverse conditions, then there is a good likelihood that it will not oscillate under almost any adverse real-life situation that it may be subjected to.

Another oscillation mechanism in some strongly nonlinear saturated PAs, particularly Class C's, is caused by variations in the internal instantaneous collector capacitance of the power transistor versus its changing voltage levels. This particular mechanism is actually modifying the value of one of the transistor's own reactive elements over time, and can cause what is referred to as *parametric oscillations*. These parametric oscillations can normally be diagnosed rapidly, since they will most often be seen on a spectrum analyzer at frequencies that are at one-half or at one-third of the PA's own design frequency.

It may be necessary to de-Q a Class AB power amplifier's base and/or collector bias inductor with a parallel resistor to force its stability. This empirically derived resistor should be of the highest possible resistance value so as not to decrease the PA's performance needlessly. Still, these resistors may not be necessary, but since this form of self-oscillation is relatively common, it is prudent to leave a set of unstuffed pads available on the PCB for possible resistor placement in parallel with the RFCs.

Decreasing the low-frequency gain of a PA stage, which is naturally at an increased level, will assist in amplifier stability. This is discussed further under Sec. 3.5.8.

The proper RF grounding of the transistor's emitter leads will help in both maintaining gain and avoiding oscillations, since even the smallest amount of inductance in this path to ground can prove disastrous to a power transistor. In fact, even the naturally occurring parasitic inductances and capacitances in the passive elements that are used for the PA's biasing, coupling, and decoupling must be modeled during the software simulation phase to prevent unnecessary and expensive tweaking of the completed physical power amplifier.

When a PA is tested alone on the bench for stability it may not oscillate, but when placed in circuit on the system's PCB, and attached to its driver amplifier, it may. This effect is caused by the increased RF feedback paths on the system board, especially those contributed by the common DC bias lines and ground returns, as well as the complex interaction mechanisms between all stages and their components. The cure for these feedback paths is better general RF decoupling, by identifying and eliminating any parasitic effects created by paralleled RF decoupling capacitors (see *anti-resonance* discussion in Sec. 1.1.3), and by removing any component-to-component electromagnetic (EM) coupling issues by physically separating all "RF hot" parts by distance.

The unavailability of a sufficient groundplane, or a groundplane that is excessively segmented, can create uncontrollable instability in a PA. The cure is a new layout, with a large, dedicated groundplane that is not broken up by any traces, and that has a large number of vias from the top ground pour, where the PA itself is located, down to the next layer of the PCB, which is where the main groundplane should be located.

Indeed, any power amplifier must, especially in today's competitive market, not self-destruct if placed in a shorted or opened state. If the power transistor has the appropriate heat sinking, it will have a much stronger chance of withstanding very poor return losses caused by missing or shorted loads.

3.5.8 Power Amplifier Gain Flattening

All wideband RF power amplifiers should incorporate some type of compensation to maintain a flat gain to within 2 dB or better across their entire bandwidth. This is needed due to a transistor's inclination to possess a higher gain of 6 dB per octave at its lower frequencies than at its higher frequencies. The increased gain can cause low-frequency

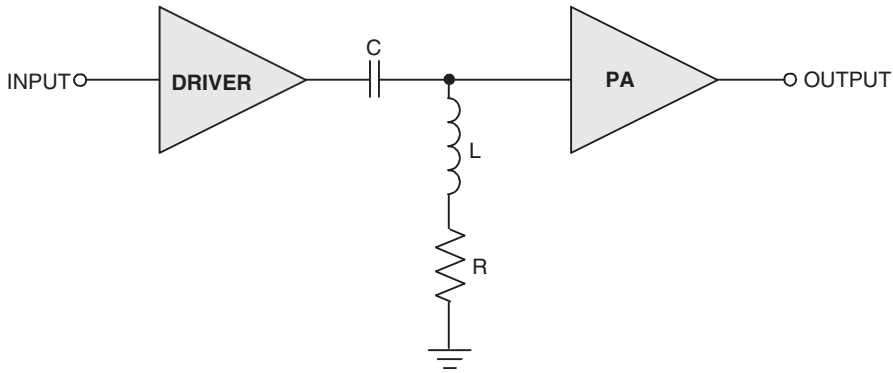


Figure 3.74 Gain flattening with an LR losser network.

instabilities, with possible transistor damage. By far the simplest method to suppress this excess gain is by adding a *losser network* (Fig. 3.74) between the PA's driver and the power amplifier itself. This will send the higher low-frequency power output to the resistor *R* at an almost perfect amplitude compensation value of 6 dB per octave, thus flattening the power amplifier's gain response.

The losser network is merely a highpass circuit with a built-in load that is designed to pass, without attenuation, the highest RF frequency of interest, and with a natural frequency rolloff that continues to flatten the lower-frequency gain. However, these low frequencies may begin to display an increasingly degraded return loss, so some empirical tweaking may be required in the RF simulation software, and on the lab bench, to optimize the values of *L*, *C*, and *R*.

Another common method is presented below, and employs a parallel bandpass resonant network also comprised of an *L*, *C*, and *R*. This circuit is tuned to the highest frequency of the passband, where the gain is low, in order to maintain its S_{21} in our band of interest, while decreasing the gain at the lower frequencies.

Frequency Compensation for Operation up to 1.6 GHz (Fig. 3.75)

To Design

1. Calculate an LC tank to resonate at exactly the highest frequency of the desired passband, and with a high L/C ratio for a low *Q* and a wide bandwidth. Start with the highest practical value of inductor that is still far below its SRF, and then calculate the required capacitor to meet tank resonance. Normally the widest obtainable passband for this circuit will be 25% of the center frequency.
2. Starting at a value of 50 Ω, adjust *R* up and down until the passband is as flat as possible, with little or no loss in amplitude at the tank's resonant frequency. If required, tune *L* lower in value, while increasing *C*, until all parameters are optimized. However, the tank's resonant point should never be moved away from the highest frequency in the desired passband.

A Quick Example Design an Amplifier Frequency Compensation Circuit (Fig. 3.76)

Goal: Create a lumped-frequency compensation circuit for an amplifier. The specifications and parameters for the circuit are:

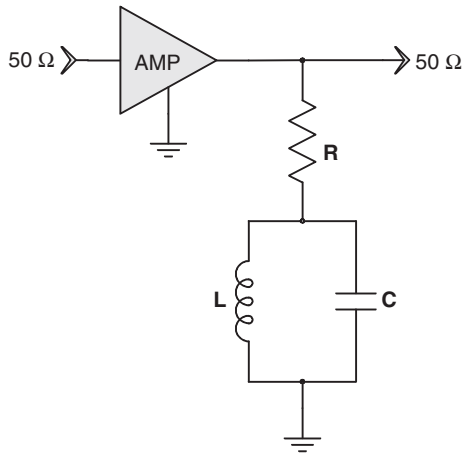
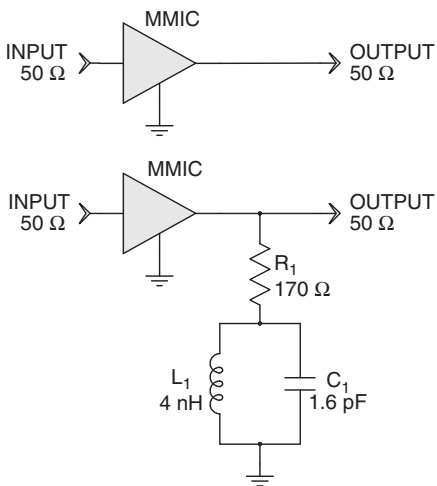


Figure 3.75 An amplifier circuit with LCR frequency compensation network for wideband gain flattening.



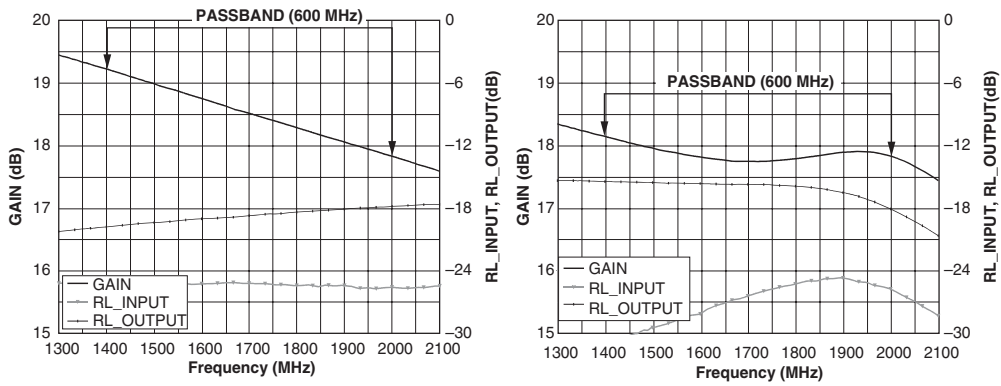


Figure 3.76 Example amplifier circuit with and without an LCR frequency compensation network, and the graphed simulation results: (*a*) before and (*b*) after passband responses. Gain has flattened from a 1.4 dB variation across the passband to 0.28 dB.

$f_{r(3\ dB)}$ = 1.4 to 2.0 GHz
Z_{IN} = 50 Ω
Z_{OUT} = 50 Ω
Passband flatness = 0.3 dB
MMIC = Mini circuit's ERA-50SM

Solution:

1. R_1 = 170 Ω
2. L_1 = 4 nH
4. C_1 = 1.6 pF

3.6 Amplifier Biasing

3.6.1 Introduction

A transistor amplifier must possess a DC biasing circuit for a couple of reasons. Foremost, we would require two separate voltage supplies to furnish the desired class of bias for both the emitter-collector and the emitter-base voltages. This is in fact still done in certain applications, but biasing was invented so that these separate voltages could be obtained from but a single supply. Secondly, transistors are remarkably temperature sensitive, inviting a condition called *thermal runaway*. Thermal runaway will rapidly destroy a bipolar transistor, as collector current quickly and uncontrollably increases to damaging levels as the temperature rises, unless the amplifier is temperature stabilized to nullify this effect.

3.6.2 Amplifier Bias Classes of Operation

Special classes of amplifier bias levels are utilized to achieve different objectives, each with its own distinct advantages and disadvantages. The most prevalent classes of bias operation are Class *A*, *AB*, *B*, and *C*. All of these classes use circuit components to bias the transistor at a different DC operating, or "*Q*", point (Fig. 3.77).

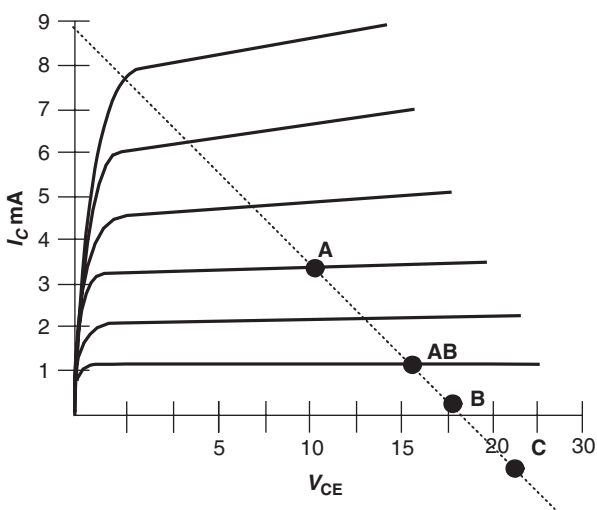


Figure 3.77 The locations of various bias *Q*-points for different amplifier classes.

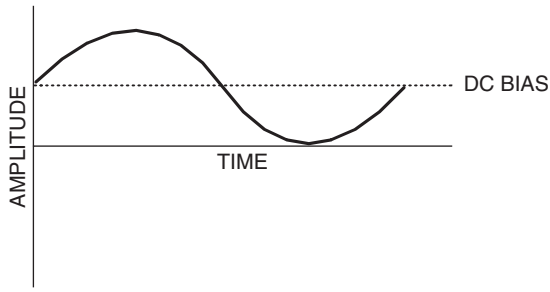


Figure 3.78 A Class A amplifier's output waveform.

As shown in Fig. 3.78, Class A bias permits a signal's amplified current to flow for the entire cycle, or 360°, of the input signal. This allows the amplified output signal to never reach saturation or cutoff, and thus stay within linear operating parameters. The output will be a relatively accurate amplified representation of the input signal.

Due to their low efficiency, Class A single-ended amplifiers are ordinarily used only in small-signal non-power applications, especially as low distortion linear RF and IF amplifiers. The lack of efficiency is caused by the large amount of continuous DC supply power required at all times, with or without any RF input signal present, to produce the constant current that is always flowing through this type of amplifier.

Simply by decreasing the *Q*-point of the amplifier by a small amount, Class AB operation can be reached (Fig. 3.79). This class of operation has a little higher efficiency than Class A, as the static output current (I_C) through the amplifier will be smaller, and will flow for something less than a complete cycle; or normally around 300° in power amplifier applications. It is a very popular RF power amplifier bias method. But any Class AB single-ended power amplifier will create more output distortion than a Class A type due the output clipping of the signal's waveform.

Class AB is also a common bias for push-pull audio power amplifiers, as well as linear RF *push-pull* power amplifiers.

Class B stage efficiency is very high. With no input signal, nearly zero power dissipation occurs within the amplifier. This is due to the almost complete absence of collector current flow, since the bias is just barely decreased to overcome the 0.6 V of the base-emitter junction. When any RF signal is placed at its input, the amplifier's output current will flow for approximately 180° of a full cycle (Fig. 3.80). This conduction will only occur when a half cycle of the RF signal actually forward biases the base, while the

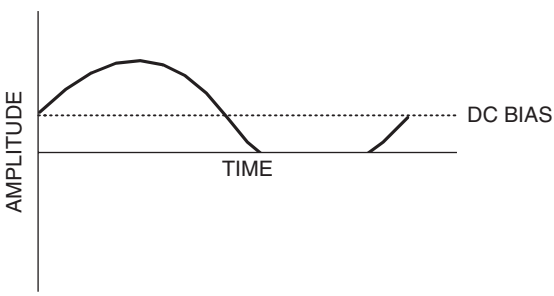


Figure 3.79 A Class AB amplifier's output waveform.

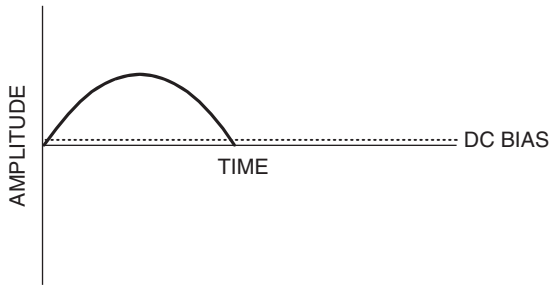


Figure 3.80 A Class B amplifier's output waveform.

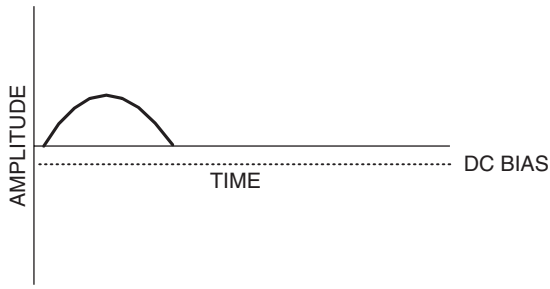


Figure 3.81 A Class C amplifier's output waveform.

other half cycle will reverse bias the emitter base, creating a complete lack of an output signal. However, considering that a Class B amplifier acts as a half-wave rectifier and amplifies only half of the incoming RF signal, it is normally found only in two-transistor push-pull power amplifier arrangements, but can also be used as a substitute for full Class C-biased single-stage PAs and multipliers.

Class C amplifiers are even more efficient than Class B bias, since they consume only a small leakage current when no RF input signal is present. When an input signal is inserted, a Class C stage will amplify for less than half of the input signal's cycle, and will really only supply a pulse at the output port. This conduction angle will be for 120° or less (Fig. 3.81), because the emitter-base junction is, in fact, slightly *reversed* biased. Many Class C schemes, however, may not use any bias at all, since silicon transistors, due to their 0.6 V emitter-base barrier voltage, will not conduct until this voltage is overcome by a high-level input signal. But, as a pulsed output is unusable for many wireless purposes, this pulse can be changed back into a sine wave by an output tuned tank circuit or filter network, which also naturally decreases the harmonic output level. With the tuned tank's flywheel effect reconstructing the missing alternation, the output of a Class C amplifier will also have a peak-to-peak voltage that is double that of the power supply.

RF Class C amplifiers are found in FM driver stages, single-ended PA stages, frequency mixers, and frequency multipliers.

3.6.3 Amplifier Bias Circuits

The most predominant biasing schemes used to obtain both temperature stabilization and single-supply operation are *base-biased emitter feedback*, *voltage-divider emitter feedback*, *collector-feedback*, *diode-feedback*, and *active-feedback* bias. All five are found in Class A and AB operation, while Class B and C amplifiers can implement other methods. Which

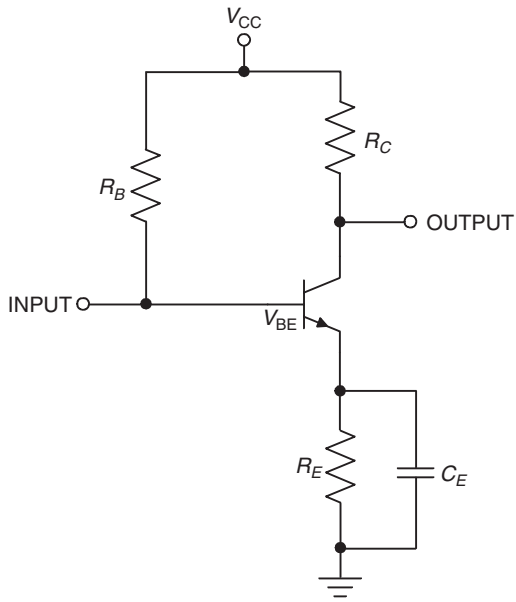


Figure 3.82 A C-E amplifier with base-biased emitter-feedback biasing.

bias circuit to adopt for a particular amplifier depends on the desired circuit costs, complexity, stability, and other considerations.

Base-biased emitter feedback (Fig. 3.82) works in the following way: The base resistor (R_b), the 0.7 V base-to-emitter voltage drop (V_{BE}), and the emitter resistor (R_E), are all in series; as well as being in parallel with the power supply (V_{CC}), as shown in Fig. 3.83.

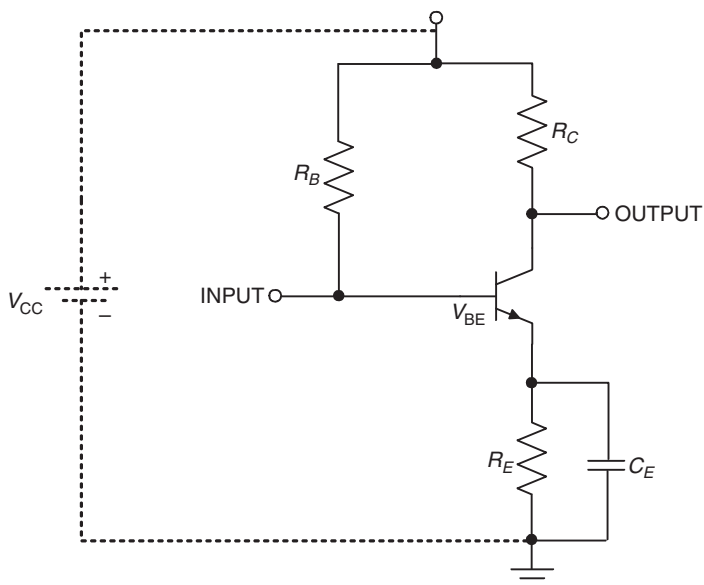


Figure 3.83 A C-E amplifier displaying its VCC connection.

As the collector current (I_C) increases due to a rise in the transistor's temperature, the current through the emitter resistor will also increase, which increases the voltage dropped across R_E. This action lowers the voltage that would normally be dropped across the base resistor and, since the voltage drops around a closed loop must always equal the voltage rises, this reduction in voltage across R_b decreases the base current, which then lowers the collector current. The capacitor (C_E) located across R_E bypasses the RF signal around the emitter resistor to stop excessive RF gain degeneration in this circuit. The higher the voltage across R_E (V_E), the more temperature stable the amplifier, but the more power will be wasted in R_E due to V_E^2/R_E, as well as the decreased AC signal gain if R_E is not bypassed by a low reactance capacitor. Standard values of V_E for most HF (*high frequency*, or amateur band) designs are between 2 to 4 V to stabilize ΔV_{BE}. However, UHF amplifiers and above will normally completely avoid these emitter resistors.

One voltage source is also supplying all the biasing required for the base-biased emitter-feedback circuit for proper operation of the negative-positive-negative (NPN) transistor, since R_b and R_C are accurately allocating the suitable voltages to both the collector and the base, with the appropriate polarity, through a single power supply. This is accomplished due to the following: the collector resistor, the collector-emitter junction, and the emitter resistor are all in series with each other, and share V_{CC}'s voltage. Thus, the collector-to-emitter voltage is equal to V_{CC}, minus the voltage drop across the collector and emitter resistors of R_C and R_E; forcing the collector to be correctly reversed biased. The base circuit is also properly forward biased by the following action: The base resistor, the emitter-base junction, and the emitter resistor are in all in series and share the V_{CC} power supply's voltage. So, the voltage drop across R_b will be equal to V_{CC} minus the normal emitter-base voltage drop of 0.7 V and the voltage drop across the emitter resistor. Since the voltage drop across the emitter base and the emitter resistor are kept relatively low, most of the power supply's voltage is dropped across R_b, properly forward biasing the transistor's base. In fact, this base current, and thus the collector current, can be increased by decreasing the value of this base resistor. However, because of the inclusion of the emitter resistor R_E and the emitter resistor's bypass capacitor C_E, and their small, but unavoidable, values of stray inductance, the base-biased emitter-feedback circuit is not normally employed in microwave amplifiers due to the associated gain reduction and possible instability problems caused by these reactances.

One of the more common of the low cost, low-frequency biasing schemes, and with a higher temperature stability than the above method, is the *voltage divider, emitter-feedback biasing* circuit of Fig. 3.84. This circuit is temperature stable because the current through the voltage divider of R_1 and R_2 is significantly higher than the base current, and any rise in the device's temperature, which will increase the base current, will not substantially vary the voltage across R_2, which is equal to the voltage at the base in respect to ground, thus maintaining a constant voltage from base to ground. As well, just as in the *base-biased emitter feedback* discussed above, when the emitter current rises with an increase in the transistor's junction temperature, the top of the emitter resistor will turn more positive. But as the base is always around 0.7 V *more* positive than the emitter itself, the base-emitter junction will now have an actual decrease in the voltage dropped across it when referenced to the common-emitter lead, reducing the I_C to its desired amplitude.

For sensitive amplifier applications, we can proceed even further to increase temperature stabilization. The most common method is *diode temperature compensation*, shown in Fig. 3.85. Two diodes, D_1 and D_2, attached to the transistor's heatsink or to

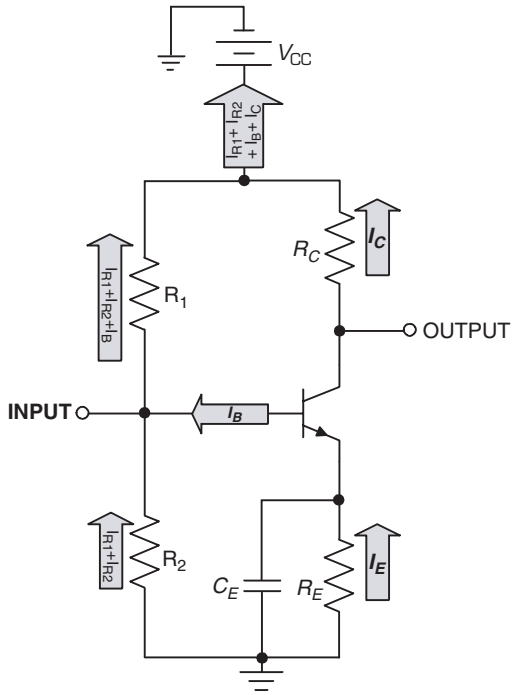


Figure 3.84 A voltage divider, emitter-feedback biased C-E amplifier with current flow.

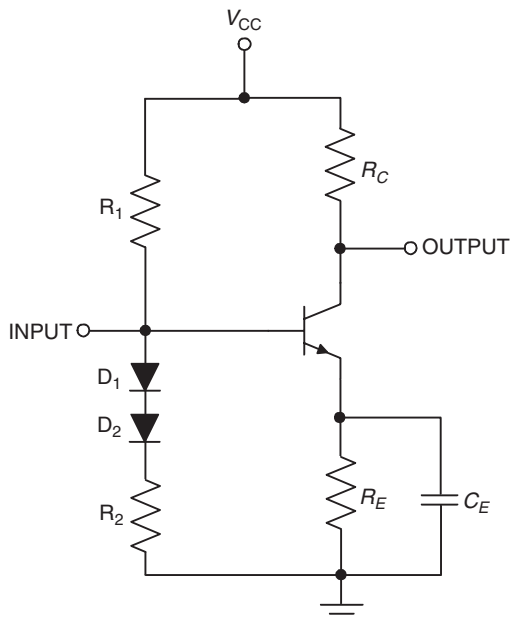


Figure 3.85 A diode temperature compensated C-E amplifier with voltage divider.

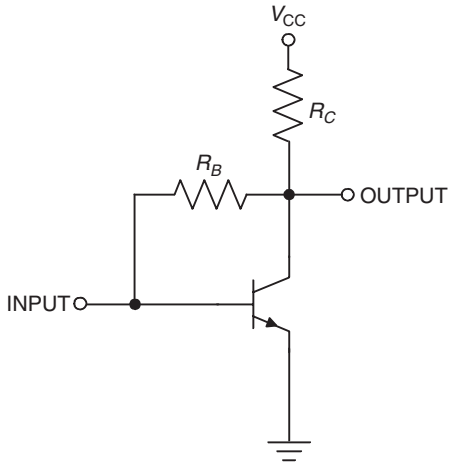


Figure 3.86 A common-emitter amplifier with collector-feedback bias.

the device itself, will carefully track the transistor's internal temperature changes. This is accomplished by the diode's own decrease in its internal resistance with any increase in heat, which reduces the diode's forward voltage drop, thus lowering the transistor's base-emitter voltage, and diminishing any temperature-induced current increase in the BJT. Employing only a single diode, or compensation by means of transistors or thermistors, may also be found in some amplifier temperature compensation circuits.

A prevalent and very low-cost biasing scheme for RF and microwave circuits, but with less thermal stability than above, is called *collector-feedback bias*. The circuit, as shown in Fig. 3.86, employs only two resistors, along with the active device, and has very little lead inductance due to the emitter's direct connection to ground. Collector-feedback bias temperature stabilization functions so: As the temperature increases, the transistor will start to conduct more current from the emitter to the collector. But the base resistor is directly connected to the transistor's collector, and not to the top of the collector resistor as in the above biasing techniques, so any rise in I_C permits more voltage to be dropped across the collector resistor. This forces less voltage to be dropped across the base resistor, which decreases the base current and, consequently, I_C.

The next method, active bias, is a common way to bias microwave power amplifiers, and the full discussion on this circuit can be found under Sec. 3.6.5, *Amplifier Bias Design*.

Common-source FETs can utilize a common Class A biasing technique called *source bias*, a form of self bias (Fig. 3.87). With field effect transistors, unlike bipolar junction transistors, no gate current will flow with an input signal present, so the drain current will always be equal to the source current. However, source current does flow through the source resistor R_S, creating a positive voltage at the top of this resistor. Now, since the FET's source is shared by both the drain and the gate circuits, and the gate will always be at 0 V with respect to ground (since no gate current equals no voltage drop across R_G), then the gate is now *negative* with respect to the common source. This allows

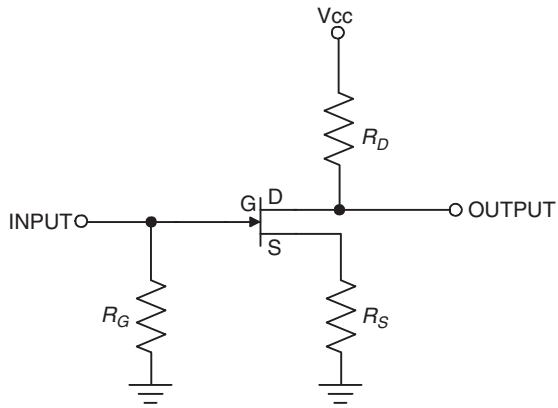


Figure 3.87 Class A source-biased FET amplifier.

the FET to be biased at its Class A, AB, or B *Q*-points, depending on the value chosen for R_S, while a capacitor can be inserted across R_S in order to restrain the bias voltage to a steady DC value.

For bipolar transistors, Class C amplifiers permit the use of three biasing techniques: *signal, external,* and *self bias*. Nonetheless, the average Class C transistor amplifier is normally not given any bias at the base whatsoever (Fig. 3.88), but in order to lower the chances of any BJT power device instability the base should be grounded through a low-*Q* choke, with a ferrite bead on the base lead's grounded end (Fig. 3.89). These biasing techniques will still require an RF signal with a high enough amplitude to overcome the reverse (or complete lack of) bias at the Class C's input.

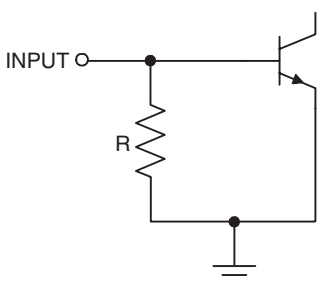


Figure 3.88 A Class C BJT amplifier showing lack of bias.

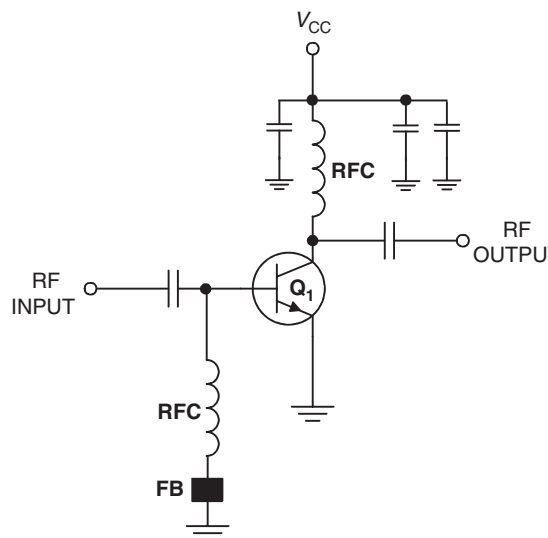


Figure 3.89 Class C power amplifier with ferrite bead on ground-end of base lead.

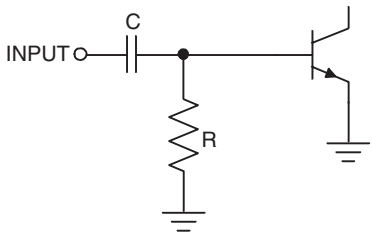


Figure 3.90 A Class C amplifier with signal bias.

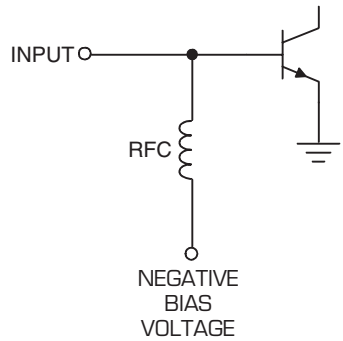


Figure 3.91 A Class C amplifier with external bias.

Signal bias (Fig. 3.90) actually makes use of the RF signal itself to obtain the negative bias at the base required for Class C operation: When a strong signal reaches the input of the transistor it begins to conduct, charging up the series capacitor, *C*. However, when the signal's voltage does not possess the amplitude required to turn on the transistor, or when the signal creates a reverse bias, *C* will then discharge through the shunt resistor, *R*. When this discharge occurs, a negative potential will form at the top of *R*, which produces the negative bias necessary for Class C operation of the amplifier. By manipulating the RC time constant of *R* and *C*, we can increase the negative bias so much that only the highest peaks of the input signal will actually turn on the transistor.

A less common method is *external bias*, shown in Fig. 3.91. This circuit uses a negative bias supply to bias the base, and a standard positive supply for the collector circuit. The RFC acts as a high impedance for the RF frequency itself so that it does not enter the bias supply.

Self bias (Fig. 3.92) uses the emitter current to form a voltage drop across the emitter resistor and, due to the direction of the current flow from emitter to collector, makes the top of the emitter resistor positive. With the emitter positive (which is the common element of the entire circuit), then the base, being at DC ground through the radio frequency choke (RFC), is now negative in respect to the emitter. This action creates Class C operation. The capacitor, C_E, placed across the emitter resistor, also has the same voltage across its terminals as R_E, and stops the bias voltage from being affected by the RF signal's amplitude swing.

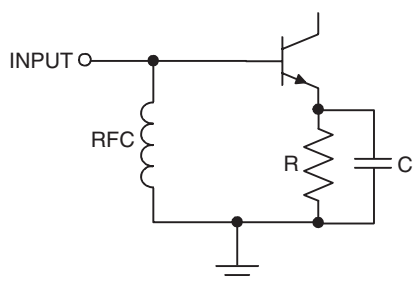


Figure 3.92 A Class C amplifier with self bias.

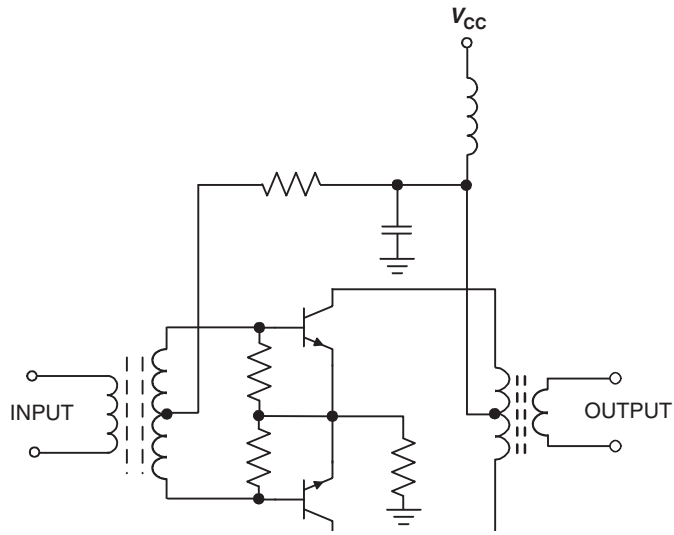


Figure 3.93 A Class B push-pull amplifier.

Class B biasing is normally utilized only with push-pull amplifiers, such as that shown in Fig. 3.93, to obtain linear amplification characteristics. Any scheme that biases the amplifiers at just above cutoff can be used for Class B operation, as in the push-pull amplifier circuit of Fig. 3.94 that utilizes biasing *stabistors*. (Stabistors are two series diodes adopted to maintain an even 0.7 V on each of the transistor's emitter-base junctions, while also helping to protect against the destructive effect of thermal runaway). But for nonlinear operation, Class B can employ single-ended amplifiers fed into RF high-*Q*-tuned circuits, along with any biasing arrangement that sets the transistor at approximately 0.7 V above cutoff.

3.6.4 Amplifier Bias Circuit Issues

To reiterate some of the more important aspects of high frequency bias design: Any emitter bias resistor and emitter capacitor can create low-frequency instability and bias oscillations, as well as increasing the NF and decreasing the gain of an amplifier. This demands that RF transistors have a directly grounded emitter lead, with no emitter feedback caused by the lead wire inductance.

A transistor bias circuit must not only supply bias voltages to the collector and to the base, but it must also control the affects of the amplifier's temperature variations, since DC current gain h_{FE} (β, which is I_C/I_B) of a transistor will increase by about 0.5% for every degree Celsius in an un-temperature stabilized circuit, as demonstrated in Fig. 3.95. This graph shows the typical temperature versus h_{FE} of a standard silicon transistor. Moreover, RF transistors can change their S_{21} (RF gain), stability, and NF quite dramatically as the bias varies due to this temperature sensitivity. In fact, bias has a very large effect on all *S*-parameters, as evidenced by S2P *S*-parameter files being taken at a certain collector-to-emitter voltage and collector current. This places special demands on LNAs, which must have a very stable bias arrangement so that NF is not degraded along with temperature. However, if the transistor is expected to operate

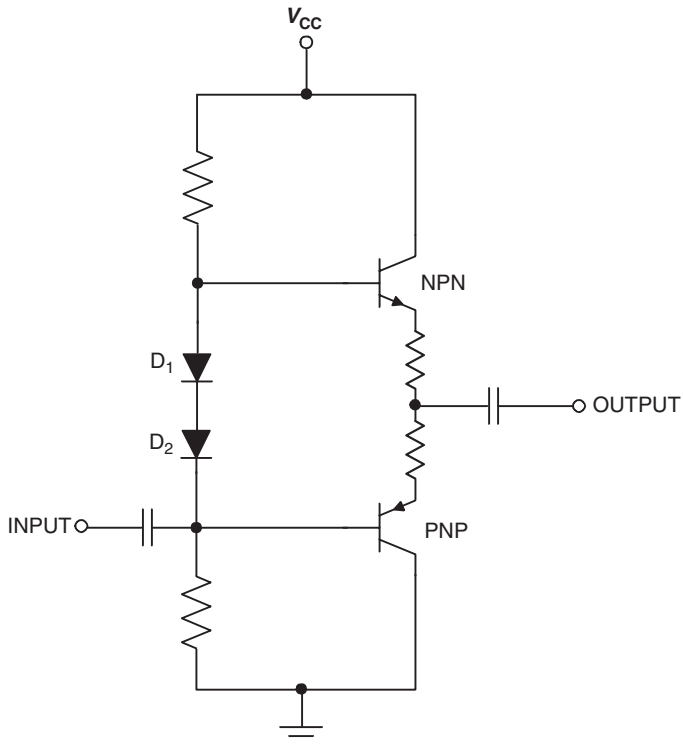


Figure 3.94 A Class B push-pull amplifier stabilized with stabistors.

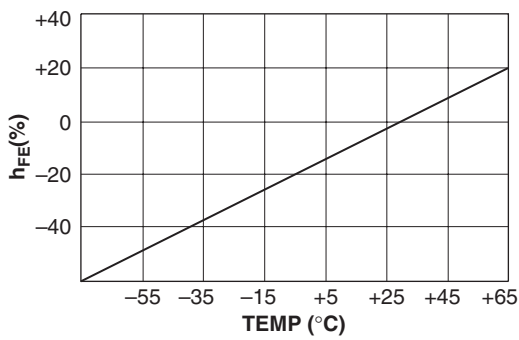


Figure 3.95 Temperature change versus *hFE* of a bipolar transistor.

only in slightly elevated room temperature environments, then relatively primitive and simple temperature stabilization bias schemes are all that may be required for most LNA and general RF amplifiers.

Looking further into amplifier temperature effects, since the two transistor characteristics that have such a large consequence on an amplifier's DC operating point over temperature are ΔV_{BE} and $\Delta\beta$, any adept temperature-stable bias design will obviously need to decrease these variations, as discussed above. With normal transistors,

changes in beta with temperature can be quite drastic, and will vary the I_C by as much as ± 25% for each temperature variation of ± 50°C. In addition, part-to-part variations in beta for even a single transistor model can be in the 10 to 1 range, with beta variations of 40 to 400. Thus, in a large manufacturing atmosphere, a method must be found to design an amplifier that ignores beta variations from transistor to transistor, as well as over temperature. The following formula can be adopted to calculate the change that can be expected in the I_C of a transistor, and to give us an idea of a proper bias design that will stabilize these beta variations. Refer to Fig. 3.96:

$$\Delta I_C = I_{C1}\left(\frac{\Delta\beta}{\beta_1\beta_2}\right)\left(1+\frac{R_B}{R_E}\right)$$

where $I_{C1} = I_C$ (at $\beta = \beta_1$), A
$\beta_1 = \beta_{LOW}$ expected of the transistor
$\beta_2 = \beta_{MAX}$ expected of the transistor
$R_B = R_1$ and R_2 in parallel, Ω
R_E = transistor's emitter resistor value
$\Delta\beta = \beta_2 - \beta_1$

We can clearly see that the entire domination over these beta variations, which strongly affects I_C, is only in the *ratio* of R_B/R_E, or:

$$\frac{\left(\frac{R_1R_2}{R_1+R_2}\right)}{R_E}$$

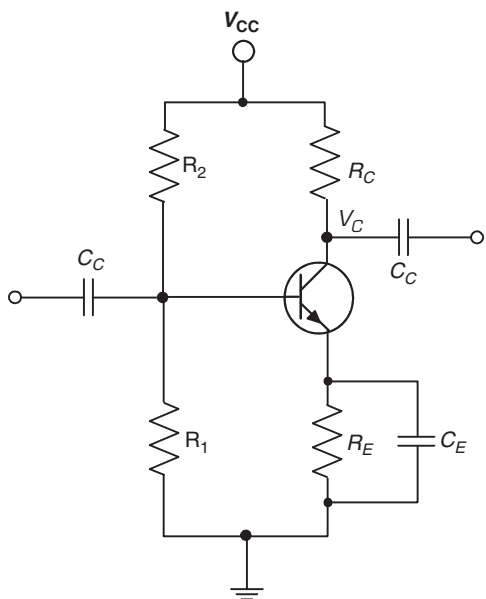


Figure 3.96 Basic C-E circuit for bias stabilization calculation.

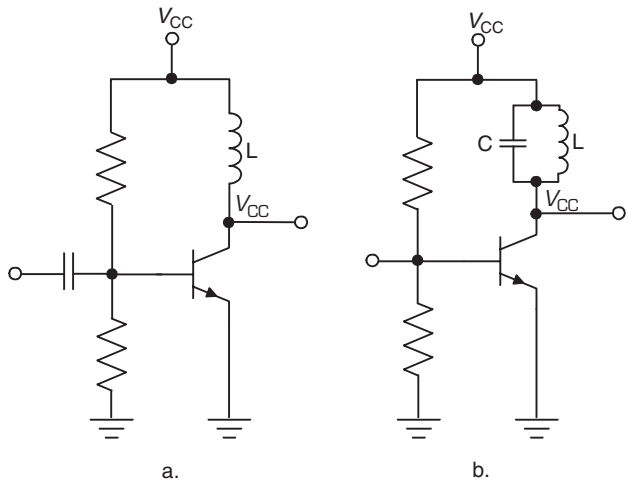


Figure 3.97 Class A transistor amplifiers with (*a*) inductor load and (*b*) tank load.

As this ratio decreases, the beta variations stabilize (but the gain of the amplifier will also decrease). A R_B/R_E ratio of 10 or less will usually give a very stable beta design.

Class A amplifiers with either inductor or LC resonant tank collector loads (Fig. 3.97) are able to have a lower V_{CC}, and less power losses, than circuits employing a resistive load at the collector. This is because the DC voltage drop across the collector load that uses an inductor is at a very low value, and is equal to the inductor's DC resistance. Since the inductor or inductor/capacitor combination "forces" the average voltage to be approximately V_{CC} at the transistor's collector, instead of around half the V_{CC} when using a collector resistor, the RF output will swing $2 \times V_{CC}$ above, and approximately 0 V below, this average V_{CC} when the amplifier's input is driven hard. This doubles the available RF voltage at the output of the transistor.

In designing small-signal amplifiers, the transistor's collector current does not necessarily have to be at the middle of its $I_{C(MAX)}$, since the stage will only be amplifying low signal levels. The I_C can be chosen to be in the most linear part of its characteristic curve, and at a low enough amplitude that DC power dissipation is at a minimum, but not so low that any RF signal will be too near cutoff, or at excess distortion levels, or where the stage gain will suffer. However, most I_C values, as well as V_{CE} values, will be chosen to conform to the *S*-parameter files available for ease of design and simulation. (It must also be kept in mind that after calculating the matching network of an amplifier with existing *S*-parameters, we must calculate the bias components with the exact same V_{CE} and I_C that were used to originally measure those *S*-parameters, and as are shown in the ∗.S2P text file, or the active device's port impedances will not be correct, since Z_{IN}/Z_{OUT} vary with changes in I_C and V_{CE}.)

Some lower-frequency RF amplifiers will split the single emitter feedback resistor into two separate emitter resistors (Fig. 3.98), with only one of these resistors having an AC capacitor bypass, while the other one is providing constant degenerative feedback to enhance amplifier stability, reducing the chance of oscillations. This also allows the

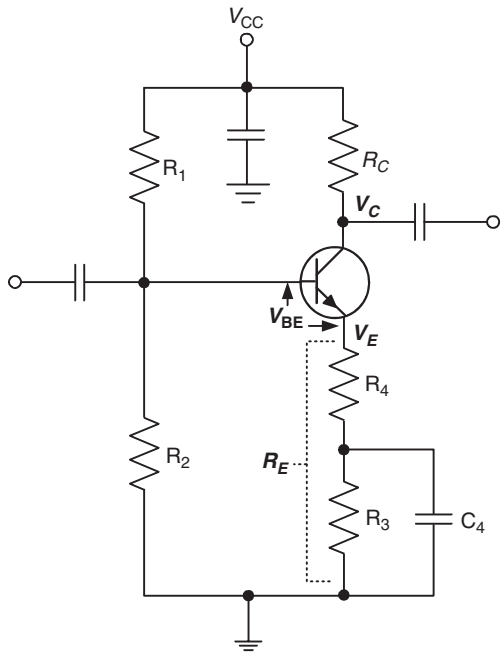


Figure 3.98 Split emitter feedback for bias and gain stabilization.

designer to solidly set the gain, irrespective of the transistor's varying batch-to-batch tolerances, to:

$$20\log\left(\frac{R_C}{R_4}\right) = \text{gain in dB}$$

3.6.5 Amplifier Bias Design

There are many different ways to bias an amplifier, depending on the required temperature stability, efficiency, cost, device, power output, linearity, and so on. The following are the most popular bias circuits and design methods. *Matching circuits not shown*.

Class AB Diode or Transistor Bias for Microwave Power Amplifiers of up to 3 W (Fig. 3.99)

Because of the huge amount of wasted power caused by high PA currents that would be dissipated across any collector resistor, standard bias techniques cannot be used for most power amplifiers. While there are highly complex active biasing schemes available, the two presented below are the most common and lowest in cost for < 3 W microwave/RF power amplifier designs.

To Design

1. $R_S = 5\ \Omega$
2. $C_{RF} = < 2\ \Omega$ at f_r

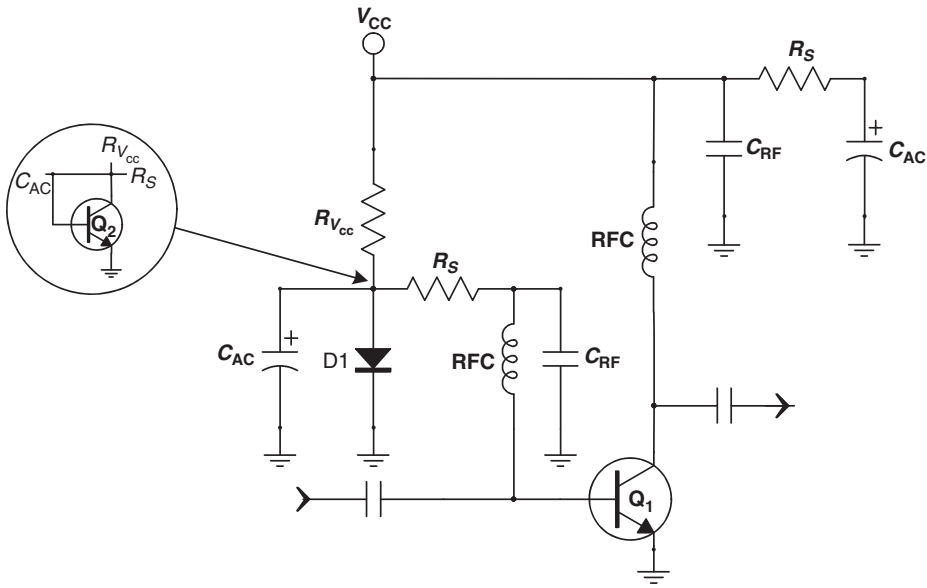


FIGURE 3.99 Diode and transistor bias for power amplifiers.

3. RFC = > 500 Ω X_L
4. C_{AC} = > 1 uF
5. D_1, Q_2 = general purpose diode or transistor
6. $R_{V(CC)} = \dfrac{V_{CC} - 0.7}{I_{C(MAX)} \div H_{FE(MIN)}}$

where V_{CC} = supply voltage, V
$I_{C(MAX)}$ = *RF transistor's* (Q_3) maximum rated collector current, A
$H_{FE(MIN)}$ = *RF transistor's* (Q_3) minimum rated H_{FE}

NOTE: *Whether employing diode* (D_1) *or transistor* (Q_2) *bias, it is essential to thermally connect these components to the RF transistor itself. This allows the semiconductor bias components to track the power amplifier's temperature variations, and thus increase/decrease the 0.7 V placed across* Q_1*'s base, maintaining the PA's collector current at a steady DC level. (As the temperatures rise, a silicon semiconductor junction's voltage decreases from its room temperature value of 0.7 V.) Since the bias components can only react to* Q_1*'s case temperature, and not its die temperature, the bias will not be completely responsive to these temperature variations, nor will the thermal bonding to the PA be 100% efficient. However, this simple bias topology will be more than adequate for low power microwave PA's, as typically found in consumer wireless equipment. Further, in order to force PA bias stability no matter what the input and output RF power levels may be, the standing current through the diode or transistor bias components must be high enough to permit a steady voltage to be maintained across the PA's base. We can guarantee this by following the above design procedure.*

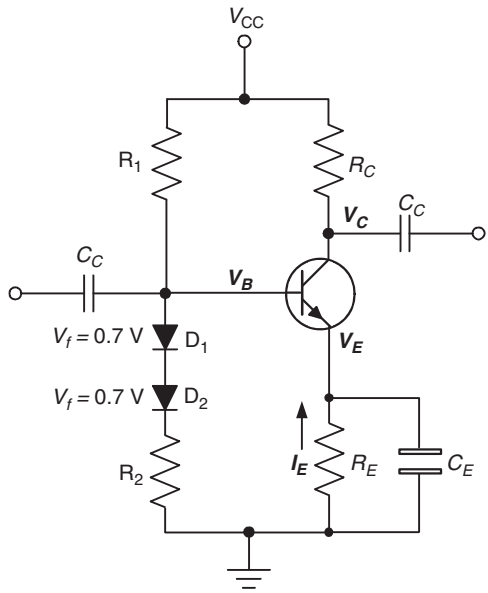


FIGURE 3.100 Class A diode temperature stabilized amplifier for bias calculations.

Class A Bias, Highly Temperature Stable Diode BJT Amplifier for HF and Below (Fig. 3.100)

Design By

1. Choose the transistor's operating point. Example: $V_{CC} = 12$ V; $I_C = 10$ mA; $V_C = 6$ V; $\beta = 50$. (I_C and V_C should be the same as the available *S*-parameter files for the active device, as found in its ∗.S2P text format).
2. $R_1 = R_2 = 10\ \text{k}\Omega$
3. $V_B = \left(\dfrac{R_2}{R_1 + R_2}\right) \cdot \left(V_{CC} - 2\,V_F\right) + 2\,V_F \quad [V_F \approx V_{BE} \approx 0.65\ \text{V}]$
4. $R_E = \dfrac{I_E}{V_B - V_{BE}} = \dfrac{I_E}{\left(\dfrac{R_2}{R_1 + R_2}\right) \cdot V_{CC}} \quad [I_E \approx I_C]$
5. $V_C = \dfrac{V_{CC}}{2}$
6. $R_C = \dfrac{V_{CC} - V_C}{I_C}$
7. $V_F = 0.7$ V
8. D_1, D_2 = silicon, small-signal general purpose diodes

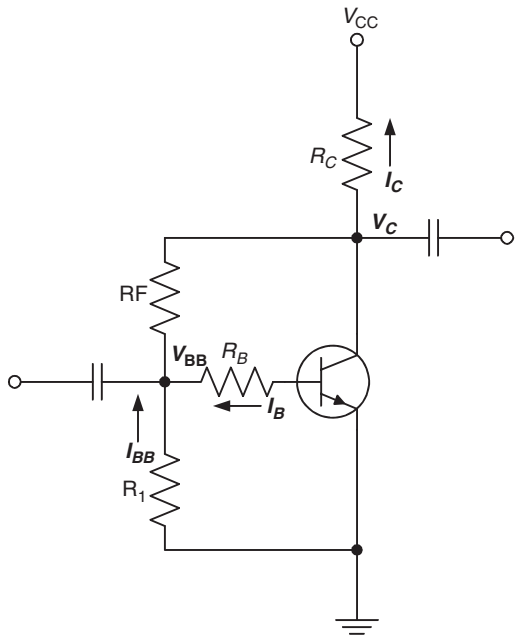


Figure 3.101 Class A HF, VHF, and UHF moderately stable amplifier design with voltage feedback.

Class A Bias, HF, VHF, UHF Temperature Stable BJT Amplifier Design (Fig. 3.101)

Design By

1. Choose the transistor's operating point. (Example: $V_{CC} = 12$ V; $I_C = 10$ mA; $V_C = 6$ V; $\beta = 50$. I_C and V_C should be the same as the available *S*-parameter files for the active device, as found in *.S2P format).
2. Use a value for V_{BB} and I_{BB} to supply a constant stabilizing current (I_B): $V_{BB} = 2$ V; $I_{BB} = 1$ mA
3. $I_B = I_C \div \beta$
4. $R_B = \dfrac{V_{BB} - V_{BE}}{I_B}$
5. $R_1 = V_{BB}/I_{BB}$
6. $R_F = \dfrac{V_C - V_{BB}}{I_{BB} + I_B}$
7. $R_C = \dfrac{V_{CC} - V_C}{I_C + I_B + I_{BB}}$

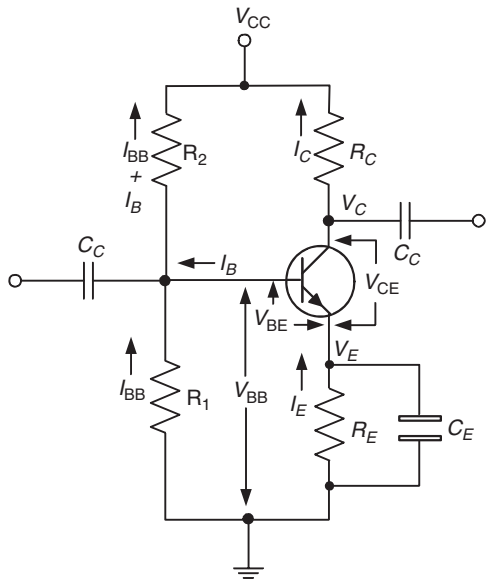


FIGURE 3.102 A low-frequency Class A amplifier for bias design example.

Class A or AB Bias, HF Temperature Stable Amplifier (Fig. 3.102)

Design By

1. Choose the transistor's operating point. (Example: V_{CC} = 12 V; I_C = 10 mA; V_C = 6 V; β = 50. I_C and V_C should be the same as the available *S*-parameter files for the active device, as found in ∗.S2P format).
2. Give a V_E value of 2 V for bias temperature stability.
3. Give $I_E \approx I_C$ for typical or high beta transistors.
4. $R_E = V_E / I_E$
5. $R_C = \dfrac{V_{CC} - V_C}{I_C}$
6. $I_B = I_C / \beta$
7. $V_{BB} = V_E + V_{BE}$
8. Give a value for I_{BB} of 1.5 mA (The larger in value I_{BB}, the higher the stability; but power dissipation increases).
9. $R_1 = V_{BB}/I_{BB}$
10. $R_2 = \dfrac{V_{CC} - V_{BB}}{I_{BB} + I_B}$
11. $X_{CE} < 1\ \Omega$

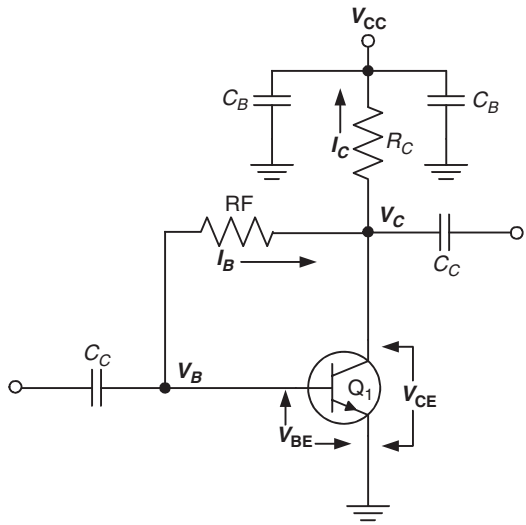


Figure 3.103 A Class A high-frequency amplifier with collector feedback.

Class A or AB Bias, Moderately Temperature Stable Collector Feedback Amplifier for HF, VHF, UHF, and Above (Fig. 3.103)

To Design

1. Choose the transistor's operating point. (Example: $V_{CC} = 12$ V; $I_C = 10$ mA; $V_C = 6$ V; $\beta = 50$. I_C and V_C should be the same as the available *S*-parameter files for the active device, as found in ∗.S2P format).
2. $I_B = I_C/\beta$
3. $R_F = \beta \cdot \dfrac{V_C - 0.7}{I_C}$
4. $R_C = \dfrac{V_{CC} - V_C}{I_B + I_C}$
5. $X_{CB} = X_{CC} = < 1\ \Omega$

Alternatively

1. Select a V_{CE} of $V_{CC}/2$ and an I_C that is shown by the transistor's data sheet to be at maximum gain, NF, or P1dB. Or choose a V_{CE} and an I_C that is the same as the *S*-parameter file available for the active device (Do not confuse V_{CE} in *S*-parameter calculations with V_{CC}. V_{CC} can be as high as desired, but V_{CE} must be as stated in the *S*-parameter file that is available for the selected transistor in order to obtain an accurate design and RF simulation).
2. $R_C = \dfrac{V_{CC} - V_{CE}}{I_C}$

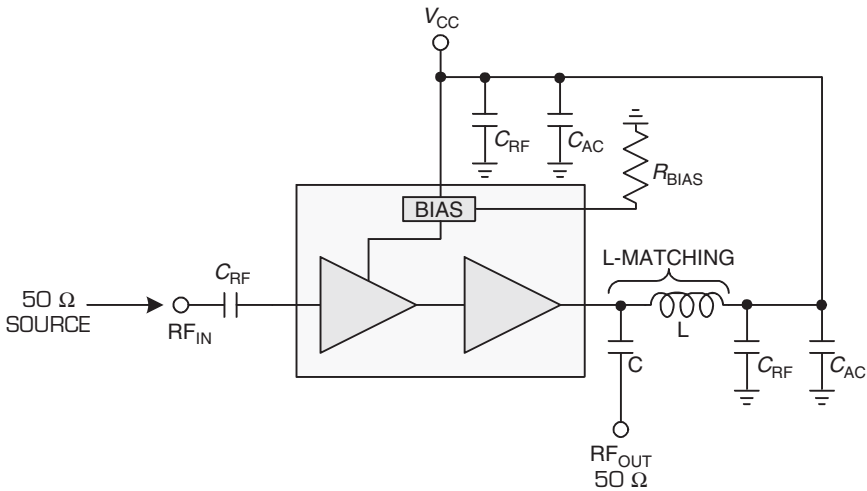


Figure 3.104 An integrated buffer amplifier for 50 Ω.

3. $R_F = \beta \dfrac{V_{CE} - 0.6}{I_C}$
4. $X_{CB} = X_{CC} = < 1\ \Omega$

Class A Bias Buffer Amplifier, Integrated (Fig. 3.104)

To Design

1. An easy to implement Maxim MAX2473 integrated buffer amplifier is available for low to medium RF output powers, and was created specifically for voltage controlled oscillator (VCO) buffer use. The chip comes as both a single buffer and as a splitter type buffer that permits a VCO to feed two other circuits (as the MAX2472).
2. The MAX2473 can be used as a high input impedance buffer for discrete VCOs designed from this book, or for any prepackaged 50-Ω VCO. The MAX2473 will function with a V_{CC} of between 2.7 and 5.5 V; has a wideband RF frequency range of 0.5 to 2.5 GHz; very high reverse isolation of almost 50 dB at 1 GHz; adjustable P_{OUT} of between −10 to −2 dBm (dependent on R_{BIAS}); has a small signal gain of 12 dB (at 1 GHz, and dependent on R_{BIAS}); has a current draw of approximately 5 mA (dependent on R_{BIAS}); an unmatched input impedance of 250 Ω at 1 GHz (with external matching, 50 Ω); and a 50-Ω output when used with a simple L network.
3. At 900 MHz and 3 *V*cc select an R_{BIAS} of 11 kΩ for a P1dB of −0.9 dBm, an I_C of 4.8 mA, and an S_{21} of 12.5 dB; select an R_{BIAS} of 23 kΩ for a P1dB of −6.0 dBm, an I_C of 2.7 mA, and an S_{21} of 8.5 dB.

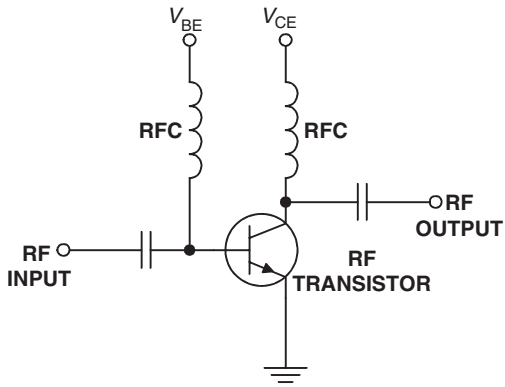


Figure 3.105 Class A lumped linear amplifier without a bias circuit.

Class A Active Bias, Highly Stable, for Microwave Amplifiers

Both the lumped amplifier of Fig. 3.105 and the distributed amplifier of Fig. 3.106 can function as linear Class A amplifiers. They can perform with high temperature stability without the assistance of the gain reducing and stability robbing emitter resistor. (The emitter resistor possesses a small value of inductance, which is an issue in high VHF and above amplifier applications). No bias resistors are required due to the inclusion of the DC active bias circuit of Fig. 3.107, which includes a positive-negative-positive (PNP) biasing transistor and its associated diode. Figures 3.108 and 3.109 show the completed and biased amplifiers, both lumped and distributed. Below demonstrates how to design the active biasing network for a high-frequency Class A lumped or distributed amplifier.

To Design

1. Select a I_D through the diode of 2 mA.
2. Select an appropriate I_C for Class A bias of the RF transistor amplifier of Fig. 3.105 or 3.106.

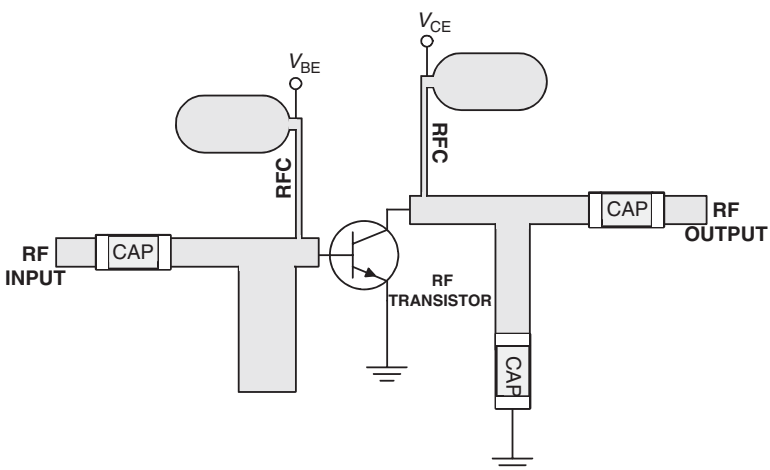


Figure 3.106 Class A distributed linear amplifier without a bias circuit.

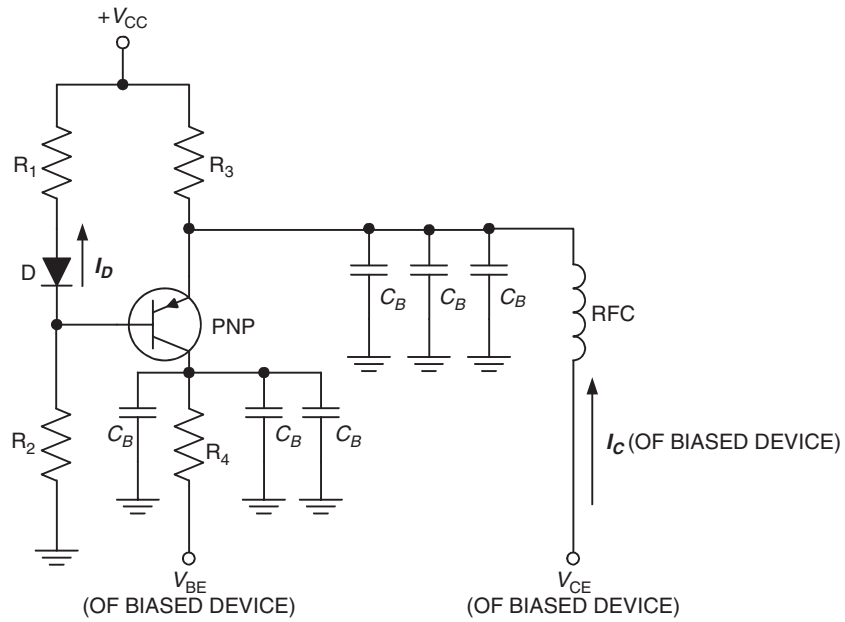


FIGURE 3.107 Class A active bias circuit.

3. Select a V_{CC} for the active bias network that is approximately 2 or 3 V greater than the V_{CE} required for the RF transistor of Fig. 3.105 or 3.106.
4. Select an RFC for the active bias circuit with the appropriate SRF (self resonant frequency) that is greater than the frequency of operation.
5. Select both a silicon PNP transistor with a beta of at least 30 and a low-frequency silicon diode (a PNP transistor is used so that the V_{CC} may be a positive voltage).
6. $R_1 = \dfrac{+V_{CC} - V_{CE}}{I_D}$
7. $R_3 = \dfrac{+V_{CC} - V_{CE}}{I_C}$
8. $R_2 = \dfrac{V_{CE} - 0.7}{I_D}$
9. $R_4 = \beta_{MIN} \dfrac{V_{CE} - 1}{I_C}$
10. Collector current of biased device will be: $I_C = \dfrac{R_1(V_{CC} - 0.7)}{R_3(R_1 + R_2)}$
11. Collector-to-emitter voltage of biased device will be: $V_{CE} = V_{CC} - (I_C \cdot R_3)$
12. $C_B = < 1\ \Omega\ X_C$

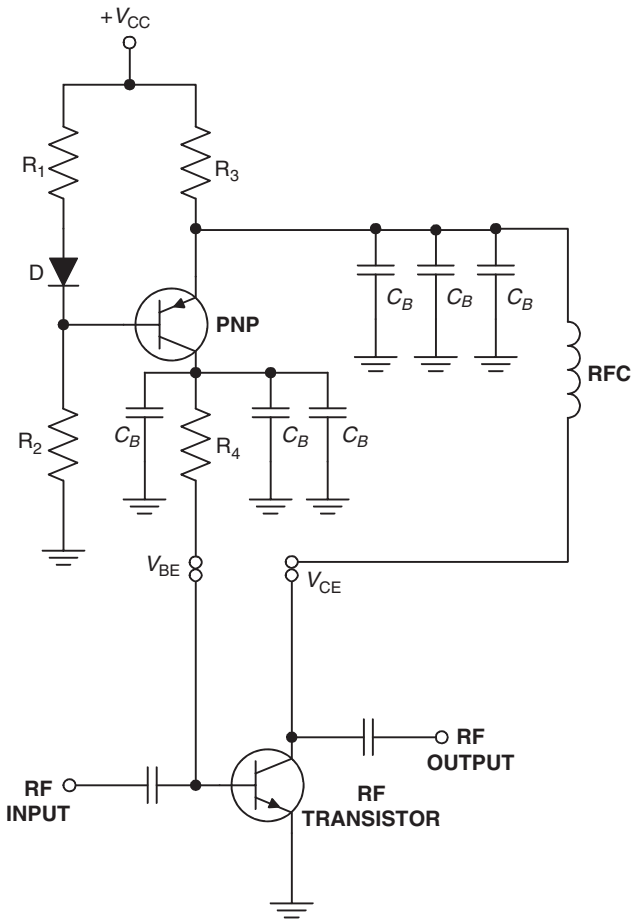


Figure 3.108 Class A lumped linear amplifier with active bias.

Class A JFET Self-Bias, Common-Source Amplifier for VHF and Below (Fig. 3.110)

To Design

1. Select a V_{dd} and an appropriate V_{gs} for Class A operation from the data sheet for the JFET selected, and note the I_d for this chosen V_{gs}.
2. $R_S = \dfrac{V_{gs}}{I_d}$
3. Select a V_{ds} of $\dfrac{V_{dd} - (2\,V_{gs})}{2}$ for a V_d of $\dfrac{V_{dd}}{2}$
4. $R_d = \dfrac{V_{dd} - V_{ds} - V_{gs}}{I_d}$ (If R_d computes to lower than 1 kΩ, an RFC must be used between the top of R_d and the V_{dd} in order to sustain a minimum RF impedance into the power supply).

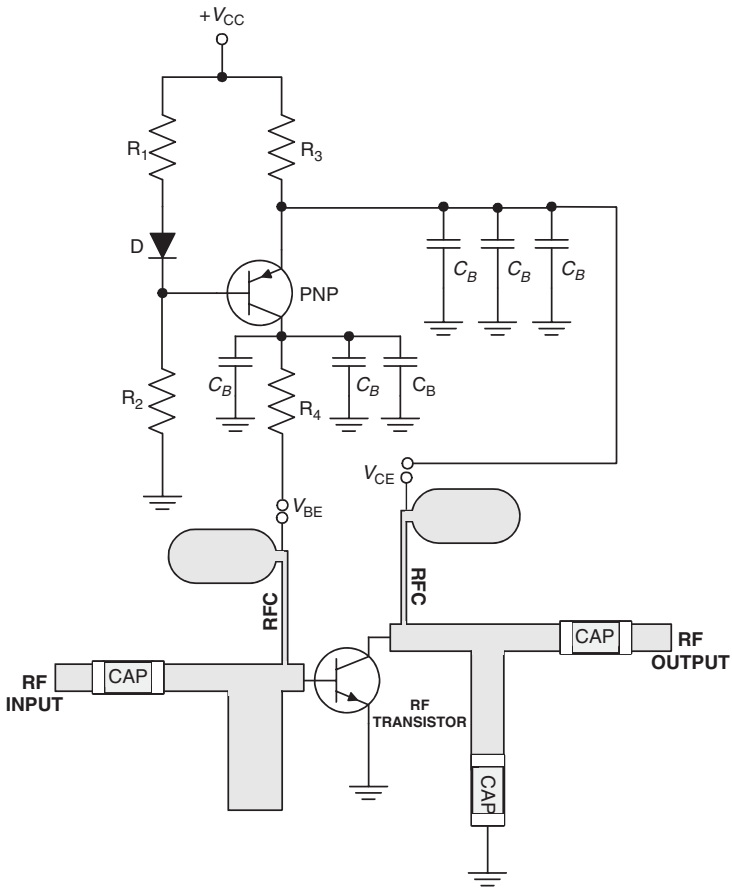


FIGURE 3.109 Class A distributed linear amplifier with active bias and matching circuits.

5. Place a high impedance RFC (with an appropriate SRF) or a high-value resistor (1 MΩ) from the FET's gate to ground.
6. $C_C = C_S < 1\ \Omega$

NOTES: *a. Always confirm that the FET will safely dissipate the power required, which is* $P = I_d \cdot V_d$.

b. Since V_{gs} *for a specific* I_d *is not always available, use the following equation to find* V_{gs} *when* I_{dss} *and* V_p *are known (look in the JFET data sheet for* I_{dss} *and* V_p*):*

$$V_{gs} = V_p\left(1 - \sqrt{\frac{I_d}{I_{dss}}}\right) \quad (V_p \approx V_{gs(off)} \text{ and } V_s = -V_{gs})$$

c. The I_d *and* V_{ds} *will normally be chosen as a duplicate of the values used in any available S-parameter file for the device to be modeled. In fact, many manufacturers of FETs will have S-parameters that are taken at different values of* V_{ds} *and* I_d*. The* I_d *can be quoted as a percent of* I_{dss} *(the maximum* I_d*) such as "50% of* I_{dss}*", which would work well for Class A bias.*

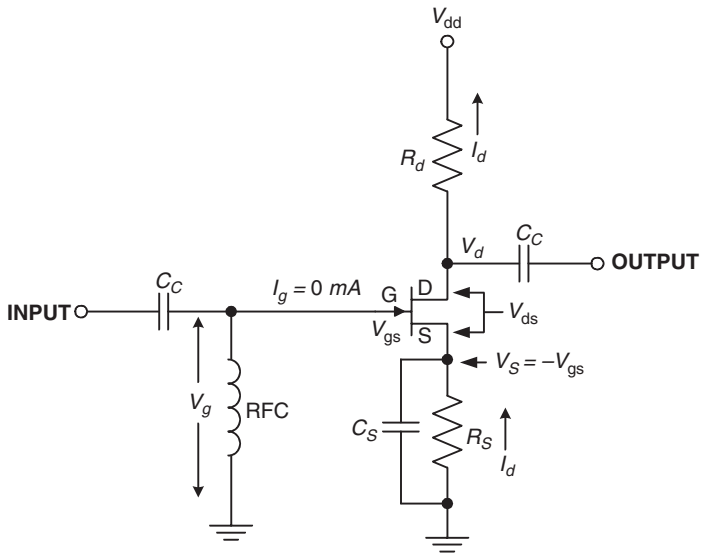


Figure 3.110 JFET biasing up to medium frequencies.

Class A Bias, JFET HF Stable Amplifier (Fig. 3.111)

To Design

1. Select a Q-point (V_{dd}, I_d, and $V_d = V_{dd}/2$).
2. $R_d = \dfrac{V_{dd} - V_d}{I_d}\left(I_d = I_{dss}\left(1 - \dfrac{V_{gs}}{V_p}\right)^2\right)$

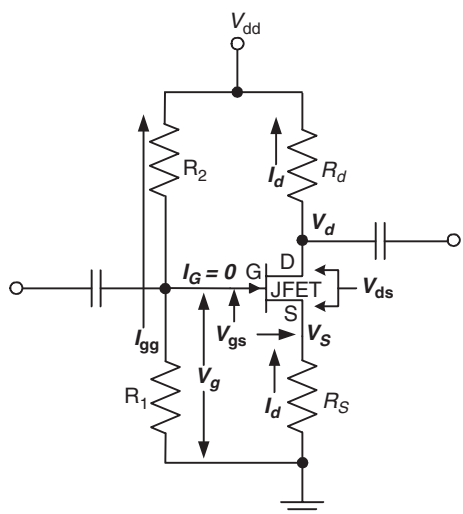


Figure 3.111 A low-frequency Class A JFET bias circuit.

3. Find V_p and I_{dss} from data sheet.
4. $V_{gs} = V_p\left(1 - \sqrt{\frac{I_d}{I_{dss}}}\right)$
5. Select V_S to be 2 or 3 V.
6. $R_S = V_S/I_d$
7. $V_g = V_{gs} + V_s$
8. Use an R_1 value of 220 kΩ (this effects the DC input resistance).
9. $R_2 = \frac{R_1(V_{dd} - V_g)}{V_g}$

NOTE: *The I_d and V_{ds} will usually be chosen as a duplicate of the values used in any available S-parameter file for the device to be modeled. In fact, many manufacturers of FETs will have S-parameters available taken at different values of V_{ds} and I_d (I_d is usually quoted as a percent of I_{dss} (the maximum I_d) such as "50% of I_{dss}," which would work well for Class A bias).*

Class C Bias, BJT Power Amplifier (Fig. 3.112)

To Design

Since the average silicon transistor will naturally run in Class C mode if no base bias at all is supplied, the bias network for such an amplifier is quite simple. In order to minimize the chances of instability, the base of this BJT Class C amplifier should be grounded through a low-*Q* choke, or a ferrite bead that is operational at the frequencies of interest should be attached to the ground end of an average RFC. An RFC or resonant LC circuit is then placed at the collector. However, certain design complications that are inherent with transistorized Class C operation are explored in depth in the Class C power amplifier design section.

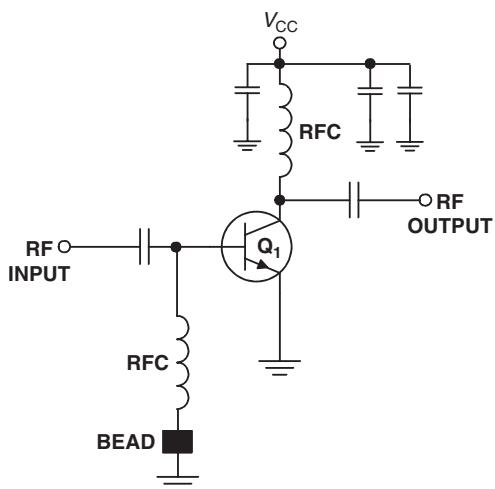


FIGURE 3.112 A high-frequency Class C power amplifier.

3.7 MMIC Amplifiers

3.7.1 Introduction

MMICs are *monolithic microwave integrated circuits*, typically comprising 50-Ω small-signal amplifiers that require very few support components for biasing, and none for impedance matching.

Figure 3.113 illustrates two very common MMIC packages, with integral microstrip leads, for high-frequency operation. Some MMICs may have a separate DC power input pin on the package itself, which may be of the eight pin DIP package variety.

Taking a simple example of a high quality and stable MMIC is the classic Agilent INA series of radio frequency integrated circuit (RFIC) gain blocks, with the internal structure shown in Fig. 3.114. This design employs a single transistor driving a Darlington pair, with a small resistive feedback to set the RF parameters. The single transistor has only a small amount of negative feedback for low noise performance, and with high gain into the DC coupled input of the Darlington pair. The Darlington pair has strong degenerative feedback, and sets the gain and matching of the RFIC, as well as the gain flatness.

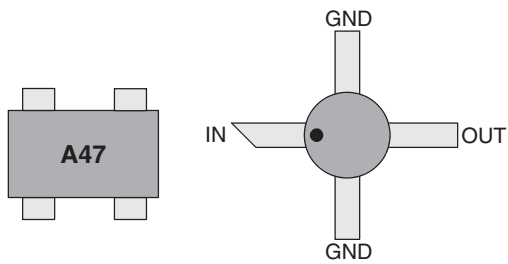


Figure 3.113 Standard MMIC amplifier packages.

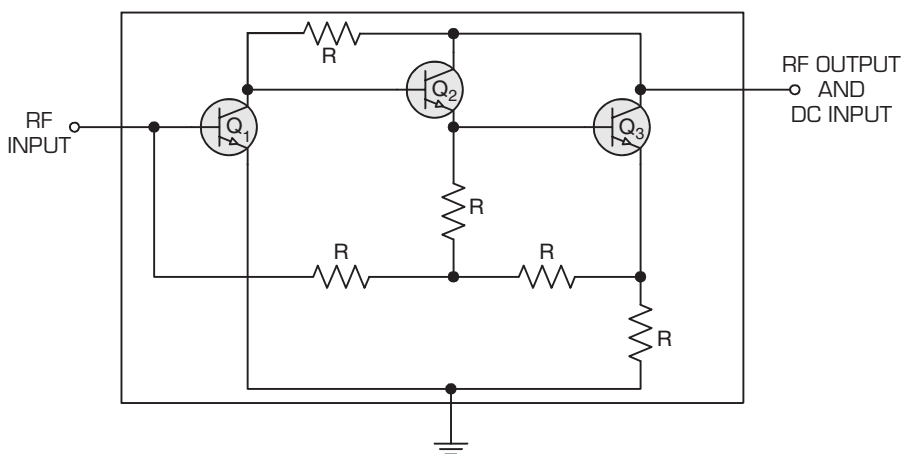


Figure 3.114 Internal circuit arrangement of an average MMIC.

Even though the majority of amplifier MMICs are unconditionally stable, it is wise not to assume that all MMICs are. However, the amplifier manufacturer will usually warn you of this fact, even if this warning is somewhere at the bottom of the data sheet.

3.7.2 MMIC Amplifier Biasing

The *current-biased* MMIC (by far the most common type), will attempt to draw more current as the temperature rises. To preserve a MMIC's drain current (I_d) so that its *Q*-point bias does not vary with temperature or large input signals, the DC bias circuitry, consisting of R_{BIAS} (Fig. 3.115), must maintain a constant I_d to the MMIC under all conditions. This bias stabilization will lower the MMIC's DC voltage (V_d) with higher device temperatures, and increase the voltage with lower device temperatures, thus preventing the bias current from decreasing; or from increasing to a level that can actually destroy the MMIC. (Any changes in I_d will vary both the gain and the P1dB of the device). However, the R_{BIAS} in the DC bias line will drop more of the MMIC's V_d voltage from the V_{CC} supply (since $V = I \cdot R$), which decreases V_d and, consequently, the device's bias current to the desired levels. Since the effectiveness of this temperature bias control is dependent on the voltage drop across R_{BIAS}, a value of up to 4 V may be required for proper stabilization across a minus 25°C to a plus 100°C temperature range. By using a higher-value resistance of R_{BIAS} along with a higher supply voltage, stabilization is improved. Nevertheless, R_{BIAS} itself should not have such poor temperature stability that it varies its resistance dramatically over temperature.

If the R_{BIAS} does not add up to 600 Ω or more, then the gain of the MMIC stage will suffer. This is because all properly decoupled power supplies are virtually a short to RF, and R_{BIAS} will decrease this gain "shorting" effect on the output of the MMIC by being at a high resistance value. However, if R_{BIAS} does not compute to be at or over 600 Ω (and it very rarely is), then an RFC should be added in series with R_{BIAS} to increase the output to this value, or approximately $R_{BIAS} + X_L > 600\ \Omega$.

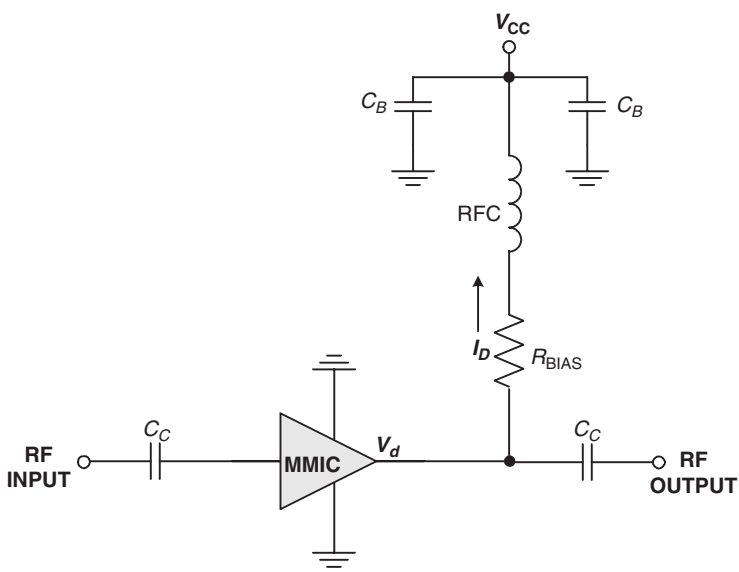


Figure 3.115 Standard MMIC gain block with biasing.

R_{BIAS}, since it drops the excess voltage from V_{CC}, also smoothes out any voltage fluctuations to the MMIC, which could cause an unstable bias point, by acting almost as a constant current source. As well, the added RFC blocks most of the RF from entering the bias V_{CC} line by behaving as a high impedance to RF, while the two $C_{B's}$ bypass any further RF to ground.

The manufacturer's approved DC bias current for the MMIC should be followed closely due to problems with decrease gain and improper matching at lower I_d levels, and device damage at the higher I_d levels, or:

$$I_d = \left(\frac{V_{CC} - V_d}{R_{BIAS}} \right)$$

Agilent, through empirical studies, recommends placing the R_{BIAS} resistor at the output to the MMIC, followed by the RFC for improved RF performance. The bypass capacitors, of course, should always be placed after the RFC, and not before, or the gain of the MMIC will decrease severely.

3.7.3 MMIC Biasing Procedure

To Design

1. Choose a V_{CC} that will allow at least 2 V, and preferably 4 V, to be dropped across R_{BIAS} for stability (Fig. 3.115), while also supplying the MMIC with the proper V_d level. (If R_{BIAS} does not reach 600 Ω, use an RFC for a combined additive impedance of 600 Ω for both R_{BIAS} and RFC):

$$R_{BIAS} = \frac{V_{CC} - V_d}{I_d}$$

where V_d = DC voltage at the MMIC's power pin, V
I_d = DC current into the MMIC's power pin, A
Vcc = power supply voltage, V

2. Check the power dissipation within the bias resistor R_{BIAS} to allow for the appropriate safety headroom of at least double the calculated R_{BIAS} wattage, or $P = 2(I^2 \cdot R)$.
3. Use coupling capacitors at the MMIC's input and output as described under Sec. 3.7.4.

NOTE: *The above describes biasing and operation of the most prevalent MMIC, the current-biased MMIC. However, some MMICs, such as Agilent's MGA-85563 LNA MMIC (Fig. 3.116), are voltage biased. This type of MMIC operates quite well when only low values of V_{CC} are available, since no R_{BIAS} is required, at low current draw levels, making it perfect for portable battery powered applications (see Sec. 3.7.5).*

Some MMICs can be adopted to limit output signal amplitudes for modulations that employ a constant modulation envelope, like common FM. A MMIC with a hard saturating characteristic, as well as high gain, is required for this application (such as in the INA series of MMICs). Since almost all MMICs will vary in both gain and saturation

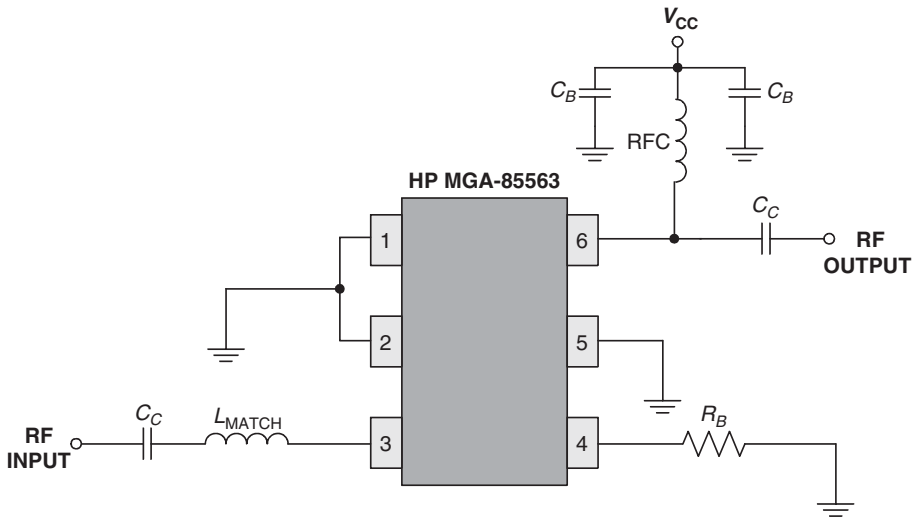


Figure 3.116 Voltage-biased LNA MMIC.

level depending on current draw, the bias point of these MMIC limiters must not be allowed to vary with large RF drive transitions, and the factory recommended bias current levels should be maintained to limit harmonic output. Maintaining this constant bias point in limiter applications can best be accomplished by using the biasing circuit as shown in Agilent's application note AN-S003.

3.7.4 MMIC Coupling and Decoupling

MMIC coupling and decoupling is just as important as it is in discrete amplifier circuit design. As shown in Fig. 3.115, two C_C's are utilized at both the input and output of the MMIC to not only couple RF with no voltage drop, but also to block DC from reaching any other device, which would disrupt the biasing of the next stage (or simply be shorted to ground). The capacitors are typically chosen to supply 1 to 5-Ω X_C at the lowest frequency to be passed, while the highest desired frequency should not be close to the capacitor's parasitic *parallel resonant frequency* (*PRF*). For narrow frequency use, the capacitor's own *series resonant frequency* (*SRF*) is sometimes chosen to be the same as the amplifier's RF signal frequency, thus allowing lower-value capacitors to be selected for microwave coupling, while also minimizing undesired lower frequencies from passing to the next stage.

Any RF that is permitted to enter the MMIC's bias power supply can cause various circuit instabilities throughout the system. To decouple, or stop, AC from entering the supply (while permitting DC an unimpeded flow) we can use the RFC and the C_B of Fig. 3.115. Normally more than one value of C_B will be selected so that a wide band of frequencies will be blocked (shunted to ground), while also filtering any power supply ripple or EMI from entering the MMIC stage itself.

As stated, it is important that decoupling and coupling capacitors not be near their parallel resonant mode, or they will act as a high impedance to the RF, instead of as an RF short. Further, any decoupling inductors should not be close in frequency to any *series* resonant modes, or they will begin to function as shorts and not as high impedances.

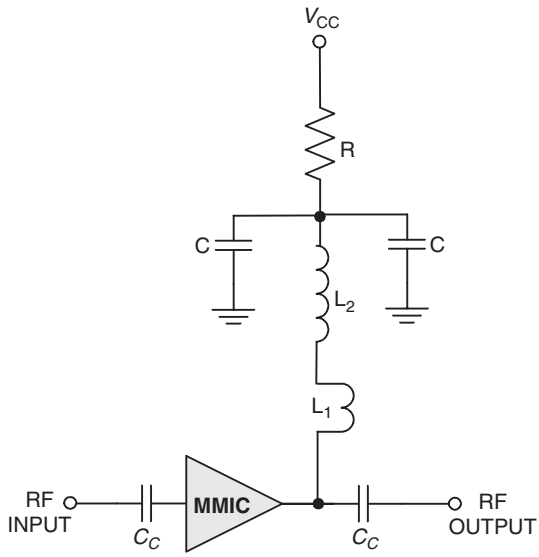


Figure 3.117 Decoupling of a MMIC at both high and low frequencies.

When a MMIC (or discrete) amplifier must be capable of operating properly across a very wide bandwidth, two RFCs may be required in its decoupling circuit (Fig. 3.117): A low impedance coil (L_1) that functions suitably at very high frequencies, without hitting any series resonances, and a high impedance coil (L_2) used to block lower frequencies. The high impedance coil will begin to lose its ability to block the upper frequencies of the passband due to its inherently high turn-to-turn capacitance in such a large-value coil. In fact, at elevated frequencies this high impedance coil begins to look more like a short. The smaller in value, but much higher frequency, coil of L_2 now takes over. An added bypass capacitor (C) to ground may also be placed between the large coil and ground to further decouple any RF from the supply. These precautions will permit a very wide RF passband for a relatively flat gain over a large frequency area.

3.7.5 MMIC Amplifier Circuit

The Agilent voltage-biased MGA-85563, as shown in Fig. 3.116, is capable of operation from 800 MHz to 5.8 GHz, with a V_{CC} of 3 V at 15 mA and an NF of approximately 1.6 dB. It has 18 dB of gain with unconditional stability.

Looking at the MGA-85563 circuit, we find at the RF input a DC blocking capacitor C_C, which is required only if DC is present at the input from the prior stage, while the inductor L_{match} is chosen to cancel the natural capacitive reactance of the device's input in order to give a $50 + j0$ input. However, if the MMIC will be used in receiver front-end applications, then the match should be chosen to supply an optimum source impedance for the lowest possible NF. For optimum NF, some simulation runs and tweaking on the prototype PC board is usually required for any matching network. Agilent's recommended L_{match} value for 800 MHz is 22 nH; 900 MHz is 18 nH; 1.5 GHz is 8.2 nH; 1.9 GHz is 5.6 nH; and 2.4 GHz is 2.7 nH. Above 3 GHz no inductor at all is required. The RF output of the MGA-85563 is 50 Ω, so no matching network is required at that port. Pin 6 outputs the RF through the DC blocking capacitor C_C, while the DC

(voltage-based) bias is also injected into pin 6. The RFC blocks the RF from entering V_{CC}, and the two capacitors, C_B, bypass any RF that makes it through the RFC, while also filtering power supply EMI from entering the MMIC. Pin 4 of the MGA-85563 can be utilized to increase the IP3 at the RF output by increasing the MMIC's bias current from its normal 15 mA up to 35 mA. Since this mode obviously consumes more current, it is only used when higher output powers are required. With pin 4 left floating, the device will have an IP3 of +12 dBm, while an R_B of 15 Ω will cause an I_d of 30 mA, and raise the IP3 to +17 dBm.

3.8 Wideband Amplifiers

3.8.1 Introduction

A wideband amplifier is designed to have an extremely broad band of frequencies that it can pass with flat gain and a good return loss response, along with perfect stability.

3.8.2 Wideband Amplifier Stability

To properly design a wideband RF amplifier we must suppress the lower frequencies where the gain is the highest (Fig. 3.118*a*). One way to do this is by showing these low frequencies a poor impedance match, while with the higher frequencies, where the gain is much less, we can supply a perfect match. Using these methods will flatten the gain of the amplifier (Fig. 3.118*b*), but will not provide a high return loss at all frequencies. In other words, matching the transistor at high frequencies will supply more gain at these particular frequencies, while mismatching at the lower end of the spectrum will decrease their gain due to mismatch losses. However, stability is always of prime importance, and since we now have various impedance mismatches across a very wide bandpass, we will want to be assured of stability at *all* frequencies. We can do this by checking with our supplied *S*-parameters, and see if the small-signal active device will be unconditionally stable at all measured frequencies. This will not, however, include the very lowest of frequencies, since these may well be below the measurements available within the *S*-parameter files.

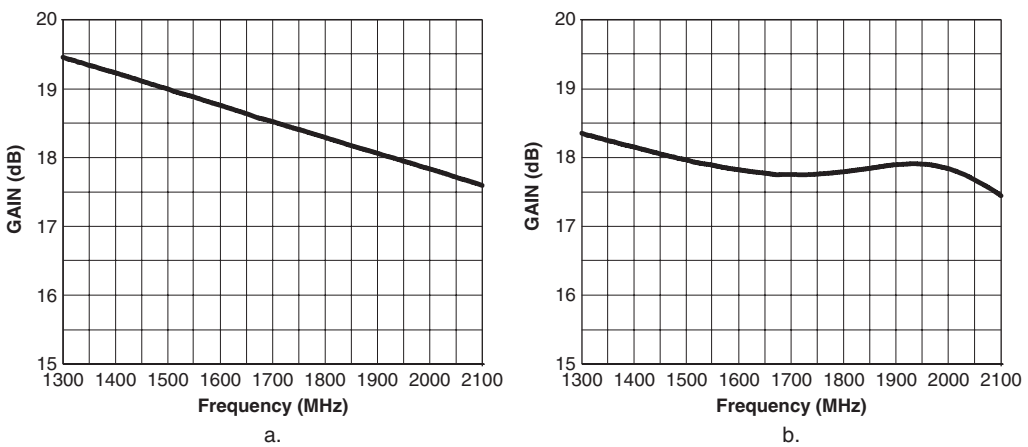


Figure 3.118 Gain flattening of an RF amplifier: (*a*) not compensated; (*b*) compensated.

Huge instability problems internal to a wideband amplifier can occur at frequencies between 1 and 20 MHz, where the transistor's gain may be as high as 40 dB. When combined with even the slightest of internal or external in-phase feedback, this level of gain is an obvious recipe for instability and oscillations. These oscillations are viewable on a spectrum analyzer as a single carrier surrounded by sidebands, with the injected carrier modulated by the low-frequency oscillations. We must find a way to lessen either the gain or the feedback, or both, of the unstable power amplifier at low frequencies. Two helpful methods of accomplishing this are by selecting a transistor with a low h_{FE}, and/or by using the lowest value of output collector choke we can that will still supply a short for low-frequency AC, but a virtual open to RF (Fig. 3.119). The choke itself should be paralleled by a low-value resistor *Rp* of between 300 to 600-ohms for de-Q'ing purposes, which will help prevent parasitic oscillations. Either a high-value capacitor must be attached to the top of the choke, or to the top of a low-value R_S, to shunt any of these lower frequencies to ground.

Another extremely potent method of decreasing low-frequency gain, which is especially suitable for wideband power amplifiers, is by negative feedback (Fig. 3.120). The capacitor *C* in this circuit is adopted to block the DC bias, while easily allowing the dangerous low-frequency AC to pass back to the base. The resistor *R* element controls the amount of feedback to the base, and must be found empirically (50 to 500 Ω will be the typical values). The inductor *L* is the vital component that actually controls the feedback. It functions by having a very high reactance to the amplifier's desired frequencies, so as not permit degenerative feedback that would lower the gain levels at higher frequencies. However, as the frequency is decreased, the reactance of the inductor will, of course, also decrease. This action increases the level of degenerative feedback to the transistor's base, decreasing the gain of any low-frequency signals.

Feedback of any kind may sound dangerous, but a common-emitter amplifier will have a perfect 180° phase shift at its lower frequencies of operation ($<<f_T$), thus assuring that *regenerative*, or oscillatory, feedback will not occur there. The only caution would be

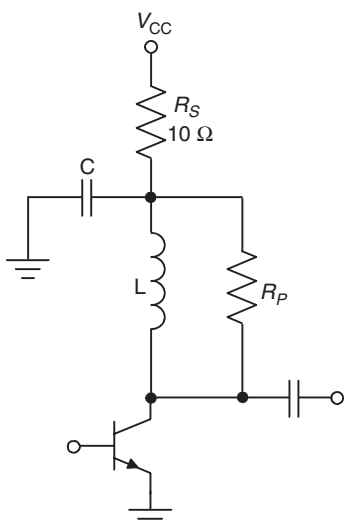


Figure 3.119 A collector inductor load for decreasing low-frequency oscillations.

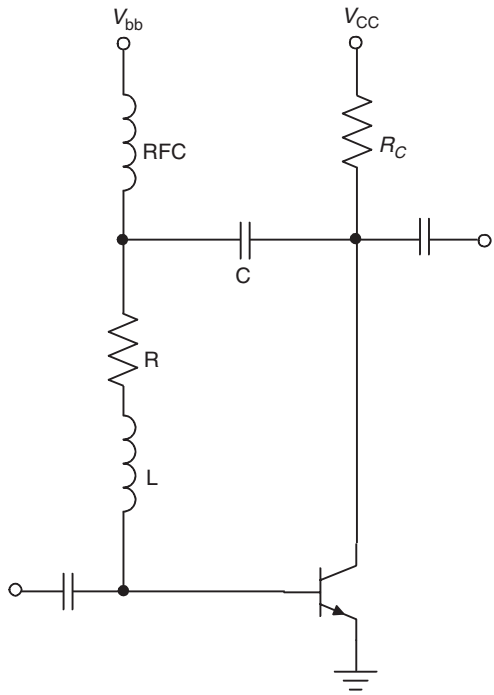


Figure 3.120 A negative feedback circuit.

to confirm that the inductor is actually capable of blocking the RF at its highest frequency, and without allowing the RF to pass through due to its internal parasitic capacitances.

If we keep in mind that almost any amplifier will be completely stable if all of the frequencies it passes will see a perfect 50-Ω resistive impedance at both the input and output of the stage (a difficult objective with many power amplifiers), then we can use this knowledge to our advantage. Excellent stability, even at frequencies above or below our interests, can be assured by supplying the active device with a good 50-Ω source and load. This can be in the form of a 51-Ω resistor that will only be seen by the unwanted lower frequencies through a one-pole lowpass filter, or by the use of a diplexer, or by employing a 2-dB 50-Ω pad at the amplifier's output.

Another viable wideband amplifier design technique is by utilizing resistive components to create a match over the entire bandwidth. This design is referred to as a *resistive negative feedback amplifier*, and can be exploited at almost any frequency, but is more common at 600 MHz and below where gain is relatively inexpensive. This technique has the disadvantage of producing much lower gain as compared to the typical LC matching method, along with supplying less reverse isolation and a higher NF. But its greatest advantage is that it has a very wide bandwidth, with a decent return loss, and much increased stability. As shown in the wideband design example, L_F is a peaking coil that decreases the negative feedback at higher frequencies, holding up the gain of the amplifier just as it begins to fall. However, the insertion loss will not be quite as good at these particular peaked frequencies.

Another technique to increase the stable high-frequency gain, while maintaining a flat, wide bandwidth, is to use a low value of bypass capacitor that in parallel with the

amplifier's emitter resistor. This will increase the emitter resistor's degenerative feedback as the frequencies are decreased, leveling out the gain. The application of emitter components of any kind is viable only up to 2 GHz and below, due to the stability-robbing presence of their added lead inductance.

3.8.3 Wideband Amplifier Design

A stable wideband Class A amplifier can be quickly designed by following these simple steps.

Wideband Resistive Feedback Small-Signal Amplifier (Fig. 3.121)

To Design

1. Select a suitable transistor for the desired frequency range, gain, NF, cost, package, and the availability of *S*-parameter files at different bias levels.
2. Select the appropriate bias for the transistor in your application by studying the transistor's data sheet. For instance, a transistor for an LNA may need a collector current of 5 mA (or less), while a transistor for other uses may require a higher level of bias.
3. Insert the transistor's *S*-parameter model, which should be at the closest expected bias level of the device to be used, into the simulator.

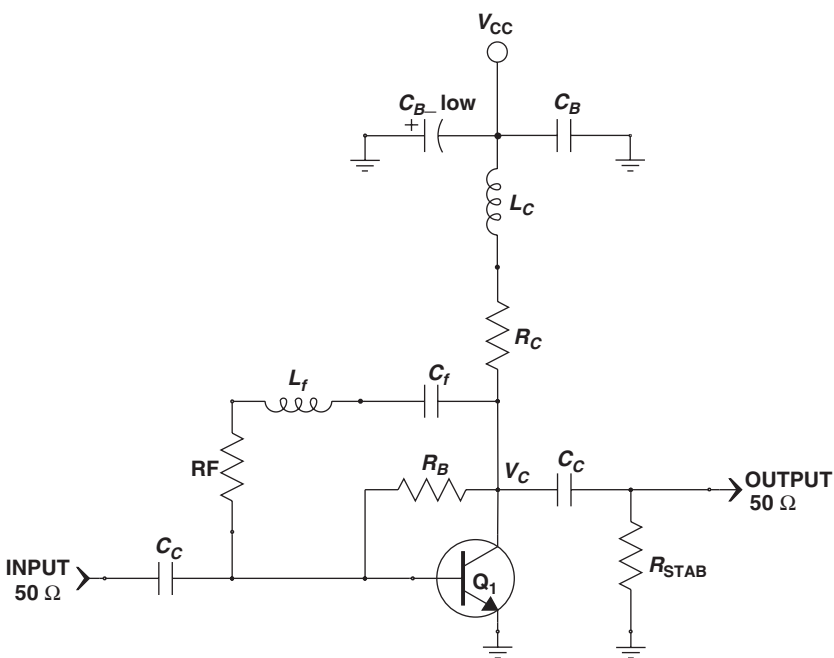


Figure 3.121 Resistive wideband RF amplifier with 50-Ω input and output.

4. Design C_f and L_f to be series resonant at, or just below, the lowest frequency in the band of interest.
5. R_f will be adjusted empirically for the appropriate feedback to optimize gain, return loss, and stability.
6. All coupling (C_C) and RF bypass (C_B) capacitors will have a reactance across the entire band of less than 3 Ω. L_C is an RF choke, required only if R_C is less than 500 Ω.
7. Design the amplifier's DC bias network for a stable temperature operation of the transistor over its specified temperature range, as detailed elsewhere in this chapter.
8. Run the circuit in the simulator, and check for stability all the way from the lowest to the very highest frequencies available within the *S*-Parameter file. If there are any frequencies that have a *K* that is less than 1, or a B_1 below 0, then stabilize the transistor, while sacrificing gain and NF as little as possible. We can accomplish this by inserting and varying the value of a resistor that is first placed in series, and then in shunt, at the transistor's collector output port. Perform the same action at the device's input port, all the while checking for improvements in *K*. In another window, confirm that gain and return loss are nor being too adversely affected, but only across our band of interest. The shunt resistive components will typically range between 100 and 500 Ω, while the series resistors will extend from 3 to 10 Ω. Most active devices will be stabilized with a single output resistor. (Any shunt stabilization circuit must have a DC block if it will short the bias to ground).

A Quick Example Design a Wideband Feedback RF Amplifier (Fig. 3.122)

Goal: Create a stable wideband 50-Ω Class A RF amplifier with LCR feedback and collector bias. The specifications and parameters for the amplifier are:

$V_{CC} = 3.3$ V
$I_C = 12$ mA
$V_{CE} = 2$ V
$S_{21} = 21$ dB
Transistor = NXP BFG425 ($H_{FE(LOW)} = 50$, $H_{FE(MID)} = 80$, $H_{FE(HIGH)} = 120$)
Passband = 430 to 930 MHz

One possible solution:

1. C_C,$C_b = 100$ pF
2. $R_B = 8.2$ kΩ
3. $R_C = 110$ Ω
4. $R_f = 960$ Ω
5. $R_{STAB} = 130$ Ω
6. $L_f = 150$ nH
7. $L_C = 160$ nH
8. $C_f = 1.4$ pF

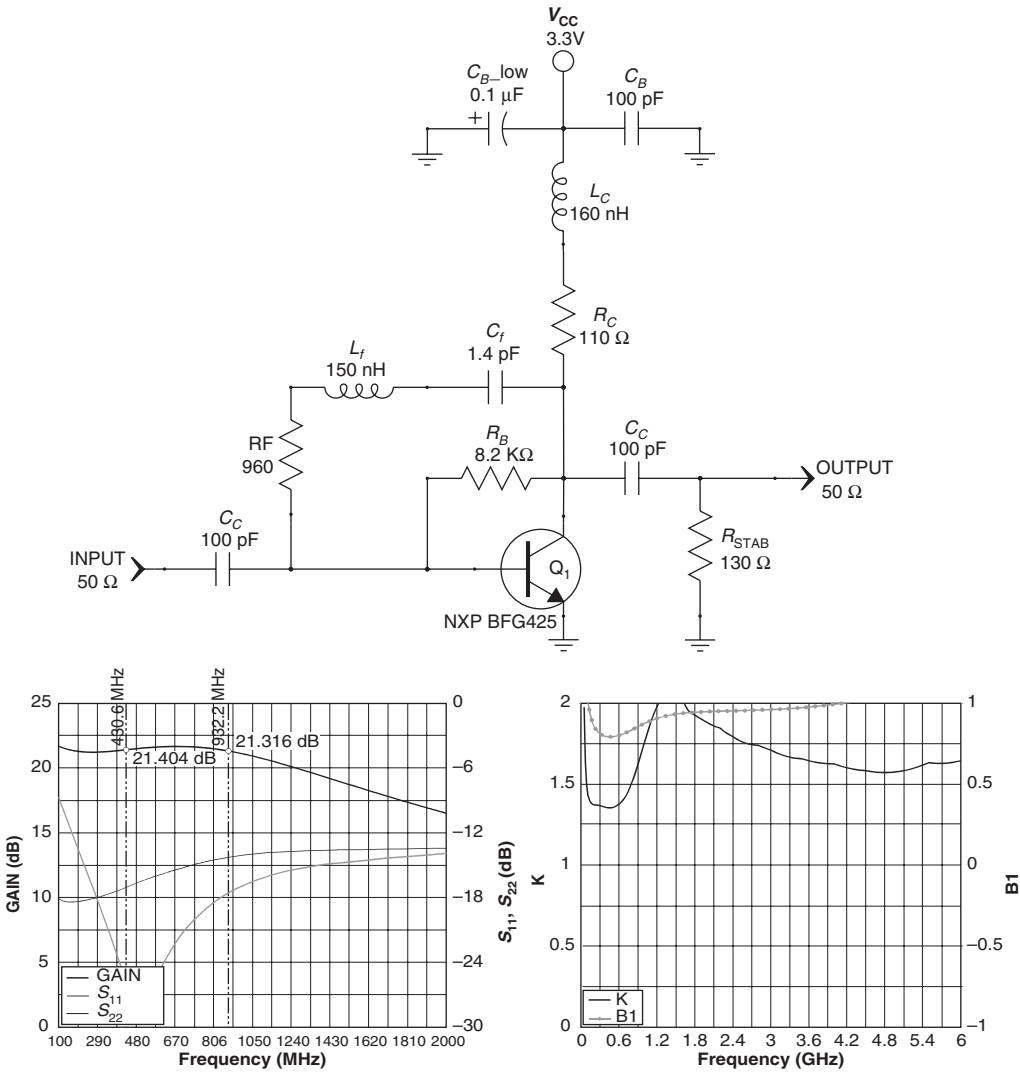


Figure 3.122 A complete example Class A wideband amplifier with bias, matching, and stabilization circuits.

3.9 Parallel Amplifiers

3.9.1 Introduction

Single-ended amplifier configurations cannot always supply us with all the RF power we need, since we may require up to several hundred watts of output power for certain applications. This can be accomplished with RF parallel amplifiers (Fig. 3.123). While the design of such ultra high-powered amplifiers is beyond the focus of this book, lower-powered MMIC and discrete amplifiers are relatively easy to realize (<4 W).

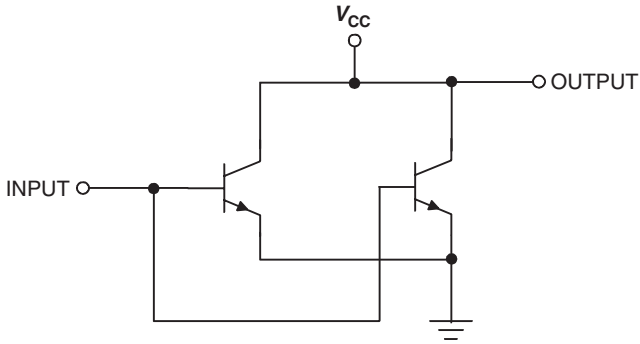


Figure 3.123 A parallel amplifier circuit without bias components.

With twin parallel amplifiers, each transistor is either on or off at the *same* time period, unlike push-pull, which sequentially distributes the power back and forth for equal, but different, time periods. Since the output current of a parallel amplifier circuit is shared evenly between the transistors when they are perfectly matched, this action will double the amplifier's power handling capabilities ($P = I \cdot E$) as compared to a single-ended configuration. Parallel amplifiers allow the entire circuit to function as if it were a single high-powered transistor, and at any bias desired (Class A, AB, B, and C). Figure 3.124 demonstrates one type of parallel amplifier circuit, along with impedance matching and Class C biasing.

Gain will stay the same whether a single power amplifier or a parallel power amplifier configuration is used. The real advantage of paralleling amplifiers is that the RF output power *capability* (P1dB) will increase by 3 dB for two stages (Fig. 3.125), and by 6 dB for four. However, the input drive power into these paralleled stages must also increase to take advantage of this attribute. Also, each amplifier must have excellently matched active devices, and the input and output capacitances of two paralleled amplifiers will be double that of a single device, which can be problematic with high-frequency operation.

Directly paralleling discrete high-frequency power transistors makes their already very low input/output impedances even lower, so effective low-loss matching can become a problem. With these types of discrete designs, it is better to first match each active device, preferably to 100 Ω in a twin parallel amplifier, and then directly combine. Or, we can match each active device to 50 Ω, and then utilize a combiner to blend the single amplifiers into a complete parallel amplifier stage.

Formerly, indiscriminant direct paralleling of active devices would lead to uneven current distribution among the transistors, and the thermal destruction of the transistor with the most current draw. Due to emitter ballast resistors that are placed internally within the semiconductor die itself, this is not a problem with the vast majority of modern power transistors.

3.9.2 Parallel Amplifier Design with MMICs

The input and output impedances of the combined amplifiers will decrease as more devices are paralleled, as stated above. Therefore, when working with 50 Ω MMICs, the Z_{IN} and Z_{OUT} of the total paralleled stages can be computed by $50/n$, with n being the

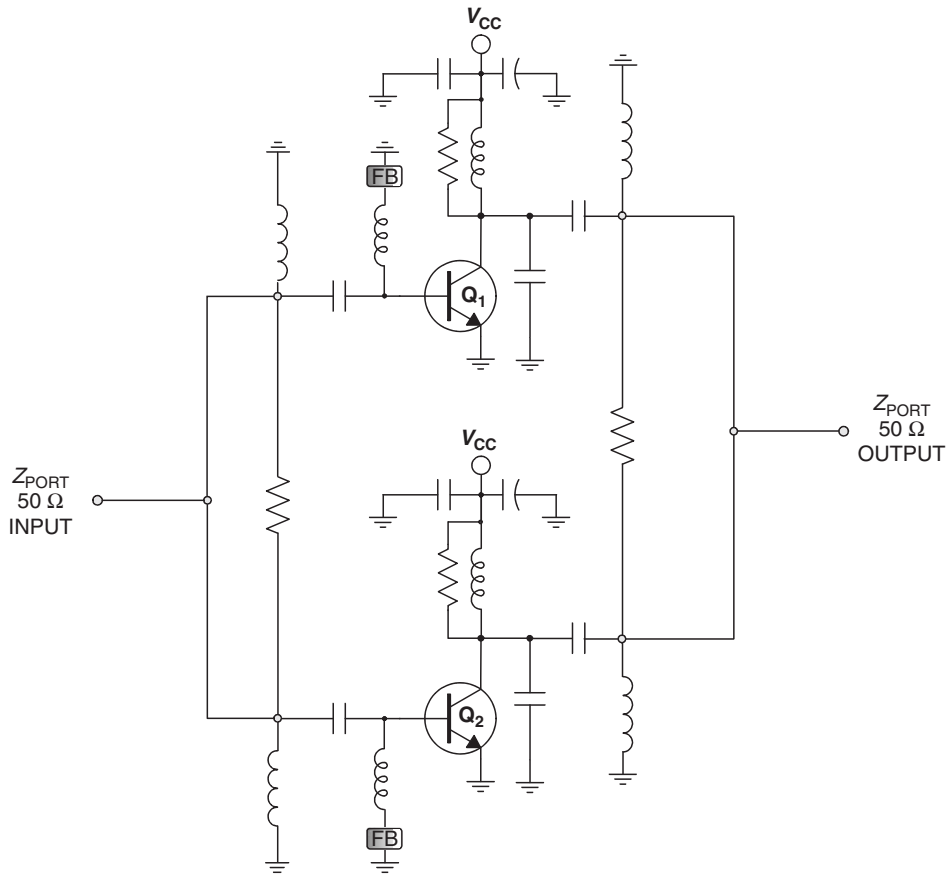


Figure 3.124 A parallel power amplifier with bias components, Class C.

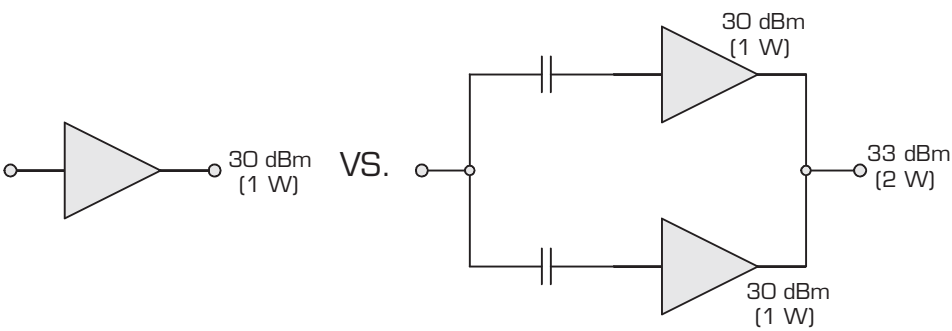


Figure 3.125 The output power of a single MMIC as compared to two paralled MMICs.

number of amplifiers in parallel. As a result, a matching network for the MMICs would be necessary if we must integrate them into a 50-Ω system.

Paralleling two MMICs will increase their P1dB by 3 dB as compared to a single device, and will maintain the MMIC's input and output impedances by using a lumped *Wilkinson power divider/combiner network*, along with a single DC bias line. The Wilkinson network, unfortunately, will also make the amplifiers much more narrowband, which may not be desired in very broadband applications. And since MMICs are not always completely resistive, and each lumped circuit element and layout has its own parasitic reactances, some tuning will be required in order to peak the gain at the desired frequency of operation.

A Parallel MMIC PA Amplifier Circuit (Fig. 3.126)

To Design

1. Bias as instructed under Sec. 3.7.2.
2. Couple as in Sec. 3.7.4.
3. $L = \dfrac{50}{1.4\pi f}$
4. $C = \dfrac{1}{2.83\pi f 50}$
5. $C_2 = 2 \cdot C$
6. $R = 2 \times Z_{\text{PORT}} = 100\ \Omega$
7. $C_C = < 1\ \Omega$

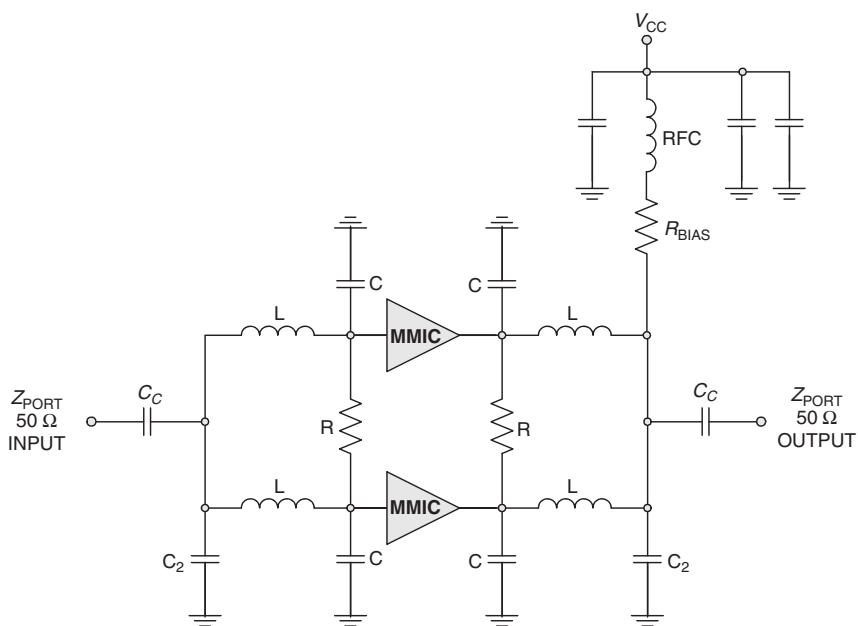


Figure 3.126 Paralleled MMIC amplifiers with splitter and combiner.

3.10 Audio Amplifiers

3.10.1 Introduction

To maximize the output voltage of a cascaded system, many non-RF, low-frequency circuits need to be terminated in a high-load impedance. We also speak of voltage or current gain, instead of power gain, in such applications, since the cascaded impedances are rarely the same.

Many non-power audio amplifier stages need not be matched to their source nor to their load, since conjugate matching is used to maximize power transfer between stages and to reduce standing waves, audio amplifiers are much more concerned with reducing distortions and isolating each stage from the effects of the next. An acceptable single stage audio voltage amplifier can be designed using the low-frequency bias design formulas as presented under Sec. 3.6.5.

Operational amplifiers are far more common for voice frequency amplification for both low-level voltage signals, as well as high-level power signals. They can be selected from many manufacturers, and are obtainable in an optimized single-voltage supply package for ease of biasing.

3.10.2 Audio Amplifier Design

The National LM386 is a low-voltage audio amplifier that is perfect for low-frequency amplification. In voice-band radios this IC can amplify the input signal all the way from the detector stage to the 8-Ω speaker or headphones. This device has very low quiescent current drain (4 mA), accepts a wide variation of V_{CC}'s (4 to 12 V), has adjustable voltage gain (20 to 200), adequate distortion levels (<10 THD), and a output driving power of 700 mW with a 9-V supply into 8 Ω.

IC Audio Amplifier (Fig. 3.127)

To Design

1. C_B will bypass any RF that escaped the detector, while it can also be chosen to limit the audio frequency response of the amplifier. (Many demanding cases will require either an active op-amp or a passive RC lowpass or bandpass filter

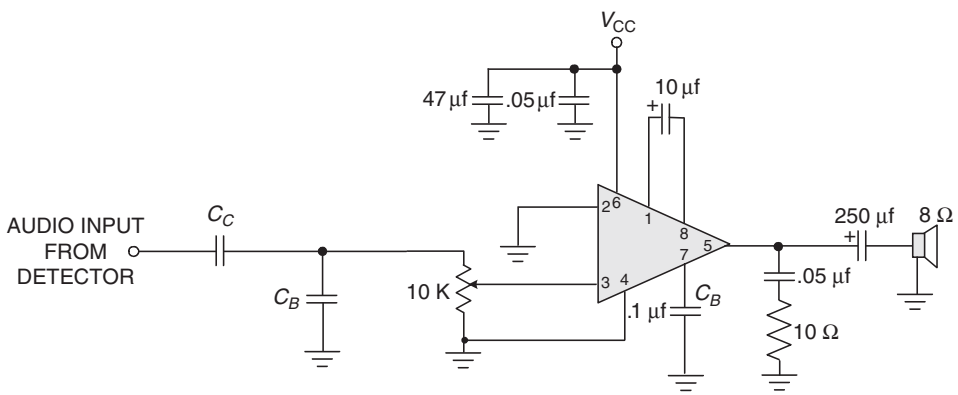


Figure 3.127 An integrated circuit audio power amplifier.

placed before the amplifier in order to limit the frequency response and noise even further, since voice should be band-limited to between 300 and 2500 Hz and will improve the higher-frequency noise and heterodyne outputs, as well as any low frequency 60 and 120 Hz hum.)

2. The gain for the amplifier as shown will be 200, but can be adjusted downward by the addition of a resistor in series with the capacitor between pins 1 and 8 (e.g., a 1.2-kΩ resistor will set the gain to 50). Removing the capacitor, and leaving pins 1 and 8 open entirely, will decrease the gain to 20.
3. Adjusting the pot at the input to the LM386 will alter the amplitude of the output signal into the speaker.

3.11 VGA Amplifiers

3.11.1 Introduction

VGAs (*variable gain amplifiers*) can be created in one of two ways: either by varying a transistor's bias voltage to its base, which controls its collector current, and thus the gain of the device (see Sec. 8.3, *Automatic Gain Control*); or by placing a voltage or current controlled variable attenuator at the input to a fixed gain amplifier. Since the latter circuit usually results in a more linear amplifier response over gain, especially with large input signals, it is preferred over the variable bias design in many applications. An added disadvantage to the variable bias type is that any modification to the bias of a transistor will also alter its *S*-parameters. This means that not only will the gain be varied, but so will the return loss and stability of the amplifier. Any loss of stage stability is simply not an option, while the change in return loss can prove problematic if the VGA is attached to a filter circuit, since a filter's response is dependant on its source and load impedance.

3.11.2 VGA Amplifier Design

The following are three simple designs for RF variable gain amplifiers, the first being a positive intrinsic negative (PIN) diode attenuator type, and the second a stage bias-control type, and the third a MMIC type.

A VGA with Low Distortion for 10 MHz and Above (Fig. 3.128)

A non-reflective type of VGA is designed by employing an absorptive attenuator design in front of a fixed gain discrete or MMIC amplifier. An automatic gain control AGC voltage of zero at the attenuator's DC control input will result in low gain (even a negative gain), while a control voltage of greater than zero results in a steadily increasing gain. The return loss will remain quite usable up to very high attenuation levels. However, a rise in the NF of the VGA circuit (whether bias or attenuator controlled) is unavoidable as the gain is decreased.

To Design

1. Bias the transistor as described in Sec. 3.6.5.
2. Match the input and the output of the transistor as described in Sec. 3.4.6.
3. Design the 50-Ω attenuator pad as described in Sec. 8.4.2. (Whether the attenuator is placed at the amplifier's input or output port depends on the required NF for an LNA, or the saturated drive requirements of a PA).

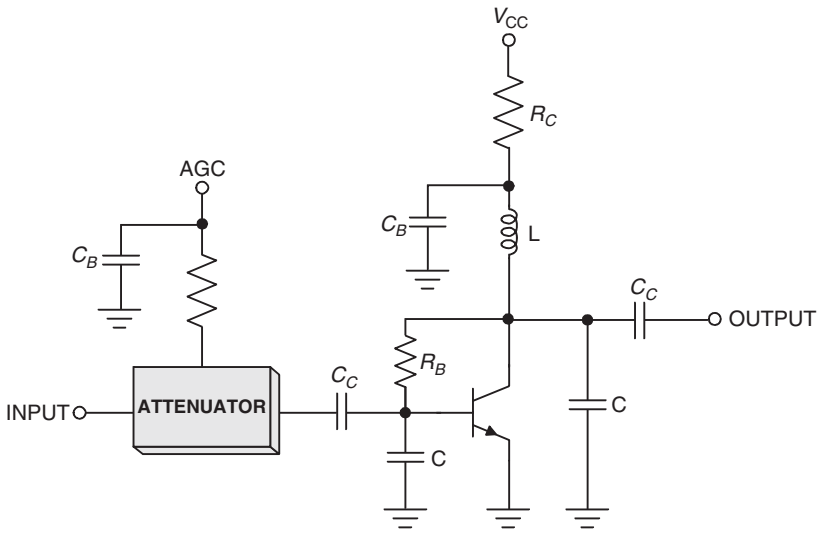


Figure 3.128 A simple and stable VGA circuit.

Low-Cost Variable-Bias VGA, Reverse Gain Control (Fig. 3.129)

To Design

1. Choose R_C to drop half the V_{CC} when the transistor is at its desired gain:

$$R_C = \frac{\left(\frac{Vcc}{2}\right)}{I_C}$$

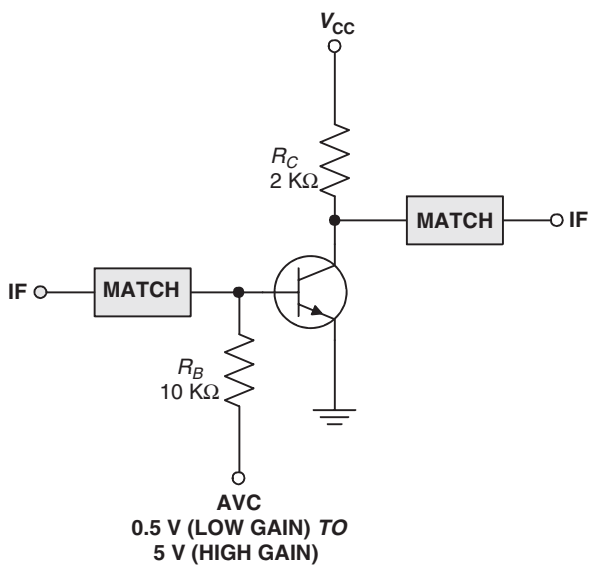


Figure 3.129 A circuit for a variable bias gain amplifier.

2. Choose an R_B of 10 kΩ.
3. Find the voltage required at the AGC (AVC) port of the amplifier that causes the base current (I_B) to create a collector current for the desired full gain. The transistor's *characteristic curves* in its data sheet will contain the information on I_B versus I_C, while the *current gain* graphs will show the I_C versus h_{FE}:

$$\text{AVC} = (I_B \cdot 10\text{K}) + 0.7$$

4. Choose the limits of AVC voltages that will supply the required range of gains by substituting the desired I_B in the above equation with the minimum and maximum I_B related gain values.
5. Design the transistor's input/output matching networks for the desired bias at maximum gain and for the frequency of operation.
6. The gain of the circuit is now controlled completely by the AGC voltage at the transistor's base. But since gain is managed by altering the I_C through this AGC at the transistor's base, then the input and output impedances, as well as the stage's stability, will also vary. Distortion and gain compression may also occur with strong input signals.

MMIC VGA (Fig. 3.130)

To Design

1. Most MMIC gain is moderately affected by a change in I_d. By looking at the device's I_d versus S_{21}(dB) curves, this susceptibility can readily be seen, and offers an easy way to operate some MMIC's as variable gain amplifiers. Gain variations of 5 to 15 dB are possible, depending on the MMIC and the frequency, by varying I_d through an AGC circuit.

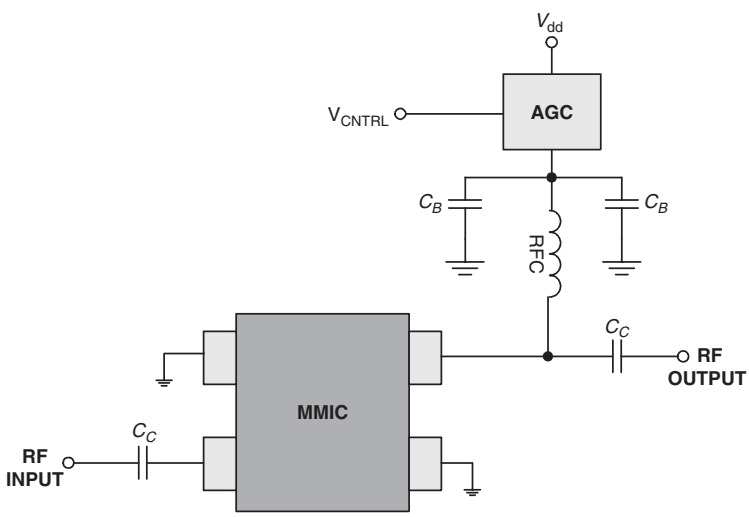


Figure 3.130 A MMIC based variable gain amplifier circuit.

2. Utilizing a MMIC as a VGA should be adopted only for low-level signals, since the P1dB will decrease along with the I_d and gain of the MMIC. The exact value of the overall circuit gain variations obtained will differ with input frequency.
3. Design the MMIC stage as described under Sec. 3.7.3, and then permit the AGC circuit to control the MMIC's V_{dd}.

3.12 Coupling and Decoupling of Amplifiers

3.12.1 Introduction

To prevent one stage's DC biasing from affecting the next stage, a method of *coupling* an AC signal into or out of an amplifier, while blocking DC, must be employed. The actual type of coupling for discrete circuits, sometimes combined with impedance matching, will depend on the sort of signal being amplified: DC, (low frequency) AC, RF, or wideband amplification.

To prevent unwanted amplifier oscillations and/or RF contamination of the system power supply, *decoupling* components, such as capacitors and chokes, that are capable of shunting the entire spectrum of RF and AC frequencies should be used for all active devices.

3.12.2 Coupling and Decoupling Circuit Design

Decoupling

Heavy decoupling of a power supply is not only required in order to prevent RF energy from entering the supply, but also to provide a reservoir of energy that may be instantly used by any active circuit on the PCB, especially high current stages such as a PA. This requirement can be better appreciated if one understands that when a 2.4-GHz power amplifier is in operation, it is actually drawing DC current directly from the power supply at a frequency of 2.4 GHz. This alternating current draw must be filtered before it can adversely affect other stages, as well as to provide instant current smoothly to the PA stage. Thus, the decoupling capacitors themselves will act as a vital charge supply for the PA, as well as function as an RF bandstop filter for any conducted EMI.

Decoupling networks must not only be comprised of RF ceramic capacitors that are capable of attenuating higher frequencies, but also of the electrolytic type, which possess a much higher capacitance value for the filtering of extremely low frequencies. In other words, so that we may send *all* frequencies to ground that may attempt to enter, or to exit, the DC power supply, the supply must be strongly RF isolated (decoupled) from the PA by a wide frequency band of various shunt capacitors of different values. Series inductors and resistors help in this filtering action as well, and can be used in concert with the decoupling capacitors.

To function as desired, decoupling capacitors must not be used near their *parallel* (high impedance) resonant mode, while decoupling inductors must not be used near their *series* (low impedance) resonant modes. If they are, then they may literally "disappear" from the decoupling circuit, and become completely ineffective.

Inductors are far from perfect components, and possess parasitic capacitances. So when an amplifier must function properly across a very wide band of frequencies, two decoupling RFCs may be required; a low inductance coil that works at very high

frequencies (without encountering any series resonances), and a high inductance coil used to block the lower frequencies. The use of two chokes can be necessary because the low frequency, high impedance inductor will begin to pass the high RF frequencies through the natural turn-to-turn capacitance inherent in any coil, but it will do so far earlier than the smaller, minimal parasitic RF choke.

To minimize the lower-frequency inductor's (L_2) significant parasitics from adversely affecting the MMIC's output, the first choke up from the amplifier stage should always be the RF inductor (L_1), followed by L_2. An additional small value RF bypass capacitor to ground can be placed between L_1 and L_2 to further decouple any RF from the power supply.

This type of wideband decoupling will permit the amplifier's complete passband to enjoy a nearly flat gain response over its entire frequency range.

Coupling

There are various coupling techniques that can be used between stages, depending on frequency, cost, performance, and impedance matching needs.

Capacitor coupling (Fig. 3.131), also referred to as *RC coupling*, is found in AC and RF amplifiers only, and is capable of amplifying over a very wide bandwidth (the amplifier's required impedance matching circuit will limit this bandwidth, however). As shown in the figure, the series coupling capacitor C_C blocks the DC bias to the next stage, but allows the RF signal to pass through unattenuated. C_C and R_6 form a voltage divider, allowing most of the RF signal to be dropped across the high resistance of R_6 located at the input to the next stage. The voltage divider functions as described because the capacitor has a much lower impedance to the RF than does the resistor. This signal across R_6 will then add to or subtract from the second stage's emitter-base junction, forcing its collector current to vary through R_7, producing an amplified output voltage.

RC coupling is a simple method to transfer energy from one circuit to another, but has great difficulty matching the stage's impedances and, unless we employ a low-value

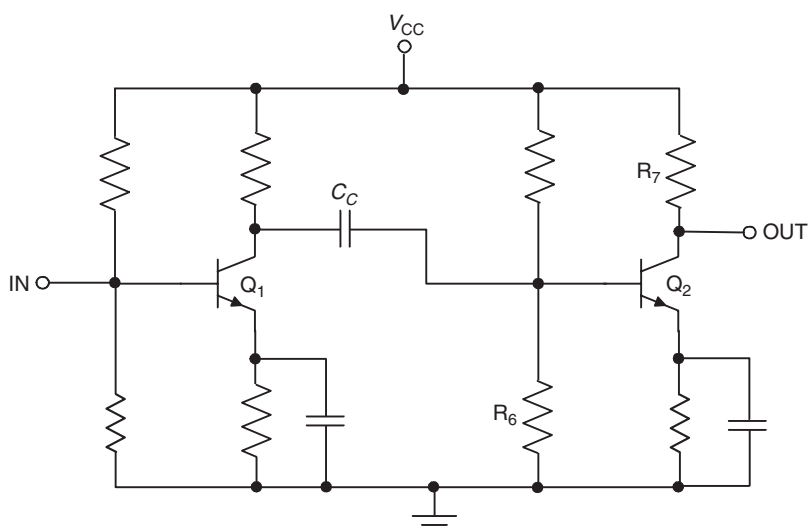


Figure 3.131 Capacitive coupling between two stages.

series resonant coupling capacitor at the stage's input and output, this coupling method does not help to attenuate harmonics from being transferred from stage to stage, and from being further amplified. The small, series resonant coupling capacitor will also assist in stabilizing the amplifier chain, which it accomplishes by somewhat attenuating the lower and higher (undesired) RF frequencies with its low capacitance value and series resonance operation, while easily passing the frequencies of interest. A L, T, or PI network is normally added for harmonic attenuation and impedance matching requirements. The coupling capacitor can be part of the matching circuit itself.

Inductive coupling (Fig. 3.132), also referred to as *impedance coupling*, is found in AC and RF circuits, and is comparable to RC coupling. However, instead of exploiting a resistor in the collector circuit, it uses a collector inductor.

Inductive coupling has the advantage in that the collector inductor wastes little DC power due to its very small DC series resistance, thus permitting for more efficient amplifier operation. This high-value inductor works as a transistor's collector load because of its high reactance to the alternating collector current, which produces an AC voltage drop. This action will subtract from or add to the voltage from the transistor's emitter-collector.

Inductive coupling is only practical over a relatively narrowband of frequencies, since X_L changes directly with frequency, and stage gain would vary as well.

Direct coupling (Fig. 3.133), also referred to as *DC coupling*, is valuable for very low frequency and DC amplification. R_3 functions as a collector resistor for Q_1, and as a base resistor for Q_2, and must be carefully chosen to function in both roles, since a small temperature induced current change in Q_1 is directly amplified by Q_2. Precision components and tight placement of parts to allow each component the same changes in temperature must be used to stabilize this circuit.

Low-frequency transformer coupling (Fig. 3.134), employing laminated iron cores, can be adopted for low-frequency AC amplifiers. This method, due to the interstage transformer, will not pass the DC bias, and can be used to match the relatively high

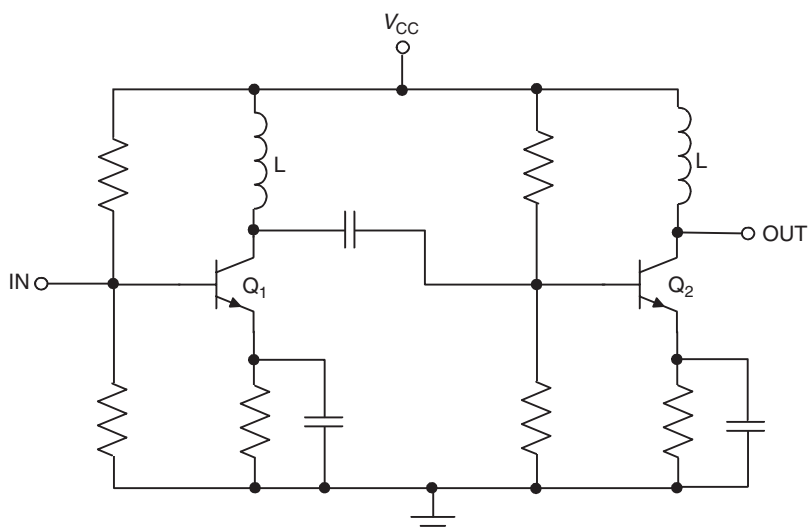


Figure 3.132 Inductive coupling between two stages.

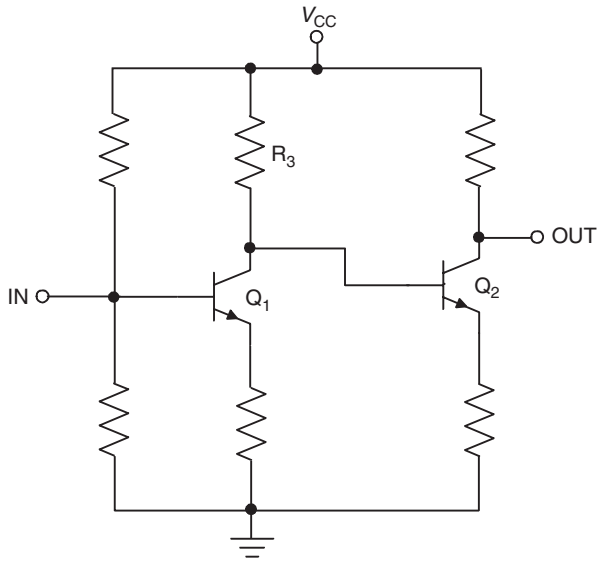


Figure 3.133 Direct coupling between two stages.

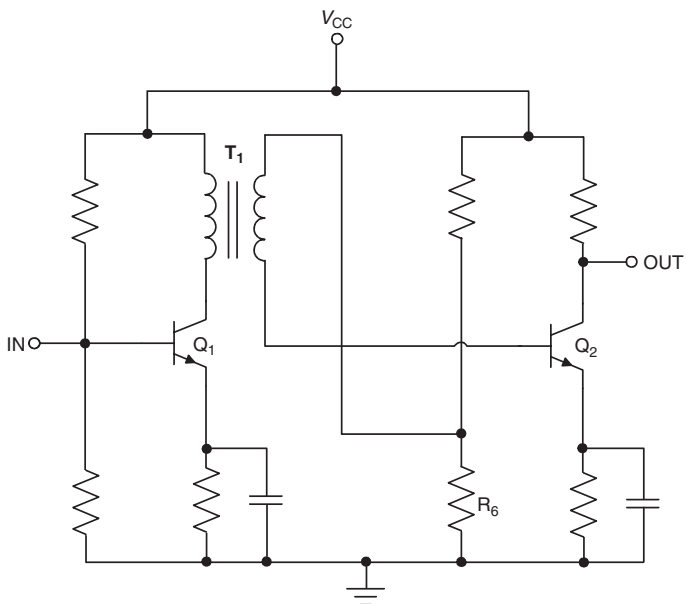


Figure 3.134 Low-frequency transformer coupling between two stages.

output impedance of Q_1 to the low input impedance of Q_2. One end of the transformer's secondary is connected to the base of Q_2, while the other end is connected to the top of Q_2's bias resistor R_6. The signal being amplified by Q_1, and being output from T_1, will then subtract-from or add-to the base bias. This results in a varying base current, which causes a much higher in amplitude collector current; creating amplification at the

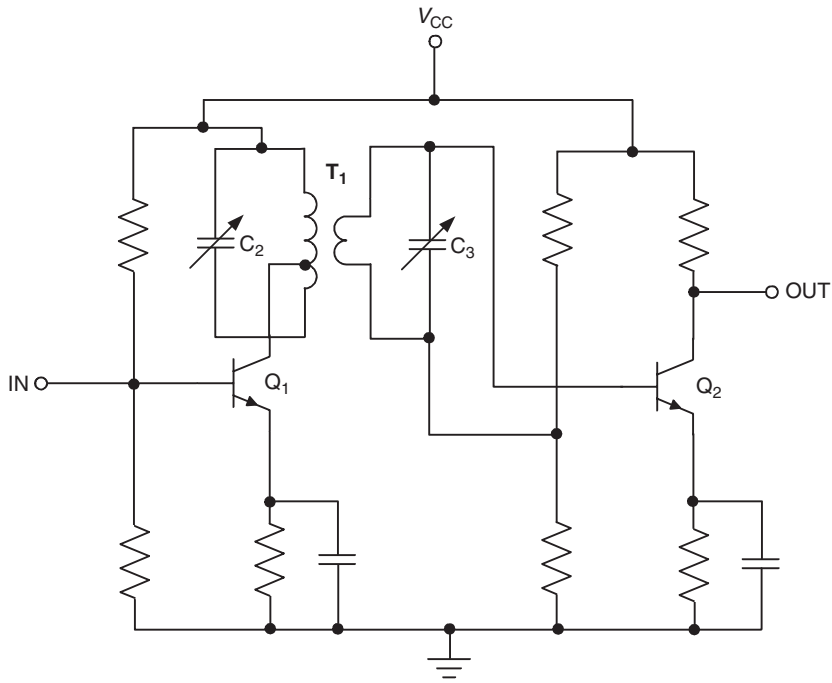


Figure 3.135 High-frequency transformer coupling between two stages.

output of Q_2. However, since a low-frequency iron-core transformer is both expensive and heavy, this coupling method is rarely found today in most consumer-grade audio applications.

High-frequency transformer coupling (Fig. 3.135) with RF tuned circuits, employing ferrite, powdered iron, and air cores, is not nearly as popular as it once was due to the expense and size of the transformer, but can still be found in some RF and IF amplifiers up to a maximum frequency of 250 MHz.

Transformers provide the required impedance match for the efficient and maximum power transfer between amplifier stages, as well as block the DC bias from stage to stage. These transformers function the same as the low-frequency iron-core transformers above, except for the frequency selective narrowband resonant tanks formed by C_2 and the primary of T_1, as well as C_3 and the secondary of T_1.

The degree of coupling between a tuned transformer's primary and secondary, which is mainly governed by the distance between the windings, will affect the signal's amplitude and bandwidth as it passes through the transformer. Indeed, as the *coefficient of coupling* increases (*over coupling*, Fig. 3.136*a*) or, in other words, as the windings are brought closer together, more flux lines from the primary will cut the secondary. This will produce a higher output voltage and a wider bandwidth over that of the *loose coupling* of Fig. 3.136*c*. The wider bandwidth is caused by the high capacitance now present between the closely spaced primary and secondary, while the high signal amplitude is due to the increased flux lines that cut the secondary. However, as the coefficient of coupling is decreased toward loose coupling, the amplitude and the

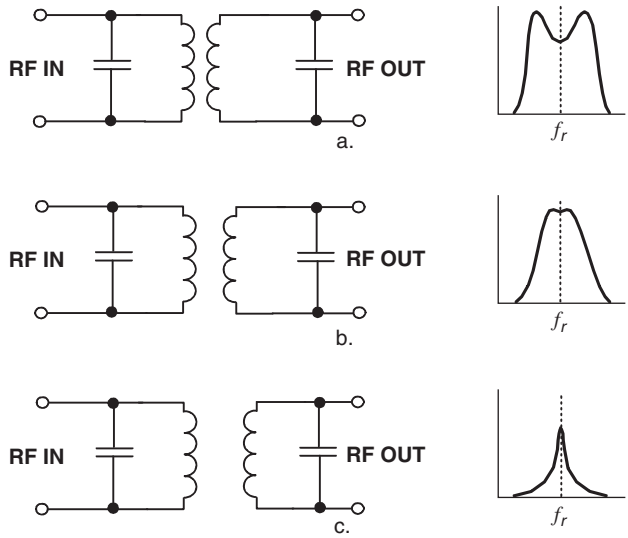


Figure 3.136 The degree of coupling and the effect on the output signal.

bandwidth of the signal diminishes. Nonetheless, loose coupling can be used to lower the capacitive coupling into the next stage, thus lowering harmonic output, and give a narrower bandwidth that may be required for certain applications. For typical narrowband uses, *optimum coupling* (Fig. 3.136*b*) will be found as a good compromise between bandwidth and amplitude.

This page intentionally left blank

CHAPTER 4
Oscillator Design

Oscillator design is perhaps one of the least understood practices of wireless engineering in general, next to RF filter design, and is considered to be among the most complex. Indeed, until Randall W. Rhea released his groundbreaking book *Oscillator Design & Computer Simulation* in 1990, oscillator design was strictly a hit or miss affair for many engineers. And as anyone in the field of RF is well aware, it is quite easy to design an oscillator; just design a poor amplifier and turn on the power, and it will probably begin to oscillate. The real problem, however, is to design an oscillator that will not only oscillate at a desired frequency and amplitude, but that will start reliably and not wander, that will not be plagued with various spurious responses and harmonics, that will not be excessively affected by normal changes in temperature, and that will be consistent in operation when built during a long production run.

This chapter will concentrate on the design, simulation, and verification of LC, crystal, and VCO oscillators over a wide range of frequencies.

4.1 Oscillator Basics

4.1.1 Introduction

Below is a quick refresher on basic LC and crystal oscillator theory, oscillator starting, noise, components, and decoupling, as well as general oscillator design considerations and the importance of bias and loaded *Q*.

4.1.2 Oscillator Operation

When a pulse is applied to a tank circuit it will ring at the tank's resonant frequency, creating a decaying sinusoidal wave (Fig. 4.1). But if amplification from an active device, such as a transistor, is used to amplify and sustain this output, then an oscillator can be formed. The natural resonant frequency of the tank circuit is established by the tank's *L* and *C* components, or:

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

Thus, oscillators will use a small part of their output signal from the active device in order to send a regenerative, or in-phase, feedback signal into their own input. This will create a continuous oscillation, with the transistor constantly amplifying its own feedback.

Considering the typical oscillator functions by feeding a 180° out-of-phase signal back to its input, with this phase shift caused by the common-emitter configuration of

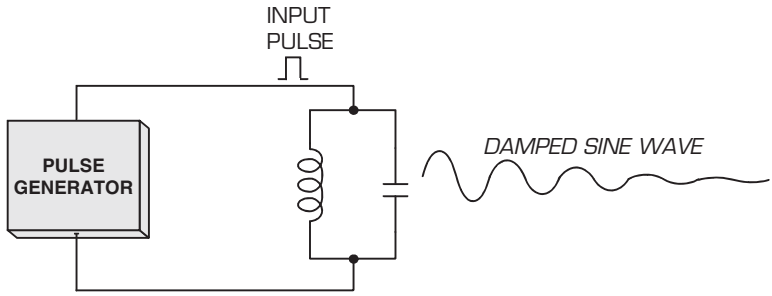


Figure 4.1 A damped sine wave output of a tuned tank after insertion of a single pulse.

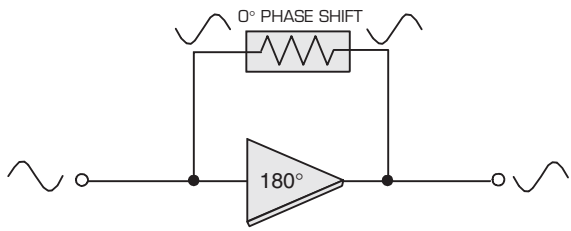


Figure 4.2 An amplifier with degenerative feedback cannot oscillate.

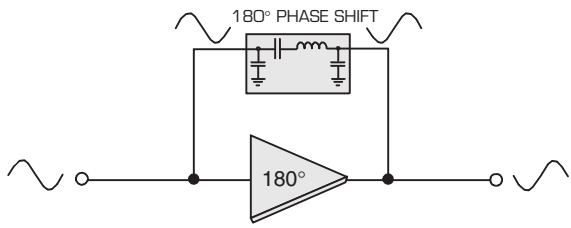


Figure 4.3 An amplifier with regenerative feedback can oscillate.

the oscillator's own amplifier (Fig. 4.2), we will obviously require a method to shift this out-of-phase output signal back to 0°, in order to obtain the necessary regenerative feedback (Fig. 4.3). Utilizing the reactance of inductors and capacitors to carry out this phase shifting, which sets the frequency of oscillation, is the easiest way to construct an RF oscillator.

An oscillator is self starting, and must be very reliable in this regard. A typical Class A sine wave oscillator starts by the following mechanisms:

1. Power is applied to the oscillator's active device.
2. Noise or transient turn-on voltages cause the oscillator to start, beginning the low-power output of sinusoidal waves, after which an oscillator is just translating its DC input power into output sinusoidal oscillations.
3. The sinusoids build to a very high level, which cause saturation of the active device and surplus loop gain is then dissipated.
4. The oscillator generates sinusoidal waves of a stable frequency and amplitude.

The output power of most such oscillators will be around 0 dB to 10 dBm, and will be biased at Class A or AB in common-emitter configuration (some higher frequency oscillators are common-base, however), though a few may be biased at Class C.

4.1.3 Oscillator Design Issues

Oscillator biasing of its amplifier section is employed for multiple reasons: To allow the use of a single V_{CC}; to positions the bias point for a certain class of operation; to swamp out any device variations in beta; and to stabilize the active device over wide temperature variations.

A vital parameter of any oscillator circuit will be its loaded *Q*. A high-*Q* feedback oscillator (such as the crystal or SAW types) will have a much more frequency stable RF output than a simple, low-*Q* LC oscillator. This is because variances in the transistor's reactances, triggered by a changing V_{CC} and temperature as well as lot variations from transistor to transistor, will cause far higher frequency shifts in a low-*Q* (LC) oscillator than an oscillator with a high-loaded *Q*. Thus, improving the loaded *Q* of an oscillator improves its frequency stability and phase-noise performance over temperature, V_{CC} variations, and changes in load impedance.

An oscillator is started by the voltage "kick" introduced by the act of simply turning on the oscillator's power supply, and not necessarily by noise. The speed at which an oscillator actually starts and produces a stable output frequency is much more dependent on the time it takes to charge up its bias capacitors, as well as the RC time constant of the V_{CC} supply itself, than on the loaded *Q*.

The proper choice of each component in an oscillator is very important, since even the passive components can have a significant impact on proper oscillator operation. Thus, unless frequency compensation is desired, the oscillator's feedback network capacitors should be of the *NPO* type for minimum frequency drift under normal temperature variations. The proper choice of the active device is also critical. Transistors with a very high f_T as compared to the oscillation frequency work much better in oscillator circuits than active devices with marginal f_T specifications. This is due to the transistor's ability, at a high f_T ($> 5 \times f_r$), to not only maintain its 180° phase shift at higher frequencies—since an amplifier's phase shift begins to drop from 180° as frequency increases—but to also have a higher feedback gain at f_r.

A shunt-C-coupled first-order bandpass filter (or *resonator*, Fig. 4.4) will supply a 180° phase shift and bandpass filtering, and so it is capable of forcing the LC oscillator's frequency of operation (Fig. 4.5) to a specific value. In other words, the resonator supplies a peak S_{21} *and* 180° phase shift for the desired oscillation frequency, and by using only a single inductor and three capacitors.

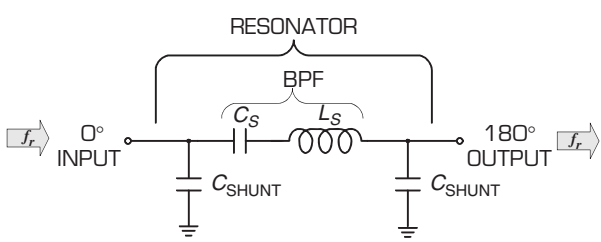


Figure 4.4 A resonator circuit, with the proper 180° phase shift from its input to output port.

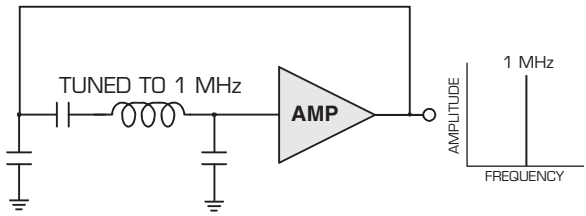


Figure 4.5 The resonantor circuit keeping oscillations on frequency.

Without the shunt capacitive coupling on either side of the series *L* and *C* elements of the resonator circuit, such a filter would simply be series resonant (X_C and X_L would cancel), and there would be no phase shift possible. This is not what we require.

Most oscillators should be heavily decoupled from any general noise and intermittent voltage variations of the power supply, and to avoid the injection of the oscillator's own RF into the system's common power supply.

4.2 Oscillator Simulation Methods

4.2.1 Introduction

The following oscillator design and simulation procedures, as popularized by *Rhea* and *Matthys*, have made oscillator design a simple and far more repeatable process than in the past. Formally, it was either an unpleasantly mathematically intensive procedure, with an uncertain outcome, or simply involved copying another particular oscillator design and empirically, via educated guessing, continually swapping out the LCR components until the oscillator functioned as close as possible to the desired specifications.

As with any design of nonlinear circuits, such as PAs and mixers, much of the final design phase must be optimized on the bench, since the simulations alone will rarely be exactly spot-on due to nonlinear active device model inaccuracies—or the models themselves may be completely unavailable.

If simulating in Berkeley Spice, instead of using the more expensive and effective RF harmonic balance or linear simulation programs, all component, trace, and general PCB parasitics will not have been accounted for, nor will the various critical phase shifts caused by trace lengths. Indeed, any *S*-parameter RF program, such as *Qucs* or *Eagleware*, will normally have the proper (linear) models available for most passive parts, as well as for the active small-signal devices; and with the added advantage over regular Berkeley Spice programs in that a linear simulator will immediately indicate whether the open loop of the oscillator's input and output ports are properly matched to each other. They can do this by displaying a Smith chart tool, which will confirm that both the input (S_{11}) and the output (S_{22}) impedances of the open loop are matched at the frequency of interest, which is critical to the accuracy of oscillator open-loop design.

A quick breakdown of the normal procedures in designing an oscillator in open loop are outlined below, and will be discussed in much more detail within this chapter:

1. Design the oscillator on paper following the specific equations given.
2. Optimize the oscillator circuit in an RF linear simulator as an ideal open-loop circuit, and without using any microstrip or passive *S*-parameter models.

3. Insert all of the microstrip traces and *S*-parameter models for all active and passive parts, duplicating all of the above tuned values, and optimize oscillator again as an open-loop circuit.
4. If a nonlinear simulator is available (if not, skip to step 5) duplicate the exact circuit layout and passive values as found in step 3, optimizing the oscillator in the nonlinear simulator as a *closed-loop* circuit, and by using the appropriate nonlinear transistor model. (Some harmonic balance software may require a technique to kick-start an oscillator design for a successful simulation, since harmonic balance simulators are used to display a steady-state condition only. This start-up technique may involve injecting a low-amplitude pulse into the oscillator's resonator, at a very high impedance, at near the frequency we expect the oscillator to operate.)
5. Lay out the oscillator's PCB with the appropriate software, and have the board fabricated and the oscillator assembled.
6. Perform physical bench tuning and testing on the fully realized circuit for final optimization of all oscillator parameters.

Can we skip any of these above steps and still be successful? Certainly: Before computers became widely available to the average engineer, oscillators were designed with paper, pencil, and slide rule, and then fabricated, assembled, and forced into submission using only the test bench. Obviously, the simulation phase of the design process was skipped entirely.

4.2.2 Oscillator Open-Loop Design and Simulation

In order to oscillate, a circuit must:

1. Have a net gain around the loop that is equal to 1 or more (or 0 dB).
2. Have a phase shift around the loop that is equal 0°.

We can validate both these requirements in our design by using open-loop linear simulation techniques, as completely described in this section.

Software open-loop design of an oscillator involves cutting the feedback loop of the oscillator from the transistor's output back to the resonant phase-shifting network (Fig. 4.6). We then insert a software tool called a *Bode plotter* within this open loop, tune the circuit for optimal performance, close the loop, and a fully functioning oscillator will now have been completed.

Indeed, much of the preliminary design optimization of an oscillator can be assisted by RF software programs, such as the included *Qucs* linear simulator, or the high-end

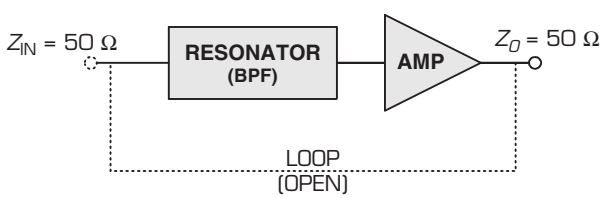


Figure 4.6 The open input and output ports of a feedback oscillator as required for simulation.

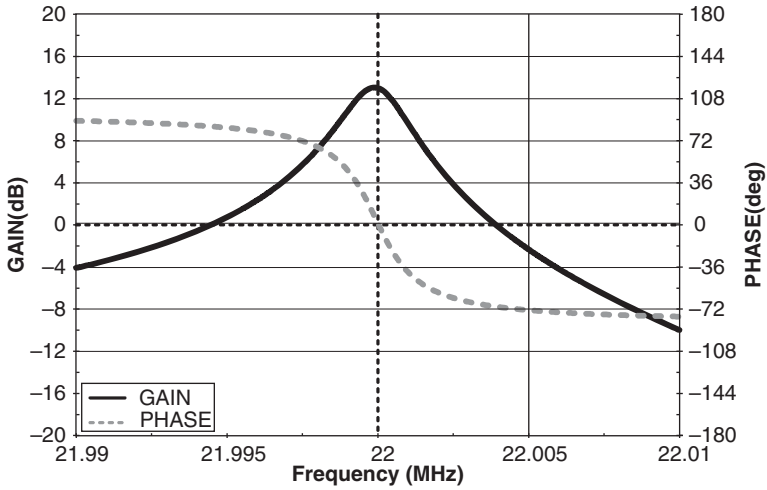


Figure 4.7 Bode plot of oscillator gain and phase.

Genesys simulator available from Agilent. Within either of these two programs, we can display the gain and phase of a signal as it passes through the oscillator's open-loop circuits (Fig. 4.7) by inserting a reference signal into the input of the circuit while sweeping through a range of frequencies. This reference signal is considered to be at zero gain and zero phase shift. Therefore any gain, either positive or negative, or any phase shift that occurs to this input frequency after it passes through the circuit, will be read in the RF software graphing window, and displayed as *frequency versus gain* and *frequency versus phase shift* in decibels and degrees. This permits us to view what happens to an RF signal at the output port of the circuit at the same instant that the input of this same circuit is swept in frequency at a constant amplitude and phase. In a nutshell, we can now see the circuit's effect on the gain and phase of a signal *after* it passes through an amplifier, filter, or open-loop oscillator.

With an *S*-parameter linear software program, we cannot only view the effect a particular circuit has on the gain (S_{21}) and phase (ANG[S_{21}]) of an inserted RF signal, but we can also observe the input (S_{11}) and output (S_{22}) return losses, the reverse gain (S_{12}), and the input and output impedances.

For accurate gain and phase responses under open-loop linear simulation, the input and output impedances of the circuit should be at some common, but real, value such as 50 Ω. This is an important concept, since it is not always possible to obtain a common input/output impedance at each port of an open-loop oscillator. And so that we may equal the true and actual input/output impedance of the open-loop oscillator circuit if we do find we have a common impedance at each port, then we must set the linear program's *S*-parameter impedances to duplicate this value. (The oscillator's projected open-loop input and output impedances will always be indicated in the design procedures of this book). If these terminating port impedances of the oscillator circuit and simulator were not taken into account, the gain and phase margins displayed on the simulator's graphs would be incorrect, as would the *Q*, but the resonant frequency peak shown would remain relatively unaffected.

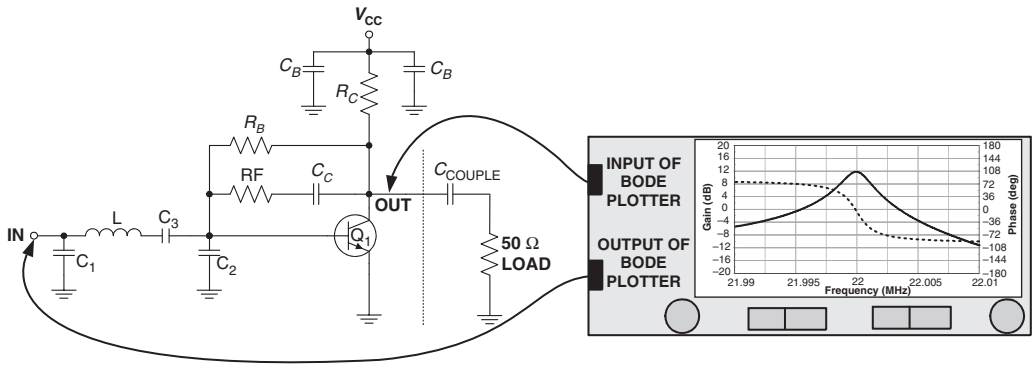


FIGURE 4.8 Virtual Bode plotter inserted into an oscillator's open loop.

The effective methods commonly used to shift the oscillator's ports to a common impedance is by using a small value of un-bypassed resistor in series with the emitter, by changing our collector-to-base RF feedback part's values, or by employing LC port matching networks.

Figure 4.8 demonstrates how to best analyze an open-loop oscillator with a linear simulator. By injecting a signal into the oscillator's input and checking the phase and gain at the oscillator's output, we will have a very good indication that our particular design is valid. This is accomplished, as described above, by breaking the feedback loop of the oscillator, and attaching the software's 50-Ω ports between the broken input/output points of the oscillator. To obtain the proper results, set the frequency and phase of the software's graph to *linear*, adjust the magnitude to display a gain of −20 to +20 dB, set the display to show phase values from −180° to +180°, and adjust the frequency sweep to approximately ±25% of the expected oscillation frequency (narrow or widen as necessary to achieve the display as shown in the figure). This open-loop response test is a good indication that the oscillator will function as intended, since the RF software is outputting a 0° phase-angle signal, at the frequencies of interest, directly into the input of the oscillator's resonator, which then changes this 0° phase by 180° before it even reaches the input of the transistor. The transistor, being in common-emitter configuration, changes the phase by another 180°, making for a total change of 360°, or 0°, for regenerative oscillatory feedback. The proper phase change, at the appropriate amplitude, can be confirmed on the graph as shown in Fig. 4.9. The graph is displaying the maximum gain peak at the frequency of the desired oscillation, which should occur at the same frequency as the phase trace when it crosses 0° from the output to the input of the oscillator (in order to sustain oscillatory feedback). To minimize noise, the S_{21} peak and the 0° phase crossing should correspond closely.

The gain trace, at its maximum amplitude, is called the *gain margin* when it is located at the same frequency as the point that the phase trace crosses the 0° phase point on the graph, and is measured in decibels.

The higher the gain margin the more tolerance the oscillator will have and still be able to start (or continue to oscillate) when components on the assembly line vary in tolerance, or the load varies in impedance. Temperature will also have far less of a deleterious effect with this higher gain margin, with 6 dB being considered optimal. The loop gain should be no more than this, however, for harmonic minimization, as well as for minimizing the phase-shift change that occurs when the oscillator is finally

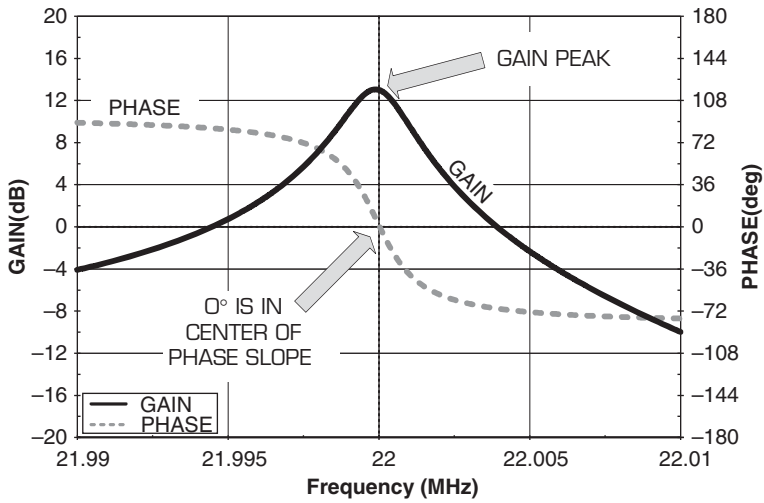


Figure 4.9 Perfect Bode plot of a correctly designed 22-MHz oscillator.

run closed loop in nonlinear mode. However, a value of no less than 3 dB should always be used to ensure reliable operation, or oscillator start-up over temperature, load, and part's variations will become erratic and/or slow for the completed closed-loop oscillator. In fact, if an oscillator has a sufficiently high gain margin, closing the loop should only cause a minor shift in the RF design frequency, with the high open-loop gain being reduced to unity when the oscillator actually reaches its steady state.

When simulating the open-loop oscillator, not only should the gain peak be at the point where the phase is zero, but it should also be as close to the center of the phase slope as possible in order to maintain the oscillator's long-term stability and low-noise characteristics. The amount of excess phase above or below this center of the phase slope is referred to as the *phase margin*, and is as important as the gain margin.

To adjust the zero crossing point of the phase to be more centered in the middle of the maximum phase slope, try varying the series V_{CC} inductor's value, or swap out the V_{CC} resistor with an inductor. The zero crossing point of the phase is critical in an oscillator, since it is not the gain peak itself that sets the oscillator's frequency, but this transmission phase zero crossover point which, in the real world, may not always be at the exact same spot as the maximum gain peak.

After the linear *S*-parameter open-loop simulation procedure is successfully accomplished, then the oscillator's loop may be closed (Fig. 4.10), and RF energy may then be tapped from the oscillator and placed into a load. This tapped energy, however, will decrease the oscillator's own available loop feedback.

Figure 4.11 is shown using a series C_{COUPLE} to remove energy from the oscillator's output port and into the load. Coupling out the oscillator's RF energy, without decreasing its feedback to excessively low levels, will be discussed later in this chapter. C_{COUPLE} may be either a series X_C or X_L of approximately 100 Ω that is placed at the oscillator's output, with a 50-Ω load attached, and a simulation attempted with the Spice oscilloscope and FFT tool. These software tools are connected across the 50-Ω load to confirm proper oscillation frequency, amplitude, start-up, harmonics, and so on. [Sometimes a Spice frequency source must be included somewhere within the Spice

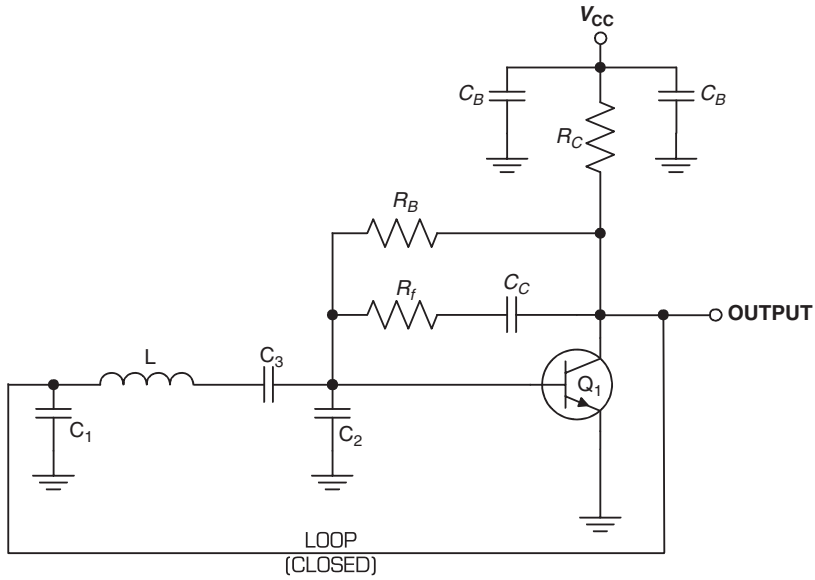


Figure 4.10 Loop closed after Bode analysis for a functioning oscillator.

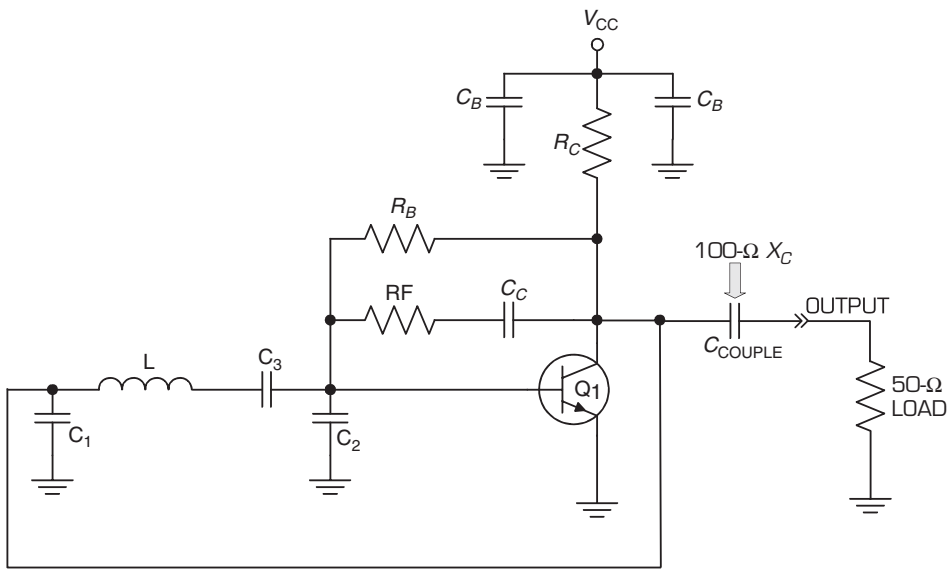


Figure 4.11 A high impedance output capacitor, C_{COUPLE}, used to extract energy from the oscillator circuit and into a load.

simulation, or the oscillator may not start. If so, simply attach the Spice frequency source to the input of the oscillator through a large-value (1 MΩ) series resistor to "fool" the simulated circuit into starting. Also, some Spice simulators on older machines may take 5 to 20 min of computer time for the oscillator to start and reach full amplitude, so be patient when employing such time-domain software].

Since the loaded Q of an oscillator will govern its phase noise and frequency drift, then we see that the higher the Q the more stable the oscillator is over temperature, along with possessing lower phase-noise figures. The loaded Q of an oscillator can be measured by employing the open-loop simulation method as outlined above, along with the following formula (this formula is only accurate if the phase slope crosses zero at the center of its drop, which is an optimum, or high phase-margin condition):

$$Q_L = \frac{f_o}{3\text{ dB BW}}$$

where Q_L = loaded Q of the open-loop oscillator, in dimensionless units
f_o = center frequency of the oscillator, Hz
3 dB BW = bandwidth of the oscillator's gain (S_{21}) at its half-power points, Hz

For a high quality and stable oscillator design, it is evident from this formula that the highest loaded Q depends on the bandwidth of the open-loop oscillator being narrow within the gain response plot.

The loaded Q must never be permitted to degrade below 5 or 10, and should preferably be much higher, in order to stabilize the LC oscillator and to lower its phase noise. By increasing the oscillator's RF power output, with a higher transistor bias current, we can also decrease these phase-noise levels, since the carrier will now be at a higher relative amplitude above the noise.

Because open-loop oscillator design accuracy depends on both ends of the oscillator loop being at the same impedance, as well as both of the resistive (real) terminating impedances in the linear simulator being equal to this same value, then ignoring the cascaded input and output impedances of the oscillator will result in a non-optimized circuit design, as discussed. Yet the input and output port's open-loop impedances need only have a return loss of 8 dB or so for an acceptably accurate open-loop analysis result. Indeed, even lower return loss values will still give us relatively accurate results, with gain and resonant frequency values being off by only about 5% for frequency, and a few decibels for gain, when the return losses are as low as 2 dB.

NOTE: *When utilizing open-loop oscillator design, it is assumed that the open loop is stable. In other words, the amplifier section (with bias) should not be unstable, since it is only when the loop is closed from input to output that our desired oscillations are meant to occur. Proper frequency stability may become quite erratic with an unstable amplifier section.*

Crystal Oscillator Simulation Issues

When simulating a crystal RF oscillator, we must first select the proper crystal by obtaining certain vital parameters (Fig. 4.12) from the crystal manufacturer, such as its *motional capacitance* (C_M), *motional inductance* (L_M), *series resistance* (R_M), and *parallel plate capacitance* (C_p or C_o) for the desired frequency of operation, holder type, and quartz cut (typically AT). The manufacturer will also need to be informed if the crystal is to be utilized in a series or parallel resonance oscillator (see *Pierce Oscillator Design for 600 kHz to 30 MHz* under Sec. 4.5.3), whether the crystal is to be run on its fundamental or on one of its overtone frequencies, the crystal's required aging specification in ppm/year, initial frequency accuracy in ppm, and frequency accuracy over temperature in ppm.

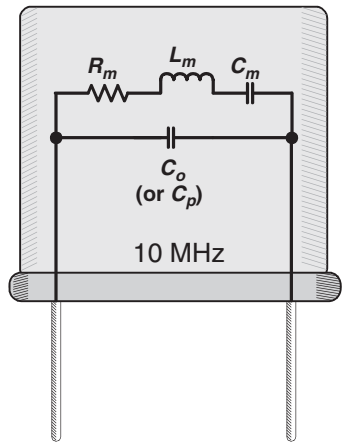


Figure 4.12 Equivalent internal structure of a crystal.

Since many RF linear simulation software packages may not necessarily have crystal models available, we can model the crystal as shown in Fig. 4.12, and place it where the crystal would be within the oscillator circuit. This equivalent LCR model, while simplistic, is more than adequate to effectively represent a typical crystal using the manufacturer's above motional specifications of L_m, C_m, R_m, and C_p.

Note that if a crystal's holder or its package are changed, as might be required when redesigning for a smaller oscillator, it may have an effect on the motional properties of the crystal, and a new simulation will then have to be performed with these latest values.

4.3 Low Phase-Noise Oscillator Design Techniques

4.3.1 Introduction

We can internally minimize all-important phase noise in an oscillator's output (Fig. 4.13) by following the guidelines below. Some of these guidelines may conflict with other low phase-noise design techniques presented, or diverge from the required oscillator specifications, so a compromise will have to be made.

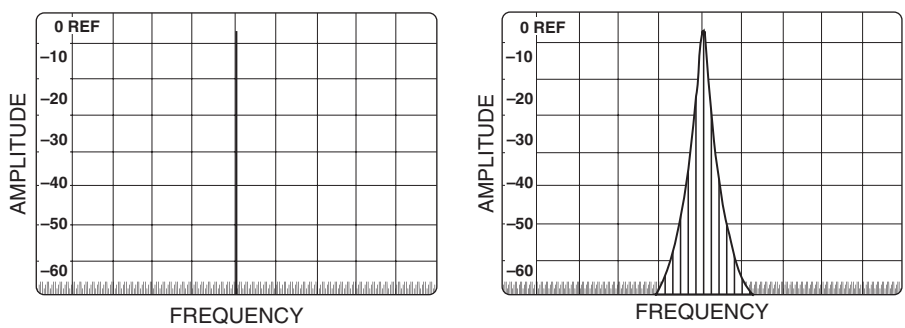


Figure 4.13 A single CW tone both without and with phase noise.

4.3.2 Low Phase-Noise Oscillator Guidelines

1. Use the least sensitive (MHz/V) varactor that will work in the particular application to lessen varactor noise modulation (see also item 7).
2. When an oscillator must function below 40 MHz, use only BJT transistors, as opposed to GaAs FETs, since BJT's will produce extremely low flicker noise.
3. Output the highest RF power possible from the oscillator, without damaging the transistor or excessively violating item 11 or 13 below.
4. Use only very high unloaded-Q capacitors and inductors in a VCO's resonator circuit.
5. Force the VCO oscillator tank resonator's loaded Q as high as possible by using the highest unloaded-Q varactors that are practical for the application.
6. Further increase loaded Q of the resonator circuit by using a high L/C ratio for series resonant VCOs, and high C/L ratios for parallel tank VCOs.
7. Strongly LC filter the varactor's DC control line in order to minimize noise modulating the varactor, and to minimize any noise coupling from anywhere into this very sensitive control line. (Any noise, glitches, or ripple on the VCO's voltage control line will cause the VCO's frequency output to vary, losing frequency stability and creating FM noise.)
8. Use only low noise figure (NF) transistors.
9. To extract the oscillator's RF energy, place its L or C output tap between the resonator circuit and the input of the transistor. (This helps to reduce phase noise due to the resonator's bandpass filtering action.)
10. To prevent reflections from entering the oscillator's output, the oscillator must only be permitted to see its own particular design impedance, and must also be fully isolated from any impedance fluctuations within the oscillator's load. (This specific effect can be one of the major causes of phase noise.)
11. Minimize the oscillator's transistor bias current as much as possible, since this can substantially lower the $1/f$ (*flicker*) noise frequency of the active device, as well as optimize the BJT's noise figure.
12. Flicker noise, and thus some of the close-in phase noise, can be further mitigated by adding a small value negative feedback resistor of around 10 Ω in series with the emitter lead, which must not be bypassed by a emitter capacitor. (However, to decrease the oscillator's other noise outputs irrespective of $1/f$ noise, a transistor should have a biasing network that does *not* use any emitter resistor—similar to standard LNA design procedures. Test the circuit with and without a resistor to see which offers less noise for your particular application.)
13. Avoid driving the oscillator's transistor too far into compression.
14. Pay particular attention to the oscillator's PCB, since poor general board layout and inferior oscillator isolation from other circuits can cause noise energy to couple into these sensitive circuits.
15. Keep the varactor's input V_{TUNE} voltage as high as possible, since the Q of the varactor diode is higher at an increased bias tuning voltage. (The Q of a varactor drops at lower bias voltages due to increased resistances of the diode at its maximum capacitance.)

16. Maintain very strong LC noise filtering at the VCO's own power supply, and across a very wide range of frequencies from low AC to high RF. (Also see item 7.)
17. If VCO output buffers are required, use only low noise types.
18. Decrease the value of the varactor's high value 10 to 50 kΩ bias resistor down to 100 Ω, and only employ a decoupling series inductor and shunt capacitors as the bias line's major noise and RF filters. (The change to LC filtering for low-noise VCOs is because any varactor bias resistor can be considered as a very real noise source that is being placed at the input terminal of the extremely sensitive tuning diode, which causes the oscillator to be frequency noise modulated, thus producing phase noise due to the varactor's MHz/V sensitivity.)

NOTE: *Another noise source that afflicts VCOs is caused by the tuning varactor's own inherent equivalent noise resistance. This particular noise contributor, out of the many that plague VCO's, is unfortunately little affected by improving the resonator or varactor's* Q *values. Therefore, a varactor diode will not only increases the phase noise of a VCO through its own low* Q*, but also due to the diode's own internal generation of noise voltages that modulate the VCO's RF output, creating further phase noise, with the severity of the phase noise being dependent on the VCO's sensitivity (in MHz/V; with the precise diode sensitivity value reliant on the particular DC tuning voltage present). Thus, high VCO tuning sensitivity results in high phase noise owing to this mechanism, as well as due to the externally coupled EM energy and the DC varactor control line noise voltage contributions. Depending on many factors, this can contribute as much as 25 dBc/Hz (measured at 20-kHz offset from the carrier) degradation in a VCO's phase noise between using extremely low sensitivity varactors and using a very high sensitivity varactor.*

Flicker noise increases the phase noise of a VCO by FM'ing its carrier through the effect of *AM-to-PM conversion*, and is increased in frequency via nonlinear mixing to the VCO's RF frequency, causing close-in phase noise. Flicker noise is a very low-frequency phenomenon, and is rarely an issue significantly above 5 kHz in a BJT, but GaAs FET's flicker noise can easily reach into the megahertz regions.

Our desire to lower the BJT's bias current to improve its overall noise performance (flicker and noise figure) is in direct opposition to our desire to maximize the oscillator's output power to improve its signal over its noise amplitude (i.e., its SNR) by increasing the stage's bias currents. Nevertheless, if possible within the constraints of the system's power supply voltage and the stage's general bias stability over temperature, we can still safely increase the transistor's V_{CE} without adversely affecting either the BJT's noise figure or the $1/f$.

4.4 LC and VCO Oscillators

4.4.1 Introduction

Today, LC oscillators are normally *variable frequency oscillators* (VFOs) of the *voltage-controlled oscillator* (VCO) type, since they can readily be tuned by adjusting the capacitance of a varactor diode in order to set the frequency of oscillation. Still, very low-cost products that utilize fixed-frequency LC oscillators can be found, sometimes

up to 2.4 GHz. But they will have very poor frequency stability over temperature, and poor phase-noise specifications—all of which is due to their very low loaded *Q*.

Indeed, for any LC oscillator to be even remotely frequency stable or to possess respectable phase-noise characteristics, it will require a high capacitance-to-inductance ratio within its LC tank (for a high *Q*), as well as a steady and clean power supply, stable temperature conditions, and strong isolation from its load. Nevertheless, LC oscillators will still drift in frequency by up to 1% or more due to aging of its components, or when conductive surfaces are in close proximity (if the oscillator is unshielded). Unless some form of frequency regulation circuit is employed, this would be considered unacceptable for any quality wireless device.

The main LC oscillator, the VCO mentioned above, is heavily utilized in frequency synthesis for phase-locked loops (PLLs), and in any application where a DC control voltage is required to alter the output frequency of an oscillator.

4.4.2 LC and VCO Oscillator Topologies

There are numerous kinds of LC oscillators. However, both the *Hartley* and the *Colpitts* oscillators are quite common, and an understanding of their function will permit a good grasp of most other LC topologies.

Hartley Oscillator

The Hartley oscillator, as shown in Fig. 4.14, exploits a tapped coil in its tank circuit, made of L_1 and C_1, to change the phase of the feedback to the transistor's base into a regenerative signal, and to set the frequency of oscillation. C_2 and C_3 block the DC, but couple the AC feedback, while L_2 and C_6 decouple the oscillator output from being injected into the power supply. L_2 also functions as the transistor's collector load, and R_1

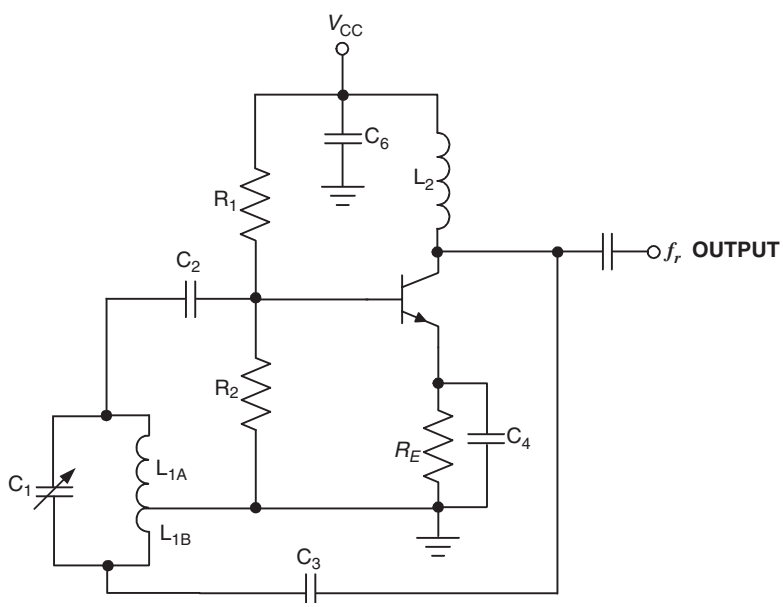


Figure 4.14 A type of LC Hartley oscillator.

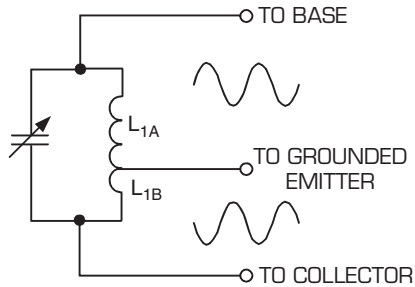


Figure 4.15 A tapped coil showing the phase relationships.

and R_2 supply the forward bias. R_E and C_4 further increase the temperature stability of this circuit, while not allowing the AC gain to be decreased by simply using R_E.

The LC tank of the Hartley furnishes the required 180° phase shift for regenerative feedback, thus allowing oscillations. This is because L_1's tapped coil forces the signal between the center tap and the top of the coil (due to the current flow with respect to the grounded tap; Fig. 4.15) to be opposite in polarity as compared to the center tap and the bottom of the coil. The location of the tap on the inductor sets the amplitude of the positive feedback.

Colpitts Oscillator

As an alternative to a tapped coil, we can use twin tapped capacitors with a parallel inductor (T1), as in the Colpitts oscillator of Fig. 4.16. This circuit performs the same function as the tapped coil by creating a 180° phase shift across each capacitor, thus furnishing the positive feedback into the transistor's base required for oscillations.

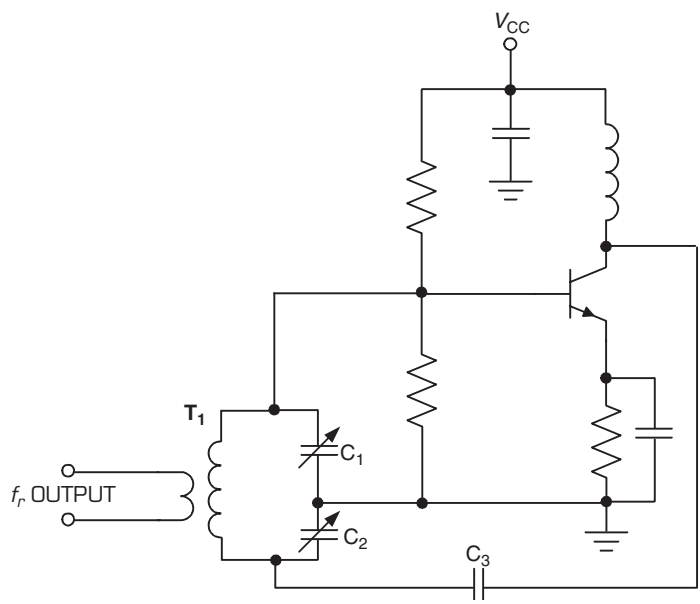


Figure 4.16 A type of LC Colpitts oscillator.

Considering the capacitance ratio changes the feedback voltage, the two capacitors should be ganged to alter the frequency by more than a few percent of its center frequency; or a tunable inductor may be used, with the capacitors at fixed values.

VCOs are simply LC oscillators, of almost any topology, that can vary their output frequency with a DC input control voltage. We can accomplish this frequency changing task by using *varactor diodes*. These diodes, in step with a DC voltage, vary their capacitance up and down. When placed in the resonator of an LC oscillator, the resonant frequency of the resonator—and thus of the oscillator itself—can be made to change either above or below a rest, or center, frequency. The rest frequency is established by the DC bias across the varactors, normally setting their initial capacitance at some intermediate value. So, by adding to or by subtracting from this rest bias value, the frequency of the oscillator can be altered over a wide range.

As an example of a VCO, we see in Fig. 4.17 that Q_1, and its associated components, are designed as a Hartley oscillator, with Q_2 acting as a buffer at the output to prevent the Hartley from being loaded down by the low input impedance of the next stage. The back-to-back varactors shown are commonly employed in a VCO so that, at low bias levels when one varactor is being affected by a strong RF signal, the other is actually being reversed biased, thus decreasing distortion products. However, this will cut the varactor capacitance values in half.

There are two different types of varactor diodes utilized in VCO circuits. The *abrupt* form, which have a high Q and thus low phase noise, and can take a wide input voltage tuning range (from 0 to 50 V) in order to travel through their full capacitance values. This means they possess low tuning sensitivity. In addition, abrupt diodes have a low capacitance range, but also have low distortion characteristics.

The *hyperabrupt* varactor type, on the other hand, has a complete tuning range of approximately 0 to 20 V (this varies with model) for increased sensitivity, so it is the

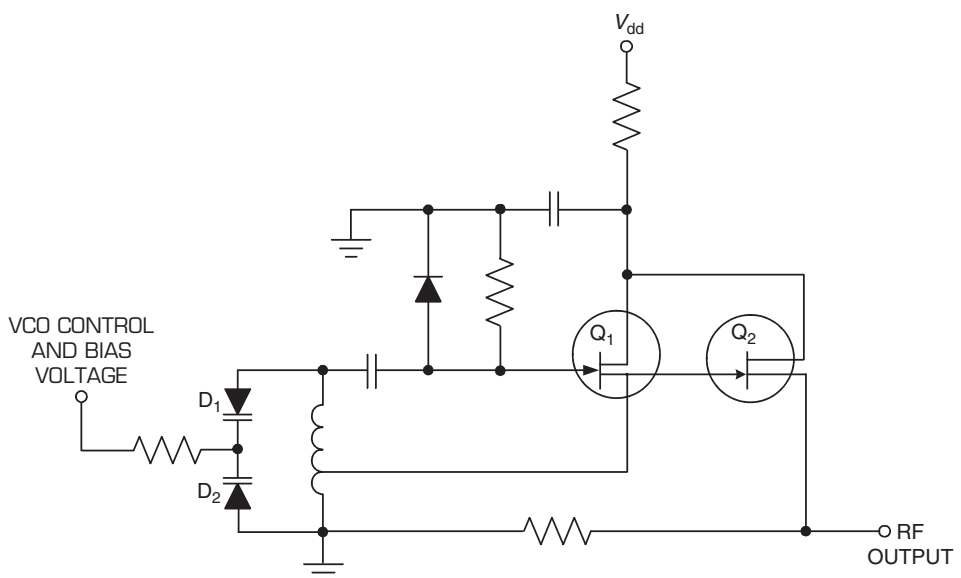


Figure 4.17 A type of Hartley VCO circuit.

varactor of choice for wideband and low voltage applications. But hyperabrupts have a lower *Q*, and thus more phase noise, than the abrupt type.

Both varactors types may have a 0-V capacitance specification, but due to nonlinearity and *Q* problems, at least 0.1 V should always be placed across any varactor, and sometimes much more.

4.4.3 LC and VCO Oscillator Design

Designing LC oscillators and VCOs with the following procedures, while verifying their operation as per Sec. 4.2.2, will permit the engineer to design and build stable and reliable circuits for a variety of requirements.

Back-to-Back Varactors

High RF voltages across the varactor diode of the VCO's resonator can create excessive output harmonics and distortion (similar to a varactor frequency multiplier) by modulating the diode's DC voltage, perhaps even forcing a single tuning varactor into forward conduction, with DC rectification, during low bias levels. This effect can also shift the expected output frequency of the VCO, even making it impossible to tune the VCO to any real extent. Therefore, using back-to-back diodes (see Fig. 4.17) will cut the voltage seen by each such varactor in half, reducing distortion levels and other negative issues when RF signal amplitudes are high, but giving us the disadvantage of halving their effective capacitance over a single diode. As a result, if a single varactor had a capacitance swing of 5 to 50 pF, then the swing for two such diodes in series would be only 2.5 to 25 pF. In other words, the single diode gave us a total capacitance change of 45 pF (50 pF minus 5 pF), while we now have only a 22.5-pF change when using two diodes. During VCO design, this must obviously be taken into account. Any DC blocking capacitor that is in series with the single or double varactor will, of course, also decrease the total available circuit capacitance, but by a much smaller amount if the capacitor is selected properly (i.e., has a much larger value versus the varactor capacitance).

Loaded and Unloaded Q

Since phase noise of an LC oscillator is controlled heavily by the loaded *Q* of the resonator tank, it is critical to employ only high unloaded-*Q* varactors within the resonator circuit. The *Q* of the varactor is given in the part's data sheet, and is normally specified at the reverse bias and frequency at which the *Q* is at its peak. However, the varactor's *Q* will vary dramatically over frequency and bias, decreasing drastically as the frequency increases or the reverse bias voltage decreases. This degradation can take one particular type of varactor's unloaded *Q* from, as a general example, a value of 3000 at 50 MHz, down to a *Q* of 40 at 6 GHz. At very low frequencies, typically well below the RF frequencies of interest, the *Q* will also drop rapidly.

To further increase the loaded *Q* within a VCO's series resonator, and thus improve resonator selectivity and phase noise, we want to maximize the inductance over the capacitance (called the *L/C ratio*) as much as possible within the size and parasitic constraints of the inductor, and within the minimum feasible value of the varactor's capacitance versus the stray real-life capacitance effects that influence its lowest realistic, and repeatable, value. For a practical inductor value used to maximize the L/C ratio when in a series 50-Ω terminated LC circuit, an X_L of 500 to 1000 Ω is considered optimum.

The reason for this selectivity improvement with high L/C ratios is because both the *L*, which is at an increased inductance value, and the *C*, which is at a decreased capacitance value, are at a higher level of X_L and X_C than if the ratio were reversed (i.e., a high C/L ratio).

However, the system's normal 50-Ω resistive terminations, which are in series with the series LC components of the resonator, are pulling down the unloaded *Q* of the *L* and *C*, causing a loaded *Q* of decreased value as compared to the individual component's unloaded *Q*.

Consequently, we see that the impedances that the oscillator's own resonator sees at both its input and output ports will be the dominant lossy element in the series resonator cascade, since these terminations will actually load down the unloaded *Q* of the inductor and capacitor, creating a lower-in-value loaded *Q* for the complete oscillator. Again, this is why we want a high L/C ratio, so as to make both the *L* and *C* impedances as high as possible *above* these two input/output terminations at the series resonant frequency, since a series LC circuit with resistive input/output terminations is really just a series RLC circuit.

As an example of the benefits of a high L/C ratio: If we have a series LC circuit with ideal components (infinite *Q*) placed between two 50-Ω ports, with an L/C ratio of 35600, then our loaded *Q* will be only around 1.9, and our frequency response will be very broad. But if we increase the L/C ratio to 39.5 million (as an example only; not feasible in real life due to component and PCB parasitics), then our loaded *Q* will increase to 63, and our frequency response would be much narrower.

We can drastically increase the resonator's loaded *Q*, as well as add the appropriate 180° phase shift needed for the oscillator's transistor, by adding input and output shunt coupling capacitors (Fig. 4.18), which will help isolate the series LC resonator circuit

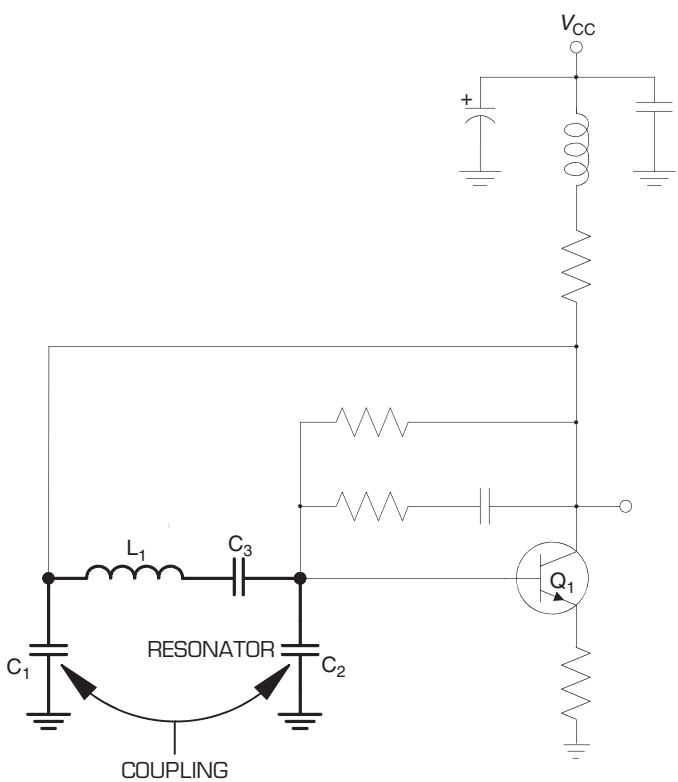


Figure 4.18 Input and output shunt coupling capacitors used with a series LC circuit.

from much of the de-Q'ing (loading) effects caused by the 50-Ω input and output resistive impedance terminations.

When using a *parallel* LC tank circuit as the oscillator's resonator, we can also significantly increase its loaded *Q* by employing a series capacitor at both its input and output ports (resembling a *top C-coupled bandpass* filter). The smaller the value of these series capacitors the better the resonator isolation will be from other *Q*-lowering circuit elements, but this will also increases resonator insertion losses.

Packaged VCOs

Most off-the-shelf VCO modules are designed to function properly only when they see a decent wideband 50-Ω match, with required return losses of better than 10 or 12 dB. Since many such oscillators will be looking directly into a highly nonlinear element (such as a mixer stage, which is one of the worst possible loads for a VCO), we must add RF isolation to the oscillator module's output. We can effectively do this by inserting a 50-Ω attenuator pad of 10-dB value at the module's output port. If the oscillator's output power is decreased too far, we can follow the pad by using a (preferably) high isolation 50-Ω RF buffer amplifier. A common MMIC or discrete transistor amplifier stage will normally suffice as a buffer, especially if the reverse isolation is better than 18 dB.

The use of an attenuator pad for prepackaged VCO's, as well as for most of the VCOs we may design, will alleviate an oscillator's poor phase-noise performance, frequency pulling due to a varying load impedance, and unnecessary RF output power variations. In fact, the sensitivity of the oscillator's own active device is responsible for most of these problems, in that the reflected power from a poorly isolated and matched load, along with a possibly varying load impedance, causes the oscillator's transistor to experience bias point variations, which alters its V_{CB} voltage. This then modifies the BJT's internal collector-to-base capacitance value, which modulates the resonator's center frequency by the shifting its LC tank's capacitive reactance, directly influencing the oscillator's own frequency and phase noise.

BJT LC Oscillator Design for 25 to 500 MHz (Fig. 4.19)

To Design

1. Choose a proper high frequency transistor with an f_T that is much higher than the oscillation frequency (5 to 10 times is a good choice).
2. Bias the active device as Class A by the following procedure:
   a. Choose the supply voltage. Select the *Q*-point for the transistor that is consistent with the available *S*-parameter file for I_C and V_C. Example: $I_C = 5$ mA; $V_C = 3$V; $V_{CC} = 5$ V. Find transistor's typical beta, such as $\beta = 80$.
   b. Calculate $R_B = \beta \times \dfrac{V_C - 0.7}{I_C}$
   c. Calculate $I_B = I_C / \beta$
   d. Calculate $R_C = \dfrac{V_{CC} - V_C}{I_B + I_C}$

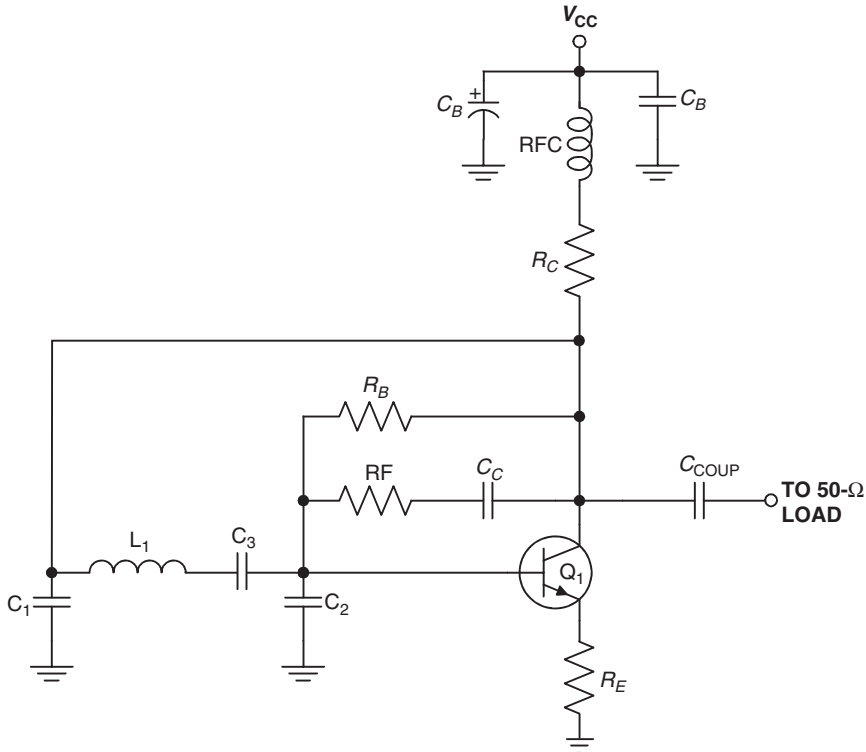


Figure 4.19 An LC oscillator design good to 500 MHz.

3. Calculate the values for the LC resonator and other components by:

$$L_1 = \frac{190}{2\pi f} \qquad C_1 = \frac{1}{24\pi f} \qquad C_2 = \frac{1}{24\pi f}$$

$$C_3 = \frac{1}{300\pi f} \qquad C_C < 1\ \Omega(X_C)\ R_f = \frac{2500}{\left(\frac{0.025}{I_C}\right)}$$

   $R_E = < 15\ \Omega$ (as required to force the oscillator's open-loop input/output ports closer to the same real impedance value).

4. Check Z_{IN} and Z_{OUT} for equality when performing the preliminary open-loop *S*-parameter analysis. Both R_f and R_E can then be tuned, if needed, until both the oscillator's input and output impedances are closer to the same real value, preferably 50 Ω.

5. $C_{COUP} \approx 50$ to 200 Ω (X_C) for a 50-Ω load. Find the necessary value of C_{COUP} by simulating the oscillator into a 50-Ω load, and then use the lowest C_{COUP} reactance value that will still allow the oscillator to maintain a decent gain margin (> 3). (If a high input impedance buffer amplifier follows C_{COUP}, then $C_{COUP} = C_C$.)

6. Simulate and optimize as explained in Sec. 4.2.

NOTE: *Increasing* C_1 *and* C_2*, as well as* L_1*, while decreasing* C_3*, will increase the loaded Q of the oscillator. However, if the* L_1 *to* C_3 *ratio is too high, the frequency tuning of the oscillator via* C_3 *can become excessively sensitive.*

A Quick Example Design a Discrete LC BJT Oscillator (Fig. 4.20)

Goal: Create a discrete LC oscillator for RF frequencies. The specifications and parameters for the circuit are:

P_{OUT} = +1 dBm (The DC bias can be raised to higher levels for more P_{OUT}.)
V_{CC} = 5 V
f_r = 400 MHz
V_{CE} = 2 V
I_C = 10 mA
Transistor = NXP BFG-425W

Solution:

1. RFC = 200 nH
2. R_C = 296 Ω
3. R_B = 10.4 kΩ
4. R_f = 500 Ω (tuned to 260 Ω)
5. C_C = 398 pF
6. C_{COUP} = 7.96 pF (50 Ω at 400 MHz)
7. C_1, C_2 = 33.1 pF
8. C_3 = 2.65 pF (optimized to 2.374 pF)
9. L_1 = 75.6 nH
10. R_E = 10 Ω (for gain reduction, improved impedance match, and a closer 180° phase shift)
11. Tune for optimal response, as required.

LC (Rhea Type) MMIC Oscillator Design for up to 900 MHz (Fig. 4.21)

To Design

1. Calculate a *T network* resonator as shown in the Chap. 3, or use the online Java matching calculator by John Wetherell, called *Impedance Matching Network Designer* (available at multiple Web sites). Select a *Q* of greater than 13, an input/output impedance of 50 Ω, and the desired oscillation frequency. A high loaded resonator *Q* is critical for proper operation of the oscillator and for low phase noise, so select a high-*Q* inductor and capacitor, as well as making sure that the T network has a high L/C ratio (which will be guaranteed by choosing a high loaded-*Q* for the above calculations).
2. A V_{CC} should be selected for the MMIC that will permit at least 2 V (preferably 4 V) to be dropped across R_{BIAS} for proper stability:

$$R_{BIAS} = \frac{V_{CC} - V_{MMIC}}{I_{MMIC}}$$

where V_{MMIC} = DC voltage required at the MMIC's power pin, V, I_{MMIC} = DC current required into the MMIC's power pin, A.

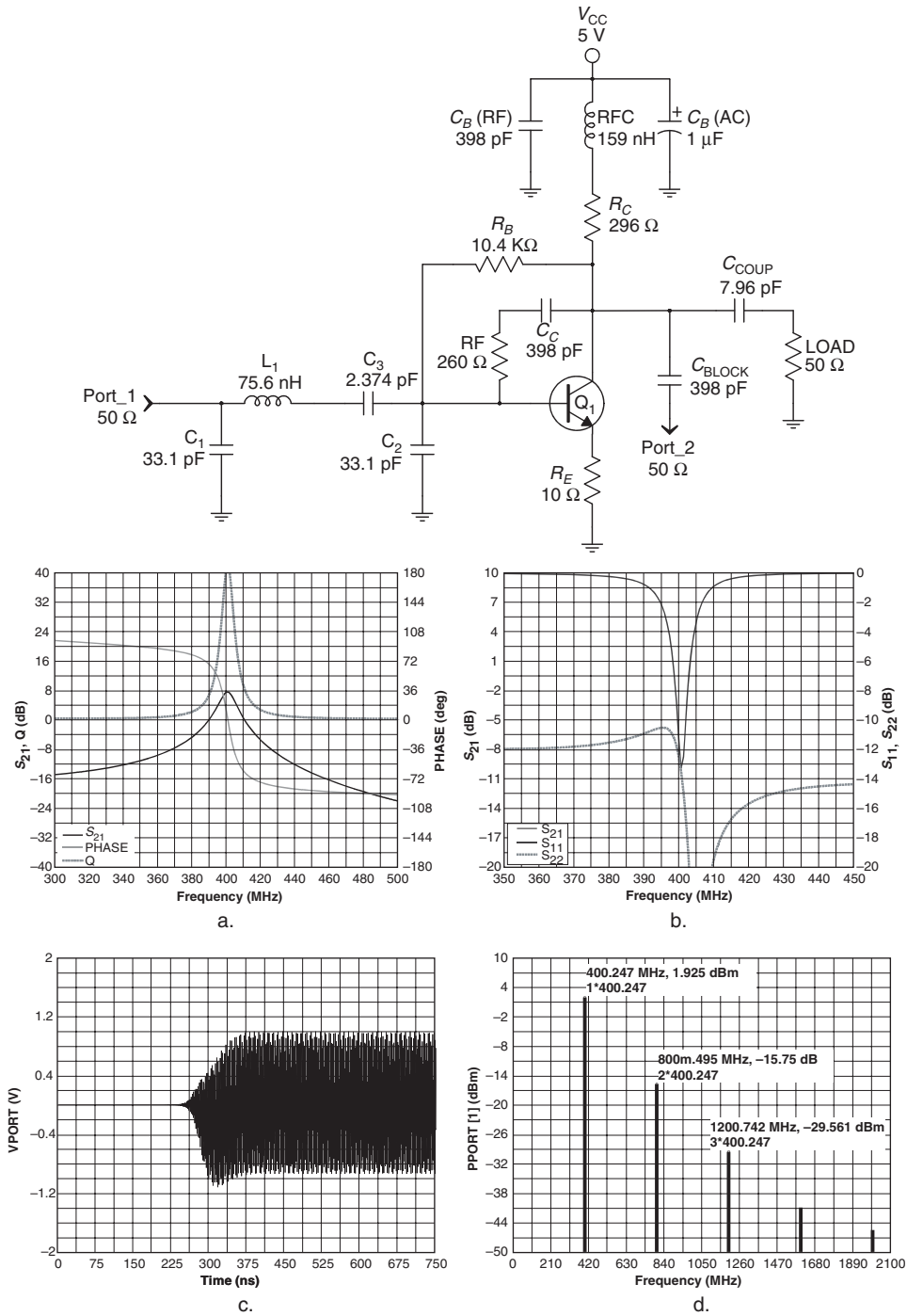


Figure 4.20 An LC oscillator design example showing excellent: (*a*) gain/phase response and loaded Q; (*b*) open-loop port-to-port matching; (*c*) startup; (*d*) output power.

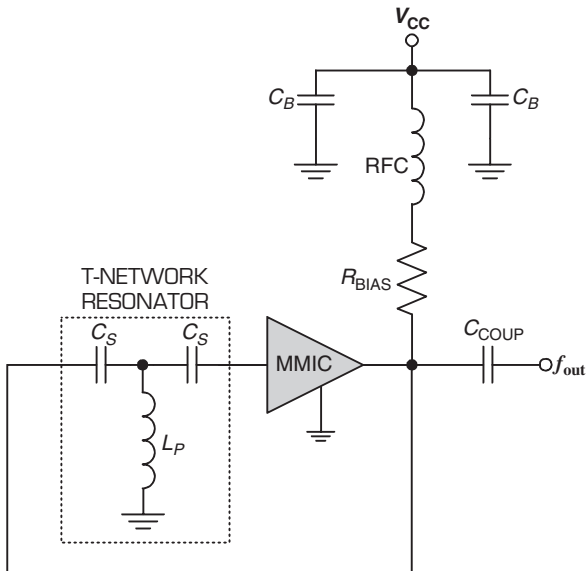


FIGURE 4.21 A MMIC Rhea type LC oscillator circuit.

NOTE: *So that a proper resistor can be selected with at least double the expected dissipation, the power within the* R_{BIAS} *resistor should be checked by* $P = I_C^2 \cdot R_d$. *Also, if* R_{BIAS} *does not reach 600 Ω, employ an RFC for a combined impedance for both* R_{BIAS} *and the RFC of > 600 Ω.*

3. $C_{COUP} \approx$ 50 to 200 Ω (X_C) for a 50-Ω load. Find the necessary value of C_{COUP} by simulating the oscillator into a 50-Ω load, and then use the lowest C_{COUP} reactance value that will still allow the oscillator to maintain a decent gain margin (> 4). (If a high input impedance buffer amplifier follows C_{COUP}, then $C_{COUP} = C_C$).
4. Simulate and optimize as explained in Sec. 4.2.

MMIC LC Oscillator for up to 900 MHz (Fig. 4.22)

To Design

1. A V_{CC} should be selected that will permit at least 2 V (preferably 4 V) to be dropped across R_{BIAS} for stability. (If R_{BIAS} does not reach 600 Ω, use an RFC for a combined impedance of 600 Ω for both R_{BIAS} and the RFC):

$$R_{BIAS} = \frac{V_{CC} - V_{MMIC}}{I_{MMIC}}$$

where V_{MMIC} = DC voltage required at the MMIC's power pin, V, I_{MMIC} = DC current required into the MMIC's power pin, A.

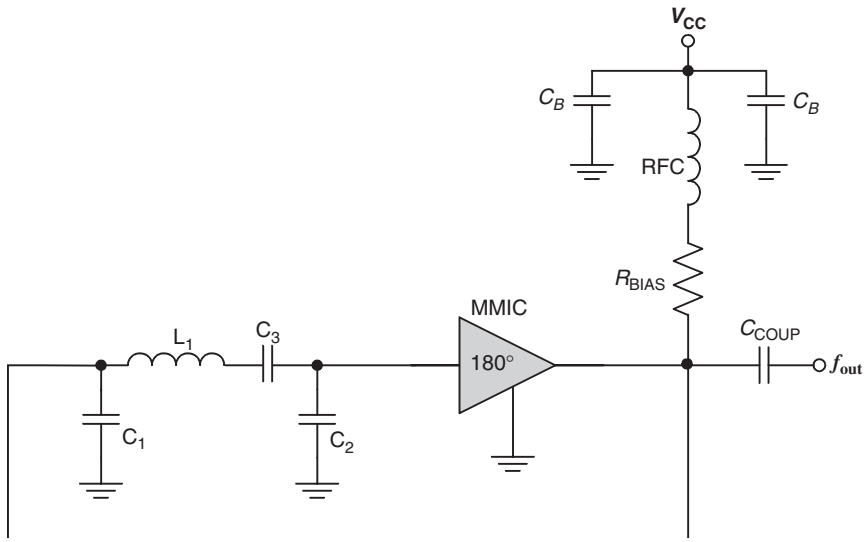


FIGURE 4.22 Another type of MMIC LC oscillator.

2. Calculate the component values for the LC resonator of the oscillator:

$$L_1 = \frac{190}{2\pi f} \qquad C_1 = \frac{1}{24\pi f} \qquad C_2 = \frac{1}{24\pi f}$$

$$C_3 = \frac{1}{300\pi f} \qquad C_B < 1\ \Omega(X_C)$$

3. $C_{COUP} \approx 50$ to 200 Ω (X_C) for a 50-Ω load. Find the necessary value of C_{COUP} by simulating the oscillator into a 50-Ω load, and then use the lowest C_{COUP} reactance value that will still allow the oscillator to maintain a decent gain margin (> 4). (If a high input impedance buffer amplifier follows C_{COUP}, then $C_{COUP} = C_C$).
4. Simulate and optimize as explained in Sec. 4.2.

NOTE: *Increasing* C_1 *and* C_2*, as well as* L_1*, while decreasing* C_3*, will increase the loaded* Q *of the oscillator. However, if the* L_1 *to* C_3 *ratio is too high, the frequency tuning of the oscillator via* C_3 *can become excessively sensitive. And the power dissipation within the* R_{BIAS} *resistor should be checked by* $P = I_C^2 \cdot R_D$ *so that a proper resistor can be chosen with at least double the dissipation expected.*

A Quick Example Design an LC MMIC Oscillator (Fig. 4.23)

Goal: Create a hybrid MMIC oscillator for high RF frequencies. The specifications and parameters for the circuit are:

$P_{OUT} = +7$ dBm
$V_{CC} = 9$ V
$f_r = 800$ MHz

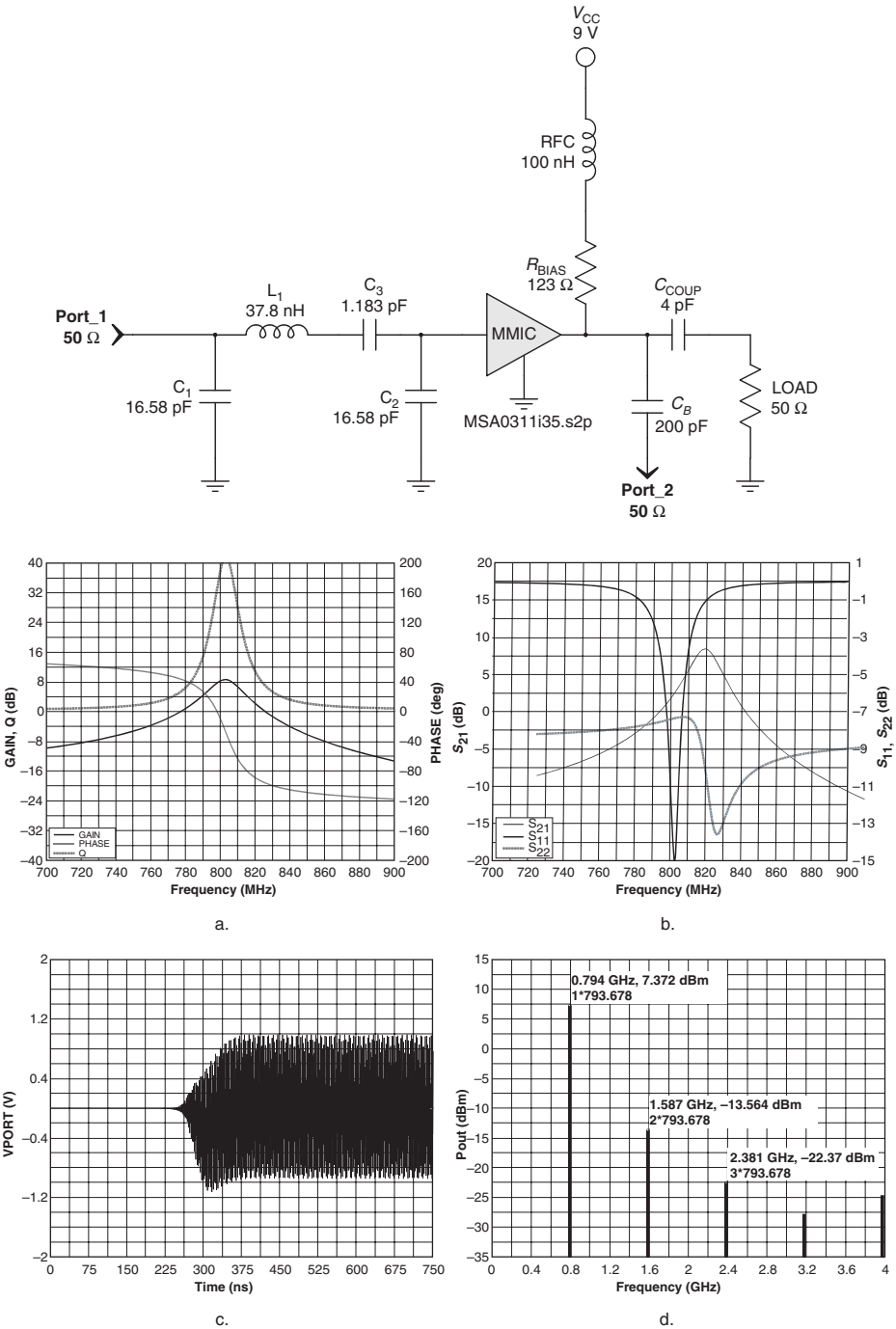


Figure 4.23 A MMIC LC oscillator circuit example showing excellent: (*a*) gain/phase response and loaded *Q*; (*b*) open-loop port-to-port matching; (*c*) startup; (*d*) output power.

V_D = 4.7 V

I_D = 35 mA

MMIC = Avago MSA-0311 (Gain = 11 dB; P1dB = 9 dBm)

Solution:

1. RFC = 100 nH (+*j*500 at f_r)
2. R_{BIAS} = 123 Ω
3. C_B = 200 pF (<1 Ω X_C at f_r)
4. C_{COUP} = 4.0 pF (50 Ω X_C at f_r)
5. L_1 = 37.8 nH; C_1, C_2 = 16.58 pF; C_3 = 1.32 pF (which was optimized to 1.183 pF during simulation)
6. Tune for optimal response, as required.

NOTE: *Decrease* L_1: C_4 *ratio for less tuning sensitivity, as required.*

JFET LC and VCO Colpitts Oscillator for up to 50 MHz (Fig. 4.24)

This oscillator has a loaded *Q* from 20 to 25, and can tune to over 100% of f_{out}. L_1 should be as high an unloaded *Q* as possible to maximize gain and phase margin. VCO circuit can easily be made into a low-cost LC oscillator by replacing the varactor D_1 with a capacitor of similar value, and removing the varactor's bias network of *R*, both C_C's, and L_2.

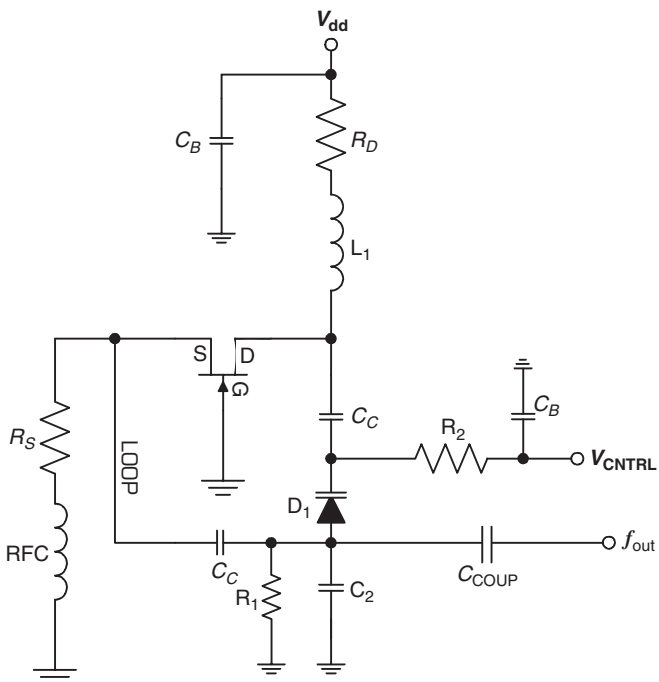


FIGURE 4.24 A JFET Colpitts VCO.

The oscillator's open-loop Z_{OUT} and Z_{IN} will be around 150 Ω, so the terminating impedances of the *S*-parameter linear simulation program should be changed to this value to obtain the proper open-loop gain and phase. However, when the mismatch is slight, the actual loop gain of the final closed-loop oscillator will be negligible. If the active device does not have *S*-parameter files available due to its low frequency of operation, then employ Spice models in a Spice or other nonlinear simulator.

To Design

1. Self-bias the oscillator:
   a. Choose an appropriate V_{GS} for Class A operation of the active device, and note the I_d for the selected V_{GS}.
   b. Compute $R_S = \dfrac{V_{GS}}{I_d}$
   c. Choose a V_{ds} of $\dfrac{V_{dd} - (2 \cdot V_{GS})}{2}$ for a V_d of $\dfrac{V_{dd}}{2}$
   d. Calculate $R_D = \dfrac{V_{dd} - V_{ds} - V_{GS}}{I_d}$
   e. Since V_{GS} for a specific I_d is not always available, use the following equation to find V_{GS} when I_d, I_{dss}, and V_P are known (look in the JFET's data sheet for these values):

$$V_{gs} = V_P\left(1 - \sqrt{\frac{I_d}{I_{dss}}}\right)$$

   f. RFC =1000 Ω
   g. C_C = < 1 Ω
   h. R_1 = 5 kΩ, or use an RFC.
   i. R_2 = 100 kΩ, or use an RFC.
2. Confirm that the JFET device will safely dissipate the power of $P = I_d \times V_d$.
3. To design the resonant LC network:

$$L_1 = 258/2\pi f\, C_{D1} = 1/480\pi f\, C_2 = 1/24\pi f$$

4. Couple the output of the oscillator to its 50-Ω load through a 200 to 600-Ω reactance (C_{COUP}), which can be either inductive or capacitive.
5. Simulate and optimize as explained in Sec. 4.2.

BJT VCO for up to 500 MHz (Fig. 4.25)

1. Choose a proper high frequency transistor with an f_T that is much higher than the oscillation frequency (Q_1 is normally selected to have an f_T of 5 to 10 times that of f_{out}).
2. Bias the active device Class A:

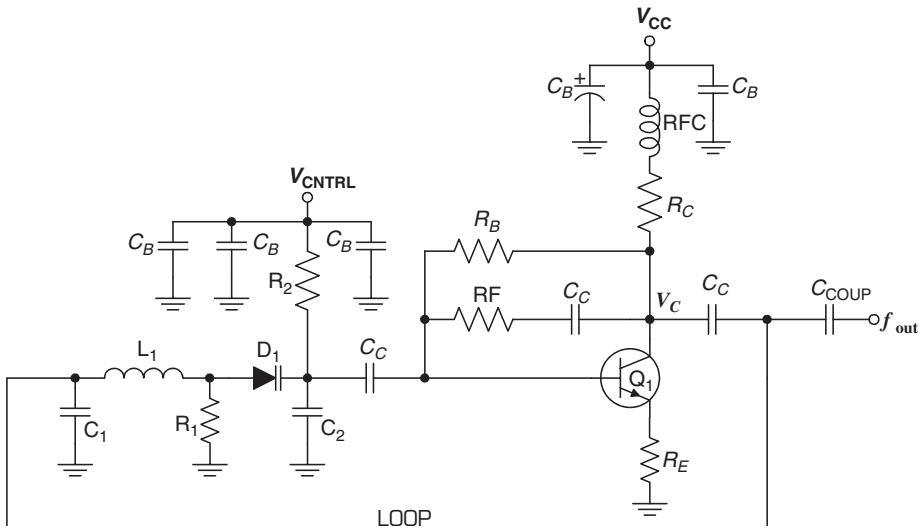


FIGURE 4.25 A complete bipolar VCO circuit.

a. Choose the supply voltage. Select the *Q*-point for the transistor that is consistent with the available *S*-parameter file for I_C and V_C. Example: $I_C = 10$ mA; $V_C = 6$ V; $V_{CC} = 12$ V. Find transistor's typical beta, such as $\beta = 50$.

b. Calculate $I_B = I_C / \beta$

c. Calculate $R_B = \beta \cdot \dfrac{V_C - 0.7}{I_C}$

d. Calculate $R_C = \dfrac{V_{CC} - V_C}{I_B + I_C}$

e. R_1, $R_2 = 50$ kΩ, or use an RFC.

f. RFC = 600 Ω X_L

3. Calculate the component values for the LC resonator of the oscillator:

$$L_1 = \frac{190}{2\pi f}\ C_1 = \frac{1}{24\pi f}\ C_2 = \frac{1}{24\pi f}\ C_C = C_B = <1\Omega(X_C)\ C_{D1} = \frac{1}{300\pi f}$$

$R_e = < 15\ \Omega$ (as required to force the oscillator's open-loop input/output ports closer to the same impedance).

4. Calculate $R_f = \dfrac{2500}{\left(\dfrac{0.025}{I_C}\right)}$

5. Check Z_{IN} and Z_{OUT} for equality when performing the preliminary open-loop *S*-parameter analysis. Both R_f and R_E can then be tuned, if needed, until both the oscillator's input and output impedances are closer to the same value, preferably 50 Ω.

6. $C_{COUP} \approx 50$ to 200 Ω (X_C) for a 50-Ω load. Find the necessary value of C_{COUP} by simulating the oscillator into a 50-Ω load, and then use the lowest C_{COUP} reactance value that will still allow the oscillator to maintain a decent gain margin (>4). (If a high input impedance buffer amplifier follows C_{COUP}, then $C_{COUP} = C_C$).
7. Simulate and optimize as explained in Sec. 4.2.

NOTES: *a. The inductance of* L_1 *and the capacitance of* D_1 *are near series resonance, while* C_1 *and* C_2 *act as coupling capacitors to obtain 180° phase shift with* L_1 *and* D_1 *(with a high* Q*).* R_f*, with its DC decoupling capacitor* C_C*, feed back some of the RF into the oscillator's input in order to diminish low-frequency gain for stabilization of the BJT, as well as to lower both the input and output impedance of the oscillator closer to 50* **Ω***. This not only makes it easier to simulate in a 50-*Ω *environment, but also in a real environment with a VNA (vector network analyzer).* R_B *and* R_C *are the BJT's bias components, and the* $C_{B's}$ *are bypass capacitors chosen to decouple all frequencies from 60 Hz all the way to* f_{out} *and beyond, with an* X_C *of less than 1* **Ω***. This necessitates using various types of capacitors, such as electrolytic for the audio frequencies and two different value ceramic (or porcelain) capacitors for low and high RF frequencies.*

b. The varactor bias voltage, selected from the varactor's data sheet for the cap acitance desired, employs R_1 *and* R_2 *to block RF, but pass the DC control voltage (*V_{CNTRL}*). Since a varactor is reversed biased, very little leakage current will flow through* R_1 *and* R_2*, so the voltage dropped across these resistors will be quite small. For lower noise design, replace* R_1 *and* R_2 *with* RFCs *(see Sec. 4.3).*

c. Q_1 *will have an* f_T *that is 5 to 10 times above the* f_{out} *frequency so as to closely maintain the common emitter's 180° phase shift across the oscillator's entire tuning range.*

d. The $C_{C's}$ *are placed to block DC, but to easily pass* f_{out}*.* C_{COUP} *will have a high* X_C *(50 to 200* **Ω***), and can be readily replaced by an equally high reactance inductor for harmonic suppression.*

e. Depending on the frequencies of oscillation chosen, a varactor of sufficiently high value may not be available. This can be overcome by increasing L_1*, which will allow* D_1 *to be decreased in value.*

f. With the proper wide tuning varactor for D_1*, a tuning bandwidth of 100% can be accomplished by employing a 10 to 1 capacitance hyperabrupt varactor, along with the proper tuning voltage range for* V_{CNTRL}*. However, when the VCO is used in this wideband mode, its output power will begin to decrease as the* f_{out} *increases. This is not a problem in less demanding VCO applications, or in a more narrowband (< 50% tuning range) mode.*

g. A MMIC version of this oscillator is shown in Fig. 4.26, and utilizes the same design equations for the resonator and varactor sections as does the above BJT design. The transistor's bias network and the BJT are simply replaced by a high frequency MMIC, along with its own bias network. The MMIC must, however, supply a 180° phase shift from its input to its output (which is the most common MMIC configuration).

h. Adding a 6 to 10-dB pad between a VCO-coupled output and its amplifier, filter, or mixer stage is normally required to prevent reflections from reaching back into the oscillator. The pad will decrease phase noise and load pulling, which are critical parameters in most LO applications.

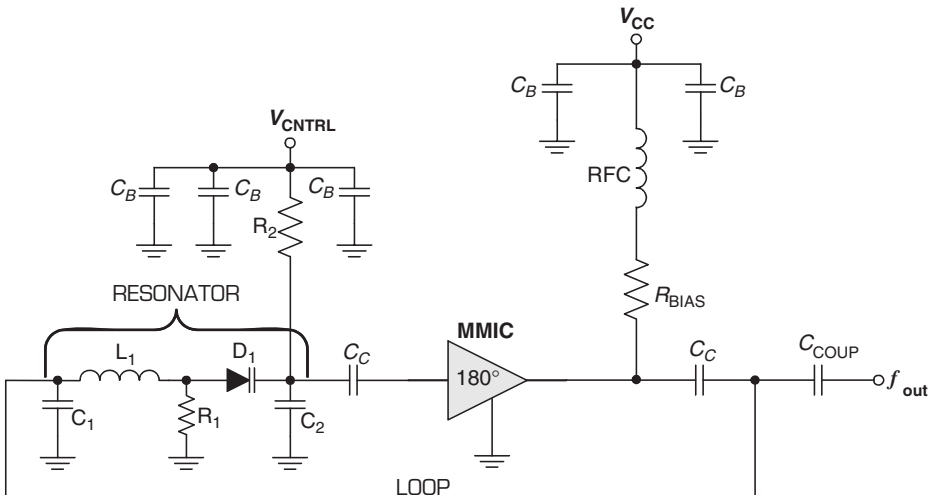


FIGURE 4.26 A complete MMIC VCO circuit.

A Quick Example Design a Discrete LC Voltage-Controlled BJT Oscillator (Fig. 4.27)

Goal: Create a discrete VCO for RF frequencies using a BJT. The specifications and parameters for the circuit are:

P_{OUT} = 0 dBm

V_{CC} = 5 V

V_{CNTRL} = 0.2 to 5 V

f_r = 340 to 460 MHz

V_{CE} = 2 V

I_C = 25 mA

Transistor = NXP BFG-425W (A transistor with a lower f_T may easily be used at these frequencies.)

Diode = Zetex ZV950

Solution:

1. RFC = 200 nH
2. R_C = 118 Ω
3. R_B = 4.16 kΩ
4. R_f = 2.5 kΩ (optimized during simulation to 540 Ω)
5. C_C = 398 pF
6. C_{COUP} = 7.96 pF (50 Ω at 400 MHz)
7. C_1, C_2 = 33.1 pF
8. C_3 = 6 pF (added to decrease the varactor's tuning capacitance for higher frequency operation)
9. L_1 = 75.6 nH
10. R_E = 7.5 Ω (optimized value: used for gain reduction, improved port-to-port impedance match, and a closer 180° phase shift)

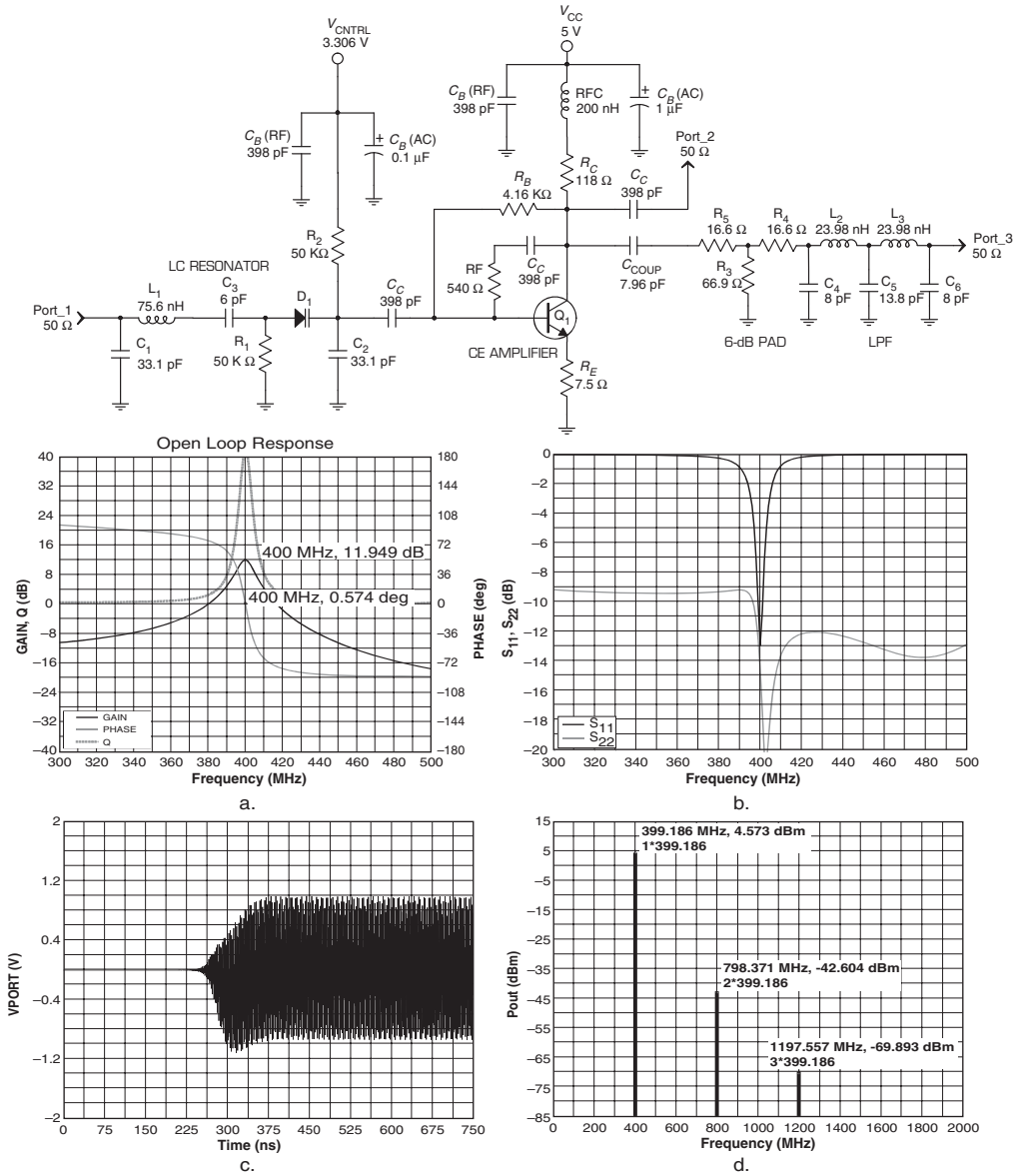


FIGURE 4.27 A BJT VCO oscillator example showing excellent: (*a*) gain/phase response and loaded *Q*; (*b*) open-loop port-to-port matching; (*c*) start-up; (*d*) output power. (Excess gain will decrease when *Q* is added to components, especially the varactor).

11. R_1, R_2 = 50 kΩ (bias resistors used to isolate varactors from other circuit reactances, and to decouple noise and RF)
12. PAD, 6 dB = 50-Ω 6-dB attenuator pad of R_5, R_4 = 16.6 Ω; R_3 = 66.9 Ω. (used for VCO isolation and to permit the LPF to see a good match)

13. LPF = fifth-order lowpass filter of L_2, L_3 = 23.98 nH; C_4,C_6 = 8 pF; C_5 =13.8 pF (used to decrease output harmonics)
14. Tune oscillator for optimal response, as required.

NOTE: *If employing a nonlinear simulator to verify the accuracy of the linear open-loop simulation, and we see that the VCO is now either tuning improperly, or not within its proper frequency range, or jumps to incorrect frequencies as compared to the linear simulation, then we must add unloaded* Q *to the resonator's ideal* L *and* C *components. This lack of any real-life resistance in the simulation of these particular parts causes high voltages to form within the resonator circuit, which then impinges on the varactor diode, producing these flawed tuning symptoms. Therefore, adding unloaded* Q *is not only a more realistic preliminary RF simulation technique, but will also obviously decrease the loaded* Q *of the entire resonator, and thus decrease the excessively high voltages that are present within the circuit.*

Be aware that in going from a linear to a nonlinear simulation, the two different methods will generally vary by a few percent in the resultant output frequency results. This is due to the transistor now running in its more accurate nonlinear mode, rather than with the original linear assumptions required for open-loop simulation.

Integrated LC and VCO Oscillator for up to 1050 MHz (Fig. 4.28)

Integrated circuits are now being manufactured that will function as an LO or as a VCO by simply attaching a few external components. One such oscillator is Maxim's *MAX2620*, which can operate anywhere between 10 and 1050 MHz, and with low phase

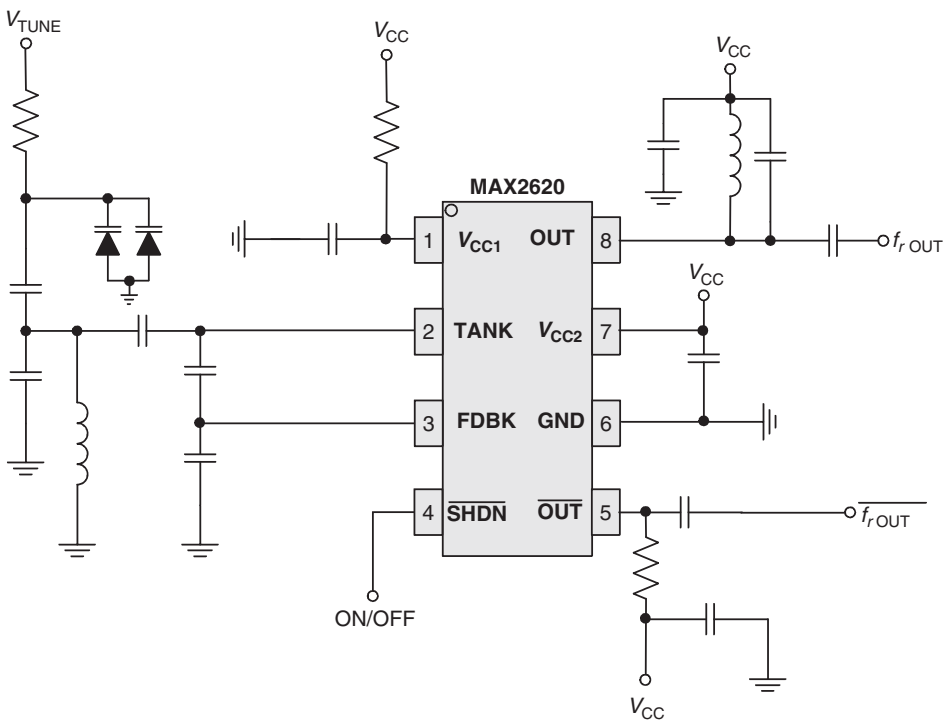


FIGURE 4.28 The Maxim oscillator integrated circuit with support components.

noise (–110 dBc/Hz @ 25-kHz offset). It has a built-in integrated output buffer, and requires only a low-voltage power supply (+2.7 to +5.25 V) for decreased power consumption (27 mW at 3 V_{CC}). The MAX2620 can be employed differentially to offer an IC mixer an LO signal, or as an unbalanced oscillator to feed a double-balanced mixer.

4.4.4 LC and VCO Oscillator Bench Testing

After any of the above oscillators have been built as physical prototypes, the following steps are helpful to confirm proper circuit operation:

1. Supply the oscillator with its nominal V_{CC}. Viewing the oscillator's output, only the fundamental frequency and its harmonics (at 10 dBc or less) should be present. Small amplitude spurii from any external EMI entering the oscillator may be present, and are usually acceptable in most applications. In a fundamental (non-multiplied) oscillator, there should be no subharmonics at the output port nor spurii located at $0.25 \times f_r$ or $0.5 \times f_r$, which if present are due to undesired *parametric oscillations*.
2. Smoothly vary the oscillator's supply voltage V_{CC} from 0 V to the maximum safe operating voltage, and then back down, while confirming that there are only uniform frequency and power output variations with no unexpected jumps in either parameter (except at some low V_{CC} value, where oscillations will then naturally cease).
3. Smoothly vary the VCO's control voltage V_{CNTRL} across its entire range, while confirming that there are only even and continuous output frequency changes across its entire monotonic tuning range, and with no severe RF power dips or peaks.
4. The oscillator should be tested over wide temperature and load variations to check that it remains within proper frequency, power, and harmonic specifications.

4.4.5 LC and VCO Output Coupling

Adding a load to the output of an oscillator can drastically affect its frequency and power, or indeed whether it will even start or not. Therefore, an appropriate RF output coupling method is important, as is the exact location within the circuit that we should be removing this power. The RF power can be coupled out of the oscillator circuit by a directional coupler, a power divider, or a reactance that is placed either before or after the oscillator's resonator.

A compromise must be reached between tapping too little or too much RF power from the feedback loop of the oscillator, since we must not load down the oscillator so much that it stops functioning, becomes unreliable in starting or during operation, or shifts its performance parameters; yet we must also provide enough RF signal to the load to meet our desired specifications.

In order to confirm that the proper gain margin will still exist within the oscillator after we place a load at its output port, it is recommended to insert the desired series output coupling circuit, along with the output load itself, into the final open-loop simulation run. Using an appropriate coupling capacitor or inductor with a high enough X_C or X_L in series with the load is recommended due to the low cost and simplicity (Fig. 4.29) of such an output coupling method. A capacitive or inductive reactance tap

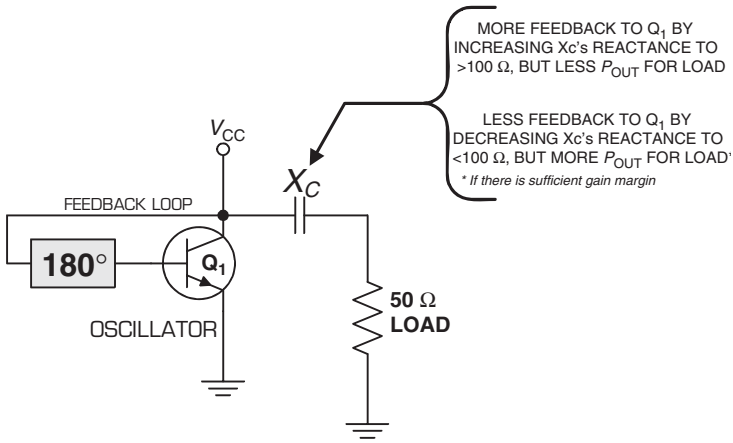


Figure 4.29 Comparison between a high and a low coupling reactance on the output power of an oscillator.

of 200 Ω, as an example, should not degrade the oscillator's vital feedback amplitude excessively, while minimally decreasing the effective RF power output of the oscillator. However, if simulations predict that a lower reactance of 60 Ω or so is possible while still maintaining the proper gain margin, then more RF power can then be safely output into the load. Therefore, within the linear simulator we must optimize the reactance of the oscillator's inductive or capacitive coupling element until the oscillator's gain margin is at a level of between 3 and 6 dB. (Adding passive circuit components with real-life unloaded *Q* will decrease this oscillator gain margin, sometimes severely). This means that a true conjugate match directly from the oscillator output to its load is generally not possible, as this would excessively load down the oscillator stage, preventing it from starting or running properly.

To reiterate and clarify this important point: During RF simulation, always attach the oscillator's expected load (normally 50 Ω), along with the oscillator's output coupling reactance (normally C_{COUP}) to the oscillator's open-loop output port as a final test that the loop gain will not be degraded excessively. However, only attach the RF simulator's 50-Ω ports between the transistor's collector and C_{COUP}, and *not* between C_{COUP} and the oscillator's load.

For insertion into a load, there are multiple methods of tapping into an oscillator for its RF output signal, with different power, phase noise, stability, and matching consequences. In view of the fact that, as stated above, it is impossible for most oscillators to supply a perfect 50 + *j*0 match directly into the next stage without decreasing the oscillator's gain margin to dangerously low levels, even to the point of quenching oscillations completely, a compromise must be found. The following examples are all viable options for coupling RF power out of an oscillator, depending on the specific requirements:

1. Tapping the output of an oscillator with a directional coupler is relatively straightforward when used for narrowband LC and VCO oscillators. This method will permit only low RF oscillator output powers (depending on the decibel output of the coupler chosen), great phase noise, and an excellent 50-Ω match at the coupler's output.

2. Tapping the output of the oscillator with a high reactance capacitor or inductor (60 to 200 Ω for a BJT), followed by a 50-Ω 10-dB pad into a discrete common-emitter or MMIC amplifier, will supply medium to high powers, respectable phase noise, and a very good 50-Ω match at the amplifier's output.
3. Tapping the oscillator's output with a low reactance coupling capacitor ($X_C < 1\ \Omega$) into a high impedance common-collector buffer amplifier will supply medium RF power, medium phase-noise performance (phase noise can be created by complex interactions between the oscillator and its less than perfect match to its buffer stage), and a very good 50-Ω match at the buffer's output.
4. Tapping the oscillator's output with a high reactance capacitor or inductor (60 to 200 Ω for a BJT) into a 50-Ω 6-dB pad, will supply low output power, good phase noise, and a medium 50-Ω match at the pad's output. This is a low cost and simple solution that is popular for many consumer applications.

4.4.6 LC and VCO Oscillator Harmonics

Depending on loop gain and where the RF output itself is taken, harmonics from an oscillator's output can reach high levels. The higher the loop gain, the deeper the oscillator travels into saturation and cutoff, and the higher in amplitude will be the harmonic outputs. However, since a decent loop gain is vital for dependable operation over temperature and part's variations, a compromise must be made. Removing the power at different oscillator circuit locations will make a large contribution to the stage's final RF output power and harmonic generation because, as shown in the standard LC feedback oscillator of Fig. 4.30, tapping the circuit at Point A will supply the highest RF output power, but will also result in significant harmonics. This is due to the resonator not having the chance to filter the RF of the amplifier before it is tapped off into the load. Taking the power at Point B, as shown in Fig. 4.31, will result in a filtered output, with the harmonics considerably lower than at Point A. But to prevent frequency changes and an excessively decreased gain margin, the oscillator circuit tap (whether located at Point A or Point B) should load the oscillator's resonator as lightly as possible. This can sometimes be a more challenging goal to accomplish if selecting Point B.

The attenuation of harmonics will also be assisted by coupling the energy out of the oscillator's output with an inductor rather than a capacitor, since the inductor's reactance will increase with frequency, thus decreasing the harmonic output. If an inductor is chosen for the output tap, then a series capacitor of $X_C < 1\Omega$ will be required in most applications to prevent DC from entering or exiting the stage. However, a single RF output coupling capacitor is cheaper and smaller than the required inductor/capacitor combination, and so this simple topology is much more common within low-cost applications.

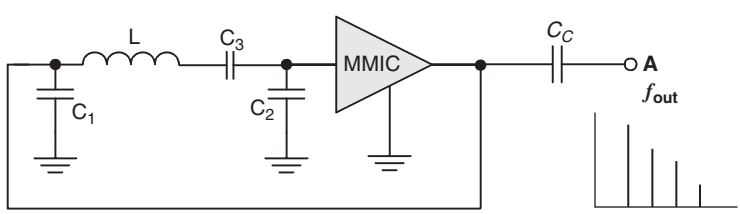


Figure 4.30 Point A of the oscillator results in high harmonic (and fundamental) outputs, but is easy to implement.

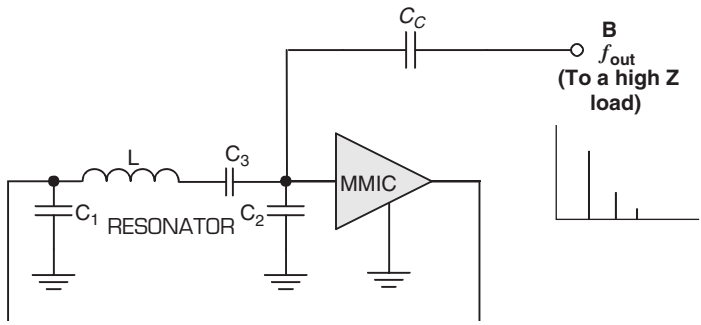


Figure 4.31 Point B of the oscillator results in lower harmonic (and fundamental) outputs, but can perturb the resonator circuit.

4.4.7 LC and VCO Oscillator Output Power

The RF amplitude of an oscillator is mainly dependent on its total loop gain, the compression level of the active device, the V_{CC} voltage, and the RF output coupling. We can quickly discover the approximate output power of an oscillator by employing the oscilloscope tool in Spice, or any other nonlinear simulator.

Place the simulator's test ports across the oscillator load, and then use the following formulas:

$$P = \frac{V_{RMS}^2}{R} \text{ (in watts)} \qquad \text{or} \qquad P = 10\log\frac{P}{1\,\text{mW}} \text{ (in dBm)}$$

Or, the RF output power of a Class A oscillator can be approximated by taking the transistor's maximum peak voltage swing of the V_{CE}, and subtracting the minimum peak V_{CE} swing, then multiplying this by the maximum I_C peak minus the minimum I_C peak, and then dividing by 8:

$$P_{RF}\left(\text{in mW}\right) = \frac{\left(V_{CE(MAX)} - V_{CE(MIN)}\right) \times \left(I_{C(MAX)} - I_{C(MIN)}\right)}{8}$$

This formula demonstrates that the wider the voltage swing between the collector and emitter, and/or the wider the collector current swing, the higher the RF output power in any Class A biased active device.

Since an oscillator rarely encompasses just an oscillator circuit alone, the coupling and any attenuator pad and/or buffer amplifier's located at the output must, of course, be added or subtracted from the above calculated RF output amplitude value.

Indeed, when the DC bias is set for maximum RF output power, an oscillator will output close to the transistor's full P1dB specification. Although the power that *can* be tapped from the oscillator will be approximately the P1dB value of the transistor, the actual RF output power that is truly available to a load will be decreased by the attenuation of the oscillator's tapped output network, such as the high impedance coupling capacitor or inductor.

If loop gain is excessively low, the available output power of the oscillator (before any RF power is tapped for the load) will be something less than the P1dB of the transistor. However, simply setting the oscillator for a high-loop gain will not always translate into higher RF output powers in a decibel-for-decibel fashion, since the

oscillator itself may be reaching hard saturation and cutoff at every cycle. But the higher the gain margin of any oscillator, the higher will be the allowed RF output power that can safely be coupled out to a load, since a high gain margin will possess more room to decrease before it becomes dangerously low within the feedback loop. This means that when the gain margin is high, a decreased coupling reactance ($X \approx 60\ \Omega$) can then be utilized for a higher RF output power level to the load. Biasing the oscillator at higher than normal current levels, and using a higher frequency transistor, will also naturally increase the oscillator's RF output power. Nevertheless, because of the increased heat buildup in the resonator and transistor circuit, the higher the RF output power of an oscillator the more difficult it may become to prevent frequency drift during its operation and warm-up.

Efficiency of an oscillator stage is also of great interest, especially within battery-operated wireless devices. When combined with the DC power dissipation created by the static bias current, and the voltage across the collector (or $P_{DC} = V_C \cdot I_C$), we can obtain the efficiency value of a Class A gain stage of the oscillator by:

$$(P_{AC}/P_{DC})\ 100 = \text{power efficiency, \%}$$

4.4.8 LC and VCO Oscillator Terms

Frequency Stability—The measurement of the frequency variation of a VCO while the ambient temperature is varied, in MHz/°C.

Harmonic Rejection (or *Harmonic Suppression*)—The measurement of all harmonically related frequency products in relation to the power level below the RF carrier, in dBc.

Input Capacitance—The measurement of the VCO's DC tuning input capacitance, in pF. (Important in PLL filter design)

Modulation Sensitivity—The parameter that specifies a VCO's frequency change per DC tuning voltage input, in MHz/V. (Not a completely linear frequency measurement vs. the DC tuning range)

Phase Noise—Oscillators are not perfect single CW frequency sources, but possess phase noise, which is similar to a modulated spectrum created by a virtual noise source that is phase modulating the desired CW signal. Phase noise is measured as the ratio between the fundamental frequency and a noise sideband, with the noise sideband located within a 1-Hz bandwidth at a particular frequency that is offset from the fundamental, in dBc/Hz@kHz.

Post Tuning Drift—An undesirable slow alteration in a VCO's frequency that occurs after the VCO changes its output frequency after commanded to do so by the DC tuning voltage, in kHz or MHz versus time.

Pulling—A parameter that specifies the shift of frequency (and/or power) when the VCO is in a specific output impedance mismatched condition, in MHz.

Pushing—The variation in frequency and/or amplitude that occurs to a VCO due to amplitude changes in the oscillator's V_{CC}, in MHz/V and dB/V.

Spurious Signals—Undesired and nonharmonically related CW power output spikes present in a VCO's output, in dBc.

4.5 Crystal Oscillators

4.5.1 Introduction

Since modern wireless communications equipment could not function properly with the extreme drift present in even the finest LC sine wave oscillator, it was necessary to develop crystal-controlled oscillators. In today's world of limited bandwidth and tight channel spacing, an RF transmitter that was equipped with an LC oscillator not only would drift and interfere with adjacent channels, but its signal would be unreadable as it moved into and out of the passband of the remote receiver, creating changes in volume, pitch, and distortion levels for analog radio, and a degradation of BER levels in digital radio. And the lack of any serious frequency stability on the receiver end of the wireless chain, caused by the use of LC oscillators in the LO, would only contribute to this problem. Even the PLL would not be able to function properly without high-quality crystal oscillators as the synthesizer's frequency reference.

A crystal oscillator requires four features to precisely oscillate at a stable frequency and amplitude:

1. The loop gain must be positive, and at +1 (but greater than +1 to start).
2. The oscillator circuit's impedance must be equal to the crystal's internal resistance.
3. The oscillator circuit must not drag down the unloaded Q of the crystal excessively.
4. The total oscillator circuit feedback phase must be 0°.

Because a crystal will vibrate at its own natural resonant frequency if an alternating signal at that same frequency is placed across the crystal's ports, thus functioning as an ultra high-Q series resonant circuit, a crystal is the perfect choice for operation in a stable oscillator. Due to its low cost, strength, and thermal stability, the most common crystal material used in such oscillators is quartz.

The maximum frequency a crystal can reach on its fundamental is 200 MHz, and this value is only attained by using specialized inverted mesa methods, with the most common crystals capable of operation to a maximum fundamental frequency of only 20 MHz. In spite of this, other crystal types can attain higher frequencies on *harmonic* and/or *overtone* operation.

For example, a *harmonic crystal oscillator*, with its output tank circuit tuned to one of these harmonics, can have an output at the crystal's 2nd, 3rd, 4th, and so on harmonic, yet the crystal itself is really only operating on its lower fundamental frequency. When operated as an *overtone crystal oscillator*, the crystal itself must actually vibrate at the high harmonic (overtone) frequencies, and will function only at one of its odd harmonics (such as the 3rd, 5th, or 7th), with the output tank circuit tuned to this particular desired frequency. Overtone crystal oscillators normally need special overtone crystals when operated in this mode.

Since the crystal functions as a very high-Q series resonant circuit, it naturally possesses a very narrow bandwidth along with high temperature stability, and therefore a crystal oscillator cannot easily alter its frequency (except by a few hundred hertz when used with a capacitive trimmer). To the rest of the oscillator circuit the crystal looks as shown in Fig. 4.32: R_s is the resistance of the crystal during series resonance,

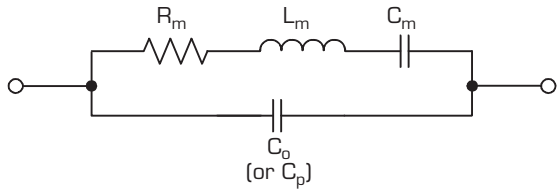


Figure 4.32 The equivalent internal circuit of a packaged oscillator crystal.

while L_m is the motional inductance, C_m the motional capacitance, and C_o is the capacitance between the crystal's holder or, in a modern crystal, it's plated electrodes. In fact, this C_o at VHF and above has so decreased in reactance that it has effectively shorted the output of the crystal to its input (this problem can be mitigated by resonating a small value inductor in parallel with C_o).

Most oscillators operate in series resonance mode, with the values of L_m and C_m governing the resonant frequency of the crystal. At series resonance the crystal is resistive with no reactances, since $X_L = X_C$, and can be described as:

$$f_s = \frac{1}{2\pi\sqrt{L_m \cdot C_m}}$$

As mentioned above, fundamental crystal operation usually peaks at 20 MHz. This is not only due to the dangerously decreasing thickness of the crystal, but also to its decreasing R_S. Indeed, the crystal's R_S can decrease to 10 Ω at 20 MHz on its fundamental, while a crystal in 7th overtone mode can reach 180 MHz with an R_S of 80 Ω . This demonstrates why many oscillators must run in overtone modes, which permits the crystal to be more easily impedance matched at much higher operational frequencies.

L_m, whose very large value is based on the mechanical mass of the quartz crystal, can vary in inductance anywhere between 3600 mH at 1.5 MHz, to 10 mH at 20 MHz. C_m, whose very small value is based on the actual stiffness of the quartz crystal, the size of the electrodes, and the size of the quartz, can vary anywhere between 0.007 pF at a fundamental frequency of 1.5 MHz, to 0.02 pF at its fundamental frequency of 20 MHz. But when a crystal is operated on an overtone, the C_m will decrease. The designer can choose the value of the C_m desired, and the L_m will then be:

$$L_m = \frac{1}{4\pi^2 f_r C_m}$$

C_o is a capacitance value that can be measured across a crystal at rest, and neither vary with frequency of operation, nor with the number of its overtone, but only with the crystal's distance between its electrodes and the electrode's area. This value will normally be between 2 to 8 pF, with the lower the value the better it is for oscillator operation.

Various crystal specifications can greatly affect an oscillator's performance. The frequency accuracy of a crystal, while at room temperature in an oscillator test circuit, can vary from ±5 ppm to ±100 ppm. The lower the number, the more accurate the oscillator's output frequency will be when at 25°C, but the costlier the crystal. And so that we may maintain frequency accuracy under varying ambient and internal temperatures, an oscillator's frequency stability over some chosen temperature range is also another important specification, such as between 0 and 60°C.

Aging affects the crystal's frequency accuracy over time, and can change this accuracy by as little as 0.75 ppm to as much as 6 ppm during a 12-month period. The type of package, crystal quality, crystal stresses, temperature, and frequency all strongly influence this specification. However, the aging of a crystal will mostly occur within its first year, after which the rate will slow down to perhaps ⅕ of this first year's value. For instance, 2 ppm aging over the first year will only result in a 4-ppm change over the next 10-year period.

4.5.2 Crystal Oscillators Types

Considering there are many different crystal oscillator designs available, we will focus only on the most common, which is the *Hartley* of Fig. 4.33, the *Colpitts* of Fig. 4.34, and the *Pierce* of Fig. 4.35.

The crystals for these three oscillators are placed in series with the transistor's feedback path, and since a crystal has a very high Q (in excess of 75,000), and will thus have a very narrow bandwidth (BW = f_r/Q), only the very narrow band of frequencies within the crystal's own natural resonance will pass onto the oscillator's phase-shifting circuits, and will be in phase at the oscillator's input. Feedback that is off frequency by even the smallest amount will be rigorously attenuated, decreasing the level of the transistor's feedback, and forcing the oscillator to return to its design frequency.

The actual phase-shifting network for these types of Hartley and Colpitts oscillators are the LC tank components, while the Pierce crystal oscillator employs a slightly different method of operation.

To form the Pierce oscillator, the inductor of the Colpitts has been substituted by a series resonant crystal. And since at series resonance a crystal will only display a small, pure resistance to the oscillator's feedback, but will exhibit either a capacitive or an inductive

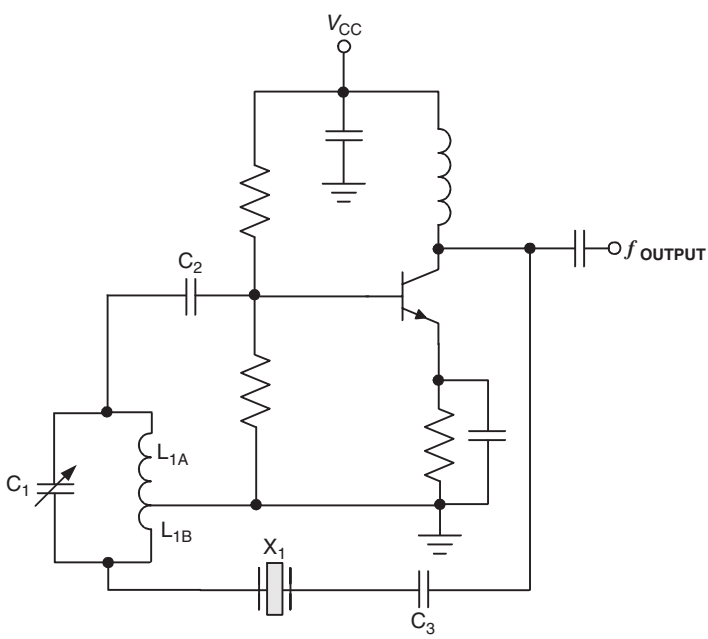


Figure 4.33 A type of Hartley crystal oscillator circuit.

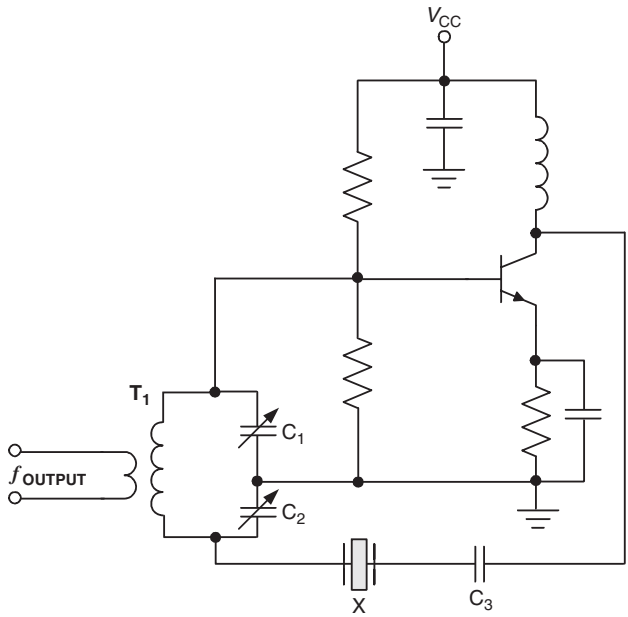


Figure 4.34 A type of Colpitts crystal oscillator circuit.

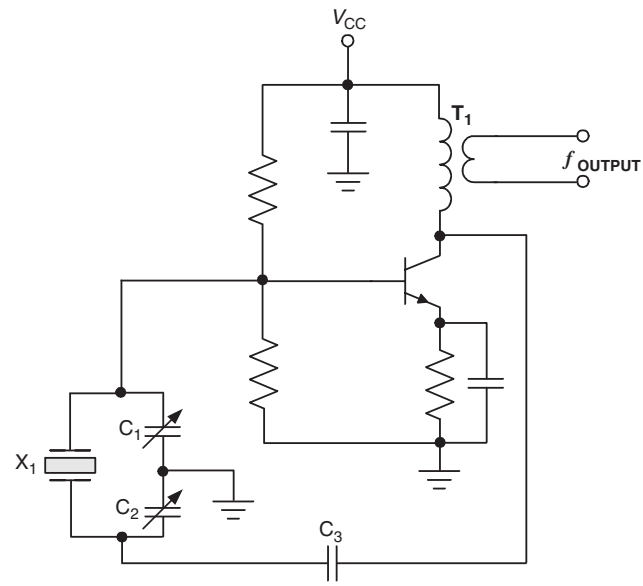


Figure 4.35 A type of Pierce crystal oscillator circuit.

element if not within this small window of series resonance, this will permit the 180° phase shift across the tank that is required for positive, oscillatory feedback. This action forces the Pierce oscillator to stay accurately on frequency in most low power, medium frequency applications. However, the feedback point is usually shifted slightly higher in frequency by a value of about 50 ppm above the value marked on the series resonant crystal itself.

In other words, the Pierce will oscillate higher in frequency, but by only a very small amount, that may be expected by the crystal's own specified frequency (see Sec 4.5.3).

4.5.3 Crystal Oscillator Design

Designing crystal oscillators by utilizing the following procedures, while verifying proper operation as per Sec. 4.2, will permit the engineer to design and construct stable and reliable oscillator circuits for a variety of uses. Due to the low frequencies of some of these oscillators, we may need to employ a Spice simulator, instead of a linear simulator, if the transistor's *S*-parameter models are not available.

Pierce Crystal Oscillator Design for 600 kHz to 30 MHz (Fig. 4.36)

The Pierce works so: R, C_1, XTAL, and C_2 form a 180° phase-shift network, while R is also the feedback control element employed to place less stress on the crystal. C_C in series with the XTAL can be used to tune the oscillator toward the series XTAL frequency, RFC and C_{BYPASS} act in the decoupling role, R_B and R_C are the oscillator's bias resistors, C_{COUP} is used to couple power out of the oscillator into a 50-Ω load without loading the oscillator down below a safe gain margin.

To Design

1. $C_{BYPASS} = C_C = < 1\ \Omega\ (Xc)$
2. $C_1 = C_2 = \dfrac{(20\ \text{nF})}{(10\ \text{uF} \cdot f_r)} \cdot C_{FACTOR}$

where $C_{FACTOR} = 0.5 \le 1$ MHz; $0.6 \le 2$ MHz; $0.7 \le 3$ MHz; $0.8 \le 4$ MHz; $0.9 \le 6$ MHz; $1 > 8$ MHz

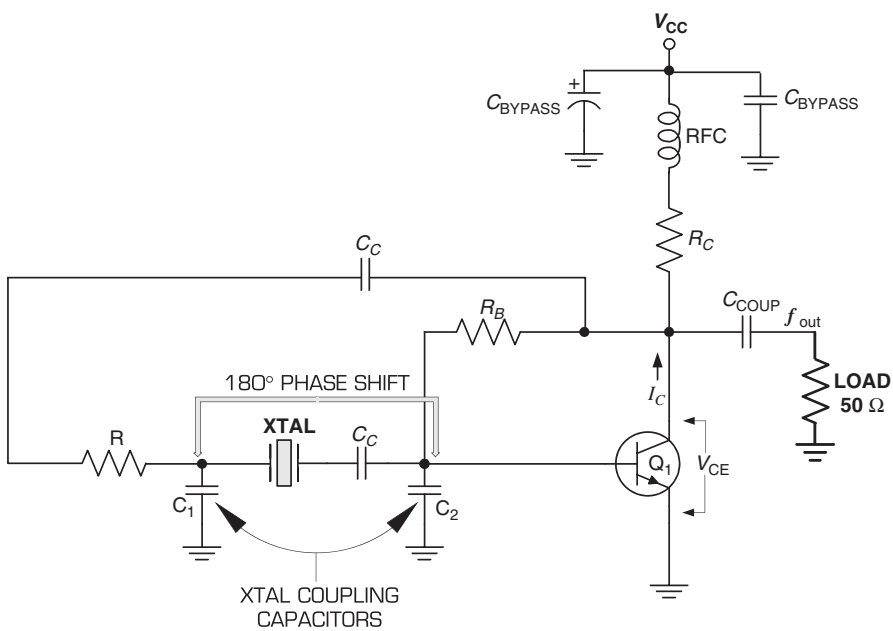


Figure 4.36 A Pierce crystal oscillator circuit showing feedback loop and load.

3. RFC > 500 Ω
4. $R = \dfrac{3}{2\pi f_r C_1}$
5. $R_C = \dfrac{V_{CC} - V_{CE}}{I_C}$

where V_{CE} and I_C = values in the *S*-parameter files for the transistor chosen, or set as desired if in Spice.

6. $R_B = \beta\left(\dfrac{V_{CE} - 0.6}{I_C}\right)$
7. C_{COUP} = start with 50 to 200 Ω X_C for a 50-Ω load, then simulate to confirm proper loop gain. (Increase this reactance value to decrease coupling power output to load; decrease reactance value to increase coupling.)

NOTES: *a. This oscillator uses a crystal that is in series resonance, but since a Pierce will actually oscillate at a frequency that is 20 to 50 ppm above the crystal's marked series resonant frequency, the crystal can be specified to the manufacturer as "parallel resonant." This will simply tell the manufacturer to build the crystal with a series resonant frequency that is approximately 90 to 100 ppm lower than a crystal specified at a series resonant frequency (both crystals are exactly the same except for this small frequency modification). The parallel resonant designation will permit the Pierce to be slightly tuned to operate at* exactly *the frequency as marked on the crystal's can by the mere inclusion of a small variable capacitor shown as* C_C *(or by a varactor circuit for a simple VCXO). This adjustable capacitor, which is in series with the crystal, can also be utilized even if a series specified crystal is employed, but in this case it will vary the frequency of the oscillator from about 5 to 50 ppm* above *the marked series resonant frequency. However, as the impedance increases when the crystal reaches closer to its true parallel resonant frequency (which is always above its series resonance frequency), whether in series or "parallel" mode, the feedback gain will begin to decrease as the frequency of the oscillator is tweaked upward. There will be a point reached where the feedback gain will decrease so much so that oscillations are simply no longer possible. This is why it is prudent to maintain the frequency of the oscillator as close to the actual series resonant frequency of the crystal as possible, without excessive tweaking, whether or not it is a series or "parallel" crystal.*

b. If the oscillator's designer does not supply the crystal manufacturer with what is called the load capacitance *(*C_L*) for a "parallel resonant" crystal, they will assume it to be around 30 pF. The load capacitance is the load that the crystal will see when placed in the oscillator circuit, and slightly affects the accuracy of the parallel crystal's marked frequency. It can easily and more accurately be computed in a Pierce oscillator by:*

$$C_L = \frac{(C_1 \cdot C_2)}{(C_1 + C_2)} + 5 \text{ pF}$$

Specifying the crystal in its series resonant mode will not require the above formula, and even a "parallel" crystal will be fine for most applications without this C_L *specification, unless extreme frequency accuracy is required.*

c. Obtain the crystal's motional capacitance C_M*, motional inductance* L_M*, series resistance* R_S*, and parallel plate capacitance* C_P *from the manufacturer for the crystal's frequency of operation, type of holder, and cut (typically AT). This will allow us to utilize the equivalent circuit of Fig. 4.32 in order to model the crystal as a simple LCR circuit within the linear open-loop simulation program.*

d. The Pierce is meant to work only on the crystal's fundamental-mode series resonance, but can function with overtone crystals if C_1 *is replaced with a parallel resonant tank that is tuned midway between the desired overtone and the overtone just below it. In this case, the crystal manufacturer must be told if the crystal is being run out of its fundamental mode.*

e. Choose a transistor with a much higher f_T *than required for the oscillation frequency (>* $5 \cdot f_T$*), as well as one with a very high gain in our band of interest. The high* f_T *is required to assure as close to a 180° phase shift from the transistor's input to its output as possible, while the high device gain is necessary because of this oscillator's rather high-loop losses.*

f. In order to obtain any substantial RF *power from the typical crystal oscillator circuit, an amplifier will also be needed at its output port. This is because of not only the high-loop losses, but also because of the power dissipation across the crystal element must be maintained to within safe and stress-free limits.*

g. The open-loop output and input impedance of a Pierce crystal oscillator are higher than 50 Ω*, so set the linear software simulator's termination impedances to about 300* Ω *for more accurate results.*

A Quick Example Design a Discrete BJT Crystal Oscillator (Fig. 4.37)

Goal: Create a highly stable BJT crystal oscillator. The specifications and parameters for the circuit are:

P_{OUT} = 0 dBm
f_r = 22 MHz
V_{CC} = 3.3 V
V_{CE} = 2 V
I_C = 10 mA
h_{FE} = 80 (of selected transistor)
XTAL: R_m = 26 Ω; L_m = 4108000 nH; C_m = 0.0125 pF; C_O = 4.185 pF
Transistor = any low-frequency transistor with an f_T greater than five times f_r.

Solution:

1. C_C, C_{BYPASS} = 7.2 nF
2. C_1, C_2 = 90.9 pF
3. L_1 (RFC) = 3.6 uH
4. R = 238 Ω
5. R_C = 130 Ω
6. R_B = 11.2 kΩ
7. C_{COUP} = 145 pF

NOTE: *If employing Spice for nonlinear simulations of the XTAL oscillator, then make sure to assign enough simulation time to permit the very high-Q circuit time to start.*

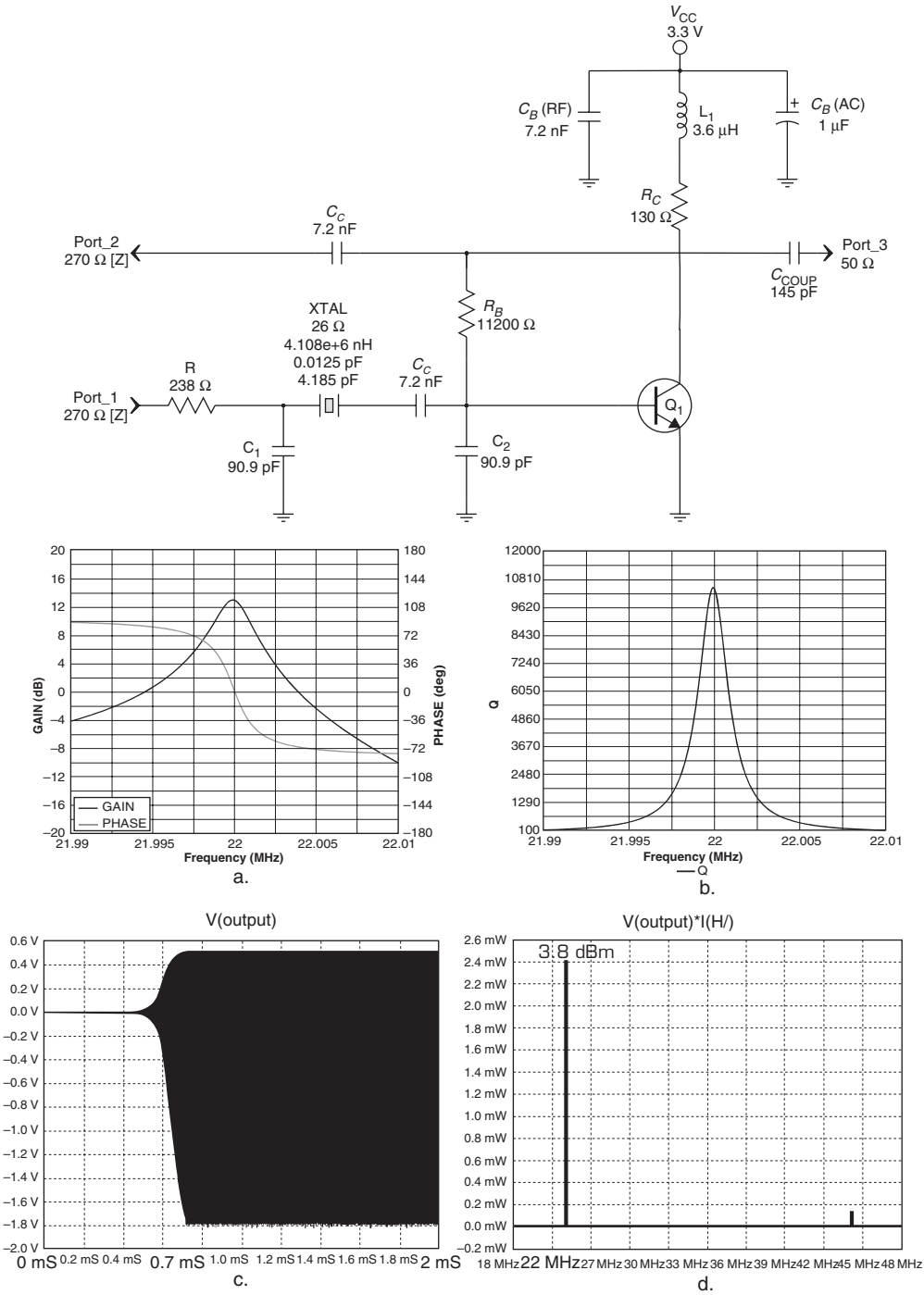


Figure 4.37 A BJT crystal oscillator example showing: (*a*) gain/phase response; (*b*) loaded *Q*; (*c*) (long) start-up; (*d*) RF output power. (Excess gain will decrease when unloaded *Q* is added to the components).

4.5.4 Crystal Oscillator Issues

Depending on frequency, the average crystal should not be subjected to more than 10 mW power dissipation. Although this value can be as low as 1 mW in some crystals before the structure becomes so heated that it may force the oscillator to drift in frequency. We can calculate the power dissipation within such a series resonant crystal with the common formula $P = V^2/R_m$.

At the crystal's normal series resonant mode, which is where the crystal's motional resistance R_m dominates, the maximum RMS voltage that should be placed across a particular crystal will range from 2 V at 1 MHz to as low as 0.3 V at 30 MHz. This means that some oscillators running with relatively sedate RF amplitudes at the crystal's series resonance may actually give the crystal a drive signal that is too high. The cure for this is to lessen the feedback loop's amplitude by:

1. Using a feedback emitter resistor.
2. Using two parallel diodes to clip the feedback voltage.
3. Using a resistor in the crystal's feedback loop.

Not only is the crystal prone to stress failure when an oscillator is asked to generate elevated RF output powers, but so is the oscillator's active device itself. Within the feedback loop the transistor must not be presented with too much power at its input port, since this can easily damage or destroy it.

Follow this formula to prevent this:

$$P_{\text{max(IN)}} > P_{\text{max(OUT)}} - P_{\text{loop}}$$

where $P_{\text{max (IN)}}$ = highest allowable device power at its own input, dBm
$P_{\text{max (OUT)}}$ = highest possible power generated at the output of the active device, dBm
P_{loop} = losses from the output of the active device around the loop to its input, dB

Because of the crystal oscillator's basic inability to generate high powers without damaging its own crystal element or its active device, an RF power amplifier can be placed at the oscillator's output port when we require higher RF output levels.

Different crystal oscillator circuit configurations are required due to the different impedance levels found at various oscillator frequencies. In fact, since a crystal's internal series resistance can be as low as 20 Ω at 25 MHz, all the way up to 0.25 MΩ at 500 Hz, special circuit designs are needed to efficiently match and drive the crystals at these diverse resistance values. Nonetheless, the Pierce oscillator circuit as outlined above will be almost ideal for the majority of crystal oscillator requirements in most wireless systems.

Oscillator start-up time is connected to the Q of the oscillator's resonator, so the higher the Q, the longer the start-up time. Crystal oscillators, with their ultra high Q's, have prolonged start-up times of up to, and sometimes surpassing, 100 ms. This start-up time will also be significantly affected by the bias network of the oscillator's own active device, as the circuit must reach a steady state value before any reliable oscillations can occur. Consequently, the RC time constant of the bias network can substantially slow down the onset of frequency-stable oscillations.

All passive and active components that comprise an oscillator circuit must, of course, be rated above the oscillator's own frequency of operation (as must their

maximum voltage, current, and power ratings). Therefore, inductors and capacitors must not have any series or parallel resonances that will interfere with oscillations, and the active element itself must have a gain and phase shift that is more than sufficient to sustain oscillations at the frequency of operation.

Depending on the application, a crystal oscillator may require higher frequency accuracy over temperature than a normal, noncompensated crystal oscillator (XO) can possibly supply. This will demand that some type of compensated crystal circuit be used, such as that within *temperature-controlled crystal oscillators* (TCXO) and *oven-controlled crystal oscillators* (OCXO). Increased size, cost, current consumption, and complexity are the tradeoffs if such a high quality oscillator is to be adopted. However, depending on the angle of the cut for AT crystals, frequency stability over a desired temperature range can be optimized for even a simple, uncompensated crystal oscillator, and thus sometimes making compensated oscillators unnecessary: frequency stabilities of ± 5 ppm from +25 to +70°C are possible with the appropriate AT-cut. Wider temperature variations than this will quickly degrade an AT-cut's frequency stability dramatically, down to ± 20 ppm from –40 to +80°C, necessitating the use of a TCXO, or even an OCXO.

If high-frequency stability of better than a few ppm is required, then both the crystal and the entire oscillator circuit itself must be ovenized within an OCXO. The OCXO ovenizes not only the crystal, but all of the temperature sensitive components. It has the highest stability commonly available in compensated crystal oscillators, with better than 0.001 ppm being common using SC-cut or AT-cut crystals over a wide temperature range. The oscillator circuit is kept in a temperature-controlled oven that maintains the crystal and circuits at a temperature that is 10° above the highest specified ambient temperature. The OCXO can even be tuned very slightly (by a few ppm) by a small screw located within the case. However, OCXOs are high in cost, consume much more current than a standard oscillator, have a certain warm-up period to reach full frequency accuracy, and may have poor aging characteristics due to the high heat that the crystal is constantly subjected to.

The TCXO is a highly temperature stable crystal oscillator at better than 1 ppm. TCXO's make use of a temperature sensor, normally a *thermistor*, which generates a correction voltage to the TCXO's compensation network, a varactor diode, which then overpowers the frequency drifting effects of the changing temperature within the oscillator's circuits. TCXOs are lower in cost, with significantly less current requirements than an OCXO, and warm up nearly instantaneously.

Most of the active and passive components making up any oscillator are temperature sensitive, especially important being the crystal and the ceramic capacitors of the resonator network. So, even the finest crystal oscillator, if built with poor or inappropriate ceramic capacitors, may have unacceptable frequency drifts. Use of incorrect temperature compensating ceramic capacitors, or capacitors with poor temperature tolerance versus capacitance, can destroy frequency stability of an otherwise good oscillator. Indeed, in a well-designed oscillator, the majority of the long- and short-term frequency drift should originate only from within the crystal itself, and any oscillator circuit that adds more than double the drift of a lone crystal is considered as being improperly designed.

4.5.5 Crystal SAW-Based Oscillators

SAW-based oscillators (Fig. 4.38) are becoming more popular in the VHF and above regions, and are similar in design and concept to crystal oscillators. However, unless the oscillator's center frequency is a common one, SAWs are somewhat limited in their

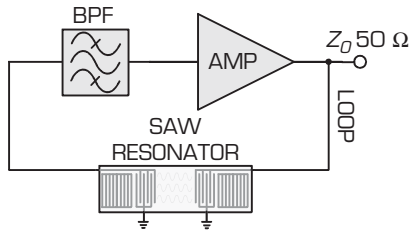


Figure 4.38 A SAW oscillator circuit, showing internal SAW resonator structure.

design usefulness, since no SAW resonators would be available (at least without an expensive custom production run). And their initial frequency stability, temperature stability, component *Q*, and aging characteristics are many times worse than the average crystal resonator. Nonetheless, replacing crystals with SAW resonators makes it possible to operate at very high frequencies of up to 2 GHz, and with powers of up to +22 dBm.

There are certain design peculiarities in a SAW oscillator that we should be aware of. When used as a resonator in an oscillator circuit, it may actually be advantageous from a *Q* and input loss perspective to *not* match a SAW to the active device's impedance, since matching the SAW can degrade the entire circuit's performance level. And to force a perfect zero phase shift around the entire loop, a phase-shifting network will be required after the SAW resonator. Depending on the direction of the phase shift needed, either a Butterworth highpass or a Butterworth lowpass filter can be used, with the component values and number of poles optimized within a linear simulator for the ideal phase shift. To avoid generating spurious frequencies other than the desired one, a SAW oscillator design may also require that an LC bandpass circuit be placed at the transistor's input port.

4.5.6 Crystal Oscillator Testing and Optimizing

We must test and optimize the completed physical crystal oscillator for proper start-up, frequency, amplitude, spurs, and the like, under real-world conditions. We may do this in the following way: Connect the closed-loop oscillator to a spectrum analyzer. Confirm that the oscillator reliably starts at room temperature by turning it On and Off a number of times, and over its full supply voltage. Cool the circuit with a canned cooler and repeat above. Heat the circuit with a heat gun set on low and repeat. Attach a stub tuner to the output, and pull the oscillator around the Smith chart to check for undesired stage oscillations and frequency stability. Is the oscillator stable over frequency at room temperature, and during and after the heating/cooling test? Are there any close-in spurious responses? Are there any wide-ranging spurious responses or excessive harmonic levels? What is the approximate phase-noise level? Are we outputting the RF amplitude we expected, and is it stable over time?

It will normally be necessary to improve the amplitude, stability, starting, and spectral purity of the crystal oscillator by tuning for optimum performance. Therefore, we should optimize the L/C or R/C ratio, the transistor's bias current, and the tuning capacitor. Tune until some best possible point is reached that satisfies the required specifications.

CHAPTER 5
Frequency Synthesis Design

A method of combining the wide tunability of LC oscillators with the high frequency stability of crystal oscillators is a necessity in modern wireless communications design. We find both of these abilities within *frequency synthesis*, which is a method of generating a multitude of exceptionally accurate frequencies from a single, low-frequency crystal oscillator. It is the dominate technique today for variable frequency production in most receivers, transmitters, transceivers, and test equipment.

By far the most widespread method of frequency synthesis is implemented by the *phase-locked loop* (PLL). A newer technique, referred to as *direct digital synthesis* (DDS), sometimes called an NCO or *numerically controlled oscillator*, is becoming increasingly more prevalent in certain limited applications. We will concentrate on the PLL, which is much easier to design, more versatile, and operates at a higher frequency.

5.1 PLL Frequency Synthesis

5.1.1 Introduction

The majority of frequency synthesis circuits are derived from the phase-locked loop. The most common type, the *integer-N* PLL, has an output frequency that is a exact multiple of the input reference frequency. In other words, a phase-locked loop multiplies a reference frequency by the value placed within a programmable frequency divider, and then outputs a stable frequency from a voltage-controlled oscillator.

Even though PLL circuits have many advantages, such as a much higher possible operating frequency than a crystal oscillator, fully tunable in discrete steps, and will be as stable as the low-frequency reference source, PLLs are still far inferior to crystal sources when it comes to phase noise specifications. This can be a problem in advanced digital wireless communications.

5.1.2 PLL Operation

Delving much further within the circuit of Fig. 5.1, which comprises a common single-loop PLL synthesizer: A low-frequency crystal oscillator, the *reference oscillator*, feeds the highly stable frequency of f_{osc} into the *R-divider*, which decreases f_{osc} to the same frequency as the f_{COM} that is coming out of the *adjustable N-divider.* This new f_{REF} frequency out of the R-divider is then inserted into the *PFD* (*phase-frequency detector)*, which compares the phase of the R-divider's f_{REF} signal to that of the N-divider's f_{COM} signal. The N-divider receives its own input frequency from the VCO's output as f_{out}, and then lowers $f_{out's}$ frequency to a value of f_{COM}, which must be exactly equal to the R-divider's f_{REF} output. As the PFD is comparing the two frequencies of f_{COM} and f_{REF} at its input from

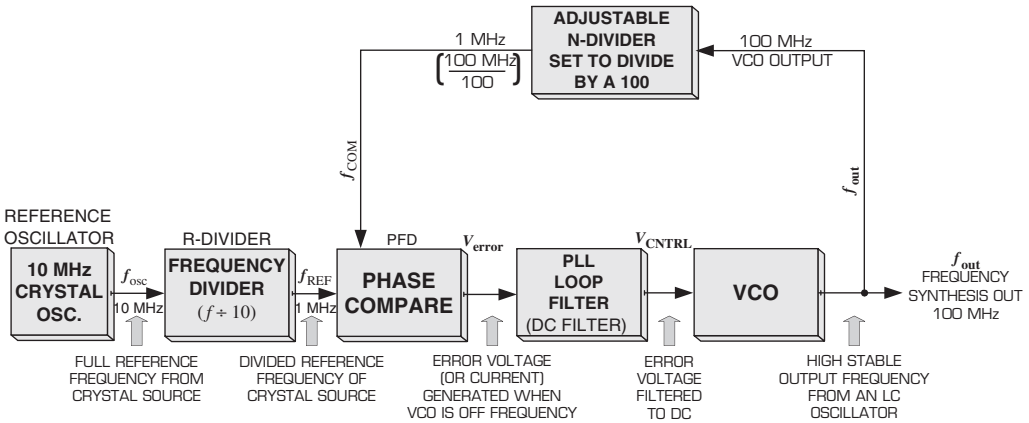


FIGURE 5.1 The phase-locked loop frequency synthesizer.

the N- and R-dividers to see if they are of the same phase, it will produce a rectified DC correction voltage, V_{error}, at its output port, which is then placed into the *PLL loop filter* if these two f_{COM} and f_{REF} frequencies differ. This lowpass filter almost completely eliminates any AC variations and noise products emanating from the PFD's output port, and places the now nearly pure DC correction voltage (V_{CNTRL}) directly into the VCO's frequency control input. The VCO immediately generates a variable and controllable output frequency that is as stable as the reference oscillator's f_{osc}.

Loop Filter

The all-important PLL loop filter is required to filter powerful phase comparator constituents at the comparison frequency of f_{COM}, and its harmonics, since if these responses actually reached through to the VCO stage they would adversely modify VCO stability. A standard three-pole loop filter (Fig. 5.2), which will form a fourth-order loop (with the other pole supplied by the PLL), is comprised of R_2 and C_2, which are the

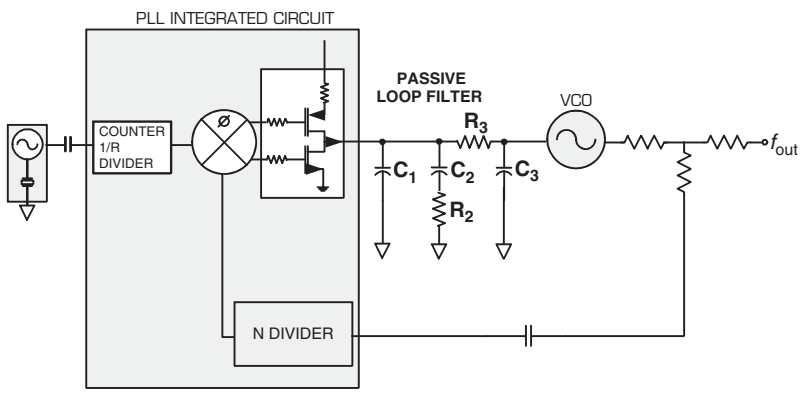


FIGURE 5.2 A standard loop filter for a PLL.

main AC-to-DC filtering elements: C_1, which aides in AC attenuation, and R_3 and C_3, used for additional lowpass filtering to further reduce PLL spur amplitudes.

The loop filter's capacitors should not be of the ceramic type due to ceramic's natural piezoelectric noise generation capabilities, with low leakage film capacitors being the preferred type. Also, the loop's resistors should not be the noisy carbon composition types, but only of the thin or thick film variety.

Voltage Controlled Oscillator

The varactor diode's bias within the voltage controlled oscillator (VCO) circuit is controlled by the PLL's DC voltage V_{CNTRL}, which immediately forces the VCO back on frequency if it has drifted off. This action permits a varactor LC oscillator to be adjustable over many discrete frequencies, and with the full stability of a crystal oscillator.

N-Divider

The N-divider is made of three separate counters, the *prescaler*, the *A counter*, and the *B counter*. By far the most common N-divider/prescaler is the *dual modulus* type, which supplies two separate prescaler divide rates, permitting the synthesizer to function at rapid tuning speeds and high frequencies, and with a relatively fine tuning resolution. It accomplishes this by internally dividing the input frequency by more than one divide ratio, such as 8/9 or 16/17. Consequently, instead of only being able to reach a specific frequency by dividing by, let's say, 8 a certain number of times, the PLL's 8/9 prescaler can now also divide by not only 8, but also by 9 for a set number of times. This gives us for more frequencies that we can synthesize, while also maintaining a decent frequency step resolution.

Therefore, prescalers (Fig. 5.3) take the high frequency of the VCO and divide it down to a more manageable lower frequency in the N-divider section: The VCO frequency is fed into this prescaler, which divides the frequency down to $P + 1$, in which P stands for the size of the prescaler. At each of these $P + 1$ cycles, $A + B$ counters decrement by 1. This creates a count of $A(P + 1)$ and $(B - A)P$, which makes $N = A(P + 1) + (B - A)P$, or $P(B + A)$. N is not permitted to be less than $P(P - 1)$, and if N is *less* than this $P(P - 1)$, then $B \geq A$, with:

$$B = \frac{N}{P} \quad \text{and} \quad A = N - (B \times P) \quad \text{and} \quad N = \frac{f_{out}}{f_{COM}}$$

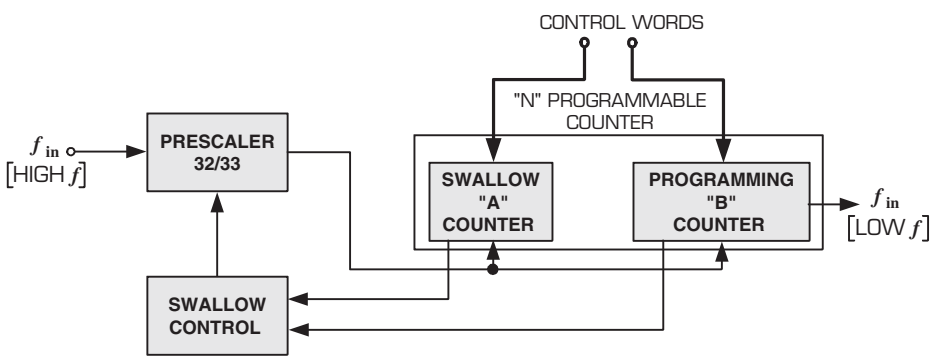


Figure 5.3 A block diagram of a prescaler.

The final outcome of using these dual-modulus prescalers in N-divider PLLs is that it becomes possible to control the division ratio into the phase comparator in steps of 1(N), as opposed to the huge steps of 32 or 33 in a fixed modulus 32 or 33 prescaler. This N value must always be an integer, with the largest N value being determined by the size of the B counter, since $N = P(B + A)$. Dual-modulus prescalers will, however, have certain illegal divide ratios, in which *specific* frequencies cannot be generated. If a particular N value results in a B register that is smaller than the A register, this will not be allowed; since B must be greater than or equal to A for a legal divide ratio. In other words, not all N values are allowed with a dual-modulus prescaler equipped PLL. The tradeoff between having certain frequencies that are impossible to generate is that we can obtain better frequency resolution at the PLL's output than would normally be possible. However, if certain frequencies must be generated by the PLL dual-modulus prescaler (since N must equal $(P + 1) \cdot (A + P) \cdot (B - A)$ to be a legal divide ratio), then a legal divide ratio check should be performed by using National's *Easy PLL* or National's *Code Loader* program.

Phase-Frequency Detectors

The PLL's phase-frequency detector detects a change in phase between f_{COM} and f_{REF}. It does this by lining up the rising edges of f_{COM} and f_{REF}, and then outputs control signals to the charge pump to tell it to sink or source current in or out of the lowpass loop filter to keep the PLL locked. The charge pump outputs a single amplitude, but changeable duty cycle, current pulse which the loop filter converts into a DC output voltage. This filtering action removes most of the charge pump's glitches and overshooting, instructing the VCO to move slightly higher or slightly lower in frequency if it has drifted or if the operator/command-signal desires a rapid change in frequency. When the f_{COM} and f_{REF} inputs to the PFD are perfectly aligned in both frequency and phase, then the loop is deemed fully locked. This locked condition is very temporary due to the extremely poor frequency stability of a VCO. During the short period of perfect lock when f_{COM} and f_{REF} are in perfect frequency and phase alignment, the charge pump will tri-state with a high-impedance output and with narrow (approximately 30 ns wide) positive and negative charge-pump current pulses at a 50-50 duty cycle.

The outputting of pulses even when in perfect lock is done to prevent *dead-band* behavior. The PLL dead-band is a condition in which the phase-frequency detector would have no control of the loop when the PLL was very close to, or is actually in, lock. In other words, when there is an almost 0° phase difference between f_{COM} and f_{REF}. (This is so, jitter is not created with the excessive PLL loop hysteresis.) And even though the PFD is still outputting current pulses when f_{COM} and f_{REF} are in perfect in-phase alignment, the charge pump itself will neither charge up nor charge down the loop filter's capacitors, since these charge-pump pulses are comprised of evenly spaced, 50-50 duty cycle, very narrow I_{up} and I_{dwn} pulses.

Unlike many of the old PLL detectors, all modern PLL chips use PFDs which will force a lock, even when the PLL itself is drastically out of lock. The PFD does this by first comparing, then finding the same f_{COM} and f_{REF} *frequency*, and only after it completes this wide frequency lock function does it then force an almost perfect "fine tune" *phase* lock for the synthesizer circuit.

Charge Pumps

Most modern PLL chips are of the *charge-pump* type (Fig. 5.4). A PLL with a charge pump permits the use of a passive filter, which is cheaper and adds little extra noise, unlike an active op-amp-based loop filter.

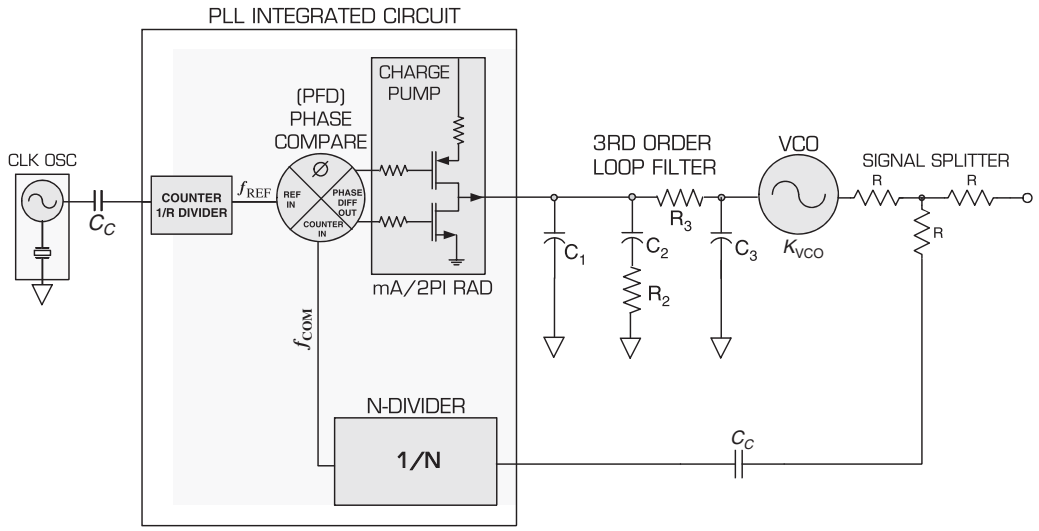


FIGURE 5.4 A charge pump PLL chip.

The built-in charge pump outputs a current of steady amplitude, but with a changeable duty cycle and polarity, into the PLL's loop filter. The charge pump's duty cycle changes depending on how far out-of-lock the VCO is: wider pulses mean that the VCO is farther out of lock than narrow pulses, and that the VCO requires a higher V_{CNTRL} level. The filter then converts this charge-pump current output into a DC control voltage for the VCO's input control port, V_{CNTRL}.

All PLL charge pumps can be considered as being comprised of two well-matched current sources that are switched fully on or fully off by the PFD. The charge pump itself can do only one of three things: it can source current, sink current, or go into a high-impedance (tri-state) mode.

A charge pump's output is connected to the simple loop filter of Fig. 5.5 to demonstrate charge-pump/loop-filter action: As soon as the PFD senses that f_{REF} and f_{COM} are not perfectly equal, a phase error is detected and the PFD sends a voltage command to the charge-pump circuit to turn on I_{up}, which begins to charge C_1 through R_1. This causes V_{CNTRL} to ramp up to a value of $I \cdot R_1$. I_{up} is then ordered by the PFD to shut *off* at the same time that I_{dwn} is ordered *on*, which draws the current back out of C_1, through R_1, and sinks it through I_{dwn}, causing the VCO control voltage V_{CNTRL} to drop to a lower value, and changing the f_{out} value from the VCO's output port. (We could actually calculate this V_{CNTRL} value by taking the average DC current that is pumped out of the charge pump and multiplying it by the loop filter's impedance.)

Low-voltage charge pumps, which are used with a typical passive loop filter, depend largely on the charge pump's voltage supply (and the charge pump's current setting) for the best lock time, since the charge-pump gain will increase in a linear manner with its voltage supply amplitude. Thus, the higher the charge pump's supply voltage, the faster the charge-pump output forces the loop-filter voltage to reach a steady-state value into the V_{CNTRL} port of the VCO, and therefore the faster the PLL lock time. This effect is due to the loop filter itself (Fig. 5.6), since as the charge pump begins to pulse current into the filter to force the VCO's voltage to some specific amplitude, these pulses begin to charge,

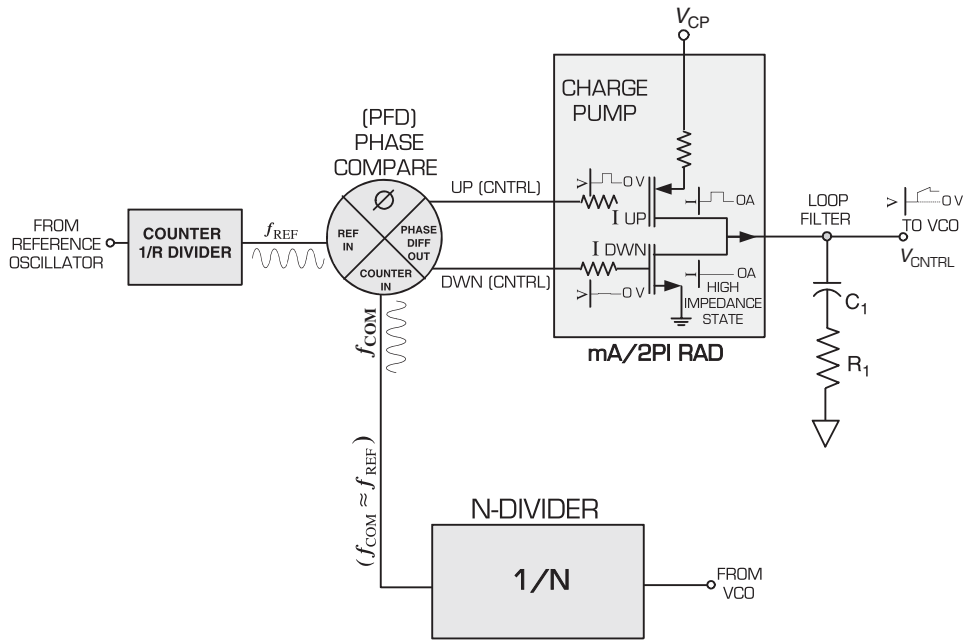


Figure 5.5 PFD, charge pump, loop filter action.

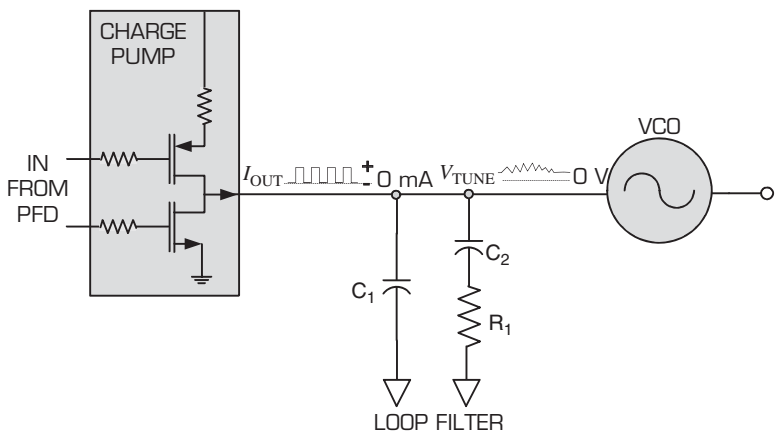


Figure 5.6 The waveform from the charge pump, and the output waveform from the loop filter.

principally, C_1, creating voltage increases at the loop filter's output. However, directly between these charge-pump current pulses, or at zero current, the charge in C_1 then discharges into R_1 and C_2, which decreases the voltage at the output of the loop filter, forming voltage charge/discharge ramps as shown. Since C_2 now has a charge, we need C_1's voltage level to be higher than C_2's if we want significantly more current flow into this larger value capacitor, C_2. But, due to the lower charge-pump voltage, the charging of

C_2 will now take longer because of the decreased current flow to C_2 caused by the lower voltage across C_1. End result: slower VCO lock time because the charging of the loop filter must trade minimum charge-pump voltage amplitude for charging time.

A typical PLL's IC charge-pump current gain setting is adjustable. Setting a charge pump to the maximum current gain setting can be advantageous, since it permits a third-order loop filter's resistors to be lower in value, creating less resistor noise. The loop capacitor that is in shunt with the VCO can also now be higher in value, which dramatically decreases the parasitic effects of the VCO's input capacitance (the VCO's input capacitance can be as high as 100 pF, with the exact input capacitance value varying over the VCO's full tuning range). Phase noise may also be improved for certain phase-locked loop chip models with a higher charge-pump current gain.

PLL Control Circuits

The adjustable N-divider of the PLL is usually controlled by the operator through a front radio panel knob, or automatically by system commands. A microprocessor will normally supply digital control words to the PLL through a serial, but sometimes even a parallel, bus to change frequency. The microprocessor can also be used to decode and drive display circuits to inform the radio operator of the exact transmit or receive channel.

A widespread manual PLL tuning scheme is shown in the circuit of Fig. 5.7. It has a shaft encoder, a microprocessor, the display with its driver, and a PLL chip with loop filter and VCO. Rotated by the radio operator, the tuning knob turns an optical or magnetic encoder that has two voltage outputs, *A* and *B*. The *A* output is a square wave in quadrature phase (90° phase shifted) to the *B* square wave's output. The *A* output is

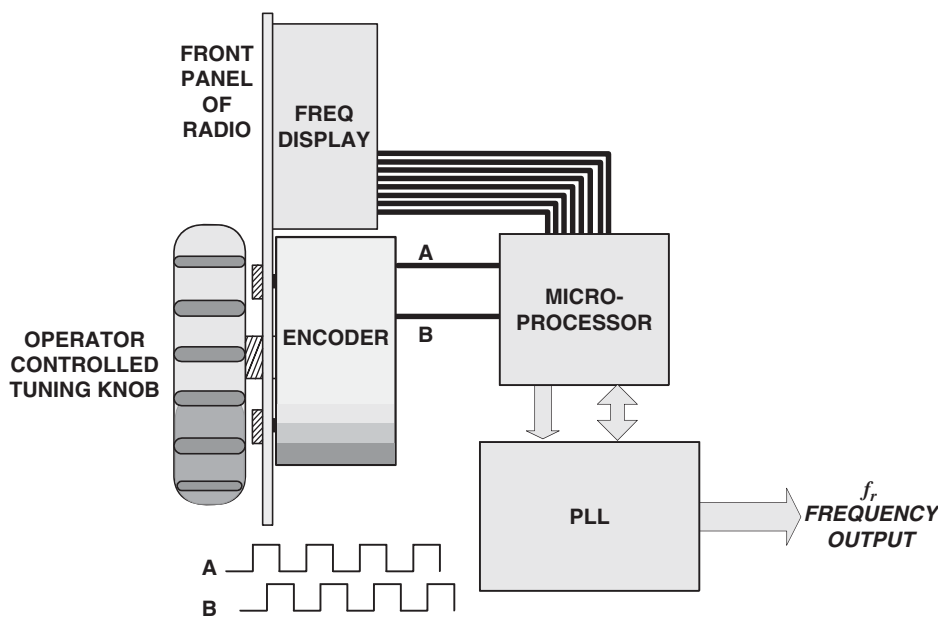


Figure 5.7 A common PLL manual tuning method.

connected directly to the microprocessor's interrupt line, so when the *A* output from the encoder produces a falling edge, an interrupt will occur. The microprocessor will then immediately look at the *B* output to see if it is a 1. If it is, then the microprocessor considers that the knob has been rotated clockwise. The microprocessor will then increment the PLL's N-divider by 1, increasing the output frequency, and update the frequency display appropriately. However, if *B* is a 0, then the microprocessor considers that the knob has been rotated counterclockwise, and the microprocessor decrements the N-divider by 1, decreasing its output frequency.

5.1.3 PLL Phase Noise

Since PLLs are used extensively as local oscillators in digital radios, and LO phase noise degrades a communications system's SNR and ACPR (adjacent channel power rejection), we see why such noise suppression is extremely critical.

The major noise sources within a PLL are the VCO's own phase noise, the reference oscillator noise, and the phase-frequency detector noise. Generally, close-in phase noise (i.e., noise within the bandwidth of the loop filter) is dictated by the phase-frequency detector and reference oscillator, while farther out the major phase-noise contributor is the VCO (Fig. 5.8).

To Minimize Phase Noise in a PLL

1. So as not to further degrade the phase noise that is generated within the PLL chip itself, use only an exceptionally clean reference oscillator, since the phase noise of the reference will be amplified by 20 log (*N*/*R*) within the loop filter's bandwidth.
2. Use a VCO designed for low noise, since this oscillator is the major wideband phase-noise contributor. (When a VCO is placed within the loop of a PLL, close-in VCO phase noise will be attenuated as compared to its free-running state,

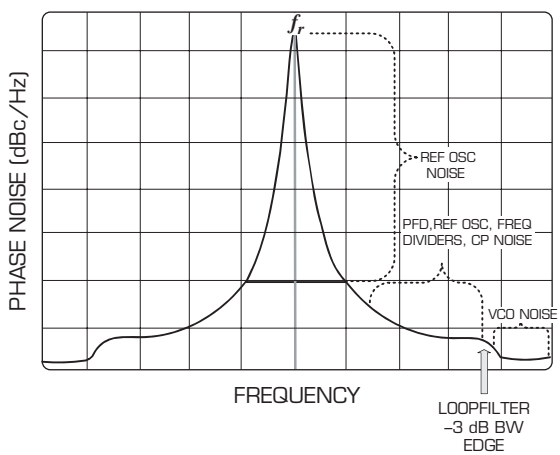


Figure 5.8 The major noise contributors and output spectrum of a PLL synthesisizer, narrow frequency sweep.

but the VCO frequencies that are farther away from the loop bandwidth will actually be increased in amplitude.)

3. Keep the number of the N-divider divisions as low as possible, as the higher the divide ratio, the higher will be the phase noise (the phase detector noise will equal 20 log 10(N), in dB). Thus, we would want to decrease the frequency of the VCO, while increasing f_{COM}, in order to lower this noise (phase noise is improved by a high f_{COM}). But this would also limit the channel spacing, since channel width equals f_{COM}. A fractional-N PLL can overcome this problem. (We can further improve phase noise, at the expense of increased cost and complexity, by utilizing multiple PLL loops to reduce the division ratio.)
4. Make the loop filter narrow, as narrow PLL loop bandwidths can provide decreased phase noise and lower spurious outputs. However, if the loop bandwidth is too narrow, the VCO's phase noise will begin to degrade close-in to the f_{out} carrier, as there will now be more of the VCO itself that is *not* within the loop bandwidth. Yet, if the loop filter is designed to be too wide, then the PLL chip's own internal noise bandwidth will be extended further outward from the f_{out} carrier, making for a less than optimal phase-noise profile for the synthesizer. So, the perfect loop-filter bandwidth is designed to be exactly at the frequency where the PLL chip's own internally generated noise amplitude equals the noise amplitude of a free running VCO. This will supply us with the best of both worlds: the lowest close-in and the lowest wideband phase-noise characteristics for a PLL.
5. Confirm that the loop filter's resistors are not of too high a value, since due to random electron motion all resistors will generate white noise. This noise power can be considered as an in-series noise generator, with the higher the resistance value the higher will be the noise produced.
6. If possible, avoid ceramic loop-filter capacitors because of their piezoelectric effects and microphonics, which can sometimes create noise transients. Also, never use carbon composition or carbon film resistors in a PLL loop filter due to their extreme random-noise-producing qualities; use only thin or thick metal film types.
7. Try increasing the VCO's output feedback signal into the N-divider port, if it is too low in amplitude, since a small feedback level back into the PLL's IC can increase noise levels.
8. Low reference oscillator drive levels into the PLL's R-divider will increase noise. Make sure the reference oscillator is inputting the recommended R-divider input signal amplitude.
9. Confirm that the phase-locked loop or VCO power supply is properly filtered by decoupling networks at both low-frequency AC and at high-frequency RF. Also, resistively isolate the phase-locked loop IC's power supply line from the VCO's power supply line by employing a low-value series resistor in each line of approximately 20 Ω.
10. Make sure that the VCO's loop filter and the VCO's V_{CNTRL} tuning line are not too exposed to EMI or digital noise, as well as to any pulse coupling from the charge pump's output. Keep the VCO tuning line short, and away from noisy circuits of any kind. These issues can be aggravated by excessively cramped PCB part's placement or a generally poor PCB layout.

Measuring Phase Noise

We can measure the phase noise of a synthesizer by attaching the PLL's control lines to a PC that is loaded with the appropriate PLL control software, then placing a spectrum analyzer's own low phase-noise 10-MHz reference oscillator output into the PLL's f_{osc} input (or REF_{IN}), and feeding the PLL's f_{out} signal back into the spectrum analyzer's 50-Ω RF input port.

Most modern spectrum analyzers will have built-in phase-noise profiles that perform this measurement at the touch of a few buttons, with results displayed in dBc/Hz in a 1-Hz bandwidth, by using the *marker noise function* at a specified frequency offset from the carrier (such as 10 kHz). See Sec. 12.2.5, *Phase-Noise Test*.

5.1.4 PLL Reference Spurs

Reference spurs are spurious PLL outputs that form outside of our desired passband. These spurs can be found on each side of the VCO's carrier, at a frequency offset that is equal to the PLL's channel spacing, and are created by PLL charge-pump leakage and mismatch. Spurs must be reduced in amplitude as much as possible to prevent interference into the passband from a strong adjacent channel mixing with these spurs, and creating in-band interference within the receiver's IF frequency. (Keeping reference spurs low in amplitude will minimize *reciprocal mixing* of these LO spurs with any large amplitude RF interferers that are within the receiver's first mixer stage, and which can overwhelm the low-amplitude desired signal within the IF. This could seriously decrease a receiver system's sensitivity.)

The spurs are produced, as stated, by the charge pump's own output pulses, which are being generated even when the phase-locked loop is in perfect lock. Indeed, while both a charge pump's I_{up} and I_{dwn} current pulses are normally designed to source and sink equal values of current, this is not really possible. There will always be a slight current offset, and any mismatch between I_{up} and I_{dwn} currents will create these reference spurs. The pulses will then modulate the VCO's DC control voltage V_{CNTRL} at a frequency that is exactly at f_{COM}, creating reference spurs at f_{out} that are at the exact same frequency of f_{COM} and its harmonics, and on both sides of the PLL's carrier f_{out}. Therefore, a value of f_{COM} of 100 kHz will give us reference spurs at +100 kHz and −100 kHz, and at integer multiples thereof, away from f_{out}.

To make sure that the reference spurs are as low as possible in divided-by-N PLLs, we should select a synthesizer chip that has low charge-pump-leakage characteristics, and that has well-matched charge-pump currents. However, the only remedial action we could easily take with an already built PLL would be to improve the filtering of the close-in spurs by tightening up the loop filter's bandwidth, as well as making sure that the V_{CC} power supplies are extremely well filtered.

Any assorted spurs that are over and above the typical f_{COM} reference spurs can sometimes be seen in the output of some synthesizers at assorted frequencies. These irregular spurs can be caused by conducted or radiated coupling of noise or RF interference into the PLL. The cure is the standard one, and can significantly reduce these atypical spurs to low levels: effective decoupling of all PLL voltage supplies, as well as proper EMI board layout procedures.

5.1.5 PLL Lock Time

PLL lock time can be shortened by raising the f_{COM} frequency, since f_{COM} controls the charge pump's pulse frequency, and will also permit an increase in the loop-filter

bandwidth. While increasing the loop bandwidth will decrease lock time, it will also increase phase-noise and spur amplitudes, as well as possibly creating loop instability if widened by more than $f_{COM}/5$. However, the lock time can be considerably sped up by never permitting the VCO's tuning voltage to be any closer than 1 V from either supply voltage rail of the charge pump. This particular effect on lock time is caused by saturation of the charge pump at its supply voltage extremes, and can be circumvented by:

1. Using a VCO with a higher tuning sensitivity
2. Using a higher charge-pump voltage supply
3. Using an active op-amp loop filter with a higher V_{CC}
4. Using loop capacitor types that do not characteristically show signs of capacitor current *soakage*, since this internal dielectric effect will prevent the capacitor from fully discharging within the short time permitted. The plastic film types are superior to other capacitors in this specification.

5.1.6 PLL Design Procedure

The design of PLL frequency synthesizer circuits, until recently, was fraught with complications and uncertain results. However, PLL chip companies, such as National Semiconductor, have released information (and improved PLL integrated circuits), that make the design of a frequency synthesizer much more simplified than in the past. Both National Semiconductor and Analog Devices have also released amazingly powerful, and completely free of charge, PLL design programs that automate the PLL design task. The National Semiconductor online design tool, *EasyPLL*, is available at www.national.com, while the Analog Devices' PLL software, *ADIsimPLL*, is included with this book's CD ROM.

To Design a PLL

A. The first place to begin in the design of a PLL frequency synthesizer circuit is in deciding on what we want our center frequency, frequency swing (minimum and maximum frequencies), speed (lock time), and channel resolution (spacing) to be, then select the appropriate PLL chip, VCO, and reference oscillator that can quickly, economically, and repeatedly meet our criteria. The following list should be completed so that all the specifications required of a PLL circuit can be clearly defined (some of the below PLL designators or their meanings may vary slightly from one company to the next):

   f_{max} = Hz. The maximum output frequency required. f_{max} must never be higher than the VCO itself was designed to handle, along with about 20% frequency excess for a VCO safety margin. (However, using a VCO with *less* tuning bandwidth decreases its phase noise.)

   f_{min} = ____ Hz. The minimum output frequency required. Must never be less than the VCO was designed to handle, along with about 20% frequency excess for a VCO safety margin. (Using a VCO with *less* tuning bandwidth decreases its phase noise.)

   f_{out} = ___ MHz. $\sqrt{f_{max} \cdot f_{min}}$ (must be a multiple of f_{COM})

   K_{VCO} = ___ MHz/V. K_{VCO} is the VCO's gain (sensitivity) measured in MHz/V, and is the amount of frequency deviation in MHz that the VCO will travel

when 1 V_{DC} is placed at its V_{CNTRL} DC control input port (normally between 5 and 200 MHz/V).

$K\Phi$ = ___ mA/2π. $K\Phi$ is the *charge-pump gain*, and is the peak amplitude of the current that the charge pump will source or sink during each pulse of the charge pump. $K\Phi$ is measured in mA/2π, and should be chosen at the highest value the PLL chip will allow, so as to obtain the lowest phase noise at the VCO's output. This value will normally be either 1 mA or 5 mA, and is typically selectable. Therefore, a charge pump with a gain of 5 mA/2π would be preferred over a charge pump with 1 mA/2π . This charge-pump gain value will also vary with the charge-pump supply voltage.

f_{COM} = ___ kHz. Normally equal to the channel spacing. f_{out} and f_{osc} must be a multiple of f_{COM}. Due to the PLL's internal prescaler, it may be required to use an f_{COM} that is a certain fraction of the channel spacing. The higher the f_{COM}, the better the phase noise.

f_c = ___ kHz. The loop bandwidth of the PLL filter. f_c should be as narrow as possible to lessen spurious noise, but this will decrease switching speed. Normally f_c should be between 1 and 20 kHz, but must be at least 1/20 of f_{COM}. (National Semiconductor recommends a 2-kHz f_c if lock time does not matter.) The choice of f_c will always be a compromise between reference sideband suppression and lock time. Therefore, select a loop f_c to just meet the lock requirement, but with an acceptable margin.

ϕ (phase margin) = ___ DEG. Normally select a value between 30° and 70° for the loop filter. The higher the phase margin, the higher the PLL's stability, but the slower will be its lock time. Choose a phase margin of 45°, which is a good compromise between loop stability and loop response.

T_3/T_1 = ___ %. Normally chosen to be 45%. T_3/T_1 is $(T_3/T_1) \cdot 100$, and is the ratio of the poles of the loop filter declared as a percentage. The higher this value (closer to 100%), the more the reference spurs will be attenuated; but peaking will begin to occur within the filter's passband, and R_3 will also increase in value and thus add excessive thermal noise.

f_{osc} = ____MHz. The frequency of the reference oscillator *before* the R-divider. Must be a multiple of f_{COM}. 10 MHz is a popular value, as applicable.

B. After filling out these required parameters, design the complete frequency synthesizer by performing the following calculations, or simply use the online version of National Semiconductor's *EasyPLL* program, or the included Analog Devices' *ADIsimPLL*:

1. $N = \dfrac{f_{out}}{f_{COM}}$

2. $\omega_C = 2\pi f_c$

3. $T_1 = \dfrac{\left(\dfrac{1}{\cos(\varphi)}\right) - \tan(\varphi)}{\omega_C\left(\dfrac{T_3/T_1}{100} + 1\right)}$

4. $T_3 = \dfrac{T_3 / T_1}{100} \cdot T_1$

5. $T_2 = \dfrac{1}{\omega_C{}^2 \cdot (T_1 + T_3)}$

6. $C_1 = \dfrac{T_1}{T_2} \cdot \dfrac{K\varphi \cdot K_{VCO}}{\omega_C{}^2 \cdot N} \cdot \left[\dfrac{1 + \omega_C{}^2 \cdot T_2{}^2}{(1 + \omega_C{}^2 \cdot T_1{}^2)(1 + \omega_C{}^2 \cdot T_3{}^2)} \right]^{1/2}$

7. $C_2 = C_1 \left(\dfrac{T_2}{T_1} - 1 \right)$

8. $C_3 = \dfrac{C_1}{10}$

9. $R_2 = \dfrac{T_2}{C_2}$

10. $R_3 = \dfrac{T_3}{C_3}$

If a wide tuning VCO is required in a broadband synthesizer design, then more DC tuning voltage for the VCO will also be needed, since very wideband oscillators may demand up to 20 or more tuning volts—but a typical narrowband PLL chip may only be able to supply 5 V or less. This increase in the necessary DC tuning voltage for a wideband VCO can be accomplished by employing a separate op-amp within the PLL filter as shown in Fig. 5.9. The VCO gain would then be:

$$\text{VCO}_{\text{GAIN}} = K_{\text{VCO}} A_v$$

where A_v = voltage gain of the op-amp, K_{VCO} = gain of the VCO, MHz/V.

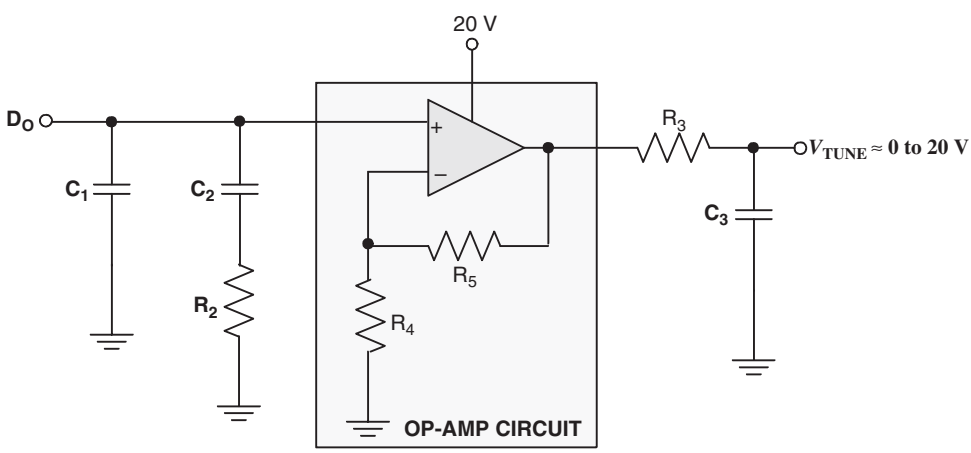


Figure 5.9 Active loop filter used to increase the tuning voltage for high V_{TUNE}, wideband VCOs.

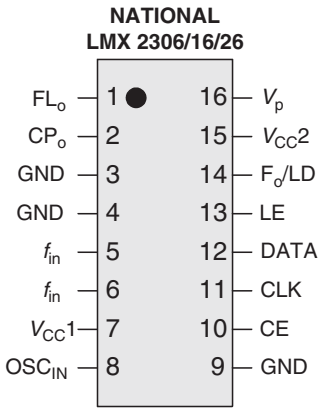


Figure 5.10 A popular PLL chip, the National LMX23XX series.

The entire PLL design will still be the same as in steps 1 through 10 above, but now simply substitute K_{VCO} for VCO_{GAIN}.

Another similar wideband tuning technique is to place a low-noise high-supply voltage op-amp directly at the DC tuning input of the VCO, with the loop filter's output placed into the input of the op-amp, and use the VCO_{GAIN} formula above to calculate the new gain of the VCO. The result of the VCO_{GAIN} calculation will then be used as the new K_{VCO} in the above PLL formulas.

C. The above completes the design of the most important part of any PLL synthesizer, the loop filter. The following will wrap up the total frequency synthesizer design by employing a National LMX23XX PLL chip (Fig. 5.10).

The complete National PLL chip's input and output pins are described in detail below for the 16 pin TSSOP package of the popular LMX2306 (which functions up to 550 MHz), the LMX2316 (1.2 GHz), and the LMX2326 (2.8 GHz):

1. Fl_o is an output pin that permits a parallel resistor to be attached between C_2 and R_2 of the PLL's loop filter. This will allow the PLL to obtain both a fast lock time and good phase noise specs by modifying the loop bandwidth on the fly. After the channel change occurs, loop bandwidth reverts back to normal.
2. CP_o is the output of the charge pump to which the loop filter is attached.
3. Pin 3 GND is for the charge pump, and can be shorted to Pin 4 and attached as directly as possible to system ground.
4. Pin 4 GND is for the analog circuits, and can be shorted to Pin 3 and attached as directly as possible to system ground
5. $-f_{in}$ should be AC shorted to ground through a 100-pF capacitor.
6. $+f_{in}$ accepts the signal from the VCO's output through a series 20 to 200-Ω resistor. This series resistor, whose value is dependent on the VCO's output power, will lower the power into the preselector, allowing most of this energy to be delivered to the load.

7. $V_{CC}1$ pin is the heavily bypassed DC analog power supply voltage input. The input voltage for this particular chip may be anywhere between 2.3 and 5.5 V, but must equal pin 15's $V_{CC}2$ voltage.
8. OSC_{IN} is the reference oscillator input for a CMOS 100 kΩ output resistance clock oscillator. A clean crystal clock input is vital for a low-phase-noise PLL output.
9. GND is digital ground. Should reach system ground by as direct a route as possible.
10. CE is the chip enable pin for power down for power saving operation. Can be tied to V_{CC} if this feature is not required.
11. CLK is an input that accepts a CMOS clock signal from the channel select microcontroller for clocking data into pin 12.
12. DATA input accepts data from the microcontroller for the R-divider, the N-divider, and the Function Latch (which controls phase detector polarity, FastLock modes, F_o/LD, counter reset, CP tri-state, test modes, and so on), with the last two bits (control bits) informing the PLL as to whether the data should be sent to the R-divider (0,0), the N-divider (1,0), or the Function Latch (0,1) on command of pin 13, LE.
13. LE (load enable) pin controls when the PLL's registers will send data to the R-divider, N-divider, or Functions Latches, depending on the control bits.
14. F_o/LD is an output pin that can typically be used as a lock detect (LD) output pin into a microprocessor, or into some out-of-lock alarm. A High will be output when the PLL is in lock. (On advanced PLL chips, such as with the National line, a trace may be taken from the LD pin back to the microprocessor. The pin, if digital lock detect is chosen by programming the proper PLL register, will output a high as long as the VCO output frequency is locked. This high or low signal can then be exploited by the microprocessor to indicate an unlocked condition by an LED warning on a display, or as an automatic shutdown of a run-a-way transmitter.)
15. $V_{CC}2$ is the digital power supply voltage input pin, and should be tied to pin 7, which is the analog power supply input.
16. V_p is the power supply for the charge-pump circuit, and must be greater than V_{CC}. (The DC control voltage into the VCO will always be a few tenths of a volt less than V_p, so V_p must have the proper amplitude to fully drive the VCO's DC control input, V_{CNTRL}.)

D. After completing the filter's design calculations for the frequency synthesizer, the following final PLL design checks must be performed to confirm that the PLL will function as desired:

1. The loop bandwidth, f_c, should be *at least* ½0 of f_{COM}.
2. Make sure that C_3 is at least five times larger in value than the input capacitance of the VCO (which is usually around 20 pF for the average VCO input capacitance).
3. Since the maximum PLL phase-frequency detector (PFD) input frequencies normally max out at 10 MHz, make sure that f_{COM} is not above this number.

4. R must generally be set to divide by at least 3 or more.
5. Check the completed PLL design to confirm that the *damping factor* (DF) is less than 1 by:

$$\mathrm{DF} = \frac{R_2 \cdot C_2}{2} \cdot \sqrt{\frac{K\varphi \cdot K_{\mathrm{VCO}}}{N \cdot (C_1 + C_2 + C_3)}}$$

6. Check that $R_3/R_2 > 2$.
7. The *optimization index* (OI) can be checked by:

$$\mathrm{OI} = \frac{\dfrac{T_2}{1 + (\omega_C \cdot T_2)^2}}{\dfrac{T_1}{1 + (\omega_C \cdot T_1)^2} + \dfrac{T_3}{1 + (\omega_C \cdot T_3)^2}} \times 100$$

NOTE: *This formula confirms that the loop is stable with a fast lock time. Any value between 90% and 100% is considered stable.*

8. A roughly estimated *lock time* (LT) can be found by:

$$\mathrm{LT} = \frac{-Ln\left(\dfrac{\mathrm{TOL}}{f_2 - f_1}\right) \cdot \sqrt{1 - \zeta^2}}{\zeta \cdot \omega_N}$$

where $\omega_N = \sqrt{\dfrac{K\varphi \cdot K_{\mathrm{VCO}}}{N \cdot (C_1 + C_2 + C_3)}}$

$$\zeta = \frac{R_2 \cdot C_2}{2} \cdot \omega_N$$

LT = lock time, s
f_2 = higher frequency, Hz
f_1 = lower frequency, Hz
TOL = frequency tolerance, Hz (acceptable frequency error at lock)

9. To check for approximate phase noise (PN) of the PLL at 150 Hz from the center frequency: PN (at 150 Hz) = 205* − (10 log f_{COM}) + (20 log (N counter value)).

A Quick Example Design a Passive Loop PLL Synthesizer (Fig. 5.11)

Goal: Create a phase-locked loop synthesizer with an RC loop filter. The specifications and parameters for the circuit are:

*205 is the average value for a typical PLL chip, and is referred to as the PLL's 1-Hz normalized PN floor, or 1-Hz PNF. This formula does not take into account an excessively noisy VCO.

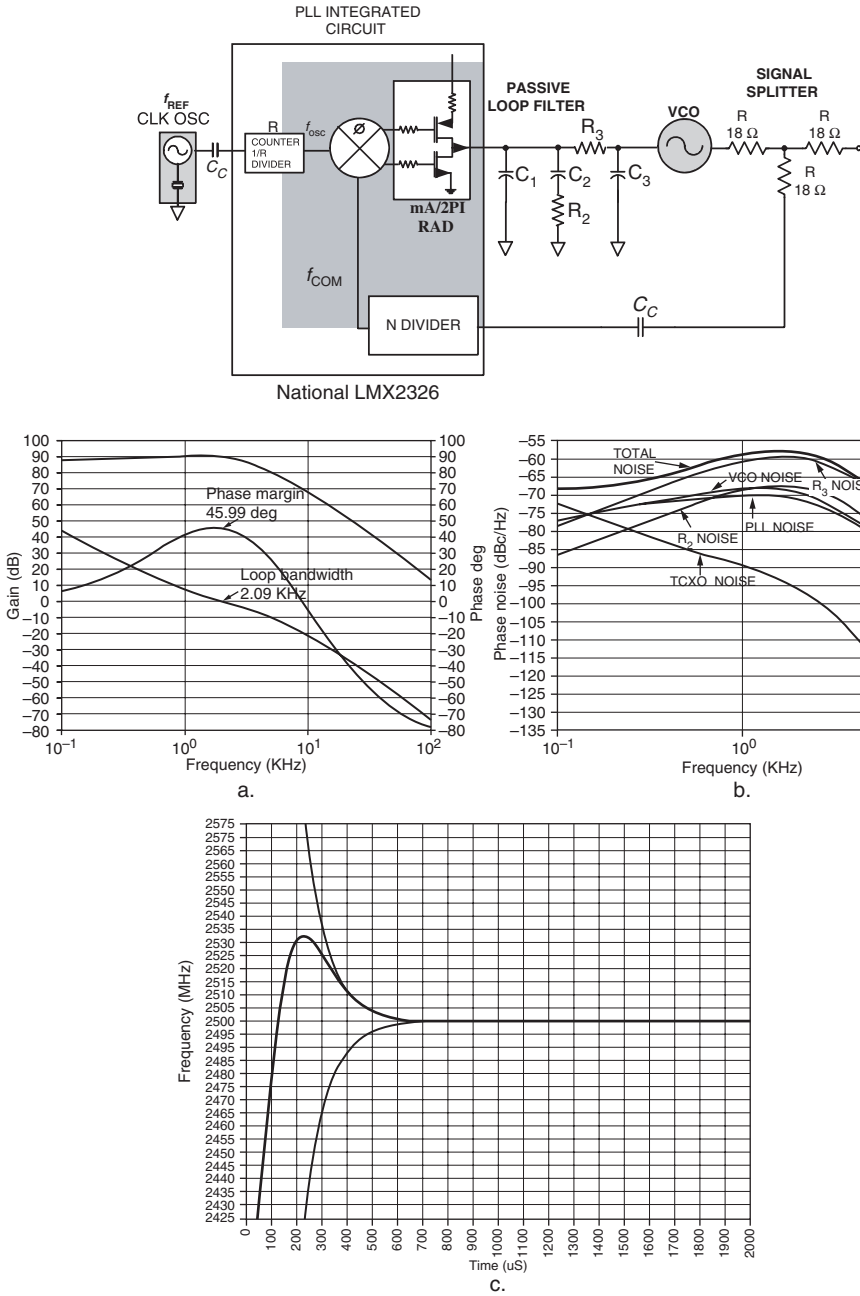


Figure 5.11 A passive loop-filter PLL design example: (*a*) loop gain/phase response; (*b*) phase noise (heavy top line is the total noise from all the PLL noise contributors); (*c*) PLL lock time.

f_r = 2.4 to 2.5 GHz (center frequency = 2.45 GHz)
V_{CC} = 5 V
f_{REF} = 10 MHz
f_{COM} = 100 kHz
Channel spacing = 100 kHz
Filter = third order
Loop bandwidth = 2 kHz
Charge-pump gain = 1 mA
Phase margin = 45°
T3/T1 ratio = 45%
PLL IC = National LMX2326
VCO = RFMD VCO790-2300T (188.5-MHz/V, 68-pF input capacitance)

Solution:

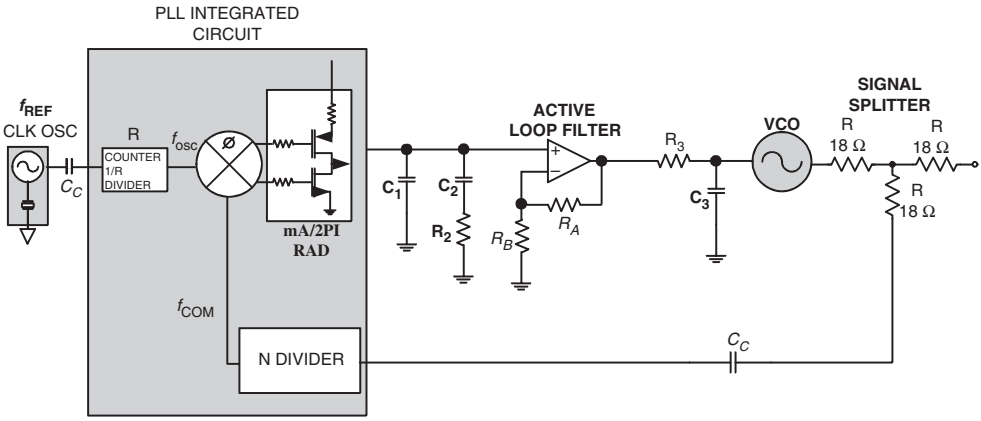
1. Using the calculated values as presented, or by quickly employing the online version of National's *EasyPLL*, enter the above desired PLL specifications.
2. Run the software, which should results in the following loop-filter values:
   a. C_1 = 10 nF
   b. C_2 = 120 nF
   c. C_3 = 1.5 nF
   d. R_2 = 1.8 kΩ
   e. R_3 = 10 kΩ
3. Analyze the design. The results should be:
   a. Phase noise
      PN@0.14 kHz = −68.60 dBc/Hz
      PN@10.00 kHz = −77.35 dBc/Hz
   b. Filter (Bode)
      Loop BW = 2.03 kHz
      Phase margin = 45.39°
   c. Lock time
      Lock time = 1289 μs (Which, in this example, is until the loop settles to within 500 Hz of the desired frequency of 2.5 GHz after being ordered to travel up from 2.4 GHz.)

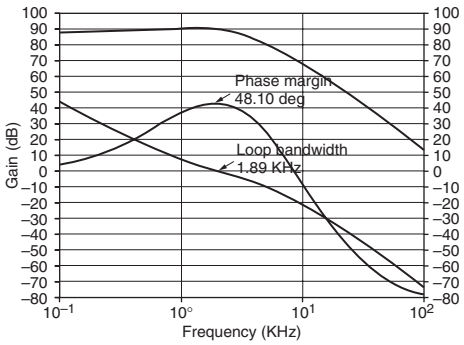
Or, if an active loop PLL synthesizer is called for, such as when a wide RF tuning range may be required, use the Quick Example below.

A Quick Example Design an Active Loop PLL Synthesizer (Fig. 5.12)

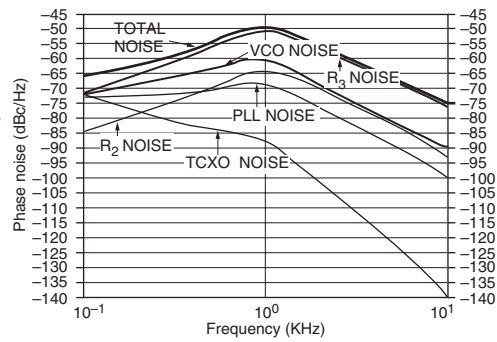
Goal: Create a phase-locked loop synthesizer with an active op-amp loop filter. The specifications and parameters for the circuit are:

f_r = 2.4 to 2.5 GHz (center frequency = 2.45 GHz)
V_{CC} = 5 V
f_{REF} = 10 MHz
f_{COM} = 100 kHz
Channel spacing = 100 kHz
Filter = third order

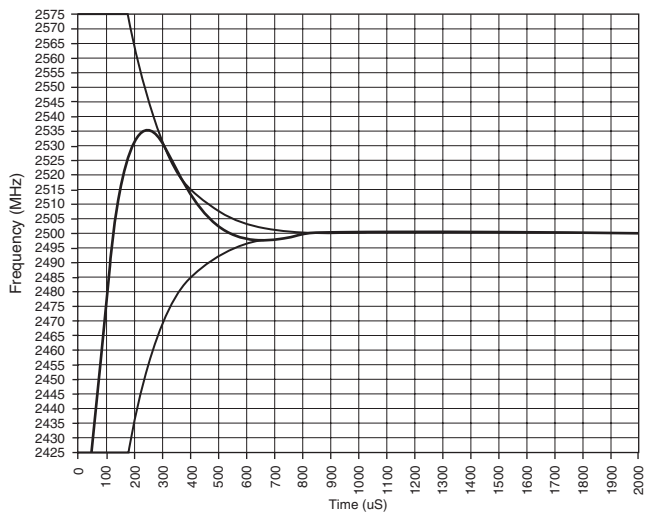




a.



b.



c.

Figure 5.12 An active loop-filter PLL design example: (*a*) loop gain/phase response; (*b*) phase noise; (*c*) PLL lock time.

Loop bandwidth = 2 kHz

Charge-pump gain = 1 mA

Phase margin = 45°

T3/T1 ratio = 100% (A high T3/T1 above 100% is commonly used in active filter designs for the best loop filtering of op-amp noise and PLL spurs.)

PLL IC = National LMX2326

VCO = RFMD VCO790-2300T (188.5-MHz/V, 68-pF input capacitance)

Solution:

1. Using the calculated values as presented, or by quickly employing the online version of National's EasyPLL, enter the above desired PLL specifications.
2. Run the software, which should results in the following loop-filter values:
   a. $C_1 = 22$ nF
   b. $C_2 = 220$ nF
   c. $C_3 = 1.0$ nF
   d. $R_2 = 820\ \Omega$
   e. $R_3 = 15\ \text{k}\Omega$
   f. $R_A = 1.8\ \text{k}\Omega$
   g. $R_B = 1.2\ \text{k}\Omega$
3. Analyze the design. The results should be:
   a. Phase noise

      PN@0.09 kHz = −65.96 dBc/Hz

      PN@10.00 kHz = −75.16 dBc/Hz
   b. Filter (Bode)

      Loop BW = 1.93 kHz

      Phase margin = 43.19°
   c. Lock time

      Lock time = 1832.4 μs (Which, in this example, is until the loop settles to within 500 Hz of the desired frequency of 2.5 GHz after being ordered to travel up from 2.4 GHz.)

Or, if a fractional-N passive loop PLL synthesizer is called for, such as when a tight channel spacing is required, use the Quick Example below.

A Quick Example Design a Fractional-N Passive Loop PLL Synthesizer (Fig. 5.13)

Goal: Using the enclosed *ADIsimPLL* software, create a fractional-*N* phase-locked loop synthesizer with an RC loop filter. The specifications and parameters for the circuit are:

f_r = 2.4 to 2.5 GHz (center frequency = 2.45 GHz)

$V_{CC} = 5$ V

$f_{REF} = 10$ MHz

$f_{COM} = 100$ kHz

Channel spacing = 100 kHz

Filter = third order

Loop bandwidth = 2 kHz

Charge-pump gain = 1 mA (938 uA)

Phase margin = 45°

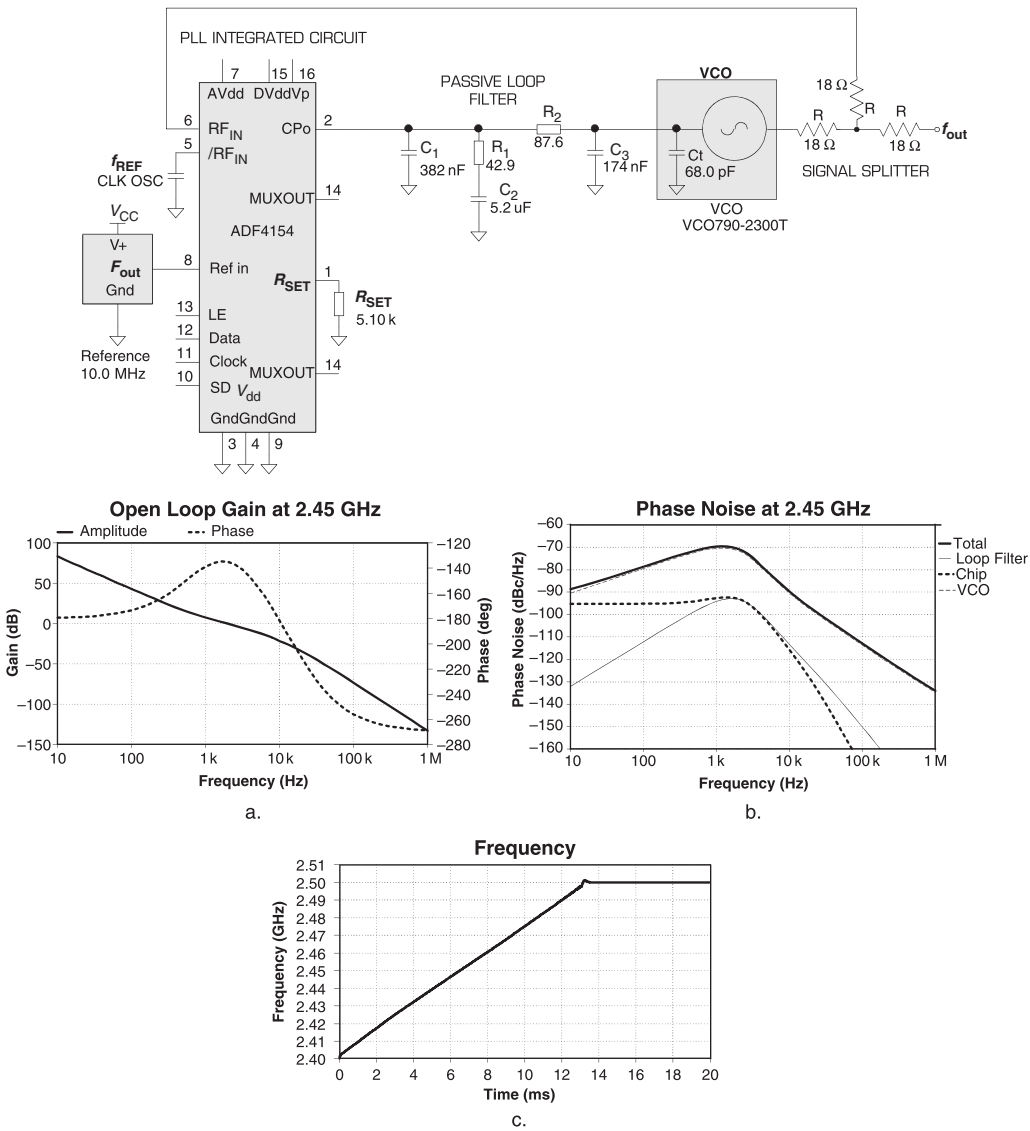


FIGURE 5.13 A passive loop-filter fractional-N PLL design example (ADIsimPLL): (*a*) loop gain/phase response; (*b*) phase noise; (*c*) PLL lock time.

PLL IC = Analog Device's ADF4154

VCO = RFMD VCO790-2300T (188.5-MHz/V, 68-pF input capacitance, PN = –90 dBc/Hz @ 10 kHz, PN = –112 dBc/Hz @100 kHz)

Solution:

1. Using the calculated values as presented in this book, or by quickly employing the enclosed *ADIsimPLL*, enter the above desired PLL specifications.

2. Run the software, which should result in the following loop-filter values:
   a. C_1 = 382 nF
   b. C_2 = 5.2 uF
   c. C_3 = 174 nF
   d. R_1 = 42.9 Ω
   e. R_2 = 87.6 Ω
3. Analyze the design. The results should be:
   a. Phase noise (total)

      PN@100 Hz = –78.66 dBc/Hz

      PN@10.00 kHz = –89.1 dBc/Hz
   b. Filter (Bode)

      Loop BW = 2.00 kHz

      Phase margin = 45.00°
   c. Lock time

      Lock time = 14.1 mS (Which, in this example, is until the loop settles to within 1000 Hz of the desired frequency of 2.5 GHz after being ordered to travel up from 2.4 GHz.)

NOTE: *ADIsimPLL ©2000-2007 by Applied Radio Labs. For latest version go to www.analog.com/adisimpll*

5.1.7 PLL Problems and Solutions

After the design and construction phases are completed, the most common issues found during the testing of a PLL are: noisy output, incorrect output frequency, spurious outputs, and an intermittent or continuous refusal to lock.

A PLL frequency synthesizer with a noisy output can be caused by multiple problems, since in a well-designed PLL circuit the highest contributor to phase noise (Fig. 5.14) will be the phase-locked loop's own integral phase-frequency detector. But this internal self-generated noise can be easily swamped out by the contributions of a noisy VCO, a noisy or noncrystal frequency reference source, low charge-pump current or voltage, incorrect

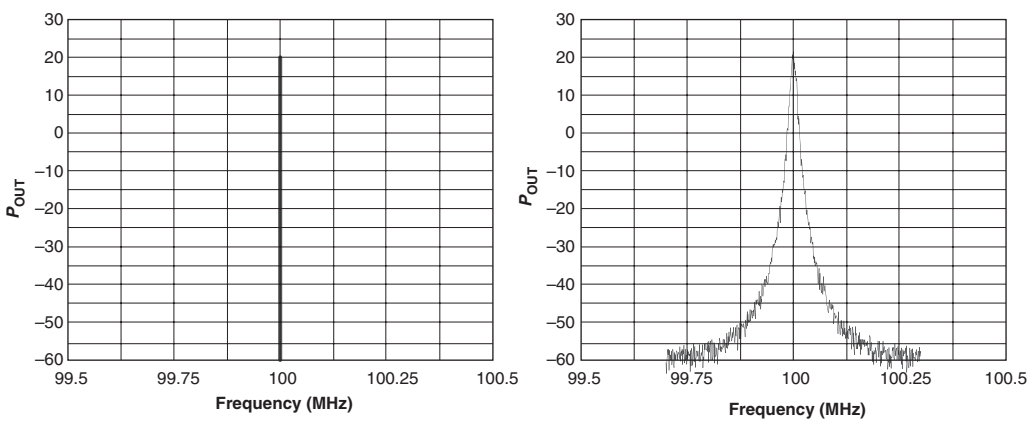


FIGURE 5.14 A PLL output without and with phase noise.

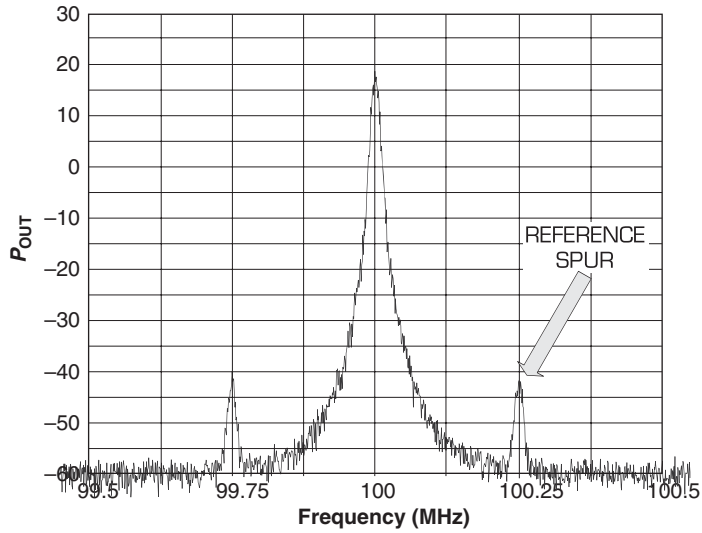


FIGURE 5.15 A PLL output with reference spurs.

signal amplitude levels into the R- or N-dividers, and a loop-filter bandwidth that is not wide enough to prevent the VCO from adding its own surplus noise.

NOTE: *VCO noise is small within the loop itself, but outside the loop the VCO noise can become quite high.*

Reference spurs will also be encountered, and are spurious signals at frequencies that are located at an interval equal to the comparison frequency (f_{COM}) away from the carrier frequency (f_{out}; Fig. 5.15). These spurs may, as well, occur at harmonics of the reference frequency, and can be created by charge-pump leakage and mismatch, PCB crosstalk, improper decoupling of DC power into the PLL, and exterior noise and signal source ingress.

A complete synthesizer, such as the design for a wideband transmitter or receiver in Fig. 5.16, will require each of its stages and circuits to be properly tested to assure reliable operation. (Not only is the PLL chip, reference oscillator, loop filter, and VCO shown, but so is the VCO gain amplifier, which is used to increase the tuning voltage for the wideband VCO; the 6-dB pad to supply a reasonable 50-Ω load to the VCO's output; a low S_{12} buffer isolation amplifier to furnish gain for a higher output amplitude; a VCO harmonic filter to clean up the synthesizer's output; and another attenuator pad to assure a decent 50-Ω termination for the harmonic filter and the next stage.)

To Test a PLL Synthesizer

A. *General VCO test:* For tests to confirm VCO will operate within the PLL as designed, check for proper:
   1. V_{CNTRL} voltage range (with ample margin), in volts.
   2. RF power output, in dBm or watts.
   3. Tuning linearity, in percent (or as a *ratio*).
   4. Tuning sensitivity, in MHz/V.
   5. Harmonic output levels, in dBc.

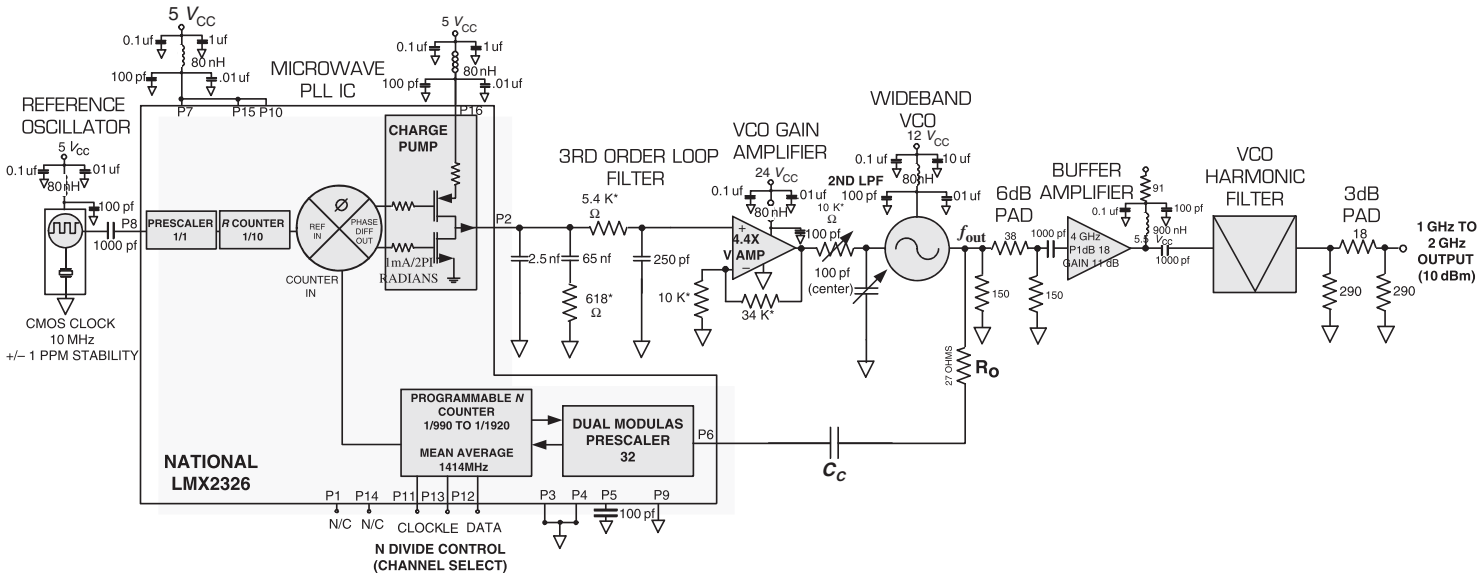


FIGURE 5.16 A complete microwave PLL for wideband tuning applications.

6. Current draw, in mA.
7. Stability over temperature, load impedance, V_{CC}, and V_{CNTRL} variations.

B. *General PLL locking tests:* For PLL locking reliability tests, confirm:

1. The reference oscillator is functioning, and at the proper power level and frequency.
2. R_o (the series resistor from f_{out} to the PLL chip) is not at too high a value back into the N-divider, and that C_C (the series coupling capacitor) still has a low reactance.
3. The data input is correct into the PLL chip, such as the load enable, the binary serial input data, and the data clock input.
4. Poor physical PCB layout is not causing reference spurs and added noise outputs.
5. The VCO will never be presented with a DC tuning voltage of 0 V (due to varactor nonlinearities), and establish that the VCO is operating within its designed tuning range, with adequate headroom due to the inevitable locking overshoots of approximately 15 to 20% above and below the required f_{min} and f_{max} frequencies.

For locking tests for A PLL with no lock, confirm:

1. All proper valid programming bits and clock signals are feeding the PLL chip, and with the proper timing.
2. The VCO is actually oscillating, and at the expected amplitude and frequency when free running.
3. There is sufficient RF signal strength at the N-divider input port from the VCO output.
4. The DC supply voltages and current levels are satisfactory in both amplitude and in waveform.

For locking tests for A PLL that is slow to lock, confirm:

1. The charge-pump current gain is not set too low.
2. The loop filter is as designed and not too narrow.
3. The loop capacitors are neither leaky, nor is there excessive flux, dirt, or condensation in or around the PLL loop-filter area that is draining off loop charge, nor are the loop capacitors prone to *dielectric soakage*.
4. The *FastLock* bit is set and is functional.

C. *General PLL spurious and phase noise tests:* For excessive levels of PLL spurious and phase noise, confirm:

1. The input reference frequency source is absolutely clean, since any undesired frequency components will be increased by the PLL loop by 20 log (N/R) within the loop's bandwidth.
2. The PLL and VCO power supplies are properly decoupled.
3. Leaky loop-filter capacitors are not used, since they will create high amplitude reference spurs caused by the necessity of the charge pump to be continually pumping current into the leaky loop.

4. The PCB layout follows proper EMI rules.
5. The VCO is effectively shielded, since any strong nearby RF energy that couples into the oscillator's resonant circuit can cause significant output spurs.

5.1.8 PLL Fractional-*N* Synthesizers

Due to their much faster lock time, very small frequency step size, wider loop bandwidth, and reduction of *in-band* spurious output levels, *delta-sigma fractional-N synthesizers* (Fig. 5.17) are an improvement over the older integer-*N* types PLLs. Fractional-*N* synthesizers can also be used with very high f_{COM} frequencies, thus giving us significantly lower divide-by-*N* numbers than integer-*N* PLLs, and therefore lower phase noise.

Fractional *N*s are referred to as "fractional" because they are not limited by their reference frequency for their minimum channel resolution, but can have a step size that is but a mere fraction of the reference frequency, or:

$$f_{vco} = (N + k/M) f_{REF}$$

where f_{vco} = frequency at output of VCO
N = integer divide-by-*N* values
f_{REF} = reference frequency at the input of the PFD
M = measure of fractionality
k = an integer between 0 and M

There are certain drawbacks to the use of fractional-*N* synthesizers. A particular noise that is intrinsic to delta-sigma fractional-*N* synthesizers is *quantization noise*, which can increase the noise level at the VCO's output unless the PFD comparison frequency

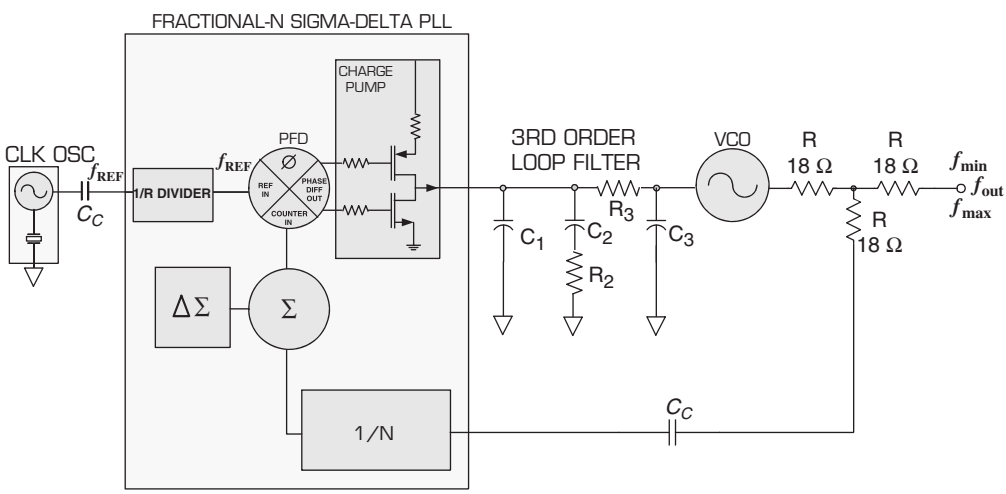


Figure 5.17 A common type of fractional-*N* PLL integrated circuit.

is quite high (> 10 MHz). Further, due to the approximately 7 to 10 dB of internal noise generated by the special circuitry within a fractional-*N* chip, the theoretical and expected improvement in phase noise over an integer-*N* synthesizer will not be fully realized. In addition, in order to even begin to realize their lower phase-noise specifications, many delta-sigma synthesizers must consume more current than standard integer-*N* types. Fractional-*N* synthesizers also have more possible frequency spur outputs, such as *fractional spurs*, *reference spurs*, and *integer-N boundary spurs*. (Fractional spurs are the most problematic of the fractional-*N* spurs, and show up close to the VCO's frequency output (f_{out}) at full and half channel offsets from the carrier. Reference spurs are, however, typically well filtered by the loop due to fractional *N*s, normally high PFD comparison frequency. The integer-*N* boundary spurs are only problematic when the synthesizer is programmed to output a frequency that is close to harmonic multiples of the PFD's f_{COM}, and are located in the VCO's frequency output, f_{out}, at carrier offsets equal to this f_{COM} frequency. There may also be other, smaller amplitude spurs present in the output of a fractional-*N* PLL. However, when PLL spur specifications are given for fractional *N*s, it is the fractional spurs that are normally being referred to.)

Fractional-*N* chips mainly differ from integer *N*s only in that the internal frequency divider can produce a fractional number, as opposed to just an integer. This fractional-*N* function is performed within the chip's divider circuitry, with the divider process carried out through averaging, which is occurring at the same time as the divider itself is switching via a fractional accumulator. Therefore, the complete PLL system design of a delta-sigma synthesizer itself is quite similar to that of any typical integer-*N* PLL. Because of the high-frequency noise created within the delta-sigma circuitry, the only obvious design difference between the two PLLs is that third-order filters or higher are commonly used for the loop filter, instead of the more basic second-order types popular in integer-*N* synthesizers.

National Semiconductor's LMX2470 is a very good example of a state of the art, high-performance delta-sigma PLL chip. The LMX2470 can function all the way from 500 up to 2600 MHz, possesses very low in-band phase noise and spurs, operates with a reference frequency of up to 110 MHz, has a maximum f_{COM} frequency of 30 MHz, consumes a very low current of 4 mA, requires a low nominal supply voltage of 2.5 V, includes an internal reference frequency doubler, and uses a straightforward 3-wire programming interface.

5.2 Direct Digital Synthesis

5.2.1 Introduction

Direct digital synthesis (DDS) is a relatively new technique, and decreases the cost and complexity of any frequency synthesizer that may need to operate with a very tight frequency resolution. DDS synthesizers can function down to 1 Hz (or even less), thus giving them the feel of LC analog fine tuning.

DDS synthesizers, as with common PLLs, are supplied in small surface mount integrated circuits, and are implemented along with one or more support ICs and multiple passive support components.

One method for generating DDS frequencies is shown in Fig. 5.18, and works by using a crystal oscillator reference as the *clock*, then calculating, in the *phase accumulator*, the discrepancy between this reference frequency and the actual frequency the wireless

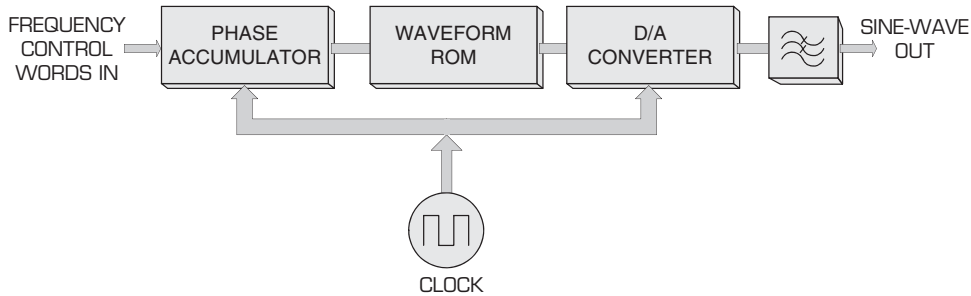


Figure 5.18 A DDS frequency synthesizer circuit.

device would like to generate. The phase accumulator will then calculate the proper address to send to the *waveform ROM* (or *RAM*) sine wave look-up table for the exact frequency we want to generate, which then forwards the appropriate discrete digital representation of the desired signal into the *D/A converter* (DAC). The still rough, stepped waveform is sent into a lowpass filter (LPF) in order to remove any spurious products, outputting a high-quality, artificially constructed analog sine wave.

All of today's DDS synthesizers have very limited maximum frequency ranges. This forces many such designs to go with a hybrid DDS/PLL scheme to increase this range to higher levels. Also, *quantization noise* can be a significant problem, and will decrease the signal-to-noise level of the synthesizer's f_{out} output. Quantization noise is a digitally produced noise, and refers to the errors made when the DDS signal is converted from digital to analog, since the analog synthesized frequency, with its infinite number of potential amplitudes, is really only being fabricated from a large, but finite, number of discrete digital levels. This makes absolute accuracy during conversion to analog waveforms impossible, resulting in a noise output riding along with the synthesizer's desired output CW signal.

Nonetheless, one method to accomplish a practical high frequency DDS/PLL hybrid frequency synthesis system is as shown in Fig. 5.19. When the radio's on-board microprocessor outputs control words into the DDS/PLL synthesizer for a change in

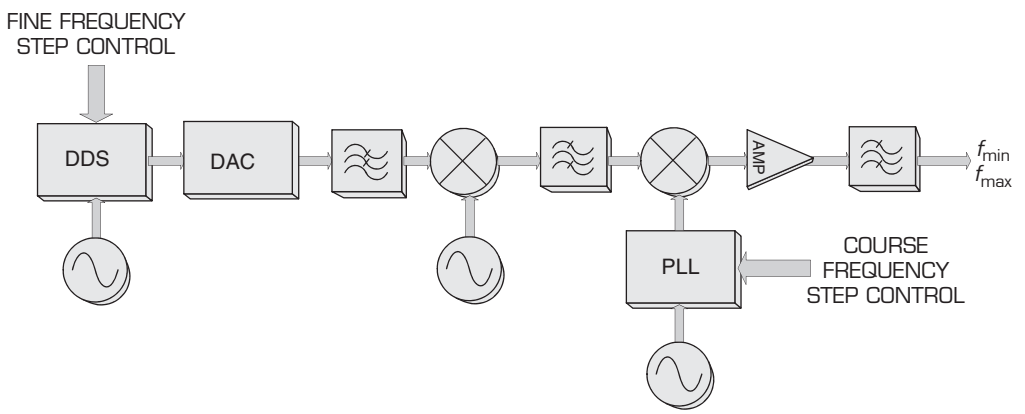


Figure 5.19 A method of synthesizing higher frequencies with direct digital synthesis.

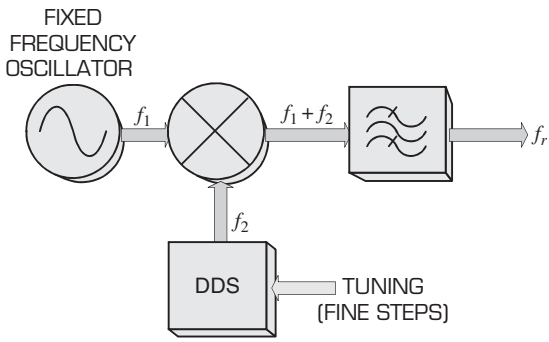


Figure 5.20 Premixing for DDS operation in the VHF region.

frequency, the actual DDS section will be able to reproduce a low, stable synthetic analog frequency (with the assistance of its DDS reference, the DAC, and the LPF), controlled by fine step words from the microprocessor. The LPF's output, a relatively clean but low-frequency sine wave, is then fed into a mixer for premixing. This Mixer1 sums the low frequency made by the DDS to the high-frequency reference of the LO in order to output an increased frequency through the BPF and into Mixer2. The PLL, with its own crystal reference (or one shared from a common clock), functions as Mixer2's LO, and is controlled by the coarse frequency digital control words from the microprocessor. The sum of the PLL and the output of Mixer2 is sent through the wide bandwidth BPF to obtain a very wide-ranging and adjustable output frequency, with decent resolution, fast lock time, and low spurious outputs.

Another method to increase the operating frequency of a pure DDS synthesizer, and allows operation of a DDS device at VHF and above, is standard premixing (Fig. 5.20). A low phase-noise fixed-frequency oscillator is inserted into one mixer port, while the other mixer port accepts the DDS synthesizer. The output of the mixer is filtered, resulting in a frequency synthesizer. This keeps DDS phase noise and spur generation to relatively low levels, while outputting a very high, but tunable, frequency selection. The premixing technique is limited only by the ability of the output bandpass filter to filter the frequencies of the LO feed through, the undesired difference frequency, and the mixer spurs. Nonetheless, DDS will usually have a high spurious output, as well as high DC current requirements, making DDS inappropriate for many RF applications.

This page intentionally left blank

CHAPTER 6
Filter Design

A filter is used to selectively pass or attenuate a particular band of frequencies, and can be constructed of *LC*, *RC*, *LCR*, *LR*, or distributed (microstrip) components, and can be either active or passive in nature. Active filters will contain some sort of amplifier combined with any of the above lumped passive components, while passive filters will simply employ lumped or distributed components, with ceramic and crystal filters found in many passband and stopband applications. Any of these filters will dramatically improve a filter's shape factor (steepness of its skirts), as well as provide a variety of bandwidths all the way from ultra narrowband to wideband. Surface acoustic wave (SAW) passive filters are also common in everyday RF applications, and are available with superb shape factors from narrowband to extremely wideband.

6.1 Filter Basics

6.1.1 Introduction

The RF spectrum contains quite a broad range of frequencies. So that we would not interfere with, or be interfered by, other communications channels, a method had to be found that would allow us to segregate a small chunk of this wide spectrum for transmission and reception. This can be accomplished with the use of untuned and tuned filters.

Due to an inductor's ability, as the frequency is raised, to increase its reactance at the same time as a capacitor decreases its reactance, a passive untuned LC filter circuit can easily function as a lowpass or highpass filter. This makes the untuned LC filter frequency selective. To act as a lowpass filter and attenuate higher frequencies (Fig. 6.1), an inductor will be arranged in series, blocking the high frequencies, while a capacitor is located in shunt, "shorting out" the higher frequencies. A highpass filter, which attenuates the lower frequencies (Fig. 6.2), has a capacitor that is in series, blocking the low frequencies, and a shunt inductor, shorting out the lower frequencies. These primitive filters may be cascaded to increase the sharpness of their skirts, as shown in Fig. 6.3. A lowpass filter can be larger or smaller than the eight-pole structure shown.

Constant-K refers to a filter that not only rejects or passes specific frequencies, but will also match impedances between the generator and its load throughout its entire operational passband. Two such filters are the *T-type* (Fig. 6.4) and the *PI-type* (Fig. 6.5) lowpass filters. The *m-derived* is another filter that displays a superior cutoff response, while maintaining a constant impedance across most of its passband.

A *bandpass filter* will only permit a certain range of frequencies to pass unattenuated between two points. Unlike the above lowpass filters, these are almost always of the

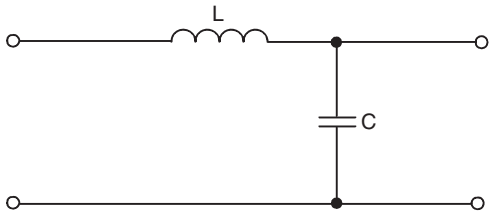


Figure 6.1 A basic LC lowpass filter.

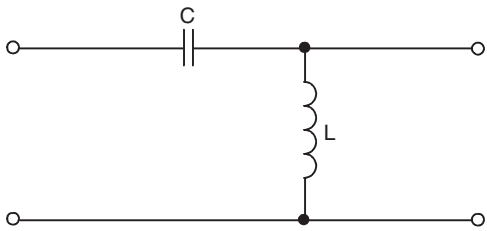


Figure 6.2 A basic LC highpass filter.

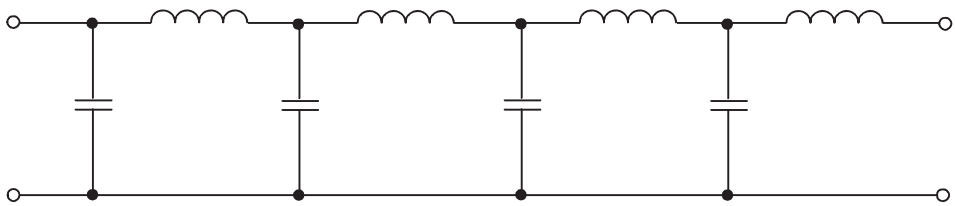


Figure 6.3 An eight-pole lowpass filter.

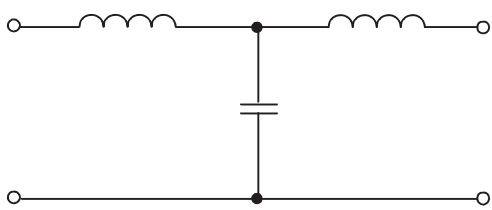


Figure 6.4 T-type constant-*K* lowpass filter.

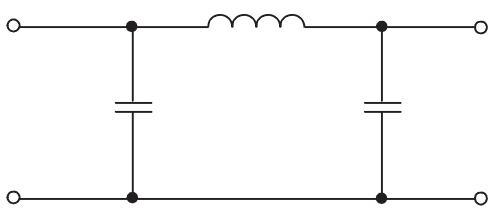


Figure 6.5 PI-type constant-*K* lowpass filter.

resonant, or tuned, form. However, by combining both a nonresonant lowpass with a nonresonant highpass filter, a bandpass attribute can be achieved. A *bandstop, notch,* or *band-reject* filter structure is the opposite of the bandpass type, and will severely attenuate a chosen set of frequencies between two points.

Tuned filters work on the principle of resonance. Since an inductor's reactance will increase with frequency, while a capacitor's reactance will decrease, and considering that these are opposing qualities, a specific frequency in which $X_L = X_C$ will soon be reached. At this point their opposite reactances will cancel. This will cause, when a capacitor and inductor are in *series*, the RF current through the circuit to be maximum, and the impedance to be at a minimum. Thus, the current will be able to reach very high levels, with the only impediment to this current being the small AC and DC resistance of the inductor. In fact, this resistance becomes very important in filter design, since the *unloaded Q* (or *component Q*) of a coil, which is X_L/r_e (the coil's reactance divided by its resistance), will severely impact the filter's insertion loss, the shape of its passband, and the return loss. The unloaded *Q* is influenced by the skin effect, the wire type, wire size, and losses within the coil forms. A very simple example of a basic series bandpass is shown in Fig. 6.6, while a bandstop is seen in Fig. 6.7.

The point that $X_L = X_C$ is called the resonant frequency (*fr*), and will vary depending on the values of the LC components. The series resonant frequency can be calculated by:

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

The formula demonstrates another aspect of resonant circuits: the higher the *L* and *C* values, the lower the resonant frequency will be; while the lower the *L* and *C* values, the higher will be the resonant frequency.

The series bandpass filter will also function as any other series LC circuit would when the input frequency is either above or below this exact point of resonance. Since the reactances will now no longer cancel the other, they will create a higher impedance and decrease the current flow.

Figure 6.6 A basic series bandpass filter.

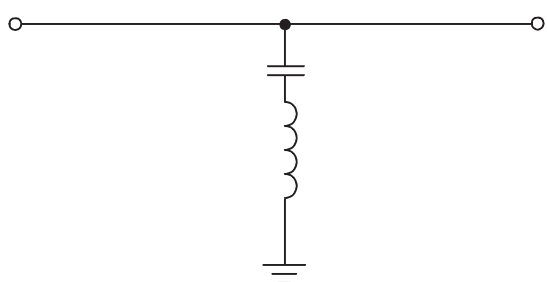


Figure 6.7 A basic series shunt bandstop filter.

A *parallel*, or resonant, tank circuit will have an impedance that is at a maximum, and current that is at a minimum, at some specific frequency. When an input frequency is inserted that is just above or just below this resonance point, the impedance of the tank will decrease while the current increases. This is the exact opposite effect as that found in the series resonant circuit above. However, this resonance point also occurs when $X_L = X_C$, and can also be calculated by:

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

The high impedance that is present only at the parallel resonant frequency will attenuate nonresonant frequencies, since only the small selected band of frequencies will be dropped across the high impedance of the tank.

Because of the finite unloaded Q of components, especially inductors, an inevitable amount of resistance is always present within a tank circuit. This will decrease the overall resonant tank impedance, which would be infinite with ideal components, and waste power, which must then be replaced. The small internal resistance that lowers the tank's loaded Q causes a modest current to flow into the LC resonant circuit. This waste of power can be minimized by using only high-Q inductors in the tank. But because the inductor and capacitor currents in a parallel circuit are 180° out of phase with the other, high circulating currents will always exist within the tank itself during resonance (Fig. 6.8). This circulating current is due to the two LC components exchanging current in a flywheeling manner. However, as these currents are completely out of phase, current flow *into* the tank is always at a minimum, and is dependent on the resistance within the tank caused by component Q limitations ($Q = (2\pi fL)/R$, with R not only being a pure resistance, but also the RF skin effect).

Another simple bandpass filter is shown in Fig. 6.9. Since it is an LC parallel filter in shunt with the output, all nonresonant frequencies will be sent to ground, while the bandwidth of interest will be passed on to the output due to the high impedance created by the tank at resonance. This creates a selective RF output voltage across the circuit, functioning as a BPF. The bandstop filter (BSF) of Fig. 6.10 has the parallel circuit in series with the output, and will pass all frequencies except the resonant frequency,

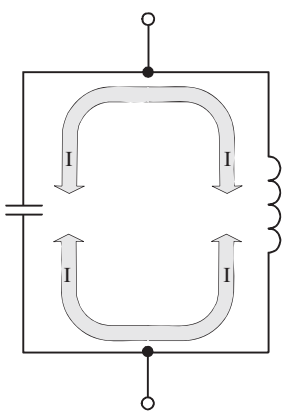


Figure 6.8 High internal circulating currents in a tank circuit at resonance.

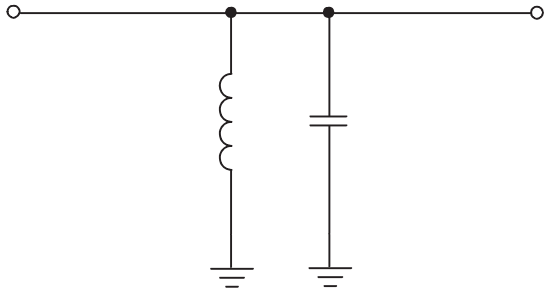


Figure 6.9 A basic parallel bandpass filter.

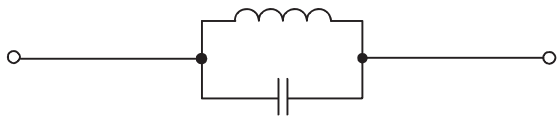


Figure 6.10 A basic parallel bandstop filter.

which is dropped across the high impedance of the tank. Since this will cause a decrease in the output amplitude at this single range of frequencies, it functions as a bandstop filter.

Most filters work by passing any frequencies within their passbands with little attenuation, while *reflecting*, but not absorbing, most of the undesired signals within its stopband back toward the source. These reflections can become a serious problem in some wireless systems design, as we shall soon see.

Filters must not only shape a signal, reject spurious frequencies, and select one frequency band out of many, but they must also maintain a specific input and output impedance through much of their passband that is identical to the system's impedance (usually 50 or 75 Ω).

Different types of filters, such as *LC*, *crystal*, *SAW*, *ceramic*, and *distributed*, will have various frequency bands in which they are most commonly employed due either to size, price, or performance. (Due to extreme cost, weight, and bulk, *waveguide filters* will not even be considered here.)

6.1.2 Filter Structures

LC filters can be utilized from 1 kHz all the way up to approximately 1.8 GHz. As the frequencies increase, however, so do the difficulties in implementation. This is because of the distributed parasitic inductances and capacitances of the components and the board layout, which all conspire to distort its expected passband and stopband responses. And because of these parasitics, we could also not possibly expect to reliably filter the 2nd harmonic, much less the 3rd harmonic, of a multi-GHz signal. Additionally, as the frequencies increase, the values of all the components decrease, causing severe part-to-part tolerance issues, and further distorted frequency responses in large production runs. Thus, at elevated frequencies (> 600 MHz), and with normal component tolerance levels, LC filters are not very repeatable in performance, especially the bandpass type. Further,

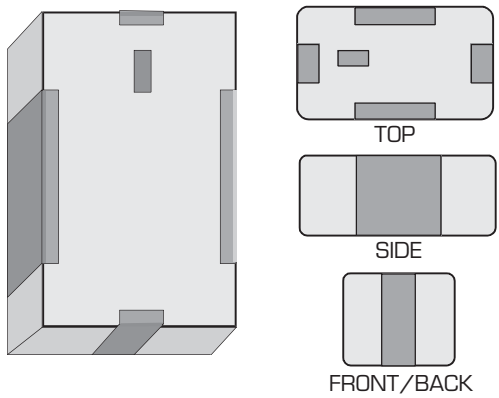


Figure 6.11 A common monolithic ceramic bandpass filter package.

LC filters are not useable for ultratight bandwidth requirements that demand steep skirts. Nonetheless, LC filters are very low in cost, can fit into a tight form factor, and are available for any frequency or band (under their maximum frequency).

The monolithic LC type *ceramic bandpass filter* (Fig. 6.11) is low in cost, has reasonable insertion losses (1 to 3 dB) and return losses (> 10 dB), permit medium RF input powers (0.5 to 1 W), are physically small, very easy to use, available up to 5.8 GHz, and are able to attenuate both the 2nd and the 3rd harmonic of a 2.4-GHz frequency signal. These filters are becoming the dominant RF bandpass filtering method for many low-power, high-frequency consumer applications in both microwave transmitters and receivers, and are constructed as multilayer planar structures with internal distributed L and C elements that are thinly deposited on a very high dielectric ceramic substrate. Ceramic filters come in both bandpass and lowpass configurations but, unlike the discrete LC filters discussed above, are only available in certain popular frequencies and bandwidths. Some ceramic bandpass filters may also have a rather weak asymmetric high-side frequency responses (Fig. 6.12).

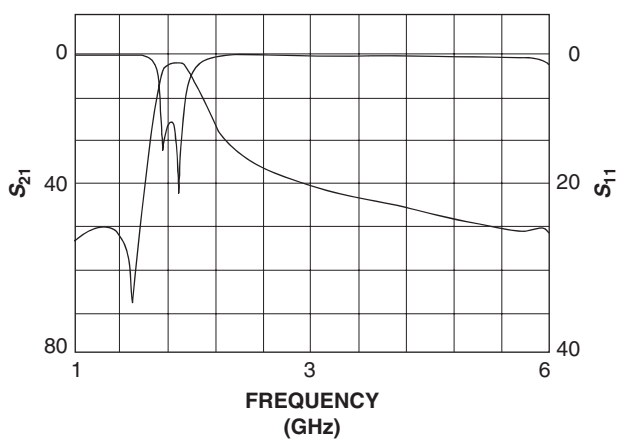


Figure 6.12 Frequency response of a typical monolithic ceramic filter.

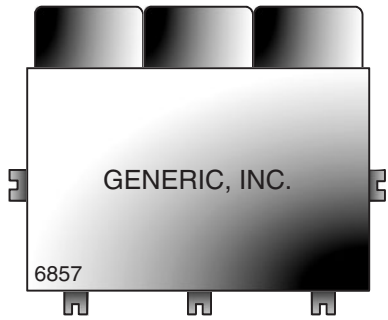


Figure 6.13 A common dielectric resonator filter package.

Higher performance ceramic *dielectric resonator* type filters (Fig. 6.13) use tuned quarter-wavelength ceramic cavity coaxial resonators that are capacitively coupled together. They offer high performance, multipoles (two to six poles), low insertion losses (approximately 1 dB per pole), high return losses (> 15 dB), narrowband operation (0.5 to 5%), and can operate with up to 1-W input powers. Ceramic dielectric resonator filters, however, can become quite physically large (and expensive), especially when combined with low frequencies and/or multiple poles. They may also have poor stopbands at or after the 2nd or 3rd harmonic, making them almost useless for applications in a power amplifier stage, or any other application that requires good stopband performance out to the harmonic frequencies. Ceramic dielectric resonator filters can be relatively sensitive to the port impedances they are presented, which must be close to 50 Ω (or passband rippling can occur along with a small-frequency offset). These filters can be readily custom made for most special applications, even in small runs.

Crystal filters are available in single monolithic packages, in multiple discrete crystal ladder and lattice topologies, and as active filter types. They are normally more expensive than ceramic filters, and are used mainly at IF (< 250 MHz) and audio frequencies. Crystal filters may have undesired spurious stopband modes that limit their attenuation of certain frequencies, and normally have a very restricted input power capability. However, most crystal filters will have excellent selectivity characteristics and an extremely high Q, along with a very wide to a very narrowband (0.001%) filtering capability (depending on the topology).

Surface acoustic wave (*SAW*) filters make use of a piezoelectric crystal substrate with deposited gold electrodes (Fig. 6.14). They are capable of replacing LC filters in certain wideband applications between 20 MHz to 2.4 GHz, and have a superb passband with an almost perfect brickwall response. Some SAWs even possess a respectable stopband rejection up to 6 GHz, and are available at medium cost, have very low ripple, and constant group delays. Nevertheless, SAWs enjoy only a limited number of off-the-shelf available resonant frequencies and bandwidths, will ordinarily experience a 3 to an extremely high 25-dB insertion loss (depending on many factors), may require input/output impedance matching components, can only survive a maximum RF input level of +15 dBm or less, and may have very poor spurious or harmonic suppression.

Distributed filters (Fig. 6.15) are comprised of copper traces etched on a dielectric substrate, such as a printed circuit board, that can act as a narrow or wideband filter structure from approximately 500 MHz up to 40 GHz. They are low in cost and can have

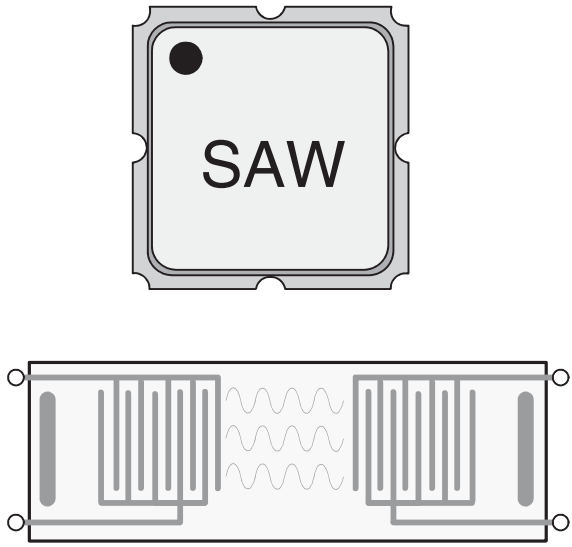


Figure 6.14 A popular SAW filter package and SAW internal structure (with surface acoustic waves).

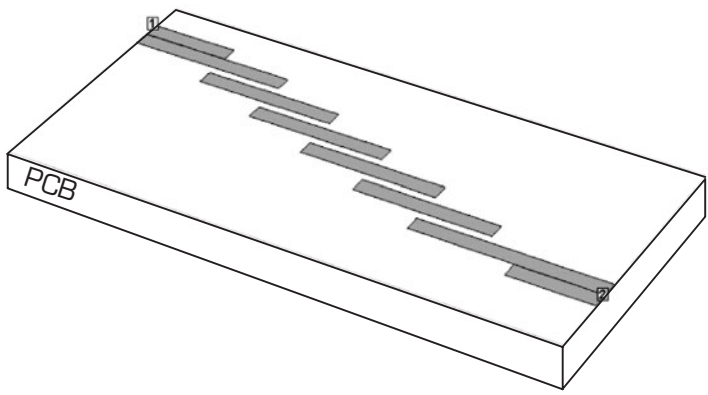


Figure 6.15 One type of distributed bandpass filter.

high *Q*s at microwave frequencies (when used with the appropriate high-quality substrate material). However, depending on frequency and design, distributed filters may take up significant board space (since they are physically larger than most other filter types), can have sizeable reentrance modes that mitigate much of their harmonic stopband attenuation, and will demand quality (i.e., expensive) PCB substrates for optimum performance. Another type of distributed filter, the *hybrid* type, is comprised of both distributed and lumped elements, with the *combline* topology (Fig. 6.16) being the most prevalent. This structure is popular due to its small size and much better reentrance characteristics than standard, nonhybrid distributed filters.

Active filters are reserved for very low frequencies, typically not much higher than baseband audio and data levels. They require a source of power, can be rather expensive

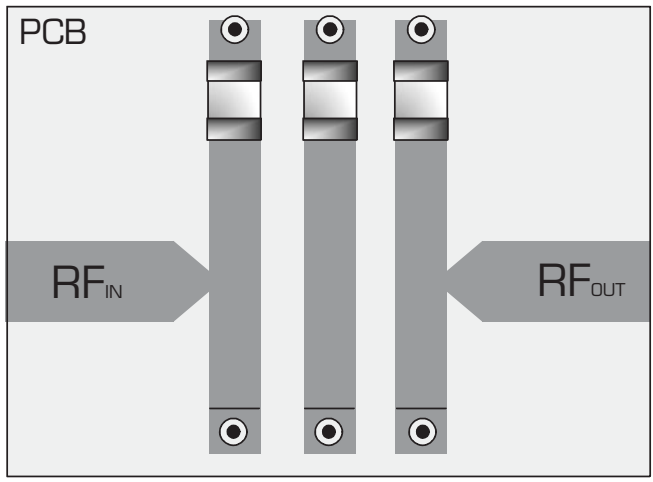


Figure 6.16 A hybrid filter structure.

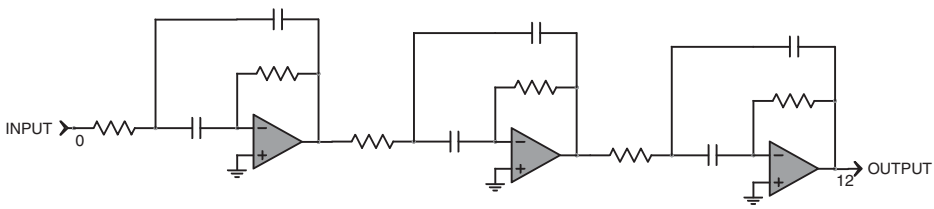


Figure 6.17 One of many common active filter network topologies.

and complex, and may take up a great deal of PCB real estate in order to obtain multipole operation (Fig. 6.17).

Mechanical filters, available as torsional, bar flexural, and disk flexure mode types in 2 to 12 poles, are expensive, physically large, and are only used for high-end applications that require a low center frequency of between 5 and 700 kHz. However, they are capable of extremely narrow bandwidths down to 0.02% (or up to 5%), and have tremendously high selectivity.

6.1.3 Filter Bandpass Responses

There are different bandpass filter responses for various requirements. For instance, the popular *Butterworth* is used when no amplitude ripple to the signal is desired within the filter's passband, has medium selectivity, medium group delay variations, and a good tolerance to component variations. (Any filter that is sensitive to component tolerances will exhibit an undesired and altered passband in S_{21} and S_{11}, due to the normal variations in *L* and *C* values.) *Chebyshev* filters will have a certain amount of passband ripple that will be forced on the input signal as it passes through to the filter's output. The Chebyshev response does, however, offer high selectivity, with high group delay variations being an unfortunate side effect of this innate superior filtering performance. (Amplitude ripple and high group delay variations can cause an increased BER in

digital signals, so are undesired.) Low ripple Chebyshevs can easily be designed, and the group delay variations can be improved by widening the filter's passband, or by using fewer poles. *Bessel* response filters will have no ripple in their rounded passbands and display very low group delay, but will have extremely poor selectivity and a inferior tolerance to component variations.

There are many different types of LC circuit filter topologies that will furnish these responses of Butterworth, Chebyshev, and Bessel. The choice depends on the shape of the desired passband, the percent bandwidth, the sensitivity to component tolerances (and parasitic distributed reactances), and the ability to obtain easily realizable component values during design.

6.2 Lumped Filter Design

6.2.1 Introduction

To start a successful filter design, we must first specify the filter's required response: lowpass, highpass, bandpass, or bandstop. Then, we need to select a particular transfer function, based on our knowledge of each filter type's particular characteristics. By far the two most common filters in RF design are the Chebyshev and the Butterworth. As stated, Chebyshev filters have amplitude and return loss ripple in their passbands, but they have an excellent amplitude rolloff of around 10 dB/octave/order, with the exact rolloff value dependant on the selected ripple design amplitude. These filter types should always be of an odd order for a predictable 50-Ω output impedance. Butterworth filters are flat in their passband response, do not possess any amplitude ripple, and have an amplitude rolloff of 6 dB/octave/order. The Bessel has a flat passband response with no amplitude ripple, but a poor amplitude rolloff of only 3 dB/octave/order. Another filter type, called an *elliptical*, has a very sharp rejection response characteristic, but sees limited use above about 500 MHz due to component variation sensitivity that can destroy its expected RF performance.

Next, we need to decide on the basic filter structure which, if we are designing our own filter, will either be microstrip distributed or lumped passive LC. If we are size constrained, and operating under a frequency of about 1 GHz, and also saddled with the typical, low cost consumer-grade FR-4 PCB, then we would normally want to select LC lumped filter types, and pass on the distributed structures. Or we may simply choose to work with an off-the-shelf multilayer ceramic type filter, for a very compact, quick solution. This can be a very attractive proposition at higher frequencies (> 1 GHz), or when development time is short.

We must have some idea as to what a specific filter type, topology, and number of poles is truly capable of. For instance, how high in frequency can the particular filter type reliably reach? How sensitive is it to component tolerances? How far do the stopbands dependably extend? What is the filter's ultimate attenuation? What will our insertion loss be within the passband? What kind of ripple will we see, if any? What percent bandwidth versus center frequency can we obtain? How steep will be our response?

6.2.2 Lumped Filter Circuit Types

Lumped filters are any filter structures that are constructed of individual, physical components, such as inductors, capacitors, or resistors, to form a network that can pass certain frequencies while blocking others. Such filters can be designed to perform in

lowpass, highpass, bandpass, and bandstop applications, and without any negative characteristics such as frequency *reentrance* (i.e., more than one passband; a problem associated with the majority of distributed filter types).

A filter works by displaying either a low or a high impedance to any undesired out-of-band frequencies, causing reflections, and thus impedance mismatch losses. This mismatched area of the filter's operation is called the stopband. The filter's passband insertion losses themselves are caused mainly by the finite unloaded-Q of the components, causing resistive (dissipative) losses within the circuit. Any passband mismatches between the filter and the adjoining stages will create losses as well.

To lessen costs, insertion losses, group delay variations, and physical space demands, filters should be designed with the minimum number of poles needed to attenuate the undesired frequencies. And rarely should we simply cascade separately designed filters to add more poles; if we do so indiscriminately, there may be undesired interactions, causing unpredictable filter responses.

There are certain topologies of lumped filters commonly adopted in wireless circuit design, and these are typically of the *all-pole* variety. As a good example, the high-end filter synthesis software program available as a module in Agilent's *Eagleware* can quickly and easily design any one of the following bandpass filters:

1. *Minimum capacitor all-pole filters* (Fig. 6.18) start with a series circuit instead of a shunt tank, and are excellent for wide bandwidth applications of greater than 30%. These filters should not be designed for narrower bandpass's due to the series inductors, which will become excessively large, and the shunt inductors, which will become excessively small. The series and shunt capacitors will react in the opposite manner. A third-order filter of this type will utilize three capacitors and three inductors.
2. *Minimum inductor all-pole filters* (Fig. 6.19) begin with a tuned tank at the input. For the exact same design limiting reasons as discussed above, these are excellent for wide bandwidth applications of greater than 30%. Because they do not require an excessive amount of different component values, they are one of the more popular filter topologies. Component values, however, can vary wildly within the same filter (e.g., some capacitors may have a value of 1270 pF, while another may have a value of 66 pF). A third-order filter of this type will employ three capacitors and three inductors.
3. *Top C-coupled all-pole filters* (Fig. 6.20) have better low-side frequency rejection than high-side, and are a good choice for applications that require bandwidths narrower than what the two above filters can produce, or under 30%. A third-order filter of this type requires up to seven capacitors and three inductors. Even though the part's count is rather high, an advantage is that all of the inductors are of an identical value and, with proper design, the two input and output capacitors will not actually be required. As well, the capacitors that are in shunt with the inductors may be close in value to each other, which can mean that only a single value of shunt capacitor is necessary.
4. *Top L-coupled all-pole filters* (Fig. 6.21) are similar to the above top C-coupled filters, except that in this case the roles of the inductors and the capacitors are reversed. This filter circuit will have better high-side rejection than low-side, and is a good choice for operation in applications that require bandwidths that are narrow, or less than 30%. A third-order filter requires up to seven inductors,

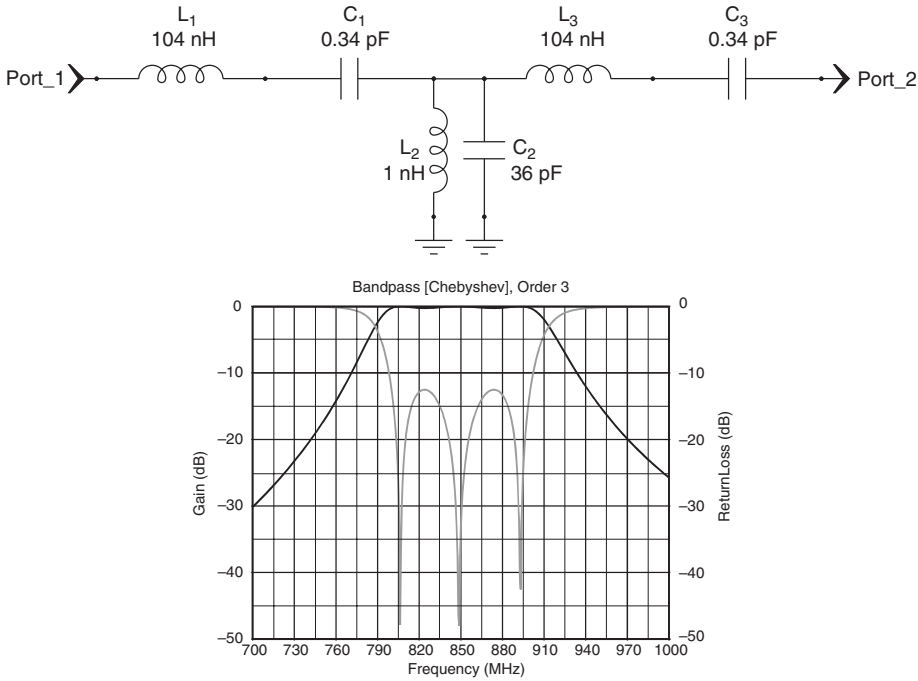


Figure 6.18 A third-order minimum capacitor bandpass filter.

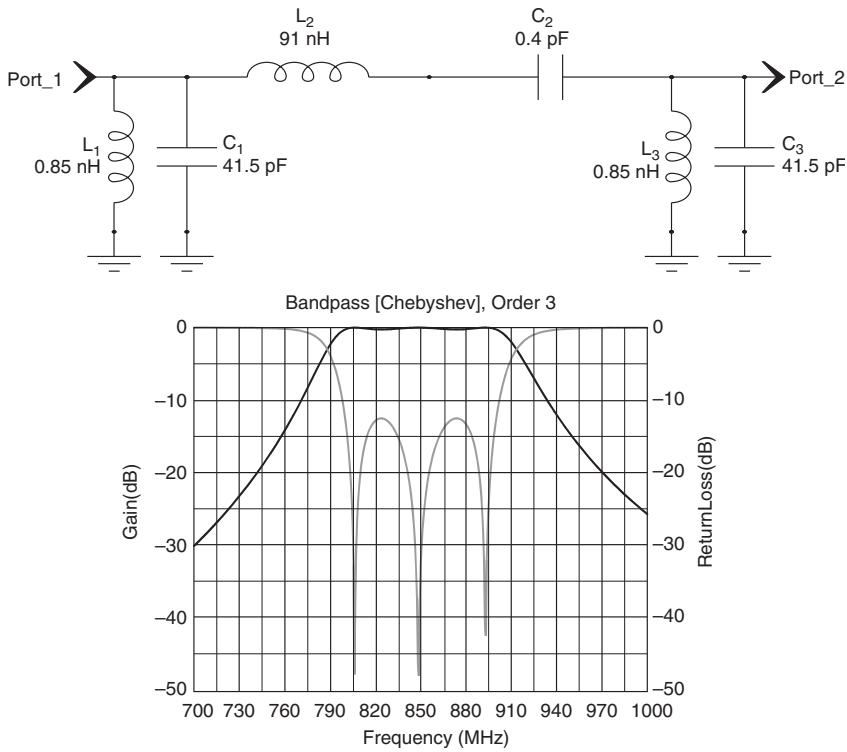


Figure 6.19 A third-order minimum inductor bandpass filter.

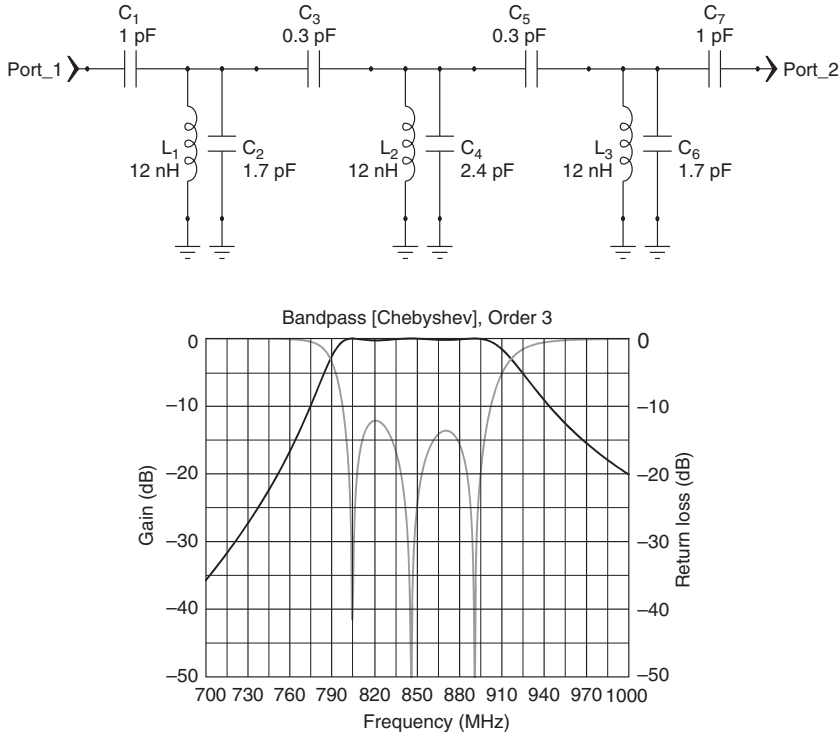


Figure 6.20 A third-order top C-coupled bandpass filter.

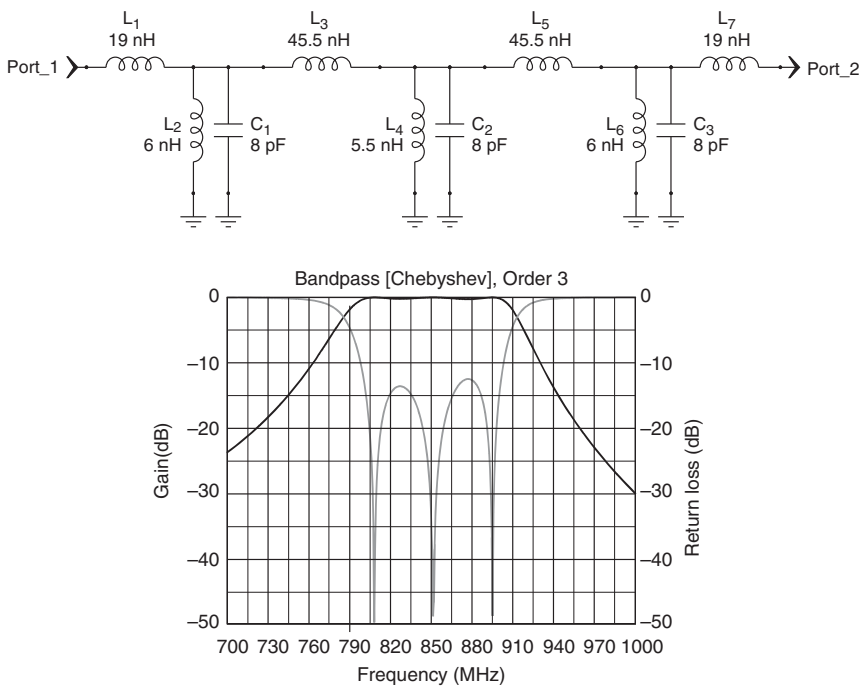


Figure 6.21 A third-order top L-coupled bandpass filter.

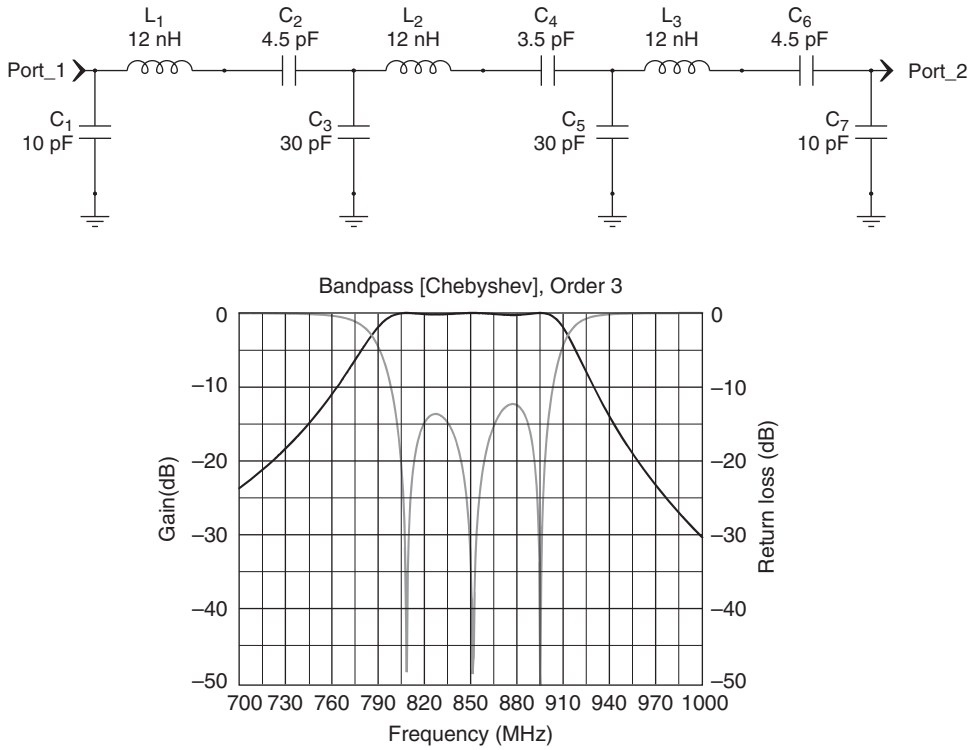


Figure 6.22 A third-order shunt C-coupled bandpass filter.

making it a rather expensive circuit, along with three capacitors. However, all of the capacitors will be of the exact same value, and the number of inductors can be decreased down to five with proper design.

5. *Shunt C-coupled all-pole filters* (Fig. 6.22) have better high-side frequency rejection than low-side, and operate at their best at bandwidths under 30%. A third-order filter of this type requires up to seven capacitors and three inductors, with all three inductors being of identical value, while the series capacitors will typically be very close in value to each other (with proper design only five capacitors will be needed).
6. *Tubular all-pole filters* (Fig. 6.23) have better high-side frequency rejection than low-side, and operate at their best at bandwidths of under 30%. This circuit requires up to eight capacitors and three inductors for a third-order filter. All three inductors are of the identical value, and with proper design only six capacitors will be required.
7. *Lowpass/highpass filters* (Fig. 6.24) are the last option when a very wideband structure is needed (> 60%), and only if all other filters are not realizable at extreme bandwidths. The lowpass/highpass filter design involves cascading a separate 50-Ω lowpass design with a 50-Ω highpass filter in order to form a full bandpass response.

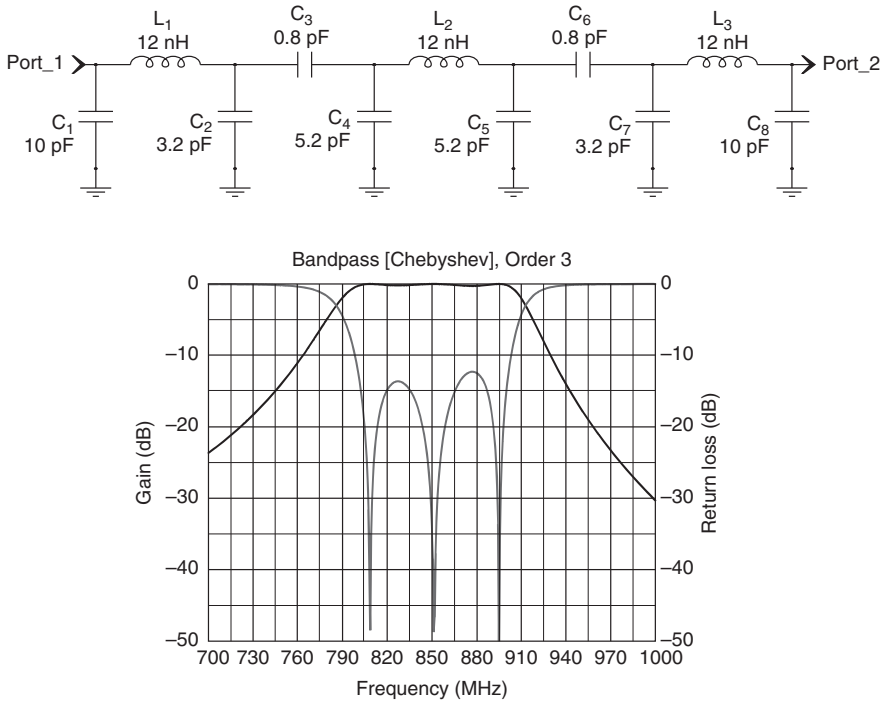


Figure 6.23 A third-order tubular bandpass filter.

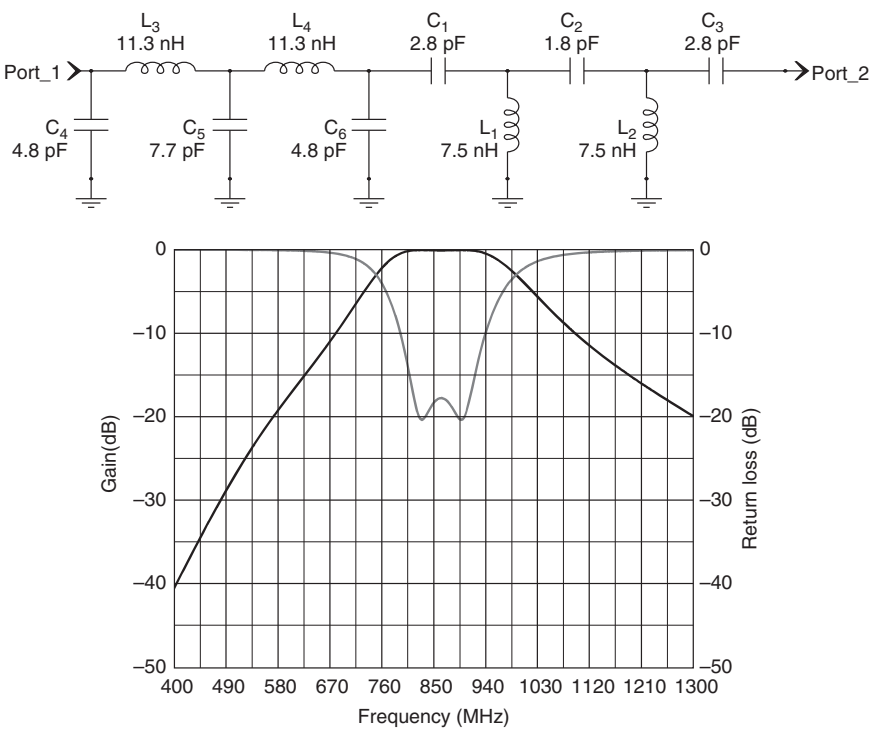


Figure 6.24 A lowpass/highpass bandpass filter.

6.2.3 Lumped Filter Image-Parameter Design

Since the proliferation of very low-cost filter design software for the average personal computer, there is really little need to hand calculate lumped filter values. However, for completeness, *image-parameter* design methods for highpass, lowpass, and bandpass filters will be presented in the popular Butterworth response characteristic, as demonstrated by Gottlieb and others.

Image-parameter design of Butterworth filters is the only design technique that readily allows filters to be calculated without computer software, or by the use of extensive and tedious look-up tables. However, sometimes impractical component values and unoptimized filter responses may result from this technique. It is therefore always a necessity that filters employed for any stringent wireless application be designed not with image-parameter methods, but with the far more complex, but more accurate, modern filter theory techniques. *Modern filter theory* calculations can realistically only be performed using a computer that is loaded with the proper design software, such as the included *AADE Filter Designer*, or with Agilent's lumped filter design module sold for use within Genesys. But for less demanding applications, or for filters with a low number of poles, image-parameter design will usually suffice.

It must always be kept in mind that whether a filter has been designed in a computer or by hand calculations, that distributed reactances at frequencies greater than 30 MHz (even when using surface mount components) will begin to noticeably decrease and/or distort the filter's passband. This is why it is absolutely necessary that after any such filter has been designed it must be transferred to an RF linear simulator (such as Qucs), and all ideal components replaced with the appropriate S-parameter or Modelithic passive models and PCB traces/pads, and then resimulated and tuned for optimal performance.

Lowpass Filters

Two basic lowpass two-pole filter types are shown; Fig. 6.25 has a series inductor at the input, and Fig. 6.26 utilizes a shunt capacitor at the input. These fundamental building blocks in lowpass filter design are also called the *half-sections*. This particular half-section is composed of one series inductor and one shunt capacitor. By cascading an increasing amount of half-sections, we can obtain a filter with any filter skirt steepness desired. This is referred to as increasing the number of poles of a filter; where four poles equal one section in a lowpass filter.

The half-sections may either be joined as PI sections (Fig. 6.27) or as T sections (Fig. 6.28). These sections will then combine with the next adjoining reactive element, creating a single inductor for the PI network of Fig. 6.29. This is because there would be absolutely no reason

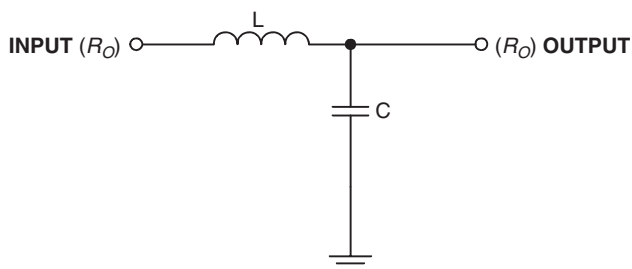


Figure 6.25 Series inductor LPF half-section.

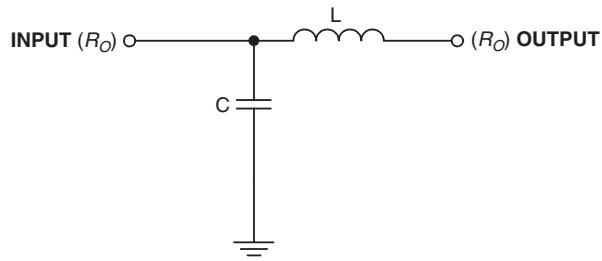


Figure 6.26 Shunt capacitor LPF half-section.

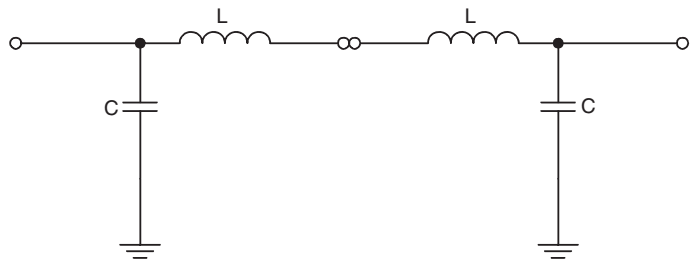


Figure 6.27 Half-sections joined as PI sections.

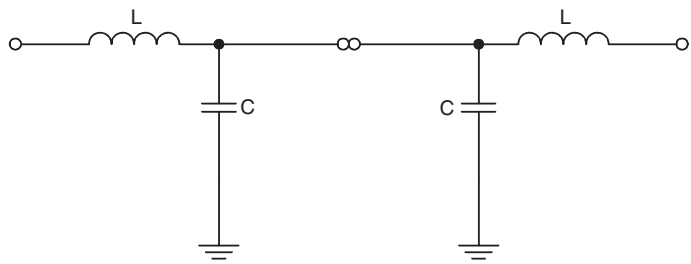


Figure 6.28 Half-sections joined as T sections.

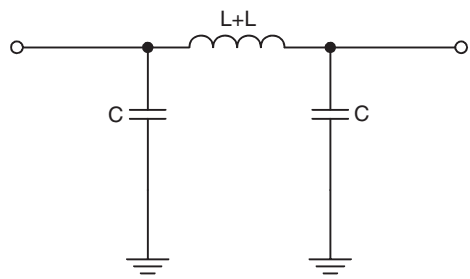


Figure 6.29 Joining two half-sections by adding series inductors.

to use two inductors in series with each other, or two capacitors in shunt with each other, when their values can simply be added to obtain the proper value (Fig. 6.30). Combining half-sections with either the PI or T technique is equally valid, but the design that results in the fewer inductors is usually preferred on a cost, loaded Q, and size standpoint.

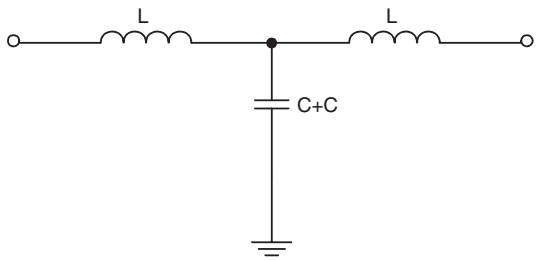


Figure 6.30 Joining two half-sections by adding shunt capacitors.

To design a lowpass filter, a half-section can be calculated by:

$$L = \frac{\left(\frac{R_O}{\pi f_C}\right)}{2} \qquad \text{and} \qquad C = \frac{\left(\frac{1}{\pi f_C R_O}\right)}{2}$$

where R_O = impedance at the filter's input and output, f_C = filter's 3-dB cutoff frequency.

After the L and C values are calculated, the actual number of poles can be increased from the present half-section's two, to any number desired. This is accomplished by combining half-sections as shown in Fig. 6.31, adding the capacitances; or by combining them as shown in Fig. 6.32, and adding the inductances.

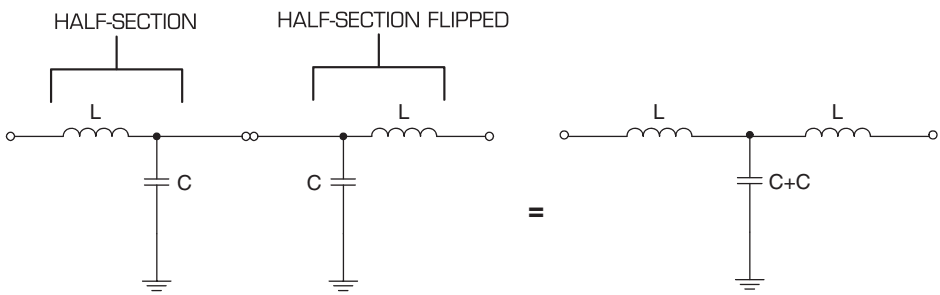


Figure 6.31 One way of properly combining half-sections.

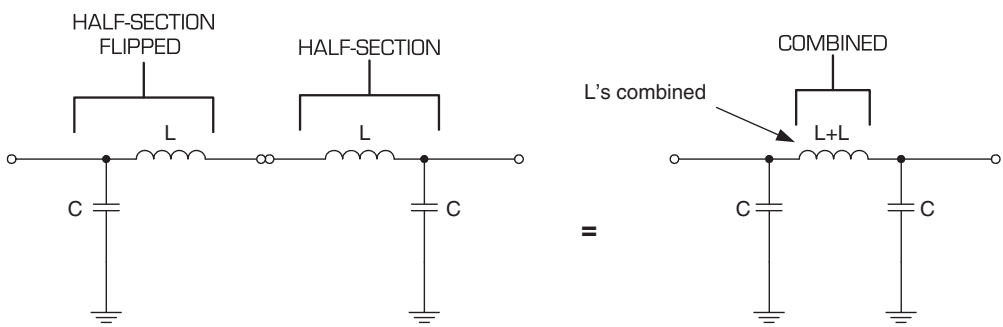


Figure 6.32 Another way of combining two half-sections.

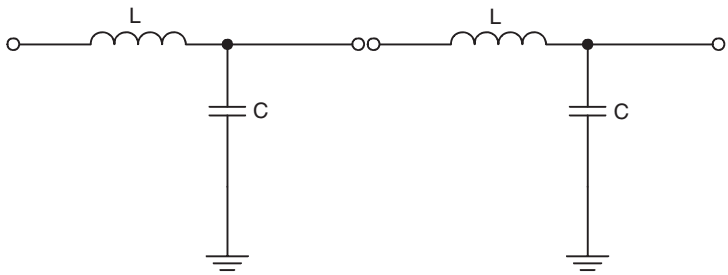


Figure 6.33 Incorrect way of combining two half-sections.

However, combining half-sections as shown in Fig. 6.33 would be incorrect, and would result in a filter with not only too many components, but one that will not function as designed. This is because only shunt elements are allowed to *blend* with shunt elements, and only series elements can *blend* with series elements.

As the number of filter sections increases, the actual cutoff frequency will begin to decrease. This is referred to as the *cascade effect*. The cutoff frequency will drop, especially when three or more sections are combined, if this is not taken into account. Table 6.1 is a list of the correction factors that must be used to obtain an accurate cutoff frequency when cascading multiple sections.

To utilize this table in lowpass filter design: If f_C of the lowpass filter is to be 200 MHz, and a two-section filter is required, then first multiply the adjustment factor of 1.15 *times* the f_C. This would be equal to 230 MHz. Now design the two-section lowpass filter as if the f_C will be 230 MHz, and a filter with a true f_C at the desired frequency of 200 MHz will now be the result.

If the lowpass filter is also required to block DC, then a large-value capacitor with an X_C of 1 Ω at its lowest frequency of interest can be inserted at the filter's input. This will have no effect to the lowpass filter's response. The ability to block DC is especially valuable at the lower RF frequencies, where ferromagnetic cores are used in the filter's inductors (the DC can effect the permeability of the inductor's cores by saturating them, and thus changing the inductance of the coils, destroying the response of the filter).

Number of Sections* (1/2 Section = 2 poles = 1 *L*, 1 *C*)	**ADJ Factor**
0.5	1.00
1.0	1.05
1.5	1.10
2.0	1.15
2.5	1.20
3.0	1.25

*Before combining of the half sections

Table 6.1 Correction Factors for Cascading Multiple Lowpass Filter Sections

The attenuation slope of the lowpass filter as presented can be approximated as an 18-dB increase in attenuation per octave for a three-pole filter. In other words, if the filter were designed to have an f_C of 1 MHz, then the lowpass filter would have attenuated a 2-MHz signal by 18 dB. By 4 MHz, the signal would be down by two times 18 dB, or 36 dB . And if the lowpass filter were increased to six poles, it would have an attenuation slope that dropped by 36 dB per octave, so with an f_C at 1 MHz, the signal is attenuated (has an insertion loss) of approximately 3 dB at 1 MHz, while at 2 MHz the signal is down to about 36 dB, and at 4 MHz it has dropped to 72 dB.

When multiple half-sections are combined with lowpass filter designs, the filter's attenuation response becomes that of the amount of the *L* and *C* components (poles) that result from the combination and joining of these half-sections. For instance, if a filter is created from three half-sections, which contain six reactive components (poles), the filter would now contain only four reactive components (poles) after the appropriate components had been combined. The components left *after* the combination of the half-sections will be the indicator of the filter's attenuation response.

Highpass Filters

Shown in Fig. 6.34 is a two-pole highpass filter that incorporates a series capacitor and a shunt inductor, and another that has a shunt inductor and a series capacitor. Both are half-sections. This is the exact opposite component layout as that of the lowpass filter style above.

For the highpass filter, as it was with the lowpass filter design, cascading these half-sections will produce a highpass filter with as many poles as is required for the particular application.

To design a Butterworth highpass filter, first calculate the *L* and *C* values of the single half-section of Fig. 6.34 by:

$$L = 2\left(\frac{R_O}{4\pi f_C}\right) \qquad \text{and} \qquad C = 2\left(\frac{1}{4\pi f_C R_O}\right)$$

where R_O = impedance at the filter's input and output, f_C = filter's 3-dB cutoff frequency.

Take the half-section as calculated, and combine by dividing by 2 the combined inductors in Fig. 6.35, or by dividing the combined capacitors by 2 in Fig. 6.36. Unlike lowpass filter design in which the combined components are added, a highpass filter's combined components must be divided by 2. As for the similar situation of the lowpass filter above, the highpass sections must *not* be combined as in Fig. 6.37.

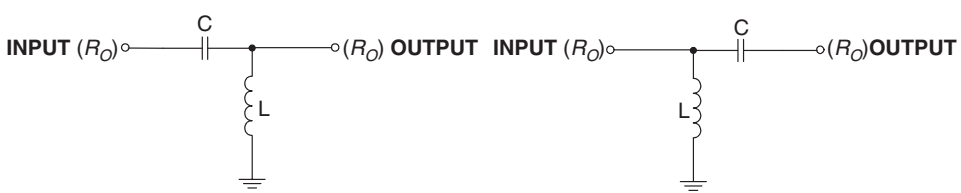


Figure 6.34 Series capacitor and shunt inductor HPF's, or half-sections.

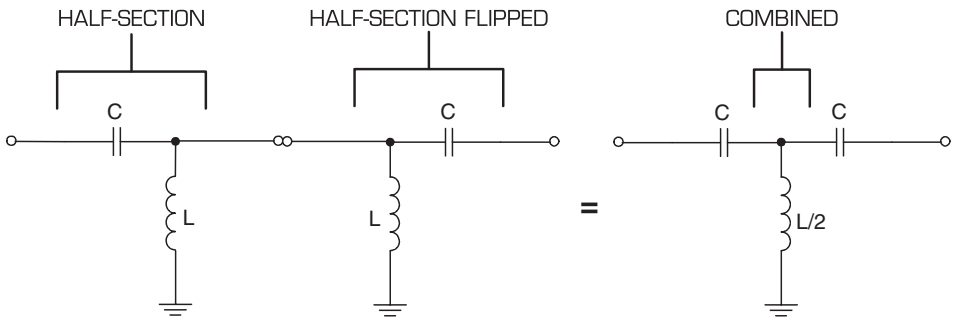


Figure 6.35 One way of combining HPF half-sections.

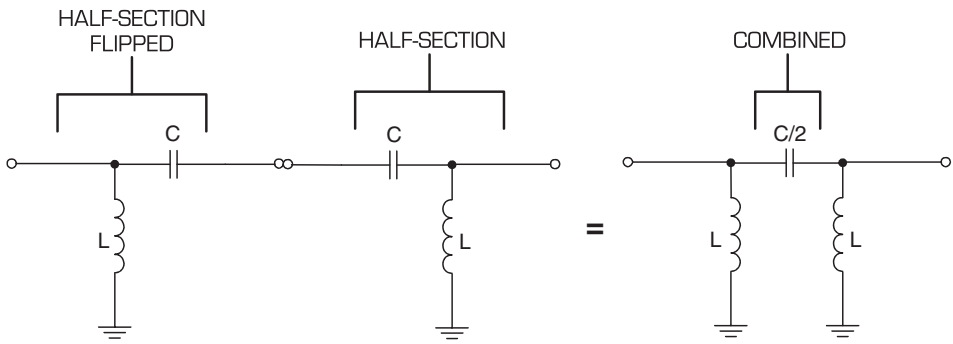


Figure 6.36 Another way of combining HPF half-sections.

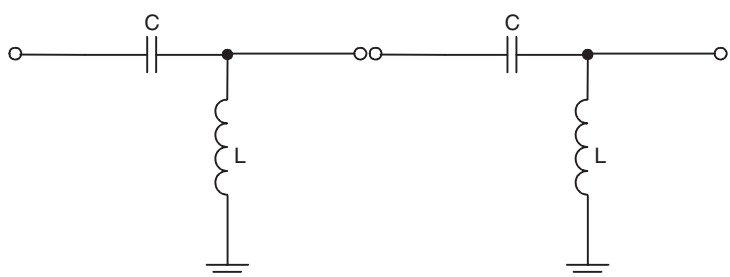


Figure 6.37 Incorrect way of combining HPF half-sections.

As with the connecting of multiple half-sections of the lowpass filter, the highpass filter will also exhibit a cascading effect. This is due to the increasing selectivity of the cascaded half-sections, and forces the f_C to *increase* in frequency above what is desired. By using the adjustment factors presented in Table 6.2, this proclivity can be compensated for.

For example, if the f_C of a highpass filter is to be 200 MHz, and a two-section filter is required, then first multiply the adjustment factor, as found in Table 6.2, of 0.85 times the f_C of 200 MHz. This equals 170 MHz. Now design the two-section filter as if the f_C will be 170 MHz, and a highpass filter with a true f_C at the desired frequency of 200 MHz will now result.

Number of Sections (1/2 Section = 2 poles = 1 *L*, 1 *C*)	ADJ Factor
0.5	1.00
1.0	0.95
1.5	0.90
2.0	0.85
2.5	0.80
3.0	0.75

Table 6.2 Correction Factors for Cascading Multiple Highpass Filter Sections

As it was with the lowpass filter design, the highpass filter attenuation slope follows the same 18 dB per octave drop for a three-pole filter, and a 36 dB per octave drop for a six-pole filter.

When multiple half-sections are combined with highpass filter designs, the filter's attenuation response becomes that of the amount of the *L* and *C* components (poles) that result from the combination of these half-sections. For instance, if a filter is created from three half-sections, which contain six reactive components (poles), it will now contain only four reactive components (poles) after the appropriate components have been combined. These components left *after* the combination of the half-sections will be the indicator of the filter's attenuation response.

An important consideration with all filters, and especially with highpass filter designs, is that such a filter is only a highpass filter up to a certain frequency, after which the stopband becomes disrupted by the eventual self-resonance of the filter's components, as well as the distributed capacitances that will inevitably let higher frequencies pass through the stopband (Fig. 6.38).

If it is required to block DC from being shunted to ground at the filter's input, a highpass filter must be designed with a series capacitor, instead of a shunt inductor.

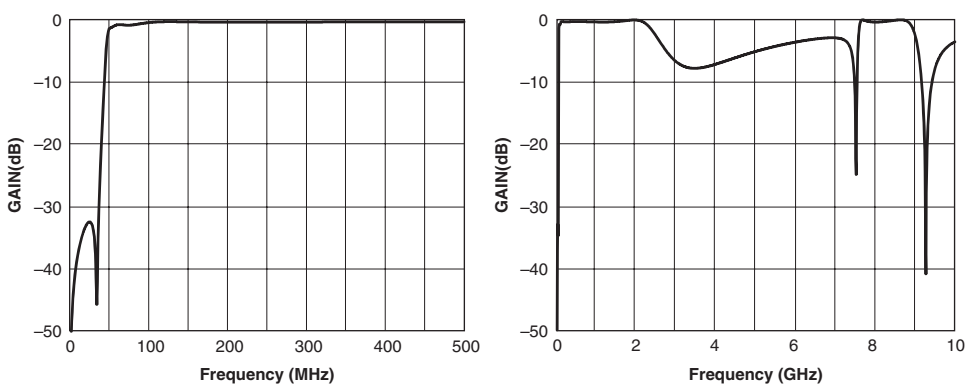


Figure 6.38 A HPF using real components, with its frequency response up to 500 MHz (left) and up to 10 GHz (right), showing the degradation of its passband at high frequencies.

Bandpass Filters

The design of a bandpass filter by image parameters is quite similar to the lowpass and highpass filter design procedure just presented, but the complexity is greater due to twice the components and twice the cutoff frequencies. As with the lowpass and highpass filter designs, we also start with a half-section (Fig. 6.39), and can add either of these half-sections together to obtain a filter with more poles.

With a bandpass filter, each LC pair is a *single* pole, so a half-section is comprised of *two* inductors and *two* capacitors, or two poles. Again, as with the above filters, only series arms of each half-section are combined with series arms (or parallel arms to parallel arms), for each half-section (Fig. 6.40); and *not* with series joined to parallel, or parallel joined to series (Fig. 6.41).

To design a bandpass filter with image-parameters, first calculate the element values for the first half-section of Fig. 6.42:

$$L_S = \frac{\left(\frac{R_O}{(f_{2C} - f_{1C})\pi}\right)}{2} \quad \text{and} \quad C_S = 2\left(\frac{f_{2C} - f_{1C}}{R_O(f_{2C} \cdot f_{1C})4\pi}\right)$$

$$L_P = 2\left(\frac{R_O(f_{2C} - f_{1C})}{(f_{2C} \cdot f_{1C})4\pi}\right) \quad \text{and} \quad C_P = \frac{\left(\frac{1}{R_O(f_{2C} - f_{1C})\pi}\right)}{2}$$

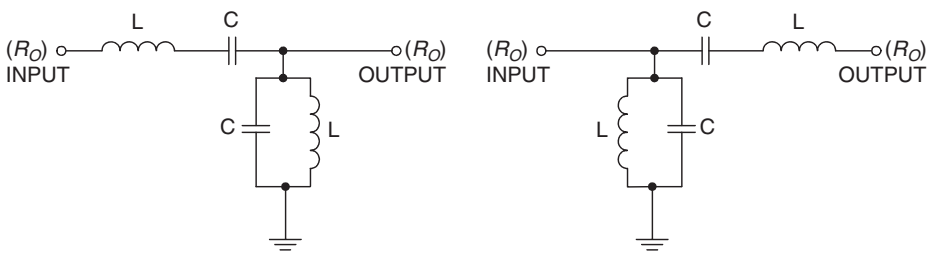


Figure 6.39 Series and tank BPF half-sections.

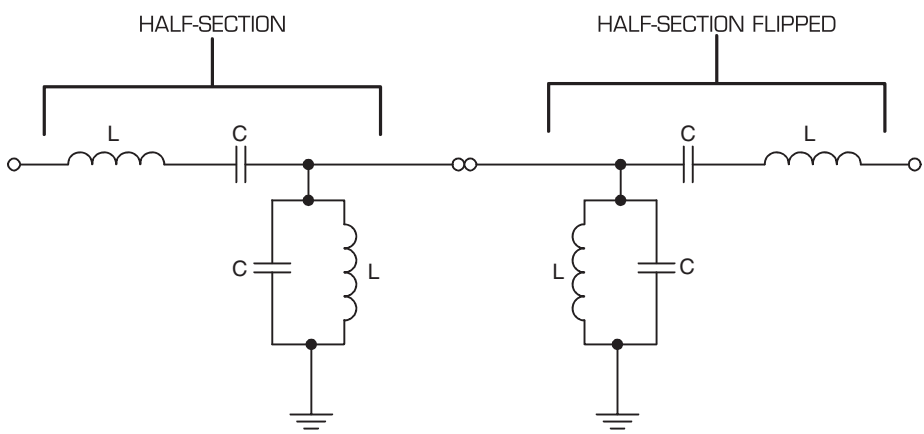


Figure 6.40 Proper way to join BPF half-sections.

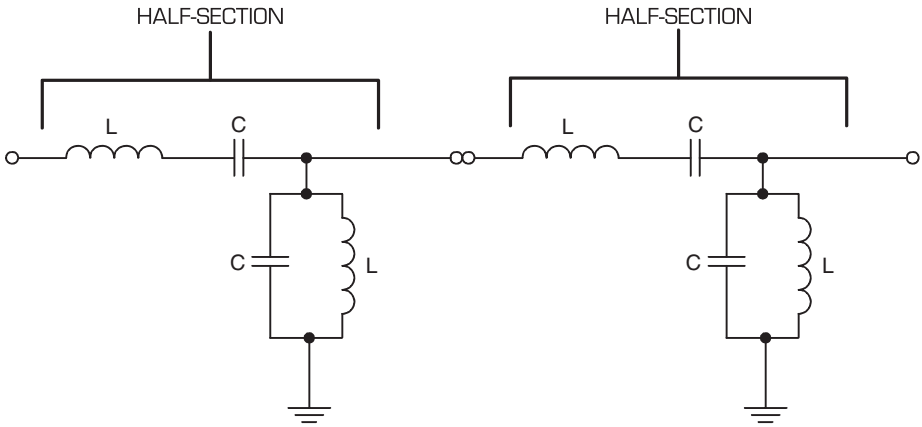


Figure 6.41 Improper way to join BPF half-sections.

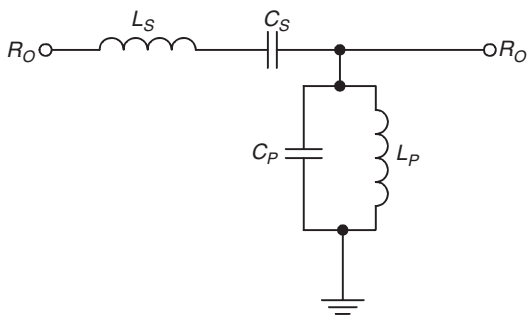


Figure 6.42 A BPF half-section.

where R_O = filter's Z_{IN} and Z_{OUT}
f_{2C} = high cutoff frequency
f_{1C} = low cutoff frequency

Then, combine half-sections to create a filter with more poles as shown in Fig. 6.43.

For an example, let us say that a six-pole filter will be required. It must have a bandwidth of 50 MHz, located between 475 and 525 MHz, and with a Z_{IN}/Z_{OUT} of 50 Ω. First calculate the proper element values for the half-section:

$$L_S = \frac{\left[\dfrac{50}{(525\text{ MHz} - 475\text{ MHz})\pi}\right]}{2} = 159\text{ nH}$$

and

$$C_S = 2\left[\frac{525\text{ MHz} - 475\text{ MHz}}{50(525\text{ MHz} \times 475\text{ MHz})4\pi}\right] = 0.64\text{ pF}$$

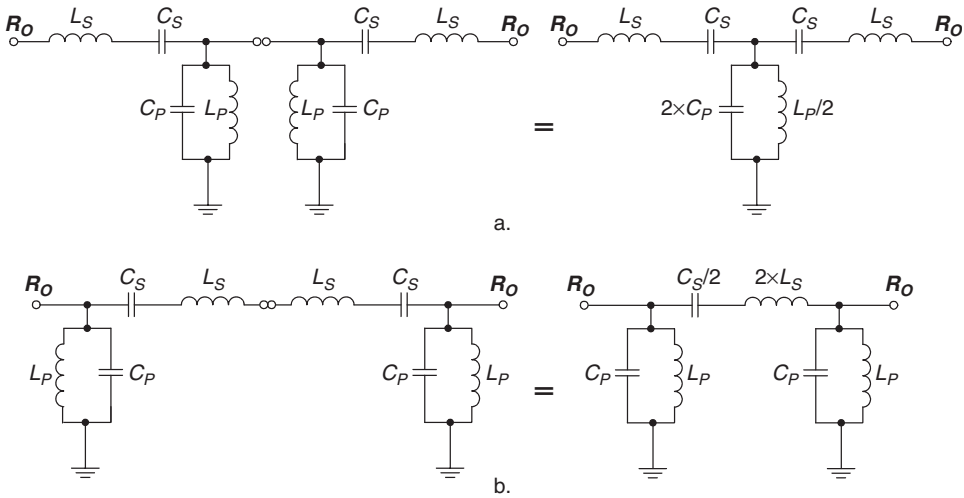


Figure 6.43 Combining two BPF half-sections when placed (*a*) tank to tank or (*b*) series to series.

and

$$L_p = 2\left[\frac{50(525\text{ MHz} - 475\text{ MHz})}{(525\text{ MHz} \times 475\text{ MHz})4\pi}\right] = 1.59\text{ nH}$$

and

$$CP = \frac{\left[\dfrac{1}{50(525\text{ MHz} - 475\text{ MHz})\pi}\right]}{2} = 63.6\text{ pF}$$

Transfer these values to Fig. 6.44, the bandpass filter's half-section. Begin adding and combining half-sections as shown in Fig. 6.45. Continue combining half-sections until the six-pole filter of Fig. 6.46 is obtained.

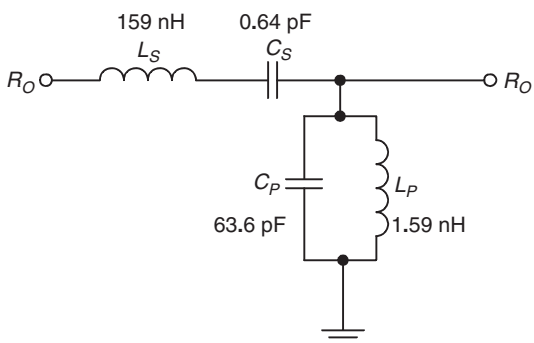


Figure 6.44 Values as calculated for BPF half-section.

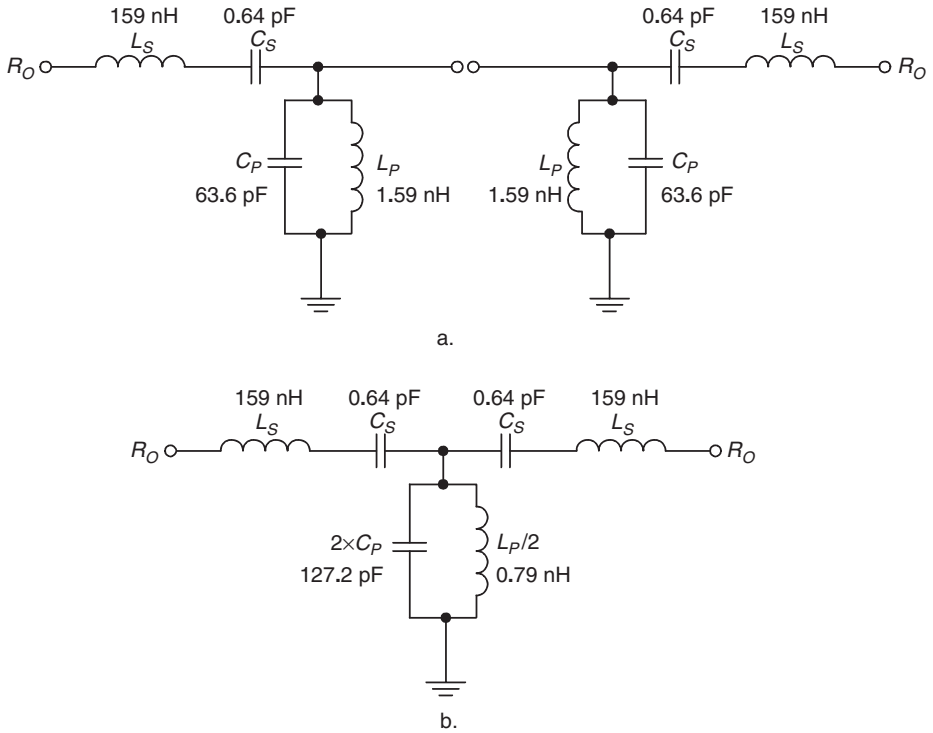


Figure 6.45 (*a*) Two single half-sections, and (*b*) combining the shunt tanks.

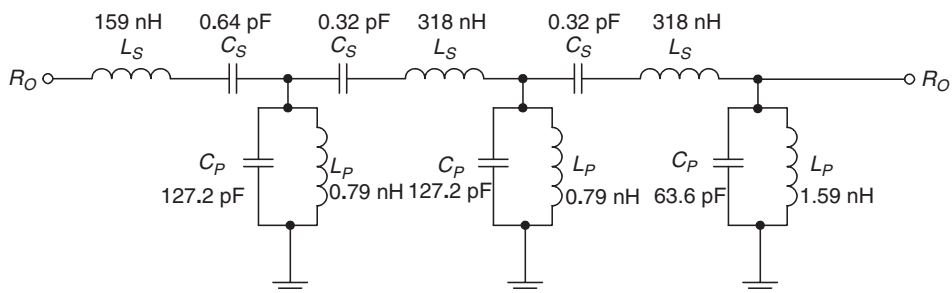


Figure 6.46 Completed six-pole bandpass filter.

6.2.4 Lumped Filter Design Issues

When designing a lumped filter, especially at higher frequencies, the highest Q inductors should be used to lessen insertion losses and the subsequent rounding of the passband edges. The capacitors, too, must be chosen carefully, since the filter's characteristics of bandpass, center frequency, and return loss will degrade if these capacitors (or inductors) have a poor tolerance value, deficient high-frequency performance, or inadequate temperature characteristics. In fact, the filters as designed may appear fine when a small

production run is tested at room temperature, but may become unacceptable when operated over temperature variations and/or over larger production runs.

An important specification for filters is their *ultimate attenuation* characteristics, which strongly depend on the number of filter sections, the unloaded Q of the components (especially the inductors), the parasitic resonances within all the passives, the PCB layout, and any RF shielding employed.

Lumped (and distributed) filters can become severely detuned if a hand, metallic object, or dielectric material is placed too close to the circuit. This is due to the *proximity effect*, and must be seriously considered when a lumped circuit design is synthesized, simulated, or built. (On the average, any RF shield or other conductive structure should be placed at least 200 mils from the top of the largest filter component within the circuit. Also, the filter's components, as opposed to the planar layout itself, own reaction to such a metallic structure cannot be readily simulated in any existing RF software.)

It is normally suggested to utilize odd-order filters in most designs, since even-ordered filters can be more vulnerable to termination impedances, possessing a degradation in their frequency responses. And many such filters cannot supply a 50-Ω output impedance when specified for even orders, with Chebyshevs being specially problematic in this regard.

6.3 Distributed Filter Design

6.3.1 Introduction

Most distributed filters are constructed on a PCB as planar microstrip structures. The actual materials to build such a filter are normally low in cost, and consist solely of copper traces and the substrate material itself. However, distributed filters can, depending on the passband frequency and the dielectric constant of the printed circuit board material, take up much more space than lumped LC filters. And, at certain repeating harmonic frequencies, pure (i.e., nonhybrid) distributed planar microstrip filters will typically display reentrant modes, which will destroy much of their vital stopband attenuation characteristics. Nonetheless, for certain microwave applications, distributed filters can be one of the more economical and practical choices for frequencies over 1.5 GHz.

Distributed bandpass filter structures cannot merely be a simple distributed copy of a lumped LC filter design. To obtain a narrow bandpass, with acceptable in-band ripple and insertion loss, specialized bandpass structures must be used. Unfortunately, the hand calculations needed to design such distributed filters are far too laborious. Therefore, most distributed bandpass filter design, as well as its optimization, is best left to a high-end software program, such as Agilent's Eagleware *M/Filters*. However, an easy to implement distributed BPF design is presented here for undemanding applications, while typical lowpass and highpass distributed structures are much more readily created, even without sophisticated synthesis software.

6.3.2 Distributed Filter Circuit Types

Due to various requirements and specifications demanded of microwave filters, such as stopband attenuation, a narrow versus a wide passband, return loss requirements, and microstrip element lengths, we see why such a multitude of distributed bandpass filter structures have been created over the years. Only the most common will be discussed here.

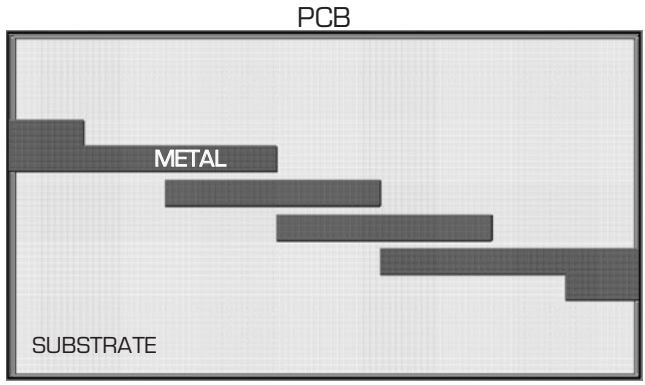


Figure 6.47 An edge-coupled distributed filter.

Nearly all of the following structures can be designed with the various general passband shapes, such as Butterworth, Chebyshev, and Bessel, as well as the less common *singly equalized transitional Gaussian* (6 and 12 dB) and the *elliptic* types of the *Cauer-Cheby*, with the use of the appropriate synthesis software:

Edge-coupled bandpass filters (Fig. 6.47) are effective in narrowband applications. They may radiate somewhat, which may necessitate RF shielding, and can be an unusually long structure at the lower microwave frequencies. For instance, edge-coupled bandpass filters will be up to 32 in long in a seven-order configuration at 1 GHz, depending on board dielectric constant and thickness. Nonetheless, the edge-coupled filter is one of the most common of all the distributed types.

Combline bandpass filters (Fig. 6.48) are hybrid types, employing both distributed (planar) and lumped (capacitor) elements, and are used in narrowband applications where size is at a premium. A seven-order configuration is only a little over 1 in long at 1 GHz.

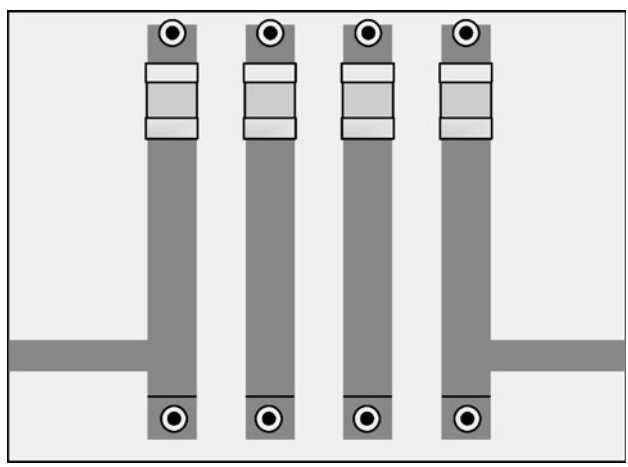


Figure 6.48 A fourth-order combline distributed filter.

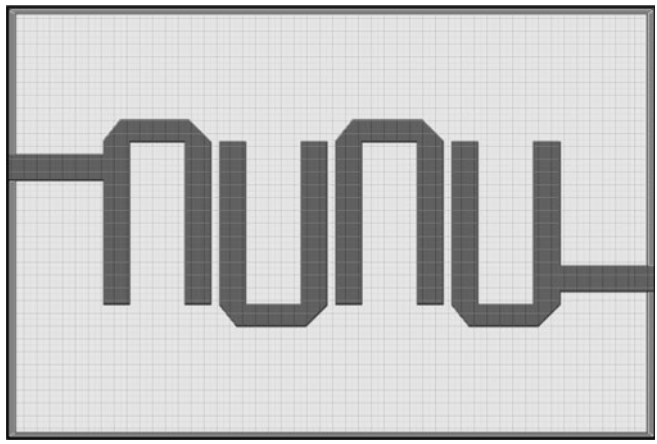


Figure 6.49 A fourth-order folded edge-coupled distributed bandpass filter.

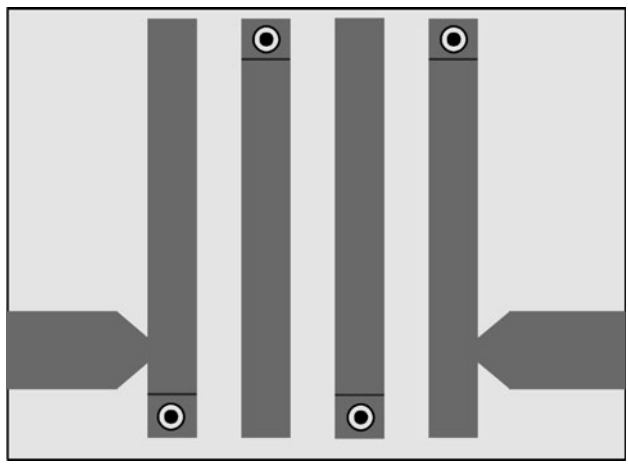


Figure 6.50 A fourth-order interdigital distributed filter.

Folded edge-coupled bandpass filters (Fig. 6.49) are utilized in narrowband filtering applications, and are very similar to the edge-coupled bandpass filters above. However, they are considerably shorter: 8 in at 1 GHz for the seven-order type.

Interdigital bandpass filters (Fig. 6.50) are also adopted in narrowband applications, and are quite compact: 2 in long at 1 GHz for a seven-order type.

6.3.3 Distributed Filter Design Methods

Microstrip Stubs

A stub (Fig. 6.51) is simply a shunt section of transmission line that is terminated in either a short or an open circuit. By employing a stub that has either a complete open or a complete short at one of its ends, the maximum peak-to-peak amplitude of standing waves will form, as well as the widest variations in impedance. The open end of a 90° stub

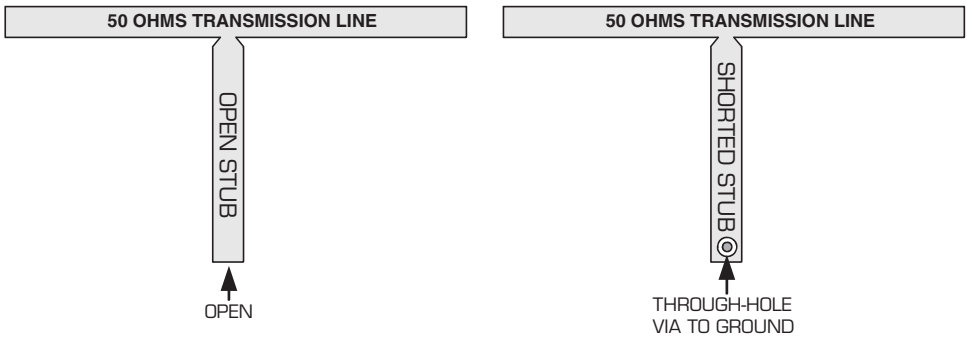


Figure 6.51 An open and a shorted microstrip stub.

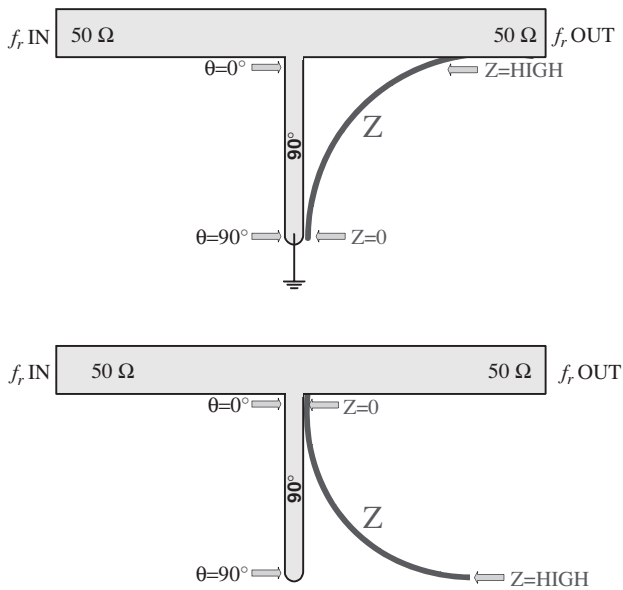


Figure 6.52 A shorted and an open stub, showing the impedances along each stub while at their 90° resonant frequency.

must be at maximum impedance (Fig. 6.52) and maximum voltage, with a current flow that is at zero (with the open end being considered simply as a very high impedance). Therefore, if we now measure the microstrip's impedance 90° backup from the open end of this stub, the impedance value will be extremely close to a direct short over a narrow band of frequencies, since the voltage and impedance must be at zero and the current at maximum. This is exactly equivalent to a lumped series resonant circuit, but will repeat for every other harmonic. When we now place a low-impedance load, such as a short, at the end of this same line, we see that 90° backup from the shorted end of this stub will have a very high impedance over a narrow band of frequencies, which can be considered as an open circuit. This is equivalent to a lumped parallel resonant circuit. And for every 90° we add to this stub, the impedance on this line would reverse.

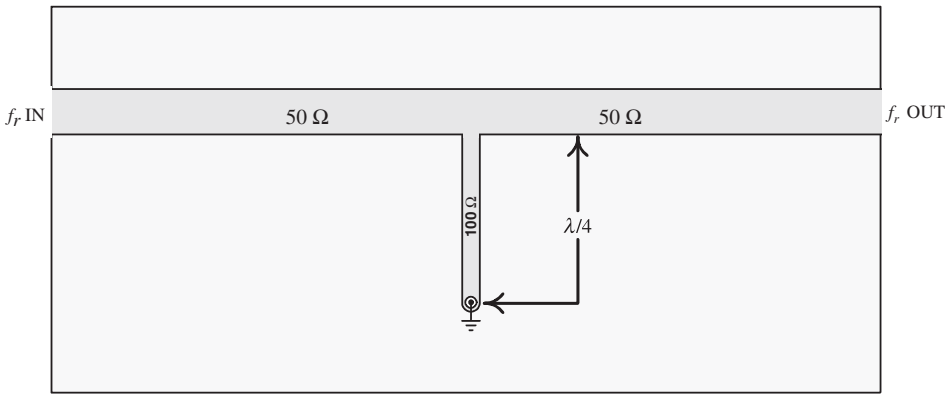


Figure 6.53 A microstrip layout for a high-impedance shorted stub.

At microwave frequencies, stubs are a very powerful and a very useful concept. Indeed, not only can we create filters with stubs, but we can also use them to form distributed components as a substitute for lumped versions, as well as construct distributed narrowband matching networks.

A Shorted Stub Bandpass Filter (Fig. 6.53)

To Design The width of the shorted stub is that of 100-Ω PCB microstrip, while the length of the microstrip is exactly a quarter wavelength, or 90°, and is grounded by a via (or RF capacitor) directly to the PCB's groundplane. This structure is used mainly as an RFC in DC bias lines at microwave frequencies.

1. To calculate how wide 100-Ω microstrip must be for the PCB substrate in use:

$$Z_0 = \frac{377}{\left(\frac{W}{h} + 1\right) \cdot \sqrt{\mathrm{E}_r + \sqrt{\mathrm{E}_r}}}$$

where Z_0 = characteristic impedance of the microstrip, Ω (in this case it must be 100 Ω)
W = width of the PCB microstrip conductor, uses same units as h
h = thickness of the substrate between the groundplane and the microstrip conductor, uses same units as W
E_r = dielectric constant of the board material

2. To calculate exactly how long the microstrip must be to be a quarter of a wavelength in length, with the chosen board material and frequency:
   a. Find the effective dielectric constant of the microstrip.

$$E_{\mathrm{EFF}} = \frac{E_r + 1}{2} + \left(\frac{E_r - 1}{2} \cdot \frac{1}{\sqrt{1 + \left(\frac{12h}{W}\right)}} \right)$$

where E_{EFF} = *effective* dielectric constant that the microstrip transmission line actually sees due to the dielectric/air interface
E_r = rated dielectric constant of the PCB's substrate material (found on the PCB's data sheet)
h = thickness of the substrate material between the top conductor and the bottom groundplane of the microstrip. Requires the same units as W
W = width of the top conductor of the microstrip. Requires the same units as h

b. After obtaining the effective dielectric constant of the microstrip's PCB substrate in use, find the propagation velocity of the microstrip.

$$V_P = \frac{1}{\sqrt{E_{EFF}}}$$

where V_p = fraction of the speed of light as compared to light in a vacuum for the microstrip transmission line, E_{EFF} = effective dielectric constant as seen by the microstrip transmission line.

c. Calculate the wavelength of the frequency of interest in a perfect vacuum:

$$\lambda = 11800 \div f$$

where λ = wavelength of the frequency of interest (f), mils
11800 = speed of light in mils for f, GHz
f = frequency of the signal of interest, GHz

d. Multiply the resultant velocity of propagation (V_p) times the wavelength (λ) of the signal times ¼ in order to arrive at the quarter wavelength of the signal of interest (in mils) when it is placed into the microstrip.

$$\lambda/4 \text{ (in mils)} = V_p \cdot \lambda/4$$

A Quick Example Design a Shorted Stub Bandpass Filter (Fig. 6.54)

Goal: Create a shorted stub BPF filter for microwave frequencies. The specifications and parameters for the structure are:

f_r = 5.8 GHz
Z_0 = 50 Ω
Substrate = Roger's RO-4003, 20-mils thick

Solution:

1. W = 10.5 mils
2. L = 330 mils

An Open Stub Bandstop Filter (Fig. 6.55)

A bandstop stub filter, which is mainly used to suppress a harmonic of the fundamental frequency, is basically just a reversed bandpass filter, so it is straightforward to design:

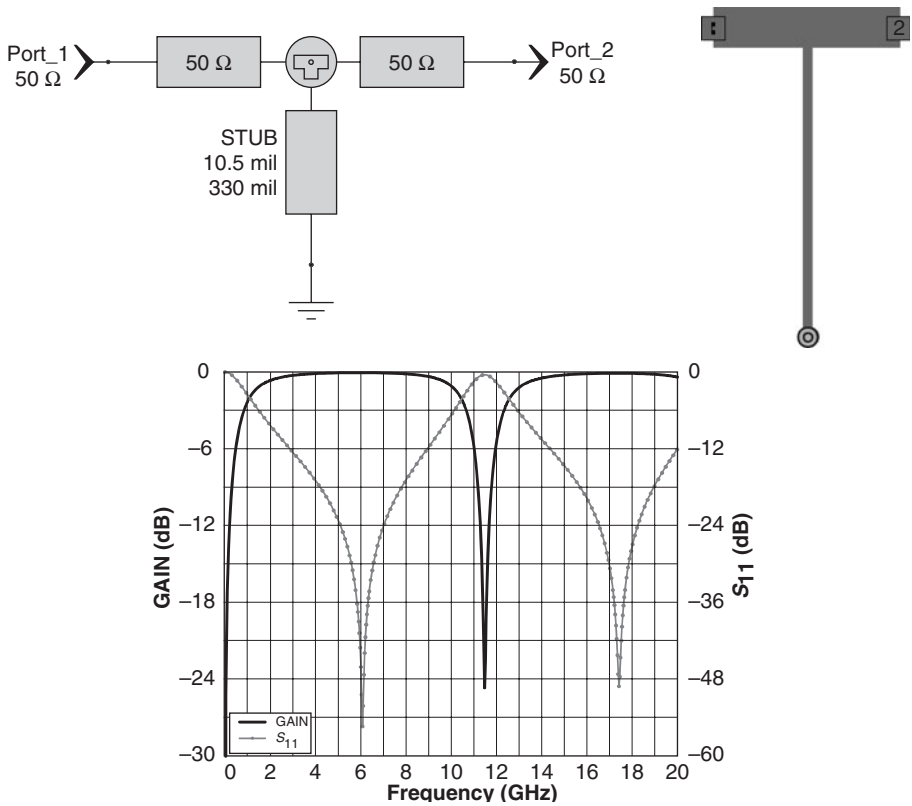


Figure 6.54 A shorted stub acting as a 5.8-GHz BPF, showing the simulated circuit, the actual physical layout, and the frequency response.

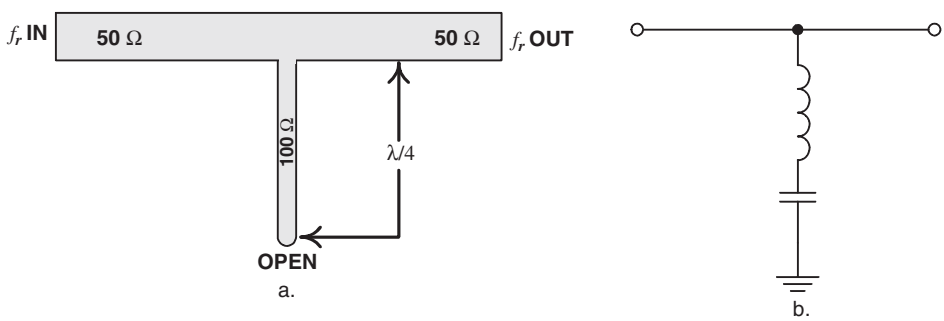


Figure 6.55 (*a*) A distributed bandstop filter with (*b*) equivalent lumped series shunt circuit.

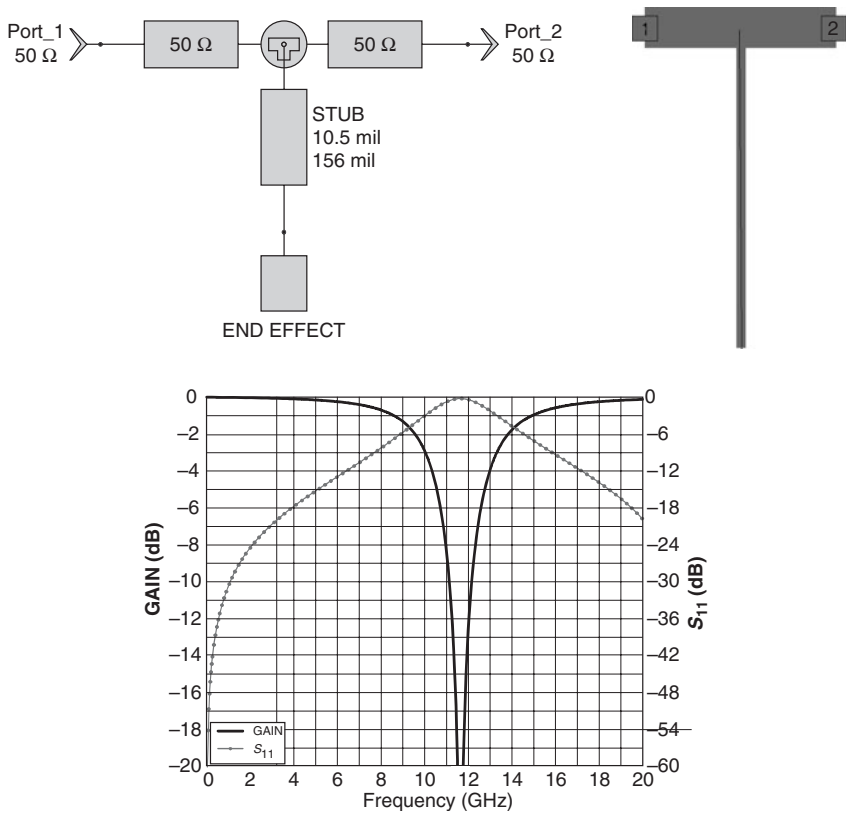


Figure 6.56 An open stub acting as a 11.6-GHz bandstop filter, showing the simulated circuit, the actual physical layout, and the frequency response.

simply duplicate the above bandpass design procedures, but leave the stub open instead of grounding it through a PCB via. Because it is now an open stub, however, the *end-effect* will demand that the length of the stub be trimmed down by approximately 5% below the calculated length. To do this, trim a small amount off the end of the open stub until the center frequency is as expected. We can also widen their bandpass by creating a wider stub (this is rarely desired in most applications).

A Quick Example Design an Open Stub Bandstop Filter (Fig. 6.56)

Goal: Create an open stub BSF filter for microwave frequencies. The specifications and parameters for the structure are:

f_r = 11.6 GHz (2nd harmonic of 5.8 GHz)

Z_0 = 50 Ω

Substrate = Roger's RO-4003, 20-mils thick

Solution:

1. W = 10.5 mils
2. L = 164 mils (optimized to 156 mils due to *end-effect*)

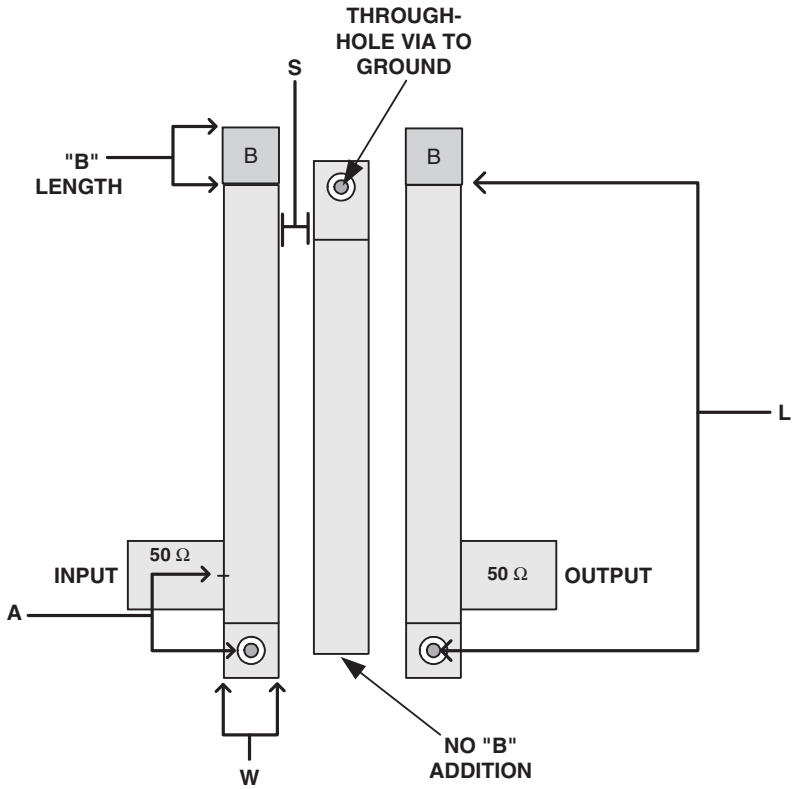


Figure 6.57 An effective third-order interdigital distributed BP filter design.

A Three-Pole Distributed Microstrip Interdigital BP Filter (Fig. 6.57)

The following Butterworth three-pole distributed filter is perfect for many basic bandpass filtering needs, but requires RF simulation to optimize for proper performance.

To Design

1. Compute the percentage of bandwidth required at the 3-dB down points at the center frequency of interest:

$$\mathrm{BW}_{3\mathrm{dB}} = \frac{(F_{u(3\mathrm{dB})} - F_{L(3\mathrm{dB})})}{f_{\mathrm{CENTER}}} \times 100$$

where $\mathrm{BW}_{3\mathrm{dB}}$ = percentage of the bandwidth at the 3-dB points, %
$F_{u(3\ \mathrm{dB})}$ = frequency of the upper 3-dB point, Hz
$F_{L(3\ \mathrm{dB})}$ = frequency of the lower 3-dB point, Hz
f_{CENTER} = filter's center frequency, Hz

2. Calculate the width W of the 60-Ω microstrip elements as in "Shorted Stub Bandpass Filter" under Sec. 6.3.3.
3. The length of dimension L will be 90° at f_r. Calculate the length L, in mils, of these 90° microstrip elements as in "Shorted Stub Bandpass Filter" under Sec. 6.3.3.

4. Calculate the length A of the two outer filter elements from the center of the microstrip input/output transmission lines to the PCB's ground via:

$$A = \text{CF} \cdot \left[L - \left(\frac{F_L}{F_u} \cdot L \right) \right]$$

where A = length to center of the input and output 50–Ω transmission line from the center of the PCB's ground via, mils
CF = correction factor. If the $BW_{3\,dB}$ as calculated above is 30% BW then CF = 1.30; 20% BW, CF = 1.35; 10% BW, CF = 1.70; 5% BW, CF = 2.0.
L = length of the 90° stub, as calculated in step 3 above, mils
F_L = lower 3-dB filter frequency (must be the same as that used for the $BW_{3\,dB}$ calculations above), Hz
F_u = upper 3-dB filter frequency (must be the same as that used for the $BW_{3\,dB}$ calculations above), Hz

5. To calculate the length required of extension "B" on each of the outer elements ("B" is the same width W as the elements themselves, or 60 Ω):

$$B = A \cdot \text{CF}$$

where B = additional length to be added to both *end* stubs, mils
A = length to the center of the input and output 50–Ω transmission line from the center of the ground, via as calculated above, mils
CF = correction factor required for various different 3-dB filter bandwidth percentages ($BW_{3\,dB}$): 30% BW then CF = 0.20; 20% BW, CF = 0.14; 10% BW, CF = 0.05; 5% BW, CF = 0.01

6. To find the proper spacing between each grounded stub section:

$$S = A \cdot \text{CF}$$

where S = spacing between adjacent stubs, mils
A = length to the center of the input and output 50–Ω transmission line from the center of the ground via, mils
CF = correction factor per various bandwidth percentages ($BW_{3\,dB}$) as calculated above: 30% BW, CF = 0.09; 20% BW, CF = 0.2; 10% BW, CF = 0.55; 5% BW, CF = 1.4

7. Ground each stub section directly to the PCB's groundplane through a via at the indicated end.
8. Place BPF design in an RF simulator, such as *Qucs*, and adjust the spacing S between elements for improved S_{21}, S_{11}, and to attain your desired bandwidth. Adjust the length L to shift the bandpass up or down.

A Quick Example Design a Three-Pole Distributed Bandpass Filter (Fig. 6.58)

Goal: Create a distributed bandpass filter for microwave frequencies. The specifications and parameters for the structure are:

f_r = 5.8 GHz
f_{PB} = 5.6 to 6 GHz

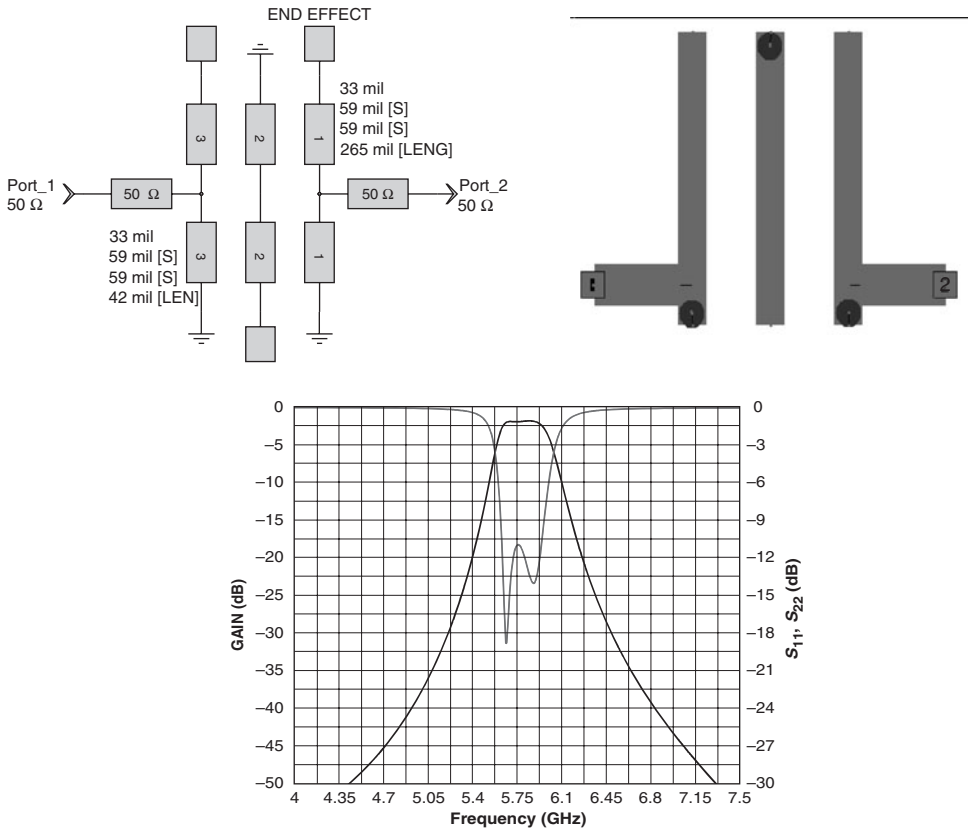


FIGURE 6.58 A worked example of the distributed 5.8-GHz bandpass filter, showing the simulated circuit, the actual physical layout, and its frequency response.

$Z = 50\ \Omega$

Poles = 3

Substrate = Roger's RO-4003, 20-mils thick

Solution:

1. BW% = 6.9%
2. $W = 33$ mils
3. $L = 316$ mils (optimized to 307 mils)
4. $A = 42$ mils (using 2.0 for CF)
5. $B = 0.42$ mils (using 0.01 for CF)
6. $S = 59$ mils (using 1.4 for CF)
7. 50-Ω microstrip = 45-mil width

Lowpass Filter

A microwave lowpass filter can be designed by using the distributed equivalents components of a lumped "prototype" filter design. Therefore, first design a lumped

filter, and then convert the lumped values to distributed equivalent components (see Sec.1.3.3, "Microstrip as Equivalent Components"). To guarantee that we will always have 50-Ω output impedances, such initial prototype filters should normally be of odd, and not even, orders.

Converting a lumped LPF into an equivalent distributed structure can sometimes be challenging due to excessively low or high line impedances required to meet a calculated lumped filter's equivalent reactances, while keeping within a line length shorter than 30°. This may be overcome, when necessary, by designing the LPF as a hybrid filter, with a combination of lumped and distributed parts. In most applications this will not be required, and a pure planar structure can be easily and quickly created.

Designing a Distributed Lowpass Filter Structure (Fig. 6.59)

To Design

1. With the desired filter response, create a prototype odd-order lumped LPF filter with image-parameter design techniques, or by using the enclosed *AADE Filter Designer*.
2. Duplicate each lumped part with its distributed equivalent: The width of each distributed component is 20 Ω for the capacitors (*C*), and 120 Ω for the inductors (*L*), as found in the enclosed *AppCAD* software for the substrate of interest.
3. The required length of each distributed element is:

$$\frac{\left(\text{ARCTAN}\dfrac{20}{1 \div (6.28 \cdot f_r \cdot C)}\right)}{360} \times \lambda = \text{LENGTH of capacitor}$$

and

$$\frac{\text{ARCSIN}\left(\dfrac{6.28 \cdot f_r \cdot L}{120}\right)}{360} \times \lambda = \text{LENGTH of inductor}$$

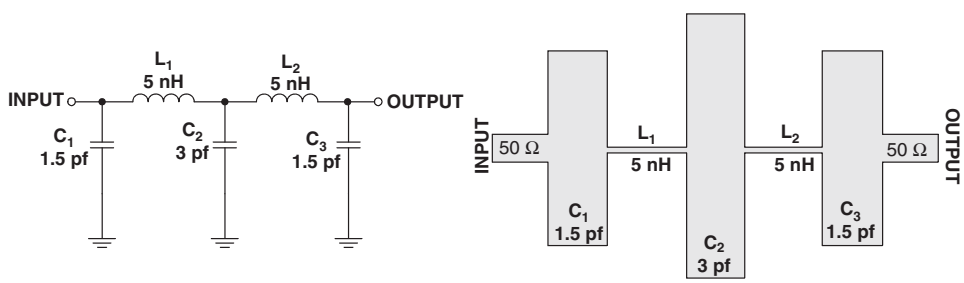


Figure 6.59 A lumped lowpass filter structure and distributed equivalent.

where λ = wavelength of the frequency of interest *using the substrate of interest* (or $V_p \times \lambda$), mils
C = value of the lumped capacitance, farads
L = value of the lumped inductance, henrys
f_r = cutoff frequency, Hz
LENGTH = length of component, mils

4. Replace each lumped part with its equivalent distributed component.
5. Tune inductor and capacitor lengths for optimal frequency response.

A Quick Example Design a Distributed Lowpass Filter (Fig. 6.60)

Goal: Create a distributed lowpass filter for microwave frequencies. The specifications and parameters for the structure are:

$f_{r(0.25\ \text{dB})} = 5.9$ GHz
$Z_0 = 50\ \Omega$
LPF type = five-pole Chebyshev
Ripple = 0.1 dB
Substrate = Roger's RO-4003, 20-mils thick

Solution:

1. Lumped prototype: L_1, $L_3 = 1.58$ nH; $L_2 = 2.72$ nH; C_1, $C_2 = 0.756$ pF.
2. Distributed: L_1, $L_3 = 6$ mils × 106.6 mils; $L_2 = 6$ mils × 208.6 mils; C_1, $C_2 = 160$ mils × 94 mils.
3. Tune for optimal response, as required.

Highpass Filter

A pure distributed highpass filter is not an easy or practical planar structure. This is due to adjacent coupling when employing a simple equivalent circuit, plus the added complications of having to employ series distributed capacitors. However, we can easily design a hybrid HPF. And to always obtain 50-Ω output impedances in all such filters, they should be of an odd, and not an even, order.

Designing a Distributed Highpass Filter Structure (Fig. 6.61)

To Design

1. With the desired filter response, create a prototype odd-order lumped HPF filter with image-parameter design techniques, or by using the enclosed *AADE Filter Designer*.
2. Duplicate the lumped inductors with their distributed equivalents. The width of each distributed inductor is 120 Ω, as found in the enclosed *AppCAD* software, for the substrate of interest.
3. The required length of each distributed inductive element is:

$$\frac{\text{ARCTAN}\left(\dfrac{6.28 \cdot f_r \cdot L}{120}\right)}{360} \times \lambda = \text{LENGTH}$$

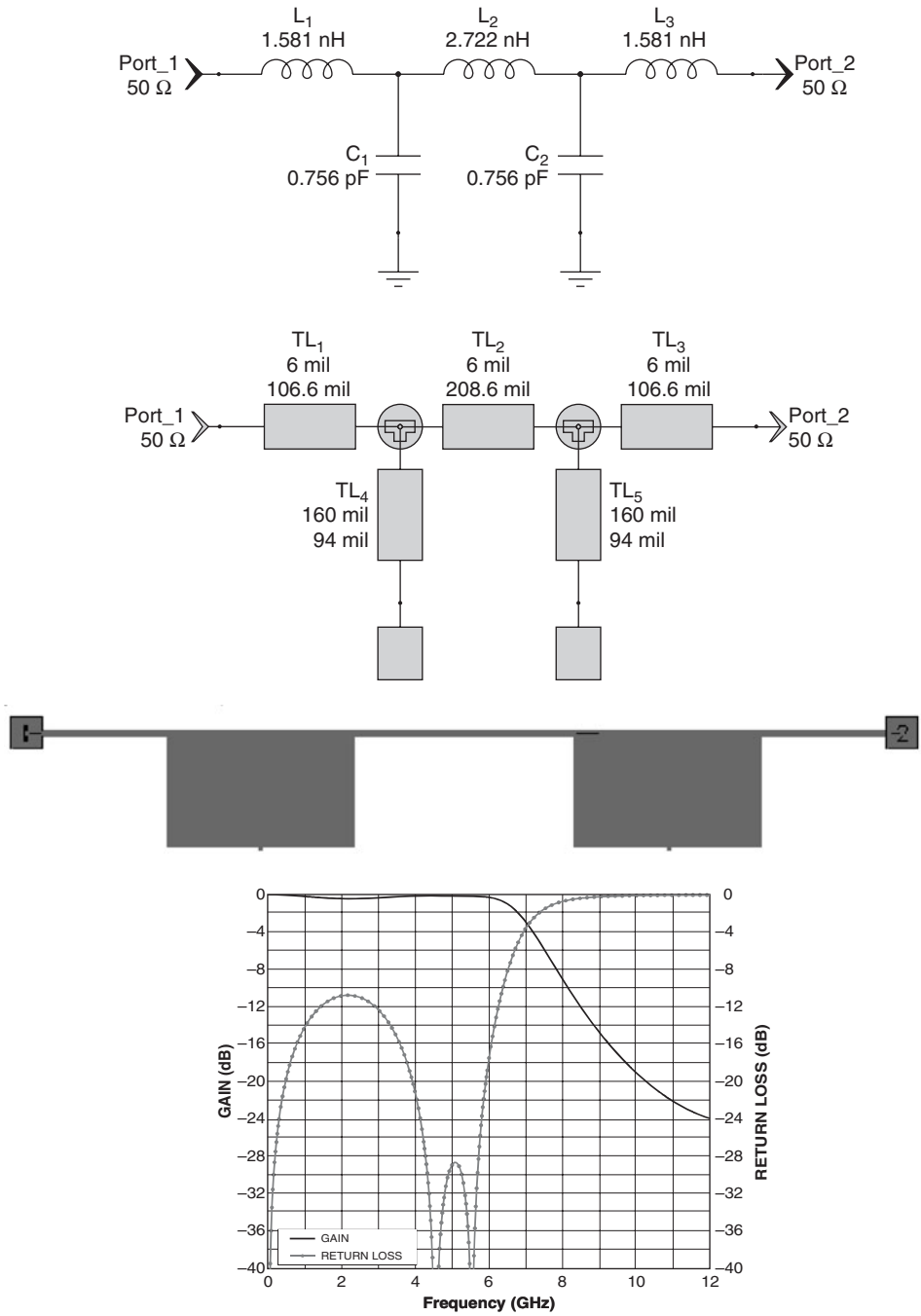


Figure 6.60 A worked example of one type of distributed 5.9-GHz lowpass filter, showing the lumped prototype, the simulated distributed circuit, the actual physical layout, and its frequency response.

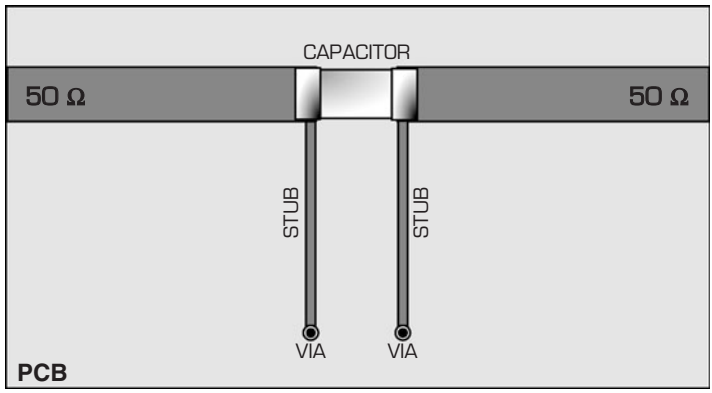


FIGURE 6.61 A distributed hybrid highpass filter.

where λ = wavelength of the frequency of interest *using the substrate of interest* (or $V_p \times \lambda$), mils
L = value of the lumped inductance, henrys
f_r = cutoff frequency, Hz
LENGTH = length of component, mils

4. Replace each lumped inductor with its equivalent distributed component, shorted to ground with vias.
5. Add the lumped series capacitors, using the same value as calculated for the above prototype lumped HPF.
5. Tune the inductor lengths and capacitor values for optimal frequency response.

A Quick Example Design a Distributed Highpass Filter (Fig. 6.62)

Goal: Create a distributed highpass filter for microwave frequencies. The specifications and parameters for the structure are:

$f_{r(0.25\ \mathrm{dB})} = 5.9$ GHz
$Z_0 = 50\ \Omega$
Type = three-pole Chebyshev
Ripple = 0.25 dB
Substrate = Roger's RO-4003, 20-mils thick

Solution:

1. Lumped prototype: L_1, $L_2 = 1.04$ nH; $C_1 = 0.47$ pF.
2. Distributed: L_1, $L_2 = 6$ mils × 65 mils; $C_1 = 0.47$ pF.
3. Tune for optimal response, as required.

6.3.4 Distributed Filter Issues

It is important for the measurement of stub lengths to be accurate, since the higher the frequency of interest, the shorter, and the less room for error, will be the distributed

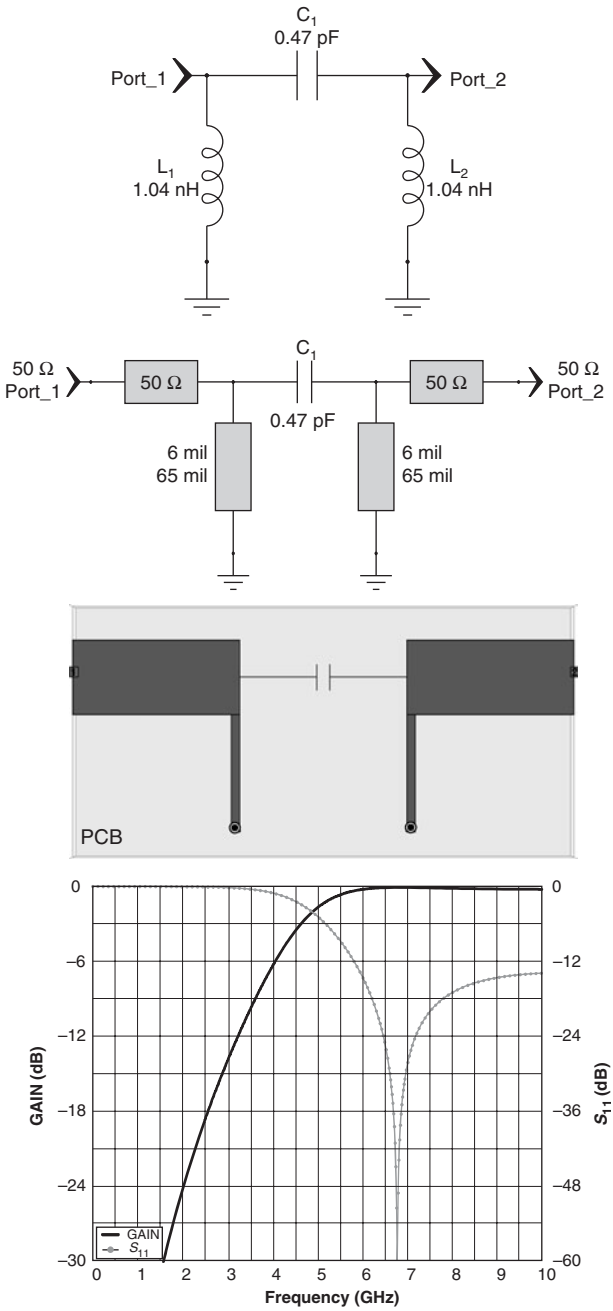


Figure 6.62 A distributed 5.9-GHz highpass filter design example, showing the lumped prototype, the simulated distributed circuit, the actual physical layout, and its frequency response.

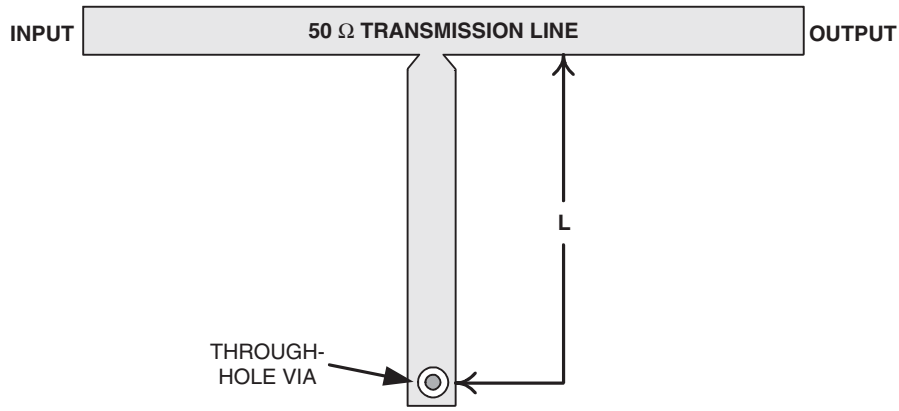


Figure 6.63 Proper distributed component length to its ground via.

stubs. An effective technique for accurate grounded stub measurement and layout is shown in Fig. 6.63, with the length of the stub measured from the "T" junction to the center of the through-hole via. PCB via length must be taken into account if the stub is short and/or the substrate is thick.

Element collisions can occur when the distributed microstrip in the filter's design become too close together, and may then actually touch or overlap. A different PCB layout or dielectric constant would then be indicated to avoid this.

Many distributed filters will have odd and even bandpass returns, called *reentrance*, that will reduce the filter's stopband attenuation to inferior levels at various frequencies. Distributed filters that integrate lumped components, such as the *combline* microstrip-capacitor filter, will have less obvious and less damaging harmonic reentrance passbands within their stopband frequencies. This same reentrance effect will also occur with lowpass and stopband filter structures.

As with most optimal microwave designs, we want to form a continuous RF connection between the top and bottom (i.e., next layer) groundplanes of the distributed filter's own printed circuit board, since this will reduce the inductance to ground and assist in reducing EMI levels and field coupling (as well as improving heat dissipation). We can do this by connecting together all such planes by the use of a substantial number of through-hole *stitching vias*. (Most microwave circuit boards will not only have a main groundplane layer located directly beneath the top component layer, but also a top groundplane that is poured in and around the periphery of the PCB's top circuit elements, and will cover most of the exposed bare surface area of the board.)

Distributed (and lumped) filters can become severely detuned if a hand is placed nearby, or if a metallic shield or dielectric material is located closer than designed. This is due to *proximity effects*, and must be seriously considered when a distributed design is synthesized or built. In fact, all electromagnetic microwave simulator programs (such as the included *Sonnet Lite*) are easily able to take many of these effects into account by permitting the engineer to set the distance from the distributed planar filter structure to the top of any RF shield that will be placed above it.

6.4 Diplexer Filters

6.4.1 Introduction

Diplexers are two or more combined filters within a single package that are used to separate two or more different bands of frequencies. This concept can be employed to split the transmit from the receive frequency in an *FDD* (*frequency division duplex*) transceiver, where in this application it is sometimes referred to as a *duplexer*. Or a diplexer can be placed at the output of a frequency mixer stage (Fig. 6.64), where it functions superbly as an *absorptive* filter. In this application, *Filter 1* of the diplexer has a passband that corresponds with the undesired frequency band, which can pass right through with little attenuation, and is terminated within the 50-Ω load. These same undesired frequencies are also blocked from entering *Filter 2* by that filter's own stopband. *Filter 2*'s passband, however, easily passes all of the desired signals onto the IF sections of the receiver with little attenuation. Therefore, because they are properly terminated into a 50-Ω resistive load, the undesired signals through the diplexer are absorbed, instead of being reflected as they would be in a typical filter. This absorption will prevent any undesired frequency products that were created by the mixer's nonlinearities from being bounced off of a reflective filter's stopbands, which would return to the mixer and cause increased IMD levels.

There are many different combinations of twin-filter diplexers: bandpass/bandpass, bandpass/lowpass, lowpass/highpass, and so on, depending on the particular wireless application.

6.4.2 Diplexer Filter Design

The design of a diplexer must be as two different frequency filters with nonoverlapping passbands. If the filter's passbands are placed too close together, they will interact adversely with each other. This harmful effect will decrease the return loss and increase insertion losses, while destroying the passband's flatness and symmetry. Further, each filter section of a bandpass diplexer should normally begin with a *series* resonant pole at the input, and not a shunt element, as a shunt element can short out the other filter's frequency of interest and completely destroy the return loss of the entire diplexer. Both bandpass filters of the diplexer of Fig. 6.65 have series resonant input poles instead of shunt, and possess a 50-Ω input. (A 100-Ω input impedance is not required for each

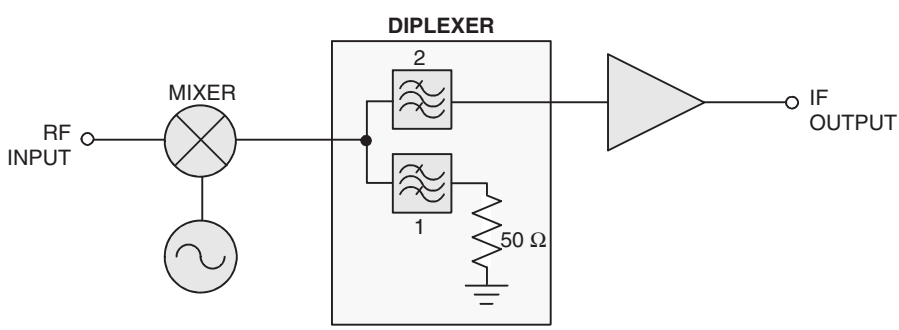


Figure 6.64 A BPF diplexer placed at the output of a conversion stage to decrease IMDs.

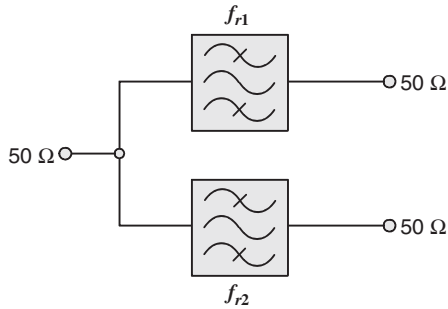


FIGURE 6.65 A BPF diplexer arrangement.

filter, as might be expected for two combined parallel stages for a 50-Ω system, as each filter is resonant at a completely different frequency, f_{r1} and f_{r2}; this allows each filter to pass only its own frequency of interest, with the other filter appearing as an almost complete open circuit to the other.)

Bandpass Absorptive Diplexer Design (Fig. 6.66)

A simple bandpass type of diplexer is shown, and can be used in many nondemanding, low-cost applications to terminate undesired frequencies without these frequencies reflecting back into the source. This is important in numerous mixer applications.

To Design

1. $Q = f_r/\text{BW}$
2. $L_2 = \dfrac{Q \cdot 50}{6.28 \cdot f_r}$
3. $C_2 = \dfrac{1}{L_2(6.28 \cdot f_r)^2}$

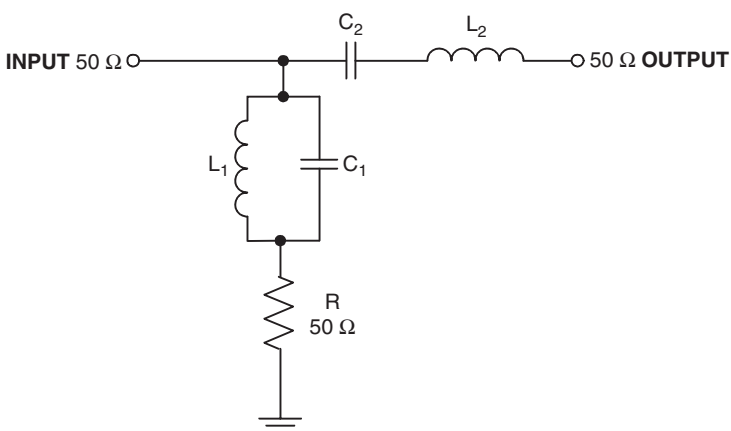


FIGURE 6.66 A type of BPF diplexer without a series input.

4. $L_1 = \dfrac{50}{Q \cdot 6.28 \cdot f_r}$

5. $C_1 = \dfrac{1}{L_1(6.28 \cdot f_r)^2}$

6. $R = 50\ \Omega$

where BW = bandwidth of the *desired* output signal at 3 dB down, Hz, fr = frequency of the *desired* signal, Hz.

NOTE: *The chosen Q for the above formulas must not be excessively high, or component values will become unrealizable.*

A Quick Example Design an Absorptive Lumped Bandpass Diplexer (Fig. 6.67)

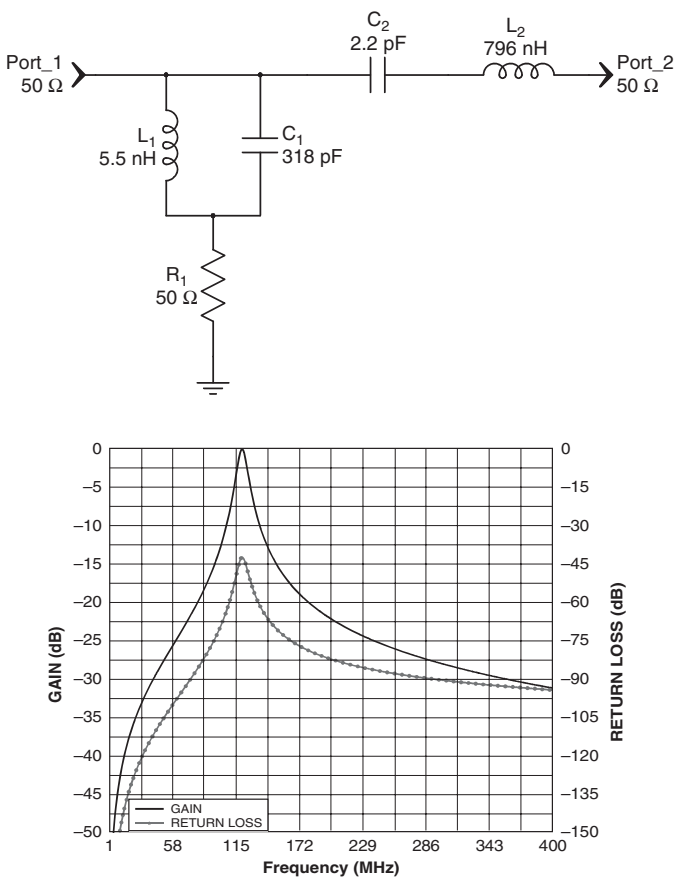


FIGURE 6.67 A diplexer filter that passes 120 MHz, but absorbs all other frequencies through the 50-Ω resistor.

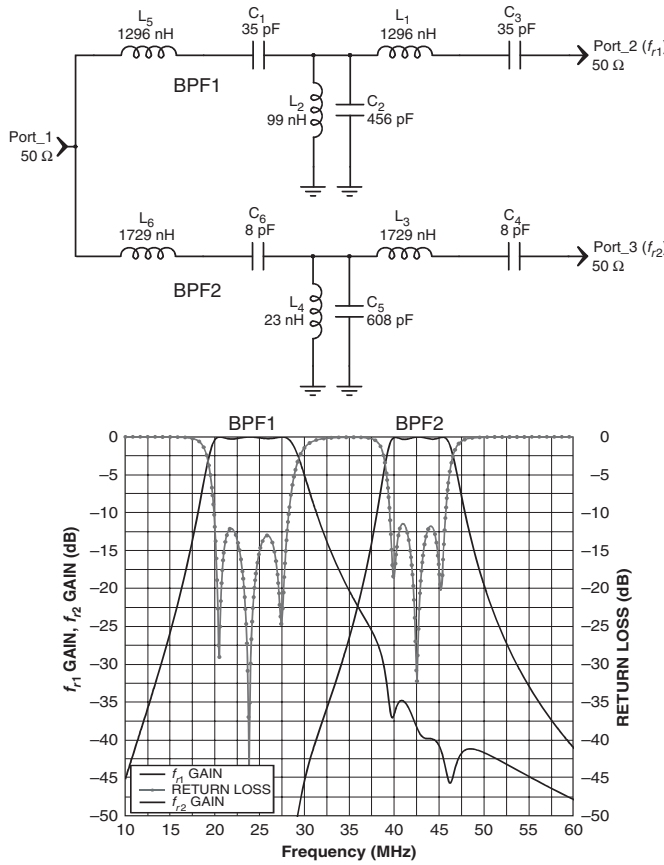


FIGURE 6.68 A diplexer filter to pass 27.12 and 40.68-MHz bands to separate ports, and its frequency response. (In this case, where the two desired frequencies are relatively close together, design each separate filter to have its own band edge as far away as possible from the other filter. Here, the desired frequency we want to pass is located at the upper band edge of BPF1, while the desired frequency of BPF2 is at its lower band edge).

Goal: Create a low RF frequency lumped absorptive bandpass diplexer filter. The specifications and parameters for the circuit are:

$f_r = 120$ MHz

BW = 10 MHz

$Z_0 = 50\ \Omega$

Solution:

1. $Q = 12$
2. $L_2 = 796$ nH
3. $C_2 = 2.2$ pF
4. $L_1 = 5.5$ nH
5. $C_1 = 318$ pF
6. $R = 50\ \Omega$

Twin Bandpass Diplexer Design

To pass two different frequencies to two different ports, other more selective diplexers with more poles can be rapidly designed by simply joining two standard 50-Ω filters together.

To Design Design two bandpass filters for two distinct frequencies with nonoverlapping passbands and series input poles using image-parameter techniques or the enclosed *AADE Filter Designer*, and then connect their inputs together.

A Quick Example Design a Lumped Twin Bandpass Diplexer Filter (Fig. 6.68)

Goal: Create a low-frequency lumped bandpass diplexer filter to pass two separate frequencies. The specifications and parameters for the circuit are:

$f_{r1} = 27.12$ MHz
$f_{r2} = 40.68$ MHz
Poles = 3
$Z_0 = 50\ \Omega$

Solution:

1. In the enclosed AADE Filter Designer, create a separate three-pole bandpass filter for each frequency starting with a series (not shunt) LC input.
2. Combine as shown.

6.5 Crystal and SAW Filters

6.5.1 Introduction

Due to a crystal's superior Q, they can be used in place of LC filters in certain low-frequency (IF) applications, especially for narrow bandwidth filtering with high selectivity.

SAW filters are also capable of replacing LC types, can operate at very high frequencies (up to 3 GHz), have excellent response characteristics, and come in a single SMD package.

6.5.2 Crystal Filters

All crystals have a series and parallel resonant mode. The parallel mode is slightly higher in frequency than the crystal's series resonance, and is due to the parallel capacitance of the crystal's holder, C_{PLATE}, as shown in Fig. 6.69. At series resonance R, which is a pure resistance of around 25 to 250 Ω (called the *equivalent series resistance*, or *ESR*), is the only impedance seen at resonance, since L and C will cancel each other. These resonant frequencies will not only depend on the thickness of the crystal, but also in the way it is cut, the crystal substance employed, and the holder capacitance.

However, additional crystal modes other than series and parallel can be used, such as the *overtone*, or *harmonic*, mode. A crystal is capable of being forced to resonate efficiently at odd harmonic intervals of its fundamental frequency, which are at the 3rd,

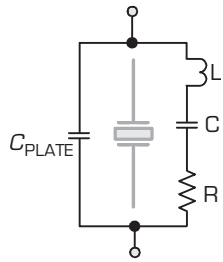


Figure 6.69 The equivalent circuit of a crystal in its holder.

5th, 7th, and up to the 11th harmonic. (To be completely correct, the overtone of a crystal is normally at a slightly different frequency than an exact odd integer multiple (harmonic) of the fundamental, and this is due to phase shifting within the crystal's structure.) The overtone modes will, unfortunately, force a crystal filter to have undesired reentrance modes at these roughly odd multiples of the series resonant frequency, causing decreased attenuation at specific points in the stopband. This can be overcome by special design procedures, and specifically by adding an LC lowpass filter at the crystal filter's output to attenuate the extra passbands created.

Crystals may be found singly, as well as in larger combinations, within RF filter packages: *crystal-lattice filters* (Fig. 6.70) contain several crystals within a single circuit, and are adopted for use as a very sharp bandpass filter. The input and outputs employ RF transformers with shunt capacitors, while each set of crystals, Y_1 plus Y_2 and Y_3 plus Y_4, are cut to different frequencies (the matched set of Y_1 and Y_2 having a lower resonant frequency than the other matched set of Y_3 and Y_4) so that we may attain the desired bandwidth and selectivity.

Another type of crystal filter structure is the *ladder filter* as shown in Fig. 6.71, and contains a stack of crystal filters, with coupling between the individual resonators being accomplished by capacitors. The coupling can also be done with shunt inductors. All the crystals in this filter are trimmed to the same series resonant frequency, and the input and output impedance matching can be achieved with an LC network.

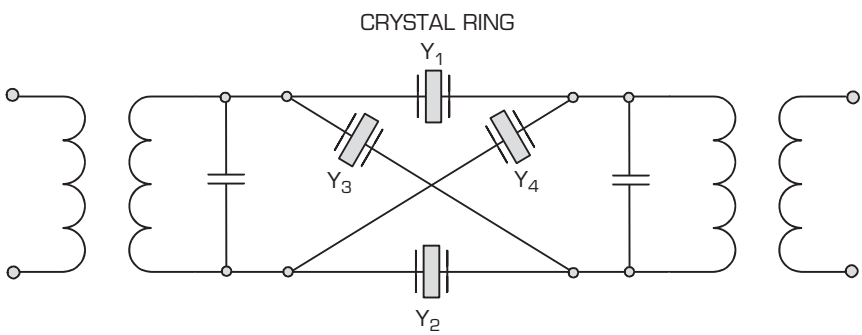


Figure 6.70 A crystal-lattice BPF circuit.

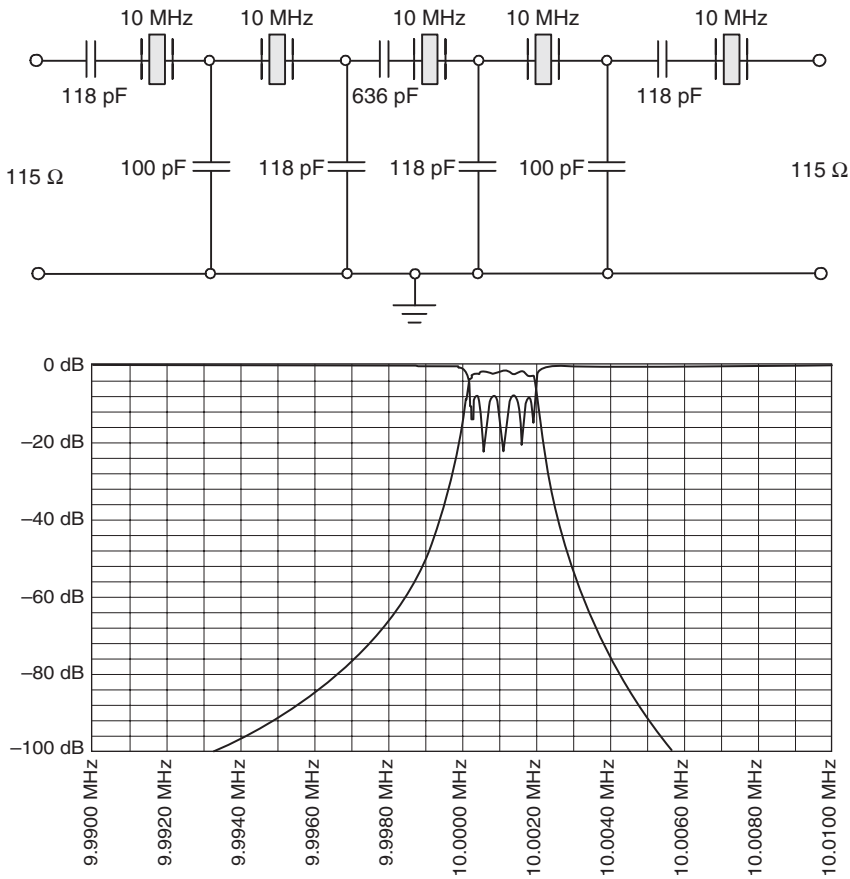


Figure 6.71 One popular type of crystal-ladder filter circuit with values and frequency response for 10.001 MHz.

Today, crystal filters are normally made to order from various specialized companies, and will comprise resonators, transformers, and trimmer capacitors all within a single, small package.

6.5.3 SAW Filters

SAWs have features that are hard to beat by any other filter type: They have as close to a brickwall filter response as can be obtained, are physically small, do not need to be tuned, and are low in cost (when purchased in large production quantities).

Using assorted substrate materials and internal circuit topologies, there are many different types of SAW filters, with various insertion losses, impedance matching requirements, frequency responses, center frequencies, bandwidths, stopband suppression characteristics, and amplitude ripple values. Selecting exactly which SAW

to use for a particular application will demand speaking directly with the SAW filter companies, or their representatives, to discuss performance requirements.

To give a general idea as to current SAW capabilities and tradeoffs: depending on the particular design, several types are capable of operating up to 3 GHz, while others can efficiently function down to 20 MHz. Some SAWs may need no input/output impedance matching at all, while others may demand external passives in order to supply a reasonable impedance match. Numerous SAWs that demand an external match may actually be forcing the SAW into a *desired* mismatched condition, instead of a perfect conjugate match, to minimize phase and amplitude ripples within their passband. Certain SAWs may have huge insertion losses as great as 30 dB, yet are capable of wide bandwidth operation (> 20%), along with negligible in-band ripple, an incredible 60-dB stopband attenuation, and an almost brickwall-like frequency response. Another type of SAW filter might have a built-in 50-Ω match, along with an extremely low insertion loss of 1 dB, yet may have a maximum possible bandwidth limitation of 4%, along with a very large amount of amplitude ripple (> 2 dB), a stopband suppression of only 40 dB, and relatively poor shape factor of 3:1. So, as with anything in life, compromises must be made. This is especially true when dealing with SAW devices.

When laying out a SAW on a PCB, use a top copper groundplane placed right under the SAW package itself, with multiple board vias stitched tightly (at better than $\lambda/20$ from via to via) from this top groundplane to the next internal PCB analog groundplane layer. It may also be advantageous to employ an isolation slot (a slit devoid of copper) in the top groundplane that is located directly under the SAW, and bisects the input from the output of the SAW package. This can help breakup the circulating RF ground current loops, preventing any undesired frequencies from simply blowing by the SAW's normally superb stopband performance. RF isolation can be further improved by keeping all the SAW's input and output ground pins away from the top groundplane pour, and from each other, to force the RF input ground currents to flow through the small inductance of the ground vias on the input side of the SAW, across the internal or bottom groundplane, and up through the inductance of the ground vias on the SAW's output side. If this particular layout method was not used, the RF ground currents could easily flow from the SAW's input ground pins, straight across the very low impedance of the top groundplane, and thus into the SAW's output ground pin.

More on SAW layout procedures, due to its importance to SAW performance, is available in Sec. 13.2.5, "Printed Circuit Board Design for Miscellaneous Circuits".

6.5.4 SAW Filter Issues

SAWs are physically small, and possess a delicate internal electrode structure that may be easily damaged by low levels of DC voltage or RF power. Consequently, all RF and DC input levels must be maintained below the SAW's maximum data sheet specifications.

While any signal pin of a SAW can be used as the input or output port, the actual impedance match for each port may be slightly different. For optimal performance, therefore, the same SAW package pins must be presented as the input or output in all subsequent wireless builds during an entire production run.

SAW's are inherently capacitive in nature, and this capacitance must be tuned out. To obtain a proper impedance match to a 50-Ω system, certain SAWs may need a series inductor placed at each port (some SAW matching circuits can be a little more complex).

There are a few limitations and problems with SAW filters that hinder their usefulness to some extent:

1. SAWs cannot have any of their filter specifications, such as center frequency, bandwidth, and insertion loss, changed during a production run, since they are manufactured in large quantities by a masking process that is similar to an integrated circuit.
2. Off-the-shelf SAWs can only be readily obtained at certain common center frequencies and bandwidths. This is because of the expense of custom manufacturing these filters in smaller quantities.
3. Some wide bandwidth SAWs can have extremely high insertion losses of up to 30 dB, which may have to be compensated for by inserting an amplifier circuit after the SAW filter.
4. Variations in temperature across some SAW devices may cause increased BER in digitally modulated radios.
5. Unpredictable discharges of energy from the SAW structure can occur with some wideband devices, causing possible damage or destruction of other components within the circuit. If this is an issue with the chosen SAW, it can be alleviated by placing shunt 5-kΩ resistors at the SAW's input and output ports.
6. SAW filters may have odd or even harmonic spurious responses. These spurious responses will degrade the SAW's stopband attenuation.
7. SAWs can have long time delays of up to one or more microseconds as the RF travels through the filter's internal structure, which can be problematic in certain systems.

6.6 Active Filters

6.6.1 Introduction

Active filters typically use an operational amplifier and an RC filter network to obtain lowpass, highpass, bandstop, and bandpass responses at low frequencies.

Passive RC networks themselves can be employed alone as a simple, nonresonant filter for certain audio applications, and can be utilized to attenuate RF, while passing only DC and low-frequency AC. As an example, the basic RC filter of Fig. 6.72*a* works as

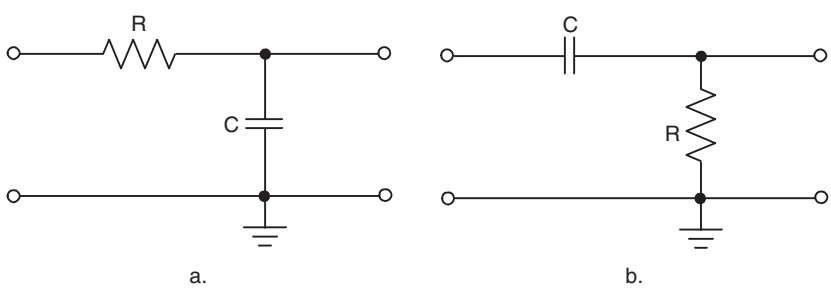


Figure 6.72 (*a*) A basic RC lowpass filter and (*b*) a basic RC highpass filter.

a lowpass filter by a voltage divider action: The capacitor *C* will have a low reactance to higher frequencies, while the resistor is chosen to be of such a value as to be of a significantly higher resistance to higher frequencies than *C*. Thus, high-frequency signals are dropped across *R*, while little RF will be dropped across *C*. However, with lower frequencies, the reactance of *C* is higher than the resistance of *R*, so the low frequencies get dropped across *C*, and are then tapped from the output with low attenuation.

Reversing the resistor with the capacitor will create the opposite effect, producing a highpass filter (Fig. 6.72*b*). Thus, any low frequencies will now be dropped across the high reactance of *C*, but not across the lower resistance of *R*. Higher frequencies will easily pass through the lower reactance of *C*, but be dropped across the higher resistance of *R*. Since the output is across *R*, a highpass filter has now been formed.

However, if we now insert a high-gain amplifier, such as an op-amp, within the above RC filter, we will obtain buffering from the effects of the filter's load, as well as a sharper filter response curve and an insertion *gain* instead of an insertion *loss*.

Figure 6.73 is a common RC active lowpass filter, and functions so: C_2 passes the higher frequencies to ground, while C_1 sends a degenerative feedback to the noninverting input as the frequency increases. This is caused by the capacitor's decreasing capacitive reactance at increasing frequencies. Thus, C_1, C_2, R_1, R_2 and the op-amp will efficiently form an active lowpass filter.

An active highpass filter is shown in Fig. 6.74, with R_2, C_1, and C_2 forming a simple highpass filter that is used to send the higher frequencies to the op-amp's input with little attenuation. At the lower frequencies, however, the increasing capacitive reactance of C_1 and C_2 attenuates, decreasing their signal at the filter's output.

The active bandpass filter of Fig. 6.75 employs a feedback network that readily passes all frequencies back to its input that are not within the filter's passband and, since this feedback is degenerative, all but a narrow passband of desired frequencies will be attenuated.

When designing active op-amp filters for use at these low, usually voice or data frequencies, it is assumed that the driving source into the filter is 0 Ω, that the filter's own input impedance is infinite ohms, that the filter itself sees 0 Ω when looking back into the prior driving stage, and that the filter itself sees infinite ohms when looking forward into the next stage. These assumptions work well, since real-life active filters

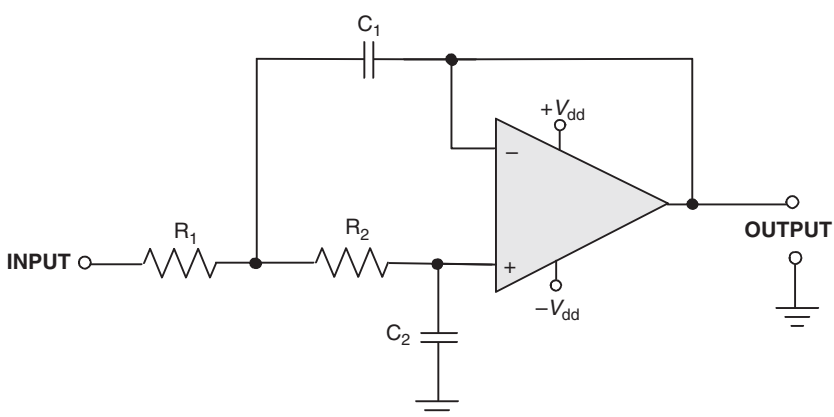


Figure 6.73 An active lowpass filter.

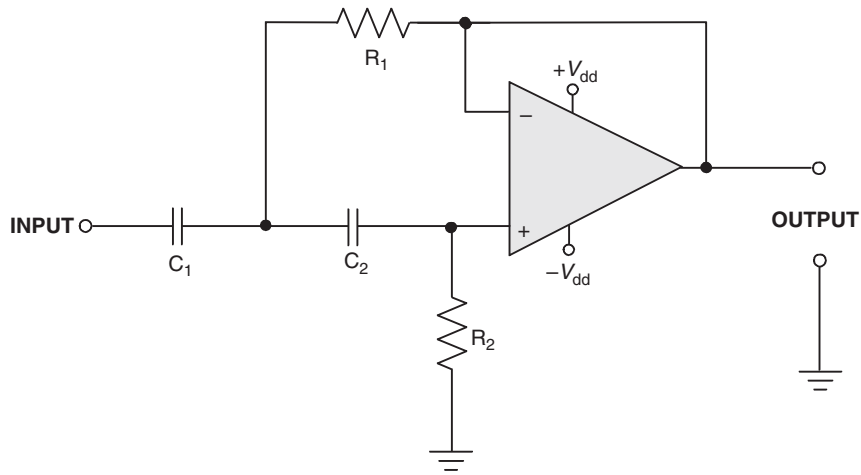


Figure 6.74 An active highpass filter.

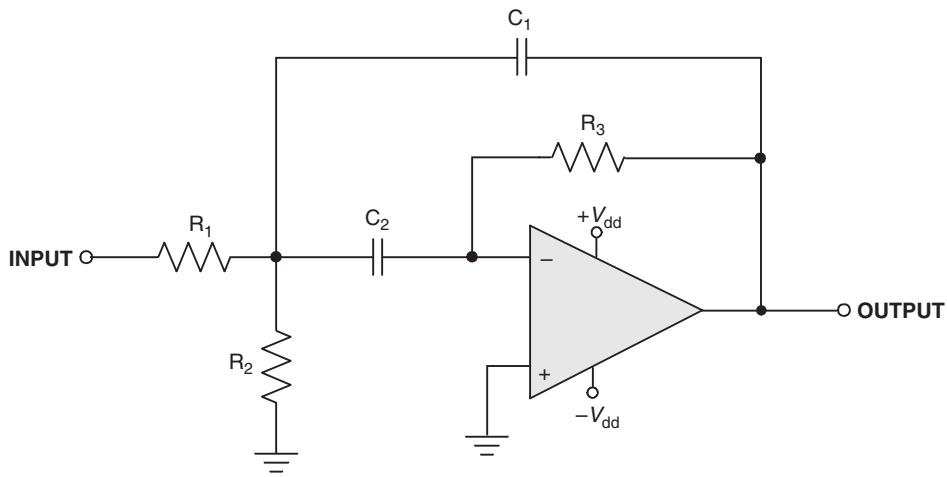


Figure 6.75 An active bandpass filter.

will easily be capable of effectively driving a load that is of far less than "infinite" impedance, and without significantly decreasing their performance. After designing such active filters with these ideal synthesis concepts, we must then simulate their design by using the appropriate linear and nonlinear Spice models.

6.6.2 Active Filter Design

All of the active filters below have a Bessel or Butterworth response. These active filters are meant to be driven from a low-impedance source that is significantly below R_1 in value, and should only be driving a high-impedance load placed at their output.

A Bessel Active Audio Bandpass Filter, Dual Supply (Fig. 6.75)

To Design

1. Select the filter's desired voltage gain (A_v). A gain of 10 (20 dB) is a safe value.
2. Select the filter's desired center frequency (f_c) and bandwidth (BW).
3. Calculate the required Q of the filter by $Q = f_c/\text{BW}$. ($2 \cdot Q^2$ must be greater than A_v, or the filter design will not function as expected.)
4. $C_1 = C_2 = 0.1$ uF
5. $R_1 = \dfrac{Q}{A_v \cdot 6.28 \cdot f_c \cdot C_1}$
6. $R_2 = \dfrac{Q}{(2Q^2 - A_v) \cdot 6.28 \cdot f_c \cdot C_1}$
7. $R_3 = \dfrac{Q}{\left(\dfrac{6.28 \cdot f_c \cdot C_1 \cdot C_2}{C_1 + C_2}\right)}$

NOTE: *Active filters can be cascaded for increased selectivity. However, the bandwidth will become increasingly narrower as more active sections are added. This problem can be neutralized by anticipating this reaction, and designing the initial filter section* wider *than we would if employing but a single filter. As a very rough guide, we should multiply the original bandwidth (BW) by 1.5 in the above equations for each section added in cascade. For example, if we require a three-stage bandpass filter with a bandwidth that will pass 500 Hz, then we would design the initial single stage to have a bandwidth of 1.5 × 1.5 × 500 Hz, equaling 1125 Hz, and then simply add two more of these duplicate 1125-Hz bandwidth stages to the first filter. The three-stage cascaded audio filter will then end up with a final bandwidth of approximately 500 Hz.*

A Bessel Active Audio Lowpass Filter, Dual Supply (Fig. 6.73)

To Design

1. $A_v = 1$
2. $C_1 = C_2 = 0.022$ uF
3. $R_1 = R_2 = \dfrac{1}{6.28 \cdot f_c \cdot \sqrt{C_1 \cdot C_2}}$

NOTE: *When cascading lowpass filter sections for added selectivity, multiply the cutoff frequency* f_c *by 1.5 in the above equation for each filter section we plan to cascade. For instance, if we are cascading two sections, and we desire a 1-kHz 3-dB cutoff frequency, then design each section with a* f_c *of 1.5 kHz. The final cascaded output will then have an approximately 1-kHz lowpass cutoff.*

A Bessel Active Audio Highpass Filter, Dual Supply (Fig. 6.74)

To Design

1. $A_v = 1$
2. $C_1 = C_2 = 0.022$ uF
3. $R_1 = R_2 = \dfrac{1}{6.28 \cdot f_c \cdot \sqrt{C_1 \cdot C_2}}$

NOTE: *When cascading highpass filter sections for added selectivity, multiply the cutoff frequency* f_c *by 0.75 in the above equation for each filter we plan to cascade. For instance, if we are cascading two sections, and we desire a final 1-kHz 3-dB cutoff frequency, then design each section to have an* f_c *of 750 Hz. The final cascaded output will then have an approximately 1-kHz highpass cutoff.*

A Butterworth Audio Lowpass Filter, Unity Gain, Second Order, Single Supply (Fig. 6.76)

To Design

1. $C_1 = \dfrac{1}{f_r \cdot 10{,}000}$
2. $C_2 = 2 \cdot C_1$
3. $R_3 = R_4 = 100\ \text{k}\Omega$
4. $R_1 = R_2 = \dfrac{1}{9 \cdot f_r \cdot C_1}$
5. $C_{\text{IN}} = C_{\text{OUT}} = 500 \cdot C_1$

A Quick Example Design an Active Lowpass Filter (Fig. 6.76)

Goal: Create an audio-frequency active lowpass filter with a positive supply. The specifications and parameters for the circuit are:

$f_{r(3\text{ dB})} = 20$ kHz
$V_{CC} = +5$ V
Order = second
Response = Butterworth
Gain = unity

Solution:

1. $C_1 = 1$ nF
2. $C_2 = 2$ nF
3. $R_1 = 5555\ \Omega$

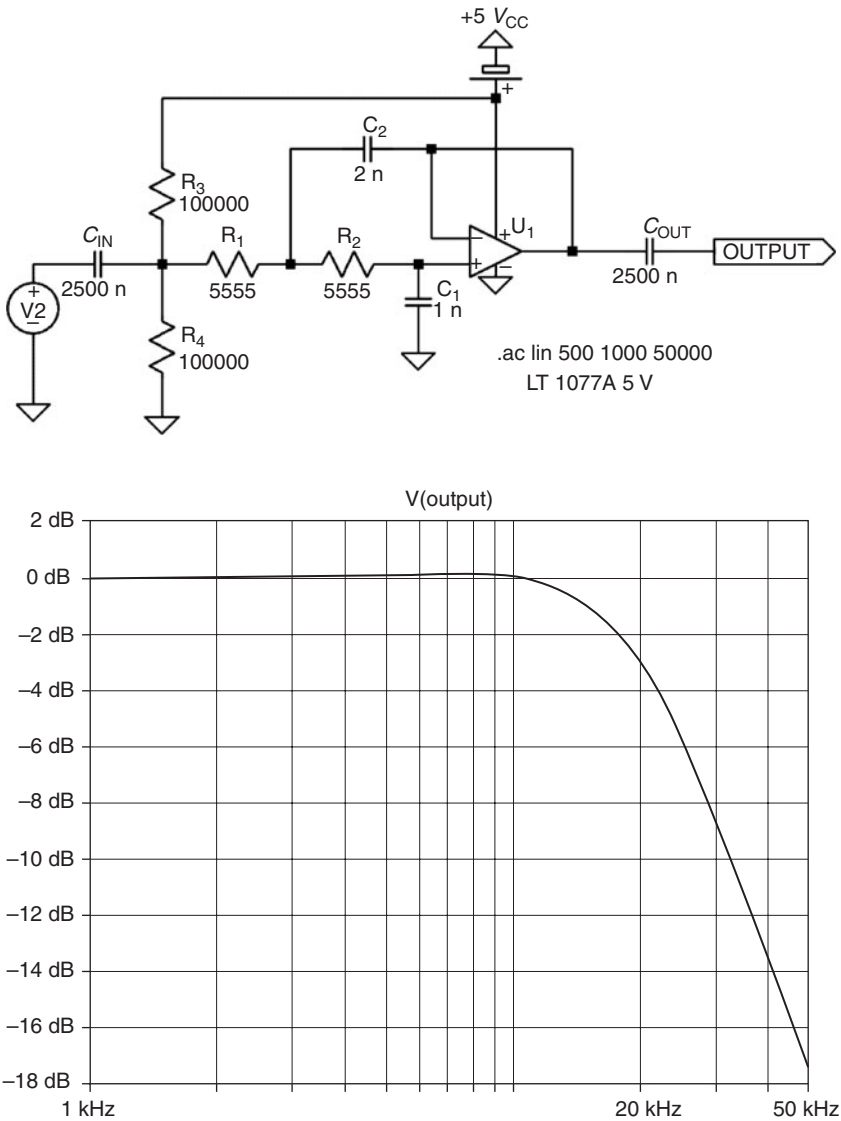


Figure 6.76 An active op-amp lowpass filter example for 20 kHz, with its (log) frequency response and part's values.

4. $R_2 = 5555\ \Omega$
5. $R_3 = 100\ \text{k}\Omega$
6. $R_4 = 100\ \text{k}\Omega$
7. $C_{IN} = 2500\ \text{nF}$
8. $C_{OUT} = 2500\ \text{nF}$

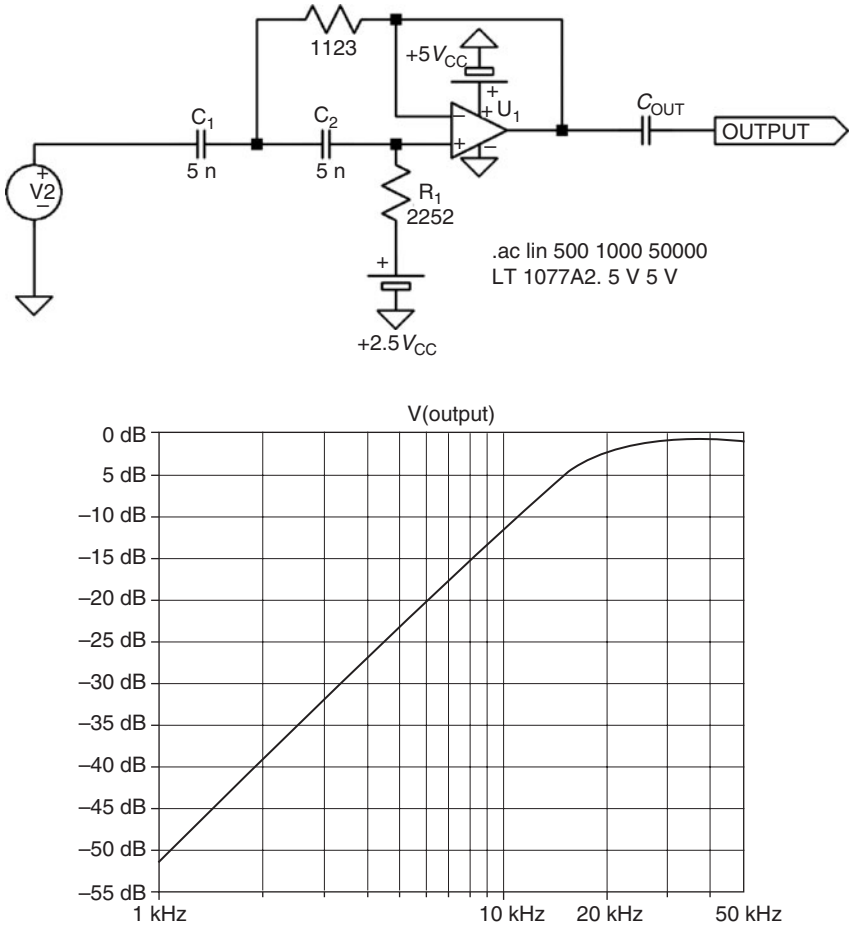


Figure 6.77 An active op-amp highpass filter example for 20 kHz, with its (log) frequency response.

A Butterworth Audio Highpass Filter, Unity Gain, Single Supply (Fig. 6.77)

To Design

1. $C_1 = C_2 = \dfrac{1}{f_r \cdot 10{,}000}$
2. $R_1 = \dfrac{1}{4.44 \cdot f_r \cdot C_1}$
3. $+V_{CC2} = +V_{CC1}/2$
4. $R_2 = \dfrac{1}{8.9 \cdot f_r \cdot C_1}$
5. $C_{OUT} = 500 \cdot C_1$

A Quick Example Design an Active Highpass Filter (Fig. 6.77)

Goal: Create an audio-frequency active highpass filter with a positive supply. The specifications and parameters for the circuit are:

$f_{r(3\text{ dB})} = 20$ kHz
$V_{CC} = +5$ V and $+2.5$ V
Order = second
Response = Butterworth
Gain = unity

Solution:

1. $C_1 = 5$ nF
2. $C_2 = 5$ nF
3. $R_1 = 2252\ \Omega$
4. $R_2 = 1123\ \Omega$
5. $C_{OUT} = 2500$ nF

A Butterworth Audio Bandpass Filter, 20-dB Gain, Narrowband, Single Supply (Fig. 6.78)

To Design

1. $C_1 = C_2 = \dfrac{1}{f_r \cdot 10{,}000}$
2. $R_1 = \dfrac{1}{6.28 \cdot f_r \cdot C_1}$
3. $+V_{CC2} = +V_{CC1}/2$
4. $R_3 = 19 \cdot R_1$
5. $R_2 = R_1/19$
6. $C_{IN} = C_{OUT} = 500 \cdot C_1$

A Quick Example Design an Active Bandpass Filter (Fig. 6.78)

Goal: Create a low-frequency active bandpass filter with a positive supply. The specifications and parameters for the circuit are:

$f_{r(3\text{ dB})} = 19.5$ to 21.5 kHz
$V_{CC} = +5$ V and $+2.5$ V
Order = first
Response = Butterworth
Gain = +20 dB

Solution:

1. $C_1 = 5$ nF
2. $C_2 = 5$ nF
3. $R_1 = 1592\ \Omega$

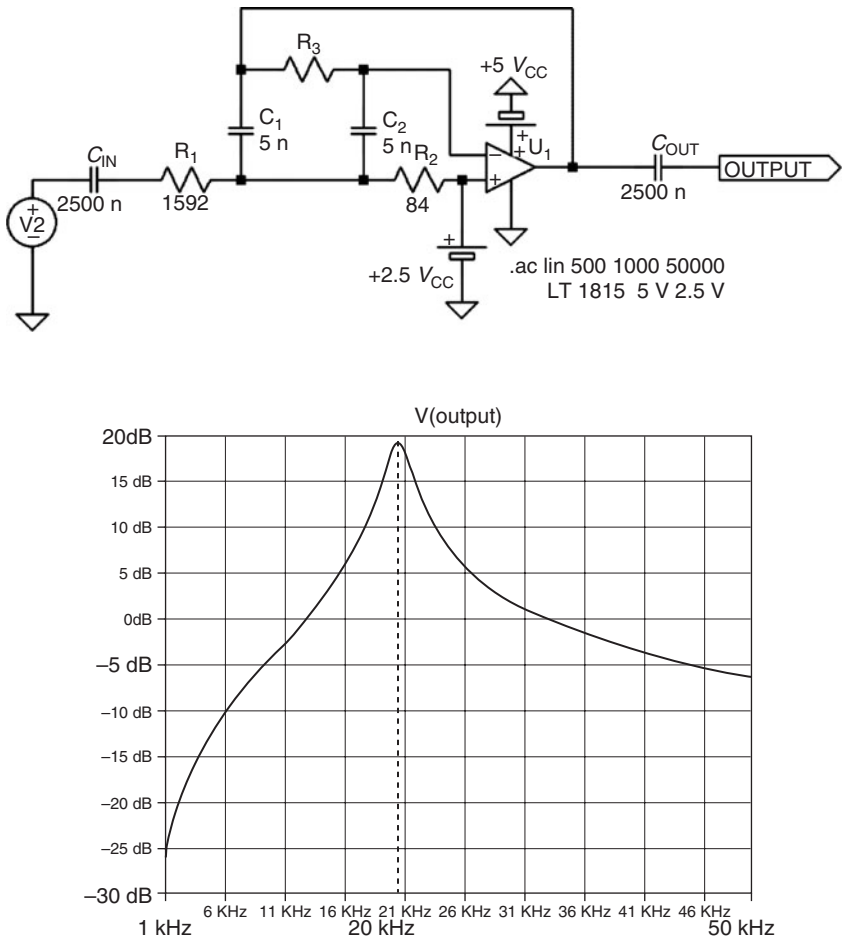


FIGURE 6.78 An active op-amp bandpass filter example for 19.5 to 21.5 kHz, with its frequency response.

4. $R_2 = 84\ \Omega$
5. $R_3 = 30.25\ \text{k}\Omega$
6. $C_{IN} = 2500\ \text{nF}$
7. $C_{OUT} = 2500\ \text{nF}$

6.7 Tunable Filters

6.7.1 Introduction

A filter that can change frequency with the application of a control voltage across it can be an invaluable asset in the design of receivers and transmitters in today's packed spectrum. Certain wideband transmitter frequency conversion stages may also benefit

by allowing the selective filtering of excessive local oscillator feedthrough from entering the IF or RF stages.

This electrical tunability is most simply and effectively accomplished with the use of varactor diodes within the filter's network. Single varactors are easily capable of changing capacitance from 0.63 to 2.67 pF, all the way up to 3.8 to 20 pF (and above), by the placement of a 0 to 20-V control voltage. This range of capacitances is perfect for most high-frequency tunable filter designs.

6.7.2 Tunable Filter Design

Below are presented two tunable RF filter designs, ideal for many diverse wireless applications. The *top inductively coupled variable bandpass filter* will be a better performer than the *capacitively coupled variable bandpass filter* in wideband applications.

Capacitively Coupled Variable Bandpass Filter, 50 Ω (Fig. 6.79)

To Design

1. Design a basic *top capacitively coupled bandpass filter* (Fig. 6.80) with any standard RF filter design program, such as the included *AADE Filter Designer* software. Select a center frequency for the top capacitively coupled bandpass filter at either the high end, low end, or middle of the tunable range of the desired bandpass frequencies, depending on the initial tuning voltages you plan to supply to the tuning varactors.
2. Now, remove C_1 and C_2 of Fig. 6.80, add the replacement components of C_V (a varactor diode) and C_T, along with the bias resistors of R_1 and R_2, as shown

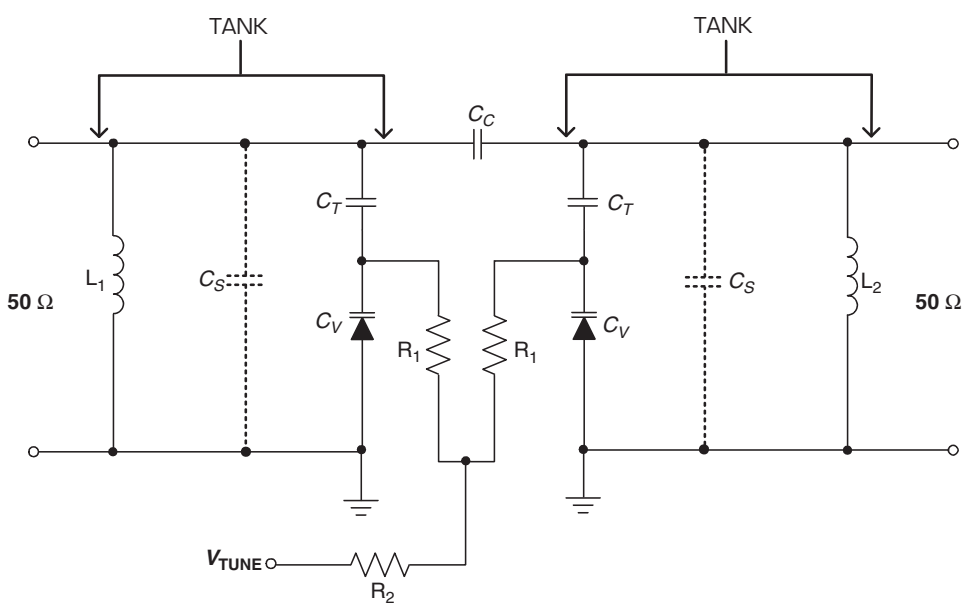


Figure 6.79 A tunable BPF employing varactor diodes.

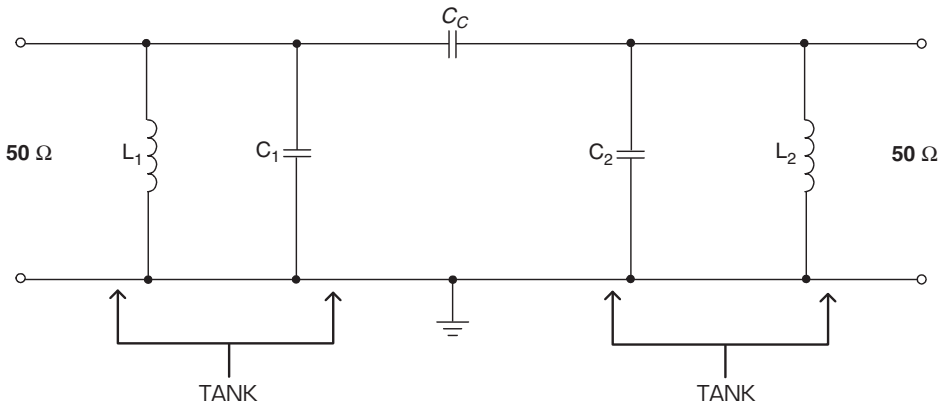


Figure 6.80 A typical capacitor-coupled BPF used as the prototype for the tunable design.

in Fig. 6.79. R_1 isolates the two varactors from the effects of each other and, with R_2, prevents a direct RF short to ground through V_{TUNE}; C_T blocks the DC inserted by V_{TUNE} from being shorted by L_1 or L_2; C_V supplies the variable tuning capacitance; C_C couples the two tank circuits consisting of L_1 and C_T/C_V, and L_2 and C_T/C_V. Select an R_1 of 24 kΩ each, and an R_2 of 100 kΩ. If V_{TUNE} is to be located at 10 V with a 20-V varactor, then choose a varactor diode (C_V) that has a value of C_1 in the center of the varactor's capacitance range, then select a C_T that is approximately 10 times this value. The capacitance of the series combination of C_T and C_V, in series, is:

$$\frac{C_T \times C_V}{C_T + C_V}$$

This is essentially the value of C_V alone due to the high capacitance of C_T. And since the value of C_1 and C_2 in Fig. 6.80 must equal this series combination of C_T and C_V of Fig. 6.79, then C_T is functioning only as a DC blocking capacitor, while C_V, the varactor, is supplying *all* of the tuning capacitance for the filter's tanks.

3. Apply the tuning voltage V_{TUNE} that will allow the varactor to either linearly tune the filter to its maximum and minimum values or, by supplying V_{TUNE} with discrete voltages, we can filter in discrete steps. Due to parasitic capacitances and inductances at these frequencies, the finished filter will have to optimize in software, using the appropriate components and pad/trace models.
4. Since varactors are extremely limited in the value of their maximum capacitance, and we may want to operate this filter at or below the VHF region, then we need some way of increasing the capacitance of the C_T and C_V combination (since C_1 and C_2 would need to be increased in value to operate at decreasing frequencies). The simple way of raising this over-all capacitance level of the

series C_T and C_V combination is to add a capacitor C_S in shunt with C_T and C_V, which will now increase the capacitance in each leg to:

$$C_S + \frac{C_T \times C_V}{C_T + C_V}$$

When calculated for a lower frequency, this will fulfill the larger required capacitance value as dictated by the C_1 and C_2 capacitors in Fig. 6.80. However, the relative tuning range will obviously decrease as the design frequency is decreased, since C_V will now have far less of an effect on altering the increased fixed capacitance of C_S (Fig. 6.81). As well, the *minimum* value of capacitance will be set by the value of C_S. Nonetheless, more

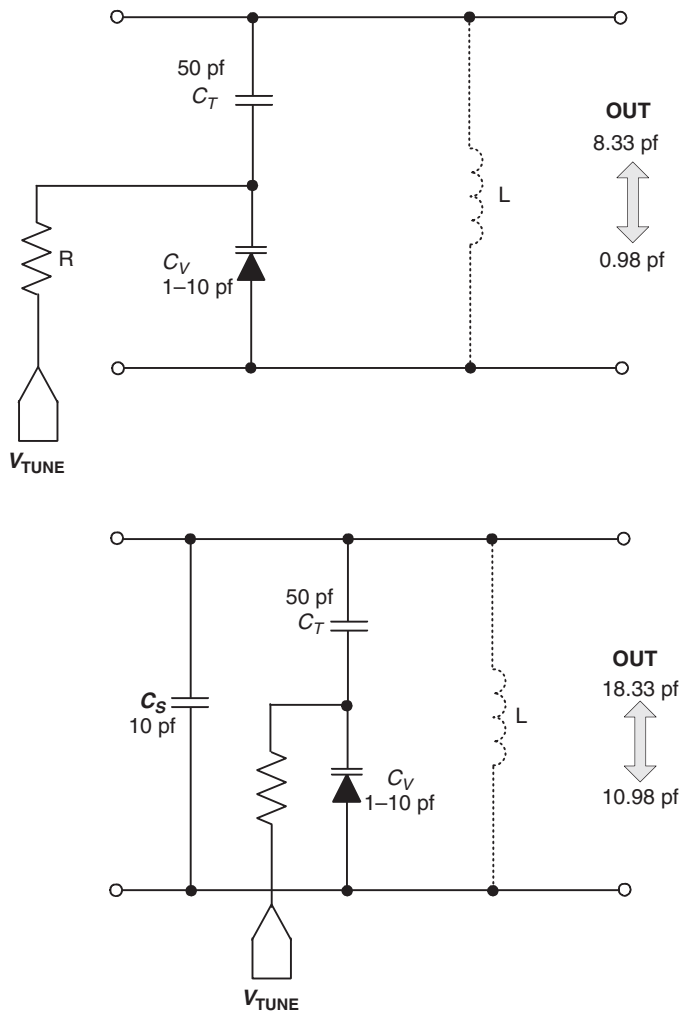


Figure 6.81 The tuning range capacitance of the varactor circuits without C_s and with C_s.

varactors can be paralleled to increase the tuning range, as required, and more varactor-tuned LC tanks can be added to increase the selectivity of this tunable filter.

To decrease any large degradation or variation in the filter's insertion loss over frequency, always select high-*Q* inductors. Since this filter as designed using software for the initial basic-coupled capacitive filter, this will permit the choice of the tank inductor's value, so select as low an inductance as the program allows (in order to decrease the filter's size and cost, as well as to increase the loaded *Q* and performance of the final design).

Over its entire tuning range, this type of filter cannot be expected to maintain its bandwidth, insertion loss, or return loss, but should be quite acceptable for most variable filtering applications when used under a half-octave tuning range. And if upper-frequency attenuation is more important than the low-frequency attenuation inherent in a top capacitor-coupled filter, then a *top inductively coupled filter* can be built as a base for this tunable filter (see "Top Inductively Coupled Variable Bandpass Filter, 50 Ω").

Top Inductively Coupled Variable Bandpass Filter, 50 Ω (Fig. 6.82)

We may use the capacitively coupled variable bandpass filter above, but as its frequency is tuned upward by its two varactor's, the filter's passband will widen

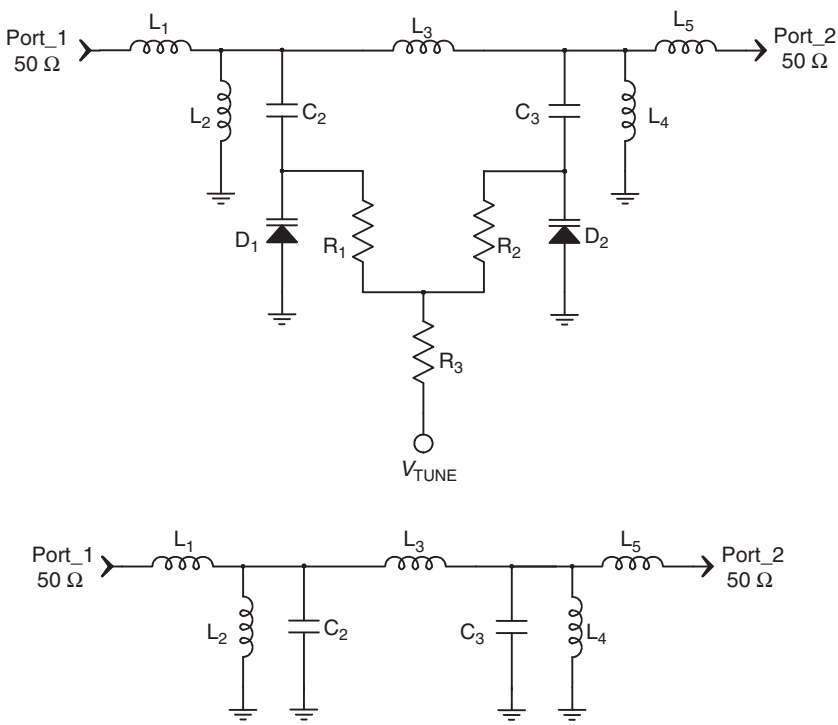


Figure 6.82 A voltage tunable inductor-coupled bandpass filter and its fixed-tune prototype.

and the insertion loss will decrease. In fact, a capacitively coupled variable bandpass filter's performance over a wide bandwidth can be somewhat poor for wider tuning ranges. However, over narrower tuning bandwidths (< 25% bandwidth), the capacitively coupled filter works well, and will still be a good choice when lower cost is an issue. A higher performing filter over wider bandwidths is the *inductively coupled* type. The inductively coupled filter will lessen the above effects, but not completely eliminate them.

To Design

1. Design a basic top inductively coupled bandpass filter circuit with any standard RF filter design program, such as the included *AADE Filter Designer* software. Select a center frequency for the top inductively coupled bandpass filter at either the high end, low end, or middle of the tunable range of the desired bandpass frequencies, depending on the initial tuning voltages you plan to supply to the tuning varactors.
2. Design exactly as with the above *capacitively coupled variable bandpass filter*, but substitute with the top inductively coupled variable bandpass filter prototype of Fig. 6.82.

A Quick Example Design a Tunable Bandpass Filter (Fig. 6.83)

Goal: Create a lumped tunable LC bandpass filter with inductive coupling. The specifications and parameters for the circuit are:

f_r = 470 to 770 MHz
BW = 50 MHz
V_{CNTRL} = 3.3 V
Z_{IN} = 50 Ω
Z_{OUT} = 50 Ω

Solution:

1. Design an inductor top-coupled bandpass "prototype" filter in *AADE Filter Designer*, with a center frequency approximately 620 MHz and a bandwidth of 50 MHz, which will then be used as the basis of the tunable filter. (The component values may differ from the presented example.)
2. Select a varactor that has the tuning capacitance within the V_{CNTRL} limitations. These varactors will replace the two-shunt capacitors. Due to the possible drop in Q at very low bias voltages, do not use a V_{CNTRL} lower than approximately 0.2 V.
3. R_1, R_2 = 24 kΩ; R_3 = 51 kΩ

NOTE: *Bandwidth will widen with increasing frequency. For the lowest possible filter insertion losses, select the highest Q varactors and inductors available.*

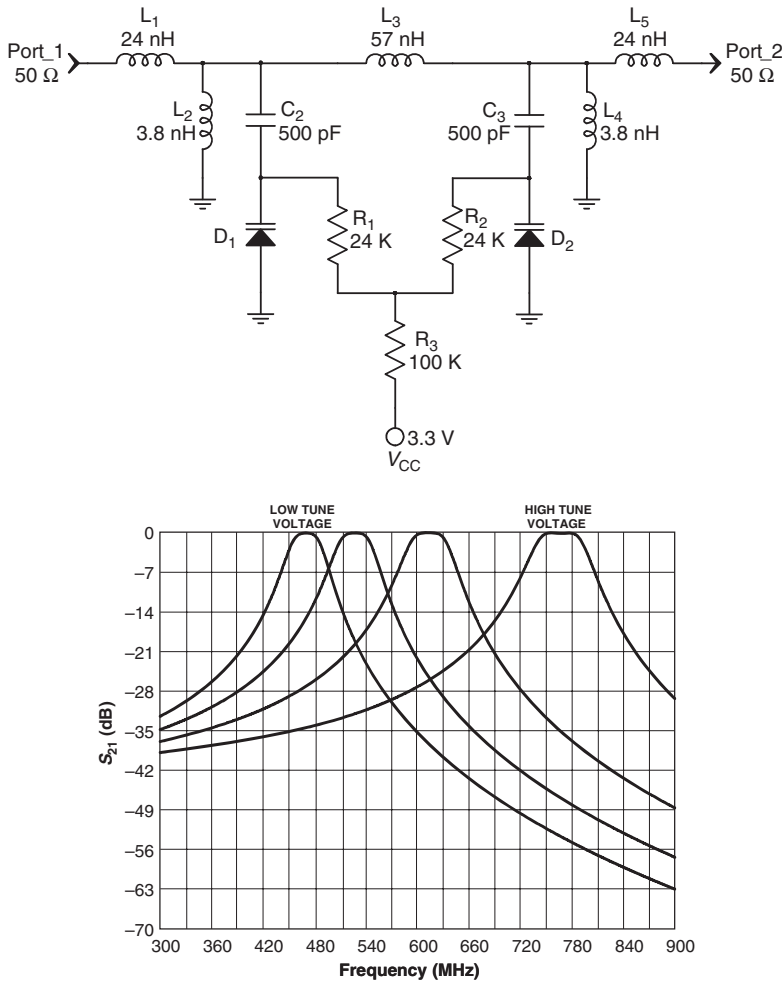


Figure 6.83 A tunable filter example for tuning across 470 to 770 MHz, with its frequency response (Note: all ideal components).

6.8 Filter Issues and Terminology

6.8.1 Introduction

Filters will have various parasitic, matching, shielding, stopband, and reentrance issues, as well as a new terminology.

6.8.2 Filter Issues

There are true bandpass filter structures that can transform the different impedances between two ports within their passband, while also maintaining their full bandpass filter characteristics. Because its passband is naturally limited in frequency at both the low and

the high end, a bandpass can perform both matching and filtering jobs. But lowpass filters will not be able to function as both a true lowpass and a true matching network and still maintain their full performance characteristics all the way from its f_r down to DC. This is not to say that the lowpass filter cannot both filter undesired frequencies and match two dissimilar impedances, but because the lowpass filter's series inductors and shunt capacitors will completely disappear at DC, the impedance at 0 Hz will be equal to that of a *completely* unmatched network. Therefore, there will not be an actual match between the two mismatched ports at 0 Hz. Nevertheless, this does supply us with a resonate peak for an adequate match and filtering of limited bandwidth, which is limited because the lowpass structure itself cannot have a low enough Q between two dissimilar impedances for a full 0 Hz and up-frequency response.

When modeling any filter type in software using ideal models, the performance will always be superior to that of the real-life circuit, since parasitics are not included in the component's themselves, nor in the PCB's pads and traces. These parasitics will begin to severely alter the filter's ideal simulated performance as the frequencies increase. In fact, these real-life component and board parasitics, as well as circuit size-versus-frequency issues, are the frequency-limiting factors to lumped circuit design. Thus, for an accurate and realistic design, the appropriate passive models and microstrip must be added to the filter's RF software simulation.

If RF shielding is employed for the filter structure it is critical to solder it fully to the PCB's groundplane, and to do so completely around the shield's entire periphery. If not, the stopband suppression, especially on the high side, can be severely degraded as the RF blows past the filter circuit itself (Fig. 6.84).

Very large or very small-value inductors or capacitors in a filter design usually mean that the circuit will not be easily realizable when placed on a real PCB, using real components. This is due to the higher internal parasitics inherent in larger-value components, which can also cause the part itself to be too close to its SRF. And if the

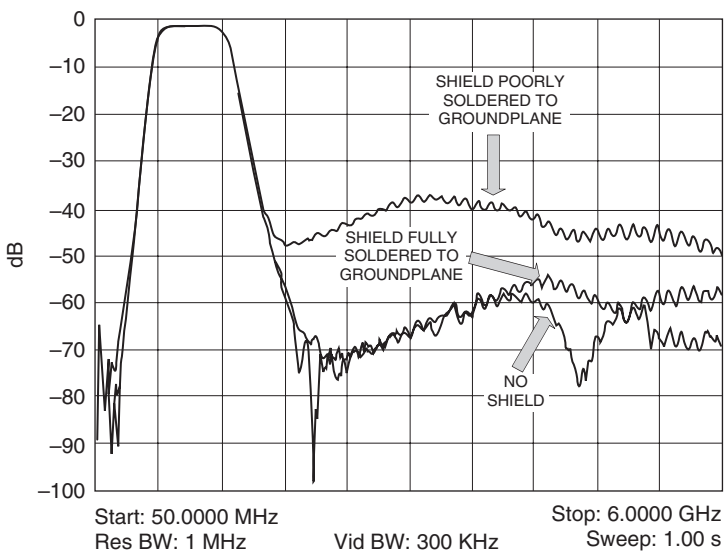


Figure 6.84 An example of the effects RF shields have on a filter's stopband performance.

component is too small in value the external parasitics, caused by PCB's pad and trace reactances, will overwhelm the part's own small capacitance or inductance value. Therefore, either a too large or a too small-valued capacitor or inductor in a filter's circuit network can cause a significant shift in the component's rated value, and is consequently highly undesirable.

All filters, at one frequency or another, will have certain reentrant modes (a decrease in stopband rejection) so all such circuits should be checked to make sure this will not devastate the filter's expected stopband characteristics. In some cases, a band-reject open stub may have to be placed on the PCB to improve rejection at one or more of these frequencies.

Any filter placed at the output of a nonlinear power amplifier to suppress its output harmonics may not be capable of supplying the rated full rejection it was designed for. This is because the filter will now not see a pure and wideband 50-Ω impedance at both its input and output ports, and is caused by the power amplifier being tuned mainly for power output, and not conjugately matched for gain and return loss. Further, the filter's output port is being forced to interface with the system's portable or mobile antenna which, due to the changing coupling effects of the antenna with the outside world, will have a variable input impedance as well. Both of these real-life effects will decrease the filter's stopband attenuation capabilities, sometimes significantly.

6.8.3 Filter Terminology

There are numerous terms used when filters are discussed. Below are the most common:

Absolute attenuation—The maximum attenuation a filter is capable of at some chosen frequency in its stopband, measured in dB.

Bandwidth—The width (f_{LOW} - f_{HIGH}) of the band of frequencies passed by a bandpass filter at its 3-dB down points, measured in Hz.

Center frequency (f_C or f_O)—The exact mathematical center of a bandpass filter, measured in Hz.

Cutoff frequency—The point in a frequency response of a filter that is 3 dB below the average passband response, and which keeps on falling, measured in Hz.

Decibels of attenuation per octave—Filters can be designed as to how rapid their frequency skirt slope falls. The decibels of attenuation per octave specification refers to a filter's steepness: If a filter is said to have a 15-dB/octave slope at a 1-GHz cutoff frequency, then the attenuation within the stopband will be 15 dB more at 2 GHz, while the filter's stopband attenuation at 4 GHz will be 30 dB. Knowing the required dB/octave fall of the filter's skirt assists the engineer in visualizing the attenuation as the frequency increases, or decreases, from its passband. Measured in dB/octave.

Group delay—The measurement of the time retardation created by a filter or circuit to any discrete signal that passes through it. Measured in nano seconds.

Group delay variation—When the group delay varies across the passband significantly, as it does in tight Chebyshev filters (and to a certain extent in Butterworth filters), this will cause increased BER in digital radios. This variation in the group delay, called GDV, is especially severe at the edges of a filter's

bandpass, so any filter that has steep skirts and/or a high pole count will have a high GDV. The problem can be lessened by widening the filter's passband beyond what is required by the signal, by using fewer filter poles, or by choosing a Butterworth type. Measured in nano seconds.

Insertion loss—The attenuation through a filter within the center of its passband when terminated with its design impedance. Measured in dB.

Insertion loss linearity—The variations in insertion loss with input power. Primarily a problem with crystal filters. Measured in dB.

Passband—The band of frequencies from f_{LOW} to f_{HIGH} that a filter passes with little attenuation, and is normally measured at its 3 dB-down points. Measured in Hz.

Passband ripple—The amplitude fluctuations within a filter's passband. A ripple greater than 0.5 dB is usually considered unacceptable in digitally modulated radios. Chebyshev is the dominant filter topology that contains ripple within its passband; however, this ripple can be decreased to 0.1 dB or less by low-ripple designs. Measured in dB.

Phase shift—The measurement of the variation of the phase of a signal as it moves through a filter from its input to its output. Measured in degrees.

Poles—Refers to the number of reactive components, inductors or capacitors, in a lowpass and highpass filter; or the number of reactive pairs in a bandpass filter (for an *all-pole* filter). The filter's order matches its poles in an all-pole filter, and the number of poles governs the steepness of the filter's skirts. Dimensionless.

Quality factor (*Q*)—The ratio between the center frequency to the bandwidth at a filter's 3-dB down points. The narrower the bandwidth for the same center frequency equates to a higher filter *Q*. Also refers to the quality factor of the individual components that make up the filter, called unloaded *Q*. This is particularly important for inductors within the LC filter circuit, since the lower the individual *Q*s are within each component, the higher will be the filter's insertion losses, the more degraded the filter's stopband attenuation characteristics, and the less sharp the filter's frequency responses at their band edges. Dimensionless.

Return loss—The measurement of the difference between the signal power sent toward a filter's input, and the strength of the RF signal power returned (reflected) from the input back toward the source. Most filters can easily be designed with a return loss of 10 dB or higher (with this value of return loss, only $\frac{1}{10}$ of the power incident to the filter's input will *not* be passed on to the load, but will be reflected back toward the original source). Measured in dB.

Ripple—The amount of amplitude variations in the passband of a filter. Excessive ripple causes high BER in digital systems. Measured in dB.

Ripple loss—The distinction between the maximum and the minimum attenuation inside the filter's passband. Measured in dB.

Shape factor—Specifies the steepness of the skirts in a filter. A perfectly square shape factor (SF) would equal 1. Measured as a ratio of the 60-dB bandwidth to the 3-dB bandwidth, or BW(60 dB)/BW(3 dB).

Spurious responses—Since the perfect component does not exist, any filter made of capacitors, inductors, and/or crystals can have areas within its stopband that

supplies less attenuation than planned, or creates ripple within the passband. This is caused by undesired reactances from the component's own end terminations, its internal parasitics, and stray PCB reactances, all of which will resonate at various frequencies. Crystal filters, especially those with the AT-cut, will have these responses at roughly odd integer multiples of the fundamental frequency, which can create unexpected consequences in the design of the communications system if not expected. Distributed microwave filters can also have *reentrant* modes, which permit a multitude of secondary passbands within the filter's expected stopband. These spurious responses can, if required, be lessened by supplemental LC filtering at the output of the filter. Measured in dBc and dBm.

Stopband (*reject band*)—The band of frequencies that a filter attenuates to some predetermined level, such as 60 dB (but can be much less). The stopband is the point on the attenuated side of the cutoff frequency. Measured in Hz.

CHAPTER 7
Mixer Design

Mixers are three port active or passive devices, are designed to yield both a sum and a difference frequency at a single output port when two distinct input frequencies are inserted into the other two ports. This process, called *frequency conversion* (or *heterodyning*), is found in most communication's gear, and is used so that we may increase or decrease a signal's frequency. One of the two input frequencies will normally be a CW wave, produced within the radio by a local oscillator (LO), while the other input will be the RF signal received from the antenna.

If we would like to produce an output frequency within the mixer circuit that is lower than the input RF signal, then this is called *down conversion*; if we would like to produce an output signal that is at a higher frequency than the input signal, it is referred to as *up conversion*. Indeed, most AM, SSB, and digital transmitters require mixers to convert up to a higher frequency for transmission into space, while superheterodyne receivers require a mixer to convert a received signal to a much lower frequency. This lower received frequency available at the mixer's output port is called the *intermediate frequency* (IF). Receivers use this lower-frequency IF signal because it is much easier to efficiently amplify and filter with all the IF stages tuned and optimized for a single, low band of frequencies, which increases the receiver's gain and selectivity.

Again, the frequency conversion process within the nonlinear mixer stage produces the intermediate frequency by the RF input signal heterodyning, or beating, with the receiver's own internal LO. This heterodyning mixer circuit will consist of either a diode, BJT, or FET that is overdriven, or biased to run within the nonlinear area of its operation. However, the beating of the mixer's RF and LO input signals yields not only the RF, the LO, and the sum and difference frequencies of these two primary signals, but also many spurious frequencies at the mixer's output port. Most of these undesired frequencies will be filtered out within the receiver's IF stages, resulting in the new desired signal frequency, consisting of the converted carrier and any sidebands, now at the *difference* frequency. This new, lower difference frequency will then be amplified and further filtered as it passes through the fixed-tuned IF strip.

There are three basic classifications for both active and passive mixers: *Unbalanced* mixers have an IF output consisting of f_S, f_{LO}, $f_S - f_{LO}$, $f_S + f_{LO}$, and other spurious outputs. They will also exhibit little isolation between each of the mixer's three ports, resulting in undesired signal interactions and feedthroughs to another port. *Single-balanced* mixers will at least strongly attenuate either the original input signal *or* the LO (but not both), while sending less of the above mixing products on to its output than the unbalanced type. A *double-balanced* mixer, or *DBM* for short, supplies superior IF-RF-LO inter-port isolation, while outputting only the sum and difference frequencies of the input signal and the local oscillator, while attenuating both the LO and RF

signals, and significantly attenuating three quarters of the possible mixer spurs at the output of the IF port. This makes the job of filtering and selecting a frequency plan a much easier task.

7.1 Passive Mixers

7.1.1 Introduction

Passive mixers permit a much higher amplitude RF input signal level than active mixers before severe distortion products within the output IF becomes unacceptable. These distortion products are in the form of *intermodulation distortion* (IMD), along with *compression distortion*. The IMDs may fall in band, or cause other signals to fall in band, possibly swamping out or creating interference to the baseband signal. This causes additional noise, which will degrade system performance and BER.

Most passive mixers also possess a lower noise figure than active mixers, which is very important for any stage within the front end of a low-noise receiver. But instead of an insertion gain as many active mixers will enjoy, passive mixers will have an insertion loss of around 7 dB.

The passive mixer conversion losses are caused by the mixing diode's internal resistance, port impedance mismatches, mixer product generation, and the inevitable 3 dB that is wasted in the undesired sum or difference frequency. (This sum or difference frequency is removed by filtering, cutting the mixer's final output power in half.)

Figure 7.1 shows a common double-balanced mixer, which utilizes a *diode ring* to achieve frequency conversion of the RF input signal. The mixer's diodes are being constantly switched *on* and *off* within the ring by the high-powered LO stage, while the RF signal is alternately sent through the diodes, mixing the two signals in a nonlinear manner, producing the IF output frequency. DBMs commonly function up to 8 GHz and beyond by using *hot-carrier* (*Schottky*) diodes, which possess low-noise and high-conversion efficiency.

DBMs made of lumped components and placed on the wireless devices' PC board as a discrete circuit are seldom utilized today. Instead, double-balanced mixers are available in a module, with the diodes and transformers already balanced and placed within a surface-mount package.

Lower performing passive mixers that are not double balanced are available that employ either a single diode or double diodes (Fig. 7.2). Unlike DBMs, they are cheap, require few components, and are relatively easy to design.

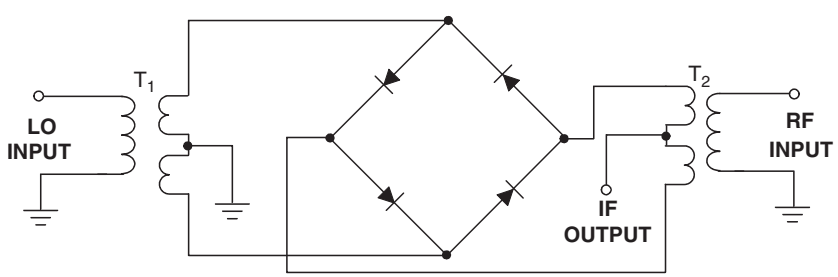


Figure 7.1 A double-balanced mixer stage.

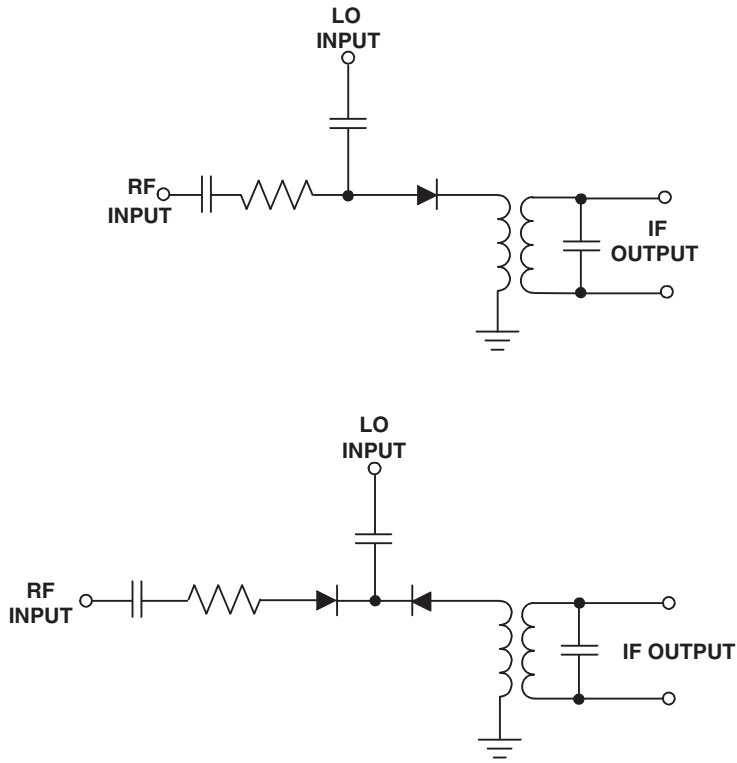


Figure 7.2 A single-diode mixer and a double-diode mixer stage.

7.1.2 Passive Mixers Types

There are several types of passive mixer designs available, depending on cost and performance levels required. Some of these passive diode mixers have already been introduced above, but will be further investigated in this section.

Figure 7.3 shows a one diode, *single-ended* mixer. This circuit is only found in very low-cost circuits, with the isolation between ports being supplied by bandpass and lowpass filters that are separated in frequency. Some of these single-ended mixer can also take advantage of a somewhat lower level of LO power needed to drive the single-diode mixing element, as compared to the often times higher drive levels required of a high-performance DBM stage. The single-ended mixer, however, has a relatively narrow

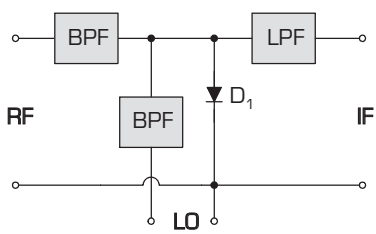


Figure 7.3 A basic single-diode mixer circuit with filtering.

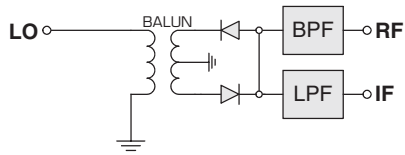


Figure 7.4 A single-balanced two-diode mixer.

bandwidth, poor port-to-port isolation, a low intercept point, and inferior intermodulation distortion suppression (their $IP3_{OUT}$ value will normally equal their LO input drive level). If we would like to increase the specifications and overall quality of this device, we will need to increase the number of diodes. This will permit a higher amplitude LO drive level input, which automatically forces an increase in the mixer's P1dB compression point. (The P1dB is normally specified at about 10 dB below the LO for all diode mixers. So, the higher the LO drive that can be inserted into a mixer, the higher the P1dB.) As we are demanding a more powerful LO for increased mixer performance, this will unfortunately not only cost more, but also radiate a higher level of EMI.

Single-balanced mixers, as shown in Fig. 7.4, are comprised of two matching diodes, a balun, and generally two filters. The balun converts the unbalanced LO output to a balanced mixer input, matches the diodes to the port's impedance, helps in port-to-port isolation, and balances the diodes. The filters, one at each of the RF and IF ports, are to improve mixer isolation.

This particular mixer type will balance out (cancel) and filter the LO power, preventing excessive LO feedthrough at the RF and IF ports. In fact, single-*balanced* mixers are superior to single-ended mixers in this LO-to-IF and LO-to-RF isolation, as well as in their wider bandwidth operation. Furthermore, intermodulation distortion suppression is increased over the single-ended type. This is because any distortion products that are made up of even harmonics will be suppressed by the balanced-circuit action and, since twice as many diodes are typically used with this circuit, along with higher LO power, the same RF amplitude levels that are inserted into the single-balanced mixer's input will create less IMDs to be generated in the first place. As compared to a single-ended mixer, the negative attributes of a single-balanced mixer would be that the LO power must often be somewhat higher (which necessitates a more expensive and power hungry oscillator), and the part's count is increased (since a perfectly balanced balun and one more matched diode must be used).

Single-balanced mixers are so named due to their single-balanced balun, while double-balanced mixers are so named for the same reason; they employ two baluns. To any significant extent, quality double-balanced mixers will output only IMD products that are constructed of both odd RF and odd LO harmonics. This action decreases the DBM's total output of mixer products to a quarter of the amount generated within any simple mixer. However, mixer products are suppressed to varying levels, strongly dependent on the quality of the diode match and the accuracy of the balun balance. So, while a DBM often may require twice the LO power as a single-balanced mixer, as well as double the number of internal-balanced diodes and baluns, a DBM will have much better IMD suppression, a wider bandwidth, and a higher intercept point.

Triple-balanced mixers (TBM, or DDM for *double double-balanced mixers*) have baluns located at all three ports, along with two complete diode rings. They have increased intercept points for decreased mixer product generation and two-tone intermodulation distortion levels, as well as better port-to-port isolation and a wider possible IF

bandwidth output. However, TBMs need higher LO power, another matched diode ring, and one more balanced balun above that demanded by the DBM type. The price will be higher.

7.1.3 Passive Mixer Design

This section provides several methods to design straightforward, but highly useful and effective, RF mixers for many low-cost high-frequency consumer wireless products.

As with any practical design calculation at these frequencies, parasitic capacitance and inductance, as well as component and PCB pad/trace length and distributed reactances, will modify ideal circuit behavior, sometimes extremely, so accurately modeling and tuning the mixer will always be necessary.

For both cost and performance reasons, design and construction of DBMs themselves is generally best left to the specialized mixer manufacturing companies. Therefore, when utilizing DBMs in a wireless design, it is easier and faster to simply purchase a completed, off-the-shelf device.

A Passive RF Diode Single-Ended Mixer (Fig. 7.5)

A huge limitation of unbalanced mixers in general is excessive LO radiation from the RF ports, limited only by any possible RF filtering. Another issue is the close-in RF feedthrough into the IF, with the stronger, but typically more distant, LO feedthrough being more easily filtered out from the IF. In fact, we could add further filtering to all mixer ports, but this would substantially add to cost and complexity, which is exactly what we are trying to avoid with the use of a single-ended mixer in the first place.

To Design

1. Select an appropriate diode for the frequencies of interest. This will normally be a silicon, low-barrier Schottky type for RF use.
2. Design L_4 and C_5 for series resonance at the IF frequency, and with a high L to C ratio for a narrow passband.
3. Design L_5 and C_6 for parallel resonance at the IF frequency, and with a high C to L ratio for a narrow passband.
4. Design a 50-Ω lowpass RF filter to attenuate the high level of LO radiation from the mixer's RF port (to decrease this LO radiation even further, additional front-end system-level RF filtering is assumed).

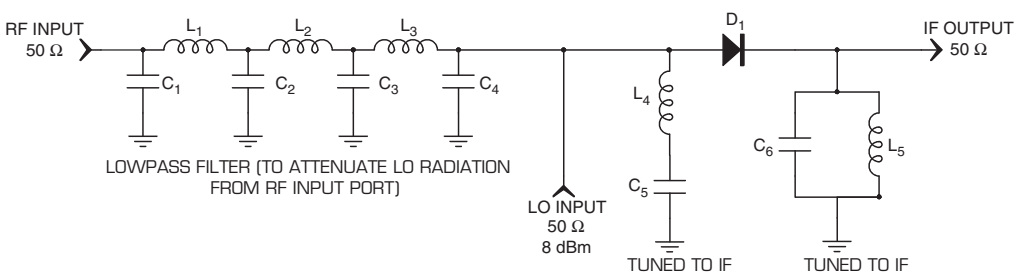


Figure 7.5 A practical single-ended diode mixer circuit.

5. In such low-cost applications, mixer port matching to the diode is rarely used. In fact, for the high LO power levels that are commonly used for this topology, the impedance will be close enough to 50 Ω to supply a decent match for the input LPF. (Increasing the LO drive level would further decrease the mixer diode's input impedance, while decreasing the drive level will increase the input impedance.)
6. Insert an 8-dBm LO signal, which will supply the mixer with an RF-to-IF conversion loss of 10-dB.

A Quick Example Design a Passive Single-Ended RF Diode Mixer (Fig. 7.6)

Goal: Create a single-diode RF mixer. The specifications and parameters for the circuit are:

f_{RF} = 60 MHz
f_{IF} = 40 MHz
f_{LO} = 100 MHz
P_{LO} = +8 dBm

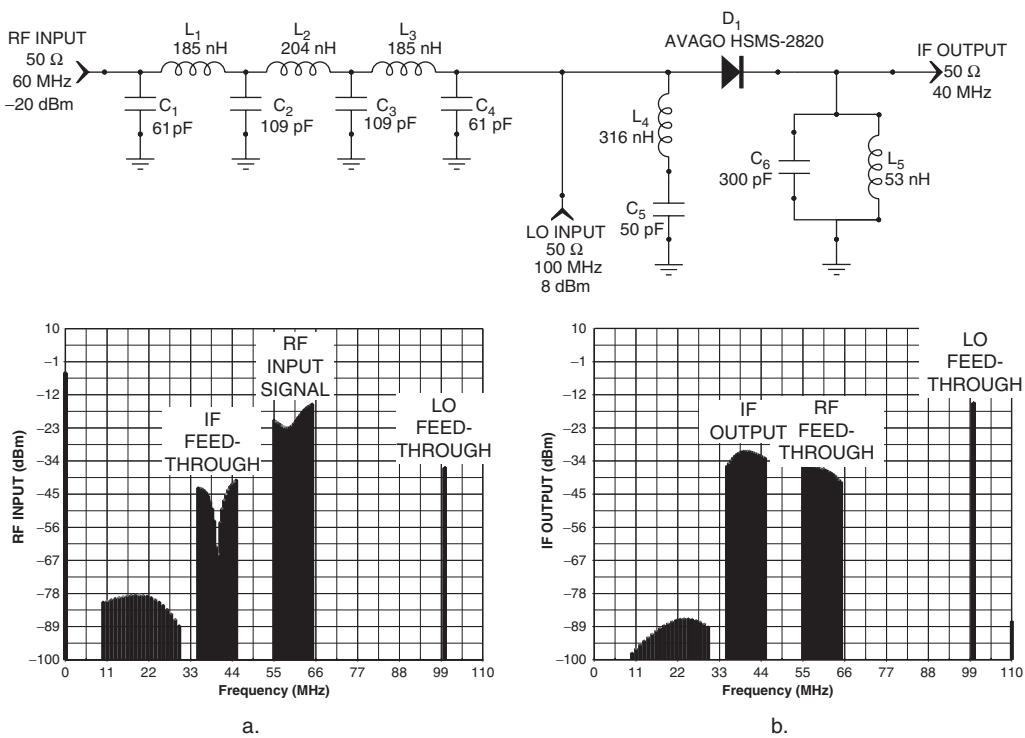


Figure 7.6 The example single-ended diode mixer circuit with part's values, along with harmonic balance frequency sweep results: (*a*) All the signals present at the RF input port; (*b*) All the signals present at the IF output port.

P_{OUT} = −5 dBm

Conversion loss = 10 dB

Diode = Avago HSMS-2820 (Schottky type, appropriate for frequency of interest)

Solution:

1. L_4 = 316 nH, C_5 = 50 pF
2. L_5 = 53 nH, C_6 = 300 pF
3. LPF component values as shown for a seven-pole Chebyshev with a $f_{r(3\ dB)}$ of 62 MHz

Distributed Diode Single-Balanced Narrowband Hybrid Mixer for Microwave Frequencies (Fig. 7.7)

This mixer structure will have decent dynamic range, and requires approximately 8 to 10-dBm LO power, with satisfactory RF/LO-to-IF and LO-to-RF isolation for most applications. It enjoys very good IMD performance, with fair cancellation of even harmonic signals. However, the IF must be no higher in frequency than 50 MHz or so, since the difference between the LO and RF frequencies must be relatively small due to the mixer's resonant distributed design, which has to be able to effectively react to *both* the RF and the LO frequencies.

For maximum LO rejection, design each microstrip section for the LO's output frequency, or for midway between the LO and RF frequency values. The mixer's conversion loss will be approximately −6 dBm.

To Design

1. A = 90° long at LO frequency or at $(f_{RF} + f_{LO})/2$, using 50-Ω microstrip.
2. B = 90° long at LO frequency or at $(f_{RF} + f_{LO})/2$, using 35.5-Ω microstrip.
3. C = 90° long at RF frequency, using 50-Ω microstrip. (C shorts RF to ground. Bends do not affect actual length, but are used for compactness.)
4. D = 90° long at LO frequency, using 50-Ω microstrip. (D shorts LO to ground. Bends do not affect actual length, but are used for compactness.)
5. E = 50-Ω microstrip, with the two traces before the D_1 and D_2 diodes being of equal length.
6. RFC = 90° long at LO frequency, using 100-Ω microstrip.
7. D_1 and D_2 = select the appropriate Schottky diodes for the frequency of operation and the application.

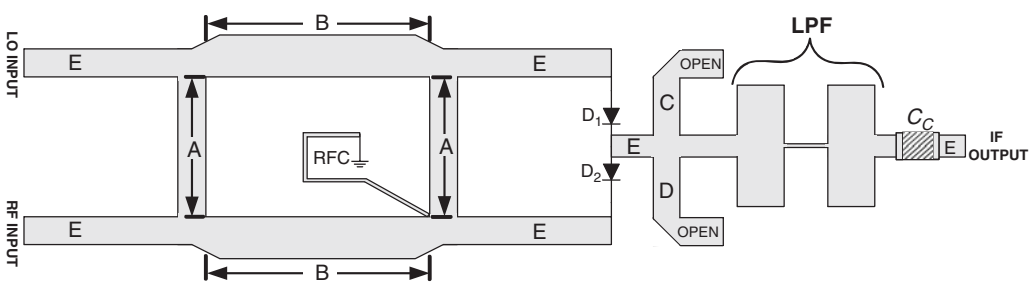


Figure 7.7 A narrowband microwave mixer for UHF and above applications.

NOTE: *The* B *section is extremely sensitive to exact dimensions for LO suppression at RF port. The LPF is used to attenuate all frequencies above the IF.*

A Quick Example Design a Passive Single-Balanced RF Diode Mixer (Fig. 7.8)

Goal: Create a single-balanced diode RF mixer. The specifications and parameters for the circuit are:

f_{RF} = 5.8 GHz
f_{IF} = 40 MHz
f_{LO} = 5.76 GHz
P_{LO} = +5 dBm
P_{OUT} = −5 dBm
Conversion loss = −6 dB

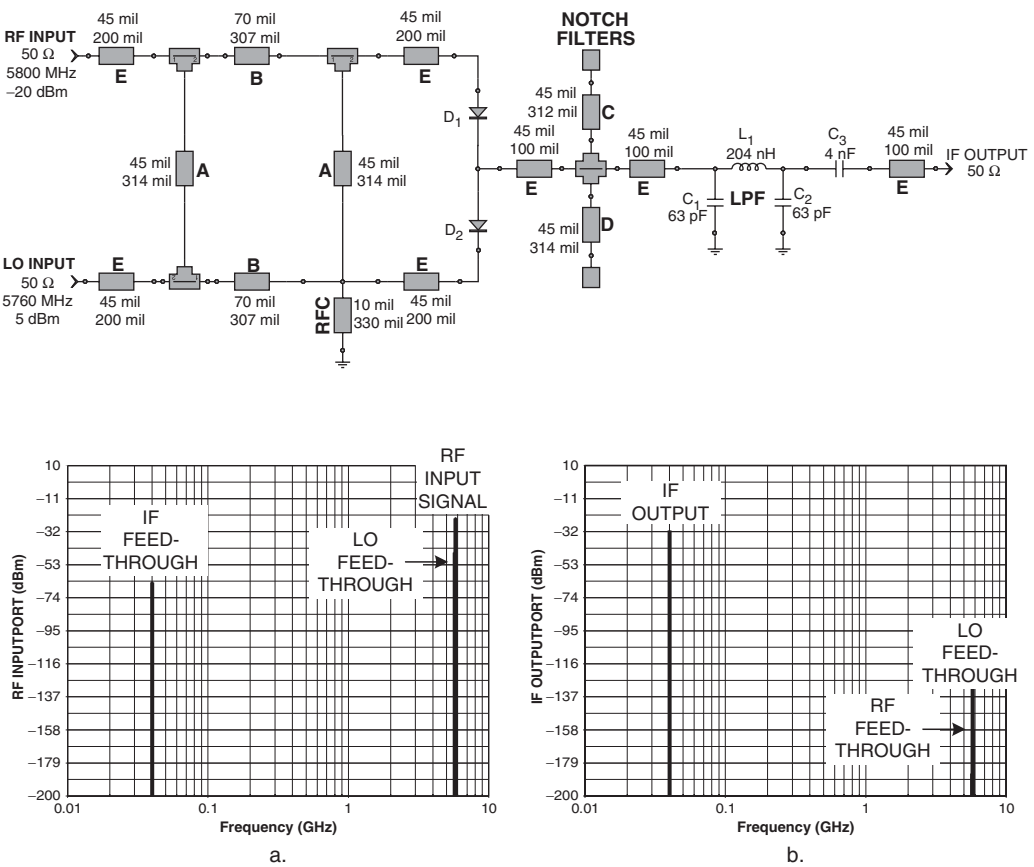


FIGURE 7.8 The example single-balanced diode mixer circuit with calculated part's values, along with harmonic balance frequency sweep results (log scale): (*a*) All the signals present at the RF input port; (*b*) All the signals present at the IF output port.

Diodes = Avago HSMS-2820 (Schottky type, appropriate for frequency of interest)

Substrate = Roger's RO-4003, 20-mils thick

Solution:

1. A = 45-mils wide by 314-mils long
2. B = 76-mils wide by 306-mils long (to decrease LO feedthrough to RF port, B section optimized in simulator slightly to 70-mils wide by 307-mils long)
3. C = 45-mils wide by 312-mils long
4. D = 45-mils wide by 314-mils long
5. E = 45-mils wide by variable length
6. RFC = 10-mils wide by 330-mils long
7. LPF = C_1, C_2 = 63 pF; L_1 = 204 nH (cutoff frequency = 42 MHz)
8. C_3 = 4 nF (coupling capacitor for 40 MHz IF)

Microwave Circular Rat Race Single-Balanced Diode Mixer (Fig. 7.9)

Very similar to the above mixer, this low-cost, distributed microwave diode mixer, when used with Schottky diodes, is an excellent choice for very high frequencies. It requires an 8-dBm LO drive level, has decent intermodulation performance, good RF/LO-to-IF and LO-to-RF port isolation and, as with all diode mixers, will have a conversion loss (in this case, −6 dB).

The IF must be no higher in frequency than 50 MHz or so, since the difference between the LO and RF frequencies must be relatively small due to the mixer's resonant distributed design; which has to be able to effectively react to *both* the RF and the LO frequencies.

To Design

1. $f_m = \dfrac{f_{RF} + f_{LO}}{2}$ (for midway between the RF and LO frequencies)
2. A = 90° long at f_m, using 70.7-Ω microstrip
3. B = 270° long at f_m, using 70.7-Ω microstrip
4. C = 10° long at f_m, using 50-Ω microstrip
5. D_1 and D_2 = select the appropriate Schottky diodes for the frequency of operation and the application

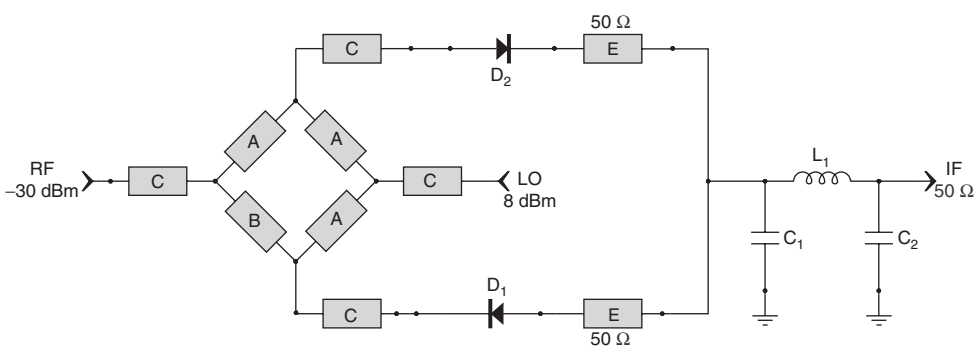


Figure 7.9 The rat race distributed diode mixer circuit.

6. LPF = in the included *AADE Filter Designer* software, design a 50-Ω lumped lowpass filter with 1.5 × IF frequency cutoff (a shunt capacitor must begin the filter for the LO and RF currents). The LPF may also be designed as a distributed network, or by using a small multilayer ceramic filter.
7. *E* = short 50-Ω microstrip

7.1.4 Passive Mixer Distortion

The undesired mixer product frequency generation, and its suppression, is important in the entire heterodyning process. Output mixer products (Fig. 7.10) are formed by the mixing in the nonlinear diode elements of the incoming single-tone RF (and its own resultant harmonics), with the single-tone LO (and its resultant harmonics). This creates high-order distortion products that are higher and lower in frequency than the desired product, with this desired product normally being the difference frequency of the LO and RF in a receiver, or the sum of the LO and IF in a transmitter.

Two-tone intermodulation products are created when two tones (f_1 and f_2) are placed at the RF input port of the receiver's mixer and, when mixed with each other and the LO, give birth to high-order in-band spurious responses at the IF output port of the mixer. While keeping in mind that the higher the possible LO oscillator power, the lower the distortion products, Fig. 7.11 demonstrates this point with three different level mixers, a *Level 7*, *Level 17*, and a *Level 23*), with each using its recommended LO input power of either 7, 17, or 23 dBm. The Level 7 mixer's IF output shows high third-, fifth-, and seventh-order two-tone IMD products for a 0-dBm RF input. The Level 17 mixer decreases these IMD products for the same 0-dBm RF input amplitude. The Level 23 mixer shows IMD products much further down than even the Level 17 mixer, at approximately 65 dBc.

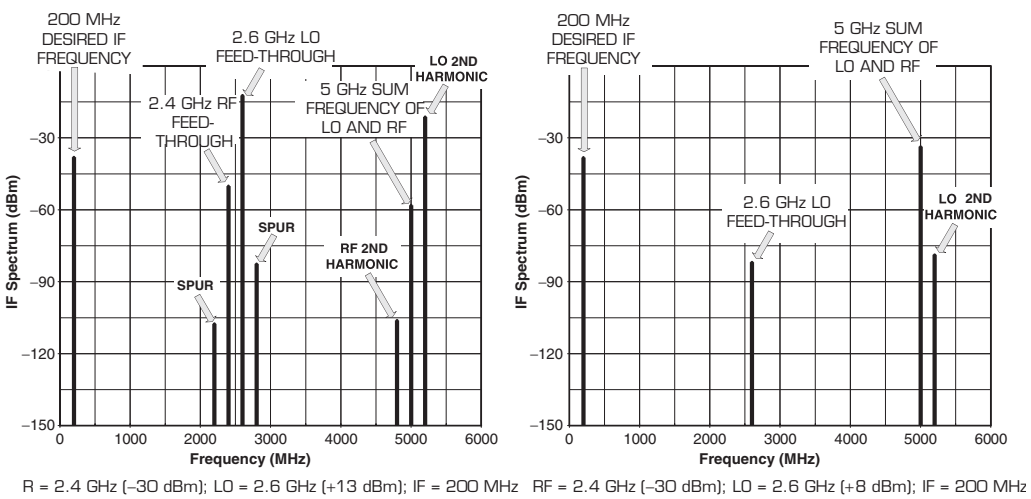


Figure 7.10 Various single-tone mixer spurs generated by the single-ended mixing of the RF (2.4 GHz at −30 dBm) and the LO (2.6 GHz at +13 dBm), and their harmonics in an unbalanced mixer, with light one-pole filtering (left); a DBM diode mixer's IF output spectrum in a well-balanced circuit (right).

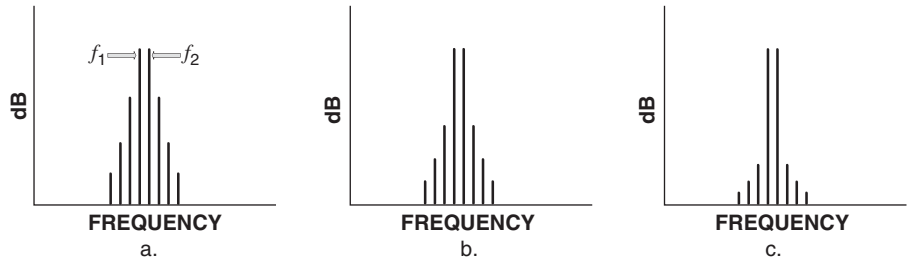


Figure 7.11 Different level mixers and their two-tone output spectra for a DBM: (*a*) Level 7, (*b*) Level 17, (*c*) Level 23.

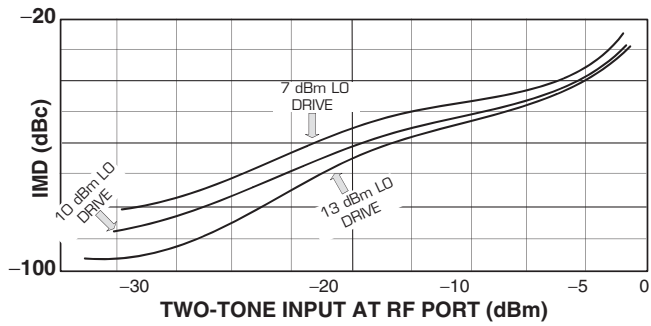


Figure 7.12 A Level 7 DBM mixer's IMD generation at various LO levels and two-tone input powers.

A further subject in mixer design is demonstrated in Fig. 7.12: Boosting a Level 7 DBM's LO drive does not in itself drastically improve the IMD product suppression. This can only be accomplished, in any significant way, by increasing the number of diodes in each mixer leg from one to two in series (as well as other techniques), and then increasing the LO drive, with the then resultant improvement in IMD suppression.

The intercept point indicates the mixer's capability to suppress intermodulation distortion, typically referring to two-tone third-order intermodulation products. A high intercept point decreases the undesirable generation of these IMDs. But in the world of DBMs, the intercept point and the 1-dB compression point do not directly correlate to each other; so choosing a mixer simply for its high 1-dB compression point as a guarantee of increased two-tone suppression could prove a poor choice.

As stated above, two-tone third-order products can be reduced by increasing the *level* of the mixer (and therefore the LO drive level itself), and/or by decreasing the power of the input two-tone RF signal. But since we will generally use the manufacturer's recommended LO drive for our chosen level of mixer, which will be selected in consideration of the maximum LO power available within our own wireless design, as well as cost constraints, then decreasing the input RF level is normally the easiest and cheapest solution for two-tone third-order product improvement.

By decreasing the input two-tone RF signal by 1 dB, we will decrease the output two-tone third-order products by 3 dB. However, the inverse is also true: increasing the RF two-tone input by 1 dB will increase the output two-tone third-order products by

3 dB. As stated, because they can fall in band at the mixer's IF port, these third-order products are the most dangerous spurious signals, and must be attenuated to the lowest level the system requires.

To assure ourselves of decent intermodulation distortion performance and conversion-loss variations, a Level 7 mixer should never be run with an RF input higher than −3 dBm (with the LO drive at the rated power level), while a Level 10 mixer should never go above a 0-dBm RF, and a Level 13 mixer never above +3 dBm, or a Level 17 mixer never higher than +7 dBm. In fact, decreasing these RF input levels to 20 dB below the LO drive is commonly adopted to reduce IMD generation to extremely low amplitudes. (This will, unfortunately, also increase the *relative* LO feedthrough level.)

Within a mixer, it is possible to calculate the highest in amplitude two-tone third-order spur level that is down from our desired RF signal by:

$$TOIM_{SUP} = 2(TOIP - RF_{IN})$$

where $TOIM_{SUP}$ = third-order intermodulation suppression down from the signal of interest at the mixer's output port, dBc
$TOIP$ = third-order (input) intercept point of the mixer, dBm
RF_{IN} = power of the RF signal at the input to the mixer, dBm

When viewing the mixer's *third-order intercept point* at its input port (TOIP or $TOIP_I$) specification on a data sheet, we may sometimes be required to convert from this input intercept to an output intercept ($TOIP_O$). This can be accomplished by:

$$TOIP_O = TOIP_I - CL$$

where $TOIP_O$ = mixer output intercept point, dBm
$TOIP_I$ = mixer input intercept point, dBm
CL = mixer conversion loss (usually 6 to 9 dB, as a positive number), +dB

7.1.5 Passive Mixer Issues

No software or mathematical calculation can design (or *synthesize*) a microwave circuit with 100% accuracy—especially nonlinear mixer networks—due to the practical inability for the program or the formula itself to take the PCB's trace/pad lengths and parasitics, the lumped component's distributed effects, and the nonlinear active models all into account. This limits the accuracy of the synthesized RF design that is possible within the calculated results. Thus, without the proper microwave design knowledge *and* the correct use of PCB traces, circuit elements, and device models within the appropriate RF circuit simulator, then no microwave design, no matter how simple or how complex, will operate as expected. Therefore, only RF software tuning, or "tweaking," of the calculated synthesized design within such a simulator will ever obtain even ballpark results, which may still require further bench optimization to complete.

During the system's frequency planning stage, an appropriate LO and IF frequency should be selected that will minimize the number and strength of mixer products present within the IF bandpass of the mixer. This is most conveniently performed by employing the appropriate software, such as Blattenberger's *RF Workbench*, or The Engineer's Club *MixerSpur*. Both of these programs will graphically indicate if there are any dangerous mixer's spurs within the IF's passband.

As all mixers have not only a nominal, but also a minimum and a maximum, LO drive level (as recommended by the mixer's manufacturer), we may sometimes want to select the minimum LO level for two reasons: one, sufficient LO power may not be readily available; and two, LO feedthrough from the RF port must be minimized. However, if we select a decreased level, then the mixer's two-tone IMD suppression, conversion losses, and return losses will all suffer as a result. Conversely, by slightly increasing the drive level above the nominal value, we will obtain a higher mixer noise figure and LO feedthrough, but will somewhat improve the mixer's two-tone IMD performance, mixer product suppression, and conversion losses across the band. Therefore, running the mixer at the recommended LO drive level will be the best compromise for superior overall system performance.

A wideband conjugate match (a good VSWR) at the RF and IF mixer ports is important for the conversion gain and intermodulation performance levels of many DBM type mixers. Therefore, nonreflective filtering is necessary since the undesired signals and products will be reflected back into the IF port of the mixer stage by the reflective stopbands inherent in a common IF bandpass filter, causing two-tone IMD performance to suffer, sometimes by as much as 25 dB. With passive mixers, there are several ways we can combat this effect. We can use a *diplexer* at the IF port of a downconverting mixer stage, which will filter and pass the desired IF but, unlike other filters, will stop the LO harmonics, the sum of the RF and LO, and the IMD products from entering the IF stages. It accomplishes this by the diplexer's ability to *absorb* rather than *reflect* certain undesired frequencies, since the reflection of such undesired frequencies could unbalance the mixer's diode ring, adversely upsetting its dynamic range and conversion loss if high enough in amplitude.

Instead of a diplexer, we may also simply pad the IF output of the mixer so that these reflections are attenuated not only as they enter the pad, but also as they are reflected back into the mixer's IF port. Attenuator pads improve the input/output VSWR by supplying the mixer with an almost pure, and extremely wideband, $50 + j0$ termination. Or we could use a wideband, high-isolation IF amplifier. This will permit all of the mixing products to pass through this amplifier and, after filtering from a normally reflective IF filter, which will bounce much of the undesired signals back toward the sensitive IF port of the mixer, these reflected signals will have been significantly attenuated by the reverse isolation of the buffer amplifier itself. And because the less the power coming from the mixer's IF output the lower in amplitude these reflections will be, and thus the less the requirement will be for their suppression, the exact choice of which of the above methods to use will depend to a certain degree on the output port power.

There may occasionally be a requirement for an *upconverting* superheterodyne receiver when working with HF radios. When designing such receivers, the incoming RF signal should be placed at the passive mixer's IF port, while the now higher frequency IF output signal should be sent out of the mixer's RF port. This is also equally valid for transmitter design, since the mixer will also be performing up conversion.

The selection of the proper double-balanced mixer for a particular receiver or transmitter application will depend on the required P1dB compression point, LO power, port isolation, device cost, and two-tone intermodulation and mixer generated product suppression. Therefore, it is important to study a part's data sheet well before making any final mixer selections.

7.1.6 Passive Mixer Terminology

Some common terminology used to specify a mixer:

Conversion compression—Specification that indicates the maximum value of the input RF signal level that will obtain a linear increase in IF output power. For example, Level 7 DBM mixers will usually have a conversion compression of +2 dBm.

Conversion loss—The rated signal level difference between the input and the output of a mixer at the rated LO input power. For instance, a Level 7 (+7 dBm LO drive) DBM mixer may have a loss in power from the input to the output of 8 dB at midband.

Cross modulation—Describes the undesired transfer of the modulation between a modulated and a CW signal within the mixer stage.

High-side injection—When the LO frequency is higher than the RF frequency in a mixer stage.

Intercept point—Superior two-tone third-order product suppression demands a high mixer intercept point. This value is approximately 10 dB higher at the mixer's input than the *conversion compression* rating discussed above. The *cross modulation* distortion and desensitization are also reduced with a high intercept point.

Interport isolation—The rating of the feedthrough between the mixer's LO, RF, and IF ports. This is the value, in dB, that one port's signal is attenuated at another port's input or output. The most important of these isolation specifications is the LO attenuation at the IF and RF ports, since LO feedthrough is a major problem in receiver and transmitter system's design, and the RF to LO isolation is normally of little concern due to the RF's low input levels. Typical mixer LO to IF isolation will range from 0 to 50 dB, depending on topology and port filtering.

Low-side injection—When the LO frequency is lower than the incoming RF frequency in a mixer stage.

Noise figure (NF)—The noise added by the mixer itself, and equals the difference between the noise at the input of the mixer and the output of the mixer, in dB. When the mixer is driven with the proper LO drive level, the NF will equal the conversion loss.

7.2 Active Mixers

7.2.1 Introduction

Active mixers vary from the passive-type diode mixers described above. Active mixers can supply a conversion gain instead of a loss, they require far less LO drive power, are much less sensitive to port terminations, have better ultimate LO-to-IF isolation, and produce less mixer spurs. However, wider adoption of certain active mixers, such as the *Gilbert cell* type, has been somewhat hindered by a poor IP3, high NF (around 15 dB), and the need for a DC supply voltage. In many high-end wireless applications, the first two problems have limited the active mixer's role to the later stages of a receiver, where the dynamic range of the signal is more under control by the AGC, and the NF matters little.

Many active mixers work by exploiting the high-level signal that is produced by the radio's LO to force the mixer's transistor to operate within its nonlinear region,

functioning only during a 180° of its conduction cycle (similar to a Class C amplifier), while the much smaller in amplitude RF signal operates within the active device's linear region. And even though the mixer is built for nonlinearity, it still acts as a far from perfect, but nondistorting, linear frequency shifter (when not overdriven by the RF input signal), and thus will produce a relatively small amount of IMD products.

7.2.2 Active Mixers Types

As with passive mixers, there are different types of active mixers. The *single-ended FET* mixer of Fig. 7.13 is comprised of a JFET, some biasing components, and two tuned tanks. The RF input signal is dropped across the first tuned input tank and sent to the JFET's gate. An LO signal is inserted into the source lead, with the resultant converted signal removed from the JFET's drain and placed across the tuned output tank. This second tank is tuned to the desired IF output frequency, with most of the mixing, RF signal, and LO frequencies being severely attenuated by this circuit. The secondary circuit of the output transformer takes this signal, reduces the high-output impedance, and places it into the IF amplifiers.

A *dual-gate MOSFET* mixer of the type shown in Fig. 7.14 employs a MOSFET, some biasing components, and a single tuned tank. The RF signal is sent through the coupling capacitor into the second gate while the LO is inserted into the first gate, with the sum and difference frequency, along with the mixer products, being sent on to the tuned circuit. Since this output tank is tuned to exactly the desired IF frequency, all other frequencies are attenuated, while the IF frequency itself is dropped across the

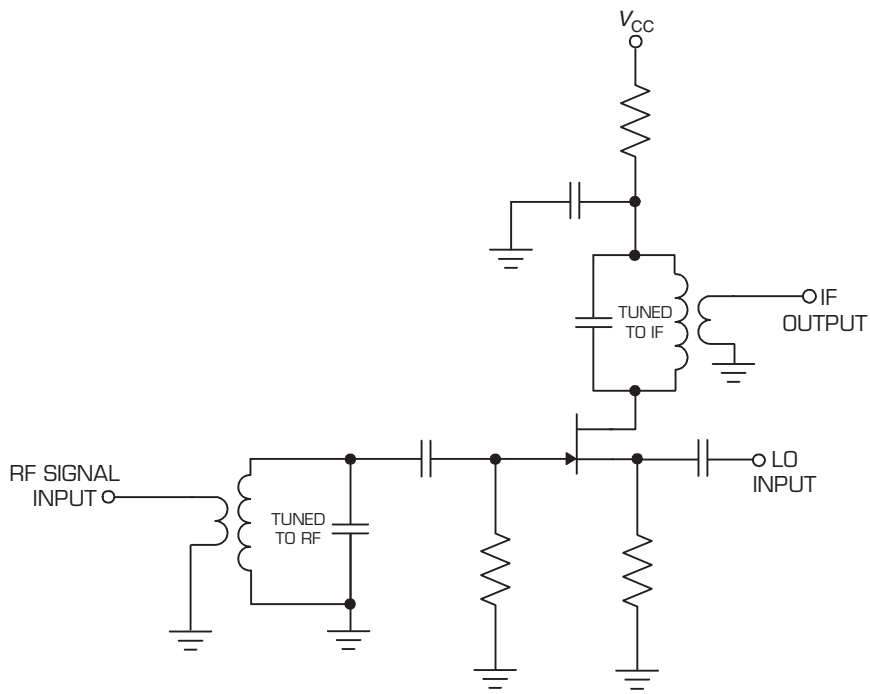


Figure 7.13 A single-ended active JFET mixer circuit.

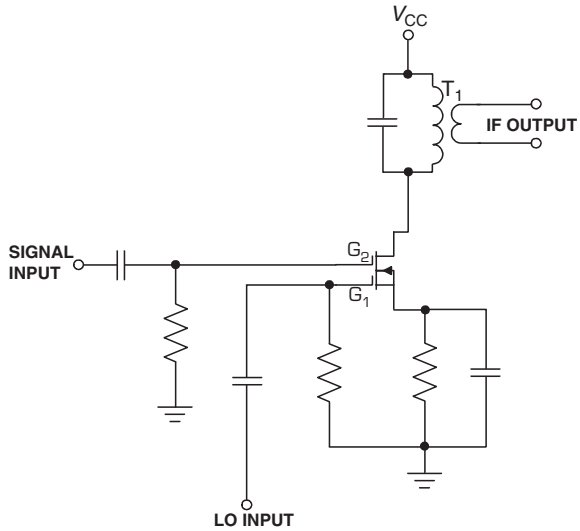


Figure 7.14 A dual-gate MOSFET mixer circuit.

transformer's primary. The IF is then removed from the transformer's secondary and sent on to the IF strip for further amplification and filtering.

Another low-cost active mixer is the *single-ended transistor* type of Fig. 7.15. Both the signal and the LO are inserted into the base and mixed together by the nonlinear Class AB biased transistor. Obviously, unless a diplexer is placed at the input, the RF and LO have no real isolation between their ports. The original RF signal and the LO frequency, as well as all mixing products, are present at the transistor's collector but, due to the

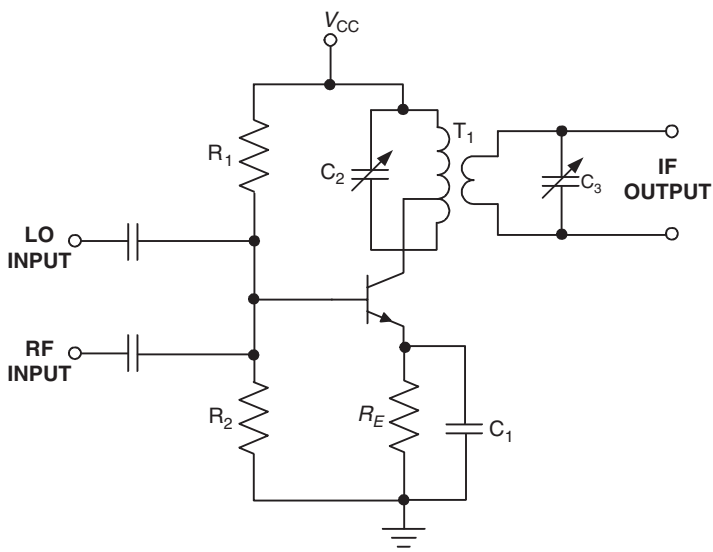


Figure 7.15 One method of using a transistor as a nonlinear mixer stage.

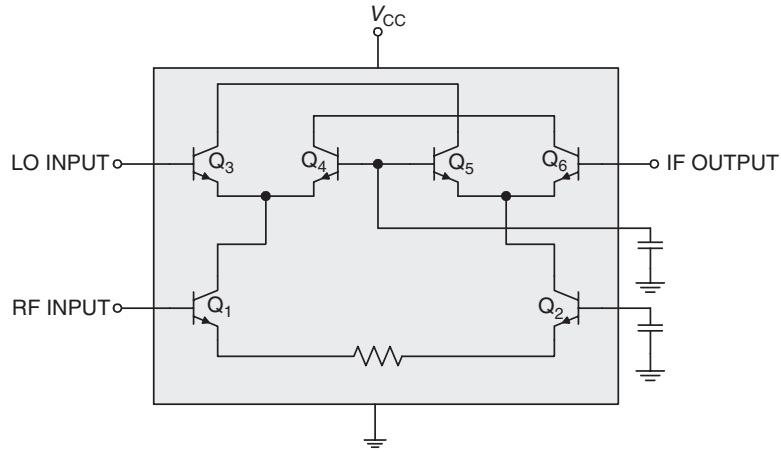


Figure 7.16 A simplified depiction of a Gilbert cell–integrated mixer.

primary and secondary tuned tank circuits, only the desired IF frequency will be of any significant amplitude.

Gilbert cell mixers, the most advanced of the active types, are commonly available up to RF frequencies of 5.8 GHz, with IF frequencies of 2 GHz, and are double balanced. A popular Gilbert cell mixer is shown in Fig. 7.16. The RF signal is inserted into the base of Q_1 of the modified emitter-coupled amplifier (comprised of Q_1 and Q_2), while the square wave LO is input into the base of Q_3. The LO causes Q_3 and Q_6 to be *on*, while next Q_4 and Q_5 transistors are turned *on*, causing mixing of the LO and the RF, creating the IF signal. When Gilbert cell IC mixers are operated at low frequencies, a square wave local oscillator must be used to decrease the steadily rising NF of the mixer, caused by the longer off times of the quad transistor's (Q_3, Q_4, Q_5, and Q_6). A square wave minimizes these off times.

Many active mixers can be operated past their maximum rated frequency, but there may then be a conversion loss instead of a conversion gain. Depending on design, these mixers are also available for both balanced and unbalanced operation, and can be found in low-cost surface-mount packages.

7.2.3 Active Mixer Design

Presented below are some low-cost discrete active mixer designs for consumer applications. However, the use of modern Gilbert cell–integrated mixers is sometimes the quickest and easiest route to take in today's fast-paced wireless market.

Dual-Gate Single-Ended Narrowband MOSFET Mixer for up to 250 MHz (Fig. 7.17)

This is a simple, inexpensive mixer that is a good choice for lower frequency RF applications in consumer wireless equipment. If the RF frequency is not exceptionally high, this MOSFET mixer will supply conversion gain. It does not require the high levels of LO drive that the passive mixers do, but has much lower intermodulation performance. The output port should be diplexed or padded to decrease IMD products. Stage gain will vary with LO amplitude and terminating impedances.

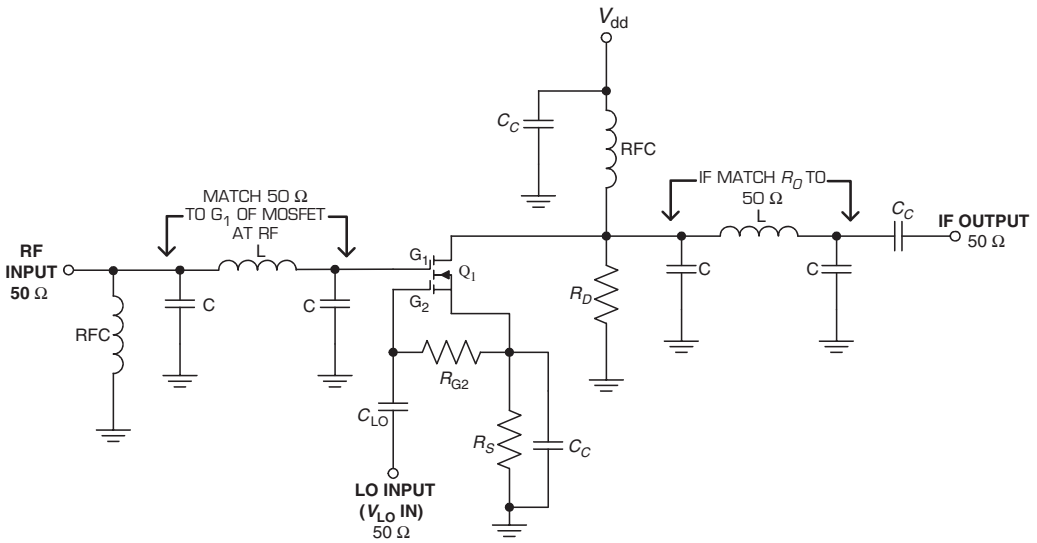


Figure 7.17 A narrowband active dual-gate MOSFET single-ended mixer design.

To Design

1. Select an RF dual-gate *N*-channel E-MOSFET that has plenty of gain at and significantly above the highest RF frequencies of interest.
2. $C_{LO} = C_C/10$ (if LO not buffered)
3. $R_{G2} = 100\ \text{k}\Omega$
4. $R_S = 560\ \Omega$
5. $C_C = < 1\ \Omega$
6. $V_{LO} \approx 6$ Vpp
7. $V_{dd} = 12$ V
8. $R_D = 2\ \text{k}\Omega$ to $5.6\ \text{k}\Omega$ (R_D pulls the MOSFET's drain down to the value of R_D, rather than to that of the MOSFET's low-frequency high-output impedance. This use of R_D helps IMD levels, but since Z_{OUT} drops as the frequency increases, R_D is not required at the higher frequencies.)
9. Match the input and output of the MOSFET to 50 Ω using *S*-parameters.
   This MOSFET mixer should exhibit the following specifications:

   $V_{G2} = 1$ V (supplied by self bias to MOSFET)

   $P_{1dB} \approx 1$ dBm (output)

   TOIP ≈ 17 dBm (output)

   $\text{RF}_{IN} < -12$ dBm (for decreased IMD levels)

   GAIN ≈ 12 dB – MAG (≈ +10 dB)

   LO-to-RF isolation ≈ 30 dB

   NF ≈ 8 to 10 dB

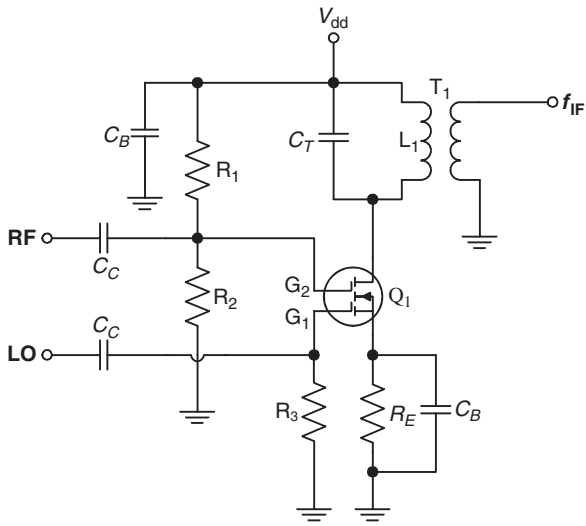


Figure 7.18 A dual-gate MOSFET mixer design.

Dual-Gate MOSFET Single-Ended Mixer for up to 400 MHz (Fig. 7.18)

This mixer is useful for undemanding consumer applications. The mixer's output should be diplexed or padded to decrease IMD's. Voltage gain of this mixer varies with LO amplitude and terminating impedances, but should be around a value of 10.

To Design

1. Select an RF dual-gate *N*-channel E-MOSFET that has plenty of gain at and significantly above the highest RF frequencies of interest.
2. C_B and $C_C = 1/(6.28 \cdot f)$
3. $C_T = \dfrac{1}{4[f_{IF}{}^2(\pi^2 \cdot L_1)]}$

   or
4. $L_1 = \dfrac{1}{4[f_{IF}{}^2(\pi^2 \cdot C_T)]}$

   where f_{IF} = frequency of the IF, Hz, f = frequency of the RF or LO, Hz.
5. $R_1 = R_2 = R_3 = 100\ \text{k}\Omega$
6. $R_E = 1.2\ \text{k}\Omega$

Distributed Narrowband GaAs FET Mixer for Microwave Operation (Fig. 7.19)

Even though a mixer is always run in nonlinear fashion, this mixer circuit can be effectively simulated using *S*-parameters in any linear software simulation program (such as the included *Qucs*). In this way the input and output impedances, as well as the stability analysis, can be roughly performed on a computer. After the circuit is computer analyzed, and then physically constructed, the microwave mixer must be further tweaked on the bench in order to provide the highest stability, conversion gain, and port isolation, as well as the lowest LO input drive requirements.

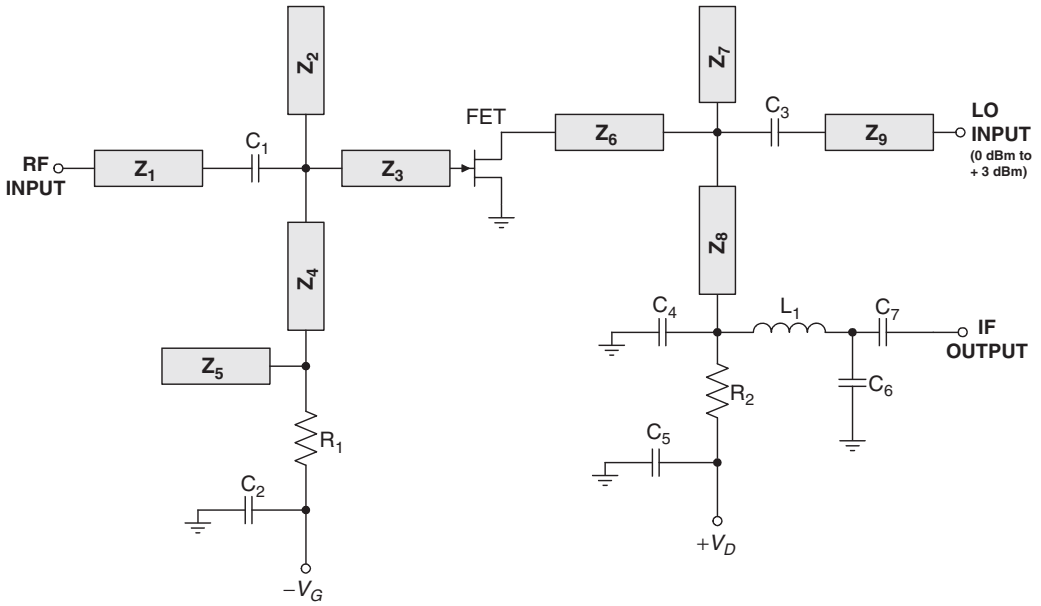


Figure 7.19 An active distributed mixer based on the JFET.

This active mixer design demands a good S_{11} match at the RF input of the FET at the RF frequency, with the FET's output being S_{22} matched at the LO frequency, along with the IF matched to the FET's output, which must form a diplexer. All of this permits the LO port to have a high return loss while rejecting the IF frequency, and allows the IF port to have a high return loss while rejecting both the RF and LO frequencies from being output, which enhances conversion gain, minimize LO drive power, and outputs a relatively clean signal from the mixer's IF port.

To Design

1. A GaAs JFET must be selected that can operate at a frequency far above the expected RF input frequency.
2. Z_1 and Z_9 are 50-Ω microstrip transmission lines.
3. C_1 will block DC, but pass the desired RF frequency with less than 1-Ω X_C.
4. Z_2 and Z_3 provide the proper input impedance match at the RF frequency for the JFET.
5. Z_4 acts as an distributed RFC to the desired RF frequency, while Z_5 functions as a capacitor. They form bias decoupling for the negative V_{CC} supply.
6. R_1 functions as a low-frequency termination to maintain mixer stability. Values of between 10 to 50 Ω should suffice.
7. C_2 is used to prevent the IF frequencies from exiting the RF port.
8. $-V_G$ should be adjusted from –5 to –1 V for best mixer operation.
9. Z_6 and Z_7 will match the S_{22} of the JFET at the LO frequency.

10. Z_8 functions as an RFC to attenuate the LO from entering the bias supply ($+V_D$) or the IF output port, but allows the DC and IF to pass unhindered. C_4 passes the LO to ground, and acts as the RF ground for Z_8.
11. C_5 bypasses the IF to ground to decouple from $+V_D$.
12. $+V_D$ should initially be set to +5 V, and then decreased for optimum performance.
13. L_1 and C_6 are chosen to match the IF frequency to the FET's output, while lowpass filtering the IF output for increased isolation.
14. C_7 is a DC block, but passes the IF with little attenuation. Could be parasitically series resonant to attenuate other frequencies besides the IF.
15. C_3 is a DC block, and should be chosen to operate at its series resonant frequency at the LO to assist in blocking the undesired IF frequency, while also increasing port isolation.
16. $R_2 = 50\ \Omega$

Integrated Circuit Double-Balanced Mixer for up to 5 GHz (Fig. 7.20)

Active mixers, such as the Agilent IAM-82028 and the IAM-82008, are useful in nonnoise sensitive applications that require a small LO input power (0 dBm). This IC mixer, unlike some active and many passive mixers, also possesses the very desirable trait of *load insensitive performance* due to its onboard buffer amplifier, and thus will have better IM suppression and conversion loss characteristics even if the RF and IF load impedances fluctuate. The IAM-82028 Gilbert cell–based mixer operates with a flat RF to IF conversion gain of 15 dB over a wide RF input range of 0.05 to 5 GHz, and enjoys an IF output capability of DC to 2 GHz. It has a maximum output P1dB value of 12 dBm, which is dependent on the V_{CC} voltage (7 V = 2 dBm, 12 V = 12 dBm), and will function with any V_{CC} between 7 to 13 V.

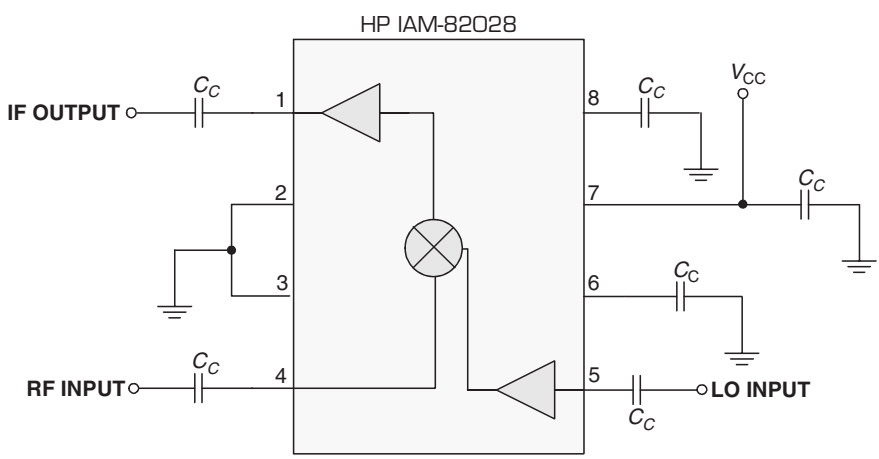


Figure 7.20 A popular IC active mixer.

To Design

1. Add the coupling/decoupling components as shown.
2. Supply the grounds and V_{CC}.
3. Done!

7.2.4 Active Mixer Issues

No circuit synthesis design software or calculation will produce a realizable mixer with 100% accuracy. This is due to the practical inability of formulas and programs to take into account PCB layout and component parasitic effects, as well as nonlinear device issues. Indeed, in order to obtain a realistic and optimized mixer stage, we must tune the as-calculated circuit values within a nonlinear RF simulator. This will also mean adding all the appropriate PCB traces, substrate, circuit elements, layout geometry, and full active/passive device models to the nonlinear mixer simulation. Therefore, unless the proper physical PCB layout, assembly, and circuit tuning have occurred within the RF software, then no microwave or RF circuit will operate as expected or as designed. Further, we will then have to perform, realistically, a certain amount of bench tuning in order to arrive at a final, reproducible mixer circuit that is fully optimized.

When an active mixer is utilized in any up conversion role, the input and output ports should not be swapped, as they would for a passive mixer (such as a DBM). The active mixer's input port will still be for the RF, while the output port will still be for the IF. However, since the IF port is generally capable of less than half, to as low as a quarter, of the output frequency that the RF input port is capable of, then up conversion, at least with any gain, will be limited to the rated output frequency of the IF port. Thus, many active IC mixers will be incapable of being used for upconverting a signal beyond 1 or 2 GHz.

Since most active mixers are much less sensitive to port mismatches than passive diode models, the LO input to the active mixer will normally not require an external buffer amplifier, nor will the IF port need a diplexer.

Many IC mixers will have DC voltages present at all ports, necessitating a series blocking capacitor at the RF, IF, and LO ports.

7.3 Image-Reject and Harmonic Mixers

7.3.1 Introduction

An *image-reject mixer* in a superheterodyne receiver can be used to phase cancel an offending image frequency and noise, instead of employing a filter for this purpose, while a *harmonic-mode* mixer permits a designer to employ a much lower LO frequency than would normally be required.

7.3.2 Image-Reject Mixers

One way a mixer can be exploited to suppress image frequencies and noise in a receiver is shown in Fig. 7.21. By using Mixer1 and Mixer2 to downconvert both the desired signal *and* the image to baseband by a 0° and +90° phase-shifted LO, the baseband *Q* leg of the signal is altered by 90°, while the *I* leg is not phase shifted at all. These two signals are then inserted into the combiner and added, which cancels the image frequency, and

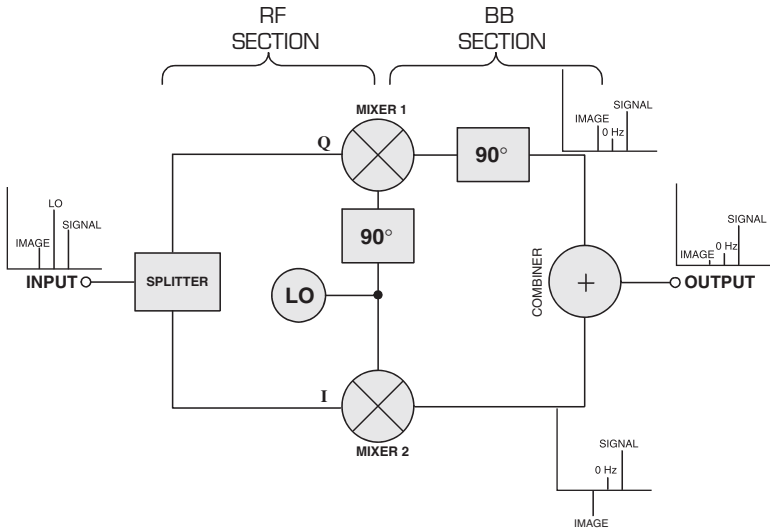


Figure 7.21 An image-reject mixer circuit.

adds the desired signal, doubling its amplitude. Image suppression is rarely better than 30 dB; however, so any high-amplitude signals present at the image frequency would still cause interference in channel.

7.3.3 Harmonic-Mode Mixers

Operating a mixer in harmonic mode allows a system designer to use a much lower LO frequency than would normally be required. Ordinarily, only the sum or difference frequencies are employed at the IF output of a mixer, but any convenient mixing product may be used for this purpose, such as $f_{RF} - 3f_{LO}$, $f_{RF} - 5f_{LO}$, $f_{RF} + 3f_{LO}$, or the $f_{RF} + 5f_{LO}$ products.

Nonreflective filtering must generally be placed at the mixer's output port, typically using a diplexer. The nonreflective filtering is necessary due to the reflective stopbands of a normal output IF filter, since the undesired signals and products would be reflected back into the IF port of the mixer, causing two-tone IMD performance to suffer.

This page intentionally left blank

CHAPTER 8
Support Circuit Design

Without electronic switching, attenuators, frequency multipliers, automatic gain control, power supplies, regulators, baluns, and so on, many modern wireless systems could not satisfactorily function. These are the essential support circuits. There are other circuits that could only be described as "bells and whistles," and are not essential for proper systems operation. These particular noncritical circuits will not be discussed.

8.1 Frequency Multipliers

8.1.1 Introduction

Because sinusoidal crystal oscillators can rarely be designed to operate reliably at frequencies above 200 MHz, even on a crystal's overtone, frequency multipliers can be employed to increase a signal's frequency.

The frequency doublers and triplers found in FM transmitters and microwave LO stages are either a basic tuned-output nonlinear (Class B or C) amplifier, or diode multiplier circuits. These multipliers are not only able to increase an FM or CW signal's frequency, but also any FM deviation that is present. This ability is required in an FM transmitter system, as a carrier oscillator's frequency and its deviation may need to be multiplied by 30 or more times. For instance, a modulated RF carrier that began at 6 MHz with an FM deviation of 150 Hz could be altered, if fed through a $30 \times$ RF chain of multipliers, into an output frequency of 180 MHz with an FM deviation of 4500 Hz.

Due to their nonlinear nature, common Class C amplifiers (Fig. 8.1) work quite well as frequency multipliers, especially when run into their saturation curve. An output of such an amplifier/multiplier can be rich in harmonics, since any distortion of a CW wave will produce multiples of the fundamental frequency. The Class C's input tank is tuned to the fundamental frequency that must be multiplied, while the output tank is tuned to an exact integer multiple of that frequency. Thus, the output tank of the active multiplier will not only send the selected harmonic on to the following stage, but will also attenuate the fundamental and the subharmonics created by the nonlinear amplifiers multiplication process.

The *push-push doubler* of Fig. 8.2 is another class of active frequency multiplier, and employs two active devices and two transformers. One of the push-push doubler's JFETs is receiving a 180° out-of-phase signal, while the other is receiving an in-phase signal, and since their outputs are tied together in parallel, twice as many output pulses will occur. These pulses are sent into the output tank circuit, which is tuned to the desired harmonic. Active frequency doubling of the original input signal has thus taken place.

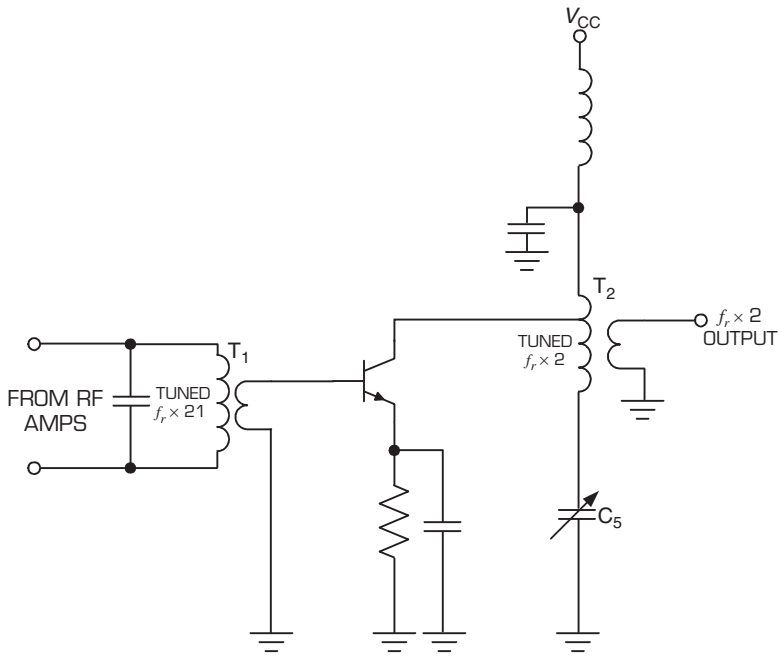


Figure 8.1 A low-frequency Class C frequency multiplier.

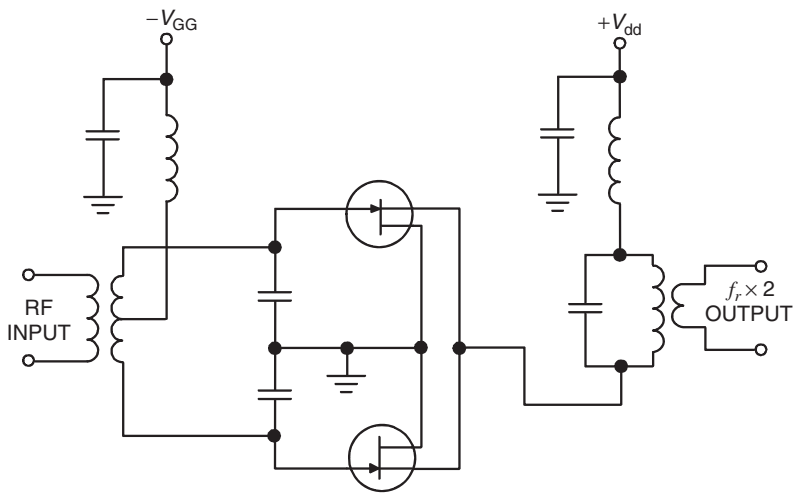


Figure 8.2 A JFET push-push frequency doubler circuit with biasing.

Passive methods of multiplying frequencies are prevalent as well. *The tripler varactor frequency multiplier* of Fig. 8.3 is one such circuit. It functions so: As the frequency to be multiplied is coupled into the input tank through the transformer, which is tuned to this frequency, the varactor will be continuously switched *on* and *off*. This action severely distorts the $1 \cdot f_r$ frequency, creating harmonics of $n \cdot f_r$. In this case we have a tripler, so

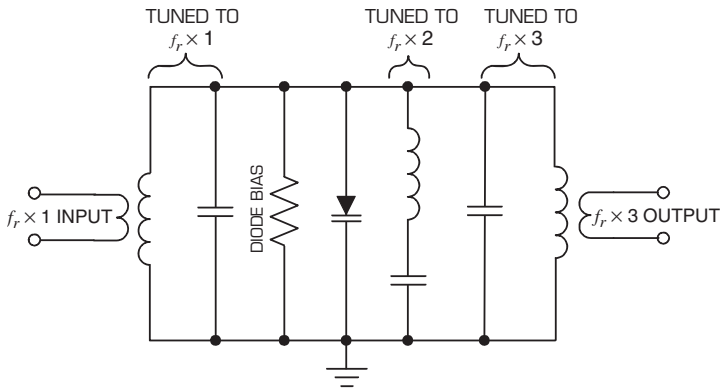


Figure 8.3 A varactor frequency multiplier circuit.

the undesired 2nd harmonic is shunted to ground through the series bandtrap filter, while the desired 3rd harmonic is sent on to the $3 \cdot f_r$ tuned tank circuit, and is output through the transformer's secondary. *R* functions as a bias resistor for more efficient multiplication, obtaining its DC bias voltage from the rectification that occurs through the diode.

In fact, there are two different diode multipliers that may be designed, the *reactive* and the *resistive* types. Reactive multipliers are comprised of either a varactor diode (such as the tripler varactor multiplier above) or a *step recovery diode* (SRD), and are very narrowband by nature, with both the input and output impedance matching circuits being effective only over a very fine band of frequencies. This is mainly due to the inherently reactive nature of these diodes. Also, while SRDs are capable of high values of frequency multiplication, and can reach up to 20 GHz in frequency, they are quite expensive. Varactor diodes are much cheaper, but have much lower levels of significant frequency multiplication (the actual conversion efficiency of any varactor multiplier is approximately $1/n$, with n being the harmonic number). However, they have a higher maximum operational frequency than the SRD, with varactor GaAs types available that can be operated into the very high microwave region.

Resistive frequency multiplication is created by the nonlinear resistance innate in any Schottky diode. Since a pure resistance is not affected by frequency, as reactive components are, then resistive multipliers will not be influenced by any variation in the input or output matching circuit as frequency changes. This allows the resistive multiplier to function over a very wide band of frequencies, and with high stability. Unfortunately, lower efficiencies prevent a Schottky diode multiplier from producing a high order of harmonics. Therefore, Schottky doublers and triplers are the most common, with the maximum output power possible from this type of multiplier calculated by $P_{OUT} = P_{IN}/n^2$, with n equaling the harmonic and P_{IN} equaling the input power in watts to the multiplier. These multipliers can be used up to very high frequencies (> 90 GHz). And while an ideal Schottky diode will produce only odd harmonics, a real-life Schottky can be used to produce both odd and even harmonics because of its small generated internal offset voltage.

If drive levels of over 10 dBm are expected into a Schottky diode shunt multiplier, such as the one in Fig. 8.4, then an adjustable DC bias circuit should be utilized for

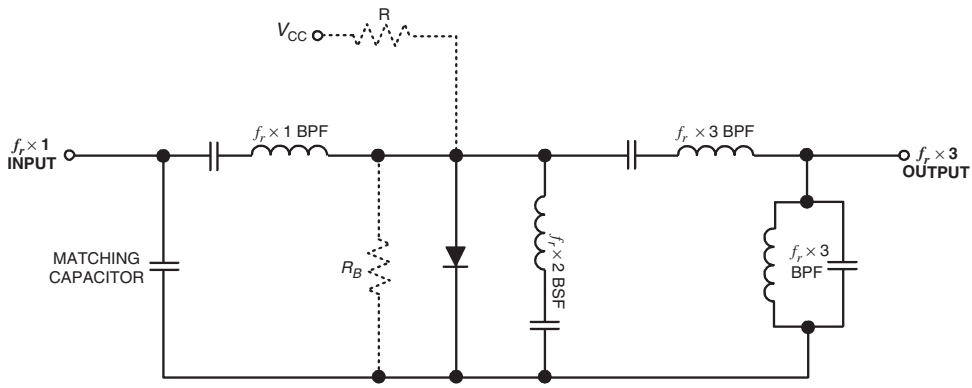


Figure 8.4 A harmonic generator for frequency multiplication showing two different diode bias methods, R_B and V_{CC}.

maximum multiplier efficiency. As shown, either a DC power supply or a simple resistor section can be added for this bias.

While multiplication is a viable and common method of increasing the frequency out of any oscillator, it should also be noted that phase noise will degrade a signal by a minimum of 6 dB per frequency doubling. Even PLL outputs can be multiplied to obtain almost any desired frequency, but at the expense of higher phase noise (and decreased frequency resolution). In fact, the phase noise of the frequency source that is feeding the input of the multiplier degrades by at least 20 log (n) at the multiplier's output, with n equaling the amount of multiplication.

8.1.2 Frequency Multiplier Selection

SRD circuits can be economically replaced by *transistor multipliers* for frequencies between 500 MHz and 4 GHz. Transistor's multipliers are superior to an SRD in cost, ease of tuning, output power level, and simplicity of design. *Gunn* and *Impatt* diodes can be used between 4 and 16 GHz, with Gunn multipliers being superior to an SRD in certain low-power applications because of their decreased noise, lower cost, and ease of design implementation. Impatt diode multipliers are cheaper and easier to design than the SRD when multiplier frequencies must reach 7 GHz and above, but SRD multipliers will have a wider bandwidth and lower noise generation. As mentioned above, *GaAs varactors* can be utilized at up to 100 GHz, are lower in cost, and can reach higher frequencies than SRDs, but cannot produce as many harmonic orders. *PIN diodes* can also be adopted in multiplier circuits to replace SRDs, with PIN multipliers having very similar characteristics to SRDs, but are much cheaper.

8.1.3 Frequency Multiplier Design

There are numerous multiplier circuit topologies that are perfect for different frequencies, costs, and multiplication levels. A few easy to design multiplier circuits are presented below.

Odd and Even MMIC Frequency Multiplier (Fig. 8.5)

By combining all of the below MMIC multiplication design techniques, a multiplication factor of up to 10 times the input frequency can be realized, and with a loss of only 30 dB below the fundamental input signal.

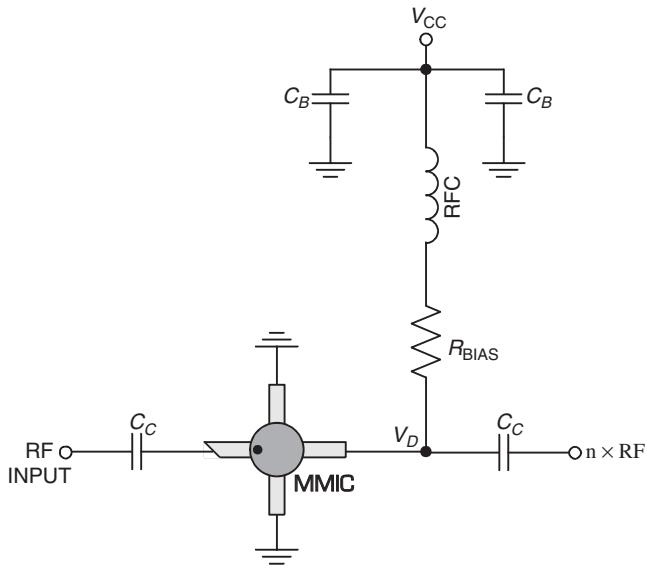


Figure 8.5 An overdriven MMIC used as a harmonic generator.

To Design

1. MMIC amplifiers can be adopted as harmonic frequency multipliers by overdriving its input into saturation (but not above its rated safe RF input level).
2. To minimize the MMIC's required drive level from the prior stage, the selected MMIC should have as low a P1dB as possible.
3. The selected harmonic should not be above the 3-dB bandwidth of the MMIC.
4. For higher RF output levels, the DC bias current should be increased by raising the MMIC's *V*cc supply voltage, or by decreasing its bias resistor value, R_{BIAS}.
5. The desired frequency must be picked-off from the comb of harmonics at the MMIC's output by a bandpass filter.
6. Proper physical spacing of the filter away from the MMIC's output port on the PCB can maximize the exact harmonic of interest, with this optimal spacing determined empirically.

SNAP Frequency Multiplier (Fig. 8.6)

Step-recovery, or "SNAP" diodes, function by switching between two extreme impedance conditions, low and high. This change of state may occur in only 200 pS or less, thus discharging a narrow pulse that is exceptionally rich in harmonics.

Due to their high cost, SRD frequency multipliers should only be selected, designed, and operated in very high-frequency circuits, as not only is the diode itself rather expensive, but the necessity to hand-tune each SRD circuit, along with the required +17 dBm input drive power, and other relatively difficult implementation issues, make the SNAP multiplier a poor choice as compared to PLLs or active harmonic multiplication.

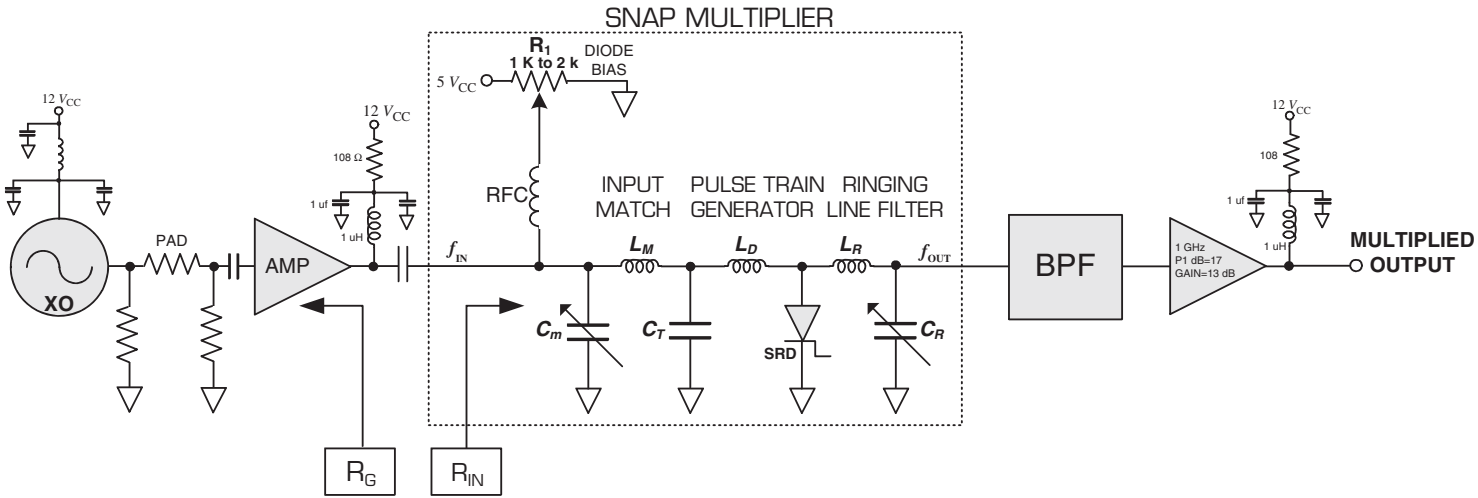


FIGURE 8.6 A complete SNAP multiplier with amplification.

To Design

1. Choose the correct SNAP diode:
   a. *Lifetime rating* on the SRD's data sheet must be *at least* 10 times longer than the period ($1/f$) of the RF input frequency, and is given in nS.
   b. C_{vr} (or C_j) = employ the nominal value on SRD's data sheet for L_D calculation below.
   c. *Transition time* (T_t) = must be less than one period ($1/f_o$) of f_o (the final RF output frequency) on SRD's data sheet, measured in pS.
2. $T_p = f_o/2$
3. $R_{IN} = 6.28 \cdot f_{IN} \cdot L_D$
4. $R_G = 50\ \Omega$
5. RFC = 600-Ω impedance at f_{IN}, but capable of attenuating f_o as well.
6. $C_m = \dfrac{1}{6.28 f_{IN} \sqrt{R_G R_{IN}}}$
7. $L_M = \dfrac{\sqrt{R_G R_{IN}}}{6.28 f_{IN}}$
8. $C_T = \dfrac{C_{vr}*}{(2 f_{IN} Tp)^2}$ (**Provided on SRD's data sheet, or as C_j*)
9. $L_D = \left(\dfrac{Tp}{\pi}\right)^2 \cdot \left(\dfrac{1}{C_{vr}*}\right)$ (**Provided on SRD's data sheet, or as C_j*)
10. $L_R = L_D$
11. $C_R = \dfrac{1}{(2\pi f_o)^2 L_R}$
12. $f_o = f_{IN} \cdot n$ (n = desired harmonic)
13. Tune the SNAP circuit: Since the diode's output is a comb of frequencies, a spectrum analyzer is utilized to tune the SRD for maximum power output at the desired harmonic. Insert the proper frequency and power (+17 to +23 dBm) at the SRD circuit's input. Adjust C_M (input match) and R_1 (diode bias) and C_R (line filter) until the desired output frequency is at maximum power, while confirming that all subharmonics and harmonics are properly attenuated.

Wenzel Tripler Schottky Multiplier for up to 1.2-GHz RF Output (Fig. 8.7)

This is one of the easier frequency multipliers to design. L_1 and C_1 form the input matching network, while also attenuating the driver's harmonics from entering the input of the multiplier, as well as increasing the voltage to the Schottky diode. D_1 and D_2 are the nonlinear, harmonic generating devices. L_2 is the DC ground for the diodes, with the DC being created by the AC of the input being rectified by D_1. L_4 and C_3 shunt

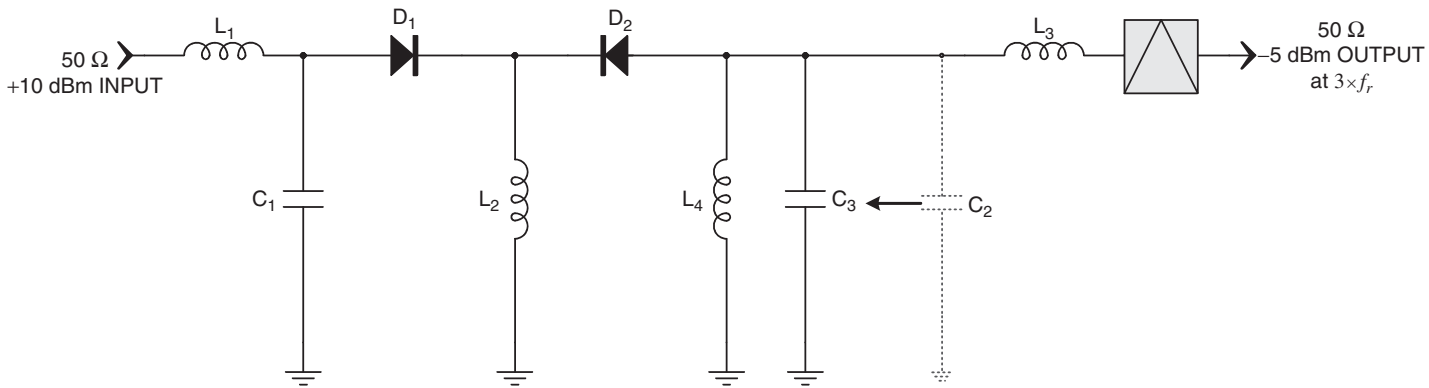


Figure 8.7 An odd-order frequency multiplier.

the undesired frequencies to ground, while passing only the frequency of interest and matching the output. As with most multipliers, the diode switching times will be degraded unless these higher frequencies are decreased. L_3, along with C_3, form the output matching network for $3 \times f_r$. RF input power should be approximately +10 dBm, which will give us a tripled output of around −5 dBm.

To Design

1. Tune L_1 and C_1 for best S_{11} at f_r.
2. D_1, D_2 = Schottky diodes of the *low flicker* type.
3. L_2 = RFC at f_r.
4. L_4 is at parallel resonance with C_3 at $f_r \cdot n$ (with n, the desired harmonic, normally equaling 3).
5. L_3, with a shunt capacitor C_2 (that has been later absorbed by C_3), are both tuned for maximum 3rd-harmonic output power and return loss.
6. Center frequency for the output BPF is $f_r \cdot n$.

A Quick Example Design a Passive Schottky RF Frequency Multiplier (Fig. 8.8)

Goal: Create a stable diode RF frequency tripler. The specifications and parameters for the circuit are:

$f_{r(\mathrm{IN})} = 200$ MHz
$f_{r(\mathrm{OUT})} = 600$ MHz
$P_{\mathrm{IN}} = +10$ dBm
$P_{\mathrm{OUT}} = -5$ dBm
Diode = Avago HSMS-2820 (Schottky type, appropriate for frequency of interest)

One possible Solution:

1. L_2 = 500-Ω X_L at $f_{r(\mathrm{IN})}$.
2. L_4 and C_3 to resonate at $f_{r(\mathrm{OUT})}$.
3. Empirically tune L_1 and C_1 for best input match and output power at $f_{r(\mathrm{IN})}$.
4. Empirically tune L_3 and C_2 for best output match and output power at $f_{r(\mathrm{OUT})}$.
5. Design an output bandpass filter (BPF) to attenuate all harmonics and subharmonics, as required.

Note: *C_2, as in the worked example, can usually be absorbed by C_3.*

RF Transistor Tripler Multiplier (Fig. 8.9)

This transistor tripler design has an RF output capability of 0 dBm at 2.4 GHz when subjected to an input of 0 dBm at 800 MHz, showing very good conversion efficiency. Typically, an output bandpass filter for the 3rd harmonic will be required to attenuate harmonics and subharmonics beyond that afforded by the filtering action of the multiplier's tank circuit.

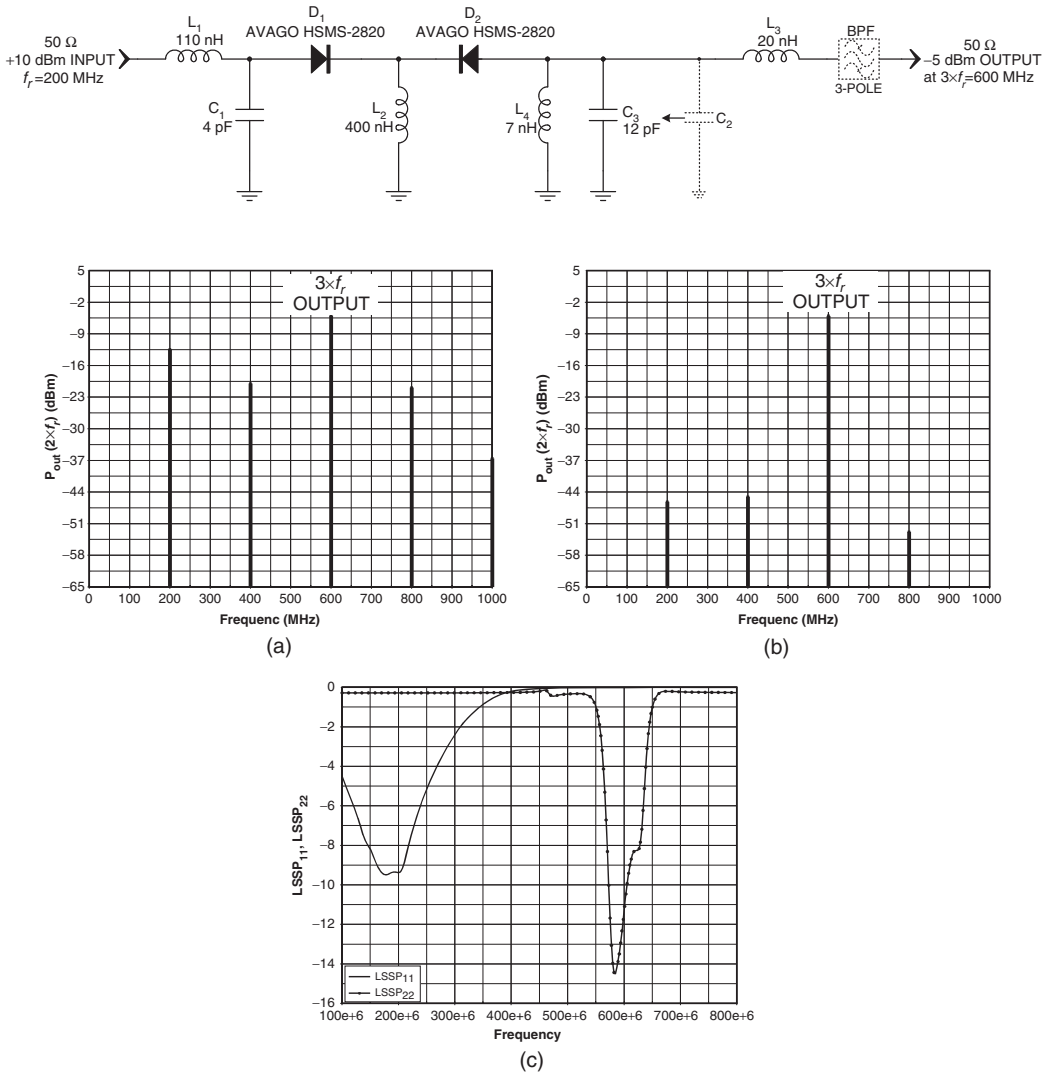


Figure 8.8 The example diode frequency multiplier circuit with calculated part's values, along with frequency sweep results: (*a*) Unfiltered output spectra; (*b*) Basic three-pole LC bandpass filtered output; (*c*) Large signal *S*-parameters of the circuit's input ($LSSP_{11}$) and output ($LSSP_{22}$) ports.

The tripler will have an input matched for a good return loss at its fundamental frequency f_r, and an output port optimized to match/bandpass $3 \times f_r$. The bias is Class AB or B.

To Design

1. Select the proper active device that fits the requirements of frequency, gain, V_{CC}, cost, package, availability, and so on. It must also have linear models available, with nonlinear models a significant bonus.

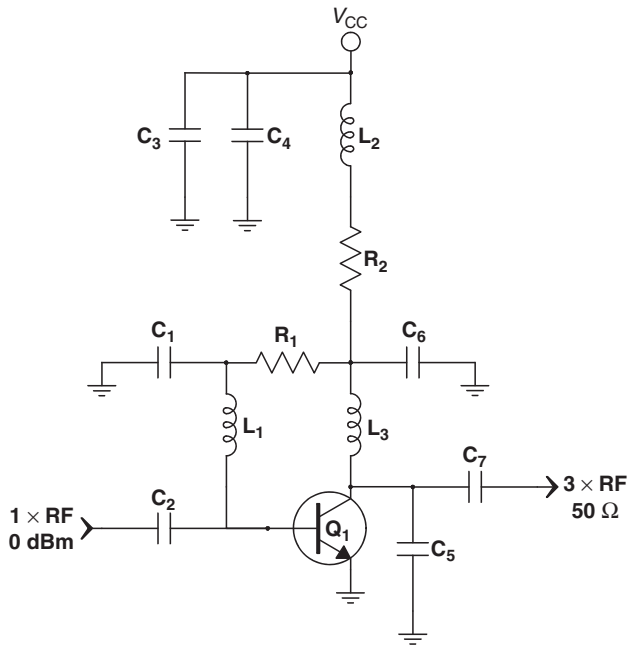


FIGURE 8.9 An overdriven discrete amplifier used as a frequency multiplier.

2. Select the linear model for the chosen RF transistor that fits our desired bias level. In this case, a low idling current of a few *mA*, and a collector voltage of $0.6 \times V_{CC}$, would be a good compromise approach for thermal stability, gain, and efficiency.
3. Design the DC bias network as presented under Sec. 3.6.5. Normally this will be of the collector-feedback type.
4. Insert the BJT model within the linear simulator and create an impedance matching network of the fundamental frequency at the transistor's input port, and of the 3rd harmonic frequency at the transistor's output port.
5. Tune the input and output network in the linear simulator to optimize S_{11} at f_r, and S_{21} and S_{22} at $3 \times f_r$.
6. Replace all ideal components, and insert accurate passive models and PCB microstrip in to the simulation, and retune.
7. If the simulator and the model are available, also nonlinear simulate the multiplier circuit, further tuning for maximum $3 \times f_r$ output power.
8. Layout the multiplier's printed circuit board as recommended under "EMI Control and PCB Layout" Chap. 13, then fabricate and assemble the board.
9. On the bench, further tune the physical multiplier for optimal $3 \times f_r$ output power, efficiency, and stability.

NOTE: *The efficiency of active frequency multipliers is highly dependent on the input/output matching networks and the transistor. For maximum harmonic generation by the transistor, the input matching network is used to assure maximum power transfer from the driving stage into the tripler's input. The output matching network needs to transfer maximum 3rd-harmonic output power to the next stage, as well as assist in attenuating all undesired harmonics.*

Due to the transistor's strongly nonlinear operation when used as a frequency multiplier, designs are normally easier to perform by tuning the input and output networks first in a harmonic balance simulator, if the proper nonlinear model of the active device is available, until maximum 3rd harmonic amplitude is attained. During this tuning process confirm that the transistor is biased and operating at far below its maximum power dissipation rating, as well as at high conversion efficiency. After construction of the tripler, we must then bench tune the physical circuit in order to optimize its output signal amplitude, filtering, and efficiency.

A Quick Example Design an Active RF Frequency Tripler Multiplier (Fig. 8.10)

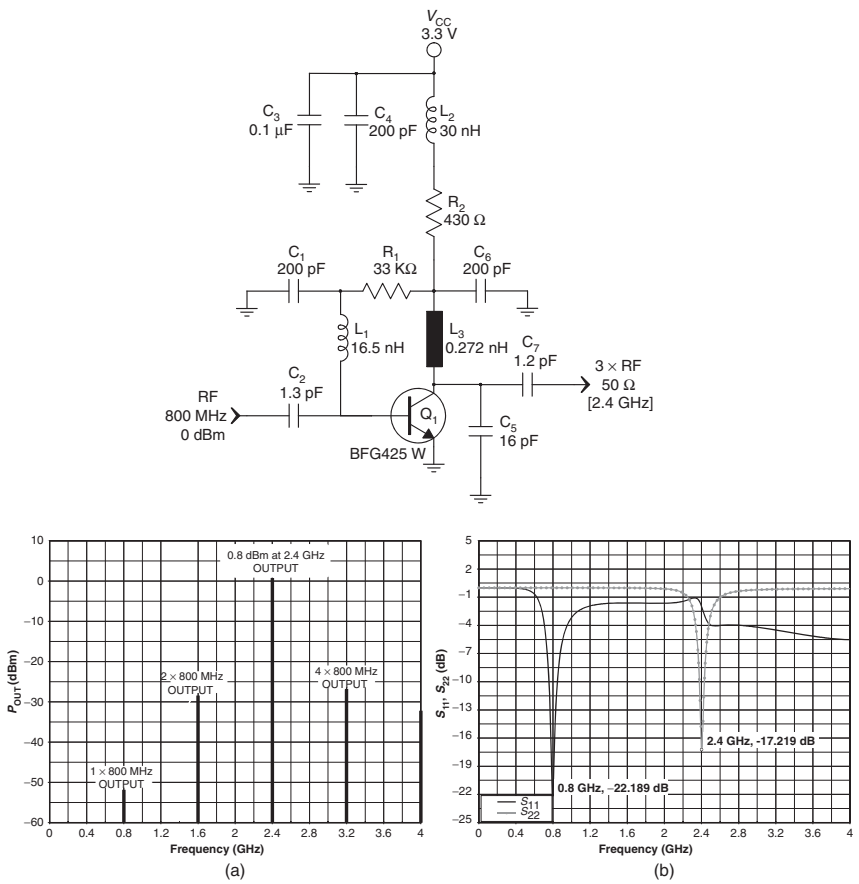


FIGURE 8.10 The example BJT frequency multiplier circuit with calculated part's values, along with graphed results: (*a*) Harmonic balance output spectrum; (*b*) Small signal *S*-parameter sweep of the circuit's input (S_{11}) and output (S_{22}) ports (initial simulations can be quickly performed with a linear simulator).

Goal: Create a stable active RF frequency tripler. The specifications and parameters for the circuit are:

V_{CC} = 3.3 V
V_{CE} = 2.0 V (for bias network calculations)
I_{CE} = 3 mA (for bias network calculations)
$f_{n(IN)}$ = 800 MHz
$f_{n(OUT)}$ = 2.4 GHz
P_{IN} = 0 dBm
P_{OUT} = 0 dBm
Transistor = NXP BFG425

One possible Solution:

1. R_1 = 33 kΩ, R_2 = 430 Ω (bias components)
2. C_2 = 1.3 pF, L_1 = 16.5 nH (input match for 800 MHz)
3. L_3 = 0.272 nH (distributed), C_5 = 16 pF, C_7 = 1.2 pF (output match for 2.4 GHz)
4. C_1, C_4, C_6 = 200 pF (X_C = 1 Ω), L_2 = 30 nH (X_L = 500 Ω)

NOTE: *In order to supply the desired SRF operation for* C_1, C_4, C_6, *and* L_2, *these decoupling and coupling component values will vary with package size and internal design.*

8.1.4 Frequency Multiplier Issues

Frequency multiplication up to high orders can be problematic for the stability of a multiplier circuit, as well as for the filtering of all the many subharmonics and harmonics (Fig. 8.11) at the output port. The difficulty in filtering can readily be seen, since the output harmonics and subharmonics of the fundamental may be spaced on either side of the frequency of interest by only a small amount, making it challenging to suppress to acceptable levels.

It has been found that multiplying beyond a tripler with a normal silicon diode may add far more phase noise than the minimum of 20 log (n) that one would expect in phase-noise multiplication. This is caused by the diode that is used for the nonlinear multiplication creating high noise floors and flicker noise. Schottky diodes

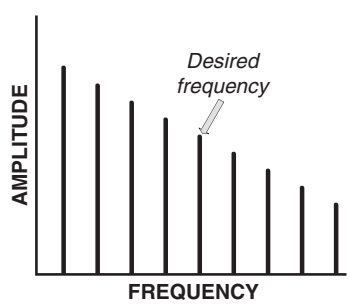


FIGURE 8.11 Subharmonics and harmonics of a multiplier stage.

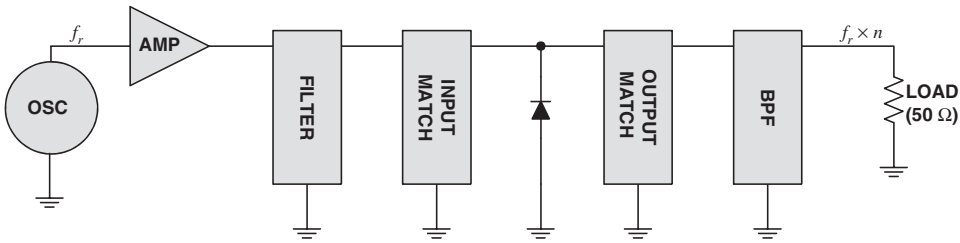


Figure 8.12 Block diagram of frequency multiplication with a nonlinear element, a varactor diode.

will not have this augmented noise problem beyond 20 log (n). Doubler and tripler topologies using low-noise Schottky diodes can usually maintain a good 20 log (n) phase-noise degradation, even up to high multiplications (such as $n = 7$). And even though varactor multipliers (Fig. 8.12) can be delicate to tune for top performance, they too add very low excess phase and amplitude noise levels beyond the unavoidable 20 log(n).

Because of the sometimes necessary use of Class A amplifiers needed to increase the low RF output power in a lossy diode multiplier's chain, active Class C multipliers may actually consume *less* DC current than many such amplified diode circuits.

Additionally, any cascading of multiple varactor diode multiplier stages opens up a severe danger of instability (oscillations), which is not as much of a risk as with the active multiplier types. However, there are many low phase-noise applications that will demand diode multipliers over the noisier active types, especially within digitally modulated communication radios.

In active multiplier design it has been recommended (by *Maas* and others) that the duty cycle of a BJT or FET amplifier be adjusted to optimize the preferred output harmonic, which can be done by biasing the amplifier close to cutoff and setting it to be *on* for 30% of the time for a doubler and 20% of the time for a tripler. So, the Class C bias point can be varied to optimize the desired harmonic for the proper *on* period for the highest drain or collector current amplitude, with the multiplier acting as a Class C amplifier and rectifying the incoming RF signal by producing only pulses, creating extreme distortion of the input sine wave, and thus significant output harmonics.

8.2 RF Switches

8.2.1 Introduction

Solid-state electronic RF switches can be utilized to remotely switch RF currents at high frequencies, where long wires or traces would be problematic due to excessive losses, EMI, and other compelling reasons.

RF switches can be either of the absorptive or reflective type. For a superior return loss during all switching states, the absorptive switch will 50-Ω terminate the RF input signal when the circuit is in the off condition. A reflective switch, conversely, will be in an open (or shorted) state when off, with a very poor return loss. In most RF applications either switch is appropriate, while in other more sensitive applications only the terminated switch would be acceptable.

The choice of exactly which active or passive RF switch circuit to adopt will depend on many things, such as the expected input signal strengths, whether the switch needs to be reflective or absorptive, the necessary output intermodulation level, the required bandwidth and frequency, and the like.

Active Switch

An elementary transistor circuit for switching RF currents is shown in Fig. 8.13. When the transistor's base is grounded through SW1, the NPN BJT will not conduct and therefore will not permit any of RF INPUT to proceed to the transistor's RF OUTPUT. However, when the mechanical switch is flipped over to engage the positive supply, a base current will start to flow through the base resistor, *R*. When this occurs the BJT's emitter-to-collector resistance will decrease, causing any collector current to flow. This allows the RF signal to pass on to RF OUTPUT.

If required for the switching of amplifier V_{CC} supply voltages, DC may also be controlled by adopting the PNP BJT of Fig. 8.14. When SW1 is set to ON, the base is grounded, which switches *on* the PNP transistor, permitting DC current to flow through

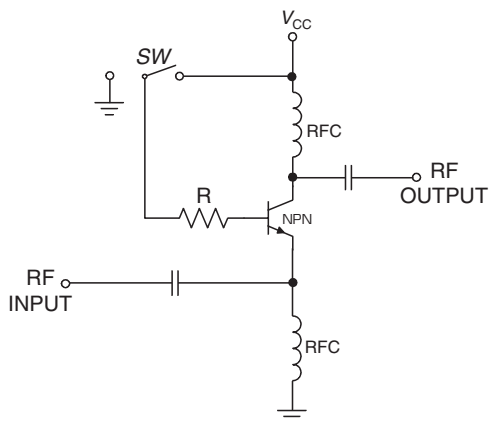


Figure 8.13 An RF switch using a transistor.

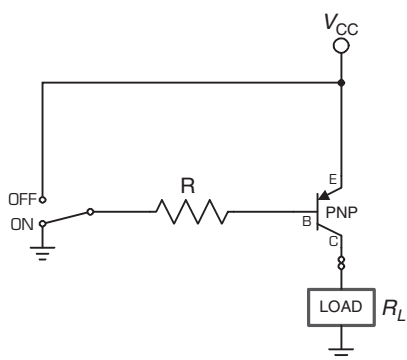


Figure 8.14 A DC switch using a transistor.

the resistive load, R_L. Turning the mechanical switch to the OFF position will stop the PNP's base-current flow with a positive bias voltage, causing an almost zero collector current through R_L.

Instead of directly placing a control voltage on the base of a BJT switch with the mechanical switch of SW1, a voltage can be sent to the base in reply to some other system condition. An example would be if excessive reverse power is detected at a high-powered transmitter's output, which should cause a solid-state BJT switch to shut down the system and/or to supply a panel warning. A similar switch can be used when a transceiver changes from receive to transmit, in which case the transmitter section, and no longer the receiver section, requires DC power to be electronically switched. This action will turn off the receiver and turn on the transmitter.

An extremely popular RF switch in common use today is the integrated circuit plastic packaged GaAs FET type, which is beginning to somewhat displace the discrete PIN switch circuits, described below, in many lower-powered RF applications. As an example of their performance level, common consumer grade plastic packaged GaAs FET SPDT integrated switches with a V_{CNTRL} of 3 V, at a frequency of 2 GHz, will have port-to-port isolation of roughly 27 dB, an insertion loss of 0.5 dB, and a P1dB of 25 dBm. Many such GaAs switches can function up to 6 GHz, and a few high-power units may have a maximum input power capability of 34 dBm. These switches are very easy to work with, trouble free to apply, have fast switching capability, consume virtually no DC power, and many can switch down to DC. Still, PIN diode switches can survive higher peak levels of RF better than GaAs switches, are still lower in cost, and create less distortion at higher RF power levels.

Passive Switches

Diodes serve a vital function in wireless communications for switching applications, and can be used to switch in or switch out various filters, crystals, tuned tanks, subsystems, active circuits, and so on, throughout the RF signal path.

A simple nonresonant diode switch is shown in Fig. 8.15. C_1 and C_2 block the DC bias, but easily passes the AC signal. When the mechanical switch (SW) is in the OFF position, a negative DC bias is placed across the diode, reverse biasing D_1 and stopping

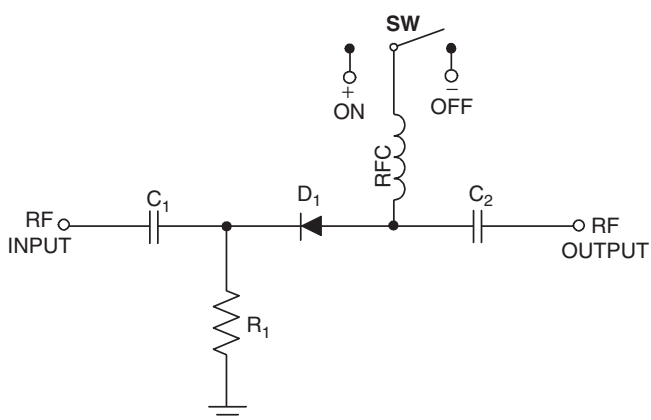


Figure 8.15 An RF switch using a diode.

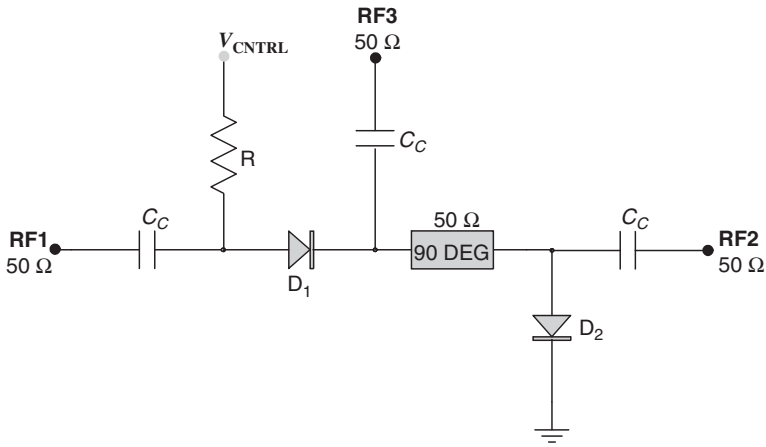


Figure 8.16 The common resonant PIN diode switch.

the signal from passing on to the output. In order to forward bias the diode and allow the RF signal to pass, the mechanical switch is turned *on*, which then places a positive voltage that forward biases D_1 into conduction. This permits the signal to cross the now low resistance of the diode to the output port. R_1 is a current limiting resistor to ground for the DC bias voltage through the diode, while the *RFC* prevents the RF from entering the power supply.

In narrowband applications, the popular resonant PIN switch circuit shown in Fig. 8.16 works so: When a signal must pass from the transmit port TX to the antenna ANT port, but not to the receive RX port, a positive voltage is placed at V_{CNTRL}. This turns *on* the two PIN diodes, which are now forward biased, making them a short circuit, and thus at a low RF impedance. The TX signal can therefore pass from the TX to the ANT port, but is blocked by the quarter-wave (90°) shorted stub, which creates an RF open through the forward biased diode D_2, and is therefore equivalent to an LC resonant parallel tank circuit at the TX frequency (and only at the TX frequency).

During RX, V_{CNTRL} is placed at 0 V, so both PIN diodes are *off* and at a high impedance, forming an open circuit at D_1 and D_2, permitting the received frequency to pass from the ANT port to the RX port, while being blocked from entering the TX port. That is, except for some RF leakage value due to the finite isolation of the PIN diodes, caused by the parasitic capacitance in series diode D_1, as well as the PCB layout. This is exactly why PIN diode D_1 should not only be selected so that it can safely handle the desired TX output power, but also to possess a low-off capacitance, which will permit the transmitter to maintain its rated RF isolation when the transmit port is *off*. This will also improve or maintain port isolation and insertion loss when the PINs are biased *off* for the RX stage as well. Moreover, D_2 must be a *low-series inductance* type to maximize RX switch isolation. In fact, the higher a PIN's series inductance when placed in shunt, the higher the value of the impedance to the RF, therefore decreasing the diode's ability to form a short circuit, decreasing the switch's important RX isolation capability.

A series/shunt PIN switch of this type, with a 10-mA bias current and a positive-only V_{CNTRL} voltage, can be expected to have an isolation between the TX-to-RX and the RX-to-ANT ports during transmit of around 25 dB at 2 GHz, and a TX-to-ANT insertion

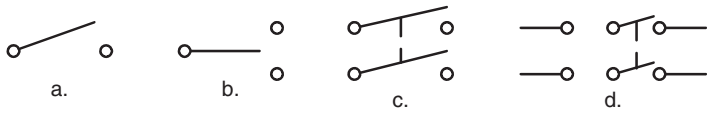


Figure 8.17 Typical mechanical switches with various poles and throws.

loss of 0.5 dB. During receive, the TX-to-RX isolation may be about 13 dB, with a TX-to-ANT isolation of about 20 dB, and an ANT-to-RX insertion loss of 0.5 dB.

Mechanical Switches

All switches are generally classified according to the number of *poles* and *throws* they possess. A pole is the number of switching contact pairs, while the throw is the number of switching positions that can actually conduct current (as opposed to just being an open). For instance, Fig. 8.17*a* shows a *single-pole single-throw* (SPST) switch; Fig. 8.17*b* is a *single-pole double-throw* (SPDT) switch; Fig. 8.17*c* is a *double-pole single-throw* (DPST) switch; and Fig. 8.17*d* is a *double-pole double-throw* (DPDT) switch.

A critical mechanical switch specification for DC and low-frequency AC is the maximum current draw that it is capable of surviving before switch damage occurs. The switch's maximum voltage rating is rarely an issue in modern wireless solid-state design, where low voltages are the rule.

8.2.2 RF Switch Design

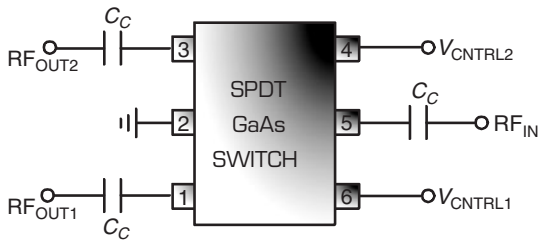
The following are a few popular and easy-to-implement electronic RF switch circuit designs that will cover the majority of common wireless requirements.

SPDT-Integrated GaAs FET Switch for up to 6 GHz (Fig. 8.18)

This is the simplest, most foolproof RF switch design possible.

To Design

1. Place DC blocking capacitors at all RF ports with a less than 2-Ω reactance (at the frequency of operation).
2. Many GaAs FET switch models recommend adding a series 100-Ω resistor and/or shunt decoupling capacitor on each control line to suppress noise inputs.
3. When laying out the PCB for the switch, keep the isolation between the RF ports as high as possible, and send all the switch's ground pins directly to the board's groundplane through multiple vias.
4. To program the switch's control lines, follow the part's switching logic table available on its data sheet.
5. When only a single control line is available from the system to control the GaAs switch, employ a CMOS or discrete BJT inverter as shown.



CONTROL INPUTS		OUTPUT PORTS	
V_{CNTRL1}	V_{CNTRL2}	RF_{OUT1}	RF_{OUT2}
LOW	HI	ON	OFF
HI	LOW	OFF	ON

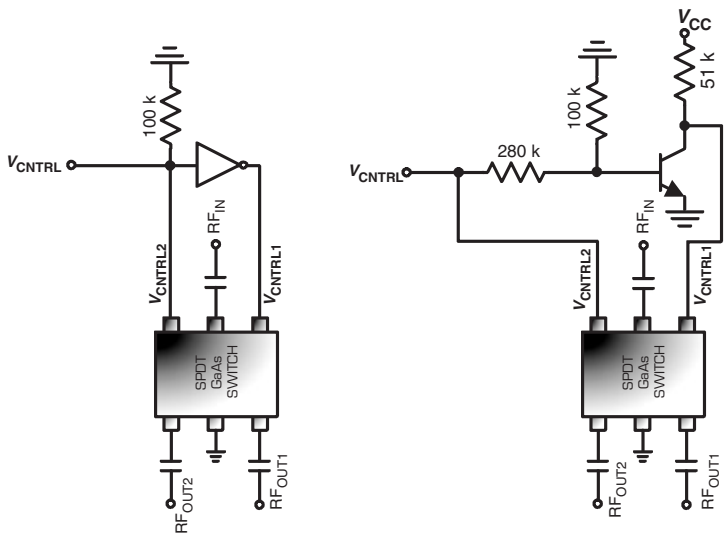


FIGURE 8.18 Typical GaAs RF integrated switch, along with its switching-logic truth table and single-line control circuits.

Small-Signal Very High Isolation RF PIN SPST Switch (Fig. 8.19)

To Design

1. RFC = 600 Ω at f_r
2. $C_B = C_C = 1\ \Omega$ at f_r
3. $R \approx \dfrac{\text{DC BIAS} - 0.9}{20\ \text{mA}}$

NOTE: *The 20-mA value is PIN3's forward current draw, while PIN1 and PIN2 will each be* on *by 10 mA, for a low value of ON resistance. Some PIN diodes may operate quite well with much less bias current, while other types may require much more. Therefore, check the PIN diode's data sheet for the appropriate bias current value that will optimize your switching circuit's needs.*

4. This design must have a positive voltage to turn *on* the series PINs and turn *off* the shunt PIN for an ON switch condition, while a negative voltage will switch *off* the series diodes and turn *on* the shunt diode (thus creating a high-isolation RF switch for the OFF condition). However, for low-level RF input signals, only

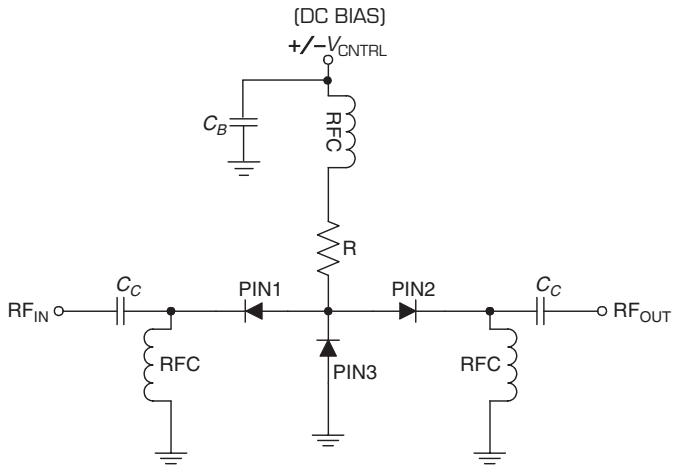


Figure 8.19 A high-isolation RF PIN diode switch.

a positive ON bias is required, while OFF simply needs 0 V. If there is no negative bias supplied, then high-level RF input signals may self bias the PINs to ON, which will lower the switch's OFF isolation drastically.

Low-Distortion High-Isolation Diode RF SPDT Switch for up to 1.5 GHz (Fig. 8.20)

Capable of operating with a very high port-to-port isolation of up to 40 dB in the OFF condition. It has a lower frequency limit of 50 MHz or so due to the PIN diode's carrier

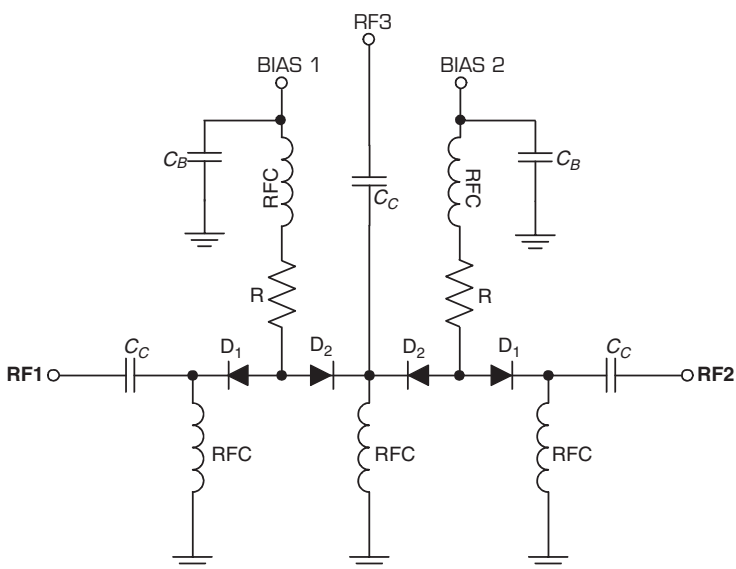


Figure 8.20 A wideband, low-distortion SPDT PIN diode switch.

lifetime limitations. A reverse bias of the OFF arm is not required, except to slightly improve harmonic distortion levels. To improve and optimize the output distortion levels it can be operated with a high 40-mA bias current for each of the bias legs (20 mA per diode). However, 20-mA total forward current per BIAS leg is sufficient to switch the diodes *on* with minimal, but not optimal, distortion of the RF input signal and a low ON resistance. (Some PIN diodes may operate quite well with much less bias current, while other types may require much more. Therefore, check the PIN diode's data sheet for the appropriate bias current value that will optimize your switching circuit's needs.) For switching low-level amplitude RF input signals from a receiver, Bias 1 or Bias 2 need only be +5 and 0 V. For higher power signals, the bias should be +5 V for ON and −5 V for OFF. *R* is the current limiting series PIN switch resistor for each diode pair.

To Design

1. RFC = 600 Ω at f_r
2. $C_B = C_C = 1\ \Omega$ at f_r
3. $R \approx \dfrac{\text{DC BIAS} - 0.9}{20\ \text{mA}}$ (DC BIAS is in positive volts.)

Shunt PIN SPST RF Switch (Fig. 8.21)

When this switch is in the ON (0 V or some negative voltage) condition the insertion loss is low, while the return loss is high. When the switch is in the OFF (positive voltage) condition, the signal is shunted to ground before it reaches RFOUT.

To Design

1. RFC = 600 Ω at f_r
2. $C_B = C_C = 1\ \Omega$ at f_r
3. $R \approx \dfrac{\text{DC BIAS} - 0.9}{10\ \text{mA}}$

NOTE: *Certain PIN diodes may require more bias current than 10 mA, while others may operate well with less; check data sheet for optimal value. DC BIAS is in volts.*

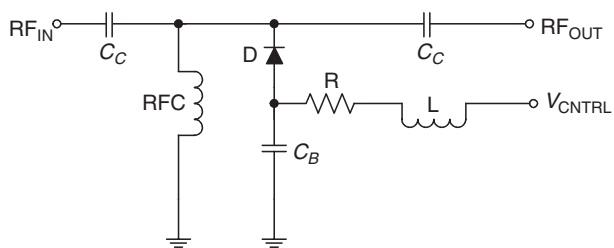


FIGURE 8.21 A shunt PIN diode SPST switch.

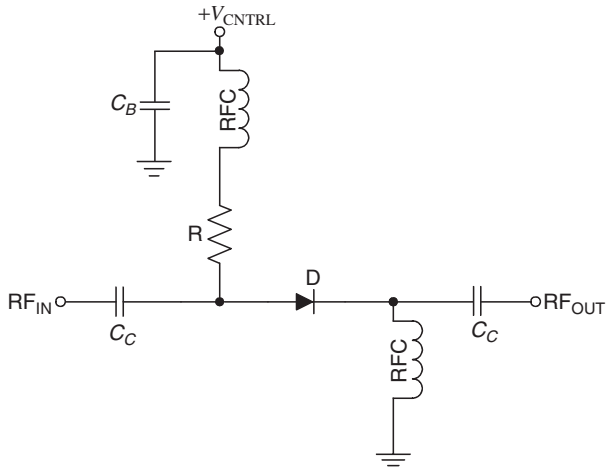


FIGURE 8.22 A low parts count PIN diode SPST switch.

Small RF Signal Series PIN SPST Switch (Fig. 8.22)

When this switch is in the ON condition (+5 V or higher at V_{CNTRL}), the insertion loss is low, while the return loss is high. When the switch is in OFF condition (0 V), the signal is severally attenuated by the diode, while also creating a low return loss.

To Design

1. RFC = 600 Ω at f_r
2. $C_B = C_C = 1\ \Omega$ at f_r
3. $R \approx \dfrac{\text{DC BIAS} - 0.9}{15\ \text{mA}}$

NOTE: *Certain PIN diodes may require more bias current than 15 mA, while others may operate well with less; check data sheet for optimal value. Higher diode bias currents can be used, up to 50 mA, to decrease distortion levels and ON resistance, if needed. DC BIAS is in volts.*

Series SPDT PIN Diode Switch (Fig. 8.23)

A positive voltage at V_{CNTRL} will steer the INPUT to RF2, while a negative voltage will steer the INPUT to RF1.

To Design

1. RFC = 600 Ω at f_r
2. $C_B = C_C = 1\ \Omega$ at f_r
3. $R \approx \dfrac{\text{DC BIAS} - 0.9}{10\ \text{mA}}$

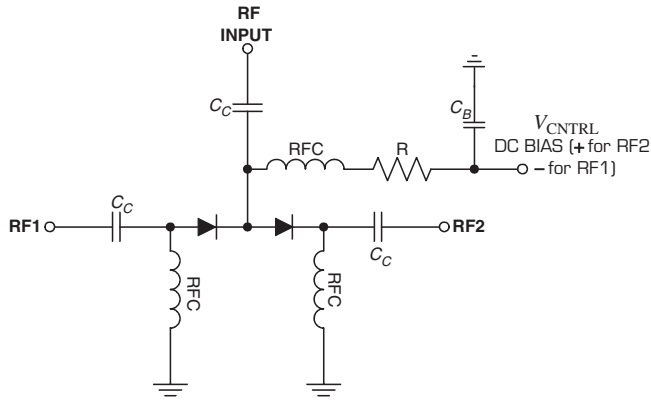


FIGURE 8.23 An SPDT PIN diode switch.

NOTE: *Lower diode ON bias currents than 10 mA can be used, but at the expense of increasing distortion levels and diode ON resistance. Certain PIN diodes may require more bias current than 10 mA, while others may operate quite well with less; check PIN data sheet for optimal value. DC BIAS is in volts.*

Narrowband RF PIN Switch for up to 6 GHz (Fig. 8.24)

In narrowband applications this resonant PIN switch circuit is extremely popular and is a good, low-cost performer. When an RF signal must pass from the TX to the ANT port, a positive DC voltage is switched to V_{CNTRL}, which turns *on* the two PIN diodes. The TX signal can therefore pass to the ANT port, but will be blocked by the 90° long

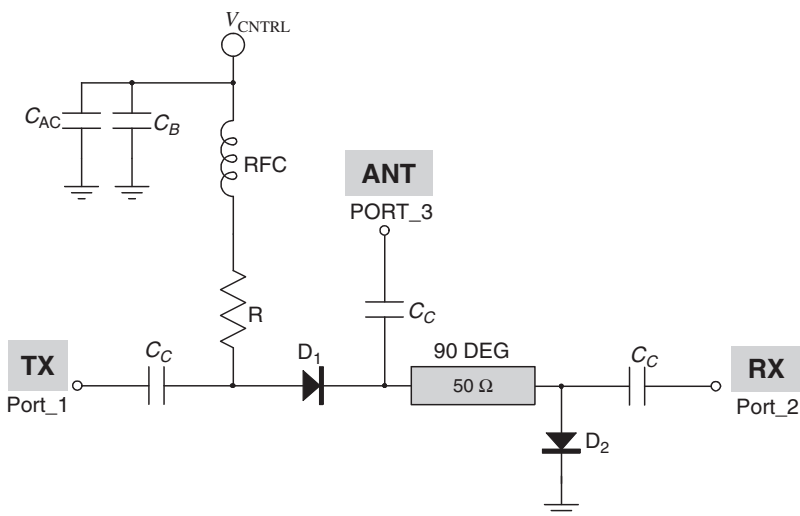


FIGURE 8.24 An SPDT PIN diode switch for microwave frequencies.

50-Ω microstrip that is now shorted through D_2, making it a resonant-shorted stub that acts as an open to the RF. During RX, 0 V is placed at V_{CNTRL}, forcing both PIN diodes *off*, creating an open circuit both at D_1 and D_2, letting the RX frequency easily pass from the ANT port to the RX port, while the ANT signal is being blocked from entering the TX port. Since the 90°, 50-Ω microstrip is now not shorted, it no longer acts as a shorted stub, but as a normal 50-Ω transmission line. Proper microstrip PCB layout procedures are needed to maintain the 90° length of the microstrip and the switch's port-to-port isolation.

To Design

1. Due to parasitic capacitance inherent in D_1, it should be carefully selected to possess a low OFF capacitance to maintain proper transmitter TX isolation when the TX port is *off*. This will also maintain the port isolation and insertion loss when both PINs are biased *off* for the RX stage.
2. To maximize RX switch isolation during TX, D_2 must be selected to be of a low series inductance type.
3. The switching PINs can be operated with a high 20-mA bias current for improved performance. However, 10-mA total forward current is normally sufficient to switch the diodes *on* with minimal distortion of the RF input signal and a low ON resistance. Some PIN diodes may operate quite well with much less bias current, while other PIN types may require much more. Check the diode's data sheet for the appropriate bias current value that will optimize your RF switching circuit's needs. For switching relatively low amplitude RF/microwave input signals from a receiver, the circuit needs only +3.3 V (or higher).
4. RFC = 600 Ω at f_r (can be a lumped surface mount component, or a 90° 100-Ω stub shorted at the top by an RF capacitor).
5. $C_B = C_C = 1\ \Omega$ at f_r
6. $R \approx \dfrac{\text{DC BIAS} - 1.8}{10\ \text{mA}}$ (DC BIAS is in positive volts. R is the current limiting PIN resistor for both diodes.)
7. Microstrip section is 90° long at f_r and 50-Ω wide.
8. $C_B = 0.1\ \mu\text{F}$

A Quick Example Design a SPDT Resonant PIN Switch (Fig. 8.25)

Goal: Create a narrowband microwave SPDT resonant PIN switch. The specifications and parameters for the circuit are:

$f_r = 2.4$ GHz
$V_{CNTRL} = 3.3$ V
$I_D = 10$ mA
$Zo = 50\ \Omega$
Substrate = FR-4, 20-mils thick ($e_r = 4.6$)

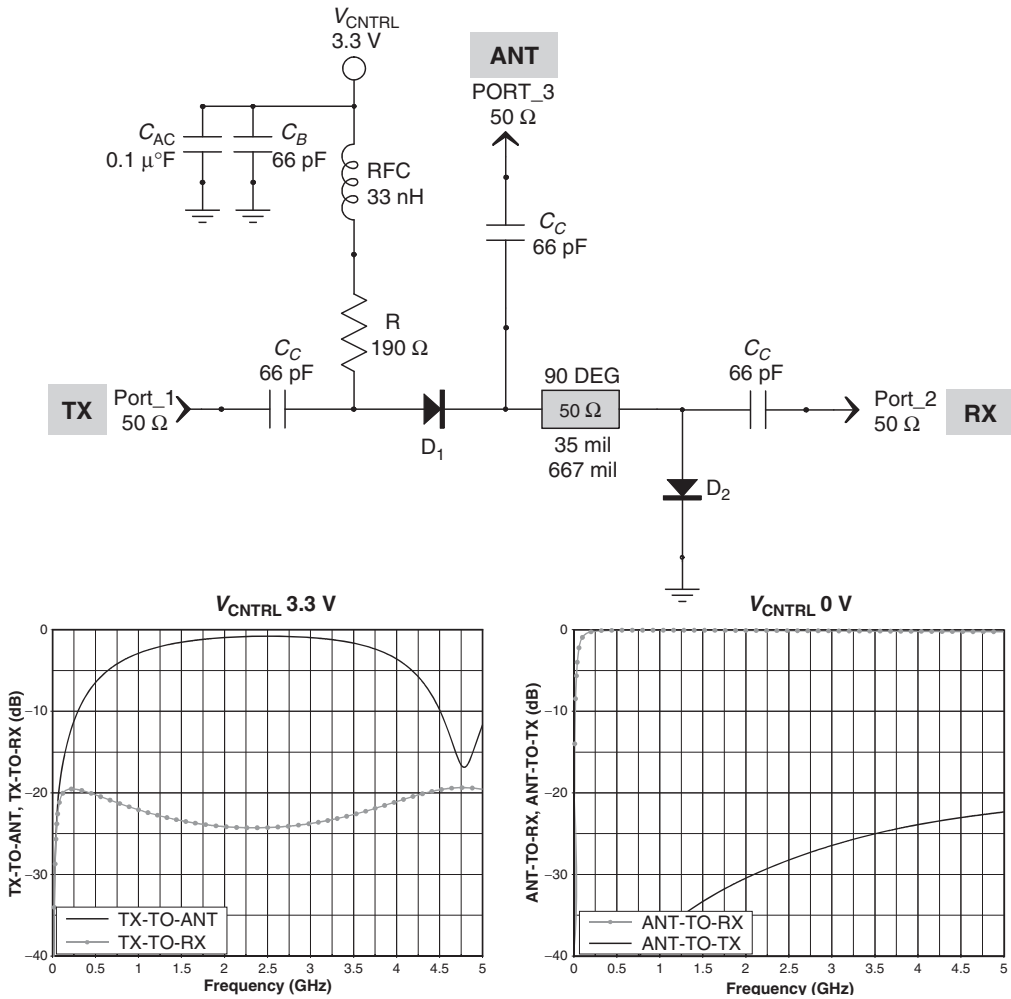


FIGURE 8.25 The example RF SPDT switch circuit with calculated part's values, along with graphed input/output port results at both 3.3 and 0V V_{CNTRL}.

Solution:

1. $R = 190\ \Omega$
2. RFC = 33 nH (500 Ω at 2.4 GHz)
3. $C_C = C_B = 66$ pF (1 Ω at 2.4 GHz)
4. 90° microstrip = 35 × 667 mils
5. $C_{AC} = 0.1\ \mu F$

DC Switch or PIN Diode Driver (Fig. 8.26)

Simple solid-state SPST transistor switches are extremely useful in turning *on* or turning *off* the positive V_{CC} supply to particular stage, such as a high-current amplifier stage or PIN switch. These DC switches, using PNP transistors, are always run in saturation

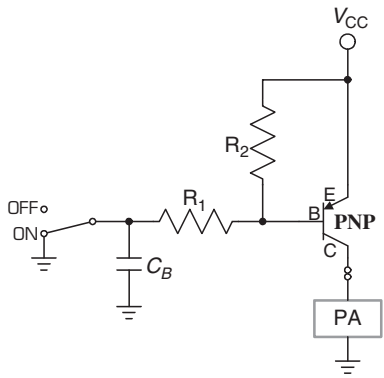


FIGURE 8.26 A solid-state SPST for V_{cc} switching of various amplifier or PIN switch stages.

mode in order to obtain a low R_{on}, which is the DC resistance between the BJT's emitter and collector. Saturation will assure us of a low voltage drop across, and a low-power dissipation within the switching transistor itself. Thus, for an efficient DC switch design, we will want to select a PNP with the lowest possible $V_{CE(sat)}$ rating.

To Design

1. $R_1 = \dfrac{V_{CC}}{(I_{CE} \div H_{FE(min)}) \cdot 1.3}$
2. $R_2 = 10 \cdot R_1$

where R_2 = resistor used to force the PNP into full cutoff, as well as into absolute stability
I_{CE} = load's required current draw, A
$H_{FE(min)}$ = PNP transistor's *minimum* DC current gain
V_{CC} = voltage required by the load to be switched, V

NOTE: *Because of the PNP's collector-to-emitter voltage drop of* V_{CE}*, the true power supply voltage of* V_{CC} *will be lower at the load by approximately 0.15 to 0.3 V, with the exact* V_{CE} *drop dependent on the PNP's* $V_{CE(sat)}$ *rating at the required* I_{CE} *current draw.*

3. C_B = bypass capacitor, with a value that will bypass noise or voltage spikes to ground over a wide range of frequencies, but that will not take a long time to charge up or charge down when the digital control signal arrives.
4. Confirm that the PNP can survive the current draw that must be supplied to the load by calculating the maximum dissipation that it will be subjected to, which must be lower than its maximum data sheet power dissipation rating.

$$P_{max} = V_{CE(sat)} \cdot I_{CE}$$

where P_{max} = actual DC power dissipation across the PNP switch, W
$V_{CE(sat)}$ = collector-to-emitter voltage drop while PNP is in saturation, V
I_{CE} = load's required current draw, A

8.2.3 RF Switch Issues

Increasing the PIN's bias current decreases IMD products, improving linearity and PIN insertion losses. A bias current of 10 mA or less may work well for low-level RF signals, but a value of up to 60 mA may be required for higher-powered applications.

In certain situations an RF switch must supply a good 50-Ω termination to an input signal, even when the switch is in the OFF position. This is referred to as an *absorptive* switch. The 50-Ω termination capability may be required if the high impedance of the OFF switch will cause the RF input power to be reflected back into a sensitive source circuit. This can severely degrade circuit performance, and will actually damage some susceptible stages, especially power amplifiers; or indeed any stage that may become unstable when presented with anything but 50-Ω. Designing a switch that switches to a wideband 50-Ω resistive termination when in the OFF position will force a desirable high return loss, thus minimizing all reflections.

RF mechanical relay switches are still seen in some specialized systems, but are far less dependable than PIN diodes, and much slower, larger, and costly. Relays are, nevertheless, better for certain high-powered, wide-bandwidth, low-insertion loss, and low IMD applications.

For basic RF switching requirements, except when switching heavily reactive loads, a single or double PIN diode configuration as presented above will serve most wireless needs. However, the inductance of a reactive load, such as an RF filter, may have an undesired inclination to resonate with the diode's own OFF capacitance, which can give the PIN switch very little of the OFF isolation that was expected.

8.3 Automatic Gain Control

8.3.1 Introduction

Automatic gain control, or AGC, is found in most modern receivers. This popularity is due to the necessity of increasing the usable dynamic range of a receiver, since without AGC powerful incoming signals would immediately saturate the receiver and create massive distortion, while feeble signals would go virtually undetected by the demodulator. Both scenarios would cause very poor BER in a digital system or unreadable and distorted signals in an analog system.

Bias-based AGC circuits exist thanks to a particular transistor characteristic: The gain of a transistor is increased when we raise the transistor's collector current and, inversely, decreasing the collector current will decrease the transistor's gain. Indeed, we can easily increase the collector current by raising the forward bias at the transistor's base, since increasing the base current will increase the collector current, and thus the gain. As shown in Fig. 8.27, however, a point is soon reached in which this capability will not only level off, but the gain will actually start to decrease slowly with any increase in collector current. The control of the base current is created by the DC bias voltage that is impressed at the base of the transistor by the AGC circuit itself. In fact, many variable gain amplifiers will depend only on this AGC voltage for their entire DC base bias. Because of this capability of a transistor to increase and decrease gain by an external circuit increasing or decreasing its own collector current, we see that there can be two methods of implementing AGC, *reverse* and *forward* AGC. Reverse AGC is by far the most popular, and can be found in the IF sections of many radios. Forward AGC may sometimes be designed into certain front-end RF amplifiers, but is undesirable for general applications since it wastes more collector current than reverse AGC, and has a much more gradual gain response.

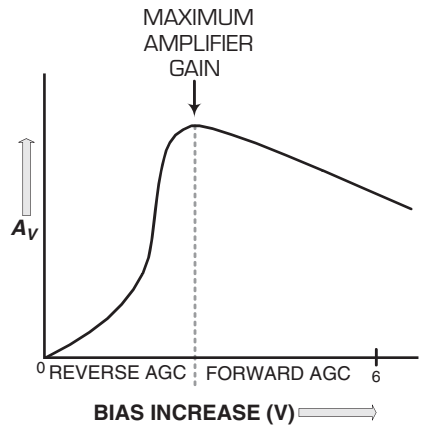


Figure 8.27 A transistor's base-bias voltage versus gain for use in an AGC loop.

With these DC bias-controlled amplifiers care must be taken to confirm that severe distortion does not occur when the gain is varied by the AGC, since the transistor can easily be biased into a nonlinear part of its operation, especially critical if the input signal is of a high amplitude. This issue is not nearly as much of a consideration when AGC amplifiers use voltage variable attenuators at their input, instead of employing AGC bias control, for this gain control function. In fact, many of the newer AGC circuits will feed the RF or IF detected and amplified control voltage back to one or more variable attenuators, which are inserted before fixed-gain amplifier stages (see Sec. 3.11 and Sec. 8.4).

The voltage needed to feed an AGC loop can be tapped off the last IF stage (Fig. 8.28) or, in some receivers, after detection by the detector. As shown in the figure, the IF signal is first tapped from the IF strip's output, RF amplified, rectified to DC, DC

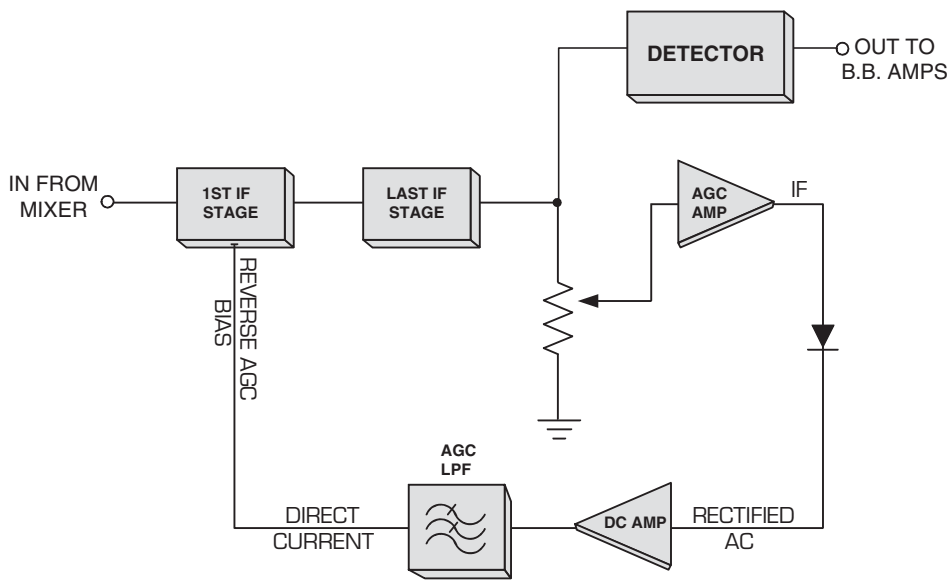


Figure 8.28 A type of AGC loop using the tapped IF signal.

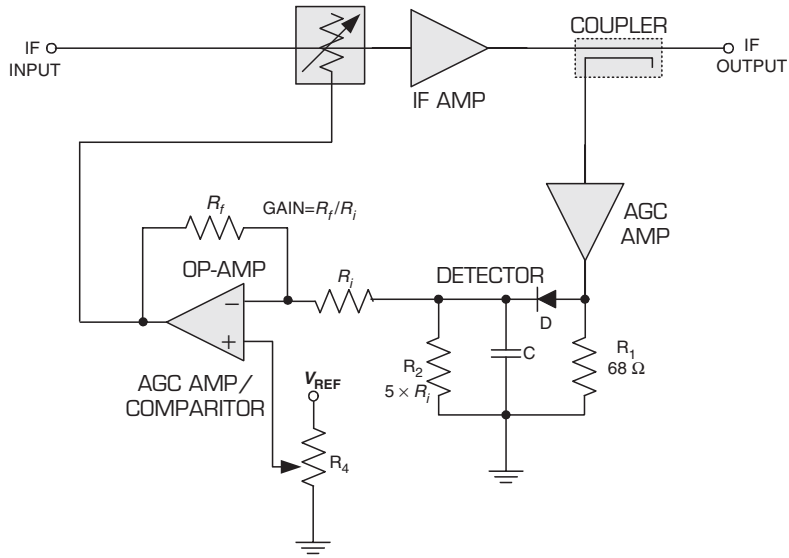


FIGURE 8.29 A common AGC circuit in a receiver's IF chain.

amplified, filtered to a steady DC, and sent to the base of the first, second, and even third IF amplifiers through a trace on the PCB board called an *AGC bias line*.

8.3.2 Automatic Gain Control Design

A complete AGC circuit is shown in Fig. 8.29, and can be designed in various ways. However, the basics still do not change: The signal to be controlled must be sampled, detected, filtered, and placed into a variable gain amplifier in order to change the stage gain, and thus the entire receiver system's total gain; and is ultimately dependent on the input RF signal's strength.

Discussed below are all of the typical stages found in an AGC circuit.

Sampling the Signal

The signal to be controlled can be tapped from the IF by one of two ways. A large resistor that is much higher in value than the 50 Ω of the IF can be exploited to remove a small portion of the signal for feeding the AGC detector, or a directional coupler can be employed to remove a small sample of the signal for AGC detection (see Sec. 8.8.2).

Detecting the Signal

Logarithmic amplifiers, or *log amps*, are adopted in some wireless applications to detect the peaks of the RF signal, then convert these peaks to a logarithmic DC output. These types of log amps are referred to as *demodulating log amps*, and can reach 2.5 GHz at their input, while still maintaining a high dynamic range of greater than 90 dB. One such amp is displayed in Fig. 8.30, and is the detector/amplifier stage in a receiver's automatic gain control feedback loop, while Fig. 8.31 shows the typical log amp's *RF input power versus DC output voltage* characteristic.

The entire log amp-based AGC circuit of Fig. 8.30 functions so: To tap off a small portion of the IF signal into the log amp, R_{COUPLE} is at a value significantly larger than 50 Ω (as stated above, a directional coupler may also be used in this role). The log amp then detects the peak IF, amplifies it, and then converts it to a log-equivalent DC output

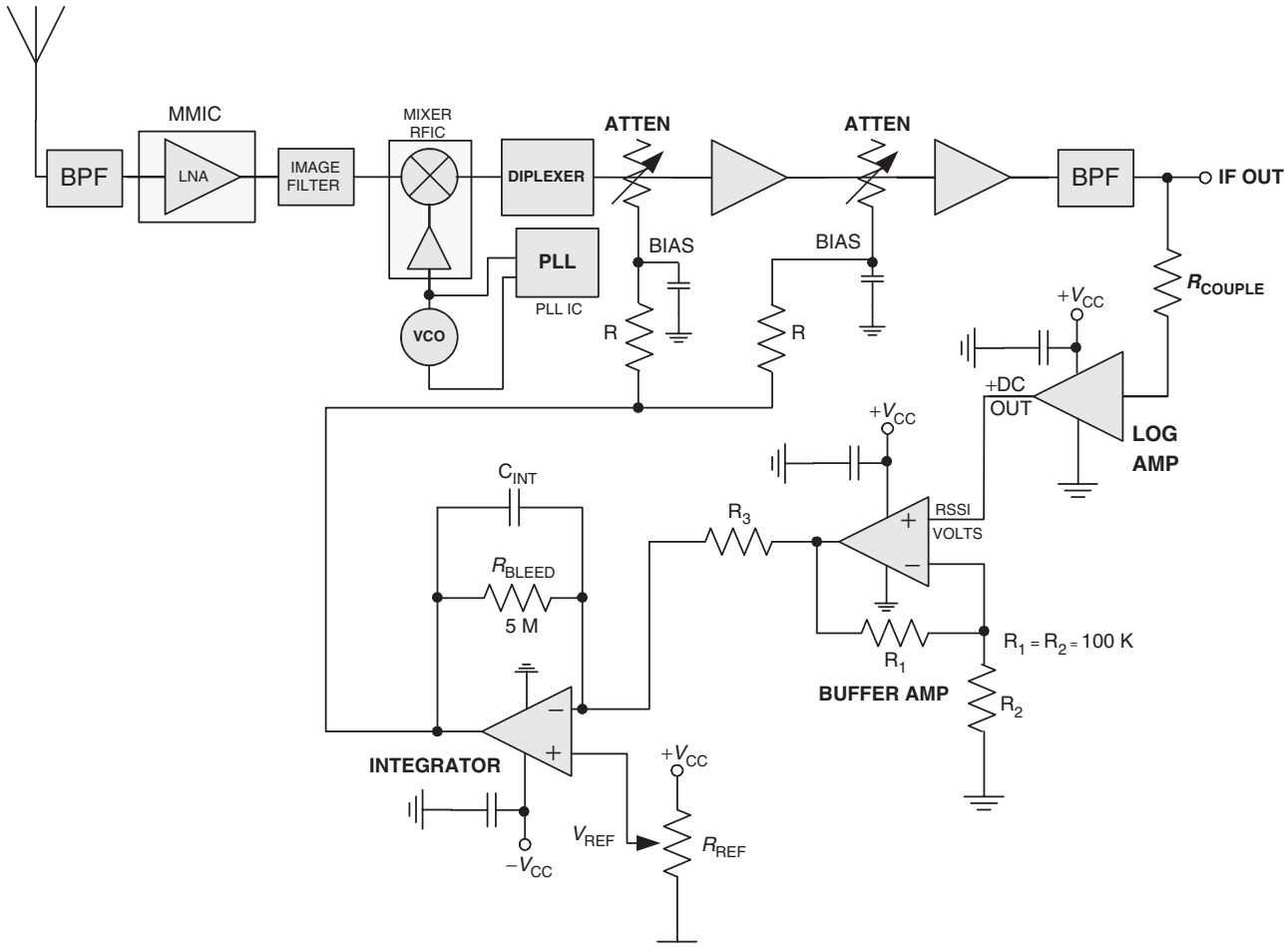


Figure 8.30 A log amp, with integrator and buffer amplifier, used in a receiver's AGC circuit.

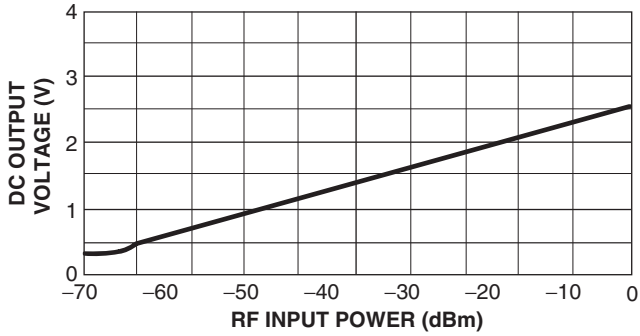


Figure 8.31 A log amp's DC output versus RF input power.

voltage, called the RSSI (*received signal strength indicator*) voltage, which is then placed into the buffer amp, and from there into the integrator, where the voltage is compared to V_{REF}. If the log amp voltage is below this V_{REF}, which is a low input signal condition, then the integrator will output nearly 0 V to the atten, or attenuator. If, however, the voltage from the log amp is above V_{REF}, then a large negative voltage will be placed at the bias inputs to the IF attenuators. Some attenuators and bias-controlled VGAs may require opposite voltages, which can be solved by using an inverting amplifier, along with a positive supply voltage for the integrator.

An AGC *detector diode* is more commonly employed to actually detect the signal out of a coupler at the IF stages (Fig. 8.32). An unbiased diode, such as a self-biased or zero-biased Schottky type, can be utilized as the detector, and converts the IF power to DC. Capacitor *C* is chosen in this diode detector circuit to have a low impedance to the RF, in comparison to diode *D*'s impedance. R_1, used in large signal envelope detectors, presents a proper impedance match for the diode's relatively high input impedance to the 50-Ω coupler's output impedance, as well as supplying a DC return for the diode (input LC matching into the diode is rarely attempted due to its high internal resistance, and thus the need for R_1). If selected, a *small-signal* square-law power detector circuit will replace R_1 with an inductor. The differences between the two types of detectors are explained below. R_2 serves as the load resistor for the RC time constant of the detector, and can be anywhere from 1 kΩ for a large signal envelope detector, to 5 kΩ for a small-signal square-law power detector. (The higher the R_2 value, the higher the output

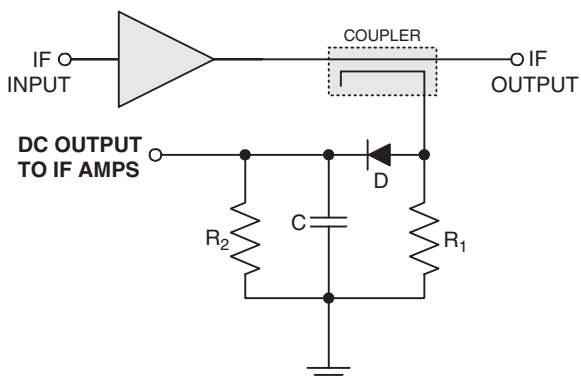


Figure 8.32 An AGC detector circuit, with coupler, in a receiver's IF stages.

amplitude.) And since we are designing with an unbiased detector circuit in which the diode's own junction resistance can be up to a few thousand ohms, its DC output must be placed into a high-input impedance op-amp, or the desired DC output voltage can drop far below design expectations (any Schottky detector's sensitivity will depend on the detector's load being much greater than the diode's own internal resistance R_J, so the input resistance of the next stage should be much greater than this R_J value).

The zero-bias Schottky diode should be chosen if small RF inputs are expected, since they will be much less sensitive to temperature variations than a DC-biased diode arrangement. Biased diodes have been employed to detect low signal levels, as the detector's circuit RF sensitivity can be dramatically improved by a small DC voltage used to bias the diode to a point just before it begins to solidly conduct, which is 0.7 V for silicon. (The detector's RF sensitivity is a measure of the input signal's amplitude in comparison to the output amplitude of the detector, measured in mV/uW, with higher values being better.) This biasing allows for less of an IF input amplitude before the diode detects the signal, and even the "zero-bias" Schottky diodes will need a *small* bias when detecting very low-amplitude RF signals. Without this bias, most of the RF signal's power will be dropped across the high internal junction resistance of the diode, and very little will actually be dropped across the desired load resistance—unless the load resistance is made to be of a very high value.

Since Schottky diode detectors operate in the square-law region at small RF signal input levels, they will detect *power*, while high-input RF signal levels will change the diode's response to a peak *voltage* detection. Therefore, two basic types of RF detectors are utilized in wireless circuits, the *envelope detector* and the *square-law detector*. The envelope detector detects the amplitude of the envelope of the RF signal, outputting a DC voltage that is equal to this value. The square-law detector detects the power of a signal, outputting a voltage that is equivalent to this power. Yet both types of detectors exploit the same basic circuit arrangement as shown in Fig. 8.32, except for minor modifications (the RF small signal power detector uses an input inductor instead of R_1). But since a diode will act as a square-law device at low input powers of between −60 to −30 dBm, and operates as a relatively accurate envelope-to-voltage converter at RF input powers greater than −15 dBm, it can be seen that both detectors may use not only virtually the same circuit, but also the very same diode—and with the exact type of operation only dependent on the actual RF input power.

However, since the envelope detector version can operate with a much lower load resistance than the square-law version and still maintain much of its sensitivity, the envelope detector is the most commonly utilized detector in most AGC circuits. This is because of the lower internal resistance of the detector diode caused by the much higher RF input levels, which force the diode to now be in almost full conduction.

Both kinds of detectors, especially the envelope type, will respond very quickly to an RF signal's increase in amplitude, but will usually have a much slower discharge time caused by the requirement that C discharge through the large value of R_2 and the load. The charge-up is only through the RF source resistance of 50 Ω in series with the diode's ON resistance, and this ON resistance can be relatively low during the IF signal's peaks. Decreasing the RC time constant of C and R_2 can alleviate this problem but, as discussed above, lowering R_2 also decreases detector sensitivity.

Designing a Large Signal Microwave Diode Detector (Fig. 8.33)

R_1 functions as a wide input match and as a DC return for the diode's own self-generated bias current, while the diode D_1 half-wave rectifies the incoming RF input signal. In other words, the RF is converted by D_1 into a rippled DC output voltage by removing one-half of the incoming RF waveform, while this rippled DC charges up capacitor C_1. C_1 is then

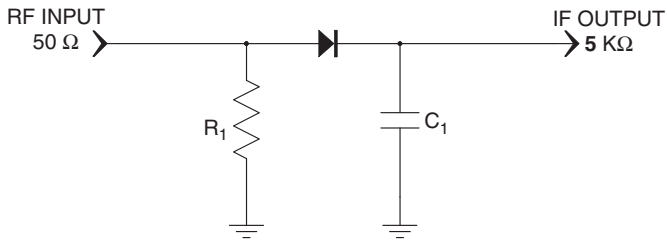


Figure 8.33 A large signal Schottky diode zero-bias detector circuit, with 50-Ω input and a required high-impedance output.

able to discharge through the high resistance of the next stage, which removes the ripple to produce a clean DC level that is a direct indicator of the detector's input RF power amplitude. For instance, with a 1-GHz RF input and a high DC output load resistance of between 1 and 50 kΩ, the 50-Ω RF input power of 0 dBm would produce a DC output voltage of around 0.4 V, and with a +20 dBm RF input, almost 5 V_{DC} would be detected.

To Design

1. Make R_1 equal to 68 Ω.
2. Select an appropriate zero-biased (or even an *n*-type) Schottky diode for the frequency range of interest, the cost targets, and for its high output voltage capability.
3. Choose C_1 with a capacitance that has an RC time constant that is sufficiently below the frequency of the RF input signal (C_1 will be in parallel with the output load resistance). The exact value of C_1 is not critical. For instance, 100 pF would work fine at 1 GHz.
4. Place the diode's Spice-based parameter values, as found in the diode's data sheet, into the enclosed AppCAD's *Diode Parameters* section (under the *Large Signal Detector* circuit selection).
5. Place the appropriate output load resistance that the detector will actually be seeing within AppCAD's *RL* box, and set *Io* (the externally applied bias current) to 0. (The load resistance generates a DC voltage, which is created by the diode's rectification of the RF input signal, and forces the diode to bias itself.)
6. Run the AppCAD simulation to obtain the results of the detector design, such as P_{IN} *versus* V_{OUT} and sensitivity.

Note: *This particular type of detector cannot be accurately used with a power/temperature combination of less than 0 dBm* RF_{IN} *at under 0°C. This is due to the diode's large internal resistance value at these low RF input levels and device temperatures which cause the detector's* V_{OUT} *to rapidly, and inaccurately, drop in sensitivity (unless the load impedance is made impossibly high). At temperatures above approximately 0°C, this should not be a problem. But to increase the sub 0-dBm RF input level sensitivity, and to also improve the detector's output response under low temperatures, we could inject a tiny 10 uA DC bias current via a high-value resistance, which is attached to an external bias voltage. However, this would also drop a DC voltage across the load resistance, which can swamp out the detection of lower amplitude RF detected signals.*

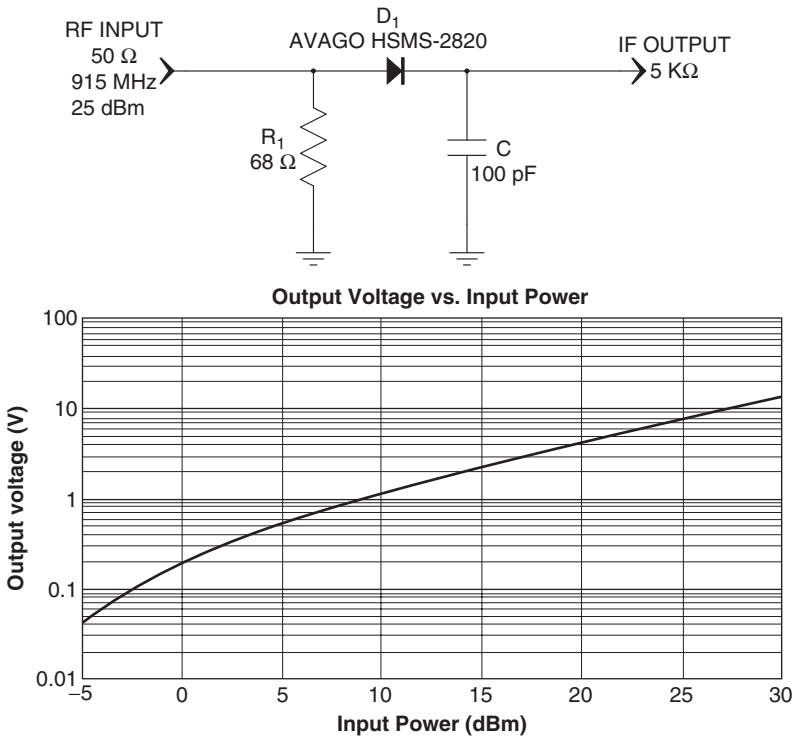


FIGURE 8.34 Example diode detector circuit, with DC output voltage level versus RF input power graph.

A Quick Example Design a Large-Signal RF Detector (Fig. 8.34)

Goal: Create a large-signal detector for the high RF frequencies. The specifications and parameters for the circuit are:

f_r = 915 MHz
Zo = 50 Ω
Z_{LOAD} = 5 kΩ
Diode = Avago HSMS-2820 (Schottky type, appropriate for frequency of interest)

Solution:

1. R_m = 68 Ω
2. C = 100 pF

NOTE: *You can readily simulate this design by opening the enclosed AppCAD software, and entering:* $I_O = 0$, $IS_O = 0.022$ *uA*, $R_S = 6$ Ω, $EG = 0.69$ eV, $N = 1.08$, *which are the appropriate parameter values for this particular Avago diode. Run the simulation. The results should be as shown in the graph.*

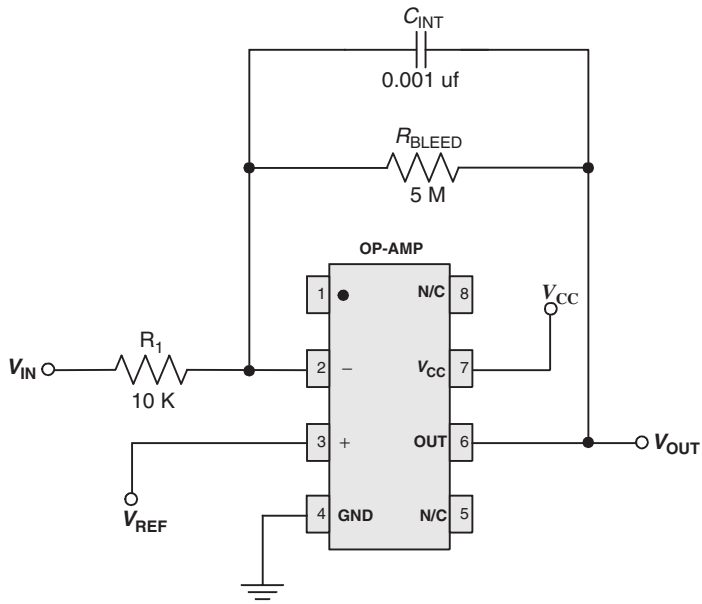


Figure 8.35 Single-supply op-amp designed to function as an integrator.

AGC Amplifiers and Integrators

DC amplifiers are normally needed to increase the AGC level of the DC signal that is fed into the gain-adjust port of the VGA. RF amplifiers may also be required on the AGC's IF end to increase the signal level into the detector. Raising the AGC's DC output level can be accomplished with standard DC amplifiers, or with the circuit as shown in Fig. 8.35, which is a single-supply op-amp set to function as an *integrator*. In this circuit, a smaller C_{INT} or R_1, or a faster change in V_{IN}, will speed up the change in V_{OUT} into the VGA's gain-adjust port. When the voltage at the inverting input is more positive than the voltage at the noninverting input, the output voltage will ramp down; if the voltage at the inverting input is more negative than at the noninverting input, the output will ramp up. The speed of the ramp up/down will depend on the RC time constant of the RC components. A large value resistor (R_{BLEED}) of approximately 2.2 MΩ or higher is usually placed in parallel with the integrating capacitor, since all op-amps have a small input bias current and random noise which will quickly charge up this small-value capacitor, and the resistor simply bleeds the current away. However, the resistor must be significantly higher in value than the input resistor R_1, or gain can be lowered excessively. Figure 8.30 demonstrates the integrator within a standard AGC circuit.

Variable Gain Amplifiers

The variable gain amplifiers in the IF strip can be of either the variable attenuator type or the variable bias type. Design of the appropriate attenuator and amplifier are explained in other sections of this book.

8.3.3 Automatic Gain Control Issues

If there are any IF filters between the AGC output coupler and the gain controlled IF amplifiers that have excessive group delay, this can cause time impediments in sensing that a signal has increased or decreased in amplitude, possibly causing AGC loop instabilities. Another major cause of AGC loop instability can be the desire to obtain excessively tight amplitude control of the IF or baseband output amplitude, since in most cases an AGC circuit with less gain and less absolute control over the output amplitude will be more stabile. It is still possible, nonetheless, to safely design and construct an AGC loop that can control the output signal to within 1 dB or better.

If the receiver's LNA stage is to be AGC controlled, it is almost always critical to maintain its noise figure within reasonable limits. Unfortunately, any AGC action will naturally decrease the receiver's NF, but by adding a *delay diode* in series with the AGC's bias line, the start of gain control to the front-end amplifiers can be postponed slightly. This will permit the LNA to maintain its NF and gain, and therefore the noise performance of the entire receiver, until absolutely necessary. Even if there will not be an AGC connection to the LNA, or to any other RF amplifier, a delay diode is still a good idea for the first IF gain-controlled amplifier of the IF strip, since this will help in some small way in maintaining a superior system noise figure.

All AGC-controlled amplifiers should be decoupled from each other by a small value of series resistor (such as 100 Ω), along with a ceramic capacitor to ground, at each VGA's DC gain control port. This will slow any undesired interactions from taking place between all of the gain-controlled stages.

In the early prototype stages of an AGC circuit's optimization, it is a good idea to utilize trimmers in place of fixed-value resistors in order to allow the AGC's loop circuit to be easily empirically optimized.

8.4 Attenuators

8.4.1 Introduction

Attenuators are either fixed or variable circuits used to reduce signal amplitudes and/or to improve return loss, while also maintaining the proper input and output impedance (normally 50 Ω) of the stages they are attached to. Indeed, RF attenuators of all types are used extensively in wireless design.

Shown in Fig. 8.36 is a *step* (or *variable*) attenuator employed for bench testing of wireless circuits, and is capable of varying its attenuation in discrete steps by manually turning a

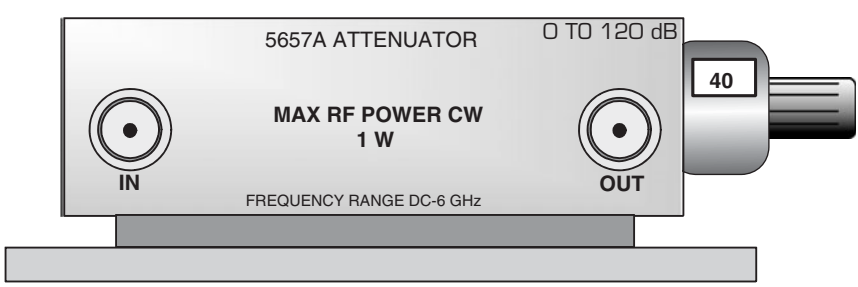


Figure 8.36 A common manually switched step attenuator.

FIGURE 8.37 A coaxial in-line fixed attenuator with SMA connectors.

knob, or by electronic control. Other variable attenuators, such as small, surface-mount types, can be inserted between stages on a PCB, and can be either analog voltage or current controlled with infinite attenuation resolution. Or, with similar SMD-packaged digital step attenuators, there may only be a limited number of discrete possible attenuation steps available.

SMA or BNC miniature coaxial in-line fixed attenuators (Fig. 8.37), for testing or general signal attenuation, are available at various values of up to 60 dB, with a maximum safe power dissipation of 25 W. (All such attenuators will be rated for the exact value of attenuation they are capable of, such as 5 dB, 10 dB, 30 dB, and so on, along with the maximum frequency of operation and the maximum recommended input signal level.)

Integrated circuit solutions for variable and fixed attenuators are readily available, but their cost, performance, and size are usually inferior to the discrete designs. However, for small production runs and for low RF signal levels, either IC analog or digital variable attenuators can sometimes be the best choice.

8.4.2 Attenuator Design

The following low-cost analog and digital attenuator circuits, as well as the fixed pad attenuator types, can quickly be designed and built for many RF applications.

Positive Voltage, 50-Ω Absorptive 1-Bit Digital Attenuator, for up to 2.5 GHz (Fig. 8.38)

This low-cost 1-bit 50-Ω attenuator can be used between a receiver's first RF filter and LNA, and will improve the receiver's dynamic range in the presence of a strong wanted signal or a close-in interferer. In this particular circuit it is critical to use very low insertion loss SPDT switches as the switching elements, since any losses in front of the LNA subtract directly from the receiver's noise figure. The insertion loss of this circuit, not counting PC board and trace losses, is 2 · IL of the insertion loss of a single switch.

To Design

1. Select the appropriate low-insertion loss, high port-to-port isolation RF SPDT switch. Hittite Microwave Corporation manufactures a GaAs switch with two SPDT switches within a single package that is perfect for this particular application, the model HMC199MS8.
2. The PI-attenuator section of R_1, R_2, and R_3 can be designed to be of any desired value up to a limit of 20 dB, and can easily be calculated by the formulas in Sec. 8.4.2, or quickly and automatically synthesized in the included *Qucs* simulator.
3. C_C, the DC blocking capacitors, are normally required for most integrated RF switches, and should be either at the SRF (series resonant frequency) of the particular capacitor model, or have a reactance under a few ohms at the input RF frequencies.
4. Keep the PCB layout for this circuit very tight, and expect that at higher frequencies (> 1 GHz) and/or attenuation values (> 20 dB), the actual attenuation obtained

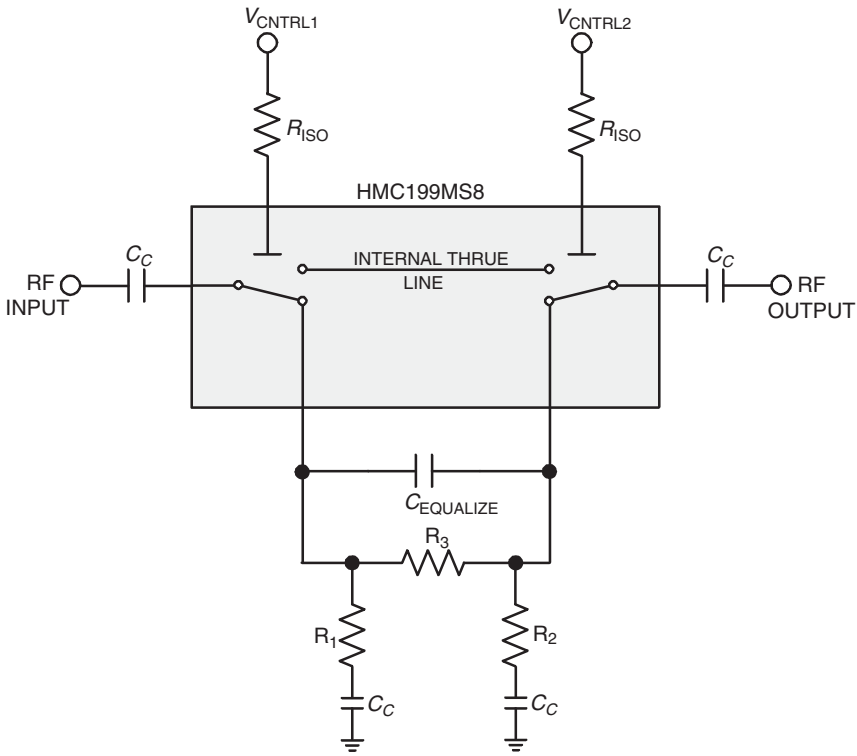


Figure 8.38 A 1-bit digital attenuator circuit.

will vary somewhat from that calculated. (This effect is caused by the different phase shifts between the two disparate RF path lengths causing different phase shifts at the combined output port, with each signal then possibly being completely out of phase with each other over a certain frequency band. Circuit and component parasitics also add to the variation in obtained attenuation. Therefore, when the attenuator pad is switched in-line, the no-loss feedthrough of *Port A* will still pass some of the RF signal to the output port of the attenuator circuit of *Port B*. Indeed, the level of switch isolation may actually approach that of the attenuation of the pads, causing severe amplitude variations over frequency due to the input and output phase recombination of both RF paths at the RF output of the attenuator circuit at higher frequencies.)

5. Add $C_{EQUALIZE}$ is to equalize the attenuation characteristics over frequency. Start with a value of 0.3 pF, and increase as needed.
6. R_{ISO} resistors isolate the control lines from the switch itself.

Positive Voltage Controlled GaAs Linear IC Attenuator for up to 2.5 GHz (Fig. 8.39)

This excellent absorptive, 50-Ω Skyworks analog attenuator, Model AV104-12, is controlled with a positive voltage (unlike many other attenuator types, which require a

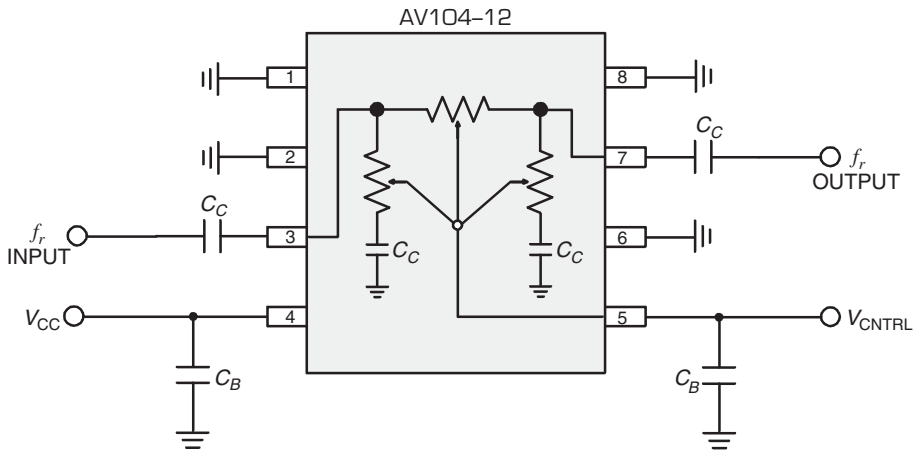


Figure 8.39 An absorptive GaAs linear attenuator integrated circuit.

negative voltage), has a 25-dB attenuation range that is highly linear over its control voltage V_{CNTRL} range, possesses a high IP3 of 20 dB at 1 GHz, consumes only 5-mA current during operation, maintains a great return loss over any V_{CNTRL} and attenuation range, comes in a low-cost plastic surface-mount package, and is an excellent choice for use in almost any low-powered (< 100 mW) application.

To Design

1. C_C = < 2 Ω at f_r
2. C_B = 0.01 µF
3. V_{CC} = 5 V
4. V_{CNTRL} = 0 to 5 V (attenuation: 0 V = 25 dB; 2.5 V = 13 dB; 5 V = 3 dB)

Fixed Attenuator Design (Figs. 8.40 and 8.41)

To design a 50-Ω pad for any attenuation value, first calculate the value of the attenuator's resistors with the following equations, then select the proper resistors for the maximum power dissipation expected.

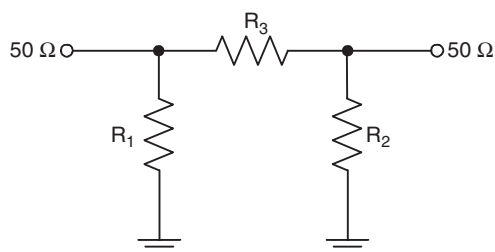


Figure 8.40 A 50-Ω PI attenuator.

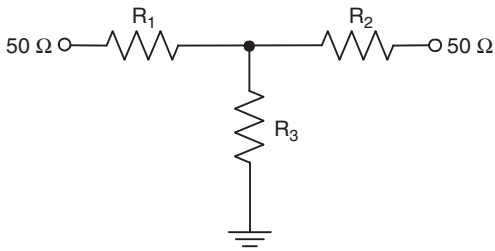


FIGURE 8.41 A 50-Ω "T" attenuator.

To Design

1. Fixed PI attenuator of Fig. 8.40
   a. $\chi = 10^{\frac{\text{LOSS [in dB]}}{10}}$
   b. $R_3 = 0.5(\chi - 1) \cdot \sqrt{\frac{2500}{\chi}}$
   c. $R_1 = \frac{1}{\frac{\chi + 1}{50(\chi - 1)} - \frac{1}{R_3}}$
   d. $R_2 = R_1$
2. Fixed T attenuator of Fig. 8.41
   a. $\chi = 10^{\frac{\text{LOSS [in dB]}}{10}}$
   b. $R_3 = \frac{2 \cdot \sqrt{2500 \cdot \chi}}{\chi - 1}$
   c. $R_2 = \left(\frac{\chi + 1}{\chi - 1} \cdot 50\right) - R_3$
   d. $R_1 = R_2$

8.5 Baluns

8.5.1 Introduction

A balun transforms a balanced line or circuit into an unbalanced line or circuit, since it is sometimes necessary to have a balanced output from an unbalanced amplifier for interfacing with other stages. The reverse is also true, and a balun can be used to interface an unbalanced to a balanced stage. Impedance matching may also be required.

A balanced stage's input or output is comprised of two parallel conductors with two input lines, one with a 0° signal and the other having the same amplitude signal but 180° phase shifted, and with each conductor having equal currents flowing in opposing directions (Fig. 8.42*a*). When a balanced source must be converted to unbalanced, this

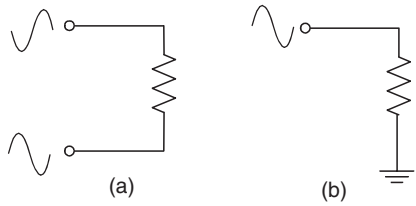


Figure 8.42 (*a*) A balanced and an (*b*) unbalanced circuit.

demands that the two differential signals be mixed (combined) in-phase in order to output an unbalanced signal.

Unbalanced outputs will have a single conductor for the current, with a second conductor used for the ground return, and is the dominant technique found in most of RF design (Fig. 8.42*b*).

For antenna use, baluns may be purchased in connectorized weather resistant packages, in which they are utilized for placing an unbalanced signal from the coax transmission lines into a balanced dipole antenna, while also matching any impedance variations. If a balun was not used in this situation, RF currents on the center conductor of the coax would pass to one leg of the dipole, while the RF current on the ground conductor would pass to the other leg of the dipole. This would result in RF radiating from the coax's ground shield, causing EMI. These balun structures, at HF frequencies, can be as simple as a wideband, untuned transformer.

8.5.2 Balun Design

Presented below are three designs, one a lumped balun for lower frequencies and the other two for narrow and wideband microwave use as distributed circuits. Employing an off-the-shelf ceramic surface-mount balun, instead of the lumped or distributed type, will simplify a design, and also offer very dependable repeatability over all frequencies and large production runs. However, this off-the-shelf solution will be more expensive.

Narrowband Lumped Balun for up to 1 GHz (Fig. 8.43)

A simple discrete balun structure can be used instead of a ceramic or distributed balun when frequencies are under 1 GHz. The lumped balun is comprised of L_1, L_2, C_1, C_2, C_3, and C_4. However, due to the required amplitude and phase balance needed in an efficient, low-loss balun, high tolerance passives should be used, as well as very careful attention to a symmetrical PCB layout.

To Design

1. $L_1, L_2 = \dfrac{\sqrt{R_U \cdot R_B}}{6.28 \cdot f}$

2. $C_1, C_2, C_3, C_4 = \dfrac{1}{(6.28 \cdot f) \cdot \left(\sqrt{R_U \times R_B}\right)}$

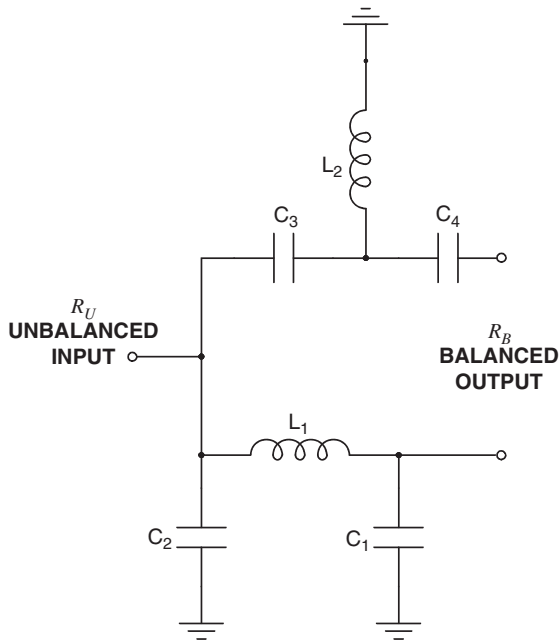


FIGURE 8.43 A lumped balun circuit.

where R_U = unbalanced source or load resistance, Ω
R_B = balanced source or load resistance, Ω
f = center frequency, Hz

A Quick Example Design a Lumped RF Balun (Fig. 8.44)

Goal: Create a narrowband LC balun for RF frequencies. The specifications and parameters for the circuit are:

f_r = 433.92 MHz
R_U = 50 Ω
R_B = 100 Ω

Solution:

1. L_1 = 25.9 nH
2. L_2 = 25.9 nH
3. C_1 = 5.2 pF
4. C_2 = 5.2 pF
5. C_3 = 5.2 pF
6. C_4 = 5.2 pF

Wideband Microwave Distributed Balun (Fig. 8.45)

This planar structure is relatively broadband while maintaining the required 180° phase shift between Ports 2 and 3 for balanced operation. When five sections are used, a 30% bandwidth is possible (more sections equal wider bandwidths).

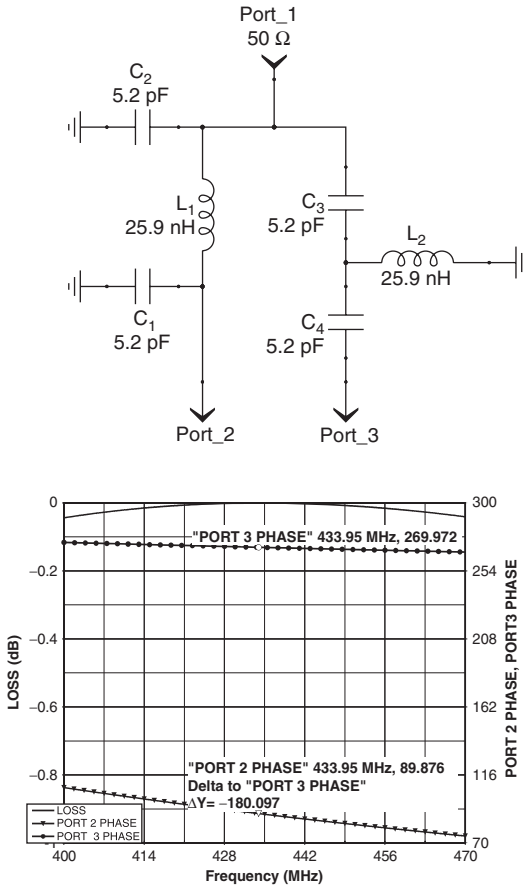


Figure 8.44 The example lumped 433.92-MHz balun circuit with calculated part's values, along with graph of 180° Port 2 to Port 3 phase difference.

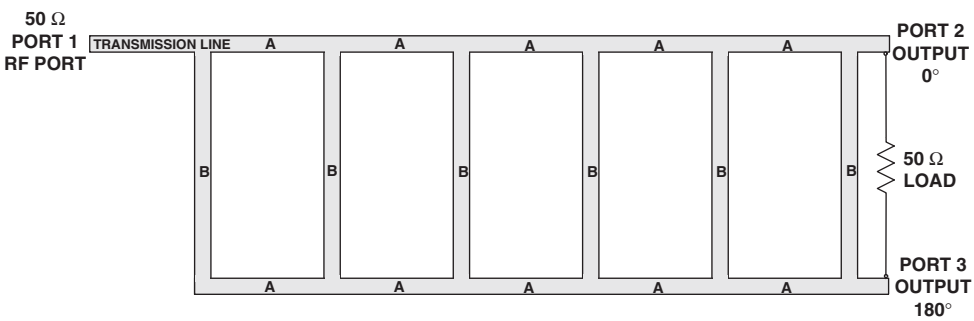


Figure 8.45 A distributed wideband balun structure for microwave frequencies.

To Design

1. $A = \frac{\lambda}{4}$
2. $B = \frac{\lambda}{2}$

where A = length of 50–Ω microstrip (Calculate as described in Sec. 1.2, "Microstrip Design".)
B = length of 50-Ω microstrip (Calculate as described in Sec. 1.2, "Microstrip Design".)

NOTE: *Attach balanced load directly to balun output, or final series section will disturb balun's frequency response and phase.*

A Quick Example Design a Distributed Wideband Balun (Fig. 8.46)

Goal: Create a wideband microwave balun. The specifications and parameters for the circuit are:

f_r = 5.8 GHz
Sections = 5
Substrate = RO-4003, 20-mils thick

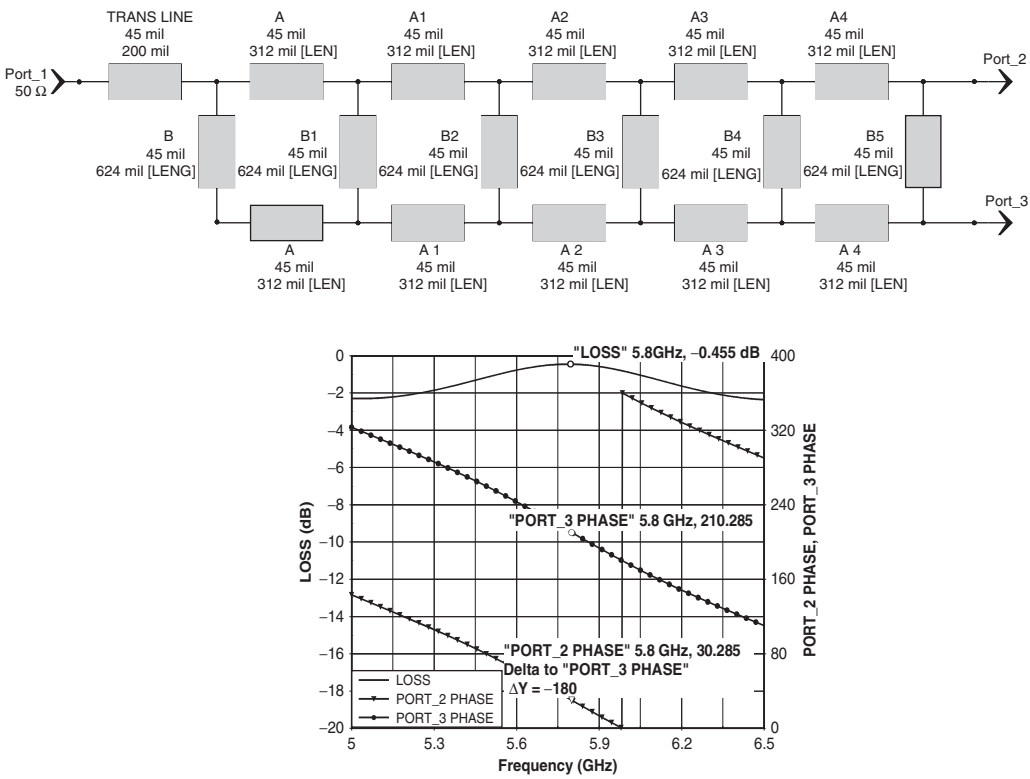


FIGURE 8.46 The example wideband distributed 5.8-GHz balun structure with calculated part's values, along with graph of 180° Port 2 to Port 3 phase difference.

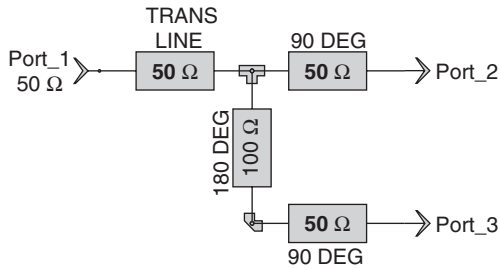


FIGURE 8.47 A distributed narrowband balun structure for microwave frequencies.

Solution:

1. A = 312 mils
2. B = 624 mils
3. Width = 45 mils

Narrowband Distributed Balun (Fig. 8.47)

This very narrowband balun uses the 180° line to delay the Port 1 input frequency before it reaches Port 3, as compared to the signal that passes through to Port 2, thus providing an exact 180° phase shift between the two output ports, and an unbalanced-to-balanced conversion.

To Design

1. The input TRANS LINE to be 50 Ω, and of any convenient length.
2. 180° line to be exactly a half-wavelength long at the frequency of interest, and 100-Ω wide.
3. Each 90° line to be exactly a quarter-wavelength long, and 50-Ω wide.

NOTE: *Attach the balanced load directly to the balun's output ports, without an intervening microstrip, or the 90° traces will no longer be a quarter-wavelength long.*

A Quick Example Design a Distributed Narrowband Balun for Microwave Frequencies (Fig. 8.48)

Goal: Create a narrowband microwave balun. The specifications and parameters for the circuit are:

f_r = 5.8 GHz
R_U = 50 Ω
R_B = 50 Ω
Substrate = RO-4003, 20-mils thick

Solution:

1. Input transmission line = 45-mils wide, any length
2. 90° line = 45-mils wide, 312-mils long
3. 180° line = 10-mils wide, 657-mils long

NOTE: *The balanced load must be connected* directly *to output of the balun.*

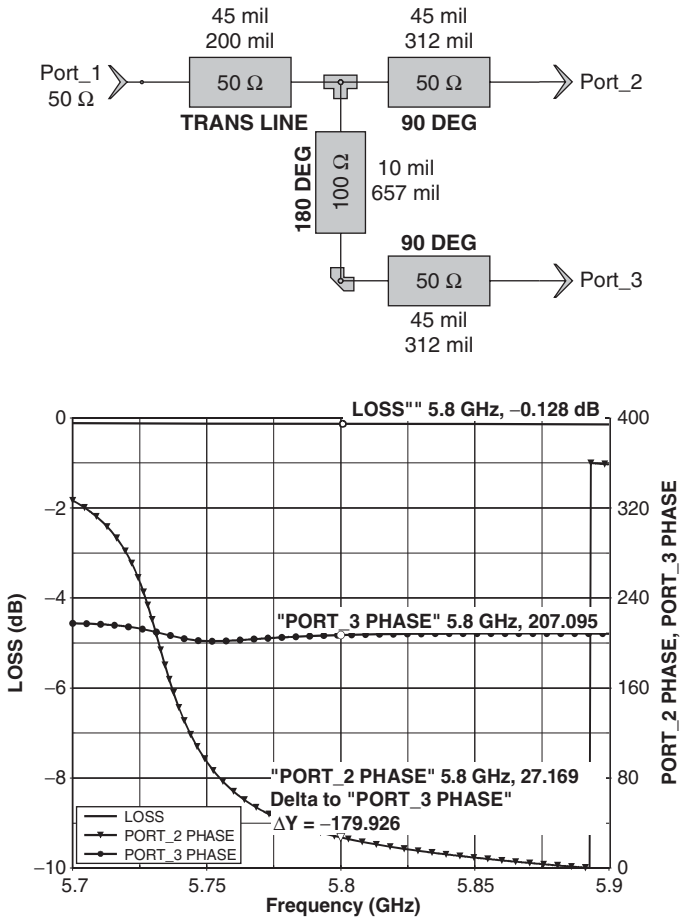


Figure 8.48 The example narrowband distributed 5.8-GHz balun structure with calculated part's values, along with graph of 180° Port 2 to Port 3 phase difference.

8.6 Splitters and Combiners

8.6.1 Introduction

Splitter and combiners are 50-Ω circuits used to "mix" different signals in a linear manner, leaving them unchanged with no new signals created. For a splitter, a signal is placed at the input, and two or more signals, usually of equal amplitude and phase, are removed from two or more separate ports at the output. For a combiner, two or more signals of equal phase are placed at the input, and a single signal is removed from the one output port that is equivalent to their vector sum. In fact, a splitter and a combiner are exactly the same circuit: to turn a splitter into a combiner, the circuit is merely reversed, with the input port becoming the output port.

The insertion loss of a combiner will be quite small, less than 0.4 dB, for two in-phase, same-frequency signals that are placed at the input. However, the insertion loss

for a two-way (three total ports) combiner will be 3 dB if the two signals are at different frequencies. With the exact same circuit reversed and setup as a two-way splitter, the insertion loss will be approximately 3.5 dB, with 3 dB of this loss simply due to the signal being split into the two output ports.

8.6.2 Splitter and Combiner Design

Three easy-to-design circuits are presented for almost any RF splitter or combiner wireless application.

50-Ω LC Power Splitter/Combiner, 0° (Fig. 8.49)

This LC circuit will have a lower insertion loss (3.5 dB) than a resistive splitter, while also maintaining at least a 20-dB isolation between output ports. However, this circuit will have a significantly lower bandwidth (BW = 20%) than a resistive design, since it is reactive.

To Design

1. $L = \dfrac{50}{4.4 \cdot f}$
2. $C = \dfrac{1}{445 \cdot f}$
3. $C_2 = 2 \cdot C$
4. $R = 2 \cdot Z_{IN}$

A Quick Example Design an LC RF Splitter (Fig. 8.50)

Goal: Create a lumped 0° LC splitter for high frequencies. The specifications and parameters for the circuit are:

f_r = 915 MHz
Z_{IN} = 50 Ω
Z_{OUT} = 50 Ω

Solution:

1. L = 12.4 nH
2. C = 2.5 pF

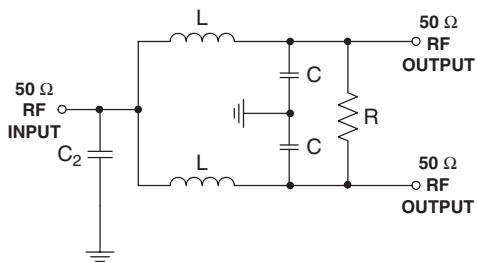


Figure 8.49 An LC discrete RF signal splitter.

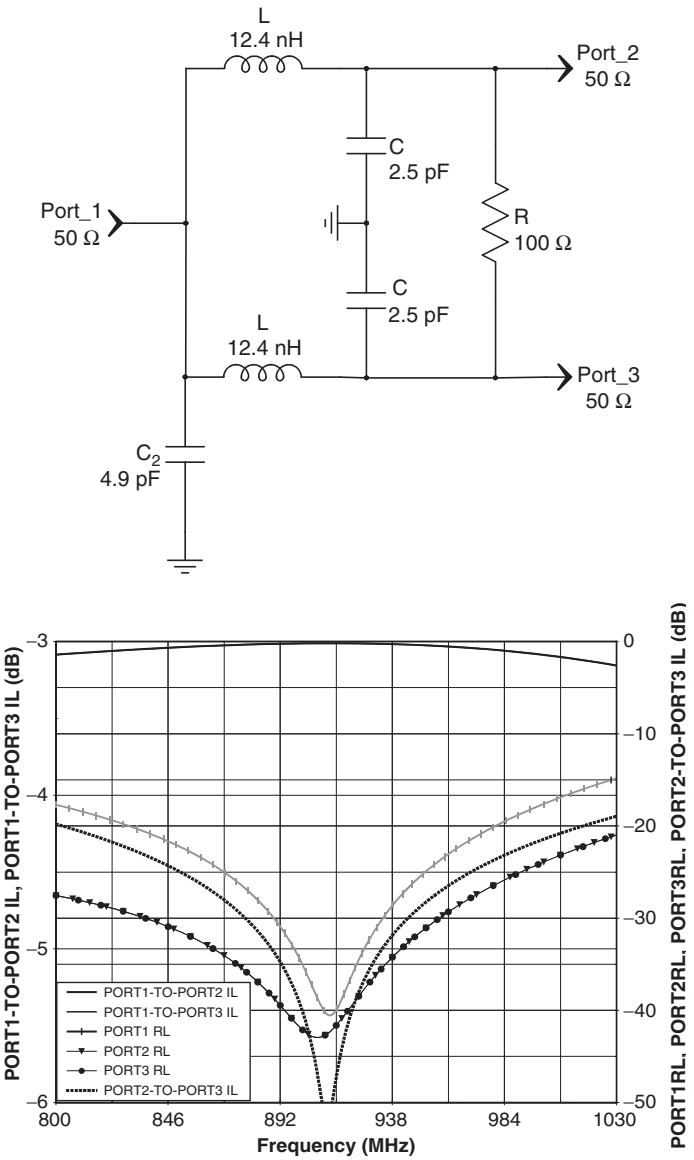


Figure 8.50 The example narrowband LC splitter with calculated part's values, along with port-to-port insertion loss (IL) and return loss (RL) results.

3. $C_2 = 4.9$ pF
4. $R = 100\ \Omega$

50-Ω LC Power Splitter, 90° (Fig. 8.51)

Similar to the above splitter, but has a 90° phase difference between its two output ports. It also has a 20% bandwidth.

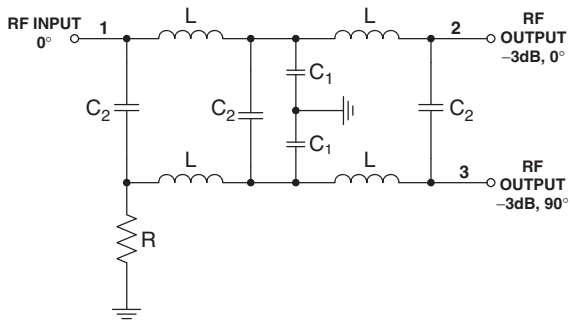


FIGURE 8.51 An LC discrete RF signal splitter with 90° phase shifts between each output port.

To Design

1. $R = 50\ \Omega$
2. $L = \dfrac{R}{2\pi f}$
3. $C_1 = \dfrac{1}{2\pi fR}$
4. $C_2 = 0.5\left(\dfrac{1}{2\pi fR}\right)$

50-Ω Resistive Splitter/Combiner, 0° (Fig. 8.52)

This simple resistive splitter is extremely wideband, but has a high insertion loss of 6 to 7 dB, along wit h a low isolation between ports that is the same level as the insertion loss.

To Design

1. $R = \dfrac{50}{3}$

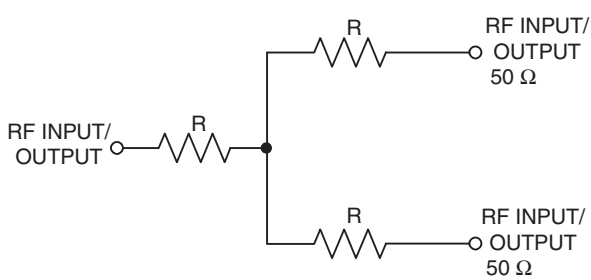


FIGURE 8.52 A very wideband 50-Ω signal splitter.

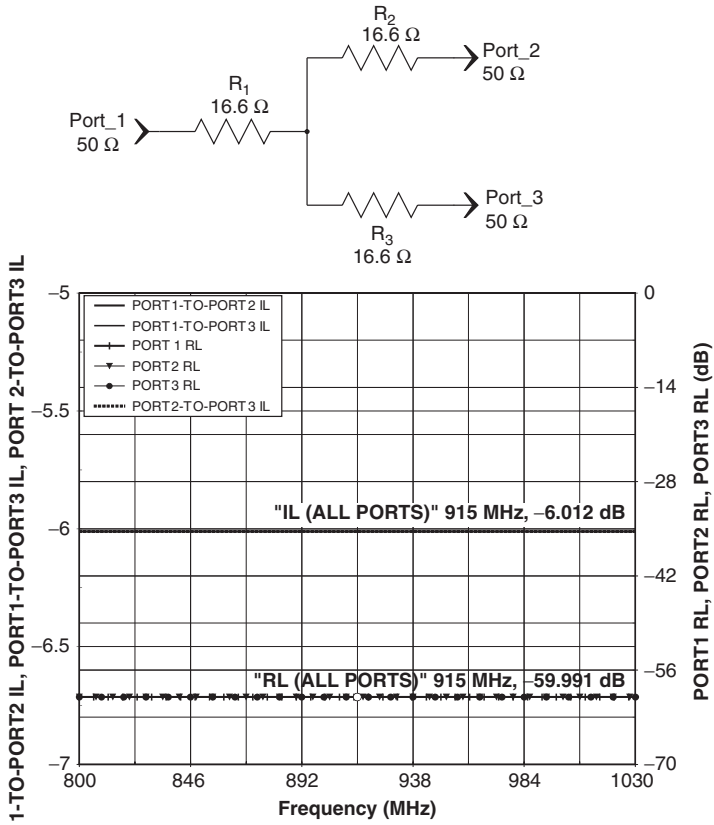


FIGURE 8.53 The example wideband resistive splitter with calculated part's values, along with port-to-port insertion loss (IL) and return loss (RL) results.

A Quick Example Design a Resistive Wideband Splitter (Fig. 8.53)

Goal: Create a lumped wideband splitter for high frequencies. The specifications and parameters for the circuit are:

f_r = 52 to 915 MHz
Z_{IN} = 50 Ω
Z_{OUT} = 50 Ω

Solution:

1. R = 16.66 Ω

8.7 Power Supplies

8.7.1 Introduction

Most wireless communications equipment run off DC power supplies that obtain their energy from the AC mains. In order to provide power at a constant and regulated voltage, this AC main's voltage is converted by a power supply into the required DC levels needed

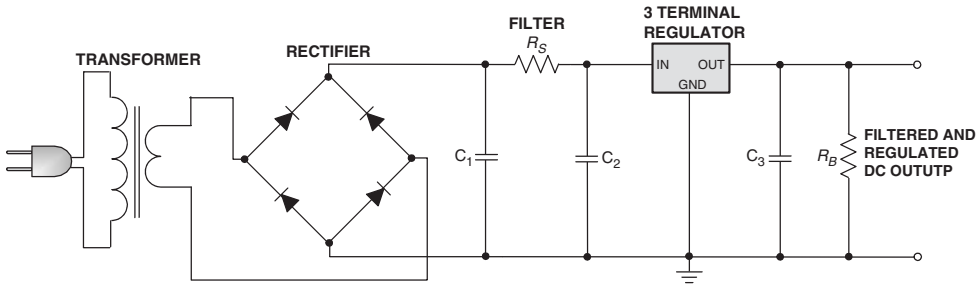


Figure 8.54 A complete linear power supply.

by the RF system. Batteries cannot be used for power in many applications, since they will only furnish current for a limited period of time, and are thus reserved for certain devices, such as portable equipment, that will not consistently or conveniently be near AC sources.

The basic power supply is shown in Fig. 8.54, and consists of a two- or three-pronged plug, a transformer, a rectifier, a lowpass filter, and a regulator. The transformer changes the AC mains voltage of 120 V_{RMS} into any desired voltage, either up or down; the bridge rectifier circuit turns the AC into a pulsating DC; the lowpass filter converts the varying DC into a steady DC; the three-terminal regulator maintains the output voltage within tight specifications.

In the figure, C_3 suppresses any output oscillations and helps in regulation, with R_B being a *bleeder resistor* that drains a fixed current from the regulator to help stabilize the output voltage, as well as drain hazardous voltage levels from the filter capacitors when the power supply is shut off. More on each of these circuits is given below.

Transformer Since a transformer conveys AC energy from one circuit to another by electromagnetic induction, we can increase or decrease the current or voltage by changing the ratio of the windings between the primary and the secondary. A low-frequency transformer is made up of a primary coil which obtains energy from an alternating current source. The primary's expanding and contracting magnetic flux lines flow through a core made of steel plates, which concentrates this flux with the least amount of losses. The secondary coil is cut with the primary's flux lines, inducing an AC voltage, producing a current that flows through the transformer's load.

Rectification Rectification is the first step in obtaining a smooth DC output voltage. AC power can be changed into pulsating DC by employing one of three general rectification circuits.

The most basic rectification technique is called *half-wave rectification* (Fig. 8.55), which has a peak voltage that is almost equal to the input AC peak voltage, while also

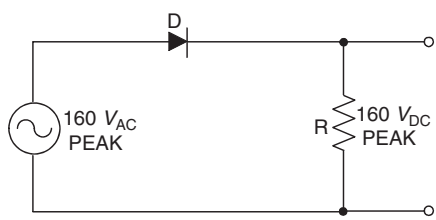


Figure 8.55 A half-wave rectifier circuit.

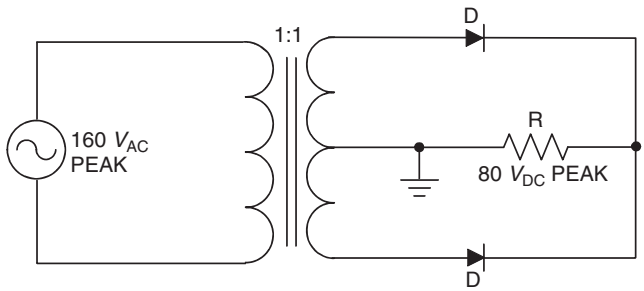


Figure 8.56 A full-wave rectifier circuit.

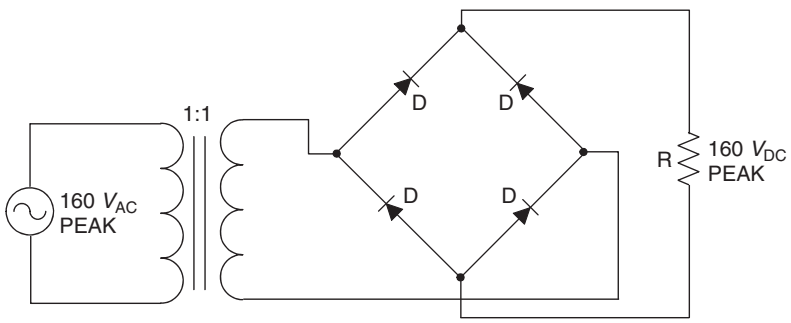


Figure 8.57 A bridge rectifier circuit.

demanding few components (a single diode). However, this method gives us a difficult-to-filter 60-Hz output.

The second method is *full-wave rectification* (Fig. 8.56), and has a simple-to-filter 120-Hz output. Unfortunately, only half of the input's peak AC voltage is actually available to the load due to the transformer's center tap.

The third, and most dominant, method in modern power supplies is *bridge rectification* (Fig. 8.57), which not only furnishes us with an easy-to-filter 120-Hz output, but also the full input AC peak voltage level is available at the output.

Filtering In order to smooth out the pulsating DC power that results from rectification, a lowpass power supply filter is necessary, as any amplitude variations would be unacceptable for most electronic circuits. Filtering is used to eliminate this pulsating current, while giving us a constant, almost ripple-free output voltage.

One such filter that will attenuate any AC component riding on the rectified DC output is shown in Fig. 8.58. C_1 filters the majority of the ripple, with R_S and C_2 functioning as an AC voltage divider. The small amount of ripple left over from C_1 will be dropped across RS, with very little across C_1. This is due to R_S's increased resistance over C_2's decreased reactance to the relatively high ripple frequency. Swapping out R_S with an inductor (Fig. 8.59) would allow the filter to continue to function properly at high current drains.

Regulation Modern equipment and circuits will only perform reliably when they are furnished with a constant supply voltage. If we consider that power supplies without

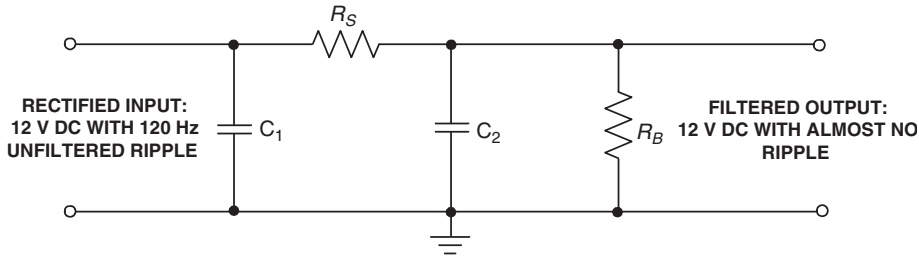


Figure 8.58 A type of power supply filter using a series resistor.

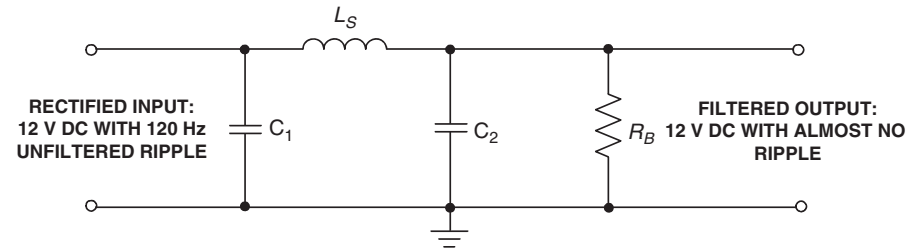


Figure 8.59 A type of power supply filter using a series inductor.

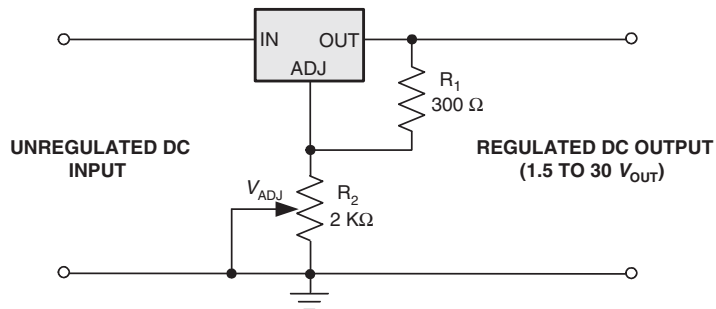


Figure 8.60 The common three-terminal (3-T) regulator with adjustable output voltage.

any regulation will shift their output voltage when the voltage varies at the AC mains, or even when the resistance of the load changes, we can see that regulators are required to avoid, or at least decrease, these undesired effects. Regulators will also assist in smoothing the output voltage, thus helping the power supply's filter section. Figure 8.60 illustrates one of the most common and easy to implement regulators available, the *3-T* type. Much more on regulators will follow.

Another kind of power supply is extremely popular, and is called a *switch-mode power supply*. Figure 8.61 demonstrates one widespread method for implementing a complete switch-mode power supply (SMPS). The AC mains inserts a 120-V_{RMS} AC signal into the SMPS input, where high amplitude transients attempting to enter the supply and cause damage will be shorted to ground through the MOV (*metal oxide varistor*), thus imparting limited protection from any voltage surges or lightning strikes. The bridge rectifier, consisting of the four diodes, rectifies the 60-Hz mains AC. Thermistors *TH1* and *TH2*

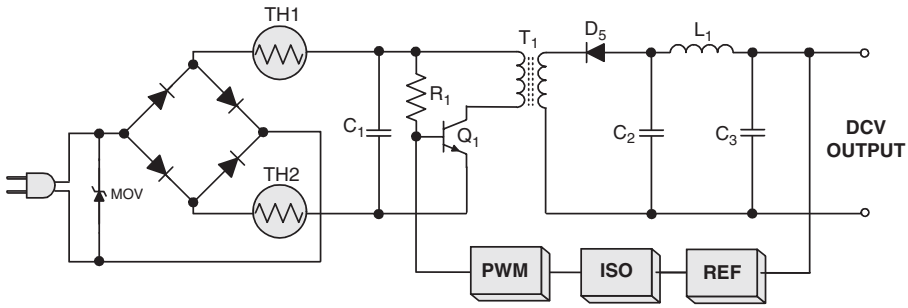


Figure 8.61 A type of switch-mode power supply.

perform current in-rush limiting by forcing the current, when the thermistors are cold, to slowly flow through their high resistance. After the thermistors have warmed up due to this current flow their resistance falls, allowing almost complete current flow. Capacitor C_1 filters much of the rectified AC to DC. The chopper transistor Q_1 is switched on and off by the PWM (*pulse width modulator*) which, when turned *on*, will change the width of its output pulses. The pulse's width is dependent on the required amplitude of V_{DC}, which is governed by the REF voltage. In other words, V_{DC} must not fall below this REF voltage or the PWM increases its duty cycle to compensate. R_1 is the Q_1 start-up resistor, while ISO (the *isolator*, usually an *opto-isolator*) supplies isolation between the low voltage secondary and the higher voltage primary. The chopped-up (by Q_1) direct current is sent through transformer T_1, rectified by D_5, lowpass filtered by C_2, L_1, and C_3, and is then placed at V_{DC} output as regulated DC. Much more will be discussed about this widespread power supply in Sec. 8.7.2.

Never run a switch-mode power supply without a load attached at its output, or it may become damaged or run improperly. Care is also always warranted when probing the circuits of any SMPS, since very high voltages can exist within some portions of the switching circuit.

8.7.2 Power Supply Regulators

The 3-T linear regulator, both the regular and the *low dropout* (LDO) type, as well as switching regulators, have become the most familiar regulator types in electronics. However, for certain applications and cost constraints, there are still many different voltage regulator circuit designs in common use.

For instance, the lowest cost regulator circuit is the simple *Zener shunt voltage regulator* (Fig. 8.62), and works owing to the following: Due to reverse diode action when the Zener

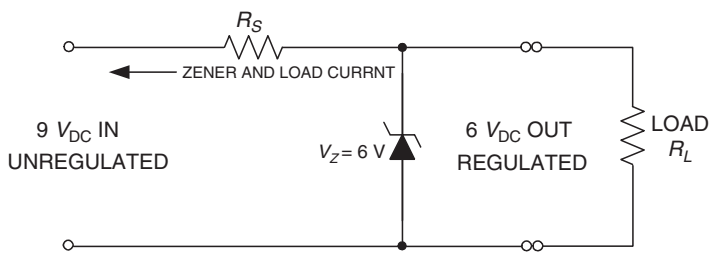


Figure 8.62 A simple Zener shunt regulator.

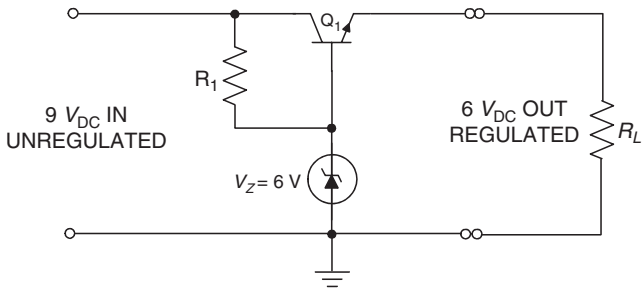


FIGURE 8.63 A series-pass transistor regulator.

hits its avalanche knee voltage, which ensures that the reverse voltage across the device will change very little for a large increase in Zener current, the load is in parallel with this diode and is therefore well regulated (V_{IN} must be a little higher than V_Z for this type of Zener regulator to remain in regulation). And as the current rises, the circuit forces the surplus voltage that is not dropped across the diode to be dropped across R_S, since the sum of the Zener voltage and the voltage drop across R_S must equal the input voltage. In essence, the Zener alters its resistance as the current changes in order to keep its Zener voltage (V_Z) constant to the parallel load. But since the Zener is in shunt with the load, this current through the diode can be considered wasted. In many applications, this is unacceptable. The following regulators solve this problem.

Another low-cost, but higher performing, regulator than the simple shunt Zener type above is the *series-pass transistor regulator* of Fig. 8.63. In this circuit, if the voltage across the load attempts to increase for any reason, the regulator's output voltage will stay nearly steady. This is due to the following action: The voltage across the load, R_L, must equal the voltage drop across the Zener minus the voltage drop across the E-B junction of the series-pass transistor Q_1. The transistor's base voltage is set by the Zener and, since the voltage dropped across the Zener cannot change, then any rise in the regulator's output voltage will force Q_1's emitter to be that much more positive than the base, which is the equivalent of making the base *less* positive. This ends in a smaller base-to-emitter voltage, resulting in less emitter current through the transistor, and thus an increased voltage drop across Q_1. Any attempted rise in voltage across the load R_L is substantially lowered in value, maintaining a steady output voltage for undemanding applications.

Figure 8.64 shows a *series-pass regulator with feedback*, which maintains a far more consistent and steady output voltage than the two regulators discussed above. This

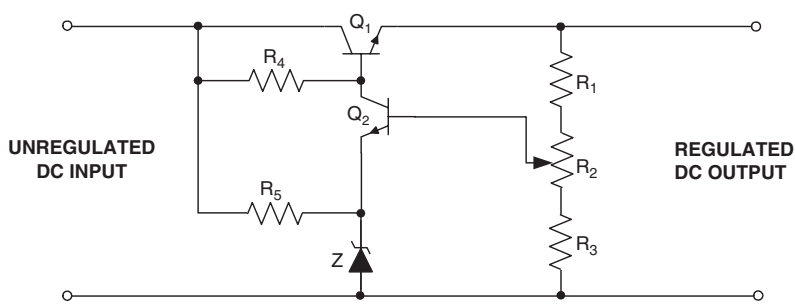


FIGURE 8.64 The series-pass regulator with feedback.

circuit contains R_1, R_2, and R_3, which monitor the output voltage as a voltage divider network. And since these resistors form a voltage divider, we can also set the required output voltage to a wide range of values by simply moving the wiper of the adjustable R_2 all the way from a little above the Zener's V_Z to just under the unregulated supply voltage. For instance, if we wish to increase the output voltage, we can move the potentiometer's wiper downward, and the voltage on Q_2's base decreases, lowering its forward bias so that this error detector/amplifier transistor will now conduct less. This forces Q_2's collector voltage to increase, which then forces the base voltage of Q_1 to rise, making this series-pass transistor conduct harder. Now, increased current will flow through any load placed at the regulator's output, thus increasing the output voltage. Q_2's emitter is clamped at a reference voltage by the Zener, with R_5 setting the diode's idling current, while the collector resistor for Q_2 and the base bias resistor for Q_1 is supplied by R_4.

3-T three terminal regulators of Fig. 8.60 above are, due to their low cost, size, weight, high efficiency, and great simplicity, a common type of regulator package for almost any circuit or system requiring 3 amperes or less (much higher currents are possible with additional external components). There are many other popular regulator IC packages available with more than three pins. Some may have up to eight pins, while countless others will possess four pins.

As shown in the figure, the programmable 3-T standard NPN type regulator circuit will employ only two simple components, R_1 and R_2, which are needed to set the output voltage level. R_2 can also be varied in order to change this output voltage for almost any requirement. C_1 and C_2 are not even needed for undemanding applications, but they do assist in the stability, as well as the transient response, of the regulator. These integrated circuit regulators also include full internal current limiting and thermal protection circuits. Thus, if a regulator's internal power dissipation rises excessively, the chip will shut itself down before burning out.

Most modern linear regulators are now of the *low dropout* (LDO) type. The LDO regulator can stay in regulation even when the input voltage is just a few tenths of a volt higher than its output voltage. Some LDO's, in fact, may only drop 50 mV across the regulator IC itself. And there are a few low voltage LDO's that can stay in regulation with output voltages of only 1.1 V (with an input of 1.5 V). Other, more common, LDOs may have a minimum 2.8 V output with a 3.0-V input. This is quite unlike the older linear regular NPN regulators above, in which some required a 2-V difference between the input and the output ports to remain in regulation. Additionally, these non-LDO regulators had a high-quiescent (no-load) current draw, and many such older regulator ICs did not even possess internal shutdown capability.

As mentioned, very efficient integrated nonlinear *switching regulators* are found in an increasing number of power supplies. These devices accomplish regulation by outputting a variable duty-cycle pulse into a lowpass filter in step with the output load requirements, and perform this switching with speeds of between 20 and 500 kHz. Due to the fast rise and fall times of these waveforms, however, strong switching noise constituents will be produced within the regulator circuit. These must be heavily filtered externally to reduce excessive hash output into sensitive radio gear. Indeed, switching power supplies require not only filtering of all the input and output leads, but also shielding and very short trace runs to minim ize EMI generation. Thus, their use is still somewhat limited in RF communication's gear, where we may favor the very quiet linear power supply regulators, such as LDOs.

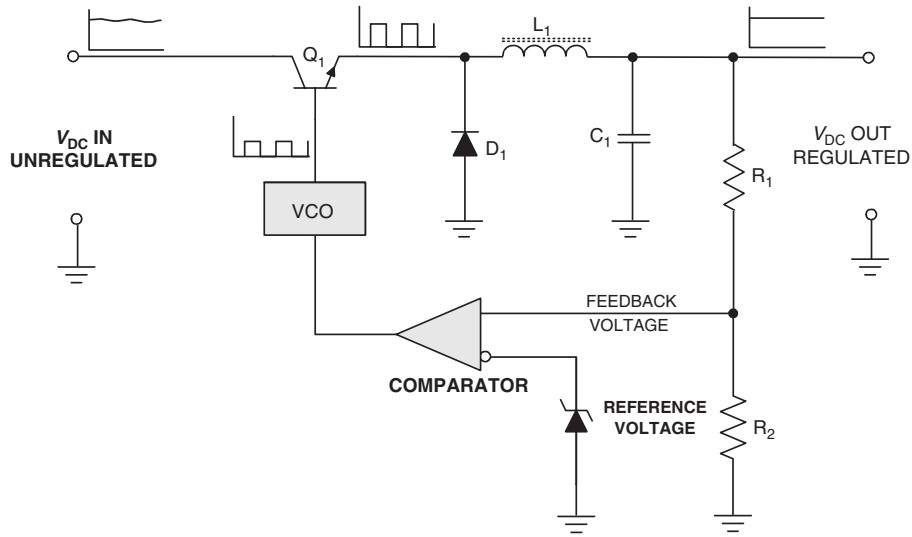


Figure 8.65 Switching regulator operation with waveforms.

Figure 8.65 is a common architecture for a switching regulator. Q_1, a switching pass transistor, receives pulses into its base from a VCO that is controlled by a *comparator*. Q_1 then outputs a switching voltage in step with these commands into the lowpass filter of L_1 and C_1, which filter the pulses into DC. By increasing or decreasing the pulse's duty cycle into this LC filter, the average DC output voltage will increase or decrease. D_1 protects Q_1 from the inductive kickback of the stored charge within L_1, which would normally produce a very high spike of voltage into the emitter during switching. The voltage divider of R_1 and R_2 program the desired output voltage of the switching regulator by using the voltage dropped across R_2 as a comparison to the Zener's reference voltage, which then turns *on* or *off* the comparator, and thus controls the VCO. Most of the above circuit is found in integrated circuit form, with built-in current limiting and thermal protection.

An integrated version, along with the typical support components, is displayed in Fig. 8.66, and is a common regulator arrangement in many production switch-mode

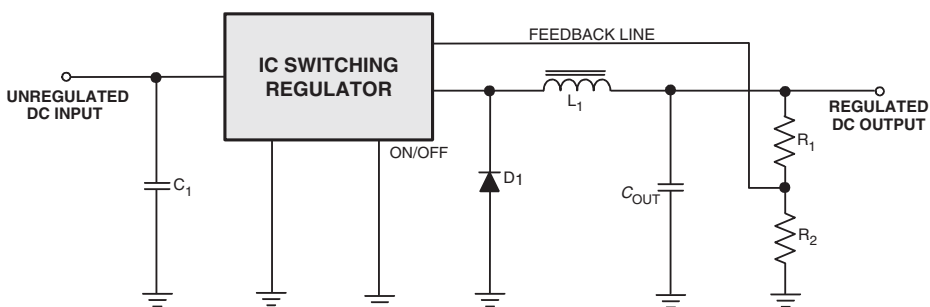


Figure 8.66 A switching regulator integrated circuit with support components.

power supplies. As above, R_1 and R_2 program the desired output voltage, with the voltage across R_2 being fed back by the *feedback line* to the internal comparator circuits. The lowpass filter is comprised of L_1 and C_{OUT}, with D_1 shunting the inductive kickback of a discharging L_1 to ground when the IC switches off, instead of across the IC. C_1 is used to give the regulator stability at higher current draws. Further filtering at the regulator's output may be required if undesirable ripple amplitudes are still present.

8.7.3 Power Supply Regulation Selection

There is, frequently, confusion as to when to select a linear regulator, such as an LDO, verses a switch-mode SMPS type. A switch-mode power supply is far more efficient than a linear regulator-based power supply. In fact, an SMPS can easily have an efficiency unheard of in linear supplies, reaching up to 94%. And since power lost in a linear regulated power supply is approximately $I_{OUT}(V_{IN} - V_{OUT})$, this can start to become substantial in high current draw applications, as can the attendant critical thermal issues. But for low-ripple and low-noise applications, linear power supplies are irreplaceable.

1. *Use an LDO linear regulator when you need*:
   a. The lowest cost regulator solution available.
   b. Low-noise and low-ripple output voltages.
   c. A low part's account.
   d. A bulletproof design that (almost) cannot go wrong from a design and realization perspective.
   e. A highly efficient regulator for battery operation when the input voltage amplitude is close to the output voltage amplitude.
   f. Compact layout on the PCB.
2. *Use a switching regulator when you need*:
   a. To step an input voltage to a higher value.
   b. Low-power dissipation, thus low heat generation, for decreased thermal problems.
   c. The highest efficiency regulator available (very important in battery-powered applications for long talk times).
   d. A negative voltage output.
   e. The lowest number of batteries possible.

Indeed, we can also combine a switch-mode with an LDO regulator within the same system. For instance, we can place a switching regulator at the input to a high-current transceiver, with the switcher being fed by 3 to 4.5-V battery cells, and generating an output voltage of 5 V for the PA section, while also feeding a linear regulator, with the linear regulator then converting the noisy 5 V switch-mode voltage into a low-noise 3.3 V for the sensitive, low-power RF receiver circuits of the transceiver. There are many other such LDO/switcher combinations possible in order to maximize efficiency, while minimizing noise and complexity.

8.7.4 Power Supply Regulator Design

This section describes designing with both the older NPN standard and the newer low dropout linear LDO regulators, as well as switch-mode integrated circuit regulators.

Designing with Linear Regulators

Low dropout regulator design is relatively trouble free, but a few precautions must be kept in mind. Positive oscillatory feedback will occur, with the LDO becoming unstable, unless the loop response of an LDO has negative gain (< 0 dB) *before* the regulator circuit's loop phase reaches 180°. However, due to the LDO's large external output filtering capacitor, the regulator's loop should normally remain stable, and without any need to simulate the regulator circuit itself. But if the ESR of this output capacitor is excessively high (i.e., above the LDO's recommended data sheet value), then whatever phase margin we had before to maintain a stable loop may be lost, and the LDO's loop gain will now not cross 180° phase before it reaches the safety of negative loop gain. It will thus become an oscillator. Since some types of electrolytic capacitors, such as aluminum, can dramatically raise their ESR as the temperature decreases, we must always design with sufficient ESR (and capacitor value) margins. This clearly demonstrates that we should always study the data sheet specifications for all components used in a circuit, even for such humble low-frequency parts as an electrolytic filtering capacitor; or we may receive an unpleasant surprise at the worst possible time, which is when there are thousands (or even millions) of your wireless devices already sold to customers.

Some LDOs may even require a *range* of ESR values. In other words, instability will occur not only if the filter capacitor has too high an ESR, but also if the capacitor has too low an ESR. For this very reason we must never use ceramic output capacitors in many LDOs, since they may have ESRs values as low as 0.005 to 0.010 Ω (at 5 μF), unless that particular regulator was specifically designed for use with ceramic capacitors.

Even while employing the exact recommended electrolytic filter capacitor, with its exact recommended capacitance and ESR values, we may still find that there is still another common source of LDO instability that can be created. This instability is caused when we attempt to filter out slightly higher frequency–noise constituents, which we normally do by adding a 0.01-μF (or some other such value) ceramic capacitor directly to the LDO's output. But, even though these 0.01-μF capacitors may be 200 or 300 times smaller than the main electrolytic output filter capacitor, they can still cause serious regulator stability problems. This is because these two different types and values of output capacitors, along with the LDO circuitry itself, may cause the regulator's loop phase to reach 180° before the loop gain drops below 0 dB. This will, as stated above, now cause the loop to oscillate. We can avoid this significant problem by using only smaller value ceramic output capacitors, which are at least 1500 times smaller than the main electrolytic; or by using larger value ceramics only if the LDO's output port and the smaller-value ceramic capacitor itself are separated by some small series lumped or distributed inductance.

The older, non-LDO NPN standard regulators do not have this instability problem, nor do they specifically require input or output capacitors at all. However, low value decoupling capacitors that are placed at their input and output ports will assist in the regulator's transient response, as well as decrease any possible chances (no matter how remote) of the NPN regulator ever becoming unstable. Figure 8.67 demonstrates a typical battery-powered NPN regulator circuit that converts 13.7 V_{DC} input to a 5.5 V_{DC} output, along with polarity (D_2), transient (TVS), overcurrent (F_1), and overvoltage (D_1) protection. A power-on indicator (LED) and the noise and ripple filtering (L_1, C_1, and C_2) are also shown.

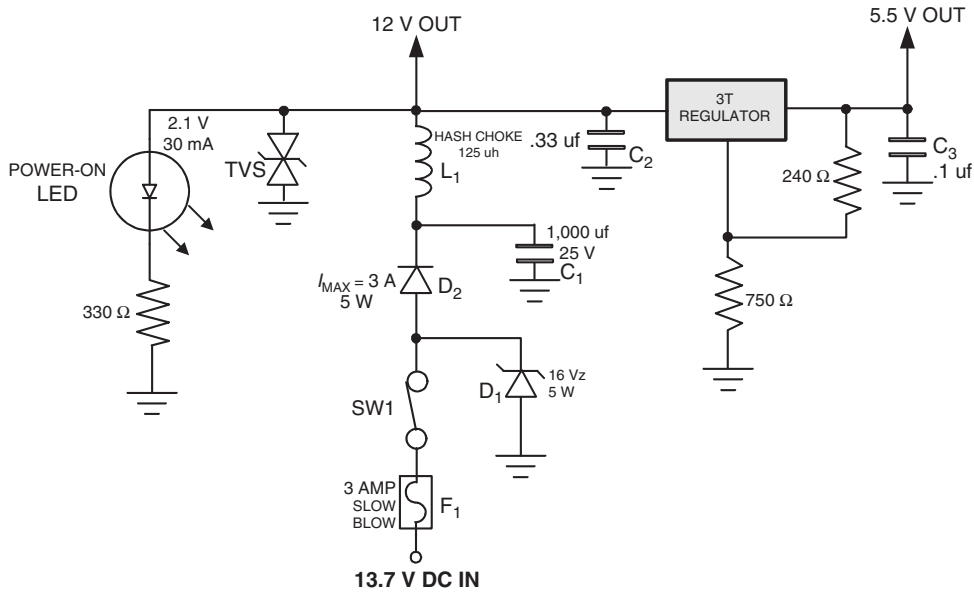


Figure 8.67 3-T regulator with input protection, filtering, and power-on LED.

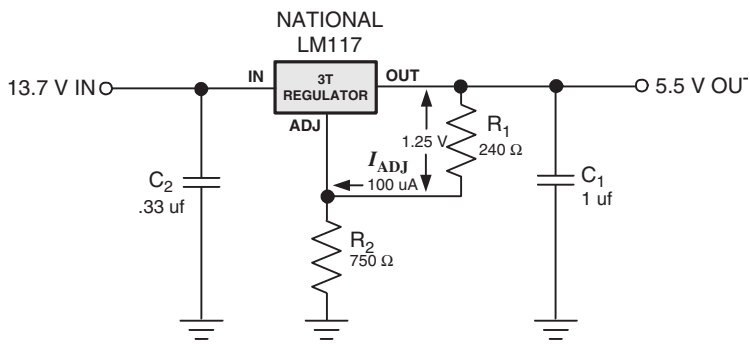


Figure 8.68 3-T regulator set for 5.5-V output.

NPN Standard Linear Regulator Design (Fig. 8.68)

Linear NPN standard regulators, which are not nearly as common as they once were due to their relatively large and wasted voltage drop across the regulator circuit itself, are still the easiest type to design with and implement. Indeed, the total parts count for a fixed voltage version may merely comprise the NPN regulator IC itself.

Certain precautions for the protection of the NPN regulator IC should be employed if voltage outputs in excess of 25 V are necessary into high-capacitance loads. This is due to the load capacitance actually discharging stored energy back into the regulator's output, causing high amperage spikes to damage the 3-T when it

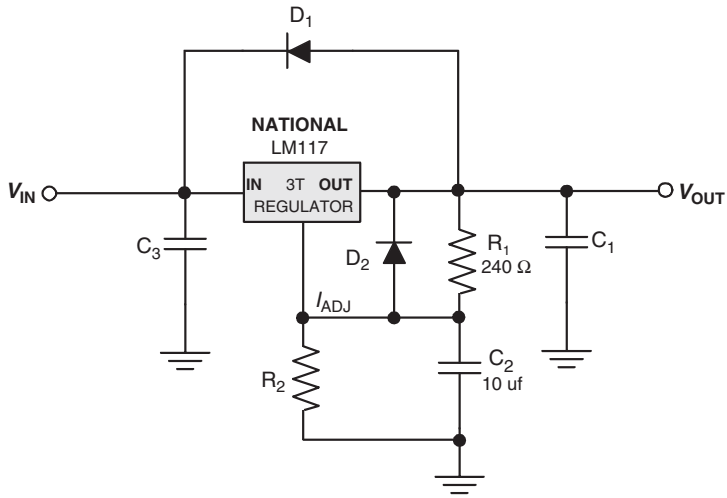


FIGURE 8.69 3-T regulator discrete protection circuit.

outputs high voltages. The circuit of Fig. 8.69 will perform this protection function. Diode D_1 is shielding the IC from the discharge of C_1, while D_2 is protecting the IC from the discharge of C_2.

To Design Figure 8.68 is a complete 3-T NPN regulator circuit using the National LM117 chip, and with all the required components of an adjustable voltage type.

1. C_2 is only mandatory if the regulator IC will be six or more inches away from the input filter capacitors of the power supply. If used, it should be between 0.1 to 1 μF. Solid tantalums are normally recommended.
2. C_1 is only necessary in order to reduce transients and prevent too much ringing with certain capacitive loads. It may be eliminated in low-cost designs but, if used, can be any value between 0.1 and 100 μF. Solid tantalums are normally recommended.
3. R_1 is usually chosen to be 240 Ω, and only R_2 must be calculated:

$$R_2 = 188.5\ (V_{OUT} - 1.25)$$

NOTE: *Formulas shown are only entirely valid for an* R_1 *value of 240* Ω*, and only for the National LM117 and LM317 series. The initial* R_1 *value may be recommended by the manufacturer of each NPN IC regulator model, and may be between 120 and 240* Ω*.*

4. Since V_{OUT} and R_2 are the only variables, it can be seen that making R_2 variable will permit V_{OUT} to be fully adjustable, or:

$$V_{OUT} = 1.25\left(1 + \frac{R_2}{R_1}\right) + (I_{ADJ}\ R_2)$$

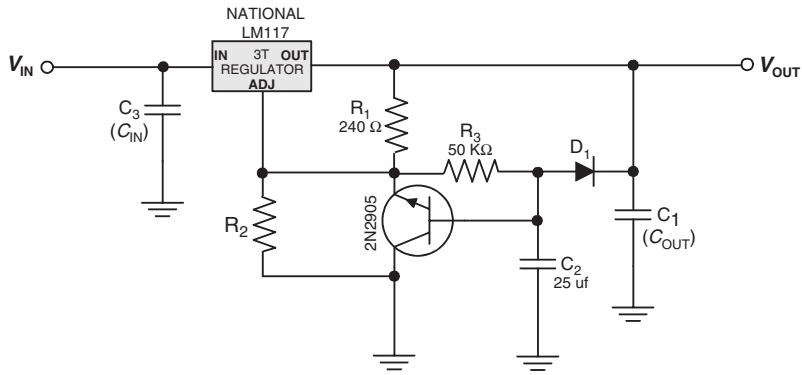


Figure 8.70 3-T regulator for slow turn-on applications.

Note: I_{ADJ} *can be found in the voltage regulator's data sheet. If a regulator must turn on more slowly, Fig. 8.70 displays the appropriate circuit values (with* R_2 *still calculated as above).*

LDO Linear Regulator Design (Figs. 8.71 and 8.72)

Depending on the package, options, and thermal issues, some LDO's will have anywhere from 3 to 8 pins. The 3 pin packages are 3-T types, with an input, output, and ground. Higher pin packages may have added functionality or performance enhancements, such as a pin for turning the LDO off (and therefore the load as well); a pin to decrease the LDO's output noise by using a shunt bypass capacitor to ground; a pin that permits the correction of the regulator's output voltage to compensate for any resistive voltage drops between the LDO and the load; and extra ground pins for added thermal dissipation into the PCB. An adjustable LDO regulator type will also have an *ADJ* pin for adding two external resistors, which will program the regulator to output any discrete voltage within its design range, as opposed to the single voltage value of a fixed LDO. And unlike the older non-LDO NPN standard linear regulators, LDOs are very selective about the value and type of output capacitor. These capacitors must be chosen as per the LDO's data sheet, or the entire regulator circuit may become unstable.

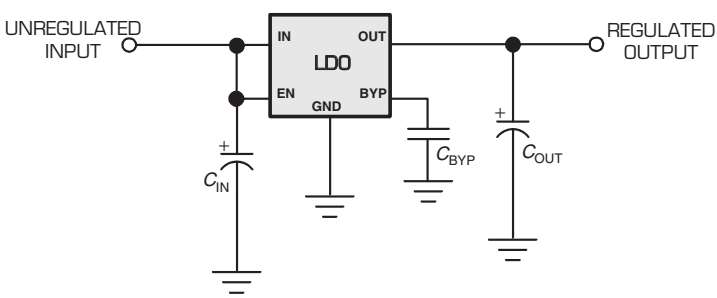


Figure 8.71 A fixed-voltage low-dropout (LDO) regulator circuit.

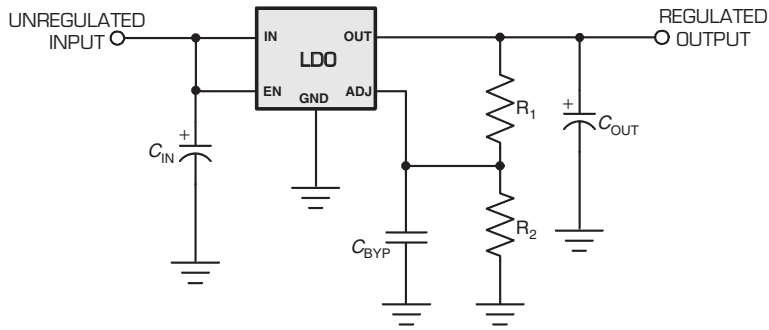


FIGURE 8.72 An adjustable-voltage low-dropout (LDO) regulator circuit.

A. Fixed Voltage LDO Regulator (Fig. 8.71)

To Design

1. C_{IN} = 1 to 10 µF (Depending on the specific LDO model)
2. C_{OUT} = 2.2 to 28 µF (Depending on the specific LDO model. The ESR value of C_{OUT} is critical for stability, and must be kept within the range as specified in the LDO's data sheet.)
3. C_{BYP} = 470 pF (Or as recommended in the LDO's data sheet)

B. Programmable Voltage LDO Regulator (Fig. 8.72)

To Design

1. C_{IN} = 1 to 10 µF (Depending on the specific LDO model)
2. C_{OUT} = 2.2 to 28 µF (Depending on the specific LDO model. The ESR value of C_{OUT} is critical for stability, and must be kept within the range as specified in the LDO's data sheet.)
3. C_{BYP} = 470 pF (Or as recommended in the LDO's data sheet)
4. Since the V_{OUT} value is programmed by the ratio of R_1 and R_2, select an R_2 that is between 4 to 470 kΩ, depending on the specific LDO model:

$$V_{OUT} = 1.23\left(\frac{R_1}{R_2} + 1\right)$$

Switch-Mode Regulator Design

Important Switcher Design Issues and Cautions In order to suppress damaging EMI, and to act as a heat sink, proper PCB layout for the switcher circuit is absolutely critical. Indeed, if trace inductance within the switcher circuit is not minimized, the rapid current switching rates of these regulators can cause severe voltage transients, which means keeping C_{IN}, SW, D_1, and C_{OUT} (see Fig. 8.73) traces as short and as fat as possible within the layout. Routing the sensitive feedback line to a quiet inner or bottom PCB layer would also be very beneficial to shield this trace from possible coupling through

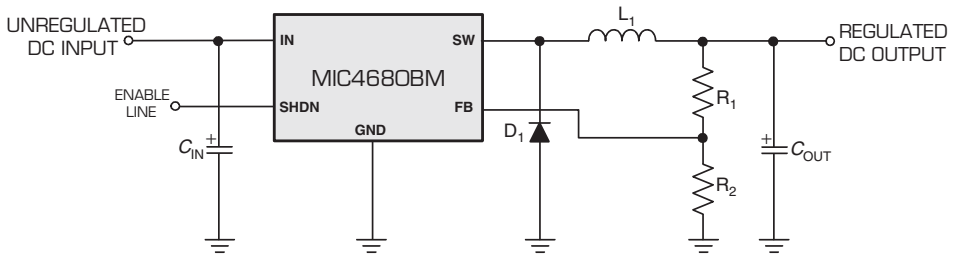


Figure 8.73 Switch-mode regulator design example circuit.

SW, L_1, D_1, C_{OUT}, and their associated traces, while C_{IN} should be as close as feasible to the IN pin.

For electrical and thermal reasons, all the regulator's ground pins should be soldered directly to the top groundplane, and then drop immediately to the bottom PCB groundplane(s) through multiple stitching vias. Further, to dissipate the heat generated by the switcher, this top groundplane should cover as large a surface area as possible. For this particular switcher model, at its maximum output current draw and voltage input level, 1.5 in² of top copper should be adequate to keep the die sufficiently cool for long-term reliability. However, the larger this copper area, the cooler the regulator die will be.

A point that is commonly overlooked in switcher design is that switch-mode regulators must obtain their rapid transient current demands from large-value capacitors, and these components must have low ESR and ESL. The ESR and ESL requirements are due to the large value of peak currents that these capacitors must source and sink every time the regulator switches, which can be as fast as 4 MHz. This obviously demands quality components, and of a value and type as recommended by the manufacturer of the switch-mode regulator chip. Therefore, for EMI control, the external input capacitor of a switcher must not only confine the rapidly switching currents to a localized, tight loop area on the PCB through a compact, low-impedance layout, but this capacitor must also be of a sufficient value to supply the transient peak current needs of the rapidly switching regulator at both low and high frequencies, and with low ESR and ESL. However, since this input capacitor is normally an electrolytic, it cannot be expected to operate effectively above a few megahertz because of the high value of parasitic series inductance it possesses to any high-frequency noise: a noise which is trying to escape from the regulator's own input port. Thus, for the particular noise-sensitive applications that are so common to wireless communication circuits, we will need to shunt an RF capacitor in parallel with the high-value electrolytic input capacitor. This higher frequency capacitor will normally be of the ceramic type, with a value as recommended by the switcher's manufacturer. Additionally, we may want to insert a series input inductor, followed by another bypass capacitor, to further reduce switcher generated EMI from entering the main unregulated DC supply of the entire system. This inductor may in fact only need to be a thin, meandering trace to supply the low value of added inductance needed to assist in the suppression of this high-frequency RF noise. The inductive trace must be sufficiently wide to safely carry the required current draw without burning up.

The input and output electrolytic capacitors must not only be selected for their low ESR/ESL and particular capacitance value, but also for a sufficient *ripple current* rating

(RCR). Both ESR and RCR are strongly related, since the constantly charging and discharging of the electrolytic capacitors in a switching power supply will cause power to be dissipated within the capacitor's ESR, producing internal heating and possible capacitor failure, as well as low switcher efficiency. Therefore, if the switcher is run in a high-temperature ambient environment, and either the input or the output capacitor is excessively increasing in temperature during regulator operation, then the capacitor(s) should be replaced with models than enjoy better RCR. (As a general rule, the capacitor's ripple current rating should never be below 80% of the expected maximum current expected from the switch-mode regulator.)

There must be minimal distance and low-trace inductance between the ultra fast Schottky diode of D_1 and the L_1 inductor, and with both placed as close as possible to the switcher's output pin. This is needed so as to reduce the generated EMI caused when the internal FET of the regulator turns *off*, which produces a very rapid transient voltage overshoot and strong ringing effect that is created by the diode's nonzero switching time and the parasitic inductance of the components and traces between these two parts.

Even with a good layout, some of this EMI is bound to occur through the above mechanism, but can be further suppressed by employing a *series RC snubber network* placed *directly* across the switcher's external output pins for the IC's internal FET drain and source. This snubber will absorb much of the EMI generating voltage overshoot and ringing caused by the rapid FET switching speeds, and consists simply of a resistor and capacitor in series. The value of the snubber's *R* and *C* must be optimized for maximum transient suppression, while minimizing switcher losses, but can be a value of between 10 Ω to 1 kΩ for the *R*, and 0.01 to 1 μF for the *C*.

Designing with the Micrel MIC4680 SuperSwitcher (Fig. 8.73)

As an excellent demonstration of the procedure to design a modern, low-cost switch-mode regulator, the Micrel MIC4680 *SuperSwitcher* is a perfect example. This particular device requires only four external parts when utilized as a fixed voltage regulator; by adding two resistors we can construct a fully voltage-adjustable switcher. The MIC4680 comes in a low-cost plastic-surface-mount SOIC-8 package, and is capable of supplying up to 1.3 A output, has a 4 to 34 V safe input voltage range, a 1.25 to 6 V output voltage range, is thermal and overcurrent protected, and can be shut down with a logic high signal.

Pin 1 (SHDN) of the MIC4680 is the enable/disable port, with any voltage greater than 1.6 V placed at this pin forcing the switcher to turn *off*, and any voltage below 1 V turning it *on*. Pin 2 (IN) is the port for the unregulated input voltage of anywhere between +4 to +34 V (this value should be 15 V if we demand maximum switcher efficiency). Pin 3 (SW) is the internal switching transistor's output port, with a frequency of 200 kHz. Pin 4 (FB) is for the feedback line back from the two external programming resistors, with this pin internally connected to a comparator circuit. All other pins are for electrical ground, which also remove heat from the chip's die junction.

To Design

1. $R_1 = 3.01\ \text{k}\Omega$ (1% tolerance)
2. $R_2 = \dfrac{3010}{\left(V_{OUT}/1.23\right) - 1}$ (1% tolerance)

3. For proper operation, the minimum permitted input voltage should be calculated as: $V_{IN(min)} = V_{OUT} + 2.5$ V (which must never be below 4 V).
4. L_1 should be a shielded magnetic of 68 μH with a 1.5-A capability.
5. D_1 must be a high efficiency Schottky type capable of 60 V and 2 A.
6. C_{IN} must be a low ESR type of 15 μF at 35 V.
7. C_{OUT} must be low ESR type of 220 μF at 10 V.

8.8 Directional Couplers

8.8.1 Introduction

Directional couplers are 50-Ω circuits designed to allow an RF signal to pass from their input port through to their output port (called the *mainline* of the coupler) with minimum insertion loss, while permitting a small part of this signal to be tapped (Fig. 8.74). Depending on the particular type of coupler, the tapped signal may be 30, 20, or 10 dB less than the coupler's mainline output, so the actual loss at this mainline output port is quite small. The low-power-tapped signal can be a sample from the forward (incident) signal that was injected into the coupler's input port, or it may be the power from any reflected (reverse) wave present, depending on the coupler's orientation. There are also dual directional couplers that will measure both the forward and reverse power simultaneously through two separate ports. Most couplers are relatively narrowband, and must be chosen or designed with a certain frequency range in mind.

Figure 8.75 shows a typical microwave distributed directional coupler. The coupler's *mainline insertion loss* is the loss from Port 1 to Port 2 (such as 0.4 dB); the

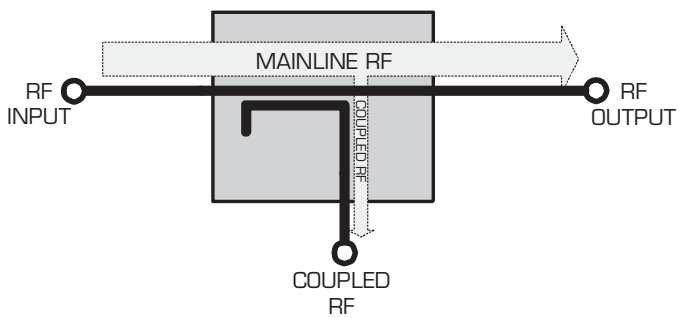


Figure 8.74 A directional coupler.

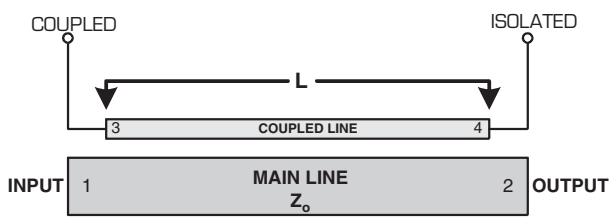


Figure 8.75 A distributed directional coupler circuit.

coupled port loss is the amplitude at Port 3, in decibel, relative to Port 1 (such as 6, 12, 20 dB, and so on); while the isolated Port 4 may well have up to 60 dB less signal than that at Port 1.

8.8.2 Directional Coupler Design

The lumped low loss *LC directional coupler* shown below is a narrowband structure, while the *Adam's 50-Ω resistive coupler* will work in very wideband applications. Miniature surface-mount single-chip ceramic directional couplers are also available that operate up to 2.4 GHz, with low insertion loss and extremely simple implementation, but their cost is higher.

A Lumped LC Directional Coupler for up to 1 GHz (Fig. 8.76)

To Design

1. $R = 50\ \Omega$ (By replacing R with a 50-Ω amplifier, or by placing a high-impedance circuit in shunt with R, will permit the sensing of reverse power.)
2. $L = \dfrac{50}{2\pi f}$
3. $C_1 = \dfrac{1}{2\pi f 50}$
4. $C_2 = \dfrac{10^{(CF/20)}}{2\pi f 50}$ (This C_2 value must be less than $\dfrac{0.18}{2\pi f 50}$.)

where CF = coupling, in dB, desired at Port 3, which must be less than −15 dB for this type of coupler, and must be in the form of $-X$ dB in the equation, and not as $+X$ dB or X dB. This CF value will remain within specifications over a maximum 10% bandwidth.

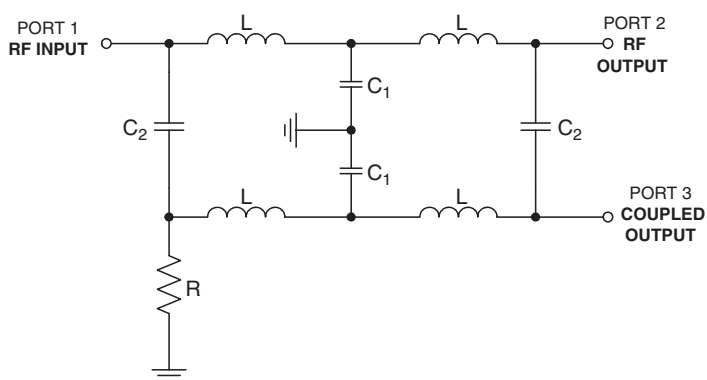


Figure 8.76 An LCR lumped directional coupler.

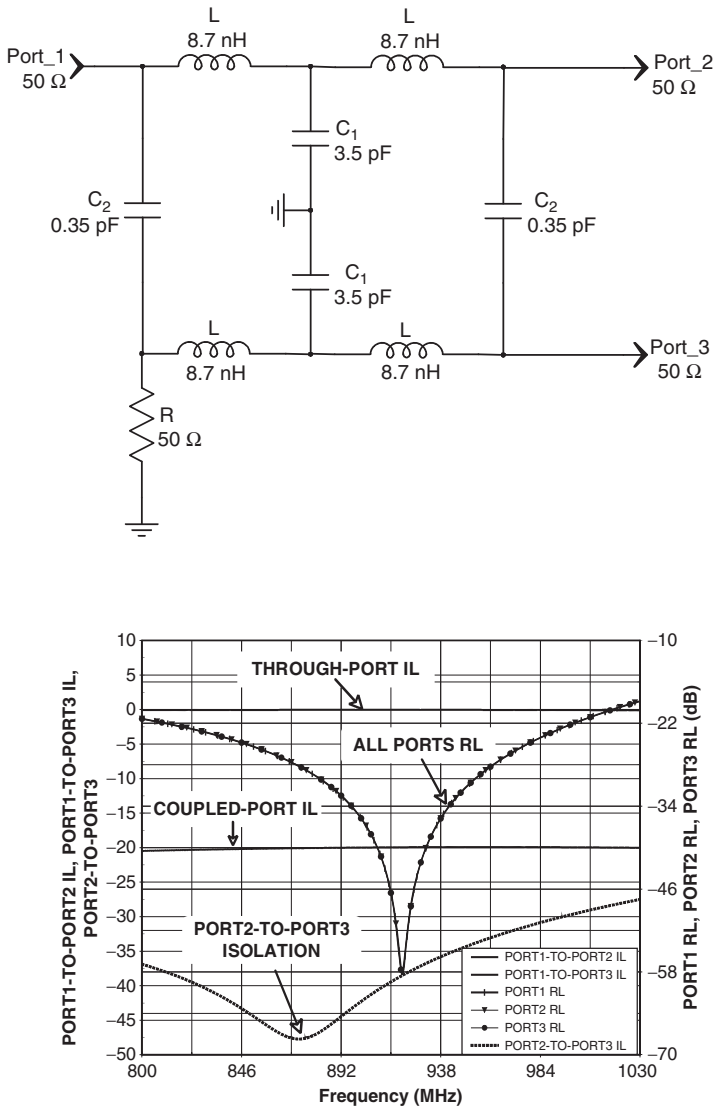
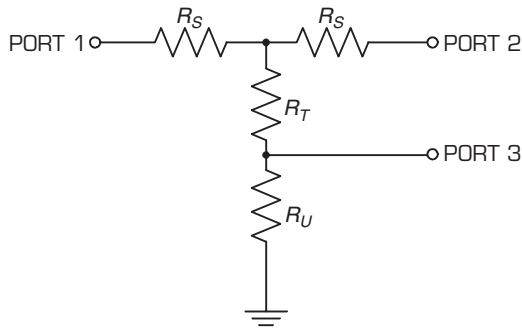


FIGURE 8.77 The example narrowband LC directional coupler with calculated part's values, along with port-to-port insertion loss (IL) and return loss (RL) results.

A Quick Example Design a Lumped Directional LC Coupler (Fig. 8.77)

Goal: Create a low-loss lumped LC directional coupler. The specifications and parameters for the circuit are:

$f_r = 915$ MHz
$Z_{IN} = 50\ \Omega$
$Z_{OUT} = 50\ \Omega$
Coupling = 20 dB



PORT 3 (COUPLED LOSS)	PORT 2 (MAINLINE LOSS)	R_S (Ω)	R_T (Ω)	R_U (Ω)
30	0.5	1.6	763	53
25	1.0	2.9	407	56
20	1.7	5.0	222	63
15	3.0	8.6	112	78
10	4.9	13.8	47	136

FIGURE 8.78 An Adam's wideband directional coupler, with different design resistor values for various Port 3 coupling output powers.

Solution:

1. $R = 50\ \Omega$
2. $L = 8.7$ nH
3. $C_1 = 3.5$ pF
4. $C_2 = 0.35$ pF

The Adam's 50-Ω Resistive Coupler Design (Fig. 8.78)

A very wideband, low-cost RF coupler.

To Design

1. Design by following the recommended part's values as shown for the needed thru and coupled attenuation requirement.

8.9 Assorted Support Circuits

8.9.1 Introduction

The following introduces assorted support circuits that are used frequently in wireless design, but are beyond the scope of this book to go into any design detail. Today, due to their complexity and cost if implemented as discrete circuits, most of these support

circuits are usually placed within IC chips, and may actually be but a small part of a much larger transceiver RFIC.

8.9.2 Speech Processing

Speech processing is a general term for a circuit that adjusts an input audio signal in amplitude, frequency, or both, before it is placed into a transmitter's modulator. *Speech compression* and *compandering* are the most common of these processing techniques. A special form of speech processing, called *ALC,* or *automatic level control*, affects the RF, instead of the audio, of a transmitter system.

Considering that the modulation frequency of FM, SSB, and AM transmitters will influence the transmitted bandwidth, a method to limit the maximum baseband frequency must be utilized. This can be accomplished by an active lowpass filter placed within the audio sections.

Speech compression prevents a much wider bandwidth from forming outside of the desired AM passband due to the deleterious effects of overmodulation, which produces spectral *splatter*. This splatter is constructed of the additional harmonics created in the baseband signal due to this overmodulation, which further modulates the RF carrier, causing extra sideband components and a widening of the bandwidth. An overdriving of the IF or RF amplifiers of the transmitter may also occur, creating IMDs. Splatter and IMDs generate *adjacent channel interference* (ACI) and a less intelligible baseband signal. A speech compression circuit decreases these negative effects by amplifying a signal normally up to a predetermined level, but then will begin to reduce gain by 1 dB for every 2 dB of audio input signal. These basic speech compression circuits will simply confine the maximum AM or SSB audio amplitude to some maximal value, while *dynamic compression* helps intelligibility by increasing the smaller baseband amplitude levels as well. Compression schemes in general can be quite capable of raising the average output power of an SSB transmitter, while also decreasing distortion and splatter and limiting modulation to 100% or below. Human speech has voice amplitudes that are so highly complex and irregular that the transmitter must always be prepared to transmit the highest voice peak, which may be 10 to 12 dB higher than the average energy contained within the entire speech waveform, and compression smoothes over much of these amplitude variations in order to allow a much higher average output power, and thus an increase in the RF range of the wireless link. Simply put, basic speech compression acts in the same way as standard AGC, but is located in the audio stages (Fig. 8.79).

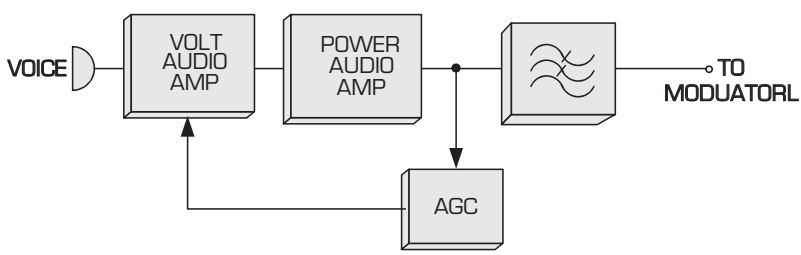


Figure 8.79 Transmitter speech compression in the audio section.

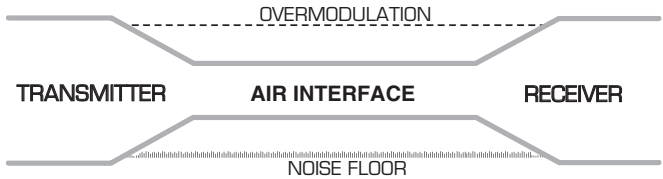


Figure 8.80 Companding action between a single-sideband transmitter and receiver.

A technique called *compandering* can take compression to outrageous levels by almost completely compressing all of the voice waveform's peaks close to the same level as the valleys at the transmitter, and then expanding them back to normal amplitudes at the receiver (Fig. 8.80). This permits a high dynamic range, better signal-to-noise ratio, and much higher average output powers.

Many lower cost FM, and some AM and SSB, voice transmitters may even employ a form of processing called *speech clipping*. If a voice signal reaches a high amplitude, the clipper circuit actually hard limits it to a certain value. To remove the harmonics produced by this clipping action and to limit the maximum frequency possible of the baseband signal, a lowpass filter is placed at the output of the clipper. A similar concept is an *audio clipper* circuit (Fig. 8.81), which can provide a degenerative out-of-phase feedback signal for any audio peak voltage over a certain preset level. This stage consists of an operational amplifier with back-to-back Zeners and a resistor that sets the audio gain, with the peak value of the baseband waveform being governed by the value of the diodes.

Another form of compression is called *ALC* (*automatic level control*, or *RF compression*), and is found in SSB, AM, and even some FM transmitters. It is, as well, essentially a standard AGC circuit, but designed to operate on the transmitter's IF (Fig. 8.82). ALC decreases high amplitude transmitted RF signals without effecting nominal or lower signal levels, and controls the gain of the IF stages by sending a voltage to the gate or base of the transistor in the IF's *variable gain amplifier* (VGA), which then decreases or increases the input power into the RF's PA. It does this by tapping the output of the PA, rectifying and filtering this RF to a DC voltage, and then placing this voltage at the input of the IF's VGA. Thus, no matter what the amplitude of the original baseband signal, the linear power amplifier cannot be overdriven, which would create excessive adjacent channel interference.

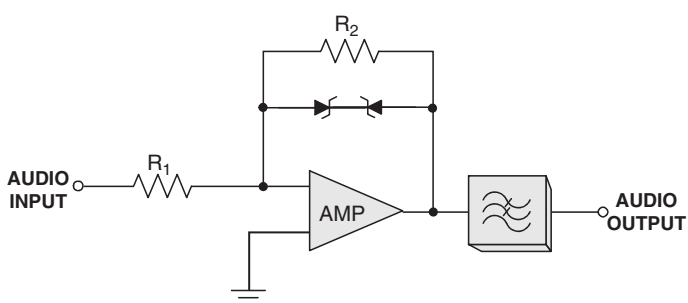


Figure 8.81 A voice clipper in the audio section of a transmitter.

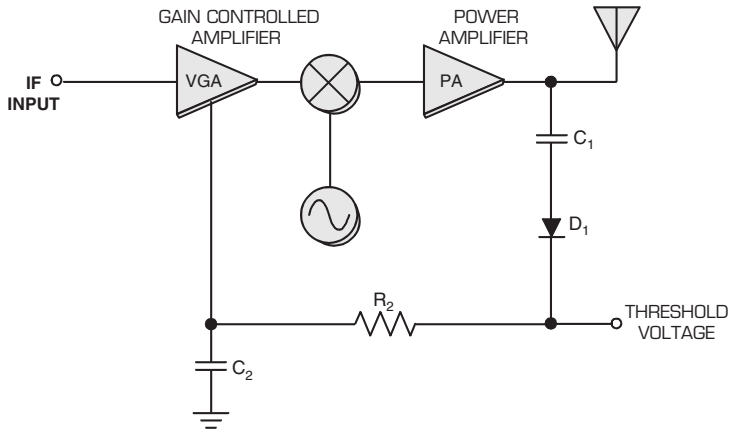


Figure 8.82 A transmitter's ALC circuit.

8.9.3 Automatic Frequency Control

In low-cost, non-PLL- or noncrystal-controlled transmitters or receivers, automatic frequency control (AFC) may be employed to steady the frequency of the system's LC local oscillators.

General frequency stability is obviously required for any modern receiver, especially when internal and external temperature changes occur, or the resultant frequency drift would convert any incoming RF signal to an improper IF, causing attenuation and distortion of the baseband signal. For the same reasons a transmitter must also maintain its stability, as well as to prevent interfering with other wireless devices within adjacent channels, along with the negative legal ramifications of a wandering transmitter.

A basic example of one type of AFC is shown in Fig. 8.83 for an FM receiver. The signal is tapped from the last stage of the IF strip and sent into an FM demodulator, which outputs a voltage in step with the IF's frequency drift. This drift is due to the unstable LC local oscillators. The demodulated signal voltage is then inserted into the lowpass filter to obtain a DC control voltage for the varactor stage. This control voltage adds to or subtracts from the varactor diode's center frequency DC bias, and since the varactor is also placed across the tuned circuit of the LC oscillator itself, the diode has some control over the output frequency of the LO. Thus, if the LO begins to

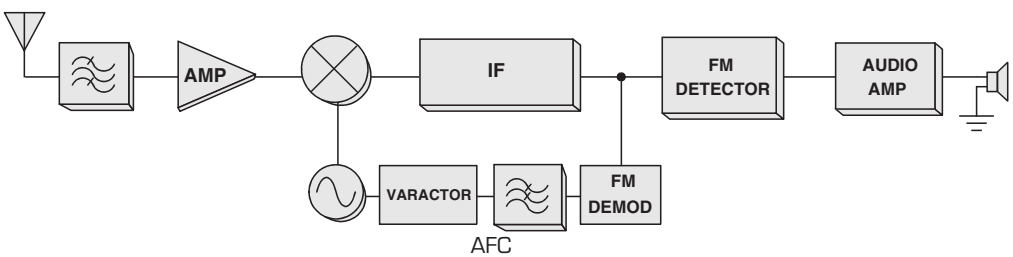


Figure 8.83 An FM receiver's AFC circuit for frequency stability.

move off frequency, the FM demodulator/filter stages will change from its center frequency output voltage level, and place the appropriate correction voltage into the varactor, altering its capacitance, and therefore coercing the LO back to its proper frequency.

8.9.4 Squelch

Squelch circuits mute the annoying heavy static that occurs when an AM or FM receiver has no RF input signal, an effect due to normal AGC action peaking the IF's stage gain under low-signal conditions. Another positive attribute to squelch circuits is that they can save a significant amount of battery power in a portable device, since an unsquelched receiver may use 100 mA or more, while a squelched receiver may draw only 15 mA or less.

A squelch circuit functions by stopping static from reaching the radio's output speaker, and can be realized in one of three ways: the supply voltage to the audio amplifiers can be switched off; or the audio amplifiers can be disabled by furnishing a reverse bias to their base; or the noise energy is prevented from reaching the audio amplifiers by either blocking the audio voltage with a series pass transistor or by shorting this voltage to ground.

The majority of all squelch circuits will detect the actual presence of a signal by simply looking at the DC output of the AGC loop. One common squelch circuit is as shown in Fig. 8.84. When the receiver is not receiving an RF signal of the proper input level, or perhaps no signal is present at all, then the IF stages of the receiver will be biased by the AGC loop for maximum amplification. Therefore, the static level output from the radio's speakers would be very high indeed if the squelch circuit were not present and shunting the noise energy to ground through the *squelch gate* (a transistor switch). But if we couple this IF AGC bias voltage into the AGC *input* of the *squelch amp* and on to the squelch gate, then when an RF signal of the proper strength is finally received it will generate an AGC voltage of a sufficient amplitude that will cutoff the squelch gate, switching it into a nonconducting state. This action permits the detected baseband audio to proceed into the *audio amp* to be amplified and sent onto the speaker.

Another squelch system is shown in Fig. 8.85, and taps the IF's noise frequencies located between 5 to 6 kHz with a bandpass filter, passing on only these particular noise components to the *static amp*, and then into the *rectifier/LPF* stage to be changed into a DC voltage used to control the squelch gate, which then dictates whether the audio signal is sent on to the audio amp.

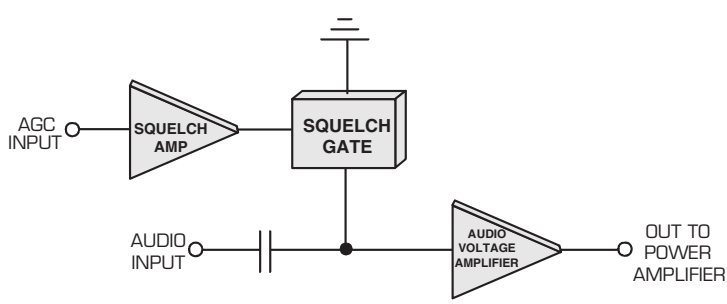


Figure 8.84 An AGC-based squelch circuit.

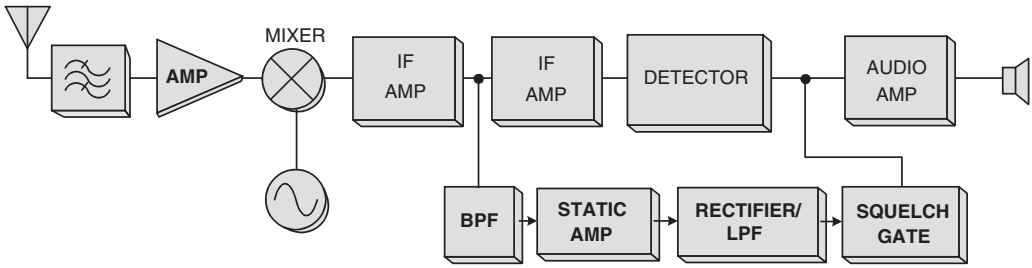


Figure 8.85 A noise-based squelch circuit.

For instance, due to the AGC the noise amplitude in the receiver will be quite high when no RF signal is being received. This undesired noise will be filtered to obtain a specific band of noise voltages, further amplified by the static amp, changed to DC by the rectifier/LPF, and placed at the squelch gate input to bias it *on*. This shorts the output of the *detector* to ground, preventing the audio amp from receiving this noise, and thus quieting the audio speaker. However, if an RF signal of sufficient strength is received, the noise levels would naturally decrease due to the AGC, biasing *off* the squelch gate, and thus permitting the baseband signal to reach the audio amp, and then on to the radio's speaker.

We can employ integrated circuits to assist in performing many of the squelch functions (Fig. 8.86). In some wireless devices a fully integrated circuit IF strip (IF IC) may be used, permitting us to simply tap this IC's *received signal strength indicator* (RSSI) output pin in order to obtain a DC voltage level that corresponds to the RF received signal strength. We may then feed this RSSI voltage into a *comparator* that is adjustable, through R_1, for the desired squelch threshold. As an example, if IN1 receives a voltage from the IF IC's RSSI output that is lower than the reference level at IN2, as set by the voltage divider of R_1, then the comparator will swing near to its

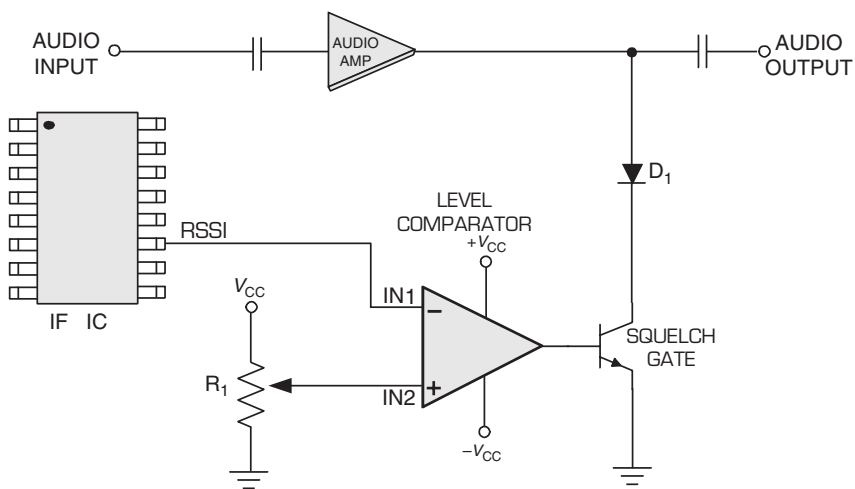


Figure 8.86 A squelch circuit using an IC's signal level output with a comparator.

positive $+V_{CC}$ rail, turning *on* the squelch gate. This will short the audio amp's collector to ground, preventing the static signal from reaching the power amplifier and speaker of the next stages.

There are highly integrated RFICs that employ their own internal squelch circuits, with an internal voltage audio amplifier, and that only demand a discrete potentiometer at the appropriate pin to fully adjust the squelch to any desired level. There are also highly advanced RFICs that include almost an entire radio on a single chip, and will also naturally include a squelch circuit, if squelch is required within the particular system.

This page intentionally left blank

CHAPTER 9
Communication System Design and Propagation

Any wireless design will surely fail without a solid understanding of the complete communications system as a whole, all the way from the transmitter's modulator input to the receiver's demodulator output, and everything in-between. The interrelationships of the transmitter, the receiver, the antenna, the air interface, and the modulation, and the selection of the system's various components, stages, and specifications can make or break an entire wireless design. For a high data rate at the required bit error rate (BER) for digital radio, or with the expected voice quality with analog radio, understanding the complete system is indeed critical to a dependable radio frequency (RF) link.

9.1 Receivers

9.1.1 Introduction

The most difficult to design element in most communication systems is the receiver. A receiver is a maze of conflicting specifications and requirements. It must have a low noise figure [at very high frequency (VHF) and above], low group delay variation (GDV) and intermodulation distortion (IMD), high dynamic range, stable automatic gain control (AGC), suitable RF and intermediate frequency (IF) gain, excellent frequency stability, good gain flatness, low phase noise, small in-band spurs, sufficient selectivity, appropriate BER and, sometimes the most critical r equirement of all, it must be under a certain cost target.

9.1.2 Receiver Image

An important concern of any superheterodyne receiver is the *image frequency* (Fig. 9.1), since any signal received within this image band will be amplified by the receiver's IF stages, to be sent on to the demodulator and output as interference. This image frequency can only be eliminated at the front end of a receiver, before down conversion, by a filter that blocks this interfering frequency from entering the first mixer of the receiver, called the *image filter*.

When the receiver's local oscillator (LO) is higher in frequency than the incoming RF signal (*high-side injection*), the image is any frequency that is at twice the IF plus the desired RF signal frequency (2 × IF) + RF, or at the IF plus the LO frequency LO + IF. Conversely, if the local oscillator is lower in frequency than the incoming RF signal (*low-side injection*), we can find the image frequency at (2 × IF) − RF and LO − IF.

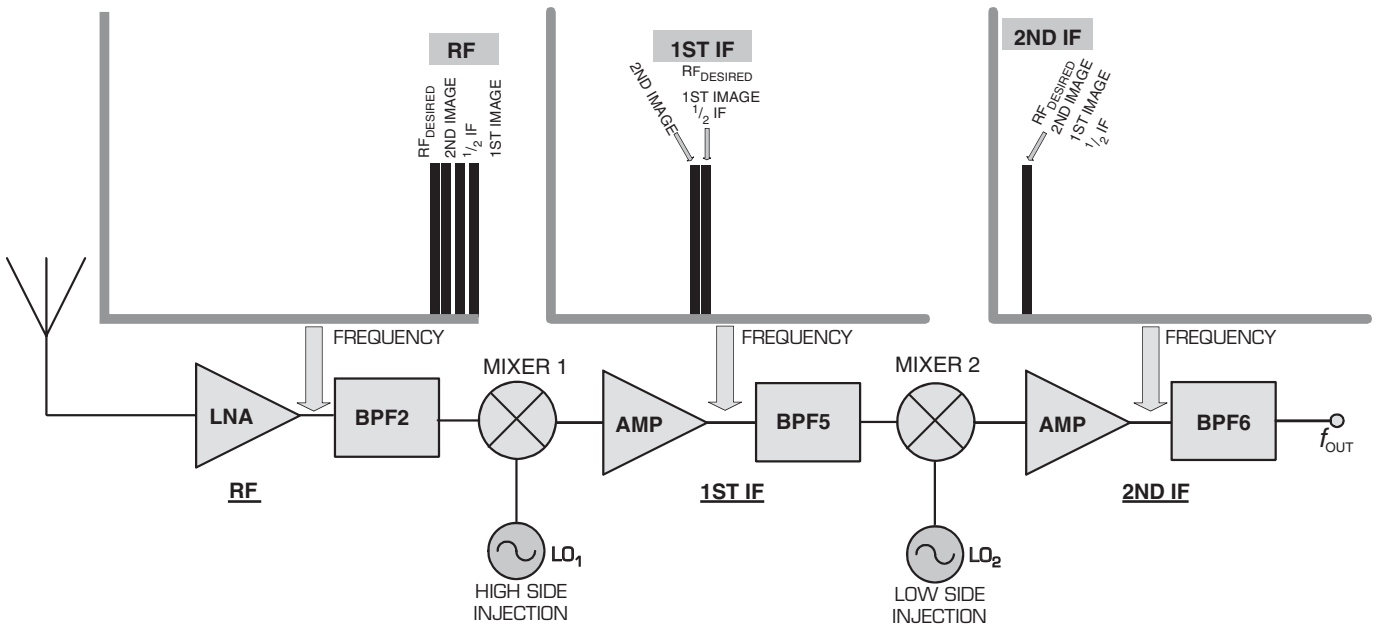


Figure 9.1 Relative locations of common receiver spurious responses for the 1st and 2nd image and the half-IF frequencies in a double conversion receiver, with the LOs set as shown.

Taking the first case of high-side injection, the image is any signal or noise at the frequency that differs from the LO by the amount of the IF, just as the desired signal does, but is higher instead of lower than the local oscillator frequency. Subtracting the desired signal from the local oscillator frequency will give the IF which will, of course, easily pass through the receiver's IF amplifiers. But any frequency (the image) that is higher than the LO by the same amount that the signal is below the LO will also give us this same frequency. This frequency will easily pass through all the IF amplifiers, and create receiver interference and a decrease in the signal-to-noise ratio (SNR).

As mentioned, the dominant technique for attenuating this undesired image frequency is by front-end filtration. The image filtering can be further assisted by using a high IF in order to remove the image as far away from the desired frequencies as possible. This will make the filtering of the image a much easier task, and with a decreased risk of excessive group delay variations caused by an exceptionally tight filter. Maintaining this first image far from the desired signal frequencies is assisted by utilizing double or even triple conversion receiver designs. With these multiple conversion receivers, the first IF is at a high frequency, while the second and third IFs are progressively lower. These lower IFs will supply most of the selectivity and gain, since the lower the frequency the cheaper, and the more stable, sensitive, and selective, the amplifiers will be. This is due to the lower frequency circuit's decreased stray reactances, fewer feedback paths, and lower cost components.

If left unfiltered, the image would contribute up to 3 dB of excess noise. So, the more we can reject the image frequency the more the image noise will decrease. For instance, if we attenuated the image frequency at the mixer's input by 10 dB, then our receiver noise figure would only increase by 0.4 dB (instead of 3 dB). By attenuating the image frequency by 20 dB, the image noise will be removed almost completely, with a noise figure increase contribution of only 0.04 dB. If we chose to not filter the image at all, and simply permitted the image noise to enter the receiver unattenuated—and if we wanted to maintain the same level of system sensitivity as a fully image-filtered radio—that would require that we compensated for the added 3 dB of excess noise. We could do this by doubling the system transmitter's RF output power, which is very expensive in both price and DC current demands. So, we can see why it is vital to filter the image band before it ever reaches the IF stages.

The first IF is not the only area of the receiver that has image frequency issues. Any receiver with a second conversion stage may suffer from the effects of an image, in this case the *second image*. Thus, sufficiently tight first IF filtering must be used in order to remove spurs that impact the second IF's image frequency band. Therefore, first and second image rejection of at least 50 dB into the second IF is desired. To clarify this further: the first image must be filtered at RF, as stated, but the receiver's second image frequency must be filtered within the first IF stages. If this is not done, the first image will be spot-on the first IF's frequency band, while both the first and second images will be within the frequency band of the second IF. And since the second image will always be within twice the frequency of the second IF (at $2 \times f_{2ND-IF}$) of both the RF and the first IF frequencies, and the second IF is normally quite low in frequency, we can see that the filtering of the second image at RF is impossible, and that a tight first IF filtering scheme is the only option.

Indeed, due to the second image frequency's closeness to the first IF frequency, sometimes the only way to effectively attenuate the second image is to utilize a surface acoustic wave (SAW) filter. This particular filter must be laid out on the printed circuit board (PCB) for maximum input-to-output port isolation, which is required to prevent

the second image from simply blowing-by the SAW's high insertion loss and steep stopbands via the circuit board's many sneak RF paths.

When manipulating a spreadsheet (or formulas) to calculate the noise figure of a receiver, an important thing to keep in mind is that such a spreadsheet will not supply us with the receiver's *worst-case* noise figure, since it cannot take into account the image noise contribution (nor will it pay any attention to the consequences of bandwidth, voltage standing wave ratio (VSWR), and sneak noise paths). Thus, if the image noise is not attenuated, 3 dB must be added to any noise figure obtained via a spreadsheet calculation. This added noise figure (NF) contribution is significant, and can literally destroy an otherwise perfect system design if not expected.

The removal of the image frequencies and the image noise is typically taken care of by two separate filters in many superheterodyne receivers, especially in higher end units. These particular filters are called the *image frequency reject* (IFR) and the *image noise reject* (INR) filter. An IFR filter will normally be placed just before the receiver's low-noise amplifier (LNA). However, it cannot remove the wideband LNA-generated noise that would still be injected directly into the mixer's port. The mitigation of this particular noise source (as well as some of the image frequency) requires an INR filter, which is placed between the LNA and mixer stages (Fig. 9.2).

9.1.3 Receiver Noise Figure

Low noise figures for a receiver means utilizing a low NF and high-gain amplifier in the receiver's front end, since any losses located before the receiver's LNA will correlate dB for dB in an increased NF. Therefore, 3 dB of filter loss before the LNA equals 3 dB more NF, which translates into a far less sensitive receiver. Unfortunately, there is a slight compromise that must be made in LNA design: for the highest front-end gain, the LNA must be matched to the receiver's front-end filters. But this matching can increase the noise, since an optimum NF match will rarely coincide with a high return loss (or low mismatch loss) match. So, considering the first stage of a receiver is required to be designed for the lowest NF and the highest gain, we will normally match for the best NF that the amplifier can provide, while providing acceptable gain.

The importance of a low NF in a receiver is demonstrated by its linear relationship with SNR. For instance, if a receiver has a NF of 12 dB, and the SNR at its output is 20 dB, we can improve the SNR to 30 dB, while still using the same input signal power and receiver gain, by decreasing the receiver's NF to 2 dB. This is a 10 dB improvement in NF *and* in SNR. Thus, if a certain signal power is placed at the input of a 5-dB NF receiver, and this creates an output with an SNR of 20 dB, then the actual *input* SNR to the receiver will have been 25 dB. Therefore, the receiver added 5 dB to the noise at its output, and the receiver's NF directly correlated into a decrease in the output signal's SNR. This also influences the power that the transmitter must send to the receiver's antenna to make up for the decreased SNR caused by a receiver's poor NF, with a relationship that is also dB for dB.

As any stage placed after the LNA will only add to the receiver's overall noise figure by the second stage's NF *divided* by the first stage of gain, we can see by glancing at the formula below the importance of LNA gain:

$$\mathrm{NF}_{\mathrm{OVERALL}} = \frac{\text{2nd stage NF}}{\text{1st stage gain}}$$

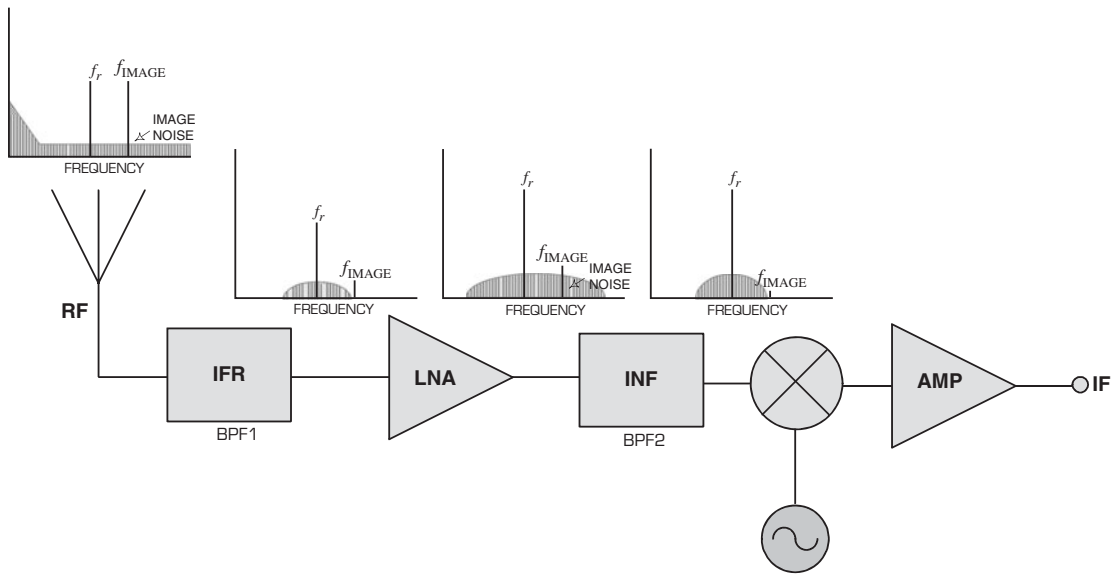


FIGURE 9.2 Filters used to suppress image frequencies (IFR) and image noise (INF).

9.1.4 Receiver Dynamic Range

Operating in a metropolitan environment will mean that there will be many other radios broadcasting, with some at very high powers, and a few of these sources will have fundamental and harmonic frequencies that fall quite close to our own receiver's frequency. And since the receiver may also be of a portable type, the fixed transmitter our receiver is linked to may be transmitting from a distance that is anywhere from 35 miles to 35 ft away. This can place extreme stresses on the receiver's dynamic range, possibly causing high levels of odd-order IMDs (especially the third order) to drop into our receiver's IF bandpass. Poor BER performance will be the result in a digital radio, or decreased fidelity in an analog radio. Therefore, we would like a receiver that has both maximum, large, and small signal performance. In other words, a high dynamic range.

But small signal performance is restricted by noise, while large signal performance is limited by distortion products that are generated by the receiver's active devices and their noninfinite linearity. Consequently, small signals benefit greatest from decreasing a receiver's noise figure, while large signals benefit greatest from decreasing nonlinearities in the receiver. Unfortunately, optimizing one of these parameters will typically degrade the other. For instance, for best noise figure we would ideally want to place the receiver's first bandpass filter just *after* the LNA, but to prevent large out-of-band interfering signals from creating strong intermodulation products in the receiver, we ideally want to place this same filter *before* the LNA. In the end, however, small signals benefit most from a well-balanced design, since even the lowest noise figure receiver in the world would be almost useless if every in-band or out-of-band interferer was able to easily enter the receiver, causing massive distortion products that would wipe out our desired, but very weak, RF signal. We can lessen this problem by using an LNA and first mixer with a high third-order intercept point (IP3), as well as a front-end attenuator that is controlled by the AGC. The front-end attenuator is used only to reduce the desired input signal levels to an amplitude that will not overdrive the receiver's LNA or first mixer. This solution would be unacceptable for attenuation of undesired adjacent channels, since the frequency of interest, especially if at a low amplitude, would also be attenuated. The noise figure will, of course, increase as attenuation is increased.

9.1.5 Receiver Gain

The complete RF and IF gain for the entire receiver, from its front-end antenna input all the way to the output of the last IF, will typically be in the neighborhood of 125 dB. A minimum value for such gain would be 90 dB, but in some specialized down-conversion applications the gain may be as little as 18 dB. The receiver's cascaded gain figure includes gains as supplied by the amplifiers, and losses as caused by the filters, diplexers, switches, attenuators, passive mixers, and so on. The antenna, cable, and receiver gains/losses will finally be combined during the planning stage to confirm that there will indeed be a signal of sufficient amplitude available at the receiver's IF output, with enough power to drive the detector or modem at the lowest received RF signal levels expected, and with sufficient *fade margin*.

9.1.6 Receiver Reciprocal Mixing

A common problem in receiver design is called *reciprocal mixing* (Fig. 9.3), and will significantly decrease a receiver's SNR. Reciprocal mixing places the noise sidebands of the local oscillator directly into the IF of the receiver, and is caused by the heterodyning of any out-of-band interfering signals with the LO's own phase noise. There are two

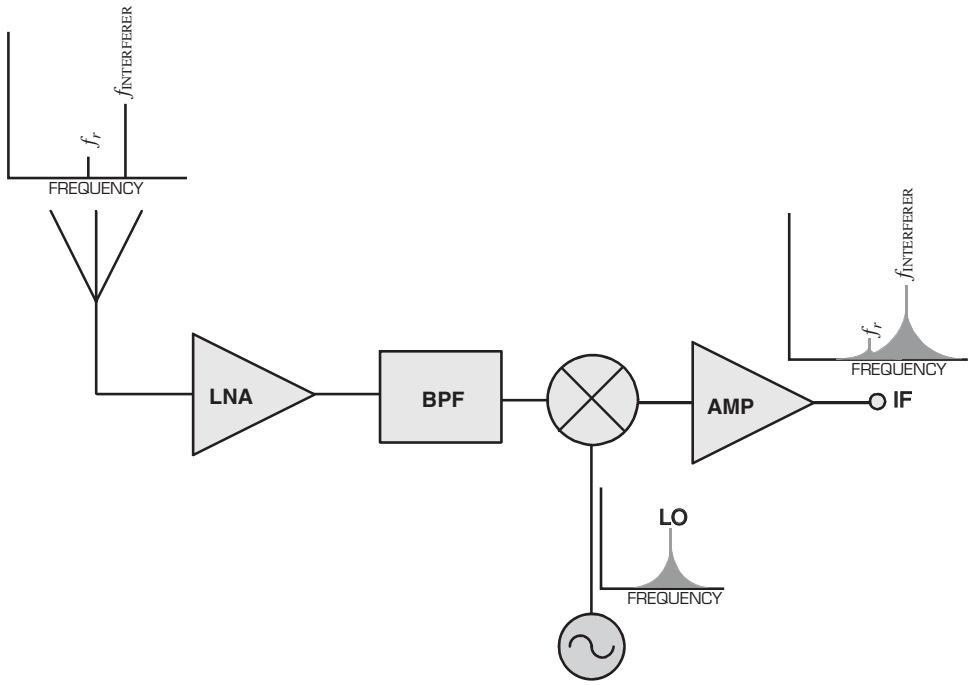


FIGURE 9.3 The effects of reciprocal mixing, increasing the noise floor for the wanted signal (f_r) through the mixing of the LO's phase-noise with the high-level interferer signal ($f_{INTERFERER}$).

ways in which we can reduce this reciprocal mixing: lower the amplitude of the interfering signal, and/or decrease the local oscillator's phase noise.

The reciprocal mixing action, which takes place within the receiver's mixer, will normally be a problem when the interfering signal is close to the desired signal:

$$f_{IF} = f_I + f_N \qquad \text{or} \qquad f_{IF} = f_I - f_N$$

where f_I = interfering signal, which is close to the desired signal, *Hz*
f_N = frequency of the LO noise sidebands, which are very close to either side of the LO frequency, *Hz*
f_{IF} = frequency of the receiver's IF, *Hz*

9.1.7 Receiver Half-IF Spurs

Half-IF spurs (as shown in Fig. 9.1) can be a problem when a receiver possesses a low IF frequency relative to its RF, and are created by the nonlinearities within the mixer stage permitting spurs into the receiver's IF by

$$\text{Half-IF (Hz)} = (0.5 \times \text{IF} + \text{RF})\ 2 - (2 \times \text{LO})$$

This formula indicates that any RF frequency that is at half the IF *plus* the RF frequency, and that substantially gets past the first RF filters before the first mixer stage, will be able to cause IF inband interference. This half-IF interference is due to the second

harmonic of this half of the IF frequency plus the RF frequency mixing with the LO's second harmonic, with the resulting difference frequency falling dead in the receiver's IF band. To mitigate this, any RF signals that are at half the IF plus the desired RF should be properly attenuated before they exit the first front-end receive filter.

These half-IF spurs in a receiver are fourth-order spurious or noise responses that are generated by the nonlinearities of the mixer and the frequency mixing of an interferer. The RF frequencies that cause half-IF spurs are located in frequency midway between the receiver's LO and our wanted RF signal, producing a second-order spur that heterodynes with the second harmonic of the LO, creating an inband spur in the IF, and can be an issue both with high-side and low-side LO injection schemes. When using low-side injection, any RF interferer at the receiver's front-end that is below the receiver's wanted frequency by $f_{IF}/2$ [or $f_{HALF-IF} = f_r - f_{IF}/2$ or $f_{HALF-IF} = \frac{1}{2}(f_r + f_{LO})$]. For high-side injection, any RF interferer at the receiver's front-end that is above the wanted received frequency by $f_{IF}/2$ (or $f_{HALF-IF} = f_r + f_{IF}/2$) can cause a spurious response in the IF passband. For instance, if we had a low-side injection receiver that was tuned to 2 GHz (f_r), and the receiver has an IF of 100 MHz (f_{IF}), then we could expect that any signal at a frequency of $f_r - f_{IF}/2$, 2 GHz – 100 MHz/2 = 1.95 GHz, will cause a signal spur to be created directly on the receiver's 100-MHz IF frequency. We can check this result by

$$f_{IF} = (2 \times f_{HALF-IF}) - (2 \times f_{LO})$$

Any mixer with a high *IIP2* specification (greater than 45 dBm) will probably not generate a significant half-IF spur nor half-IF noise, since the spur is, as mentioned, caused by second-order distortions within the mixer.

9.1.8 Receiver Phase Noise

Phase noise (Fig. 9.4) is a noise similar to a modulated spectrum created by a virtual noise source that is phase modulating the desired signal. All real-world receiver local oscillators are not perfect single continuous wave (CW) frequency sources, but possess noise sidebands, called phase noise. Any receiver's modulation scheme that uses phase variations to communicate, especially dense modulations such as QAM-16 and above, are severely affected by the presence of this phase noise, since low phase noise is vital to maintain a receiver's SNR and BER (See "Receiver Reciprocal Mixing").

Phase noise is measured as the ratio between the fundamental frequency and a noise sideband (in dBc), with the noise sideband located within a 1-Hz bandwidth at a

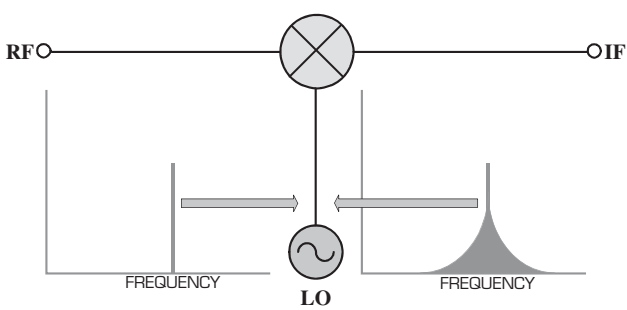


Figure 9.4 A mixer's local oscillator without phase noise (left), and with phase noise (right).

particular frequency that is offset from the LO's fundamental (in kHz). This phase noise specification may look like "–95dBc/Hz@10kHz", which in this case means that the receiver's LO phase noise is measured at a 10-kHz offset from the LO's fundamental output frequency, and the noise power (in dBm) within a 1-Hz bandwidth is 95 dBc below the amplitude of the LO.

9.1.9 Receiver Systems Design

We must decide on the optimal system specifications during the wireless receiver's planning phase, such as the most favorable IF/LO frequencies, mixer type/performance, filter topology/bandwidth, RF and IF amplifier IP3/NF/S21, and the best possible arrangement for all the required stages within the cascade. This demands the use of a receiver or system simulation program or spreadsheet, such as RFcafe's *RF Workbench*, which is included on the enclosed CD, or *RF Cascade Workbook,* also by RFcafe.

Receiver Cascade

Most receivers are of the *down-conversion* type. These receivers will take the RF input and immediately begin to convert the RF to either a single lower IF, or down to two or more increasingly lower IFs. The other type of receiver, called the *up-converting superhet*, is operated in wide tuning-range applications, and is especially common in certain HF SSB (single-sideband) ham radios. The up-converting receiver takes the incoming RF and converts it to some higher frequency, typically about twice the highest expected receive frequency, in order to assure simple RF front-end filtering by distancing the image frequency as far as possible from the LO. Up-converting is not seen at VHF and above.

In most superheterodyne receiver designs, an RF-to-IF ratio of no larger than 10:1 is used when converting the RF to the IF. Thus, if the incoming RF frequency is 1 GHz, then the IF should be at or near 100 MHz. However, we do not want an IF that is located at the same frequency as a harmonic of the system's digital or PLL clocks. Nor can we simply select just any IF frequency we desire. We would normally attempt, for cost and availability reasons, to only select popular IF frequencies that are in common use.

A standard double down-conversion superhet receiver block diagram is shown in Fig. 9.5. First, the antenna picks up the electromagnetic waves from the environment and, due to certain antenna's natural passive gain in a specific direction, may amplify any signals that are within its bandwidth. To prevent static buildup on the antenna from entering the receiver, the inductor L_1 is shorting any such charge to ground, since the static could possibly damage the delicate LNA of the receiver's front-end. L_1 can also be an inductor within (band-pass filter) BPF1 itself, or as part of an input matching network. The RF signal is placed into the input of the receiver's BPF1 from the antenna. BPF1 is a filter, sometimes referred to as a *preselector*, utilized to select a band of frequencies and reject any out-of-band signals, therefore minimizing IMD product generation. However, this filter must not have so tight a passband that insertion loss (IL), NF, and GDV are increased excessively. BPF1 will also help reject the first LO frequency from radiating its CW signal back through BPF2 and the LNA. The LNA, due to its reverse isolation, can significantly attenuate this reradiation by itself, and by also selecting the proper mixer with an adequate amount of port isolation. The added ability of a discrete LNA to be designed and tuned to assist the receiver front-end in the rejection of interferers and noise should also not be ignored. Using a wideband monolithic microwave integrated circuit (MMIC) in this application would be far too broadbanded to be of any help, but the high-Q input and output impedance of most

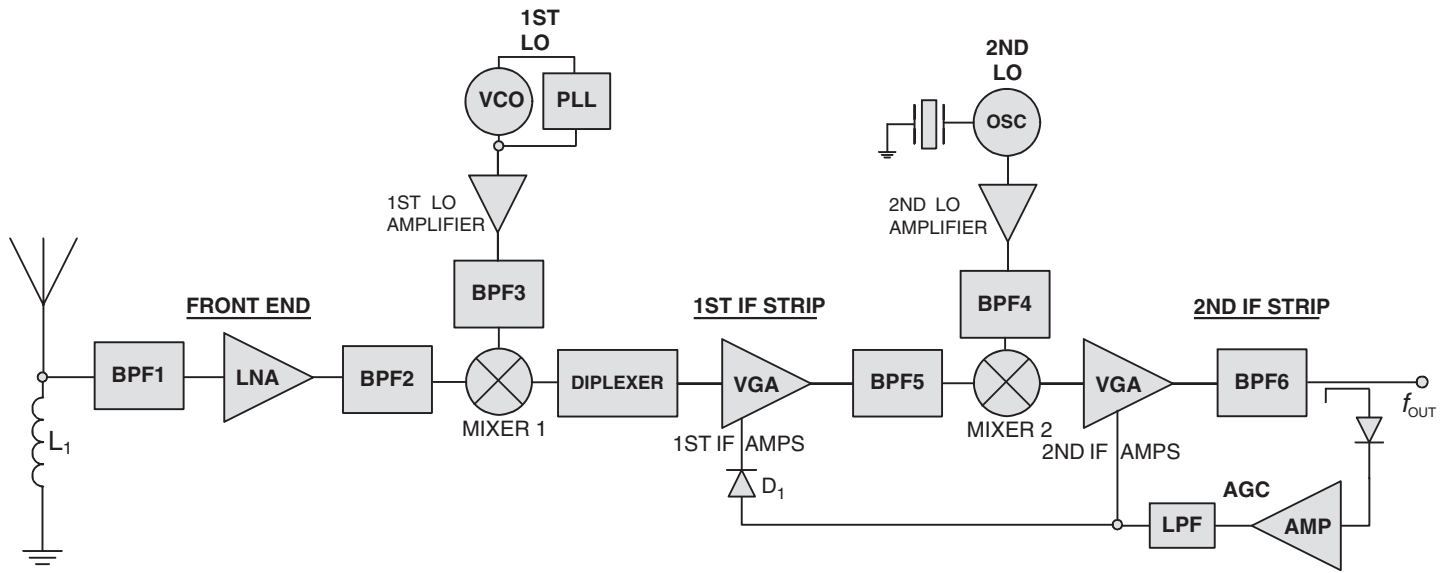


Figure 9.5 A superheterodyne dual-conversion receiver.

discrete LC-matching networks can significantly assist in rejecting these out-of-band signals. BPF1 will further aid BPF2 in the attenuation of image frequencies and image noise located at (2 · LO) – RF, or LO – IF. Due to the predominant noise contribution of these first stages of the receiver, image noise at any second or third mixer stage can normally be ignored, but not the image frequencies. Therefore, sufficiently tight first IF filtering must be used in order to remove spurs that impact the second IF's image frequency band. In fact, the frequency of the IF itself is chosen to permit a receiver to reject this first RF image frequency without requiring excessively expensive, complex, and tight image filters before the first down mixer, MIXER1.

The next stage after BPF1 is the LNA, which will set most of the receiver's noise figure and IMD performance, and gives us approximately 20 dB of front-end RF gain, along with a low NF of less than 2 dB. Next, BPF2 attenuates any harmonics created by the nonlinearity when large signals impact the LNA, as well as attenuating the added image noise caused by the LNA circuit itself (no amplifier can be completely noiseless). Further, BPF2 will assist in rejecting some of the out-of-band signals and LO feedthrough. If the receiver is required to operate only within the HF region or below (< 30 MHz), then both the LNA and BPF2 can be dropped from the entire receiver design, as noise figure is not very important due to the truly huge natural and man-made noise that is generated within this region of the frequency spectrum. This will decrease HF receiver sensitivity and increase NF, but will give far better IMD immunity in the crowded and noisy HF bands. In low-cost receivers that may have but a single image filter in their front-end, BPF1 would be used to decrease the amplitude of out-of-band signals, as well as provide all of the image frequency filtering, without the assistance of BPF2. But, considering that filter insertion loss located *before* the LNA translates directly into an increased NF, we see why another image filter, BPF2, is usually placed between the LNA and the 1st mixer to decrease the higher IL below that of what a lone BPF1, if it were indeed the only filter in a single filter design, would need to possess. The two filter front-end design does this by sharing the image filtering chores, and thus the insertion losses, between BPF1 and BPF2; and with the added advantage, as mentioned above, of almost completely removing the LNA's generated noise.

MIXER1 will normally be subjected to high amplitude input signals, so a high compression point is mandatory here to decrease mixer-generated intermodulation distortion. This requirement will generally demand the use of a high-level diode double-balanced mixer (DBM), instead of an active mixer stage, in this sensitive receiver location. For decreased IMD generation, the RF signal into the DBM should be at least 10 dB less than that injected into the DBM's LO port. Therefore, a Level 10 mixer could safely accept a maximum 0 dBm signal at its RF input port before excessive IMDs started to became a problem. A separation of 15 to 20 dBm between the two signals would be even better, and would be necessary for higher quality, lower distortion receivers. DBMs also have the added advantage in that they suppress even-order LO and RF mixer-generated harmonic products, as well as the RF and LO frequencies themselves, at the IF output. Attenuation of these frequencies is far from complete, however.

The DIPLEXER, placed after MIXER1, will filter and pass all the desired IF frequencies but, unlike other typical filters, a diplexer will stop other frequencies from entering the IF's bandwidth through *absorption* rather than *reflection*. Reflection of the undesired frequencies, such as LO harmonics, the sum of the RF and LO, and the IMD products, would cause RF power to be reflected back into the mixer's IF port, which can unbalance its diode ring, causing increased IMDs, as well as adversely affecting the mixer's dynamic range and conversion loss. Indeed, many practical receiver architectures

may simply pad the output of the mixer ports so that the reflections are not only attenuated as they enter the pad, but also as they are reflected back into the mixer's IF port. As well, the pads will lower the nonlinear mixer's input/output VSWR by supplying each port with an almost pure $50 + j0$ termination. Any pads placed at the RF port will, however, introduce thermal noise at the image frequency, which would normally have been removed by the image noise filter. To improve the rejection of reflections back into the MIXER 1 port, a wideband, high isolation amplifier may also be used at the mixer's IF output, as shown in the second IF strip (the variable gain amplifier [VGA]). This technique will permit all of the mixing products to pass through this isolation buffer amplifier and, after filtering from a normal, reflective IF bandpass filter, the filter will then naturally "bounce" much of the undesired RF signals back toward the sensitive IF port of the mixer. However, these reflected signals will have been significantly attenuated by the reverse isolation of this wideband amplifier. Any such buffer amplifier used for this purpose must have a high $P1$dB, so as to linearly accept the sometimes high-powered out-of-channel signals that can occur, and without producing significant distortion.

To mitigate BER and adjacent channel selectivity degradation, the LOs of Fig. 9.5 must generate low phase noise. This can be a difficult requirement in phase-locked loop-based LOs due to their low Q, but is relatively easy in high Q crystal oscillators. In addition, to minimize receiver spurious responses, all LO-generated spurii must be as low as possible. To minimize VSWR, the LO AMPLIFIERS are broadband types that are used to buffer the local oscillator from the nonlinear LO mixer port, and thus lower mixer-generated IMD products. These buffer stages will also increase the LO's output power to the nominal levels needed to maintain the mixer's NF and conversion losses. When designing the LO chain, the LO AMPLIFIER should be somewhat overdriven for maximum power output and flatness, since the LO signal itself will only be a single CW frequency. Therefore, no IMDs will be generated; only harmonics, which are easily filtered out.

Both BPF3 and BPF4 are present to reduce wideband LO noise, harmonics, and possibly any subharmonics that may be present in the oscillator's outputs, since reduction of generated noise improves the mixer's NF and the receiver's sensitivity, while harmonic suppression prevents a decrease in the mixer's IP2 (second-order intercept point) and the ensuing generation of increased LO second harmonics.

The first IF chain of the receiver will furnish some gain and filtering, along with delayed AGC control of any IF amplifiers, before injection into MIXER2. If desired, this first IF filtering of the DIPLEXER and BPF5 will also remove the second image frequency from reaching MIXER2, but the filtering itself must not be so tight as to introduce excessive group delay variations, which can cause increased BER. The first IF filtering is present mainly to provide some channel selectivity and the rejection of spurious signals, since image noise into MIXER2 is not a big concern if relatively high gain stages were adopted in the receiver's front end, which is where the NF is set. Indeed, the noise floor at MIXER2 will mainly be noise that was amplified by the LNA. The band-limiting first IF filter cascade must possess a bandwidth and rejection that will only just suppress external and internal spurs, images, harmonics, and noise, but not be so tight as to create excessive amplitude distortion, group delay spread, nor high insertion losses.

Since lower frequency circuits are cheaper and more stable, the second IF amplifiers and filter stages supply most of this receiver's gain and selectivity. However, to prevent amplifier oscillations and receiver interference, such high-gain stages should be shielded

against EMI, which can be emitted from other internal stages or from external sources. By the use of VGAs (*variable gain amplifiers*) within the second (and first) IF strip, the receiver's AGC loop will be able to increase the receiver's total dynamic range. The AGC will not only control the gain of the second IF stages, but also set the gain of the first IF strip, as well as the LNA (in some receivers). BPF6 will filter the spurious output of MIXER2, outputting f_{out} into an internal detector stage or external modem.

Frequency Planning

The receiver's internal frequency plan can make or break a design. The proper RF and IF bandwidth and the exact LO and IF frequencies must all be selected carefully, or serious interference and mixer-generated spurious responses can greatly decrease the expected receiver performance, sometimes rendering an entire design almost useless. In order to catch most mixer-generated spurs, such as $f_{RF} - 3f_{LO}$, $f_{RF} - 5f_{LO}$, $f_{RF} + 3f_{LO}$, $f_{RF} + 5f_{LO}$ and so on, it will be necessary to display these spurious output frequencies, and their amplitudes, by employing the proper RF system simulation software, such as the enclosed RF Workbench. If it is found through simulation that the specifications cannot be met with the current receiver design, then more IF filtering, new LO and IF frequencies, and/or a new mixer topology may be required. This is because we do not want any strong spurious frequencies to fall within the bandwidth of the receiver's IF due to the many generated nRF ± nLO mixer products, which would cause the BER to be degraded in a digital receiver, while even the reception of undesired signals and interference of an analog receiver is possible within our desired channel. Since these internally generated spurs will degrade the highest potential receiver sensitivity by masking our desired low-level signals, their in-band amplitudes should be no more than

$$P_{SPUR}\,(\text{dBm}) = -171 + \text{NF} + 10\log\,(\text{BW})$$

where P_{SPUR} = Amplitude of spurious frequencies, dBm
NF = receiver's input noise figure, dB
BW = bandwidth of the receiver's IF, Hz

As stated above, high-side injection is when the LO frequency is greater than the RF frequency in a conversion stage, while low-side injection is when the LO frequency is actually lower than the incoming RF frequency. The choice of whether to operate a superheterodyne receiver with a high- or low-side injection will depend on whether the final demodulated signal needs to have the sidebands with an inverted or noninverted frequency spectra. In digital communications, this frequency inversion can be a large consideration, since the output of the IF is typically sent straight into a modem, which may or may not require inversion. But even if we do not have a choice as to whether we can frequency invert or not, we may still select our high- or low-side injections throughout both the transmitter and receiver stages, and base our high- or low-side preferences on the availability of oscillators, multipliers, and phased-locked loops (PLLs), as well as their cost and design complexity, along with the location of our undesired image frequency and the undesired mixer spurs.

Receiver System Calculations

There are several important design formulas to assist the designer in calculating the receiver's specifications, and these are presented below. However, it will be far easier and more accurate to obtain answers to these important receiver calculations by utilizing

the included AppCAD or RF Workbench programs, or the appropriate receiver cascade worksheet. Such software will, within seconds, effortlessly compute total receiver NF, SNR, IP3, minimum discernable signal (MDS), gain, sensitivity, noise floor, dynamic range, and the like.

a. The total gain required of a receiver can be calculated by finding out what the lowest expected RF signal level will be after the antenna (i.e., at the receiver's input port), and deciding on what the minimum receiver output signal requirement is into the modem or detector:

$$G_{dB} = P_{OUT} - P_{IN}$$

where G_{dB} = required gain of the receiver, dB
P_{OUT} = lowest acceptable signal output level of the receiver, dBm
P_{IN} = lowest expected RF signal level into the front end of the receiver, after the antenna, dBm

---

NOTE: To find the receiver's AGC range needed to control the gain for the IF strip, and in some cases for the RF stages, calculate the difference in dB between the lowest RF signal expected that will still be able to supply the desired amplitude at the receiver's output, and the highest RF signal expected. This will be the range, in dB, that the AGC must lower *the gain of the receiver to maintain the nominal output signal level, even as the RF signal amplitude increases drastically.*

b. Minimum discernable signal (MDS) is a sensitivity rating for receivers, and is the lowest signal detectable. This can be at 0 dB (sometimes at 3 dB) above the receiver's noise floor, and can be calculated by

$$\text{MDS (dBm)} = -174 \text{ dBm} + 10 \log_{10} (\text{BW}) + \text{NF}$$

where BW = noise bandwidth of the receiver, or approximately the 6-dB down bandwidth (instead of the typical 3-dB bandwidth), Hz
NF = receiver's noise figure, dB

c. The IP3 is approximately 10 to 15 dB above the *P*1dB compression point, and is the location where, if the gain slope of the receiver could continue, the undesired output third-order frequency products would be at the same amplitude as the output two-tone fundamental test signals that had been originally placed at the receiver's input. To compute the total cascaded input IP3 (IP_{TOT}) of multiple stages of a receiver:

$$IP_{TOT} = \frac{1}{\frac{1}{IP3_1} + \frac{GAIN_1}{IP3_2} + \frac{GAIN_1 \times GAIN_2}{IP3_3}}$$

where IP_{TOT} = receiver's total IP3 from its input to its output, dBm
IP3*n* = IP3's linear term, not in dB
GAIN*n* = gain's linear term, not in dB

d. The NF of a receiver can be calculated by using *FRIIS' equation*, with the noise figure itself not requiring a reference to any particular bandwidth, since it is a ratio between the input and the output of a receiver (or amplifier) over the same bandwidth:

$$NF_{TOTAL} = 10 \log \left[(10^{NF1/10}) + \frac{(10^{NF2/10}) - 1}{(10^{G1/10})} + \frac{(10^{NF3/10}) - 1}{(10^{G1/10}) \times (10^{G2/10})} + \frac{(10^{NFn/10}) - 1}{(10^{G1/10}) \times (10^{G2/10}) \times (10^{Gn/10})} \right]$$

where NF_{TOTAL} = total NF for the entire receiver, dB
NF1 = noise figure for the first stage, dB
NF2 = NF for the second stage, dB
NF*n* = NF for the *n* stage, dB
*G*1 = gain for the first stage, dB
*G*2 = gain for the second stage, dB
Gn = gain for *n* stage, dB

or

To calculate the required NF for a receiver if we wish a desired receiver output SNR. These calculations call for the receiver's planned bandwidth:

1. Calculate the thermal noise power within a certain receiver's bandwidth.

$$KTB = -174 + 10 \log(BW)$$

where KTB = thermal noise power within the specified BW, dBm, BW = receiver's bandwidth, Hz.

2. Calculate the lowest sensitivity that will obtain a signal at the specified SNR at the receiver's output:

$$SENS_{LOW} = (-KTB + SNR)$$

where SNR = receiver's desired output signal-to-noise ratio, dB, KTB = thermal noise power, dBm.

3. The NF required to *just* meet the sensitivity specifications:

$$NF_{MAX} = SENS - (-SENS_{LOW})$$

where SENS = required sensitivity of the receiver, dBm
$SENS_{LOW}$ = sensitivity of the receiver at 0 dB NF, for the specified SNR, dBm
NF_{MAX} = maximum noise figure that the receiver can be and still satisfy the sensitivity requirement of SENS, dB

As an example of a system's NF calculation using the formulas above: If a receiver under design has been specified to have a sensitivity of −100 dBm, what is the maximum NF we can permit the receiver to have, with zero sensitivity margin, if the receiver's bandwidth is 100 kHz?

A. First, calculate: KTB = −174 + 10 log(100,000) = −124 dBm.

B. The modulation to be received will need a minimum of 18-dB SNR for a certain desired BER. Calculate what the sensitivity of the receiver would be with 0-dB NF: $SENS_{LOW}$ = −124 + 18 = −106 dBm.

C. What is the maximum NF allowed for this receiver with zero margin? The answer is: NF_{MAX} = (−100) − (106) = 6 dB

NOTE: In reality, a safety margin would have to be added to the above calculation, the reasons for this are presented later. To add a safety margin, the NF of the receiver must be reduced even further. So, to increase the sensitivity margin by 3 dB, the NF would have to be lowered to 3 dB (6 dB − 3 dB = 3 dB). Depending on the particular radio service, more margin may be required, especially if the implementation margin *is taken into account. (The implementation margin is the loss in system specifications that occur from the design stage to the actual implementation stage, as well as due to mass production variances).*

9.2 Transmitters

9.2.1 Introduction

If not properly designed, a transmitter may create RF interfere with not only other wireless devices, but with many different types of non-RF electronic equipment. Harmonic and spurious outputs, wideband and close-in noise, frequency and amplitude stability, and peak and average output powers, are but a few of the critical parameters that must be addressed before any transmitter design can begin.

Transmitter spurious signals, generated by the mixing of its LO and IF, along with their harmonics, are of particular concern, as are two-tone intermodulation products created by two or more frequency components mixing together within any stage's nonlinearities, especially up to the fifth order ($3f_1 \pm 2f_2$ or $3f_2 \pm 2f_1$). Other undesired transmitter output signals, such as harmonics of the RF carrier signal and the feedthrough of the LO and IF frequencies, can all cause further interference. Transmitted noise, especially in a multipoint environment, will raise the noise floor of the receiver at the other end of the link, therefore lowering its SNR, which will decrease the distance the RF communication link can reach. Depending on the frequency, power, band, modulation, and wireless service, certain frequency stability requirements are mandated by law, or are simply required for proper demodulation at the distant receiver.

9.2.2 Transmitter Systems Design

The generic double-conversion transmitter of Fig. 9.6 could be a single-conversion unit if the RF output frequency was not too high. In fact, the choice as to whether to employ single or double conversion transmitters is based on the final transmit RF frequency versus the much lower baseband modulator or modem's input frequency. Since at higher output frequencies double conversion is normally required to properly suppress, through the IF filtering, the close-in sum or difference signals of the input baseband signal mixing with the LO stage, as well as the undesired LO feedthrough, while also suppressing all mixer spurious responses. Therefore, if we have a low input baseband

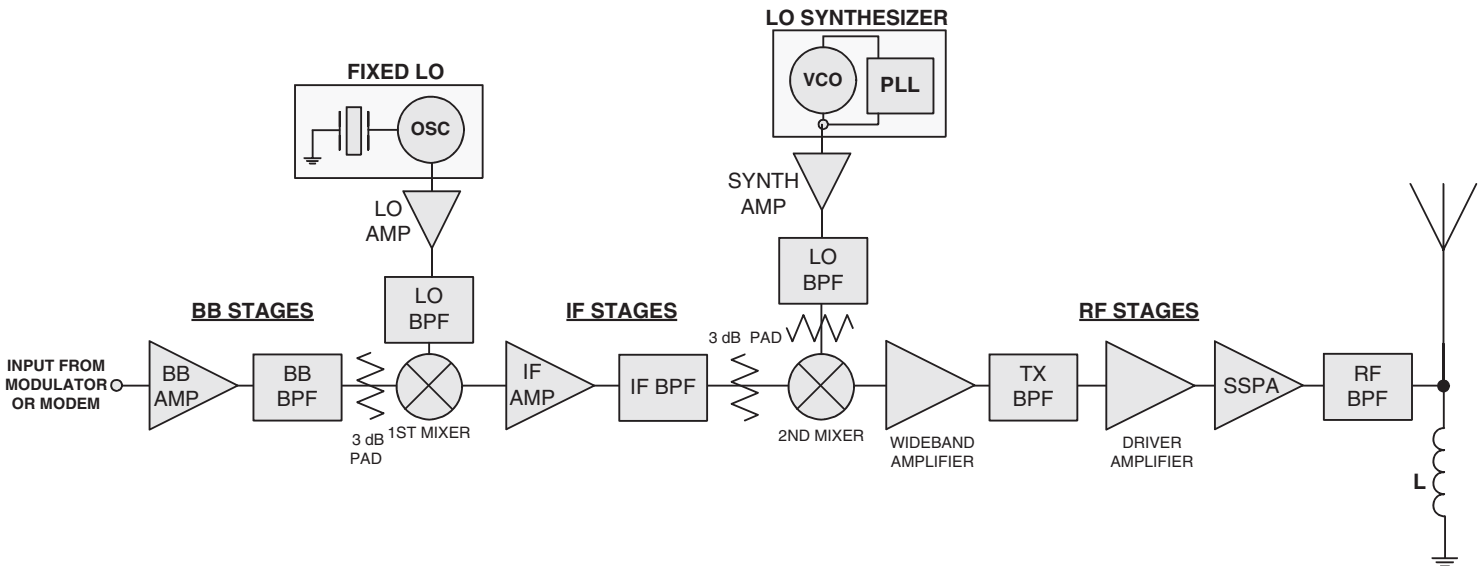


Figure 9.6 One type of linear RF transmitter.

frequency that must be converted to a much higher RF carrier frequency, yet we are only willing to employ a single conversion transmitter architecture instead of a double conversion design, we would need to suppress these frequencies by employing IF filters that have very difficult to realize narrow passbands and ultrasteep skirts which, even if actually achievable and affordable, would also increase the problem of high group delay variations at the IF filter's band edges, possibly distorting the entire modulated RF signal.

Taking the first element of the cascade of Fig. 9.6, the transmitter's antenna, we see that it is at DC ground through the RF inductor, and is used to protect the output filter RF BPF and the solid-state power amplifier (SSPA) against static buildup discharge damage. This chore may be part of the RF BPF or antenna-matching network. The RF BPF suppresses much of the transmitter-generated harmonics, wideband noise, IMD products, and out-of-band conversion frequencies. If the transmitter is to be employed in FM or FSK service, the SSPA must be run saturated for not only maximum efficiency, but also to assure a constant output power from transmitter to transmitter. However, this nonlinear operation will create large harmonic output levels, which must then be sufficiently attenuated by the RF BPF output filter. In fact, since a typical RF filter is highly reflective within its stopbands, most of these undesired harmonic frequencies will be reflected back into the SSPA's output, which will then create significantly higher than expected harmonic outputs from the power amplifier. This problem is further exacerbated by the lack of a perfect 50-Ω match at the output of the nonlinear power amplifier, with both of these issues necessitating a filter that must possess a much higher stopband attenuation level than that originally calculated by approximately 15 to 20 dB.

If the SSPA is to be operated at less than saturation, as required for digital or SSB voice communications, the amplifier must be designed to maintain low distortion levels and the proper RF output power. This means we may have to run the power amplifier up to 10 dB (or more) under its maximum RF output power rating. To put it another way, the SSPA will have to be backed off in output power by up to 10 dB to maintain the required linear operation that a particular modulation technique may demand, which will lessen spectral regrowth (a form of IMD), and maintain an acceptable BER. The SSPA stage, whether in linear or nonlinear operation, must also be exceptionally stable, and not begin oscillating nor decrease its output power with the wide impedance variations that it will encounter in mobile or portable operation. These impedance variations are caused by the antenna being constantly presented with many different conducting and dielectric structures that are passing nearby within the portable environment.

The DRIVER AMPLIFIER stage must supply the necessary RF input signal amplitude to the SSPA, without significant distortion level and with complete stability. The TX BPF filter must be tight enough to attenuate the LO feedthrough, the undesired sum or difference frequency, and other assorted mixer products, but not contribute significantly to the transmitted signal's group delay variations nor its amplitude ripple. The WIDEBAND AMPLIFIER stage is not always required, but will permit all the undesired nonlinear 2ND MIXER products to pass through to the TX BPF without being powerfully reflected by the TX BPF filter's stopbands and back into the 2ND MIXER. These mixer products that are reflected back into the diode ring of a passive 2ND MIXER stage will cause mixer diode imbalance, with the result of increased spurious outputs, and decreased third-order intercept points. Instead, when these products reach the stopbands of the TX BPF, they will still be reflected back toward the mixer stage, but the WIDEBAND AMPLIFIER, due to its high isolation, will attenuate most of these reflected signals, preventing the formation of increased IMD levels within the 2ND MIXER. Many other

transmitter designs may suppress reflections from the TX BPF's stopbands by the use of a diplexer or attenuator pads, or may even simply dispense with all of the above and place the TX BPF directly at the output of the 2ND MIXER port. This choice will depend on cost as well as the actual output power expected from the mixer stage. The less the output power and the lower in amplitude their reflections, then the less will be the requirement for their suppression.

To suppress IMD the 2ND MIXER itself can be a double-balanced type of a level 10 or higher. The LO port must typically have an amplitude that is 10 to 15 dB higher than the signal placed at the IF port, or excessive intermodulation products will result. As an example, if the LO port is at 10 dBm, then the IF input must be at or lower than 0 dBm. The 3-dB pad located at the LO output helps to present a 50-Ω impedance to the termination sensitive LO BPF. The LO BPF is used to suppress LO harmonics and wideband noise, improving the mixer's NF. The synthesizer's SYNTH AMP buffers the output of the LO SYNTHESIZER, minimizing VSWR, as well as increasing the LO SYNTHESIZER's output power to the mixer's nominal level needed to maintain its rated NF, IMD, and conversion loss specs. So as not to degrade the downstream receiver's SNR and BER, the LO SYNTHESIZER stage itself will have high stability and low-phase noise output. The 3-dB pad at the IF input to the 2ND MIXER may be used to maintain a more constant 50-Ω input impedance for the IF BPF, which will shift in characteristics if not presented with its exact design impedance. The IF BPF itself should be tight enough to reject the 1ST MIXER's products, the sum (or difference) frequency, the LO feedthrough, and excess noise, yet loose enough to minimize group delay variations and sideband cutting. Selection of the proper IF BPF topology and filter order is vital. The wideband IF amplifier stage of IF AMP is not only furnishing amplification, but will also have high reverse isolation to prevent the mixer products from the 1ST MIXER from reentering the mixer's IF output port. Moreover, the IF AMP will provide a reasonable 50-Ω termination for the IF BPF, since the ports of most mixers, because of their diode's continuous switching action, have difficulty maintaining $50 + j0$. The 1ST MIXER converts the low-baseband input signal up to the IF frequency, while the LO BPF filters wideband noise, harmonics, and spurii created by the FIXED LO and/or its LO AMP. The FIXED LO is a high-Q crystal oscillator, and thus is low in phase-noise and high in stability. The BB BPF and/or the BB AMP may or may not be present, depending on the modulation source and its requirements. This modulation source will typically be an I/Q modulator or a modem.

To maintain proper transmitted output levels by the automatic level control (ALC) circuit, the transmitter's RF output may be tapped by a directional coupler, which is utilized in order to feed this amplitude information to a microprocessor. To prevent SSPA damage or destruction, a temperature sensor signal and reverse power level signal may also be output to a controller, which would shut down the PA in case of high operating temperatures or excessive VSWR.

So as not to degrade the overall system's signal-to-noise ratio, a digital transmitter's PA stage should have an SNR of greater that 65 dBc, while the phase noise of the LOs for a typical QAM transmitter should be better than 95 dB/Hz@10 kHz. The digital transmitter's IF and RF filters must also pass the entire signal with no passband cutting or in-band ripple in excess of 0.5 dB, and with an amplitude tilt of less than 2 dB. In many of the radio services, the transmitted spurious output signals into all adjacent channels may also need to be 65 dBc or less, and must radiate as little noise as possible (any transmitted noise can swamp-out a low-amplitude desired signal in local adjacent receivers).

9.3 The Complete Communications System

9.3.1 Introduction

Complete communications system design must take into account many variables, such as the RF link itself, general transmitter and receiver specifications, types of modulations, data rates, BERs, and so on. Most of these issues have been covered in other sections of this book. This section, however, will concentrate on the overall technical issues and system architectures that affect wireless system performance.

At the beginning of any wireless design project, certain system parameters must be established. In a digital data link, we would need answers to some of the most basic of questions: What will be the system's frequency of operation? Will the system need to be full duplex or half duplex? Will transmit and receive be separated in frequency [(*frequency division duplex* (FDD)] or in time [(*time division duplex* (TDD)]? What is the required bandwidth, modulation, fade margin, gain, BER, SNR, phase noise, group delay, transmit power, and receiver NF? What number of conversion stages will we need for the transmitter and receiver sections? What is the transmitter and receiver's internal frequency plan, frequency stability, frequency inversion, dynamic range, and third-order intercept point? Modern radio systems must always be designed for low levels of phase noise, group delay variation, intermodulation distortion, amplitude ripple and shape, and frequency stability so as not to adversely affect the BER of their phase/amplitude modulated digital signal.

After all specifications and requirements have been completed for the entire communications system, a detailed block diagram should be created showing the gain, frequency, bandwidth, and signal levels for the system's receiver. This should be done at both its highest expected RF input signal level (to confirm that no internal stage is being overdriven) and at its lowest expected RF input level (to confirm that the output into the detector/modem is adequate). Transmitter section block diagrams must be suitably created as well. The system diagrams will assist in verifying that the internal gain, IP3, and bandwidth distribution is appropriate within the receiver and transmitter, as well as confirming if there is adequate spurious and harmonic suppression. Only after all this is accomplished can the actual circuit design for each stage begin.

9.3.2 TDD Transceiver

Fig. 9.7 illustrates a generic digital linear *time division duplexed* (TDD) transceiver with its own built-in modulator and demodulator. Other digital transceiver systems may drop internal modulation and demodulation entirely, and begin and end with a modem.

On the transmit side of the block diagram, the baseband digital data is input into the DAC, converting it into serial analog data which is filtered and sent into the I and Q inputs of the MODULATOR. The MODULATOR takes the I and Q signal and mixes them within their own separate DBM, with one DBM being fed an in-phase LO signal and the other a 90° out-of-phase signal. The DBM outputs are added in a LINEAR MIXER and output as a single IF signal, along with attenuated sidebands and the undesired carrier that were not fully suppressed. This lack of complete suppression is due to imperfect phase and/or amplitude mismatching between the I and Q legs. The IF is then amplified, filtered, upconverted, and amplified again before being sent out of the antenna. In certain lower frequency RF systems, the signal coming directly out of the modulator can sometimes be used for transmission into space, and would only need to be filtered and amplified by the transmitter before being sent into the antenna.

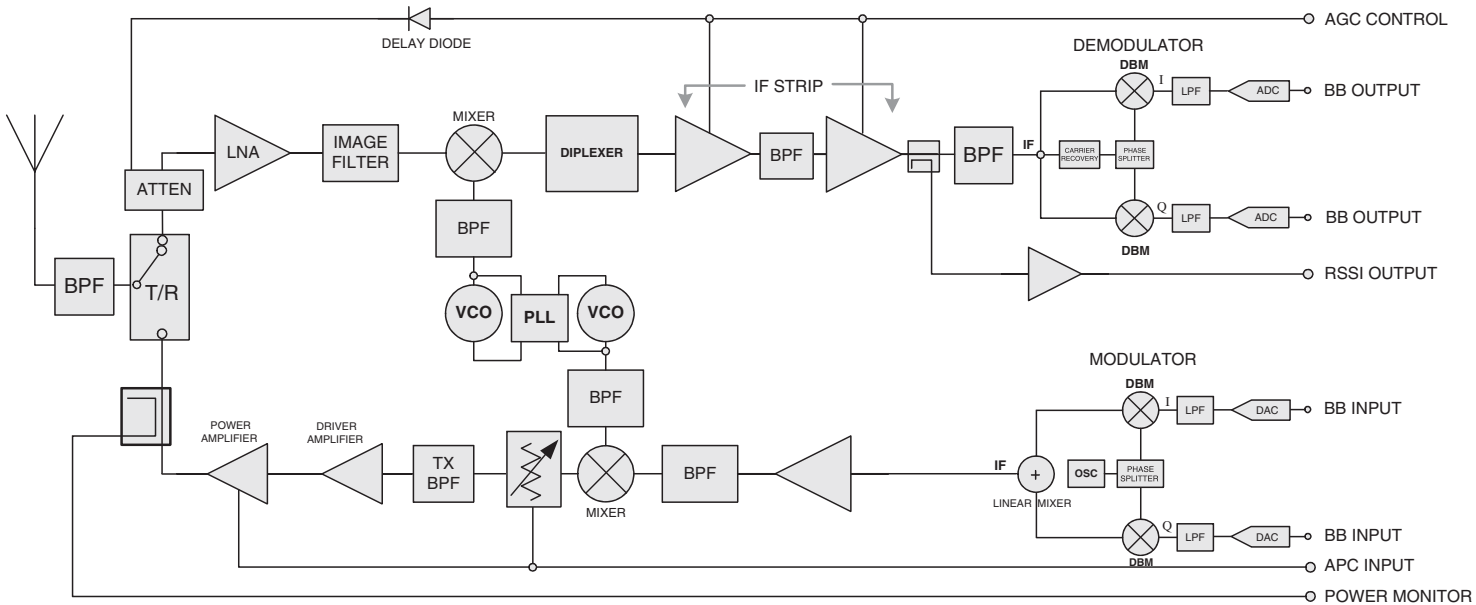


FIGURE 9.7 A complete TDD radio.

On the receive side, the transmit/receive (T/R) switch is switched over to route the incoming RF signal through the initially low insertion-loss front-end attenuator, which is employed to lessen very high-level RF input signals, and into the LNA stage, used to increase the signal power and lower the receiver's total NF. The RF signal then travels through the IMAGE FILTER, used to reduce image noise and interference, and into the down conversion MIXER stage to heterodyne the RF to the lower IF frequency. The resultant IF is then sent through the DIPLEXER absorptive filter to decrease reflection-generated mixer IMDs, and into the AGC'ed IF STRIP, employed to decrease/increase the signal level and supply selectivity. Finally, into the DEMODULATOR and ADC stages (or modem), where it is output as digital data.

9.3.3 FDD Transceiver

The following is an overview of the operation of a FDD transceiver.

In the full duplex FDD radio of Fig. 9.8, a DUPLEXER is placed at the transceiver's front end to allow simultaneous TX/RX on different frequencies while using the same antenna. This DUPLEXER must have a high enough level of attenuation to prevent the high-powered POWER AMPLIFIER's transmit frequency and noise from negatively affecting the very sensitive receiver section, while also attenuating all transmitter-generated harmonics and some of the receiver's own image frequency—and at the same time not adding excessive group delay variations.

On the FDD receiver side, the LNA will set most of the receiver's NF, and thus sensitivity, and must be high in gain and low in internal noise generation. However, simply deciding to design an LNA with the lowest NF and the highest gain will result in a receiver with poor intermodulation performance, since the 1ST MIXER's IP3 will be reduced by the gain of this LNA stage (indeed, this would be true even if the LNA itself had an infinite IP3). This would make the 1ST MIXER, and thus the entire receiver, very intolerant of any strong RF input signals. Consequently, we see that the mixer is basically determining the entire receiver's IP3. As the 1ST MIXER is such a critical element in receiver IP3 performance, high-level DBM diode mixers are usually used for this purpose in better devices, since DBM's have a far higher IP3 than most active mixers circuits. Therefore, the DBM permits more gain in the LNA stage, which lowers the effect of the noise contributions of all following stages, including the high-noise figure 1ST MIXER stage. Obviously then, to maximize receiver IP3 and minimize NF so that we may obtain a highly linear and sensitive receiver, the LNA gain and 1ST MIXER IP3 must be selected with care. Most quality communications receivers will also have an IMAGE FILTER placed just before the 1ST MIXER, which will further assist the DIPLEXER's filtering action, and will almost entirely eliminate the LNA's own self-generated image noise from severely affecting the SNR. The WIDEBAND AMPLIFIER that follows the 1ST MIXER has a high reverse isolation, which prevents the large mixer-generated sum frequency from reflecting off the stopbands of IF BPF and directly reentering the mixer's IF port, which would increase the mixer's IMD. The IF BPF rejects all signals that are not on channel, which is an important task, as the gain of the IF strip can be quite high (> 90dB). In some radios, rejection of the second image band is necessary to attenuate damaging spurs, and this is done by the IF BPF also. The 50-Ω attenuators shown into and out of the 2ND MIXER will help to increase the return loss of this conversion stage, thereby decreasing IMDs. This is especially vital at the IF port, where reflections off the filter's stopbands can be severe. In this specific nonlinear FM receiver case, a LIMITER sets the IF_{OUT}'s amplitude into a discriminator or modem, with

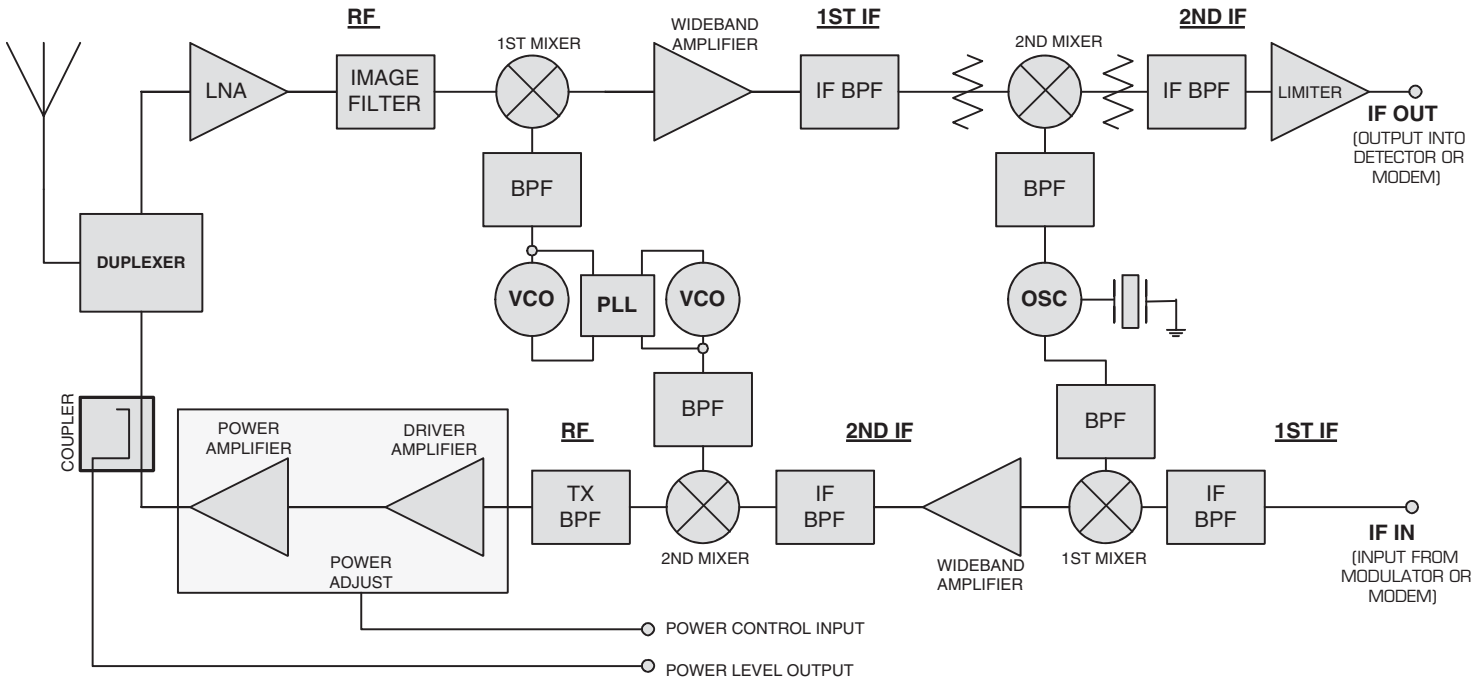


Figure 9.8 A full-duplex FDD radio.

the then demodulated baseband flowing into signal processing circuits, and on to some type of speaker or other device.

So that we may economically filter out the close-in sum (or difference) frequency and LO feedthrough from exiting the antenna in strength, two frequency conversion stages were necessary in this particular high-frequency superheterodyne transmitter. On this transmitter side of the transceiver, the signal from a modem or modulator is placed into the IF IN port, and thence into IF BPF. The modulated IF is converted up in frequency by the 1ST MIXER, amplified and filtered by the second IF stages (a high isolation WIDEBAND AMPLIFIER located at the 1ST MIXER's output port is used here to dampen reflections back into the mixer), and mixed up by the 2ND MIXER to RF frequencies. The TX BPF suppresses wideband transmitter noise, as well as harmonics and assorted mixer products. The DRIVER amplifies the signal to an input level that is acceptable to properly drive the POWER AMPLIFIER (PA). In this FM case, the PA will run saturated, which demands that the DRIVER has enough gain and power to keep it so. The power amplifier stage can be the most difficult part of a transmitter to design. The PA must be very tolerant of low and variable output return losses, which are caused by the large impedance variations created by the mobile antenna and its rapidly changing physical environment, as well as reflections off of the RX/TX duplexer's reflective stopbands. Therefore, not only must the PA not be destroyed by these VSWR variations, but its performance must not degrade excessively, and must remain completely stable at all times. A nonlinear PA for FM use demands that it also be highly efficient, generate minimal harmonic levels, and have enough gain for efficient saturated operation. Some of the RF signal out of the PA is tapped by the COUPLER for RF output power confirmation to a microprocessor, with the majority of the signal being sent to and filtered by the DUPLEXER, and broadcast out of the transceiver's antenna.

9.3.4 RFIC Transceiver

Most wireless system designs now rely heavily on the ever-increasing use of RFICs. These integrated circuits may contain the LNA/mixer stages, the entire IF stages, or even a complete transceiver. In fact, RFICs have already replaced many discrete stages, as high levels of integration permits a single chip to replace dozens, if not hundreds, of components. Currently, however, a complete transceiver on a single RFIC chip is usually available only for low data rate, low-power applications. Nonetheless, higher integration levels are slowly becoming a reality as more and more companies attempt to make a complete wireless *system on a chip*. Depending on the frequencies and specifications required, using RFICs in a wireless device can significantly lower design and production costs, as well as the physical size, of the completed system.

RFICs are available for each stage of a radio, with the level of the desired integration depending on the design flexibility required. The design of Fig. 9.9 demonstrates relatively low levels of RFIC integration, but still permits a far more rapid time-to-market as opposed to many discrete designs. Most of the RF chips for this, and similar radio designs, are readily available from companies such as Maxim, RFMD, Mini-Circuits, National, and so on.

The transceiver design of Fig. 9.10 exhibits a much higher level of integration. Using a single Sirenza/RFMD ML5800 integrated circuit that is operating as a full TDD frequency shift keying (FSK) radio with a built-in demodulator, running in the industrial, scientific and medical (ISM) band of 5.8 GHz at a data rate of 1.5 Mbps. It is, literally, a highly sophisticated RF-in/bits-out RFIC transceiver. The discrete circuits stages that

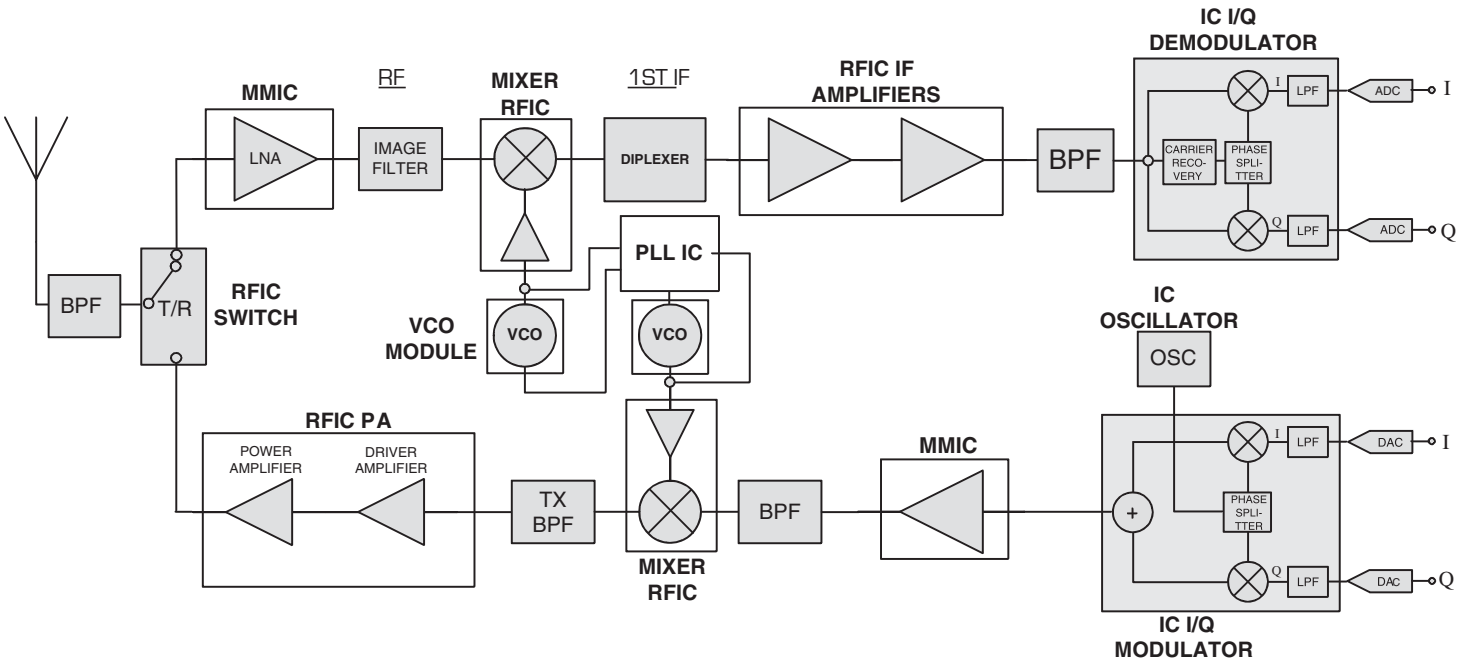


Figure 9.9 Using RFICs and MMICs for a complete wireless transceiver.

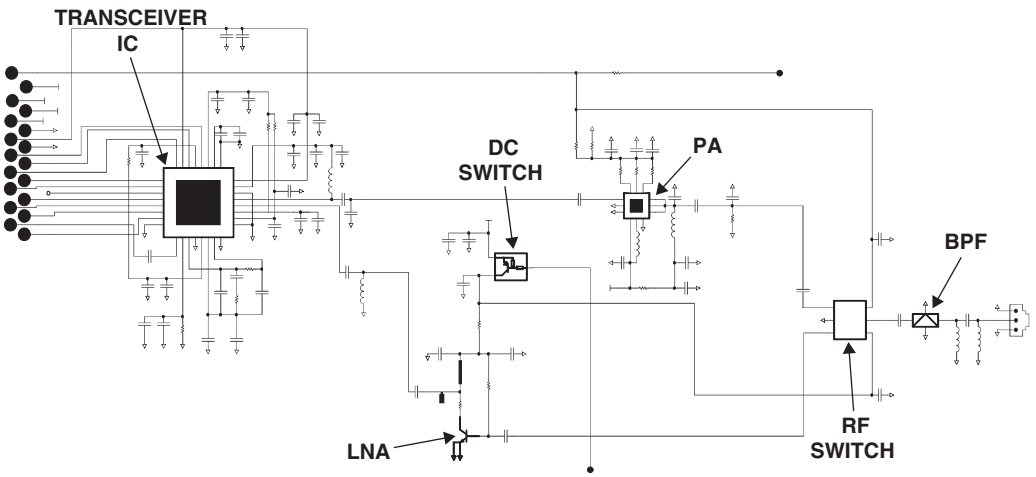


Figure 9.10 A complete data radio with an extremely high level of integration, requiring only an external LNA, PA, RF and DC switching, and RF filtering.

surround this chip are used to increase receive sensitivity and transmit output power, and comprise an LNA, RX/TX switch, and a PA. This is a very good example of the type of common RF design task seen in today's communication radios, which can regularly consist of but a single integrated transceiver chip, along with surrounding discrete (or MMIC) support circuits that are used to improve the RF performance of the central RFIC over the inherent limitations within all integrated circuits, which is low *Q* and low RF output power capability.

Many wireless designs must still rely on discrete, or at least individual MMIC, amplifiers and mixers. This is because the available RFIC's may not be compatible with a particular design goal, while other radios may have to employ an almost entirely discrete design for price/performance issues. Indeed, the RFIC solution is only compatible with any particular wireless design if the chip itself already exists to perform the required functions within our wireless specifications, or if volume production makes it economically viable to actually design the RFIC and have it manufactured. This is obviously an extremely expensive, and a very risky, option.

9.3.5 System Design Issues

The *implementation margin* is an important aspect of transmitter/receiver design. It is the decrease in SNR, and the corresponding increase in the BER, that naturally occurs within a system from the design stage to the final building of the radio itself. The implementation margin losses must be accounted for early in the design stage by increasing the required SNR for the receiver section to properly compensate for this effect. In fact, when adding up all the modem and radio impairments within a real-life and imperfect wireless link, and excluding the *fade margin*, an implementation margin of up to 6 dB is quite common in high data rate radios. These hardware-related impairments can be caused by excessive stage-by-stage amplitude errors, group delay variations, jitter, excess bandwidth noise, phase noise, thermal noise, nonlinearities, frequency instabilities, adjacent channel interference, noisy carrier recovery, and so on.

Another issue in digital wireless design is *group delay ripple* (GDR), which fosters intersymbol interference. GDR must be no more than 75 ns for QAM-64, and under 200 ns for QPSK. Simple analog design considerations, mainly in the system's bandpass filters, and digital adaptive equalizing, will keep GDR in specifications. Therefore, when exploiting phase/amplitude modulations, it is essential to keep the IF filters from introducing too much group delay variations (and amplitude ripple) within the radio filter's passbands, since this will degrade the entire system's bit error rate, sometimes severely. This is especially important when using tight bandpass filters, combined with strong out-of-band attenuation performance.

Any multipath-created *amplitude notches* or *amplitude slopes* in the radio's passband can cause increased intersymbol interference (ISI). To decrease the resultant BER degradation, amplitude ripple, and tilt across a wideband digital channel must be minimized to a level not to exceed 0.5 dB for QAM and QPSK modulations. This multipath brings about phase cancellation, which causes both amplitude and phase distortions of the desired signal. This will decrease the received signal's strength, thus decreasing SNR and increasing BER. We can mitigate, but cannot eliminate, many multipath problems by boosting the gain (and thus the directionality) of the system's antennas, by selecting the proper location for the transceivers themselves (such as not placing them in front of a tall building, mountain, or reflective wall), and employing equalizer stages.

To minimize the *near/far effect* in receivers, some compromises in transmitter and receiver design may need to occur, since a receiver that is physically close to its remote transmitter may become saturated by its constant RF power output in a multiradio environment. This can be significantly reduced by using front-end receiver attenuators, the natural shadowing effect created by a high-mounted transmit antenna, lower gain antennas for the close-in receivers, switchable in/out bypassable LNA front ends.

To improve an already constructed wireless system's bit error rate, we can increase the received SNR by any of four common methods: by increasing transmit power, by boosting the gain of the transmitter and/or receiver antenna, by decreasing the noise figure of the receiver, or by improving the modulation's SNR by processing the incoming baseband signals. Typically, the only alternative that does not involve either significantly increased cost, additional battery power, running afoul of government regulations, longer development times, or some other major impediment, is by simply decreasing the receiver's noise figure. This is accomplished by lowering the noise and increasing the gain of the receiver's LNA, as well as by decreasing any losses that have been placed before the LNA stage.

As some receiver LNAs are AGC controlled, and have their gain modified by the decrease/increase of the amplifier transistor's collector current, it should be confirmed that the LNA's strong signal-handling capability is not excessively compromised when the AGC is enabled. Further, in order to prevent the RF band-limiting filters that are located before or after the LNA from being adversely affected by amplitude ripple or a bandpass frequency shift, it should be checked that the LNA's input/output return losses are not decreasing excessively with this AGC bias change. Obviously, when the AGC lowers the gain of the LNA during these strong signal conditions, and even if this AGC action didn't actually increase the LNA's own noise figure (which it normally will), the NF of the entire system would still be increased due to the reduced front-end gain. This should not be an important consideration at these high signal levels.

Detector-based sampling used for a receiver's front-end AGC control cannot possibly detect strong out-of-band interferers, which may form severe intermods and

even full LNA and receiver blocking. This is because the downstream detector will never actually see these out-of-band signals, since they will be filtered out by the receiver's tight IF before they even reach the detector.

For maximum ADC (and thus receiver) performance, the signal levels into its input must be consistently below the ADC's maximum nonlinear amplitude rating, but within its full-scale optimal range, making a tightly controlled AGC action especially important for digital ADC-equipped radios.

In a time division duplex (TDD) transceiver design, the entire receiver section should always be switched Off during the transmit period. This not only saves DC current, but also prevents the power amplifier's RF feedthrough, which travels across the limited RF isolation of the transmit/receive (T/R) switch and into the receiver chain, from causing multiple issues. The PA feedthrough would travel around the transceiver's On receiver section, into the main radio chip, and out back around into the transmitter section, causing possible feedback oscillations of the PA. An Off receiver solves this concern because it will have attenuation, and not gain, protecting not only the PA from any appreciable feedback, but also the radio chip itself from overload. Conversely, during the TDD's receive period, turn Off all transmitter stages. This not only saves significant amounts of DC current, but prevents the transmitter's PA from leaking noise, spurs, or LO feedthrough past the T/R switch's limited RF isolation and straight into the extremely vulnerable LNA.

Phase noise, created by the imperfect nature of a real local oscillator, will degrade the BER and increase the ISI of a digital radio. A typical digital wideband receiver should possess better than –85 dBc/Hz@10 kHz phase-noise spec to lessen the BER degradation of the incoming RF signal. Such communications systems will also require that the receiver have a dynamic range, with its AGC, of about 100 dB, with a minimum receiver design requiring at least 80 dB. And if wideband cable modems are employed as the system's modulators and demodulators, then frequency stability can also be critical. In such cases, stability with some modem types must be better than ±12 kHz to maintain the BER over time (or even simply to capture the modem's initial start-up preamble). To further maintain digital radio performance, we can improve adjacent channel interference specifications by increasing the IP3 of the transmitter's power amplifier and the receiver's LNA, and/or by backing off the PA driver amplifier's RF power into the PA's input (when using modulations with a high peak-to-average ratio). Increased system linearity, through hardware design and/or the appropriate PA back off, will lessen BER degradation of a digital signal by decreasing intermodulation distortion levels.

9.4 RF Propagation

9.4.1 Introduction

An electromagnetic wave propagates through a vacuum at approximately 300,000,000 m/s. This speed decreases as it passes through any type of dielectric material, even air. The *E-field* (electric) and the *H-field* (magnetic) of the electromagnetic wave are not only at 90° angles to each other, or *orthogonal*, but they also increase and decrease in amplitude together over time, with one field regenerating the other as they travel through space. This is referred to as *transverse electric mode* (TEM) propagation.

Radio wave propagation is highly conditional on the frequency of the RF carrier, and occurs through three main modes: The *ground wave*, which travels on top of and

through the surface of the earth at frequencies below 1 MHz; the *surface wave* (also called *space* or *direct* waves), which travel through the atmosphere in an almost straight line from transmitter to receiver, and is the primary form of propagation for RF signal's over 30 MHz; and the *sky wave*, which is an RF signal that refracts and reflects off of the Earth's ionosphere, and is the method that HF low and high powered simplex long-range RF communications occur under 30 MHz.

Even when employing direct line-of-site surface wave communication paths between the transmitter and the receiver, natural signal losses will begin to decrease the transmitted RF power to a level that reaches closer and closer to the noise floor of the remote receiver's input. This will obviously decrease received SNR, which increases the BER of a digital system or the noise level of an analog system.

9.4.2 Multipath

Multipath fading effects, especially problematic at microwave frequencies, occur when a transmitted RF signal bounces off a conductive object, such as a building's pylons, streetlight poles, or even the earth itself, and reaches the receiver at a slightly different time than the directly transmitted RF signal. This can produce an out-of-phase reception condition, or phase cancellation, causing fading of the received signal, with its severity dependent on antenna height, frequency, gain, and side lobe suppression.

Multipath fading is a huge problem in HF communications as well, since it also creates an intermittent or continuous decrease in the received signal's amplitude. However, it is produced by another highly reflective surface that is not as close to the surface of the earth as buildings, light poles, and mountains: the changing conditions within our planet's ionosphere. This HF multipath is caused by multiple-path reception, which can occur when a HF receiver is accepting both ground-wave and sky-wave propagation, or sky wave and surface wave, or by one-hop and two-hop paths off of the ever-changing and continually undulating ionosphere. These two HF signals are, as above, alternately in phase and out of phase with each other, causing severe fading conditions.

Multipath can create high intersymbol interference and BER in digital communications systems, and can be magnified by misalignment of the transmit or receive antenna, a receive location that is near an RF reflective site, or the nearby movement of vehicles or humans. In digital communications systems, multipath can be examined as a frequency-selective amplitude notch and/or a tilting and misshaping of the received wideband waveform when viewed on a spectrum analyzer. Without an *equalizer* built into the receiver's baseband circuits to fill in for these multipath effects, small tilts and notches of even 1 dB (or less) could quickly render many digital signals unreadable.

An important term when discussing multipath is *delay spread*. Multipath signals take longer to reach the receiver's antenna, with the time dependent on how far they have traveled as compared to the direct wave; the time period between when the direct signal arrives and when the last significant multipath signal arrives is the delay spread. This effect can cause increased ISI due to the received data actually overlapping. Equalizers are, again, typically employed in an attempt to mitigate this problem, but if the delay spread is excessive, then most equalizers will not be capable of properly handling the multipath effects.

In an outdoor microwave link, the basic free-space pathloss and multipath effects will not be the only losses encountered; other impairments must be added to this scenario to establish a worst case within the communication's link. Therefore, outdoor propagation environments will create an effective radio range that can be far less than

that predicted by simple free-space pathloss calculations. These impairments are very similar to indoor propagation issues, but are not as severe. They involve signal blocking caused by shadowing of the receiver or transmitter by large buildings; attenuation of the RF signal caused by passing through structures; and the general multipath effects, as just discussed, of time domain delay-spread (causing bit and symbol smearing) and frequency-domain phase cancellation (causing frequency selective signal attenuation). The more crowded the outdoor environment is with conductive surfaces, the worse these effects become. Thus, a large city filled with many tall office buildings will be the most problematic, while open country areas will be far less so. However, even an open grassy field contains a huge conductive and reflective surface—the earth itself—as well as various small RF-scattering structures, such as plants, grass, and leaves. As a result, it too will have a propagation loss that is much higher than would be predicted using simple free-space calculations.

Rain begins to dramatically increase microwave path losses at frequencies above 10 GHz. A heavy rain shower may attenuate a 10-GHz signal by 2 dB/km or more. As these frequencies increase, so do the losses. It is, however, only the rain that is located directly between the two TX/RX antennas that will add attenuation to the signal, and much of this path may (or may not) in fact be clear. Wet snow has a similar attenuating effect, while dry snow will have little significance even at the higher microwave frequencies. Dense fog can attenuate a 10-GHz signal by up to 1 dB/km, with much smaller attenuation levels as the frequencies decrease. Water vapor and atmospheric oxygen absorption create attenuation at 18 GHz and above, with various high attenuation peaks at several microwave frequencies. *Fresnel zone* clearance inadequacies, atmospheric reflections, and general scattering will also conspire to increase path losses. But not all impairments are continuous attenuation mechanisms, especially the atmospheric effects causing reflection. Nonetheless, such losses can still be compensated for by increasing the power from the transmitter, decreasing the receiver's NF and increasing its gain, and increasing the transmitter and receiver antenna gain. The end result we desire is to guarantee that there will be enough signal amplitude out of the IF of the receiver to sufficiently drive the modem or demodulator, and that the receiver's output signal has the SNR required for reliable demodulation at a low BER.

9.4.3 RF Link Budgets

When designing a communications system we need to know what the maximum RF transmit power can be, the available DC power, any official wireless standards used, cost requirements, complexity issues, compatibility problems, and so on. Next, we need to find out how much bandwidth our system may possess, which will be controlled by similar issues as above. Lastly, we need to be aware of what is the expected reliability of the system as characterized by the BER. These factors will set the limit on the range and capacity of the link, and will allow us to calculate the system's link budget. Consequently, before any piece of RF hardware is ever designed, a system's link budget analysis must be performed. This will tell us how much NF and gain the receiver requires and how much power the transmitter must output in order to reach our desired transmission path range, at a specified receiver BER and SNR, for the entire system.

Fade Margin

Due to the vagaries of wireless communication propagation caused by multipath fading, a certain amount of fade margin will be necessary to make sure the RF link

remains operational for a certain percentage of the year, even under infrequent but severe absorptive weather conditions. This margin is a safety allowance of excess receiver NF, receiver gain, transmitter power, and/or antenna gain that is inserted into our RF link budget to be assured of a dependable wireless connection, and with only a certain maximum amount of permitted downtime per year (normally measured in seconds). For instance, a 20-dB fade margin for digital communications systems can be added to our link budget, a figure that will not only cover most atmospheric anomalies and multipath, but also equipment aging. With this 20 dB of fade margin we might see a certain link that is dependably up and running, throughout an entire year-long period, for 99.99% of the time with a BER of 10^{-8} or better. Even a fade margin as high as 30 dB is a relatively common figure for dependable operability for certain critical digital systems.

E_b/N_0

The answer to the question that will most drive the calculations when performing a link budget analysis is "what is the actual E_b/N_0 (or SNR or C/N) required by the particular modulation, at our required BER, used within our system?". This information is supplied in the form of a BER versus E_b/N_0 graph or table. BER versus SNR and BER versus C/N (carrier-to-noise) graphs and tables are also readily available (See Chap. 2, "Modulation").

E_b/N_0 is an indicator of the necessary system energy per bit as compared to the noise power, and is completely independent of data rate. However, it is often more convenient for the analog engineer to convert E_b/N_0 to SNR. We can accomplish this by

$$\text{SNR (dB)} = (E_b/N_0) \times (R_S/\text{BW})$$

where E_b/N_0 = Energy needed per bit of information over the thermal noise within a bandwidth of 1 Hz, dB
R_S = data rate of the system, bps
BW = bandwidth of the system, Hz

or

$$\text{SNR (dB)} = 10 \log (P_S/P_N)$$

where P_S = signal power at a particular point in the spectrum, W
P_N = noise power at a particular point in the spectrum, W

Outdoor Link Budget Design

To begin our outdoor link budget analysis we must take the range, in kilometers or miles, that the communications link must reliably transmit information, and calculate the free-space path loss. This is the loss, in dB, that occurs to an RF signal at a specific frequency over a specific range, but does not account for any impairments. Another figure we must obtain is the fade margin, as mentioned above. This will be required to assure the link of reception reliability caused by any unexpected and unpredictable, but persistent or periodic, atmospheric anomalies and multipath effects. We would also like to find the minimum power that must be available at the receiver's output port for proper demodulation by the modem or demodulator, and what SNR is required for the particular modulation and error correction in use.

In the following case, it is being assumed that the RF designer will be told what the maximum transmitter equivalent isotropic radiated power (EIRP) will be. But in order to obtain the necessary BER expected at the receiver's output port, a compromise will normally have to be made between transmitted RF power, which must generally be minimized in certain bands or applications to the lowest level possible, and the receiver's NF and gain versus IMD. However, higher gain antennas and lower loss coax cable on either the transmit and receive side, or both, can sometimes be an easy way to increase the required margins.

Performing an Outdoor Link Budget Analysis (Fig. 9.11)

A. First, calculate the free-space path loss over the desired link distance. As mentioned, free-space path loss does *not* include any losses caused by the atmosphere or by multipath, but merely accounts for the inverse square law signal-spreading of the RF wavefront as it leaves the transmitting antenna. (For instance, if the distance from the transmitter is doubled, the receiver would only receive a quarter of the RF energy it would have received at half this distance.) So no matter how directional, and thus high-gained, an antenna may be, the RF power will still drop off commensurate to the inverse square law, decreasing in field strength over distance. This means that if a transmitter is sending at equal powers into two different antennas—one low gain and the other high gain—the high gain directional antenna is merely *starting off* with a higher field strength on its surface than would have been possible with the use of the low gain omnidirectional antenna. By following the free-space path formula below, and working it out for various distances, we can obviously see that for every time the distance between the receiver and transmitter is doubled, a further 6 dB of path loss occurs:

$$L_p = 32.4 + 20 \log (f) + 20 \log (d)$$

where L_p = free-space path loss, dB
f = frequency, MHz
d = distance, km

B. Decide on the fade margin value. The higher the frequency, the longer the link path, and the greater the desired reliability, then the more fade margin is required. Fade margins of 10 to 20 dB in 20-Km digital microwave point-to-point links are quite common, while higher frequencies and longer links may require up to a 30-dB margin.

C. Since the magnitude of the receiver noise floor will set the minimum signal level that will still be easily detectable above the noise (a signal that will have a *positive* SNR), we can calculate the signal strength that is required at the receiver's antenna inputs at its rated NF to obtain a desired SNR at the receiver's IF output port:

$$S_{\text{dBm}} = -174 + \text{SNR} + \text{NF} + [10 \log_{10} (\text{BW}_N)]$$

where S_{dBm} = equals the signal strength needed at the receiver's antenna inputs, to obtain an output IF signal at a desired SNR in a 50-Ω system, dBm
SNR = required signal-to-noise ratio, for the type of modulation used, at the receiver's output port (i.e., at the detector or modem input), dB

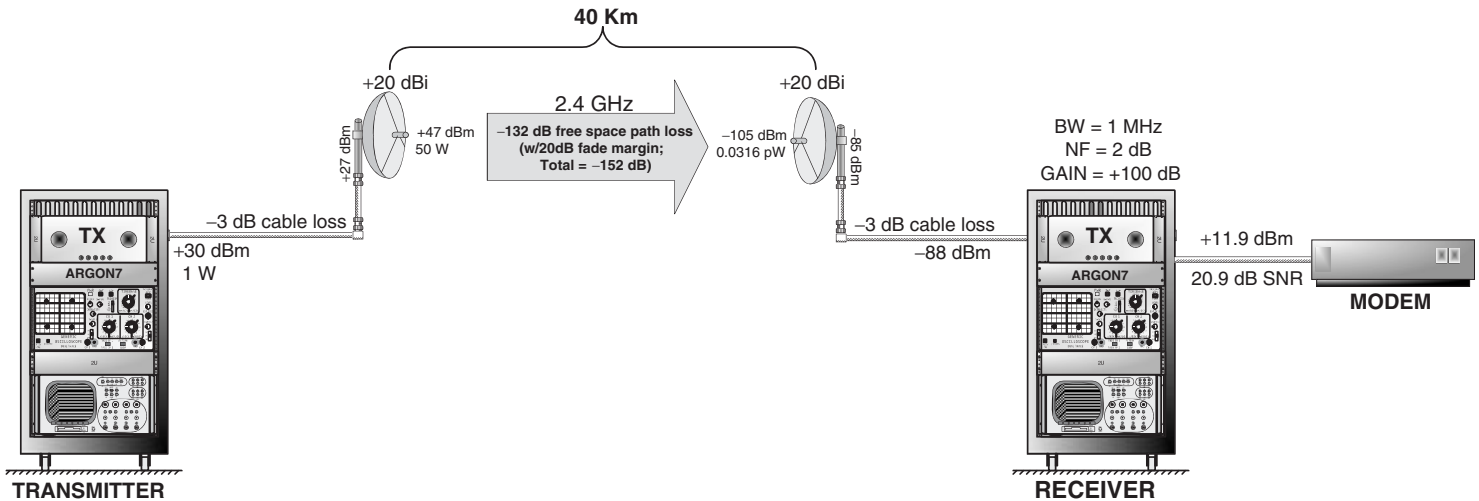


FIGURE 9.11 A wireless point-to-point communications link example for a link budget analysis.

NF = noise figure of the receiver, dB
BW_N = The 6-dB bandwidth (or, more accurately, the *noise bandwidth*) of the IF, Hz

D. Calculate the power that will be placed at the receiver's input port (i.e., after its own antenna) after its trip across the link from the transmitter:

$$P_r = P_t + G_t + G_r - \mathrm{PL}_{\mathrm{dB}}$$

where P_r = power present at the receiver's input, dBm
P_t = power delivered by the transmitter into its own antenna, dBm
G_t = gain of the transmitter's antenna, dB
G_r = gain of the receiver's antenna, dB
$\mathrm{PL}_{\mathrm{dB}}$ = free space path loss between the transmitter and receiver antennas, dB

E. Now, confirm the transmitter is outputting enough power to overcome the free-space path losses and to account for the desired fade margin, and/or that the receiver will have low enough NF for a desired SNR, and enough gain for a proper output signal strength into the modem or detector.

The following formula will tell the designer the power at the receiver's output port, into the modem or detector:

$$P_{\mathrm{OUT}} = P_t - L_p + G_t + G_r - L_t - L_r + \mathrm{RX}_{\mathrm{dB}}$$

where P_{OUT} = power at the receiver's output into the modem or detector, dBm
P_t = transmitter's power output into its own antenna coax, dBm
L_p = free-space path loss as calculated above, dB
G_t = transmitter's isotropic antenna gain, dBi
G_r = receiver's isotropic antenna gain, dBi
L_t = loss of the transmitter's coax, dB
L_r = loss of the receiver's coax, dB
$\mathrm{RX}_{\mathrm{dB}}$ = gain of the receiver section, including conversion and filter losses, and the front-end and IF amplifier gains, dB

The following formula will tell the designer the signal strength output of the receiver stage when the signal strength at the receiver's antenna is known:

$$P_{\mathrm{OUT}} = P_{\mathrm{SIG}} + G_{\mathrm{dB}} - L_r + \mathrm{RX}_{\mathrm{dB}}$$

where P_{OUT} = signal strength available from the receiver's output into the modem or detector input, dBm
P_{SIG} = power picked up by the receiver's antenna from the transmitter's antenna, dBm
G_{dBi} = gain of the receiver's antenna, dBi
L_r = loss of the coax cable between the receiver's antenna and its front end, dB (as a positive number only)
$\mathrm{RX}_{\mathrm{dB}}$ = gain of the entire receiver stage—including all filters, conversions, and amplifiers—from the front end to the output port into the modem or detector stage, dB

This next formula will give the designer the signal-to-noise ratio at the output of the receiver stage:

$$\text{SNR} = P_{\text{OUT}} - (-174 + 10 \log (\text{BW}_N) + \text{NF} + L_r + \text{RX}_{\text{dB}})$$

where SNR = signal-to-noise ratio present at the output of the receiver, dB
P_{OUT} = signal strength available from the receiver's output into the modem or detector input, dBm
RX_{dB} = gain of the entire receiver stage—including filters, conversions, and amplifiers—from the front end to the output into the modem or detector stage, dB
L_r = The loss of the coax cable between the receiver's antenna and its front end, in *dB*. Can be presented as a *positive* number only.
NF = noise figure of the receiver, dB
BW_N = 6-dB bandwidth of the IF (or, more accurately, the noise bandwidth), Hz

The above calculations should give us an excellent indication of the signal power, noise figure, and gain required across any particular outdoor link.

Indoor Link Budget Analysis

If we could actually see the electromagnetic energy at a single transmitted frequency within a typical dense office environment with our own eyes, it would visually illustrate that the air around us is alive with thousands of EM waterlike ripples and wavelets of RF signal energy, varying in strength by 35 dB or more (Fig. 9.12). Reflection, diffraction, and scattering cause these chaotic variations in signal power, called multipath effects.

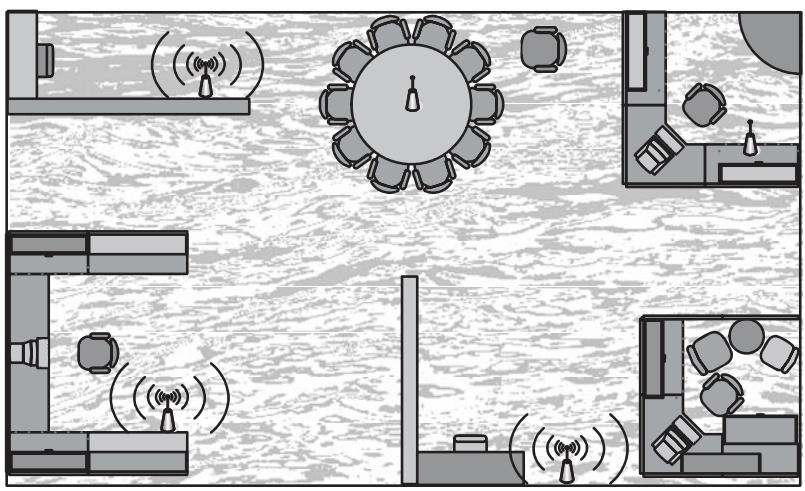


Figure 9.12 Depiction of the microwave wireless signal strength in a typical office space, looking down, varying in intensity due to refraction, interference, reflection, and other effects (dark wavelets are weak RF signal strength areas, light wavelets are strong RF signal strength areas). If all the transmitters, frequencies, and objects within the office remained static, then so would the field strength over time, remaining exactly as shown (assuming no outside EM influences).

Indeed, if a transmitter were to send a 2.4-GHz modulated bit stream to a receiver located only 30 ft away, our worst-case indoor path loss may be as high as 70 to 80 dB, while at 300 ft we may lose over 105 dB! On the other hand, in a static office environment, shifting our transmitter or receiver hardware by but a mere few inches one way or the other may actually increase the received signal strength by up to 40 dB. So, we can easily understand the problematic and challenging nature of indoor wireless propagation at microwave frequencies.

For RF signals that are of a small wavelength versus the indoor environment's own obstructions, which we can consider as being at UHF and above, we may expect to continually experience these severe multipath effects. When such high-frequency RF signals impact a "large for its wavelength" office surface, be it a cubicle's partition, structural wall, file cabinet, ceiling, floor, and so on, then part of that wave will be reflected and part of it will be absorbed. How much of either is mainly dependent on the signal's angle of impact and its frequency, as well as the structure's own conductive/dielectric properties, shape, and roughness. In fact, 100% of the RF energy would be reflected if the structure were constructed with an ideal conducting surface. Other, less ideal office obstructions will reflect less of the impinging signal, while passing through and absorbing the rest of the incoming RF energy.

Diffraction of a wave occurs when an RF signal strikes an abrupt conductive edge within the building—and there are hundreds of these types of edges in the typical dense office environment—causing the RF energy to bend downward and actually filling in physical areas that would normally be blocked from receiving that individual wave.

Scattering, an effect created by RF energy striking any large and multifaceted conductive surface, or multiple small surfaces that are tightly packed together, causes the impinging RF energy to disperse into many assorted directions; very much as if a multiplicity of tiny antennas were rebroadcasting the single incoming wavefront in multiple directions.

But the limiting range factor in any indoor microwave environment is multipath and absorption, not simple free-space path loss, due to all the objects in the way of the propagating signal. For instance, a 2.4-GHz radio that may have the power and sensitivity to reach out 2 mi in an *outdoor* multipath environment, may actually only perform reliably out to 200 ft indoors. This is because of the normal, severe multipath and absorption effects that are present within an ordinary business environment.

To reach our desired communications range at a specified BER, the indoor link budget calculations will need to supply us with information as to how much NF and gain the receiver requires, and how much RF power the transmitter must output. Performing this basic link budget analysis is then completed by calculating the final SNR at the output of a wireless receiver, after the desired signal is sent from the transmitter, across the entire transmission path, and into the receiver.

Approximate Indoor Range Calculations

As stated above, the free-space path loss calculations presented at the beginning of this section would give us a good indication of the signal power, noise figure, and gain required across a particular line-of-site outdoor link. But in an indoor environment with its inherent RF absorption by equipment and walls, range becomes severely decreased, and is extremely hard to predict. Nevertheless, we can approximate indoor range (with zero link budget for multipath effects) in a typical, harsh office environment by a simple rule of thumb, as recommended by Infineon (Intersil):

1. Calculate the free-space path loss for the first 20 ft (at 2440 MHz this will equal 56 dB).
2. After the initial 20 ft that was calculated as free-space loss above, add up indoor path losses at the rate of 30 dB for each and every 100 ft of indoor range.
3. Add the two path losses as obtained in steps 1 and 2.

Therefore, by using this rule of thumb, if we are transmitting at an equivalent RF output power of +22 dBm (the sum of a + 20 dBm transmitter along with a +2 dBi antenna), and we have a receiver possessing a cascaded sensitivity of −104 dBm and using a receive antenna gain of +2 dB, then we will have a maximum 128-dB signal path loss permitted in order to obtain a desired SNR of 3 dB. So, our maximum indoor range will be, with zero added fade margin to compensate for multipath fading, about 240 ft. (Outdoor ranges at these power and sensitivity levels, in a perfectly open environment, would be measured in miles).

However, due to the vagaries of indoor (as well as outdoor) wireless communications caused by both static and dynamic multipath fading, a certain amount of fade margin will be necessary to make sure that the data link will remain open for a given percentage of time. This fade margin is a safety allowance of excess receiver NF, transmitter power, and/or antenna gain that is placed within our link budget to assure a dependable wireless connection, with a certain amount of permitted downtime, even under most severe multipath conditions. For instance, adding a 25-dB fade margin to a digital communications system's link budget will not only cover the majority of in-office multipath fades, but also most interference and equipment aging problems. Nevertheless, some indoor fades can cause a signal cancellation of up to 40 dB, and this may demand that we look into using antenna diversity, adaptive equalization, or signal spreading to safely decrease our fade margin requirements and increase reliability. The system's antenna pattern is also critical for the easing of multipath, and since an omnidirectional antenna can receive equally well in all directions, it is the most susceptible of all antennas to such effects. Omnidirectional reception capability can be, of course, imperative for proper system operation, so there may be little we can do about this particular concern.

We can also perform some indoor range calculations with the following formula. This formula can be relatively accurate for 2.4 GHz, which is a very common frequency in today's business environment, but the path loss results are still an approximation:

$$\text{Path loss (dB)} = 40 + [35 \log (D)]$$

where D = Indoor distance between the transmitter and the receiver, m.

Will the Communication's Link Work?

For the following example of a simple outdoor link budget analysis of the communication's system and path, see Fig. 9.11.

A 40-Km link, with no obstructions, is needed to operate dependably at 2.4 GHz with a transmitter/antenna combination that can output a + 47-dBm (50-W) EIRP signal with our modulation of choice. We find that the receiver must, through all atmospheric conditions, maintain a signal into the detector or modem of + 10 dBm, with a minimum SNR of 20 dB, at a bandwidth of 1 MHz. What would be the specifications of the receiver necessary to meet this requirement?

First, we must calculate the free space path loss between the transmit and receive antennas for a frequency of 2.4 GHz, then add 20 dB for the link margin to be assured of reliable operation. We can now subtract the calculated path loss, with link margin, from the EIRP of the transmitter system to arrive at the power level that will be present at the receive antenna under a worst-case situation of –20 dB over and above the free-space path loss. We find this figure to be –105 dBm. The receive antenna, with its 20 dB of gain, will take the –105 dBm signal and amplify it by 20 to signal level of –85dBm, while the 3-dB of cable losses will drop the –85-dBm signal level to –88 dBm. Thus, –88 dBm of signal is finally placed at the receiver's front-end input from the original transmitter signal. We can see that a total gain of 100 dB will be required of the receiver to meet or exceed the +10 dBm output power requirement into the modem or detector, or 100 dB plus –88 dBm. We have left ourselves a small gain implementation budget of about +2 dBm to cover design-to-realization (real-world) losses, for a total signal level out of the receiver of +12 dBm. We then employ the SNR formula as shown above to find that we have a good SNR of over 20 dB at the receiver's output, or a very small 0.9 dB for the SNR implementation budget, for a total of 20.9 dB. We have thus found that this outdoor link will be quite dependable over almost all atmospheric conditions, and during most parts of the year.

NOTE: The final value of the communications receiver's output signal above the noise is dependent on the receiver's bandwidth, its NF (which is why receiver NF is so important, since a 1-dB improvement in NF translates to a 1-dB improvement in SNR at the detector), antenna temperature, antenna gain, cable losses, and transmitted EIRP. However, the actual signal amplitude at the receiver's detector input is dependent only on the receiver's antenna gain and cable losses, receiver RF and IF gain, and the transmitted power.

9.4.4 RF Link Issues

Fresnel Zone

Fresnel zone clearance must be considered when setting up a microwave link close to the earth. This is because optical line-of-site between the transmitting and the receiving antenna is only one factor in most microwave links, the other being the first Fresnel zone clearance.

The Fresnel zone is referred to as the *radio line-of-sight*, and will need more clearance that the optical line-of-sight our own eyes can see: a radio wave, when it passes near an object—such as a building or a mountain—will become defracted or bent, even if that obstruction is many feet below, or to the side, of the radio wave.

To avoid attenuation of the transmitted signal at the receiving station, we must confirm mathematically, and not optically, that we have sufficient Fresnel zone clearance above any mountain or building. The following formula can be used in conjunction with Fig. 9.13 to verify this:

$$h = 72.1\sqrt{\frac{d_1 \times d_2}{fr(d_1 + d_2)}}$$

where h = clearance between the top of any obstruction, and the direct line-of-sight, that is required for zero attenuation to the transmitted signal, ft

d_1 = distance between the transmitter and the obstruction, mi

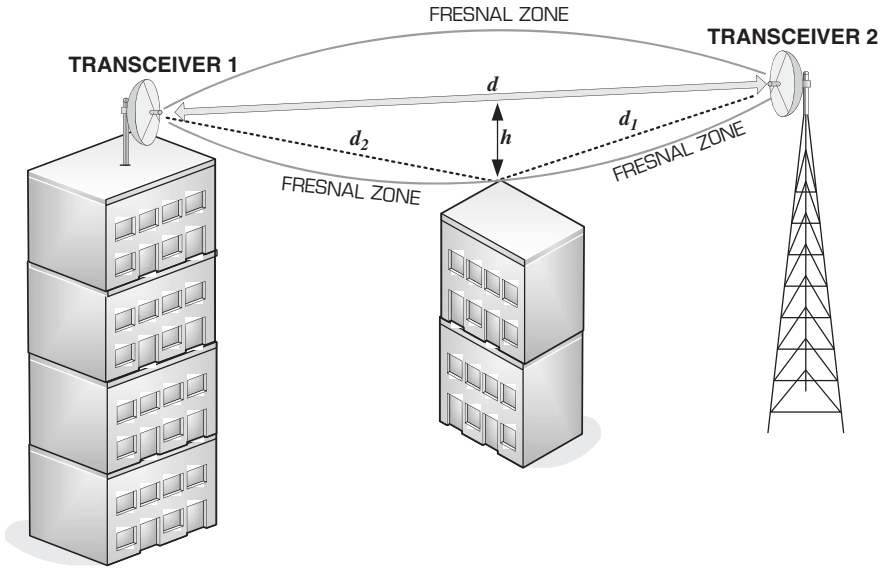


Figure 9.13 Measurement of Fresnel zone clearance in point-to-point wireless link.

d_2 = distance between the obstruction and the receiver, mi
fr = frequency of the transmitted signal, GHz

or

Alternately, if the answer is required in meters:

$$h = 17.3\sqrt{\frac{d_1 \times d_2}{fr(d_1 + d_2)}}$$

where h = clearance between the top of the obstruction and the direct line of sight required for no disruption of the link, m
d_1 = distance between the transmitter and the obstruction, Km
d_2 = distance between the obstruction and the receiver, Km
fr = frequency of the transmitted signal, GHz

We could ignore the Fresnel zone clearance entirely but, depending on the frequency and the geometry of the obstruction, and if the obstruction is close to the optical line-of-sight, the additional path losses can be anywhere between 6 and 20 dB.

Antenna Issues

As we have discovered, the noise floor of a receiver system is a critical issue. To calculate this, including the antenna's own NF as well as the receiver's NF, we can use the formula below. This formula works accurately for terrestrial receiver system NF calculations, but assumes a receiving antenna's noise temperature to be 290 K, which is a good antenna temperature estimate for all antennas in an earthbound link environment. And since the formula calculates the signal strength, in dBm, required at the input of the

receiving antenna for the receiver to output at a 0-dB SNR, we would have to confirm that the transmitter on the other end of the link has the power to permit us, with sufficient fade margin, to obtain this SNR (which is the value as required by the receiver's modem or detector in use at the output of the IF):

$$S = 10\log(\text{BW}) + 10\log\left(290\left[10^{\frac{\text{NF}}{10}} - 1\right]\right) - 198.6 - G_A$$

where S = entire receiver's sensitivity with a 0-dB SNR at its output, dBm
BW = system's IF bandwidth, Hz
NF = receiver's noise figure, dB
G_A = receiver's antenna gain, dB

Another factor that affects the noise of the receiver system is the antenna's orientation. If the receiving antenna is pointed more toward the sky (but *not* toward the sun) in order to receive, as an example, a signal from a remote transmitter that is placed on a mountain top, then its noise temperature, and thus its NF, will be less than if it were pointed at the ground. (The ground has an approximate noise temperature of 290 K, as inferred above). However, even if the receiver antenna were pointed at the coldest regions of space, as it would be if communicating with a satellite transmitter, then its side lobe reception would still increase its noise temperature somewhat, depending on the side lobe's gain, angle, and other factors.

CHAPTER 10
Communication Antennas

Antennas are designed to efficiently transform alternating current into electromagnetic (EM) waves, and to then send these waves out into free space. The electromagnetic waves are caught by the receiving antenna, where they are converted into an alternating current. The actual design of the antenna depends on the frequency of its operation, output power, directivity, robustness, cost, and space limitations. However, any resonant antenna can be considered as a series resonant circuit, and when cut to a quarter wavelength for a vertical monopole or half for a dipole (Fig. 10.1), maximum current will be allowed to flow through its elements. This will give the maximum signal strength possible for that particular antenna design.

Simply, an antenna is a device that accepts high-frequency alternating current directly from a radio transmitter, translating it into a radiating electromagnetic field. An antenna can be thought of as an impedance-matching structure that matches the radio transmitter's 50-Ω output to the characteristic impedance of free space, while also focusing its radio frequency (RF) energy in a preferred direction. This same antenna is also equally capable of taking an incoming electromagnetic field and translating it into high-frequency alternating current for direct insertion into a receiver. Thus, as an electromagnetic wave cuts the antenna's elements it induces a current flow, and a difference in potential will now be created between each end of the antenna. In other words, any passing electromagnetic wave in the air around us will dump some of its RF energy behind into the antenna as it propagates past its conductive elements. However, the actual amplitude of this extracted and received RF energy is relative to the antenna's *aperture*, which is the useful area that the metallic structure of the antenna embodies to any passing electromagnetic wave. This is analogous to a larger fish net being capable of catching more fish than a smaller net.

The electromagnetic fields that are passing by or being created by an antenna will have both intensity and direction, with the electric field measured in volts, and the magnetic field, measured in gauss. Moreover, when one of these fields starts to move, it will repeatedly recreate the other type of field; which is why it is referred to as an *electromagnetic* field. Therefore, a current will flow within a conductor when any magnetic field travels past it; and when current flows, a magnetic field will then be instantly produced.

The antenna structure itself can consist of any geometric assembly that is capable of conducting RF energy. How efficiently the antenna actually converts received or transmitted energy rests strictly on its size, shape, materials, frequency of operation, and nearby structures. And any passive antenna, whether used for receiving or transmitting, will have the same gain, impedance, and directional pattern during both receive and transmit. This is called *reciprocity*. Consequently, any efficient antenna will

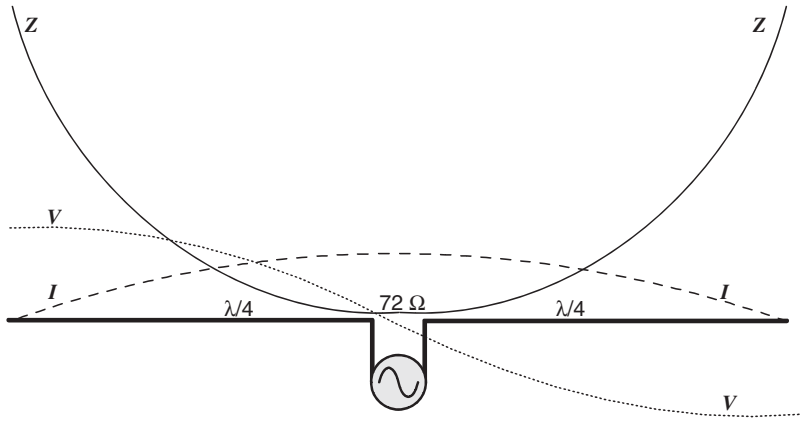


Figure 10.1 Current, voltage, and impedance along a half-wave resonant dipole.

radiate virtually all of the RF energy sent into it, no matter what the frequency, and the electromagnetic radiation will then break away from the antenna's conductive elements and be radiated into space—if the signal's frequency supplied by the transmitter is greater than 10 kHz.

The actual *gain* of an antenna is the gain as compared to that of an isotropic source, measured in dBi, but it can also be measured as dBd when being compared to a resonant dipole. (There is a 2.15-dB difference in gain between an antenna that is rated with dBd versus dBi. This is because a dipole possesses a superior gain of 2.15 dBi against an isotropic's 0 dBi). Therefore, the gain specification is the measure of how well an antenna amplifies an RF signal over a specific reference antenna, such as the dipole or isotropic radiator standards, in a particular, favored direction. Depending on many factors, this gain may be higher or lower as compared to these reference antennas. But even though many microwave antennas can supply higher (positive) gain during transmit and receive than can an isotropic or dipole structure, these antennas still cannot actually amplify a signal in the sense that a transistor can; antennas can only exploit the RF energy supplied at their input port, and then pass this same energy out into space in a certain direction with a certain gain over (or under) that of the reference antenna. The total radiated power from an antenna, no matter how directional and how highly gained, will never be greater than the energy placed into its input by the transmitter. In fact, there will always be a certain loss in total power within the antenna, since radiation efficiency can never be 100% in a practical antenna system due to losses within the antenna structure itself, its matching network, its substrate [if a printed circuit board (PCB) antenna], any dielectric housing (radome), and surrounding structures. Even as the antenna gain is increased, its beam-width must also decrease to produce this extra gain. Therefore, high gain, high directivity, and large electrical size to wavelength are all interdependent. Indeed, it is impossible to have a high-gain omnidirectional antenna, or a high-gain antenna that is electrically small.

The *beam-width* of an antenna is the number of degrees, in the horizontal, of the *main beam* at its 3-dB down points. Intimately linked with the antenna gain above, and is measured in *degrees*. The main beam of an antenna is the dominant lobe of a directional antenna where the majority of the output power radiates, while the antenna's *side lobes* are normally wasted and undesired emission areas which, unfortunately, emit electromagnetic

radiation at other directions than from the main beam's lobe. Due to smaller side lobes collecting less temperature, and thus less noise, from the earth, minimizing such lobes will increase antenna performance. High-powered RF outputs from an antenna may actually contain dangerous levels of EM radiation within these side lobes. The *directivity* of an antenna is the power in the main beam of an antenna when compared to an isotropic, and is a ratio that compares the antenna's radiation strength in an explicit direction in space to that of an isotropic source of the same radiated power. It is the directional attribute of the emitted antenna energy, and is analogous to the gain specification, but the efficiency (heat losses) is completely ignored. Directivity is measured as a ratio.

The antenna's *efficiency* is another critical specification of the antenna system, which includes all losses that occur within the antenna's general structure, such as dielectric, copper, and mismatch losses. Perfect radiation efficiency would mean that when 100 W of RF energy is inserted into the antenna's input, then 100 W of RF energy would be radiated. However, real-life antennas are never 100% efficient, so any of the RF energy that is not actually radiated by the antenna has been wasted in the form of heat. (Efficiency of the antenna itself is included within the antenna's basic specification of *gain*.)

This brings us to an important concept that is often overlooked is antenna *radiation* and *loss resistance*: It is convenient to think of an antenna as composed of a *radiation resistance*, which is the value of a theoretical substitute resistor that could take the place of the entire antenna, and that would also dissipate the identical magnitude of RF energy that the antenna itself radiates into space. Conversely, any wasted RF energy that is given up as heat in the entire antenna system can be compared to some other theoretical value, called the *loss resistance*. We can then state that an antenna's radiating efficiency is solely the ratio of this radiation resistance *versus* loss resistance. In order to obtain high antenna efficiency, we must keep our loss resistances as low as possible as compared to the magnitude of the antenna's radiation resistance. This is exactly why it is very difficult, and very expensive, to maintain high efficiencies in very electrically short antennas, since an antenna has increasingly lower radiation resistance with decreasing physical size. With very electrically short antennas, we must be constantly lowering the antenna's loss resistance to keep its efficiency as high as that of a full quarter- or half-wave antenna. (Significant losses that decrease antenna system efficiencies can and do originate at the antenna's feedline and matching network, as well as in any surrounding objects.) If there were no matching losses, skin effects, transmission line losses, or coupling losses into other structures, we could, with 100% efficiency, use an antenna that was only 1000th of a wavelength long, and yet all of the RF power fed into it would be completely and efficiently radiated.

An antenna's *polarization* is the orientation of the *electric* field of the electromagnetic wave as it travels through space, and when an antenna's elements are parallel with the ground it is referred to as being a *horizontally polarized* antenna. Such an antenna can only receive an electromagnetic wave from a *vertically polarized* source through the small shift in polarization that takes place over distance. If it were not for this slight electromagnetic wave change, a perfectly horizontal antenna would not be able to induce a voltage into a perfectly vertical antenna, and vice versa.

During transmit, the actual physical size of the *resonant* antenna does not really matter. So, a resonant antenna that is cut for 5.8 GHz will efficiently radiate all of its input power, just as a much longer resonant 10-MHz antenna of the same type would. However, during receive, the short (but resonant) 5.8-GHz antenna simply will not have the same effective surface area as the lower frequency and much longer resonant

10-MHz antenna, and is therefore completely incapable of extracting the huge amount of RF energy from space that the lower frequency structure is capable of. But this size discrepancy between high-frequency antennas and the low-frequency antennas can also help mitigate much of the disparity in this receive efficiency. Because of the smaller physical dimensions of an antenna cut for microwave frequencies, we can now very easily design and deploy them as much higher-gained structures. For example, a resonant 1-MHz multielement high-gain Yagi antenna is virtually unheard of due to its enormous size, making only a single (normally random length) wire or loaded dipole antenna truly practical. But at the microwave frequency of 900 MHz, such an antenna would be ridiculously small, and extremely practical to deploy.

An antenna's radiation pattern will vary if the antenna's length stays the same, but the frequency injected into it changes (Fig. 10.2). These shifts in the radiation pattern will alter the antenna's directional characteristics. Now, instead of an omnidirectional antenna radiating in all directions with equal signal strength when run on its fundamental frequency, it will have far more power nulls and power lobes when run on the harmonics. Nonetheless, since the antenna's center-fed feedpoint will remain almost at the same low input impedance when run at odd multiples of the fundamental, this makes operation at the third harmonic possible with a single antenna.

Just as with any RF circuit, a good impedance match between the transmitter, the feedline, and the antenna is important for the maximum transfer of power. A proper match will also stop the energy generated at the transmitter's output from being reflected back from the antenna's input, which can produce transmission line insulation breakdown, as well as serious transmitter damage due to the mismatch between the transmitter's powerful power amplifier (PA) and the antenna structure. This perfect match between the transmitter, its feedline and the antenna will occur when the inductive reactances in each circuit equals the capacitive reactances and cancel, leaving only the resistances, which must equal each other, therefore allowing maximum power to be radiated from the antenna. The stage-by-stage match will prevent the RF power from being wasted as heat as it is bounced back and forth between the unmatched antenna and the transmitter, which would quickly decrease the RF signal's amplitude through dissipation within the unavoidable resistive elements of the coaxial cable. Therefore, matching the antenna to the transmitter's own output impedance is always the desired goal, and with the antenna resonant and the impedances all matched between the transmitter, feedline, and antenna, maximum alternating current is inserted into the antenna elements, and the maximum electromagnetic wave is broadcast into space.

As stated above, antenna power can only be radiated as an electromagnetic wave or consumed by a resistance as heat. As a result, if there were absolutely no losses in the entire antenna system, then all of the inserted transmitter power would eventually be radiated by the antenna, even if that antenna caused an extremely high voltage standing wave ratio (VSWR). Nevertheless, even small runs of transmission line placed between the transmitter and its antenna can have very high losses at microwave frequencies, so this real-life fact alone would be our major compelling reason to supply a perfect transmitter-coax-antenna match. Further, a transmitter's power amplifier (or its output filter) is normally designed to see roughly 50 Ω, and it may markedly decrease its efficiency, power, gain, or even stability if the impedance varies excessively from that figure at the antenna's input port. Any of these consequences would be completely unacceptable in any properly designed transmitter system. The majority of higher powered transistorized transmitters have protection circuits that prevent damage to the PA if presented with an impedance that differs significantly from the system impedance,

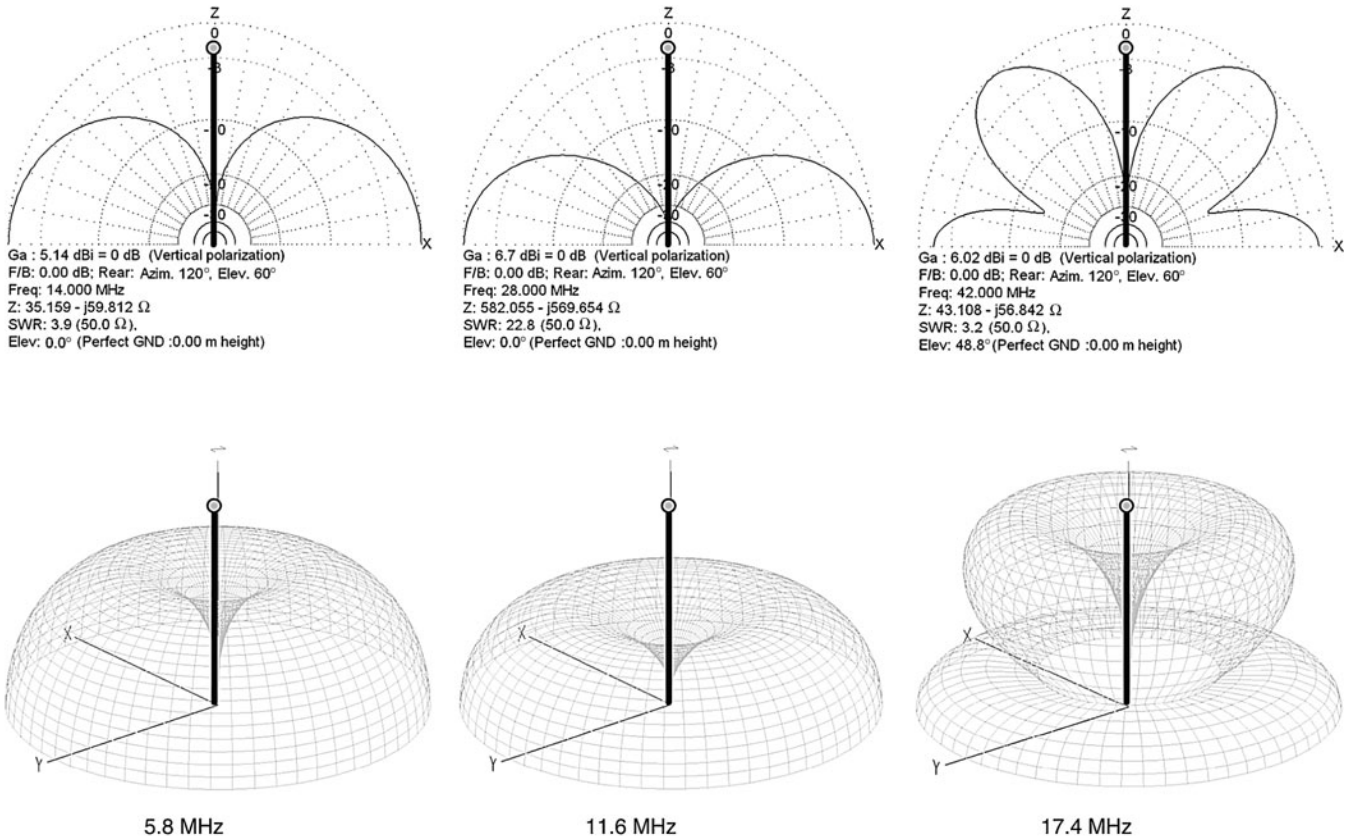


Figure 10.2 Top, showing the change in a 5.8-MHz monopole's radiation pattern (over an ideal groundplane) as the input frequency is increased in harmonic increments (11.6 MHz = second harmonic; 17.4 = third harmonic). Bottom, a 3D representation of the changing gain patterns of a monopole with frequency.

but most of the amplifiers we are concerned with here, which are all under 3 W, will not have this protection capability.

As we are aware, making an antenna physically half-wavelength long is a common method to cancel any reactances at its input port. But there is another method that is commonly used to remove this undesired reactance at our desired operational frequency, and this is by exploiting the ordinary method of impedance matching. Thus, the reactance of an antenna can be cancelled by either utilizing an antenna that is created to be a physical half-wavelength long, or an antenna that is at some completely random length, but has its reactive components tuned out (and its resistances matched) by external components. This also infers that a highly efficient radiator can be an antenna of *any* length, not just a half-wavelength long. And such a nonresonant structure *will* operate with the same efficiency as its resonant half-wavelength case and, in fact, the only valid reason we may want a full classic half-wavelength dipole or quarter-wavelength vertical is to obtain an easily predictable radiation pattern, an undemanding impedance match, and a classic current and voltage distribution. A resonant antenna structure simply does not have any extraordinary characteristics that make it a superior radiator over some random wavelength antenna, except for the legitimate need for a standard radiation pattern. Indeed, the only situation where a resonant antenna is truly superior to a random length type is when we compare very electrically small antennas to the half-wavelength case: As the antenna's size is *dramatically* decreased, its input impedances will begin to decline severely, making real-life, efficient matching difficult and lossy. In reality, any antenna input impedance of less than 5 Ω will start to noticeably degrade antenna radiation efficiency.

Confusion between an antenna's resonance and the antenna's input impedance is common. The distinction is that a half-wave antenna can be resonant at a certain frequency, since it is at the exact length required for resonance at that particular frequency, but still be a relatively poor match to the transmitter. A good example would be an ideal dipole antenna which, at resonance, has a resistive feedpoint impedance of 73 Ω, yet the transmitter may require 50 Ω for a 1:1 VSWR. And with the insertion of a simple L-matching network at the input of the antenna we can easily supply this expected design impedance to be exactly 50 Ω. The same can be said for an ideal vertical antenna with a resonant impedance of 37 Ω resistive: for a 1:1 VSWR, the vertical will also need a matching network for a 50 Ω system. Thus, antenna resonance is decided by the antenna's exact physical length, while the antenna's input impedance, as seen by the transmitter, is established by its resonance *and/or* by a matching network.

Dipole

Taking the case of a classic resonant dipole, the RF currents will enter the dipole's feedpoint from the transmitter, and will begin to move toward the ends of the antenna elements (Fig. 10.3). After the current reaches the ends of the antenna's elements it now

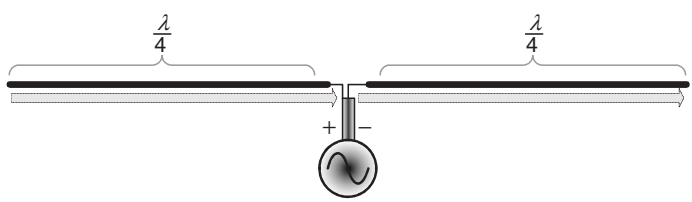


Figure 10.3 Resonant dipole antenna with current flow.

has, literally, nowhere to go, so it is reflected back toward the antenna's feedpoint where the next RF cycle appears from the transmitter to support this initial cycle. Accordingly, the transmitter must continually supply the periodic charging signals in order to continue these RF oscillations, or the initial charge would quickly be dissipated as heat within the resistive elements of the dipole structure itself, as well as by the desired RF output radiation. The cycle continues for as long as RF energy from the PA is fed into the antenna.

Since the voltage is low and the current is high at the half-wave dipole's center-fed feedpoint, its input impedance will also be low, while the impedance at the antenna's ends will be extremely high (both due to $Z = V/I$). Therefore, the current and voltage distribution along the dipole will establish its feedpoint impedance; and since a half-wave dipole is fed at the center, its ideal input impedance will be, as mentioned above, approximately 72 Ω. But maximum output power will only be possible if there are maximum standing waves on the antenna itself, thus creating the maximum voltage at the ends and the maximum current in its center. And since the power of the RF signal radiated from an antenna is also contingent on the RF currents within the antenna, then the more the current the higher the dipole's output power (for a structure of equal size). But for the energy to actually break free of the dipole antenna and be radiated far into space, the frequency must reach a point where the field lines (created by the RF currents) cannot fall back into the antenna's elements before the current changes polarity. The minimum frequency at which this can occur is considered to be approximately 10 kHz. And as the electromagnetic energy travels through space after leaving the dipole, it will eventually cut the elements of a receiving antenna inducing a tiny voltage, which must then be heavily amplified and filtered by a receiver in order to obtain the desired demodulated signal output, at the required amplitude and SNR, for use in a modem or detector.

Antenna Performance

Cutting an antenna to perfect resonance and matching the resistances for an excellent VSWR does not necessarily guarantee that the antenna structure is now a high-gain, high-efficiency radiator. Only far-field range testing can tell us that. Still, if we did not have the appropriate RF test equipment, such as a spectrum analyzer and a fully calibrated log-periodic antenna, along with a sufficiently reflection-free environment—all of which are needed to properly and accurately perform basic EM range testing—then the only way to be relatively assured that a particular antenna structure will operate with maximum down-range field strength is by not only possessing a good understanding of basic antenna design, but also running a realistic and accurate EM simulation, in conjunction with vector network analyzer (VNA) testing of the completed antenna's S_{11} (with matching circuit, if required). Nonetheless, even a perfect S_{11} is not necessarily a good indication that the antenna is radiating at a maximum level, as even a 50-Ω dummy load will provide us with a superb VSWR, and yet the dummy load is far removed from being a resonant or efficient antenna.

However, if the antenna can be designed, simulated, and tested for maximum field strength in our favored direction (indicating maximum gain), with an excellent input match (to prevent the RF energy from bouncing back to the transmitter), with high radiation efficiency (to prevent the RF from being dissipated as heat), and there are not an excessive number of reflective or absorptive structures within the antenna's near field (to prevent the disruption and weakening of the far field), then we can be assured that we have created a high-performance antenna structure.

10.1 Antenna Types

10.1.1 Introduction

For all portable wireless devices there are two general classifications of RF antennas, which can overlap: the *internal* and the *external* antenna structure. Sometimes this generalization may end in uncertainty, as some internal-type antennas may be found outside of the radio's housing itself, while what would normally be considered as an external antenna may be found located within the wireless device's plastic case. But external antennas are visibly attached to the outside of the radio's main plastic or metal case, or remote from the device itself. The antenna is usually connected to the radio's circuit board via coaxial cable or an RF connector, but the antenna may be just a simple piece of wire that extends from the PCB to above the wireless device's housing, and encased in its own radome (Fig. 10.4). However, there are many external antennas today that are found formed on their own external printed circuit boards.

Internal antennas can be classified as any radiating structure that does not extend above the wireless device's main enclosure (Fig. 10.5), and is either directly attached to the radio's printed circuit board (such as a piece of wire), or is actually printed on and is a direct part of the radio's own PCB.

10.1.2 Internal Antennas

An internal antenna, especially of the printed circuit board variety, will not have the excellent efficiency, high gain, or uniform radiation pattern of the typical external antenna. This decreased performance of most internal antennas is due to their proximity to the PCB's noise-producing circuits, as well as to the nearby energy absorbing metallic structures and lossy-dielectric objects within the wireless device's own plastic housing. In addition, because antennas store their energy in the near field, and this near field is where the far field radiated EM energy is acquired to form the wireless link itself, then

Figure 10.4 External antennas on a portable wireless device.

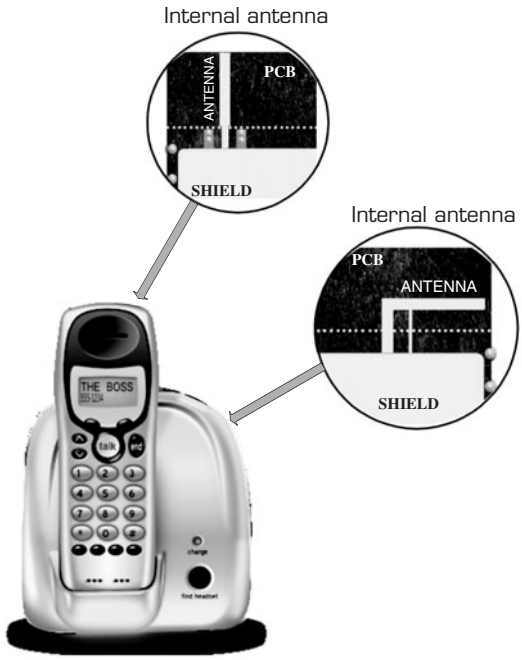


Figure 10.5 Internal antennas on a cordless telephone, of the printed PCB type.

any such energy drained from this near field will decrease the efficiency of the entire antenna system, decreasing the range of the wireless link. Indeed, when objects are placed within the near field's bubble of stored energy, a loss in gain as severe as 15 dB can occur to the outlying far field, as well as detuning the antenna. Because of near-field coupling and dielectric losses, as well as antenna size and form-factor constraints, it is not unusual for an internal antenna to actually contribute a total *loss*, rather than a *gain*, to the transmitter's RF output signal. But the use of these inexpensive, compact internal antennas can be necessary in portable and handheld transceiver or transmitter applications, and with a slightly more powerful PA making up for the lower antenna gain; but at the cost of decreased talk-time due to the unavoidable increase in battery power consumption.

When any full-length, resonant antenna is shortened to fit into a limited form-factor, as many internal antenna's are, the Q will increase and its bandwidth will decrease. Therefore, due to normal part-to-part tolerances variations of the antenna itself and its matching components, somewhat expensive production tuning may sometimes become necessary with very compact structures. A further, and quite significant, issue with such sensitive, high-Q antennas in portable wireless devices is that the user's hand and head capacitances may severely detune this internal antenna, and in these cases it may not be possible to reliably cover the full band of frequencies necessary, at least without excessive mismatch losses at the band edges. But shrinking an antenna to not less than half its full-size can and does give very good, trouble-free antenna performance, but in those extreme cases of element shortening where the Q becomes exceptionally high, we can place resistive loading within the antenna circuit to widen its bandwidth. This will,

indeed, widen the bandwidth to repeatable and manufacturability levels, but the losses introduced by this method may be unacceptably large.

Skillful antenna design is the most crucial aspect in achieving superior range and throughput in any wireless system. Especially challenging is the area of small, handheld RF devices, where both the antenna and the RF output power are always a compromise.

To improve compact internal antenna design, follow these guidelines:

a. The antenna's ground area (if required) should be as large and as uniform as the available space permits.
b. The antenna must be placed far from the PCB's noise-producing circuit elements.
c. The antenna should not have other conducting traces or wires placed in parallel nearby.
d. The antenna should be fed with coax or PCB microstrip and not simple wire.
e. If the antenna is of the printed type, it should be composed of wide traces.
f. The antenna should be as close to full size as possible, or not less than 50% of its resonant size.
g. The antenna should possess a low-loss LC match that is as close to the system impedance as possible.

Internal Monopole Antennas

In order to keep its length to a minimum, the internal type of monopole is normally a loaded antenna structure (Fig. 10.6). However, monopoles employed for the higher microwave frequencies may commonly utilize full-length antennas, due to their small size. Such full-length, resonant monopoles have an ideal radiation resistance of about 37 Ω, and an unbalanced, single-ended, input.

When the quarter-wave monopole is placed over an infinite groundplane, it replicates the two poles of a full dipole antenna, mirroring the missing quarter-wave pole through the reflective ground surface, and will therefore possess the classic ideal characteristics of a theoretical monopole. If the monopole is placed at right angles over any uniform groundplane area that has a radius of greater than one wavelength, the monopole's input characteristics and radiation pattern will still be very close to what it possessed with the hypothetical infinite groundplane. But when the groundplane is decreased to a quarter wavelength or under, the monopole's characteristics start to become noticeably affected, and begin to depart from that of the ideal structure. And because of the lack of any significant board space in almost all handheld wireless devices, this will be the normal state of the groundplane under a monopole.

The internal monopole's groundplane normally comprises, in a practical handheld implementation, the dedicated PCB copper groundplane, the RF shield, board traces, and even the operator's own hand. But due to the size constraints, the monopole's groundplane may be under a half-wavelength radius, far from uniform in copper area, and not at right angles to the antenna structure, all of which will begin to adversely influence its efficiency and input impedance, as well as making the antenna hard to tune. Add the PCB's own dielectric losses, in addition to the necessary matching lumped components and their losses, along with the antenna's proximity to nearby conductive coupling structures and EMI producing circuits, will make this a much less effective antenna than the full-size, external type. It is, however, an extremely low-cost and

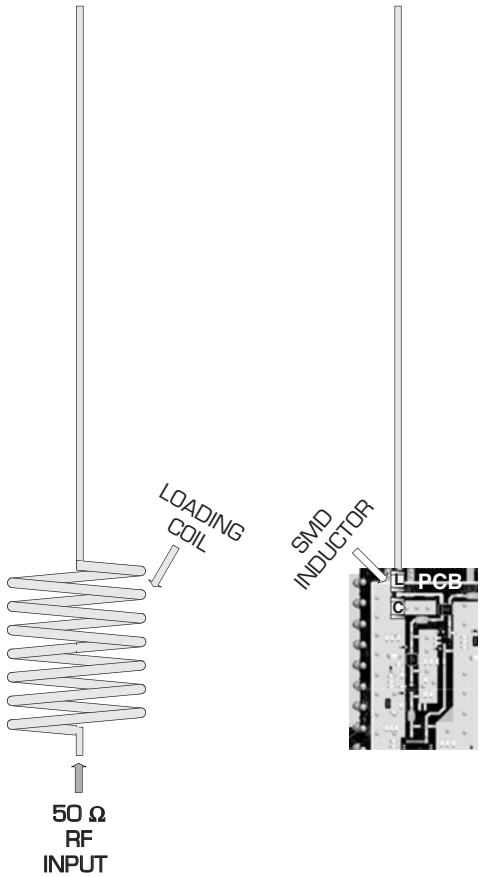


Figure 10.6 Two different types of loaded monopole antennas.

practical antenna: a piece of wire, and perhaps an LC matching network, are all that are required.

Internal PIFA antennas Antennas

The *printed internal-F antenna* (PIFA) of Fig. 10.7 is a printed trace on a PCB that is essentially a quarter-wave vertical antenna, but that has been bent horizontally in order to be parallel with the substrate's copper ground pour, and then fed at an appropriate point that will supply a good input match.

PIFA is an excellent choice for small, low-profile wireless designs, and is not as adversely affected by tiny, poorly shaped groundplanes as that of the monopole above. The PIFA also supplies decent efficiency, is of a compact geometry, and has a relatively omnidirectional radiation pattern (with some deep nulls). However, PIFA antennas do have somewhat of a narrower bandwidth than the average monopole.

Internal Loop antennas Antennas

Many compact, low-cost, low-power wireless devices use heavily loaded, low-efficiency, and electrically small loop antennas that can be printed on a PCB (Fig. 10.8).

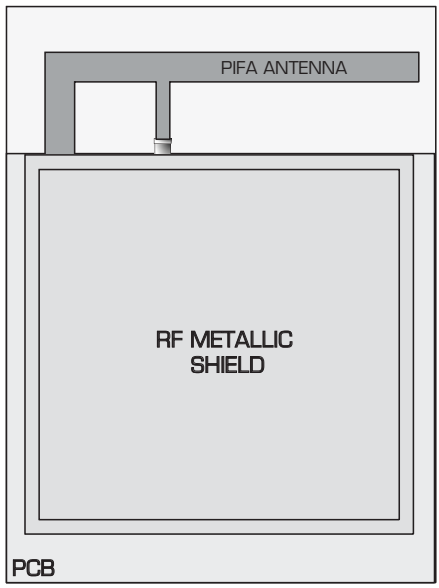


Figure 10.7 A printed PIFA PCB antenna.

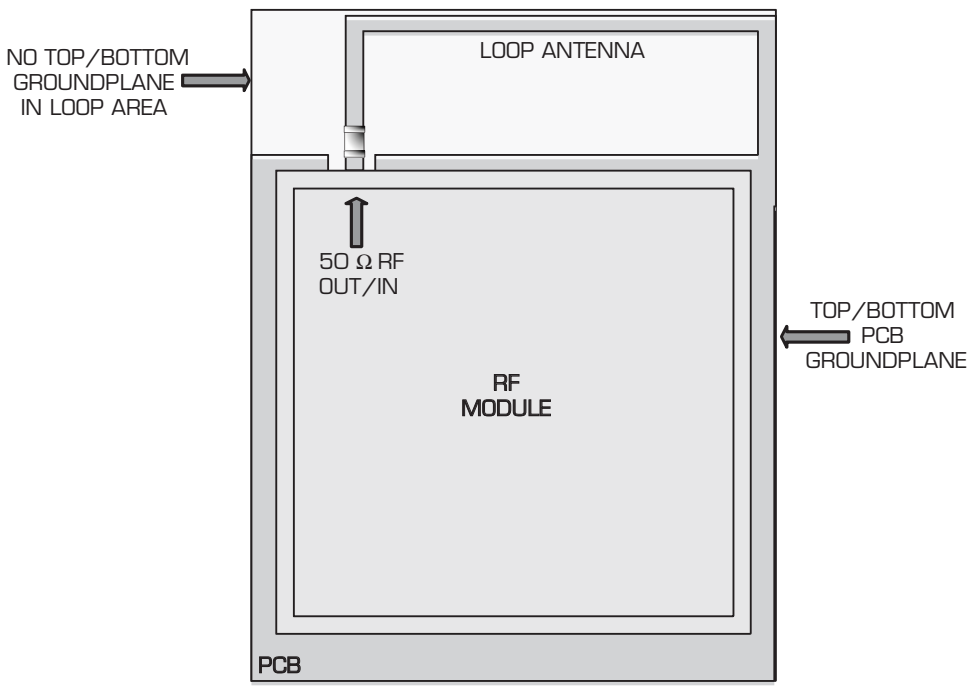


Figure 10.8 A printed PCB loop antenna.

These tiny loops can have a very low radiation resistance of 1 Ω or less, and with a very high loss resistance. This means that only a small portion of the loop's RF input power will be radiated, with the rest of the signal being consumed as heat. The low efficiency of this type of structure can be improved by increasing its loop area. Because of the loop antenna's generally small size, the *Q* will be extremely high when the impedance is matched, therefore making the bandwidth extremely narrow. In fact, unless we compromise and make the antenna's losses even higher than they already are, which we can do by broadening the input impedance match with resistive loading, or by performing a poor, lossy, and detuned LC match, then the narrow bandwidth will make tuning necessary for each and every PCB on a production run. This is not the appropriate antenna for most voice or data applications, but is a good alternative when the transmission range is very short and antenna size is a major issue.

10.1.3 External Antennas

There are obviously dozens, if not hundreds, of external antenna types that can be exploited for wireless applications, depending on the particular system requirements and antenna cost. Only the most common such structures will be covered below, and only those that are appropriate for portable applications.

External Dipole Antennas

One of the most prevalent antennas in RF design is the horizontal half-wave dipole. It can be utilized as the only element in an antenna, or as the driven element in tandem with other gain-enhancing conductive structures (such as that of the common Yagi). Dipoles are found in common operation from the lowest HF band (as 200-ft long wire elements strung between one or two tall towers), all the way up to a frequency of 60 GHz (as a planar metallic structure a few 100th of an inch long, printed on a special substrate). For portable use, a dipole antenna can also be mounted vertically (Fig. 10.9) for an almost null-free, omnidirectional groundplane-independent performance application.

External Monopole Antennas

As stated above for the internal monopole antenna, when this efficient quarter-wave structure is placed over an infinite groundplane, the external monopole antenna (Fig. 10.10) replicates a dipole antenna at infinite height, thus mirroring the missing quarter-wave pole of the dipole. If the monopole is placed over a groundplane with a radius greater than one wavelength, the antenna's input characteristics and radiation pattern will be close to that of the infinite groundplane case. If the groundplane is shrunken to quarter wavelength or under, the monopole's characteristics become strongly, and negatively, affected (such as its efficiency and input impedance), also making the antenna hard to tune. In a practical handheld monopole implementation, its groundplane may comprise only a small dedicated PCB copper groundplane, a RF shield box, assorted circuit traces and ground floods, and the user's own hand.

While the monopole is quite similar to a dipole, its ideal radiation resistance is about 37 Ω, or half the 72 Ω of the ideal dipole, and possesses an unbalanced input, making it perfect for most RF single-ended design applications (as opposed to the inconvenient and cost-enhancing balanced input needs of the dipole). Even without a matching network, a monopole can supply a decent interface to a 50-Ω system.

A monopole can be created by forming a quarter-wavelength long wire vertically over a groundplane, and then continually decreasing the wire's length until the VNA's S_{11} notch is on the frequency, and/or until the antenna's field strength is maximized.

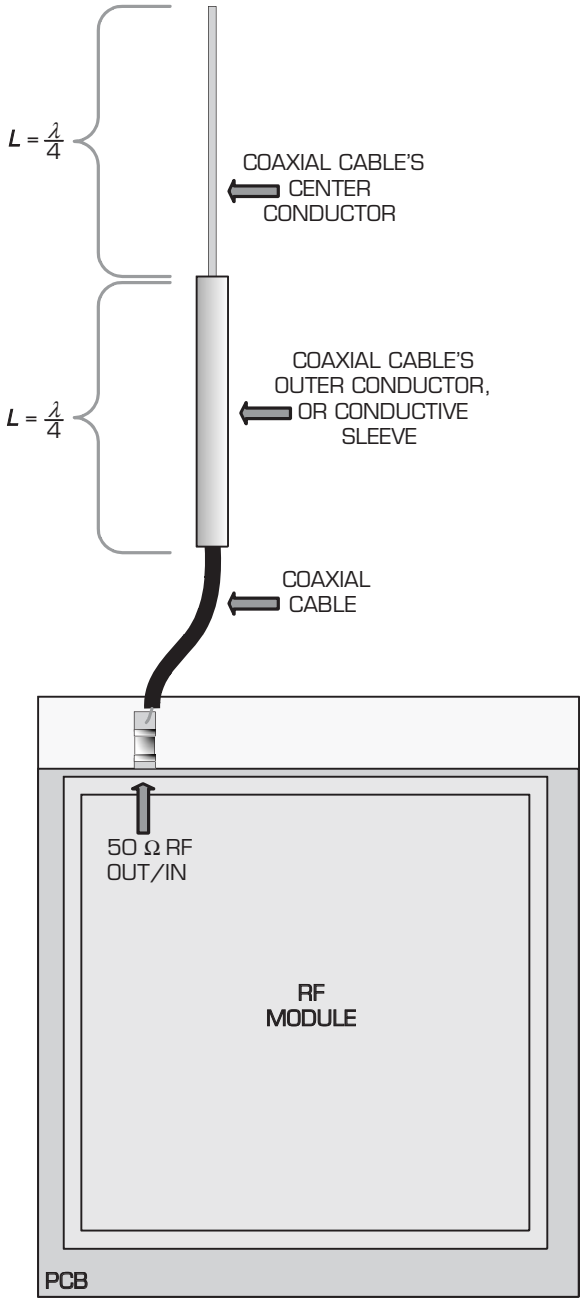


Figure 10.9 A vertical dipole sleeve antenna.

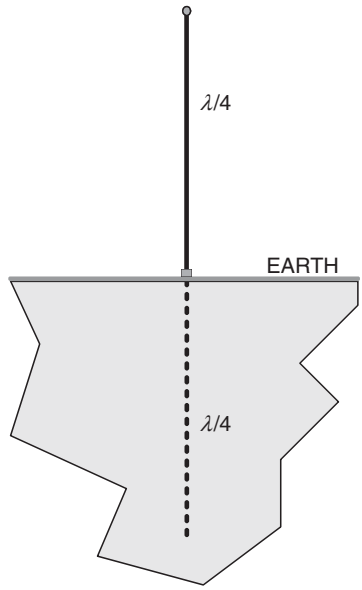


Figure 10.10 A quarter-wave monopole (vertical) antenna, with a groundplane supplying the other quarter wavelength.

10.1.4 Antenna Design

The following antennas are by far the most popular, the most straightforward to design and optimize, and are commonly used in small, portable or fixed, low-power wireless devices. They can be slightly modified to be either internal or external types.

Wire Monopole Antenna (Fig. 10.11)

Characteristics:

Gain = 1dBi; pattern = omnidirectional with medium null; polarization = vertical. (For vertically oriented handheld devices. Antenna located within radio's plastic housing.)

To Design

1. $L = 2950/f_r$

   where f_r = input frequency, MHz.
   L = length of wire monopole and minimum preferred radius of the monopole's groundplane, in

2. Length of the wire monopole must be tuned for S_{11} or VSWR dip, and/or monitored and tuned for maximum field strength. Must be fine-tuned with plastic housing and any PCB RF shield in place, as well as with any expected hand and head effects. Length of monopole can be made shorter or longer with the appropriate change in the LC matching. If groundplane's geometry is irregular, or if antenna is not at right angles to major surface of groundplane (both a common and sometimes unavoidable problem with internal verticals), then tuning may differ significantly from expectations.

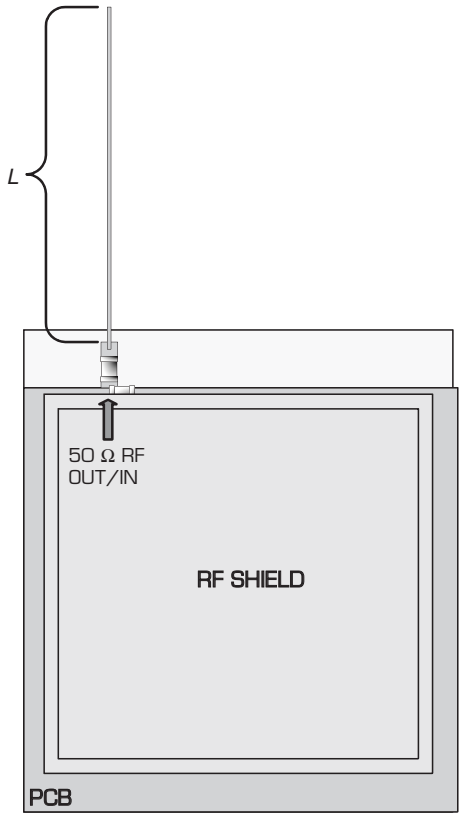


Figure 10.11 A wire monopole antenna, working against an RF shield as its groundplane.

PCB Monopole Antenna (Fig. 10.12)

Characteristics:

Gain = 0 dBi, pattern = omnidirectional with medium null, polarization = vertical. (For vertically oriented handheld devices. Antenna located within radio's plastic housing.)

To Design

1. $L = 2100/f_r$

   where f_r = input frequency, MHz
   L = approximate trace length of monopole and minimum preferred radius of the monopole's groundplane on the PCB, in

2. Length of the PCB printed monopole must be tuned for S_{11} or VSWR dip, or monitor and tune for maximum antenna field strength. Length can be decreased for a shorter antenna structure, but antenna will then have to be loaded with a high-Q series inductor at its feedpoint to obtain resonance (the gain and efficiency will decrease somewhat). Length is also dependent on the thickness and dielectric constant of the PCB material, as well as the wireless device's own plastic housing and hand/head effects.

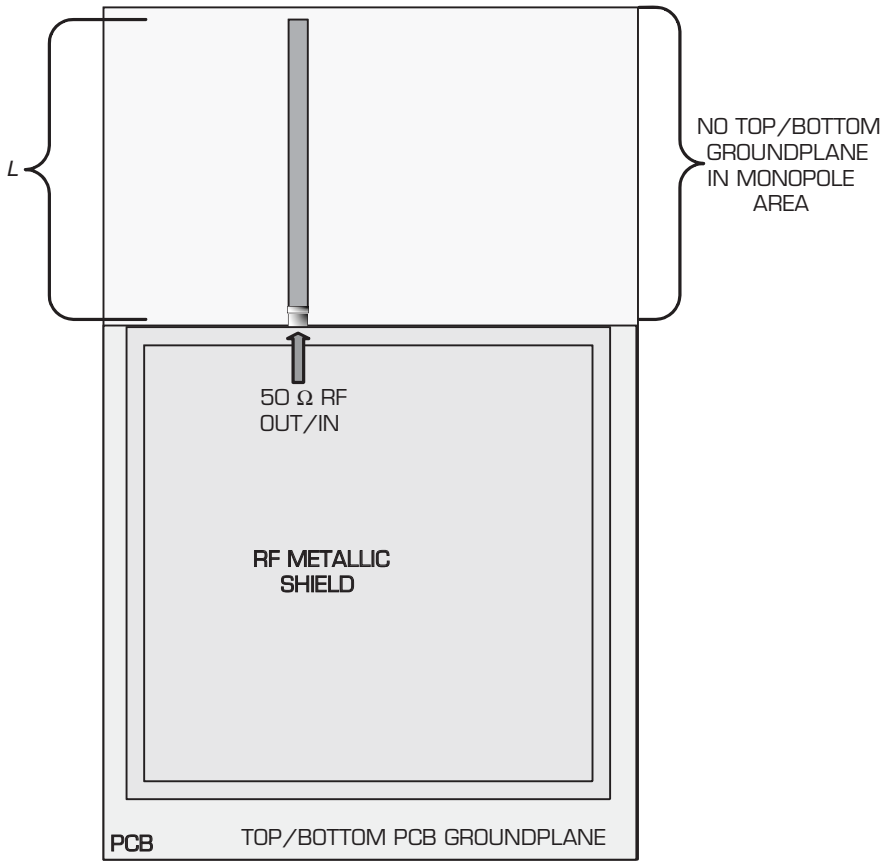


Figure 10.12 A printed monopole PCB antenna, working against an RF shield as its groundplane for maximum performance.

Wire Compact Helical (Fig. 10.13)

Characteristics:

Gain = –4 dBi, pattern = omnidirectional with medium null, polarization = vertical. (For vertically oriented handheld devices. Antenna is located within radio's plastic housing.)

To Design

1. $L = 7400/f_r$

   where f_r = input frequency, MHz
   L = approximate length of wire *before* coiling, in

2. Using initial length of wire L, wind a loose helical. Tune antenna by trimming length and/or by spreading turns and/or by tuning with LC-matching network. Final length will be about one-eighth to one-tenth of a wavelength long. Helical is sensitive to hand/head effects.

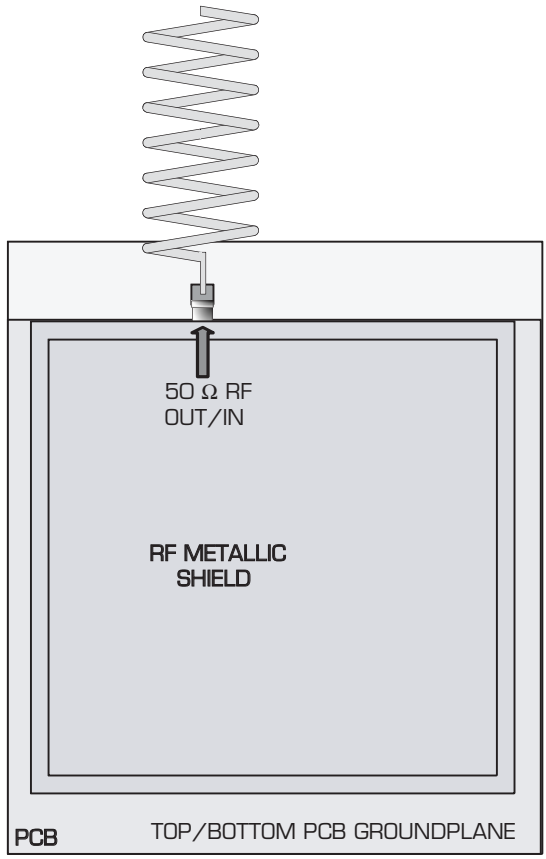


Figure 10.13 A helical antenna structure.

PCB Printed PIFA (Fig. 10.14)

Characteristics:
Gain = 1dBi, pattern = omnidirectional with strong null, polarization = linear. (For vertically oriented handheld devices. Antenna located within radio's plastic housing.)

To Design

1. $L = \lambda/4$
2. $H = L/3.5$
3. $W = L/10$

where L = length, in
H = height, in
W = width, in (trace width W is wide for increased PIFA bandwidth and lower resistive losses)
λ = wavelength of input signal on the PCB's substrate, in

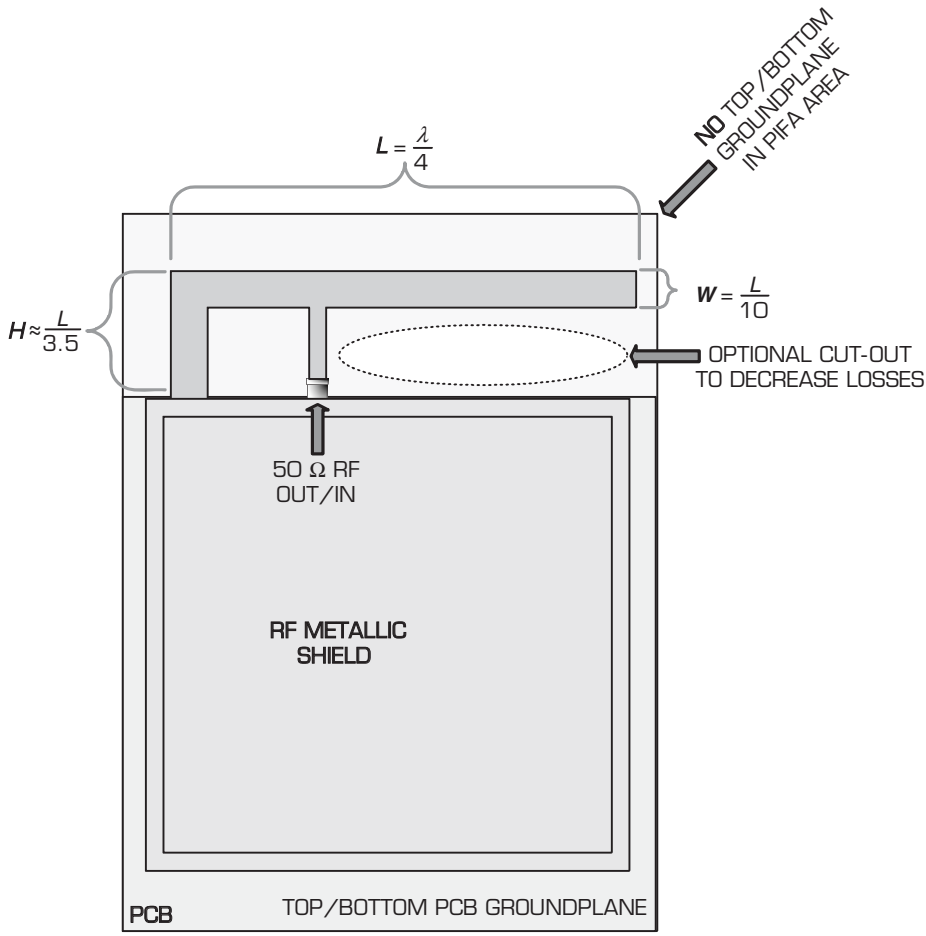


Figure 10.14 Designing a PIFA antenna structure, working against the module's RF shield as its groundplane for maximum performance.

4. Antenna's 50-Ω feedpoint tap to the main antenna structure must be found empirically for a good S_{11} match to system impedance.
5. Optimally, the groundplane should be part of the radio's RF shield structure, and the substrate can be cut out to decrease dielectric losses (as shown).

SMD Chip Antenna (Fig. 10.15)

Characteristics:
Gain = Varies with type/model (average –2dBi), pattern = varies with type/model (there may be a very deep null of up to 35 dB in the radiation pattern on some models), polarization = varies with type/model. (For vertically oriented handheld devices. Antenna located within radio's plastic housing.)

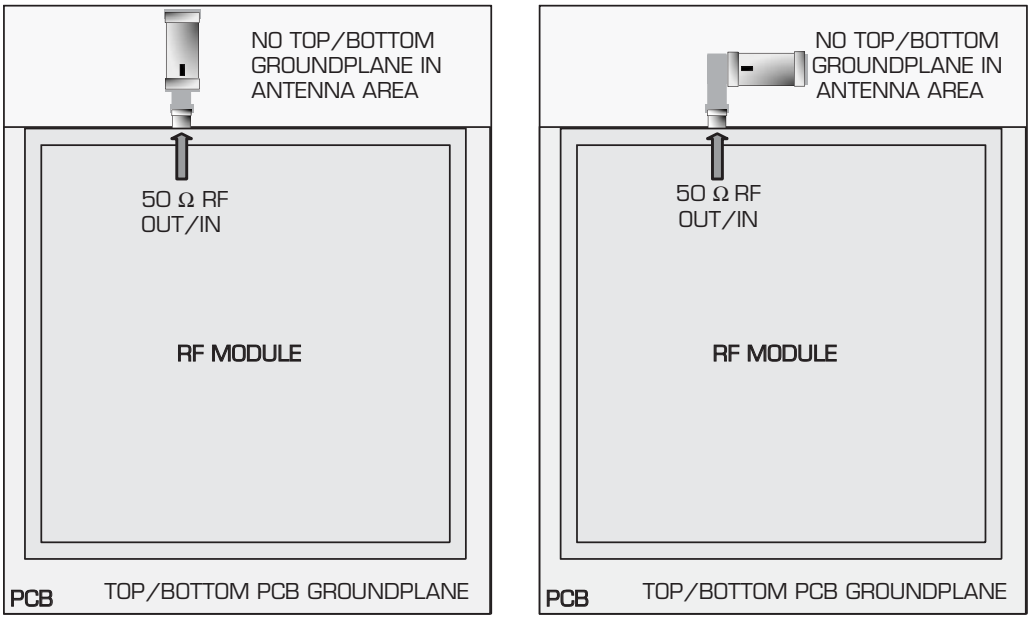


Figure 10.15 Two common ways of mounting surface-mount chip antennas.

To Design

1. This type of ultraminiature antenna is created for UHF and VHF frequencies. Due to their very small size, these antennas with their integrated planar radiating structures will have an extremely high Q. Thus, hand effects, dielectric thickness, and groundplane area can severely detune the antenna. The recommended PCB layout as shown on the chip antenna's data sheet must be followed implicitly. The antennas may also require a rather large *keep-out area* so as not to be detuned by any surrounding objects. However, if the keep-out area is too large, serious thought should be given to selecting another type of antenna which may have better performance at a lower cost and still fit within the same PCB area.
2. Tune antenna with a VNA for maximum S_{11}, with any added hand, head, and case effects added.

CHAPTER 11
Radio Frequency Simulation

During the past 20 years RF simulation software and the computer have become invaluable tools in wireless circuit design. Indeed, there are many helpful RF/microwave software programs available that will materially assist the engineer in optimizing a circuit or system for increased gain, output power, linearity, stability, return loss, and so on. Without computer simulation capabilities, many complex or highly optimized RF designs would simply not be possible, especially within the shortened time frame demanded of modern industry. This is not only true on the discrete circuit design level, but can also hold true for the overall wireless system's design.

11.1 RF Simulator Design Software

11.1.1 Introduction

Computer circuit simulation can involve one or more of the following diverse program types:

Spice

The most common electronic simulation software is *Berkeley Spice*, developed at Berkeley university some 30 years ago. A very powerful and effective, but easy to use, Berkeley Spice-based program is included with this book, Linear Technologies' superb *LTSpice*.

Spice permits simulation of linear and nonlinear circuit behavior within the time, frequency, and transient domains, while also allowing us to view the real-life effects that occur when biasing a transistor or diode. Unmodified Spice works extremely well at frequencies below 20 MHz, with this maximum frequency highly dependant on the particular lumped models employed for the Spice simulation itself. Some Spice models are optimized for much higher frequencies (however, high-frequency component models can be difficult to find), or can be made to operate more accurately at RF by inserting parasitic lumped components.

Spice can be very slow. A relatively simple nonlinear circuit may take up to an hour to simulate, depending on various factors. Microstrip transmission lines are also not available, which is a key limitation to the accuracy of all microwave circuits simulations. And quite often *convergence*, or the ability of Spice to accurately arrive at the correct answer, can become a major problem, since whatever arcane Spice convergence issue occurs during the simulation will have to be hunted down and corrected, and the simulation then rerun.

Nonetheless, even when using other and more RF-accurate simulation methodologies, Spice is still utterly invaluable in the simulation of many circuit designs,

especially for confirming proper device bias, and to examine the lower frequency output waveforms of any lumped circuit within the time and transient domains.

Linear Simulators

Linear circuit simulation programs, such as Agilent's *Genesys* or the included *Qucs* being two excellent examples, are the dominant program types employed in the RF and microwave world today. Linear simulators operate by exploiting *S-parameter models* for both the active and passive devices. (*S*-parameters are the most accurate method of representing components or circuits within the RF and microwave regions).

Linear simulators can simulate both lumped and distributed circuits, supplying us with forward and reverse gain results, as well as return losses, input/output impedances, phase angles, stability factors, and noise figures for any small-signal circuit (i.e., circuits that are not forced into nonlinear operation). Linear simulation software is also extremely fast and exceedingly accurate, with the necessary high frequency models being readily available, and real-time circuit tuning and rapid optimization very straightforward.

There are certain limitations to the linear *S*-parameter file-based simulation approach. For instance, the ability to look at DC or RF power, harmonics, and spurious signals is not possible, nor are time or transient domain views. And as the linear models are all text-file based, and are therefore already biased with a specific collector current (I_C) and collector-to-emitter voltage (V_{CE}), *S* parameters do not permit the actual biasing of active devices by the user within the simulator itself. (These bias values were previously selected when the physical component was captured into *S* parameters by the manufacturer. However, there are normally many active component *S*-parameter text files available that have an assortment of various biases for almost any application).

To better understand *S*-parameter simulations, we need to better understand the actual contents of a typical text-based *S*-parameter file. Studying Fig. 11.1, we see at the top of the file the *Comment* lines, which are inserted to add explanatory information to assist anyone who is viewing th e file's text. These lines will not be used by any part of the simulation software itself, and are indicated by a "!". After one or more comment lines indicating the device's manufacturer, model number, bias levels, and column heading (FREQ, S_{11}, S_{21}, S_{12}, S_{22}), the next will be the *Option* line, designated by a "#". This comprises GHz or MHz, indicating the frequency units used in the file; "S" for *S-parameters*; "MA" for *linear magnitude and angle* (or "RI" for *real and imaginary* or "DB" for *decibel magnitude and angle*); and "R 50" for the impedance value under which the *S*-parameters were measured. Next, the *Data* lines contain the frequency (in the same units as specified under the *Option* line) that was used to take all of the *S*-parameters on that particular line for that particular frequency. The next higher frequency immediately follows on each line, until the maximum test frequency has been reached.

General *circuit-theory simulations*, as are utilized in any basic *S*-parameter linear simulator, are adequate for virtually all RF circuit chores. However, with certain tricky simulations or at very high microwave frequencies, such software may become somewhat less accurate. This decreasing accuracy is caused by the fact that all RF components will store energy and produce overlapping fields, while even simple printed circuit traces are reactive and accumulate energy. Also, linear simulators will be incapable of simulating arbitrary planar structures, while an electromagnetic (EM) simulator—if given enough time and computer power—can simulate an entire complex RF printed circuit board (PCB) layout, along with all of the component's pads, traces,

```
! FILENAME (21bfg425.S2P)
! MODEL OF TRANSISTOR (BFG425W)
! TRANSISTOR BIAS (V2 = 2.000 E+000 V, I2 = 1.000 E–003 A)
# GHz  S  MA  R 50
!          S11          S21          S12          S22
!Freq(GHz) Mag   Ang     Mag   Ang     Mag   Ang     Mag   Ang
  0.040  0.950 –1.927   3.575 177.729 0.003 83.537  0.996 –1.116
  0.100  0.954 –5.309   3.518 175.247 0.007 87.057  0.996 –3.082
  0.200  0.951 –10.517  3.504 170.441 0.014 82.341  0.991 –6.343
  0.300  0.947 –15.891  3.496 166.534 0.020 78.681  0.988 –9.405
  0.400  0.941 –20.987  3.493 161.221 0.027 75.109  0.982 –12.576
  0.500  0.935 –26.297  3.476 156.531 0.033 71.254  0.974 –15.593
  0.600  0.928 –31.508  3.433 151.954 0.040 67.636  0.965 –18.605
  0.700  0.919 –36.669  3.384 147.515 0.046 63.875  0.954 –21.674
  0.800  0.910 –41.871  3.350 143.152 0.051 60.357  0.943 –24.600
  0.900  0.898 –46.948  3.317 138.801 0.057 56.929  0.930 –27.559
  1.000  0.886 –52.161  3.272 134.309 0.062 53.488  0.916 –30.396
  1.100  0.874 –57.181  3.223 130.114 0.067 50.181  0.903 –33.098
  1.200  0.861 –62.218  3.171 125.837 0.071 46.955  0.888 –35.859
  1.300  0.849 –67.154  3.119 121.786 0.075 43.791  0.873 –38.531
  1.400  0.835 –72.157  3.072 117.682 0.079 40.631  0.857 –41.151
 11.500  0.845 –2.938   0.375 -130.163 0.134 -104.397 0.607 14.337
 12.000  0.848 –9.981   0.326 -139.789 0.124 -115.184 0.658  4.326

!BFG425W_noise.xls                          !Philips part #:BFG425W
! Bias condition: VCE = 2V, IC = 1 mA       # MHz dB Ratio deg Ω

! Freq. Fmin Gmag Gangle Rn
900 0.7 0.67 19.1 0.40                      2000 1.3 0.56 57.5 0.36
```

Figure 11.1 The text contents of a device's *S*-Parameter file, used for linear simulations.

vias, and ground pours. Consequently, some circuits may need to be looked at electromagnetically within EM simulators employing Maxwell's equations. The excellent and very user-friendly Sonnet EM software, enclosed in this book's CD, is ideal for this application.

Harmonic Balance Simulators

Harmonic balance (HB) simulation methods are employed to model both linear and nonlinear circuit effects. They are capable of displaying a linear or nonlinear circuit's RF output power, harmonic levels, large-signal gain, intermodulation performance, *P*1dB, and so on, and can simulate closed loop oscillators and nonlinear mixers.

Unfortunately, when compared to linear analysis, HB component models are not only difficult to obtain, but the simulations themselves may take a long time to run. Harmonic balance techniques also cannot determine transient effects, and are rather poor at accurately measuring the intermodulation distortion (IMD) products and higher order harmonics of mixers and saturated amplifiers. The software itself is also expensive to purchase, and requires a much more powerful computer than a linear simulator. Yet, if this type of simulator can be afforded, it is nevertheless an incredible and enlightening look into the functionality of many RF circuit designs that contain nonlinear components.

A cousin to harmonic balance is the *Volterra series*, which accurately models the slightly nonlinear effects present within all linear circuits. This method is not meant to model grossly nonlinear circuits such as mixers, Class C amplifiers, oscillators, or frequency multipliers.

Electromagnetic Simulators

Planar electromagnetic analysis software employs the *Method of Moments* or the *Method of Lines* technique to linearly simulate microstrip, stripline, or arbitrary 2-D metallic and dielectric structures at RF and microwave frequencies. Most modern EM simulators permit adding lumped *S*-parameter or ideal component models into the planar circuit as well. This category of simulator is able to accurately display the gain and return loss of distributed filters, transmission lines, spiral inductors, patch antennas, and more, in addition to presenting the actual RF current flow and current density running through these metallic structures. The common *circuit-theory* based algorithms of other simulation programs are typically inadequate to accurately model these types of microwave structures, and their stray coupling interactions and box modes. In fact, the EM simulator supplied with this book, *Sonnet Lite*, is the light version of one of the premier EM simulation packages on the market today, *Sonnet Suite*, and is perfect for this type of wireless application.

System Simulators

System simulators are outstanding at reproducing an entire transmitter and/or receiver system in software. The system chain comprises blocks of amplifiers, filters, mixers, local oscillators, and the like, with each block being assigned a specific stage gain, insertion loss, noise figure, *P1*dB, third-order intercept (IP3), temperature, and so on. Such simulators quickly give the engineer the overall transmitter or receiver results of his paper design, and whether the design will meet all the desired specifications, such as cascaded gain, noise figure (NF), linearity, sensitivity, output power, passband shape, in-band spur levels and their locations, transmit-to-receive path losses, and the like. The enclosed *RF Workbench* program by RFCafe.com will easily supply us with all of these answers.

11.1.2 RF Simulator Overview

Since various RF circuit simulators have different strengths and different weaknesses, here is a quick rundown on each of the general types discussed above.

A. *Linear Simulators*:

Pros

1. Very fast simulation speeds.
2. Instant visual feedback information during circuit tuning.
3. Models available for most RF parts.
4. Simple to learn and use.
5. Needs modest computer power to run rapidly.
6. Always converges to an accurate answer.
7. The software is relatively low in cost.

Cons

1. Gives no information on any nonlinear effects, such as harmonics, spurious, power levels, compression, or mixing products.
2. No data possible on current or voltages.
3. Inc apable of displaying transient effects or time-domain waveforms.

Conclusion

A linear simulator can be utilized to accurately and rapidly simulate over 95% of all linear RF/microwave circuit designs. Not owning and using an RF linear simulator is simply not an option in the wireless design field.

B. *Non-linear Simulators*:

Pros

1. Can model all linear and nonlinear effects (at steady state).
2. Can display currents, voltages, and RF power levels.
3. If special high-frequency nonlinear models are not available, common Spice models can be used.

Cons

1. Simulation times are much longer than linear simulators.
2. The software is very expensive.
3. Fast computers are needed to run effectively.
4. Convergence is not always successful.
5. Cannot simulate oscillator or amplifier transient behavior.
6. Displays only spectral lines, without bandwidth.
7. The active devices require nonlinear models, which are not always available.

Conclusion

Nonlinear simulators are considered necessary for power amplifiers, mixers, or other nonlinear circuits, but are still regarded as somewhat of a luxury for use in the linear world.

C. *Berkeley Spice Simulators:*

Pros

1. Very low purchase price (some are even free).
2. Excellent at displaying nonlinear transient and steady state time-domain information.
3. Readily gives us current and voltage amplitude levels.
4. Nonlinear models available for many active devices.

Cons

1. Cannot model distributed circuits easily.
2. May be quite slow during simulation runs.
3. Does not always converge on an answer.
4. Spectral frequency-domain information is only available through fast Fourier transform (FFTs).
5. Part's models may not always be accurate above 30 MHz.
6. There is no built-in way to add microstrip effects.
7. Modeling general high frequency effects can be challenging in unmodified Spice.
8. Does not natively display watts (or dBm) in FFT graphs (must typically form a linear *V-I* expression to obtain FFT power).
9. Slows dramatically when employing multitone excitation with two diverse frequencies (multitone excitation to measure two-tone intermodulation products can prove excessively time-consuming to simulate).

Conclusion

Spice is best used to simulate real-life, submicrowave circuit behavior, and delivers an invaluable insight into all levels of AC/DC circuit operation.

D. *Electromagnetic Planar Simulation Software*:

Pros

1. Supplies very accurate linear data and current-density information of arbitrary 2D (planar) printed circuit structures, with all electromagnetic coupling effects accurately taken into account.
2. Many modern EM simulators can also simulate an entire linear circuit with *S*-parameter and/or ideal lumped component models included. (The lumped components themselves are not actually EM simulated, but are sent into a linear simulator engine, while only the planar PCB copper itself is EM simulated. The two diverse results are then accurately recombined and presented.)

Cons

1. Extremely expensive to purchase.
2. Very slow simulation runs for medium and complex structures.
3. Limited in utility unless designs are above 6 GHz, or distributed planar filter structures are involved, or we have other 2D metal structures that are of an arbitrary shape.
4. Requires high-end computer power to run effectively for larger or more complex planar circuits.
5. Due to the software's gridlike simulation nature, inaccuracies will occur unless EM cell size is extremely small, which slows the simulation (unless *conformal meshing* is available, which is an expensive option, if offered).

Conclusion

An EM simulator can prove invaluable if planning to design accurate planar circuit structures at microwave frequencies, but at a higher simulation software/hardware cost than other, somewhat less precise, solutions.

11.1.3 RF Simulator Software Programs

There are many free or low-cost RF simulation programs that will assist the engineer in the successful design of a high-frequency wireless circuit or system. There are also more expensive and complex programs available for more intricate and powerful simulation requirements. All of the programs as presented below will run quickly and reliably on almost any modern personal computer.

LINC2 Pro [Applied Computational Sciences (ACS)] is a very affordable, but extremely powerful, alternative to the more expensive linear simulators. LINC2 Pro not only supplies the standard graphing, simulation, and optimization abilities of much higher priced packages, but also possesses circuit synthesis capabilities for generating automatic lumped and distributed matching networks, single-ended and differential attenuator pads of any input/output impedance, microstrip transmission lines, baluns, and couplers. Additionally, there are LINC2 synthesis tools available that will automatically design lumped and distributed filters, as well as single and multistage linear amplifiers. Indeed, LINC2's LNA design module is the most powerful synthesis tool for RF amplifiers available within any software at any price. Another impressive feature is the fully automatic exporting of all LINC2 Pro circuit schematics directly into AWR's *Microwave Office*, as well as automatic export of all LINC2 Pro distributed filter designs straight into *Sonnet EM* software. And, incredibly, LINC2 also comes with free lifetime technical support. (This support, which must normally be paid yearly, can become a significant cost with virtually every other RF simulator software).

Sonnet Lite, by Sonnet Software, is an advanced electromagnetic simulator that permits a microwave designer to simulate planar (microstrip) circuits and patch antennas, and displays a graphical output of linear gain and return loss. Lumped components can also be easily added to the simulations. A very important software tool in all high-frequency microwave planar design. Sonnet Lite is included within this book's CD.

AppCAD is a free program from Agilent, and is included for your convenience in this book's CD ROM. AppCAD helps the engineer to instantly design bias networks for bipolar junction transistors, field-effect transistors, and monolithic microwave integrated circuits (MMIC), as well as complete detector circuits and microstrip/stripline/ coplanar waveguide transmission lines. AppCAD also has a reflection calculator to compute voltage standing wave ratio , return loss, and mismatch loss for any desired input and output circuit impedance, a noise calculator to compute a receiver's NF, a standard value calculator for passive components, a section that displays all the common engineering constants (Boltzmann's, the speed of light, Planck's, and the like). Even more functionality to this program is expected to be added by Agilent in the future.

Qucs circuit simulator, with the full version included in this book's disk, currently performs linear *S*-parameter, as well as time-domain transient analysis, in a simplified, user-friendly GUI reminiscent of Agilent's *ADS*. Qucs has a built-in optimizer, microstrip discontinuities, and multiple selections of various RF sources. It can also synthesize lumped filters, microstrip, lumped-matching circuits, and pad attenuators. Qucs is an enormously powerful and stable RF circuit simulator.

PLLs Made Easy (or *Easy-PLL*), by National Semiconductor, makes the design of phase-locked loops (PLLs) a much more accurate, speedy, and repeatable process. It is available on the National Semiconductor's Web site as a continually updated online tool. By simply placing the desired reference frequency, tuning range, and comparison frequency within the appropriate blocks of the program, Easy-PLL will actually select the proper voltage controlled oscillator and National PLL chip, as well as design the entire loop filter, and then provide the engineer with the phase-locked loop's complete performance characteristics, such as phase noise and lock time. Easy-PLL will also perform a comprehensive design check that guarantees PLL stability and proper component values.

AADE Filter Designer, by Almost All Digital Electronics (AADE.com), is a free program included on the enclosed CD, and helps the engineer rapidly design almost any type of lumped passive lowpass, highpass, bandpass, and bandstop filters, such as Butterworth, Chebyshev, Elliptic, Bessel, Gaussian, crystal ladder, and the like, while also displaying the filter's insertion loss, return loss, group delay, input impedance, and the like. As with all lumped circuit design synthesis programs on the market today, some costing many thousands of dollars, AADE Filter Designer uses ideal components (except for inductor *Q*) and cannot model the influences of interconnecting microstrips or traces. Therefore, stray coupling, distributed reactances, and electrical length will all begin to undermine the synthesis accuracy as the frequency increases above 50 MHz or so. Thus, any ensuing filter design must be placed in an RF linear simulator, with the appropriate missing microstrip elements and models added and the filter tuned for the desired performance level. (This need for *all* circuit synthesis programs to utilize only ideal components is simply a fact of life, no matter what the cost of the synthesis software.) AADE Filter Designer is indeed a must-have tool for any advanced lumped filter design effort.

Impedance Matching Network Designer by John Wetherell is a free online program that assists in the rapid design of 16 various (ideal) RF-matching networks. Simply enter the source's resistance and reactance, along with the frequency of interest and the desired

circuit *Q*, and this application will immediately display the values of all the appropriate matching topologies.

MixSpur, by the Engineer's Club, is a program that will graphically and in tabular format display the spurious output frequencies, and their amplitude, as emitted from a mixer and local oscillator (LO) stage. Use MixSpur to confirm that both in-band and out-of-band mixer-generated spurii are below specifications for receiver and transmitter design; or will more filtering, a new LO frequency, or a different mixer be required? Available for sale online.

RF Workbench, sold by RFCafe.com, is a low-cost system design program that will assist the RF designer to define the specifications of a transmitter or receiver chain's amplifiers, filters, and mixer stages, and then view the resultant spectrum in the frequency domain, with bandwidth. RF Workbench will not only permit you to see if the system's sum frequency or LO feedthrough will be troublesome, but if any mixer spurs are in or out of band, and what their amplitudes may be. It also shows you if any particular stage is going into compression, what the system's *gain*, *P*1dB, IP3, P_{OUT}, and NF values are, calculates the complete transmitter-to-receiver path, and creates a frequency and cascaded amplitude budget plan. A truly invaluable program, the fully operational shareware version is included in this book's CD. (If Excel is available, *RF Cascade Workbook* is another outstanding program available on RFcafe.com for systems cascade analysis.)

A *Spice* program of some sort will always be required for certain AC/DC time-domain circuit simulation tasks. Linear Technology's excellent *LTSpice* is included in the enclosed CD. Also, Beige Bag software's *B2 Spice* is a well-known and very effective Spice-based simulator with a multitude of powerful features and capabilities. Numerous other Spice-based programs can be found on the market today.

There are additional RF software tools available and in common use among most RF design engineers. Such as the expensive, definitive software suites that model almost all linear, nonlinear, and electromagnetic effects of all kinds, such as the first-class *Eagleware Genesys* by Agilent; the ubiquitous *ADS*, also by Agilent; and the outstanding *Microwave Office*, by AWR.

11.1.4 RF Simulator Accuracy and Guidelines

Many subtle factors can lower or even destroy an RF circuit simulation's real-world accuracy. This section will go into many of these issues.

Model Accuracy

A major impediment to accuracy in RF linear and nonlinear simulators is the inability to easily export the entire circuit layout, as simulated, directly into a PCB layout package (the PCB layout itself is a major part of any RF design above a few 10s of MHz). Indeed, if the final PCB's layout is different from the RF simulator's circuit layout, then the real-world results will differ, sometimes quite drastically. We must perform this time-consuming and difficult duplicate layout task by hand, and as accurately as possible.

However, a second major limitation, which can drastically affect simulation accuracy as well, is the lack of truly accurate RF passive component models. This is because any *S*-parameter model, which is typically used for all RF simulations, cannot be expected to be completely accurate for every PCB substrate's dielectric constant and board thickness. Therefore, if the RF designer selects a board material or thickness that differs from how the *S*-parameters were taken, then simulation accuracy will suffer, especially as the frequencies increase. In other words, all available *S*-parameter

models are highly substrate dependent. Also, some of these *S*-parameter models may have PCB footprint pad effects included, while others may not; which is another substrate dependence that can significantly influence simulation accuracy. Further, *S*-parameter models are not the optimal method to use in RF simulations, since they cannot easily be changed in value during the software tuning of the circuit design itself. They also cannot be automatically optimized by the simulator, nor can we perform a Monte-Carlo tolerance evaluation. Nevertheless, this significant RF simulator limitation in the lack of adequate RF passive component models for the microwave region has finally been successfully addressed by Modelithics, Inc. Modelithics creates passive models that are invaluable for use in all accurate RF simulations above 900 MHz, and which have the full capability to factor-in a component's PCB footprint parasitics, permits full circuit optimizations to be run on each part's value, and (most importantly) the board's substrate thickness and dielectric constant are all automatically factored into the RF simulation. The resultant simulation *versus* real-world results is incomparable, with previously unheard of accuracies over anything ever obtained in the RF/microwave simulation world. In today's highly competitive wireless microwave market, Modelithics models are no longer simply an option when designing at high frequencies; they have now become a requirement for highly accurate, first-pass RF design.

Accurate RF Simulation

If some type of valid RF computer simulation is not performed after the paper circuit design has been completed, much tuning and PCB rework will have to be performed to force a microwave or RF design to function properly, even if we follow an extremely cautious wireless design and PCB layout methodology. This is due to the profusion of undesired, real-world component and board parasitics, along with the various component/trace tolerances and temperature affects, which will all conspire to alter the expected performance. These influences cannot be realistically calculated during the design stage, either on paper or within current synthesis software packages. However, they can be easily and automatically reproduced in any common RF simulator program, and the synthesized circuit values then appropriately tuned and optimized in software.

As a result, RF circuit design and simulation, to be truly accurate and useful, must be carried out in certain, fundamental ways. As all initial RF designs must first be calculated with the assumption of ideal components and zero length traces, we must first model and then tune our design under these unrealistic conditions. Then, after the circuit operates as we had originally anticipated, the next step is to replace all the ideal components with the appropriate and more lifelike active and passive models, and add all PCB traces, microstrips, pads, and bends. (This is not of any concern in a circuit or section of a circuit that will only be seeing low frequency AC or DC waveforms, as trace/component parasitics and length will have little influence.) Next, we will tune this more realistic circuit until we obtain the performance desired. Due to significant real-life implementation losses, such performance will generally be somewhat inferior to our original paper calculations.

After this simulated circuit is sent out for fabrication and assembly, further performance degradation may occur due to the inaccuracy of the original *S*-parameter models used in the simulator, with this model inaccuracy caused by the parasitic effects as outlined above, and also by the component manufacturer's own *S*-parameter

measurement errors and/or the unavoidable necessity to supply a single model that is actually a convenient statistical average of multiple tested components (since real components will always have a particular tolerance, with values that vary from part-to-part and lot-to-lot). The consumer-grade substrate material itself will also possess dielectric constant and dimensional variations over that of the simulator's precise and unrealistic homogonous substrate, thus adding to a certain innate inaccuracy of any circuit simulation *versus* a real physical PCB. Moreover, the assembled PCB's components will be inserted and soldered at a slightly different distance from each other than had been specified by the simulation, causing another small simulator to real life variation. Lastly, if the particular component models used in the simulation have not been characterized at the exact low- or high-frequency band we may be designing in, then a further inaccuracy may crop-up, called *extrapolation error*. This is due to the RF simulator actually speculating as to the true value of the missing data points, sometimes causing significant inaccuracies. However, if performed correctly, linear RF simulations can be extremely accurate up to 6 GHz, even with all of the concerns expressed above. In fact, the following is an excellent step-by-step procedure on how to perform very precise circuit simulations, and with the absolute minimum of simulation to real-life error.

Accurate RF Simulation Steps

Two wideband bandpass filters were designed, simulated, and built to demonstrate the level of accuracy that can be supplied by employing the proper simulation techniques. Due to PCB size constraints, as well as center frequencies, bandwidths, performance, and cost, an LC highpass/lowpass lumped bandpass filter (BPF) topology was selected for these structures. The filters will employ high-quality consumer grade 0402-size capacitors by ATC (600L series) and inductors by Coilcraft (0402CS series).

The two BPFs were designed by using standard modern filter design methods, and assuming ideal components and perfect, zero-length conductors. The resultant circuits were then laid-out to fit within the limited PCB physical space constraints [(Fig. 11.2(*a*)]. Next, the discrete circuit was entered into Agilent's Eagleware linear simulator following, *exactly*, the PCB layout's dimensions and topology (Fig. 11.3). In other words, each of the PCB layout's copper "T" junctions, bends, width variations, spacings, lengths, significant vias, and the like, were meticulously recreated and placed in the simulator (Fig. 11.4). Indeed, if the RF *simulation* layout does not follow the physical *PCB board* layout closely, the level of simulation accuracy will be quite poor at RF and microwave frequencies. A quick method to confirm that the simulation layout is actually following the PCB layout is to make use of the simulator's own built-in EM or PCB layout capability, when available, as shown in Fig. 11.2(*b*). By commanding the program to create a full EM simulation from the linear simulator's circuit, we can now visually and intuitively see that both the physical PCB layout and the simulator's layout are virtually identical.

NOTE: *Looking closely at Fig. 11.2, we see that the component's pads are actually placed at the* edge *of the copper microstrip, rather than directly on top of the microstrip itself, which would normally be preferred. This is to avoid the need to modify the component model for a new, and in this case, negative, reference plane, and also for complex footprint pad effects.*

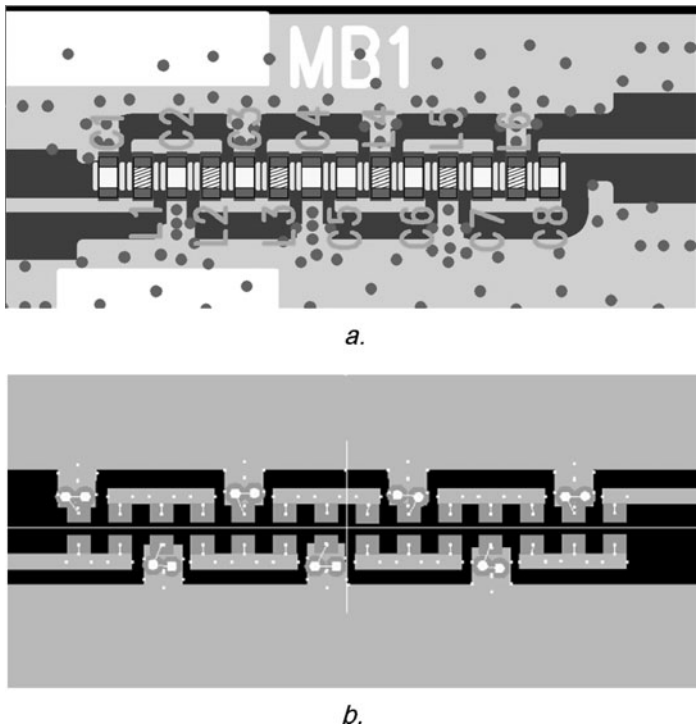


Figure 11.2 (*a*) Layout of bandpass filter's PCB test boards (with passives added for clarity), and (*b*) with the EM simulator's layout.

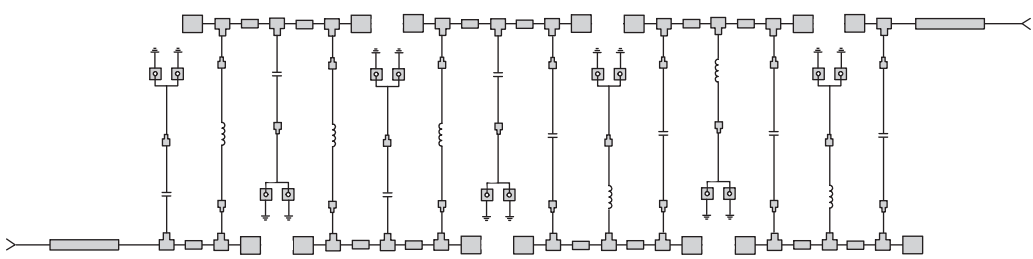


Figure 11.3 Lowpass/highpass filter structure as simulated for both the 690 to 1135 MHz and the 1105 to 1815 MHz BPF.

The Modelithic models are set to the component values as calculated above for the ideal BPF circuit, and the part's PCB footprint pad sizes are entered. The simulation is run, and the part's values consecutively tuned and optimized for the desired filter response. All final LC component values must then actually be equal to what is currently and readily available in the marketplace. This very important last requirement complicates optimization greatly, but is an obvious and often times overlooked practical necessity. The final gain and return loss graphs are then taken of both BPF circuits (Figs. 11.5 and 11.6). To make sure the filters will also operate over the normal

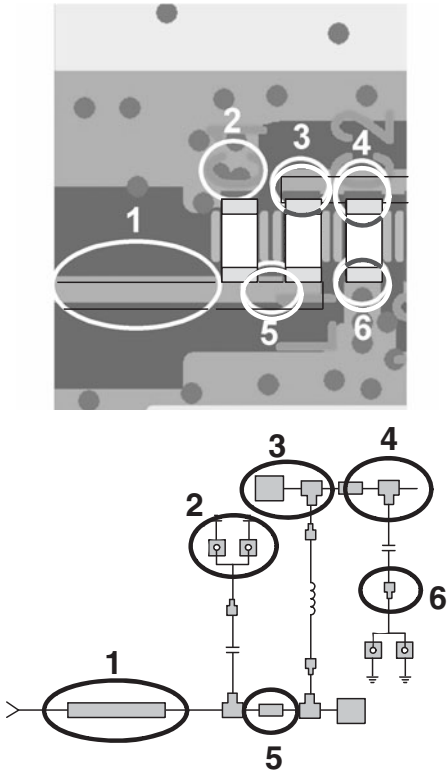


FIGURE 11.4 The simulation layout must equal the PCB layout: 1. Input trace; 2. Critical vias; 3. Right-angle bend metal (with end effect); 4. T-junction trace; 5. Inter-pad trace between components; 6. Step discontinuity used when joining two different trace or pad widths.

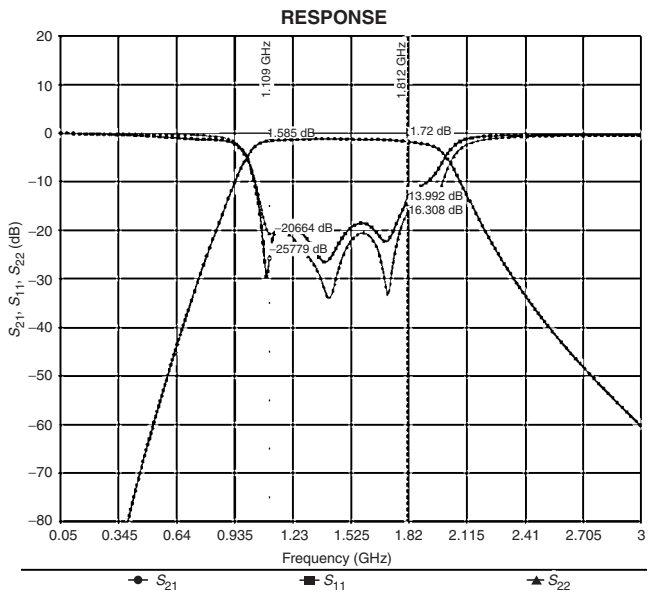


FIGURE 11.5 Gain and return loss for the 1105 to 1815 MHz simulated bandpass filter.

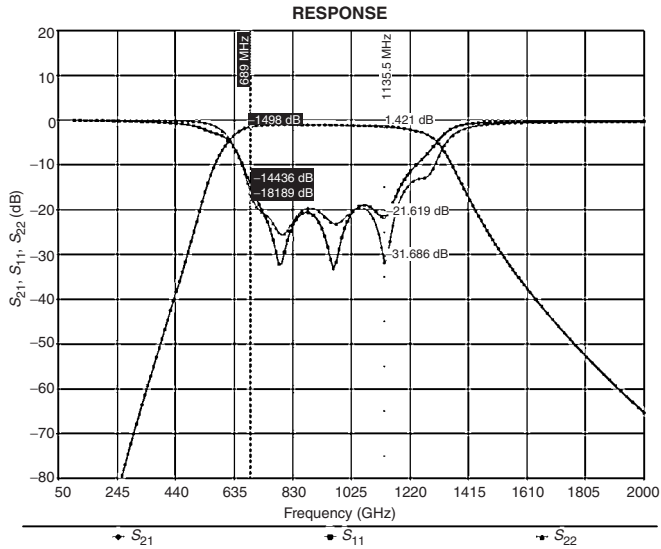


FIGURE 11.6 Gain and return loss for the 690 to 1135 MHz simulated BPF.

range of tolerances for the particular components selected, a Monte Carlo analysis is run on each circuit (Figs. 11.7 and 11-8).

Gerber files of the PCB layout were then produced and sent to the board fabrication house, and the completed PCB sent on to the part's assemblers. When the stuffed PCBs were received back from the assemblers, they were thoroughly tested with a fully calibrated *Agilent 8719ES* vector network analyzer to confirm that the real-life filters did indeed duplicate the RF simulation results. Glancing at the VNA's graphs of Figs. 11.9

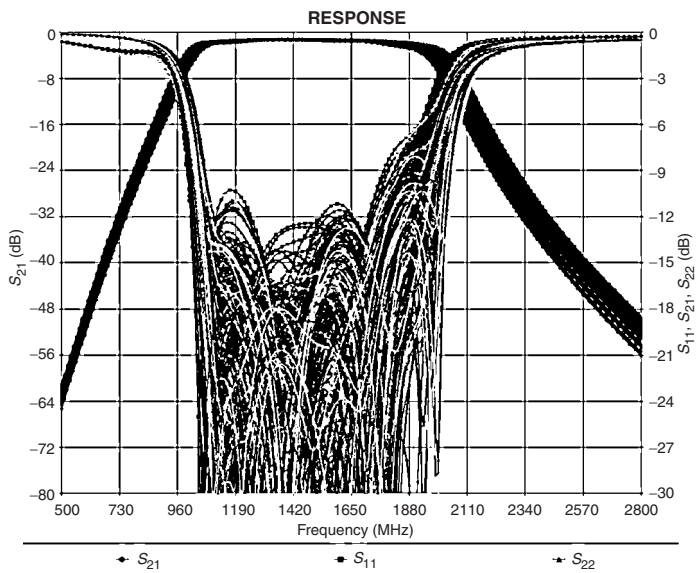


FIGURE 11.7 Monte Carlo simulation analysis for 1105 to 1815 MHz BPF.
(*Note:* Normal distribution, sigma of 10%, capacitor tolerance of ±5%, inductor tolerance of ±2%.)

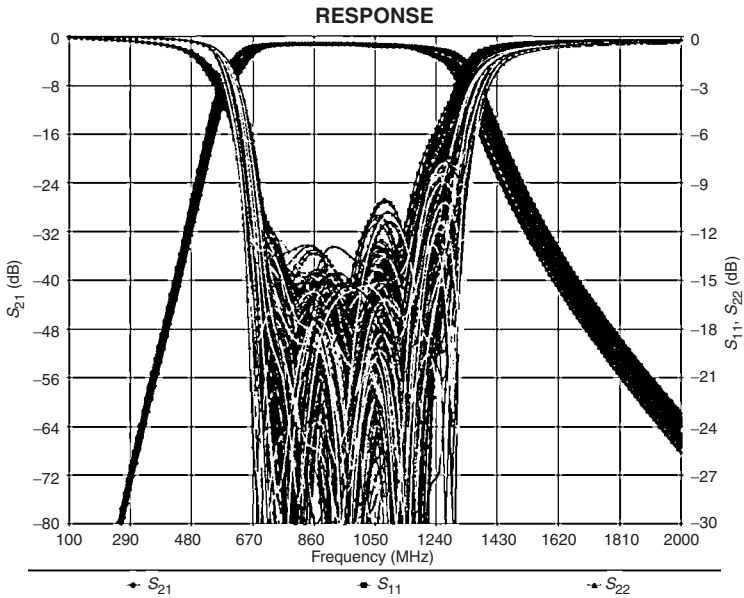
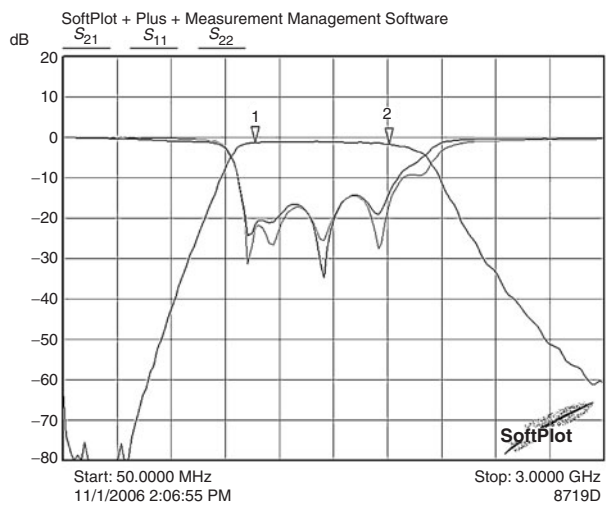
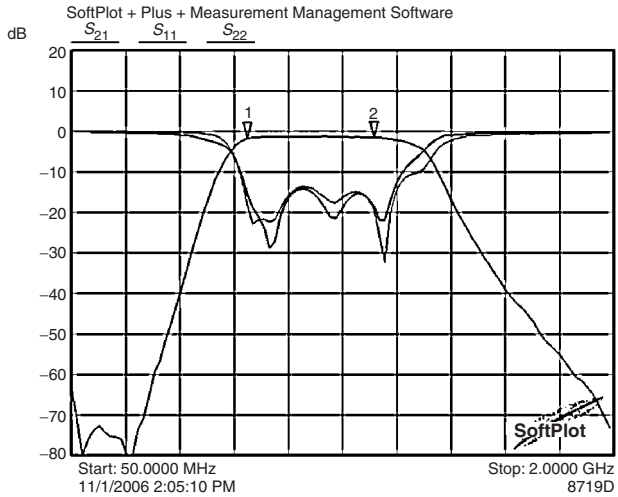


Figure 11.8 Monte Carlo analysis of 690 to 1135 MHz BPF.
(*Note*: Normal distribution, sigma of 10%, capacitor tolerance of ±5%, inductor tolerance of ±2%.)



Mlkr	Trace	X-Axis	Value	Notes
1 ▽	S_{21}	1.0968 GHz	–1.33 dB	
2 ▽	S_{21}	1.8275 GHz	–1.67 dB	

Figure 11.9 VNA test results showing gain and return loss of the real 1105 to 1815 MHz fabricated bandpass filter.



Mlkr	Trace	X-Axis	Value	Notes
1 ▽	S_{21}	682.0000 GHz	–1.73 dB	
2 ▽	S_{21}	1.1363 GHz	–1.51 dB	

Figure 11.10 VNA test results showing gain and return loss of the real 690 to 1135 MHz fabricated bandpass filter.

and 11.10, we are not disappointed. The closer view in Figs. 11.11 and 11.12 displays excellent correlation between the softw are simulation and the final, physical filter when both are plotted on the same graph, especially when employing the reliable and accurate Modelithic models. In fact, these outstanding results are due both to proper simulation layout, and to the use of these highly accurate models.

Briefly then, any part of a wireless circuit that is exposed to RF must be modeled with care if simulation accuracy is desired. This entails using the same substrate, PCB thickness, via length and diameter, trace lengths/widths, pad sizes, discontinuities (bends, T-junctions, microstrip steps, microstrip end effects, and the like), along with the proper passive and active component models, within the simulation in order to replicate, as closely as possible, the final layout and parts of the fully assembled PCB.

11.1.5 RF Simulator Issues

S-parameter Simulation

Since an *S*-parameter model of a transistor is acting as a prebiased black box, linear simulation software is incapable of allowing the model to be physically biased with a real power supply. This does not mean that a biasing network should not be added to the *S*-parameter model's input and output just as in a real amplifier. By adding the actual calculated biasing resistors and capacitors the software simulation will take into account the true effects that these biasing components will have on the final amplifier circuit.

Most *S*-parameter files of passive devices are measured by *de-embedding* (i.e., removing all of the parasitic and phase-shift length effects inherent in the part's test

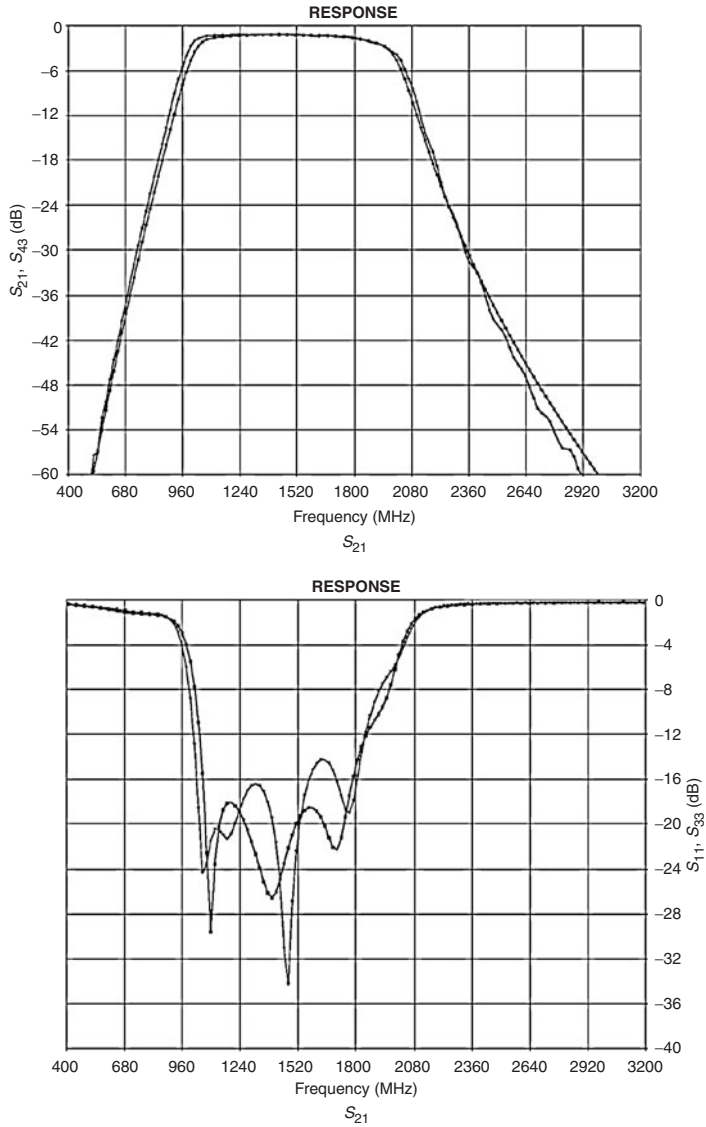


Figure 11.11 Top graph showing S_{21} for Modelithic's simulated filter along with the real 1105 to 1815 MHz fabricated PCB filter, with superb correlation. Below, S_{11} for Modelithic's simulated filter and the real fabricated filter.

fixture) up to the component's package terminations. This is highly advantageous since it also permits the package's length to be integrated into the S_{21} phase. But the component itself may also be measured over an air gap or over some nonconsumer grade (high quality) microwave substrate. Thus, when mounting the real component for our circuit on a particular PCB substrate, the parasitics of the part will be modified if our PCB has a significantly different dielectric constant or board thickness, causing inaccuracies.

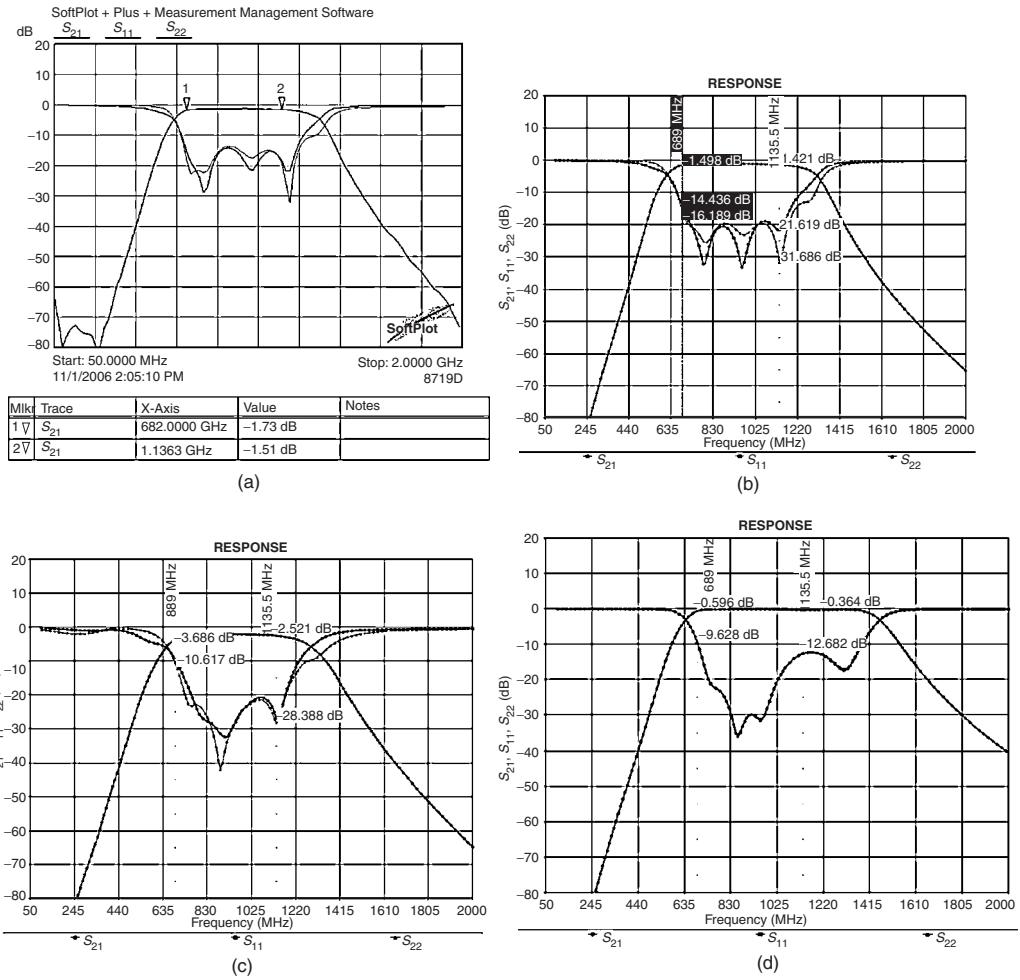


FIGURE 11.12 Gain and return loss comparison of (*a*) 690 to 1135 MHz PCB fabricated BPF; (*b*) BPF circuit using Modelithic models, with excellent correlation to fabricated BPF; (*c*) *S*-parameter models showing excessive insertion loss and less accurate return losses; (*d*) Ideal models (with infinite *Q*) displaying excessively low insertion loss, improper high-end frequency response, and less accurate return losses.
(*Note:* A LP/HP BPF is known as an extremely tolerant filter network to part's variations and parasitics. Other common BPF types would have shown a heavily distorted waveform if they had been simulated with ideal components, while the *S*-parameter model simulation results would have varied with the quality of the S2P model which, in our limited case, were quite excellent.)

This inaccuracy is exacerbated if the component is mounted in shunt, or possesses a different footprint pad size.

Further, some older *S*-parameter file models describing active components may be highly inaccurate. This is directly due to a lack of any substantial number of data points within some of these files. And if the RF simulator itself interpolates or even extrapolates this missing data, it really means that we are still missing the *real* device data for a significant portion of the active component's true performance curve, and we are now only using the

simulation software's own best guess as to where these next data point's true locations are. In fact, some simulators may merely draw a straight line between the two available measured data points, with no interpolation, causing gross inaccuracies if we actually believe that any part of this straight line is a true indication of the part's characteristics. However, the straight line should be relatively easy to spot, while interpolated graphed data, due to its realistic curves, is much more difficult to recognize. Nonetheless, for most simulations with sufficient model-supplied data points, interpolation and extrapolation are very powerful and extremely useful features within any RF simulator.

Never select any active or passive component for your design that does not have, *at least*, linear *S*-parameter models available. Resistors would be an exception to this rule, since models for these passive types are seldom found. In addition, it would be very helpful to also have nonlinear models for all the selected active devices, even when run under small-signal conditions. And obviously, when simulating under large signal conditions, finding nonlinear models is not an option, but a requirement.

A commonly overlooked RF design issue is that all high-value resistors will quickly lose most of their rated resistance at microwave frequencies, while very low-value resistors will actually increase their rated resistance. This decrease in resistance with high-value components is caused by parallel parasitic capacitances permitting the RF to "escape" around the high resistance value; while the increase in resistance of the low-value component is created by the now dominant series parasitic inductance adding its reactance to the mix. And since, as stated above, it can be quite difficult to find *S*-parameter models for most resistors, even those resistors designed for microwave service, component parasitic effects will strongly influence the accuracy of all circuits using these parts in the path of high frequency RF.

As long as the circuit simulation has the appropriate component models that possess enough data points encompassing enough bandwidth, it is essential that very wide-band frequency sweeps be performed (i.e., all the way from the component's lowest plotted frequency data point in its *S*-parameter file to its highest), and not just within the simulated circuit's own limited passband. This is so we may view any out-of-band and undesired gain peaks, any positive S_{11} or S_{22} incursions, and to confirm full amplifier stability across frequency. However it is quite possible, especially with transistors that have a very high f_T, to be unable to simulate up to the highest frequency we may feel should be tested for stability, since the other model files for the passive components may not actually extend that far. Therefore, we will need to physically bench-test the completed amplifier's stability to *at least* one-fifth of the transistor's rated f_T, which is always a very prudent idea whether the simulation is guaranteeing full stability or not. (This would mean inspecting a 70-GHz f_T transistor for complete stability up to at least 14 GHz).

Unless an *S*-parameter S2P model contains noise data, accurate LNA noise figure simulations will not be possible. (Without this noise information in the file the transistor itself will be considered as noiseless.) If the noise data is present, it will be located at the bottom of the S2P file, with a *Comment* line that states "*!NOISE PARAMETERS*". The noise data will then be displayed as "*Fmin*" or "*Fopt*" (which is the part's minimum noise figure); "*GAMMA OPT*" or "*MAGN[opt] PHASE[opt]*" (the magnitude and phase of the optimum source reflection coefficient for a minimum noise figure); and "*Rn*" (the 50-Ω normalized effective noise resistance).

A further issue to assure simulation accuracy is to be aware that with increasing frequency, a PCB's own dielectric constant E_r will drop somewhat. As a result of this, the proper E_r should be accurately specified for the simulation frequencies in use. For

example, at a low frequency FR-4 material may have a value of $E_r = 4.7$, while at microwave frequencies this may drop to $E_r = 4.4$.

Non-Linear Simulations

One of the biggest impediments to nonlinear amplifier design is finding acceptable RF models for the harmonic balance simulator; especially models that are accurate enough at microwave frequencies. Unfortunately, many times all we may have to work with are simple, low-frequency Spice models, which will then require the addition of correct package parasitics (in the form of adding external RLC components) for acceptable accuracy at RF frequencies.

The nonlinear active models available as part of the harmonic balance simulator are either pure Spice models, or what may be called *Spice-type*. These Spice-type models will typically have accurate parasitic capacitances, inductances, and other effects added. The basic Spice model file itself may possess only the data for the active device's die characteristics, and may not actually contain the vital plastic package parasitic information, as stated above. Therefore, only a *MODEL* statement may be present, without any *SUBCKT* statement, within the *.MOD or *.LIB file. This would be almost useless for RF and microwave design unless the parasitic information was added. Nonetheless, as a general rule, most Spice models are not nearly as accurate as *S*-parameter models at frequencies above a few tens of megahertz. Indeed, when combined with gigahertz frequencies and saturated signal operation, such as in PA and active frequency multiplier designs, many Spice models are so grossly inaccurate that they will only supply us with a ballpark indicator as to how our final, completed circuit may actually operate. Therefore, if at all possible, before relying solely on a Spice model for a particular RF or microwave simulation, contact the semiconductor company that supplied the model and ask about its usefulness in your particular wireless application. There are also RF/microwave modeling companies that will create precise transistor or diode linear and nonlinear models specifically targeted for high frequencies. Such custom nonlinear models can be extremely accurate, but are costly in both time and money to create.

The nonlinear *large-signal S-parameters* (LSSP), also referred to as *power dependent S-parameters*, are harmonic balance-based data files that have been measured under true, large signal conditions versus frequency. LSSP files are useful in the simulation of low power (a few watts), nonlinear power amplifiers. The information supplied by an LSSP file are the large signal gain, return loss, input/output power, input/output impedance, and compression point, but without any directly available harmonic output information.

One issue that may be overlooked in harmonic balance software when bias voltages are present is the need to DC block the simulator's own input/output ports. This is required to prevent the simulated circuit's own bias voltages from being shorted to ground through the software's 50-Ω resistive terminations.

EM Simulations

Due to an EM simulator's necessary grid structure within its layout window for all metals, simulation inaccuracies will occur unless EM cell size is set to be extremely small, or enabled with *conformal meshing* (if offered or licensed). To clarify: The typical EM simulation software will presume that a grid cell is either completely filled-in with metal, or completely empty of all metal, based on whether the cell itself is less than 50% filled with metal, which would equal a totally empty cell, or greater than 50% filled, which equals a completely filled cell. If this eccentricity were not fully understood when setting

up or running the EM simulation, then gross inaccuracies can occur if we selected a cell size that was too large. In fact, what may look to be a relatively smooth piece of metal in the EM program's layout window may actually be, in reality, an extremely rough sawtooth or stair-step of metal to the EM software's simulation engine. This will make certain trace sections of the PCB's metal much narrower or much wider than expected.

But because of the simulation time involved, all EM simulations should be streamlined for maximum speed, at least for the initial runs. If not, many such simulations might not only take a very long time to complete (a week or more), but may simply be impossible due to computer memory or time limitations. We can speed up the entire electromagnetic simulation process by using any or all of the following techniques: Use the largest cell and subsection sizes possible (while still retaining an accurate result), never add unnecessary complexity to the metal layers, use the fastest computer with the maximum amount of RAM possible, minimize the number of analysis frequency points, use all lossless metals and dielectrics, minimize circuit/PCB size by decreasing input/output port microstrip lengths, minimize the amount of metal on the PCB (more metal equals more simulation time), adjust the simulator to use single precision instead of double precision, turn off the *current density* calculation function, merge any adjoining metal polygons into a single polygon. Cover height and substrate thickness will have no effect on EM simulation time.

Circuit Synthesis Limitations

At microwave frequencies, certain limitations within all circuit synthesis design tools, such as that used to automatically create filters, oscillators, mixers, and amplifiers, can plague even the most basic high-frequency circuit. When working at low frequencies, the answers provided by common design formulae and synthesis software will match well when compared to a final, realized PCB circuit. But as the frequencies increase into the higher RF regions, these calculations can lose much of their validity. In fact, if a circuit, trace, or component is greater than one-tenth of a wavelength long at the highest frequency of interest, then such calculations will begin to noticeably sacrifice synthesis accuracy. This can be seen in circuits operating at frequencies as low as 25 MHz. Other major influences, such as component parasitic reactances and resistances, as well as stray electromagnetic coupling interactions between other components and/or transmission lines, are also not modeled by any circuit synthesis software, as they so commonly are in linear and nonlinear RF simulators. Thus, we can see that we will have some extremely complex and unpredictable RF effects that can severely degrade circuit performance, resulting in a synthesized circuit that may actually be resonant or matched for operation at the wrong frequency. The common method to solve most of the more significant issues is to place the ideal synthesized circuit design within an RF linear or nonlinear simulator, add the full models and traces, and then tune and optimize for maximum performance.

CHAPTER 12
Wireless Testing

Testing of wireless RF circuits and systems for any new radio design can be a highly complex and involved project, since different types of tests, procedures, and methods can be involved.

We may be required to set up and perform extensive hands-on bench testing of a single new wireless circuit or system design to confirm that it meets our expectations, or we may need to arrange for the efficient functioning of advanced automated test equipment (ATE) for an entire high-volume radio production line. Since ATE is an extremely complex specialty in itself, only the techniques and methods as used on the RF test-bench will be discussed within this chapter.

Typical bench test and measurement equipment will generally comprise spectrum analyzers, vector network analyzers (VNA), frequency generators, function generators, oscilloscopes, variable power supplies, RF power meters, and perhaps bit error rate testers (BERTs) and error vector magnitude (EVM) analyzers. Assorted and essential ancillary RF lab bench tools and devices should also be a part of any test setup, and will include SubMiniature version A (SMA) connectorized coaxial cables, wideband amplifiers, low noise amplifiers, splitters, fixed and variable attenuators, filters, as well as hand tools, soldering stations, desoldering stations, microscopes, RF (high impedance) probes, heat guns, cooling sprays, adjustable stub tuners, assorted LCR components, digital multimeters, and the like.

Common wireless bench testing involves inspecting a radio's RF output power, linearity, sensitivity, stability, bit error rate (BER), NF, phase noise, lock time, gain, harmonics, spurii, dynamic range, EMC compliance, adjacent channel power, bandwidth, current draw, and the like. To be assured of reliable operation over time, temperature, voltage, impedance, shock, humidity, and moisture, the initial testing of any new high-volume radio may involve months of test planning, test setup, and testing of the integrated and discrete circuits and components that are to be used in the final system, in addition to the radio's prototypes and small test-production runs. Some testing should also involve a certain amount of accelerated lifetime/reliability tests, which are performed over a set amount of hours, such as 1000, and is important in order to verify that the wireless device will not fail in the field. (Failure may involve either catastrophic failure—a filter capacitor shorts, or a PA burns up—or subtle performance changes over time, such as decreasing RF output power, sensitivity, efficiency, or changes in frequency or stability).

12.1 Wireless Receiver and Transmitter Tests

12.1.1 Introduction

New tests have been created to confirm proper radio operation while employing digital modulations, such as bit error rate, error vector magnitude, and constellation/eye diagram tests. In comparison to analog-modulated radios, today's radios may need more rigorous testing methods to detect the smaller levels of error-producing phase noise, frequency instabilities, group delay variations, and so on, that are demanded of today's digitally modulated systems. Even basic tests for a radio's modulated output power measurement are not the same as for a narrow-band analog system.

12.1.2 Wireless Receiver Tests

The following section outlines the minimum *receiver* tests and measurements that must be completed to confirm proper operation on any common digitally modulated receiver system. (Not all tests may be relevant for every digital receiver. Most tests are equally applicable to most analog systems as well.):

1. Cascaded gain, dB
2. Gain flatness across channel, dB
3. Frequency accuracy after a specified warm-up period, Hz
4. Frequency stability at three discrete temperatures, Hz
5. Frequency drift over a specified time-period from turn-on to full warm-up, at 25°C, Hz
6. P1dB (1-dB compression point) at output, in dBm
7. Phase noise (PN) of local oscillator (LO) at 10-kHz offset from carrier, dBc/Hz/10 kHz
8. Two-tone intermodulation distortion (IMD) at output, dBc
9. Internally generated in-band spurs with no input signal, dBm and Hz
10. Bandwidth at 3-dB points for a single channel, Hz
11. Minimum discernable signal (MDS), dBm
12. Sensitivity, dBm or uV (at 50 Ω)
13. C/N (carrier to noise) at $P1dB_{IN}$, measured 250 kHz from carrier at IF/BB (intermediate frequency/baseband) output, dBc
14. C/N at *n* miles at IF/BB output, dBc (Adjustable attenuator used to simulate free-space path loss)
15. Signal level at *n* miles at IF/BB output, dBm. (Adjustable attenuator used to simulate free-space path loss)
16. VSWR across channel, in dimensionless ratio.
17. Group delay variation (GDV) across channel, ns.
18. DC current draw, mA
19. Noise figure (NF), dB

20. BER, bit errors per number of bits (such as 1×10^{-9})
21. Interferer rejection at output [continuous wave (CW) directly injected into specified out-of-band channels], dB

12.1.3 Wireless Transmitter Tests

The following outlines the minimum *transmitter* tests and measurements that will have to be completed on a digitally modulated transmitter system to confirm its proper operation, and is similar to the digital receiver tests above. (Not all tests may be relevant for every digital transmitter. Most tests are equally applicable to analog transmitter systems.):

1. P_{OUT} (RF power output), dBm
2. Frequency stability at three discrete temperatures, Hz
3. Frequency accuracy after a specified warm-up period, Hz
4. Frequency drift over a specified time-period from turn-on to full warm-up, at 25°C, Hz
5. Cascaded transmitter (TX) gain, dB
6. Gain flatness across channel, dB
7. PN of LO at 10-kHz offset from carrier, dBc/Hz/10 kHz
8. Two-tone IMD at RF output, dBc
9. Internally generated in-band spurs with no input signal, dBm and Hz
10. 3-dB bandwidth of a single channel, Hz
11. VSWR across channel, in dimensionless ratio
12. GDV across channel, ns
13. LO feedthrough at TX output, dBm
14. DC current draw, mA

12.2 Wireless Test Procedures

12.2.1 Introduction

The following steps outline the major RF circuit and system bench-testing methods used for receivers, transmitters, and individual stages, and are limited to the most critical instructional steps only.

Important note for all following receiver or transmitter procedures:

a. Automatic gain control or automatic level control must be disabled.
b. Set any variable gain stages to maximum gain (unless otherwise indicated).
c. Calibrate out losses of cables, connectors, combiners, and attenuators from final results.

12.2.2 Digital Signal Power Test

A digitally-modulated signal is noise-like, so its power peaks will randomly reach high amplitudes. Therefore, in order to obtain a meaningful power value over this entire digitally modulated wideband RF signal, we must measure its average power over time. The noise-like digitally modulated signal covers a very wide band of frequencies, and we may consider that this digital signal, for clarity, is made up of an enormous number of discrete RF carriers spaced over a wide band of frequencies. Accordingly, if we measured the power over this *entire* wide band of frequencies, and the broadband test instrument itself then summed the result within its internal detector, we would obtain an accurate RF output power level for the entire signal. But, if we then measured a narrower slice of this overall wideband of frequencies by bandpass filtering the signal, the measurement instrument would actually be seeing fewer of these discrete carriers and, because the detector would now have less energy to sum together, we would also see a lower measured output power reading.

We can measure such a broadband signal on the test bench by inserting it into a wideband power meter that was filtered to cover only the actual noise bandwidth of the signal itself. Or, we could measure the power over many smaller frequency segments of the signal's entire bandwidth, and then sum all of these multiple segments together. Due to its limited bandwidth and pulse response characteristics, this is the same method that a spectrum analyzer must exploit, while also automatically applying various internal correction factors, to obtain an average power measurement accuracy of approximately ±1%. Consequently, we could use a spectrum analyzer, which has a limited bandwidth as set by its resolution bandwidth (RBW) IF bandwidth control, to measure the wideband digital signal's amplitude at various points across the signal's complete frequency span, then perform integration of the measured values to supply us with the true average power reading. As we can readily see, to be truly accurate the average power reading of a digital wideband signal must be taken within a specific bandwidth. For instance, a 3-MHz-wide digital signal's average RF power reading on a spectrum analyzer with an RBW set at 300 Hz is measured as −35 dBm. In reality, however, the RF output power is actually 0 dBm, which would have been accurately measured had the RBW setting been correctly adjusted to 3 MHz (so as to completely encompass this signal's full bandwidth).

To summarize the important concept of power in digital signals: As the symbol rate of a digital signal increases, the bandwidth also increases, and the power within the signal is stretched across a very wide frequency range. It is not localized in a narrow band around the center frequency of an RF carrier, as in AM or FM modulation. This means that a single power measurement, taken at a single discrete location of limited bandwidth within the digital channel, will give a deceptively low power output reading. Any accurate method of measuring the wideband output power with typical narrowband test equipment would involve manually taking many discrete average power measurements of the signal over its entire bandwidth, and then summing these points together. However, this is not normally necessary if we possess the appropriate power meter or spectrum analyzer, since CW meters are simply not calibrated nor designed to accurately or easily measure such transient, wideband signals.

Measuring Digital Signal Power

1. To precisely measure a digital signal's power, the bandwidth must first be found. In order to measure the power as accurately as possible, the −30 dB down points will be used instead of the normal −3 dB down points classically adopted to indicate a signal's bandwidth (Fig. 12.1). This −30 dB point will allow for most

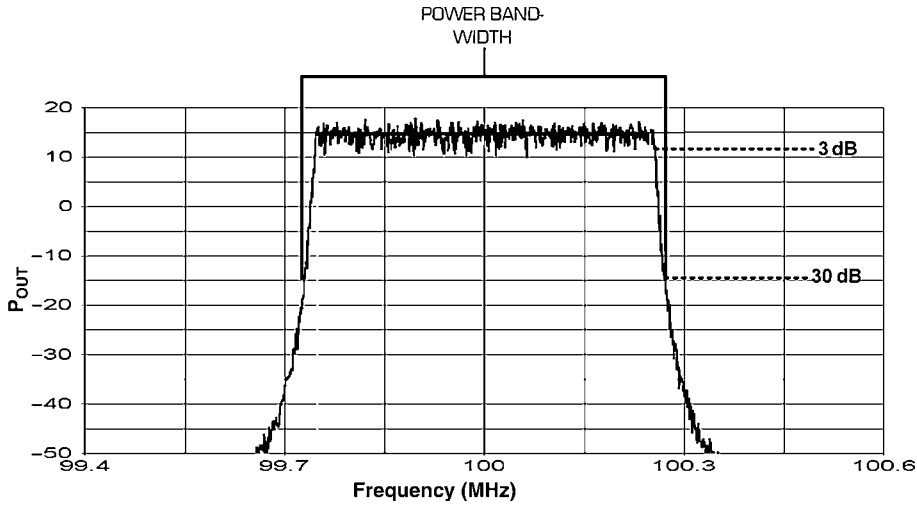


FIGURE 12.1 A digital signal's bandwidth for an average power measurement.

of the digital signal's power to be measured within its entire communication's channel, and any power below the −30 dB points can be discounted.

2. After we have obtained the signal's true power bandwidth, we can go about accurately measuring its average power level. Most quality spectrum analyzers have the capability to automatically make these measurements by
   a. Find the *Power Menu* On your model spectrum analyzer.
   b. Select the channel to be measured to *DIGITAL*.
   c. Input the digital channel's *CENTER FREQUENCY*.
   d. Input the digital channel's 30-dB bandwidth.
   e. The digital signal's true power will be indicated on the spectrum analyzer's display.

   Alternatively, another slightly less accurate, technique is utilized by spectrum analyzers that do not possess the above automatic power measurement capabilities, by

   a. Measure the digital signal's bandwidth at −30 dB down points.
   b. Adjust the analyzer's *RBW* setting to one-twentieth of the signal's −30 dB bandwidth.
   c. Set the analyzer's *SPAN* to 1.25 times the signal's −30dB bandwidth.
   d. Decrease the analyzer's *VBW* (video bandwidth) setting to reduce the signal's displayed noise.
   e. Take the digital signal's power measurement with the spectrum analyzer's frequency/amplitude *MARKER* placed at the center of the signal.
   f. Find the true total average power P_{AVG} of the digital signal by taking the power, in dBm as measured in Step e and adding in a *bandwidth correction factor* (BWCF):

$$P_{AVG}(\text{dBm}) = \text{Measured power (dBm from Step e)} + \text{BWCF}$$

where $\text{BWCF} = 10\text{LOG}\left(\dfrac{\text{SIG BW}_{30\text{ dB}}}{\text{RBW}}\right)$.

g. For a more accurate digital signal power measurement we can add another correction factor to Step f that takes into account the spectrum analyzer's internal RBW and log-detection stage losses of approximately 2 dB. Therefore, the approximate formula for the measurement of the digital average power now becomes

$$P_{AVG}\,(\text{dBm}) = \text{Measured power (dBm from Step e)} + \text{BWCF} + 2\ \text{dB}.$$

12.2.3 Constellation and Eye Diagram Test

To view the degradation created by noise and frequency instabilities in a digitally modulated signal, as well as other impairments, we can employ constellation [*Fig. 12.2*(*a*)] and eye diagrams [Fig. 12.2(*b*)]. There are also other test and measurement equipments available that are quite capable of quickly and effortlessly taking constellation and eyes diagrams, rather than using the older (but much lower cost) oscilloscope method as outlined below.

To view constellation or eye diagrams requires the ability to tap into the digital receiver demodulator's I and Q outputs, as well as the demodulator's timing clock (Fig. 12.3). The outputs of the demodulator's I and Q will then be fed into an oscilloscope with an *X*-*Y* display that has some type of analog or digital persistence function (to view the I/Q outputs over time). Such a setup will let the operator confirm the phase and amplitude differences of the output signal in the form of a constellation diagram, allowing the inspection of the digital signal's distortion, phase noise, or amplitude instabilities. An excellent constellation diagram with no major impairments is shown in Fig. 12.4(*a*).

While constellation diagrams will display the digital modulation's symbol patterns, eye diagrams will actually permit the symbol's *transitions* to be viewed over time. Yet both methods will measure the baseband signal's modulation condition, and whether impairments are truly degrading this expected pattern.

When viewing eye diagrams, the eye will be rounded, instead of square. This is due to the necessary limiting of the bandwidth by the baseband filters. The eye comprises two lines, one at a digital 1 and the other at a digital 0, and represents a series of pulses

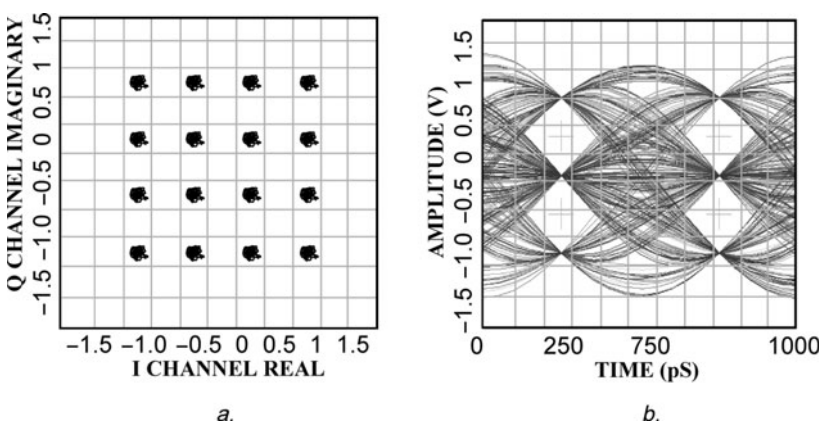


Figure 12.2 (*a*) A constellation diagram; (*b*) an eye diagram.

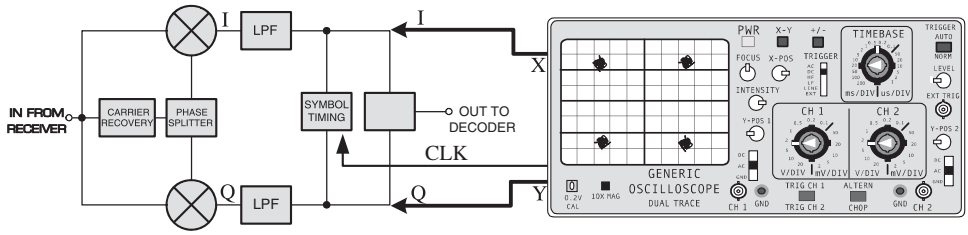


FIGURE 12.3 Test setup for constellation and eye diagrams.

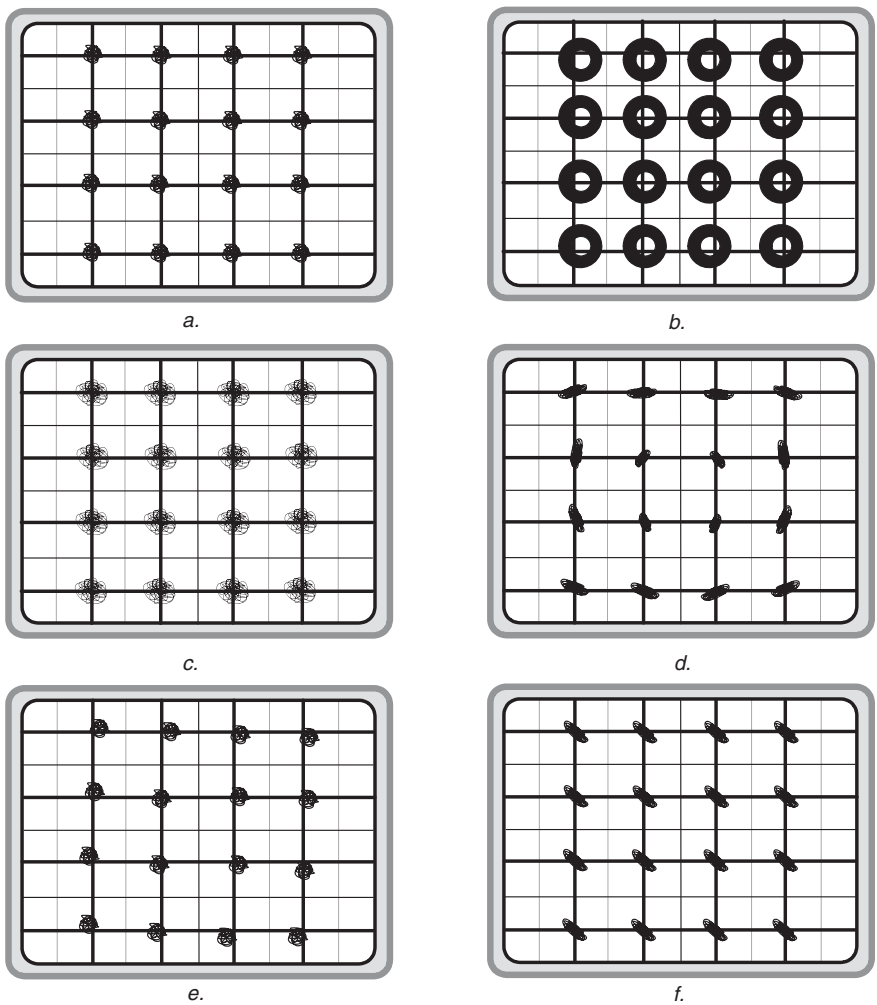


FIGURE 12.4 (*a*) An almost perfect 16-QAM constellation diagram; (*b*) a CW signal interference tone effect on the constellation diagram; (*c*) signal degradation caused by a poor SNR; (*d*) signal degradation caused by the digital radio's local oscillator instability; (*e*) slight overdriving of the transmitter's power amplifier; (*f*) multipath causing an uneven amplitude across the passband of the digital signal.

as displayed on the screen of the oscilloscope. Each of these pulses from the demodulator will also exhibit the noise and jitter as added by not only the transmitter, but also by the wireless signal path and the receiver itself. This makes each pulse slightly different from the last and, as the received signal is degraded by these impairments, the eye will begin to close, and the BER will increase.

Measuring a Constellation Diagram

1. Attach the I to the X input of the scope, and the Q to the Y input of the scope (normally, this will be Channel 1 and Channel 2), and turn on the scope's X-Y mode and the persistence function.
2. Attach the symbol clock to the external trigger (turn on *EXT TRIGGER*).
3. The constellation itself should now be visible, along with faint lines joining the various points of the constellation. (The lines are the actual symbol transitions between the constellation points, called *symbol trajectories*.)
4. To obtain only the constellation points, without the symbol transition lines, a BNC connector on the back plate of some oscilloscopes can be connected to the receiver's demodulator symbol clock to force the oscilloscope's electron beam to turn on only at the exact moment of sampling. This will blank the transitions between constellation points.
5. Common impairments, and how they look in a constellation diagram, are shown in Fig. 12.4.

To Measure an Eye Diagram

1. Maintain the same setup as above, but turn off the scope's X-Y function.
2. Set the scope's *HORIZONTAL TIMEBASE* to obtain 3 symbol periods per 10 divisions.
3. View the eye diagram. The height of the eye is the *noise margin* of the receiver's output, while the left and right corner of each eye indicating the amount of frequency jitter present. Therefore, the eye itself should always be open: the wider the eye, the less jitter, while the taller the eye, the less noise.

12.2.4 Bit Error Rate Test

The bit error rate test measures the ratio of bad bits to all the transmitted bits, often over a range of received input and output powers, for a digital communications system. When the received power is low, BER rate increases due to a degraded signal-to-noise ratio (SNR); when high, amplifier compression is the major contributor to a poor BER.

BER testing is an excellent way to check the signal quality of a digital radio system, a lone receiver, or the entire link. However, any in-line error correction and/or adaptive equalization will make the BER appear far better than it actually would be without these two processes. Thus, the normal indication of how close a system is to a complete digital signal outage may be hidden by the error correction and equalization circuits, which is why many receiver BER tests are best performed before any such correction and equalization stages (to obtain the *raw BER*).

In the field, it is still invaluable to perform final BER tests from one end of a complete communications system to the other, with correction and equalization engaged, to confirm that the entire link meets BER specs over a set time period, and that the total

system is functioning as designed. Indeed, to prevent a digital wireless link from failing periodically, it must be completely tested to check that there is sufficient link budget to overcome any impairment in the air interface between the transmitter and receiver (see Sec. 9.4, "RF Propagation"). Still, a generally poor BER in a complete end-to-end system measurement will not indicate exactly where the trouble lies. Finding the weak link in a complete end-to-end communications system is actually part of *error distribution*, and could be caused by poor antenna alignment, overdriven amplifiers, low signal output, path obstructions, high NF, multipath, cable losses, frequency drift, component malfunction, and so on. The test equipment we can utilize to assist us with tracking down BER problems are a *vector signal analyzer* (VSA), a spectrum analyzer, and standard lab equipment, while using typical system troubleshooting techniques.

The radio design phase, and not after, is the time to confirm that the BER will meet the expectations for our particular communications system. To preserve the desired path length and quality of service at our preferred BER, the only way to cure a radio's deficiencies *after* it is built and tested is through system modifications to the transmit power, antenna gain, and/or receiver NF.

Measuring Bit Error Rate

There are many different models of BERTs available, as well as test setup procedures, so the best course is to follow the operational instructions that came with the particular BER tester. Generally, the main type of BER test that a circuit designer will be involved in is for the sensitivity of a digitally modulated receiver. To perform this kind of BER test, a pseudorandom bit sequence, which is modulated through a special signal generator and onto an RF carrier, with this carrier set to the exact same frequency as the receiver, is directly injected into the receiver's RF input port through a coaxial cable of known insertion loss. The RF input level is either controlled through an external adjustable connectorized attenuator, or through the signal generator itself. The receiver's output baseband data is sent on to the BERT, which continuously calculates the errors between the originally sent data, and the currently received data. The amplitude of the RF input signal is then adjusted downward until the BERT's bit error rate display duplicates our required BER. The level of the RF signal's power (dBm) or voltage (μV) that is *directly* at the receiver's RF input port is then measured or calculated. This, then, is the sensitivity of the digital receiver at its minimum RF input signal amplitude and at the rated maximum bit error rate.

12.2.5 Phase Noise Test

Follow the test procedures as outlined below to perform phase noise measurements on any oscillator. These tests are valid only if the spectrum analyzer has a lower phase noise than the device under test (DUT) itself.

To Measure Phase Noise

Method One:

1. Set up phase noise bench test as shown in Fig. 12.5. It is critical that the power supply is a low-noise unit, since most general lab power supplies are quite noisy. Or, preferably, use a battery. The DUT's DC power cables should also be as short as possible to avoid stray noise pickup.
2. Set the spectrum analyzer to the desired center frequency, with a span of 10 MHz.
3. Activate the spectrum analyzer's phase noise measurement module or PN utility, and set to *MINIMUM FREQUENCY OFFSET* = 10 kHz and *MAXIMUM FREQUENCY OFFSET* = 10 MHz.

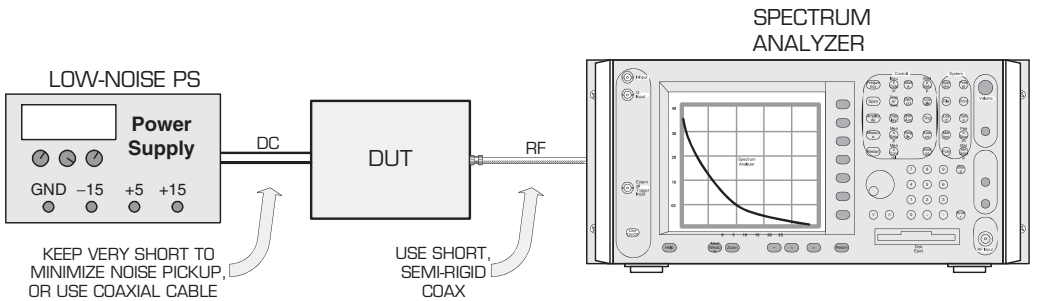


FIGURE 12.5 A simple phase-noise test setup using a spectrum analyzer.

4. Set DUT to desired center frequency (must be the same as specified in Step 2 above).
5. Activate the spectrum analyzer's *Phase Noise Utility* to begin the actual measurement.
6. With the spectrum analyzer's noise marker, measure the signal to obtain PN power density at the desired carrier offset frequencies.

Alternatively,

Method Two: Used to measure a locked PLL's phase noise

1. Set up DUT and test as shown in Fig. 12.5.
2. Set DUT to desired frequency.
3. Set the spectrum analyzer's *SIGNAL TRACK* to On; set *FREQUENCY SPAN* for 10 times the desired measurement frequency (i.e., if we want to measure phase noise at a 10-kHz offset, then set spectrum analyzer's *FREQUENCY SPAN* to 100 kHz.); set (decrease) the *VIDEO BANDWIDTH* for the clearest trace views on the display.
4. Use the dBm/Hz noise marker at the exact phase noise offset frequency of interest, or use the noise marker's delta function, placing the marker first at the CW carrier's peak and then at the particular phase noise frequency of interest to obtain dBc/Hz. A third way is to use the following formula and the normal dBm amplitude delta marker, and then calculate the 1-Hz normalized phase noise level:

$$\text{PN (dBm/Hz)} = (-\Delta_{\text{C-to-PN}}) - [10 \log (\text{RBW})]$$

where PN = Phase noise at a specific frequency offset from the carrier, normalized to 1 Hz, dBc/Hz*

$-\Delta_{\text{C-to-PN}}$ = Delta measurement of phase noise† below the carrier, dBc

RBW = The spectrum analyzer's resolution bandwidth setting, Hz.

5. Further, if we use a normal amplitude marker on the spectrum analyzer, at some non-1Hz RBW, we can also convert this into the industry standard for noise figure measurements, which are all normalized to take place within a

*dBc/Hz refers to the amplitude offset from the carrier to the phase noise point located at an explicit offset frequency, and specified in dBc within a 1-Hz bandwidth).

†Since the phase noise is *below* the carrier, a negative sign must be used while calculating the formula.

1-Hz bandwidth. (A 1-Hz bandwidth is commonly used to remove any possible ambiguities as to what spectrum analyzer RBW setting was employed during the noise measurement, since this would drastically alter the measurement amplitude results):

$$\text{dBm/Hz} = \text{PN}_{\text{dBm}} - 10 \log (\text{RBW})$$

where dBm/Hz = Noise power in a 1-Hz bandwidth, dBm/Hz
PN_{dBm} = Power of the phase noise at a specific offset frequency from the carrier, dBm
RBW = The spectrum analyzer's resolution bandwidth setting, Hz.

NOTE: *Accurately measuring the noise level of a* crystal *oscillator directly with typical RF test equipment is not possible, considering that the oscillator's noise levels should be far below that of the spectrum analyzer's own noise floor.*

12.2.6 Noise Figure Test

There are three popular methods to measure a communication system's noise figure. The first is the use of a specialized *noise figure meter*. This is an expensive piece of hardware and will have a limited upper frequency measurement capability (typically less than 3 GHz). The second technique is called the *gain method*, and has severe limitations in that we cannot measure most of the low-noise figures found in modern microwave devices and systems. For any real NF accuracy, this test also requires the DUT to have a very high gain. The third technique is called the *Y-factor method*, and is very accurate at measuring low-noise figures at high frequencies without any required high DUT gain limitations. The only specialized piece of necessary equipment for this test is an *excess noise ratio* (*ENR*) *noise head*, which is a calibrated ultrawideband noise generator.

Measuring Noise Figure

1. Setup the equipment as shown in Fig. 12.6.
2. Set the spectrum analyzer's marker function to measure dBm/Hz; calculate and set the spectrum analyzer's RBW *versus* VBW ratio to be 0.3, or RBW/VBW = 0.3; set the spectrum analyzer's input attenuation to 0 dB.
3. Turn on the ENR noise head. Take a reading on the spectrum analyzer. Now turn off the noise source and take another reading. This is the difference in the receiver's output noise power density as measured on the spectrum analyzer's noise marker function in dBm/Hz. We can obtain the true system or device's noise figure by using the following formula:

$$\text{NF} = 10 \log [10^{(\text{ENR}/10)}]/[10^{(Y/10)} - 1]$$

where ENR* = calibrated noise level of the wideband noise head at specific frequencies, dB.
Y = The difference in dBm/Hz between an on and an off noise source.

---

*This ENR noise figure for the specific frequency of interest can be found inscribed on the side of the ENR noise head itself.

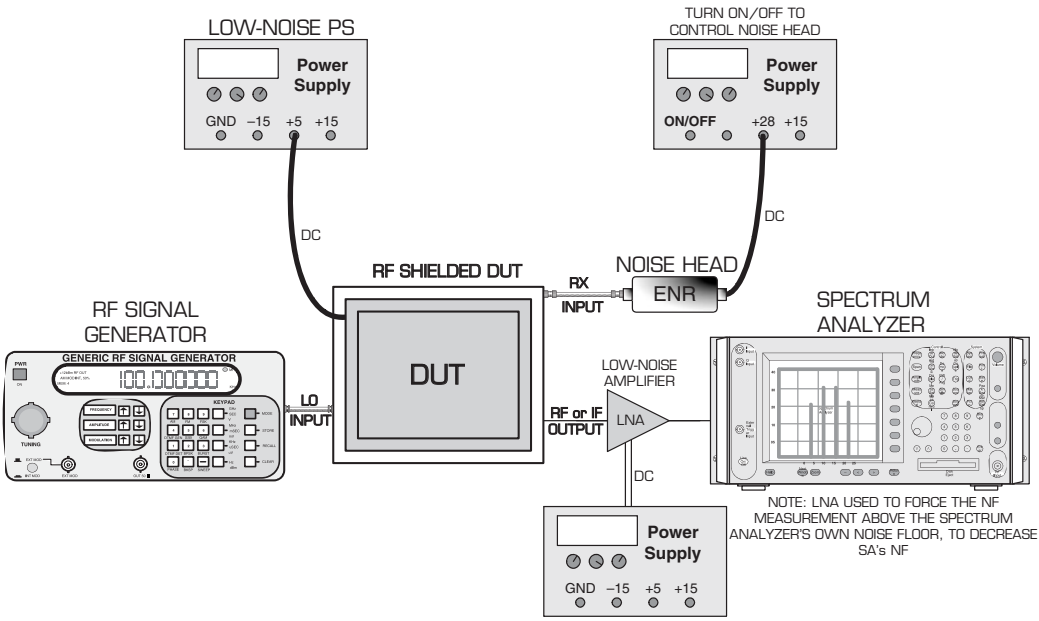


FIGURE 12.6 Common receiver or amplifier noise figure (NF) test setup.

12.2.7 Reference Spur Test

PLL reference spurs can increase the BER of a wireless device, and can be inspected at the phase locked loop's VCO output port.

To Perform Reference Spur Test

1. Attach the PLL's VCO output to a spectrum analyzer's input.
2. Set the PLL and the spectrum analyzer to the same center frequency.
3. Open the spectrum analyzer's *FREQUENCY SPAN* setting to allow viewing of all reference spurs. (Reference spurs will be located both above and below the PLL's output frequency by an offset equal to the PLL's comparison frequency of f_{COMP}, as well as occasionally at its harmonics).
4. Take the output amplitude of the PLL's center frequency and the output level of the reference spurs in dBm, and subtract the two. This will be the level of the reference spurs below the carrier, in dBc.

12.2.8 Blocking/Desensing Test

A receiver must be tested for out-of-band signal rejection, since strong off-frequency interferers can block or "desense" the receiver's LNA if the receiver's front-end filter is not sufficiently selective. (Blocking is one form of desensitization.) This creates large BER degradation problems and the overdriving of the LNA stage, generating IMD or mixing products in band.

There are two general methods to perform this test. Method two is frequently referred to as BDR (*blocking dynamic range*).

Performing Desensing and BDR Test

Method one

1. Carry out test by combining two RF signal generator outputs into a *hybrid combiner* of the appropriate frequency.
2. Set one generator to be the *desired* signal source by setting it to a center in band frequency at −80 dBm. Set the second signal generator to a frequency at either the lower- or upper-band edge, and it will function as an *undesired* out-of-band interferer at an amplitude of −20 dBm.
3. By subtracting the desired signal level located at the receiver's IF output port by the RF input signal level, in dB, confirm that the gain of the desired signal has not decreased below specifications due to amplifier desensing.

Performing a Blocking Dynamic Range Test

Method two

1. Place one 3-dB attenuator at each of two signal generator's output ports; attach these attenuator outputs to the input of a hybrid combiner; place another 3-dB attenuator at the RF hybrid combiner's summed output port; attach this combiner's attenuator output to the receiver's RF input port; attach a SINAD meter (or BERT) to the receiver's baseband output.
2. Set *desired* signal generator to 3 dB above the receiver's minimum 12-dB SINAD sensitivity at the desired channel frequency (as found in the ACR or SINAD tests above).
3. Set the *adjacent* channel signal generator to the center of the adjacent channel frequency.
4. Increase the adjacent channel signal generator until the SINAD reaches its rated dBm level as taken in the above 12-dB SINAD step (i.e., at the receiver's SINAD when no adjacent signal is present, but when the desired signal has not been increased 3 dB in amplitude); *or* the desired signal at the receiver's IF output decreases by 1 dB.
5. Record the amplitude and frequency of the adjacent RF input signal, and the desired RF input signal frequency that are all currently at the receiver's RF port (corrected for cable, attenuator, and so on losses), in dBm and Hz.
6. Calculate the amplitude of the receiver's noise floor:

$$\text{Noise floor (dBm)} = -174 \text{ dBm} + 10 \log_{10}(\text{BW}) + \text{NF}$$

   where BW = noise bandwidth of the receiver, or approximately the 6-dB down bandwidth (instead of the typical 3-dB bandwidth), Hz. NF = receiver's noise figure, dB.

7. Take the difference in the frequency between the *desired* frequency and the *adjacent* channel frequency as recorded in Step 5.
8. Subtract the amplitude of the *adjacent* frequency as measured in Step 5 from the amplitude of the receiver's noise floor as calculated in Step 6. This is the BDR value, and is combined with Step 7 and written as dB@kHz.

12.2.9 Gain and Flatness Test

Maintaining the system's gain and gain flatness is especially critical for dense digital modulation formats.

Performing Gain and Flatness Test

1. Attach a signal generator to the receiver's RF input, and a spectrum analyzer to the receiver's IF output.
2. Set signal generator's output power to –50 dBm (or to some other linear small-signal value), and set spectrum analyzer to sweep the complete IF frequency bandwidth on *MAX HOLD*.
3. Sweep the signal generator to cover the complete IF bandwidth of the receiver.
4. Run this gain and gain flatness test at low, medium, and high-channel LO settings.
5. View and log the amplitude flatness and gain values.

12.2.10 Transmitter Output Power Flatness Test

Performing Output Power Flatness Test

1. Place a sufficient attenuator pad value between the transmitter and spectrum analyzer so as to decrease the RF power level into the spectrum analyzer, thus decreasing the generation of spectrum analyzer–generated IMD products. (Use sufficient attenuator pad to decrease RF input power to a safe number for spectrum analyzer's front end, and with a pad that can operate at the particular dissipation levels.)
2. Set spectrum analyzer on a high input attenuation level (and decrease input attenuation level, as appropriate, after test begins). Set *MAX HOLD* On.
3. Sweep the output frequency using the transmitter's CW or modulated signal capabilities, or employ an external swept frequency generator source into the transmitter's IF input and placing the transmitter's LO at different channels.
4. Take measurements at the transmitter's low, medium, and high frequency band synthesizer settings.

12.2.11 SINAD Sensitivity Test

Performing a FM Receiver Sensitivity Test

1. Set receiver under test to mid band.
2. Hookup a signal generator to the receiver's RF input, and a SINAD meter to its output port.
3. Set signal generator for a 1-kHz tone with 3-kHz deviation at mid band frequency of receiver.
4. Increase the signal generator's RF output level until the SINAD meter display reads 12 dB.

5. Note the signal generator's RF signal level, corrected for any test cable, attenuator, or filter losses. This reading is the 12-dB SINAD sensitivity, in μV or dBm.

12.2.12 Adjacent Channel Rejection Test

Performing a Receiver Adjacent Channel Rejection (ACR) Test

1. Hookup two RF signal generators, one representing the *desired* signal and the other an *adjacent* channel signal, to a hybrid signal combiners. Place combiner's summed output port into the receiver's RF input, and a SINAD meter at its output. Place one 3-dB attenuator at each signal generator's output port, and at the signal combiner's output port.
2. Set the desired signal generator for a 1-kHz tone and 3-kHz deviation mid band.
3. Set receiver to mid band, on same frequency as Step 2.
4. Perform a normal 12-dB SINAD receiver test, with only the desired signal generator active: Increase this signal generator's RF output level until the SINAD display reads exactly 12 dB. Note the signal level (corrected for cable, attenuator, and so on, losses), which is the 12-dB SINAD sensitivity in μV or dBm.
5. Increase the desired RF signal generator's output level by 3 dB.
6. Set the adjacent RF channel signal generator at some low output amplitude and to the center of its own designated adjacent channel frequency.
7. Now, increase the adjacent channel RF signal generator's amplitude until SINAD reaches its exact rated μV or dBm level, as taken in the above 12-dB SINAD Step 4 measurement (i.e., at the receiver's own SINAD as measured when no adjacent signal was present, but when the desired signal had not been increased 3 dB in amplitude).
8. Record both the amplitude and frequency of the adjacent RF input signal and the desired RF input signal that are currently at the receiver's RF port (corrected for cable, attenuator, and so on, losses), in μV or dBm and Hz.
9. Take the difference in amplitude and frequency between the desired frequency and the adjacent channel frequency as recorded in Step 8. This is the ACR value, and is written as dB@kHz.

12.2.13 P1dB Compression Test

Performing a *P*1dB Test

1. Place a lowpass filter at a signal generator's output (to minimize the sometimes substantial generator-produced RF harmonics). Attach the lowpass filter with signal generator to the input of the amplifier under test. Set the RF signal generator's CW frequency to the center of the amplifier's passband.
2. Place a spectrum analyzer or power meter at the output port of the amplifier stage. Do not overload this test equipment with what may become a high-powered RF signal. If the DUT, filter, or spectrum analyzer is not DC blocked, insert DC block to prevent DC from entering the test equipment.

3. Insert a small-amplitude RF input signal into the receiver that is well below the amplifier's expected *P*1dB. Calculate the small signal stage gain.
4. Increase the signal generator's RF output power until the amplifier's stage gain decreases by 1 dB below its small signal stage gain as measured in Step 3.
5. The amplifier's *input P*1dB will be the *signal generator's* output power level, in dBm (corrected for any cable and filter losses). The amplifier's *output P*1dB will be the amplitude as measured on the *spectrum analyzer*, in dBm, corrected for cable losses.

12.2.14 Third-Order Intercept Point (IP3) Test

Performing a TOIP (IP3, Third-Order Intercept Point) Test

1. Place a 3-dB attenuator at each of two RF signal generator outputs; attach these 3-dB attenuator outputs to the input of a hybrid combiner; place a 3-dB attenuator and a bandpass or lowpass filter on the hybrid combiner's summed output port; place the output of the attenuator/filter into the receiver's or amplifier's (DUT) RF input port; place a spectrum analyzer at the DUT's output port. (Filter added to decrease the signal generator's output harmonics).
2. Turn off both of the signal generator's automatic level control (ALC) circuits (it will be a RF signal generator menu option).
3. Set *Signal Generator #1* to the center of the band, then set *Signal Generator #2* 100-kHz higher in frequency.
4. Taking into account the losses of the inline attenuators, the coax cables, the hybrid combiner, and the filter, set both *Signal Generator #1 and #2* to output into the receiver's input port a CW signal that is 20-dB below the DUT's *P*1dB. Confirm that both RF output tones are symmetrical in amplitude.
5. On the spectrum analyzer, measure from the amplitude peak of the two-tone signal to the peak of the third-order distortion tones, and subtract one from the other, as dBc. Record the solution as *IPdBc*.
6. Record down the peak amplitude of one of the two-tone DUT outputs as *P*out, in dBm.
7. Calculate the TOIP/IP3 from the answers found in Step 5 and 6:

$$\text{TOIP} = P_{\text{OUT}} + |\text{IP}_{\text{dBc}}/2|.$$

12.2.15 Spurious Free Dynamic Range Test

There are multiple techniques to measure spurious free dynamic range (SFDR). Presented here are three such methods.

Performing a SFDR Receiver Test

1. Place a 3-dB attenuator at each of two RF signal generator's outputs; attach these two 3-dB attenuator outputs to the input of a hybrid combiner; place a 3-dB attenuator and a lowpass filter(LPF) or bandpass filter (BPF) on the hybrid combiner's summed output port; place the output of the filter into the receiver's RF input port; place a spectrum analyzer at the receiver's output port with an RBW setting equal to the receiver's IF, or to some other specified measurement bandwidth.

NOTE: *To minimize the signal generator's harmonic outputs, superior stopband attenuation is demanded of the LPF or BPF filter since the SFDR test must use an almost perfect sine-wave into the receiver's input, without discernable harmonics, for sufficient test accuracy.*

2. Turn off the ALC of the signal generators, and set *Signal Generator #1* to center of band, then set *Signal Generator #2* to 100 kHz higher.
3. Set RF *Signal Generator #1 and #2* to output two equal amplitude, low power CW signals into the receiver's input port.
4. Increase/decrease *Signal Generator #1 and #2's* output power equally until the IMD spurs are at the same level as the receiver's noise floor, as viewed on the spectrum analyzer.
5. Now, measure the difference in amplitude between the receiver's output fundamental frequency *versus* the receiver's output noise floor amplitude. This is the SFDR of the receiver, in dB.

We can use the formula below to calculate receiver SFDR from known parameters:

1. $SFDR = 2/3(IIP3 - RX_{N-F})$

   where IIP3 = RF input power at the third-order intercept point, dBm
   RX_{N-F} = amplitude of the receiver's noise floor, *dBm*

   or

   $$-174 \text{ dBm} + 10 \log_{10}(BW) + NF$$

   where BW = noise bandwidth of the receiver, Hz; NF = receiver's noise figure, dB.

Measuring a Transmitter's SFDR

1. Set up a spectrum analyzer, with the appropriate high-power pad attenuator attached to its input (so as not to damage the spectrum analyzer's front end).
2. Connect the spectrum analyzer, with pad, to the output port of the low-power transmitter DUT.
3. Making sure that the spectrum analyzer will not be overdriven by the transmitter (to avoid creating its own internally generated spurs and harmonics), and turn on the transmitter, setting it to its standard RF power level to output a single, CW test signal.
4. Read the difference in amplitude between the fundamental and the second or third harmonic, whichever is greater, in dB.

12.2.16 Image Rejection Test

Performing an Image Rejection Test for a Receiver

1. Place a 3-dB attenuator at each of two RF signal generator's outputs; attach these two 3-dB attenuator outputs to the input of a hybrid combiner; place a 3-dB attenuator and a quality LPF or BPF filter on the hybrid combiner's summed output port; place the output of the filter into the receiver's RF input port; place a spectrum analyzer at the receiver's output port. (Good stopband

attenuation for the LPF or BPF are demanded in order to minimize the signal generator's harmonic outputs.)

2. Turn off the ALC of both signal generators; set *Signal Generator #1* to center of band, set *Signal Generator* #2 to the image frequency; set both generators to inject exactly −30 dBm into the receiver's RF input port (compensated for attenuator, combiner, and filter losses); and that both generator tones are equal in amplitude.
3. Turning one signal generator off, then the other, measure both tone levels on the spectrum analyzer at the receiver's IF output port.
4. The difference between these two measurements is the image rejection, in dB.

12.2.17 Frequency Stability Test

Performing a Receiver Frequency Stability Test

1. Place a signal generator at the RF input to the receiver, and a spectrum analyzer at the receiver's output port.
2. Set the signal generator's CW frequency to the center of the receiver's bandpass and the output level to around −30 dBm or less.
3. Set a very narrow frequency span on the spectrum analyzer.
4. Examine display for any frequency offsets or frequency instability over time.

12.2.18 Minimum Discernable Signal Test

Performing a Receiver Minimum Discernable Signal (MDS) Test

1. Place a signal generator at the RF input to the receiver, and a spectrum analyzer at the receiver's IF output port.
2. Set the signal generator's CW output frequency to the center of the receiver's bandpass, and the RF output amplitude to −130 dBm.
3. Adjust the spectrum analyzer for the lowest input attenuation and narrowest RBW and VBW feasible (within sweep speed time limitations). These should all be set to a value that clarifies the measured IF CW signal above the noise floor.
4. Increase the signal generator's power output from the −130 dBm until the spectrum analyzer's display shows a IF CW signal that is either at 0 dB or at 3 dB (depending on specifications) above the noise floor.
5. Now, measure the amplitude of the RF signal directly at the receiver's input port. This is the MDS, in dBm.

12.2.19 NxM Spur Test

Testing for all Self-Generated In-Band Receiver Spurs

1. Place a signal generator at the receiver's RF input port, and a spectrum analyzer at its IF output port.

2. Set the signal generator's CW frequency to the center of the receiver's bandpass, and at an amplitude of −30 dBm (as measured directly at the RX's input).
3. Vary the receiver's LO frequency from low to high.
4. Examine and log frequency and amplitude of all in-band IF spurs as shown on the spectrum analyzer display.

12.2.20 Phase-Locked Loop Response Test

Performing a PLL Loop Filter Frequency Response Bandwidth Test

1. Using a spectrum analyzer with a built-in tracking generator, connect tracking generator section to an RF signal generator's modulation input; connect the RF signal generator's RF output to the PLL's reference frequency input (this connection takes the place of the PLL's crystal reference oscillator).
2. Connect a high-impedance active FET probe to the input of the spectrum analyzer section, with the other end of this FET probe connected directly to the V_{TUNE} point (located between the VCO and the output of the PLL loop filter).
3. Program the PLL synthesizer chip to mid-band through the radio's own frequency control adjustment, or by using appropriate interface cable and programming software (available from the PLL chip's manufacturer).
4. Set the spectrum analyzer to *START FREQUENCY* = 500 Hz, *STOP FREQUENCY* = 25 kHz.
5. Set the RF signal generator to the exact PLL's reference frequency, at an amplitude of around +3 dBm, and to *EXTERNAL* FM analog modulation, with *DEVIATION* = *WIDEBAND* (or to some appropriate deviation value to obtain a V_{TUNE} trace on spectrum analyzer screen).
6. Viewing the spectrum analyzer, inspect and capture the resultant trace of the V_{TUNE} voltage output.
7. Repeat for the PLL's low, middle, and high frequencies.

12.2.21 VCO Pushing Test

To Perform a VCO Voltage Supply Pushing Test

1. Place a spectrum analyzer at the output of the VCO, an adjustable power supply at the VCO's DC power input, and another adjustable power supply at the VCO's V_{TUNE} port.
2. Set the VCO's DC supply voltage to its recommended nominal value, and measure the VCO's frequency at three different V_{TUNE} voltages (pushing can vary with tune voltage).
3. Increase the VCO's DC supply voltage by 1 V from its recommended nominal value, and measure the VCO output frequency at three V_{TUNE} voltages.

4. Decrease the VCO's DC supply voltage by 1 V from its recommended nominal value, and measure the VCO output frequency at three V_{TUNE} voltages.
5. Record the results for each V_{TUNE} value as a change in the VCO's output frequency per 1 V change in the DC supply, in MHz/volt.

12.2.22 Radiated Output Power Test

While we can easily test the conducted RF output power P_{OUT} of a transmitter's fundamental frequency (and its harmonics) by directly connecting a coaxial cable to a spectrum analyzer, it is a different story entirely when an antenna is already attached to the transmitter, and we are tasked with discovering the transmitter/antenna combination's radiated power. The only way to accomplish this assignment is by measuring the radiated field strength of the antenna's RF output at some specific far-field distance, and then calculating backward to discover the total transmitted output power from the transmitter/antenna. Most current wireless systems operate in the microwave region, and 3 m is normally considered as the standard measurement distance for the far-field. Lower frequency testing will require a much longer distance to reliably reach this far-field region.

The following test can only be expected to realistically supply us with an approximate number for its radiated output power from the system, as it must assume zero multipath phase cancellation effects. However, without specialized test facilities, such as an anechoic chamber, this would be impossible. Nevertheless, the multipath causes and effects can be minimized to a large extent by following these steps.

Performing a Transmitter Radiated Output Power Test

1. Select an outdoor test location with minimal multipath issues, such as an open field devoid of trees, buildings, automobiles, light poles, people, and so on.
2. Place the transmitter/antenna DUT combination on a nonconductive table, with the DUT's actual RF-radiating antenna element located at a height of at least 1 m above the ground and exactly 3 m away from the calibrated *field strength reference mark* of the calibrated EMI test antenna, with this antenna mounted on a nonconductive tripod.

NOTE: *The field strength reference mark is normally located at the center of the calibrated antenna. The antenna itself will generally be of a log-periodic type.*

3. Bore-site the calibrated EMI antenna directly at the DUT transmitter's antenna element. During testing we will have to move and/or reorient the DUT's antenna element until the maximum signal strength is obtained, since it is not a perfect omnidirectional radiator.
4. Set the spectrum analyzer to the transmitter's frequency, the RBW setting to twice the transmitter's own bandwidth, and switch the analyzer's vertical amplitude to dBuV.
5. Acquire the spectrum analyzer measurement, which must be corrected for the EMI antenna's own calibrated *antenna* (correction) *factors* and the coaxial cable losses, and then calculate the total ERP-transmitted RF output power (which

includes the DUT's PA output power and the DUT's antenna gain) by applying the formula shown:

$$P_{TX}(\text{dBm}) = (V_{RF} + \text{AF} + \text{CBL}_{LOSS}) - 95.23$$

where P_{TX} = total approximate transmitted output power (ERP) from the transmitter system, dBm
V_{RF} = RF voltage as measured at 3 m from DUT by spectrum analyzer, dBuv
AF = EMI antenna's specified *antenna factor* at frequency of interest, dB/m
CBL_{LOSS} = coaxial cable losses between the EMI antenna and the spectrum analyzer, dB.

12.2.23 General Precompliance Test

The most important document for a wireless system design engineer in the United States is the Federal Communications Commission (FCC) Part 15 regulations, which outlines all aspects of the RF spectrum of any radiator by means of a signal's maximum field strength in microvolts and millivolts per meter at a specified distance (usually 3 m). We can measure this radiated field strength by using a fully calibrated EMI wideband test antenna, along with an accurate spectrum analyzer that is outfitted with a wideband preamplifier.

In order to increase the radiated test accuracy, we must test in an environment with the least possible multipath effects, since excessive multipath would cause gross measurement inaccuracies and completely unrepeatable results. Open fields and empty parking lots are as close to ideal as we will be able to find with our limited testing capabilities. However, the most convenient test locations, such as your company's own indoor wireless lab, or an outdoor car-filled parking lot, will probably be the worst. We must also use the exact same antenna on the DUT that will be used on the final product. And if the DUT (normally an RF module) will be part of a larger system, then this system should also be integrated into your final precompliance tests.

The EMI test antenna should be a fully calibrated unit that covers all of our necessary test frequencies (i.e., up to at least the fifth harmonic of the DUT's fundamental). But, as the fundamental test frequency climbs in value into the microwave region, this single antenna concept will not always be able to fulfill our entire bandwidth needs, which will mean obtaining another test antenna of a different type. Typical calibrated EMI test antennas can be of the *biconical* type, which normally tops out at about 400 MHz; the *double ridge guide horn* type, with a huge maximum possible bandwidth of approximately 700 MHz to 18 GHz; the common and hugely popular 300 MHz to 7 GHz *log periodic*; or the wideband *tuned dipole* with a bandwidth of 300 MHz to 1 GHz. Frequency spans and maximum calibrated frequencies vary widely, depending on the particular EMI antenna model and manufacturer.

The spectrum analyzer used for EMI testing should either have a built-in preamplifier, or an external LNA. However, we must always be careful that the fundamental frequency of a transmitter does not produce spurious products within the preamplifier or the spectrum analyzer itself, which would make many of the test results meaningless. If this issue is occurring, we must employ either a tunable notch filter placed at the test instrument's input, which will substantially decrease the amplitude of the transmitter's fundamental frequency,

or remove the external LNA and/or switch-out the internal preamplifier of the spectrum analyzer, or increase the input attenuation by inserting an external pad attenuator, and/or by increasing the analyzer's own internal front-end attenuation.

Many EMI regulations will specify the type of the spectrum analyzer's detector to be used during EMI measurements, such as *peak*, *quasi-peak*, *average*, and so on, depending on the type of signal to be tested. The selection of the spectrum analyzer detector type can have a significant affect on measurement accuracy, but for the types of rough precompliance tests as described below, the default spectrum analyzer detector is more than adequate for most of our elementary EMI testing needs.

After these ballpark tests are completed for our wireless product, we will have some idea as to whether the unit's transmitter (and receiver) are close enough to the FCC Part 15 specifications to warrant employing a genuine, and far more accurate (but much more expensive), FCC-compliant test laboratory to perform the final testing.

However, if our own radiated RF testing quickly reveals that our product is exceeding regulatory limits, and that we are not yet ready to contract-out for true compliance laboratory testing, then we have just saved a significant amount of time and money by doing some basic testing ourselves. Indeed, we may even find that we must further attenuate the radiated harmonics or spurious outputs from the transmitter, which may demand circuit retuning, circuit/PCB modifications and additions, completely new PCB layouts, or redesign of certain stages of the transmitter. In fact, the most common design reasons for an initial precompliance failure of a new wireless design will be found in the transmitter section, and can routinely be traced to insufficient output harmonic filtering, lack of or improper RF shielding, poor RF decoupling, insufficient grounding on the PCB itself, long RF traces that are improperly terminated, incorrect part's placement, and too few fencing/grounding vias.

Performing an Approximate Radiated Pre-Compliance Test

This test will indicate whether a new RF design is near or far from compliance, follow these procedures (Fig. 12.7):

1. Place the DUT on a nonconductive, preferably rotatable, table at a height of at least 1 m from the ground's surface, and at a typical distance of 3 m from the calibrated wideband EMI antenna. Bore-site the antenna directly at the DUT,

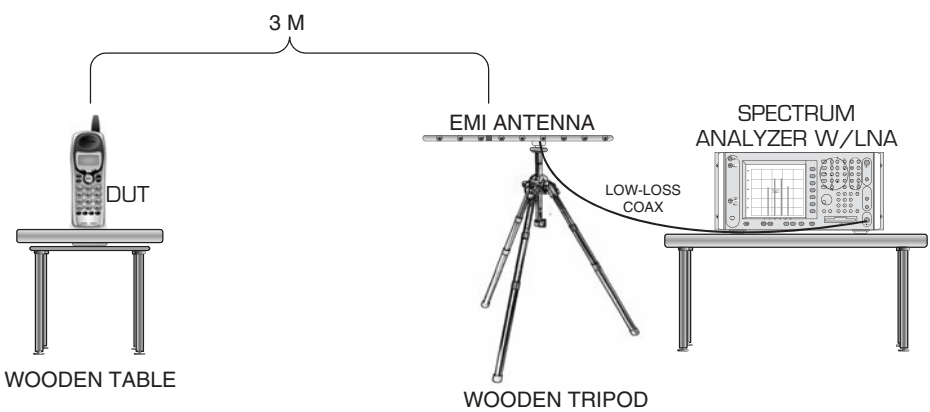


Figure 12.7 EMI or pre-compliance 3-meter test setup using a calibrated antenna in open air.

lock down the antenna on its nonmetallic tripod, and attach the antenna's low-loss coax to the input of the preamp-equipped spectrum analyzer. (For an accurate radiated test, and to prevent the creation of false frequencies within the spectrum analyzer, it may be required to utilize a notch filter to attenuate the strong fundamental frequency entering from the transmitter. This adjustable or fixed filter should be placed at the input of the test instrument, and not on the output of the DUT transmitter).

2. Turn on the spectrum analyzer and adjust for a slow, wideband frequency sweep. Check for, and log, all other local transmitters within the ambient RF signal environment. Since these signals are the *interferers*, they must not become confused with the radiated signals coming from the DUT itself. Throughout this sensitive electromagnetic compatibility (EMC) test, periodically inspect the entire measurable frequency spectrum with the DUT transmitter off, in order to confirm that no new noise or frequency sources have popped up since our last look.
3. Turn on the DUT and preamplifier, position the spectrum analyzer's RBW control to a narrow setting, adjust the input attenuator to its lowest value. Use *logarithmic power averaging*, when required, for very low-level signals. Switch the spectrum analyzer's vertical amplitude to dBuv. Move and reorient the DUT until the maximum signal strength of all radiated transmitter frequencies is catalogued. In searching for these harmonics and spurs, we may want to place the spectrum analyzer in *MAX HOLD* mode, and swivel the transmitter under test through different orientations. This is because the transmitter's antenna is not a perfect isotropic radiator, and therefore will have different gains at different orientations.
4. Obtain the true 3-m field strength, at each frequency of interest, within the 50-Ω test system by adding the spectrum analyzer's displayed voltage, in dBuv, to the calibrated EMI antenna's *antenna factor*:

$$F_S = V_{RF} + AF + CBL_{LOSS}$$

where F_S = RF field strength at 3 m from DUT, dBuv/m
V_{RF} = RF voltage as measured by spectrum analyzer, dBuv
AF = EMI antenna's specified *antenna factor* at the frequency of interest, dB/m
CBL_{LOSS} = coaxial cable losses between the EMI antenna and the spectrum analyzer, dB.

5. Since we are always looking for a DUT orientation that displays the worst case EMI spectrum/amplitude, reorient the DUT both vertically and horizontally to display the maximum worst-case amplitude of any harmonics or spurious transmitter outputs. And also reorient the test antenna from horizontal to vertical polarization to obtain the maximum, not minimum, EMI readings.
6. Place all data in a spreadsheet in the following format: Freq (MHz); Sig Level (dBuV/m); Ant & Cable Loss (dB); Corrected Reading (dBuV/m); FCC Limit (dBuV/m); Margin (dB); Result (PASS/FAIL).

As an example, one row of cells of this spreadsheet may read as follows: [500 (tested frequency, MHz); 25 (signal level, dBu/V/m); 20 (antenna and cable losses, dB); 45 (corrected reading, dBu/V/m); 43.5 (FCC's maximum limit, dBu/V/m); *+1.5* (margin of the tested signal above or below the FCC's maximum limit, dB); *FAIL* (final result at our test frequency for FCC compliance)].

CHAPTER 13

EMI Control and Printed Circuit Board Layout

Along with excellent wireless system and circuit design, sufficient EMI control through proper decoupling, shielding, and PCB layout techniques is a critical piece of the RF design puzzle.

13.1 Electromagnetic Interference

13.1.1 Introduction

Electromagnetic interference (EMI) is energy that is conducted, radiated, or both, and that harmfully influences otherwise proper circuit performance. Attenuating electromagnetic interference and *radio frequency interference* (RFI) to the lowest possible levels is a major requirement of wireless design. Undesired electromagnetic radiation escaping from a radio's enclosure or RF output port will not only interfere with neighboring wireless equipment, but small high-frequency currents and induced voltages within the radio itself can also destroy the correct operation of an otherwise solid wireless design.

EMI can be created in different ways and for different reasons. Rapidly switched DC, or any high values of AC or RF currents, if permitted to flow through a trace with impedance or inductance on a printed circuit board (PCB), will generate a voltage. If any nearby trace or component couples this energy via mutual inductance, major EMI issues can be created. Therefore, because of these transient current spikes, EMI problems can be produced by the pulsing voltage that is dropped across the unavoidable inductance of the DC supply traces located between the power supply itself and the active circuit, causing electromagnetic radiation and interference. However, decoupling components that are placed near the active circuit's V_{CC} pins will moderate this effect, as these high-value capacitors are able to immediately supply the local current required by the active circuit through a short, low impedance supply line, therefore decreasing the transient current needs of the remote power supply.

Return currents for an RF board trace, such as microstrip, will always run directly on the groundplane immediately below it. In fact, the current itself will follow any low impedance path around any obstruction to get right back under the trace. Thus, to decrease EMI, we must avoid increasing the area of this loop, caused when the return currents are diverted around some brake in the groundplane, which is created by poor PCB layout. This increased loop area will increase the inductance, creating a common frequency-sensitive impedance over the signal path as governed by $Z = 2\pi fL$. Now, due

to the increased inductance created by the larger loop area, a larger magnetic field is formed, which can couple into adjacent circuits, as well as functioning as a loop antenna and radiating EMI due to $v = L(di/dt)$. Indeed, the flowing RF signal may find, when confronted with the broken groundplane that created this high impedance loop in the first place, that is would rather flow through *any* other path that offers it less impedance to its flow. That other path may not be where we want the RF to go, since it may be through other circuits or systems. The EMI/RF coupling may also be capacitive, which means that no direct DC connection is required for the RF to be diverted to another area or component of the PCB. This is why when laying out a board it is important to completely understand the path of any return currents, so that we can confirm that these currents will be able to flow, unimpeded, back to their source.

Since a groundplane is intimately involved with this flowing current, it is essential to understand what a groundplane is and what it is not. A groundplane is not some magical current sink that sucks in and eliminates all RF and DC currents. No; a groundplane is merely a very low impedance return path that permits the current to flow easily back to the source from which it came, and with minimum impediments. Due to this exact capability, the groundplane can also function as an (almost) zero voltage reference for every circuit on the PCB.

When an RF trace must shift from its existing plane to any other plane on the PCB, we should also employ enough *stitching vias* near the microstrip's transition point to effectively connect the two different 50-Ω microstrip reference groundplanes together as a single groundplane. These multiple vias will then permit the return currents to flow easily and tightly between the microstrip's old reference ground-return layer and on to its new ground-return layer. Again, we are keeping the loop area on the PCB as tight as possible.

We can sum this all up by stating that the return currents that travel directly underneath an RF signal trace, on any unobstructed and continuous groundplane, *is* the path of least inductance, and actually possesses the tightest possible loop diameter within the current's path. This happy situation will permit the PCB to generate far less RFI, and also be better immune to stray RF fields.

13.1.2 Electromagnetic Interference Suppression

Analog Electromagnetic Interference

Suppression of EMI is necessary for all electronic designs, both wireless and nonwireless, due to rigid European and strict Federal Communications Commission (FCC) regulations, as well as to protect your design against interferers.

There are many ways to accomplish this, many of which must be used in concert: Metallic shielding of synthesizers and oscillators, which is used to prevent coupling energy into other circuits or to prevent other circuits from coupling energy into our frequency sources, is crucial. This is because any such EMI that is propagating out of a frequency source may decrease the isolation between other stages and/or cause harmonic or mixing products to appear in the rest of the radio, while EMI that successfully enters a phase-locked loop (PLL) or voltage controlled oscillator (VCO) may create undesirable spurs. Therefore, traces leaving a frequency source should be protected by placing them on a lower PCB layer or within an RF shield box, which is especially essential when dealing with high-power frequency sources such as those used to feed passive diode high-level mixer conversion stages.

Although placing RF shielding over any EMI-generating circuit (or shielding a circuit that is susceptible to EMI) will aid in decreasing this radiation substantially, the EMI emissions may still find another path: they may actually travel through the dielectric of the PCB itself. For either the fundamental or its harmonics, the method used to attenuate this particular EMI propagation mode is to utilize a barrier of through-hole vias all the way from the top of the PCB to its very bottom groundplane. This is referred to as a *via fence*, and is located around some or all of the affected circuit. At high microwave frequencies, or if high amplitude harmonics of the fundamental are present, these vias will have to be placed in a relatively dense pattern.

Preventing EMI from corrupting the system's DC power supply, and thus the entire wireless device, is critical. A generous amount of decoupling capacitors that are capable of shunting the entire spectrum of possible EMI frequencies, along with a high frequency choke, should be placed as close as possible to all active devices. This will help prevent unwanted oscillations and/or contamination of our desired signal, and is especially important for VCOs, PLLs, low-noise amplifiers (LNAs), and PAs, which must all be heavily decoupled from any noise-producing conducted emissions from the DC supply, or prevented from injecting any of their own circuit-generated EMI from reaching the common DC supply and other circuits. And since saturated power amplifiers and other nonlinear stages can produce high-amplitude harmonics and distortion products, the EMI created within these particular circuits may be the fundamental itself—or the fifth harmonic of the fundamental.

Digital Electromagnetic Interference

The electromagnetic interference created within digital circuits located in or attached to the radio system is quite similar to the above analog circuits. However, due to certain design variances, and the huge amount of harmonics generated by the high-speed square-like waveforms of a digital circuit, there are some added considerations.

If traces of any significant length are left unterminated, or terminated into a high impedance, they can be capable of radiating substantial EMI, especially at high frequencies, almost as if they were small monopole or loop antennas. The trace may also be located close enough to another trace or component to couple capacitively or inductively between the two objects, causing significant or subtle crosstalk between the circuits. The capacitive coupling can be visualized as a virtual capacitor that has been placed in parallel between the two circuits, creating a low impedance path for the EMI current to flow. Crosstalk created by inductive coupling can best be visualized as each of these traces being comparable to a series inductor, or a winding of a transformer, and thus causing induced coupling of EMI currents to flow within the victim trace. EMI radiation from traces may not even be at the expected fundamental frequency, but can be at one of its many harmonics. In fact, these harmonics are more easily radiated than the fundamental due to their shorter wavelength: as the length of the trace gets closer to one-tenth of a wavelength of the signal of interest or longer, the *transmission line effect* must be considered in order to preserve signal integrity. The transmission line effects infers that after a trace becomes longer than this one-tenth of a wavelength, the signal's response to the trace itself must be considered and, as the frequency of operation increases further, reflections will become more of a problem on a simple noncharacteristic impedance trace, sometimes causing extreme signal degradation of the original digital waveform. We can prevent much of these transmission line issues by confirming that the trace itself has the proper impedance, by using shorter line lengths *versus* wavelength,

and by employing proper characteristic impedance end terminations. Just as in analog RF design, 50 Ω is accepted as the characteristic impedance for transmission lines and terminations for most high-speed digital designs.

Power Planes, Groundplanes, and Electromagnetic Interference

PCB power or groundplanes can be a powerful tool in the improvement of EMI—or be the cause of it. By following established ground- and power-plane PCB layout procedure we can prevent a board design from becoming an unacceptable EMI radiator, and thus failing FCC radiated emission testing. Yet the visual differences between a successful electromagnetic compatibility (EMC)–compliant board design and a failed one can be an extremely small one, and may be almost invisible to even critical observers. Therefore, as with all work in high-frequency RF design, the difference between success and failure can be in the smallest of design details.

To optimize the printed circuit board's power- and ground-planes for suitable EMI suppression, we must successfully ensure that some basic layout rules have been followed: (A) Have we lessened the board trace radiating loop areas? This can be accomplished by not permitting any large slots or holes in the PCB's groundplane, and by connecting the top partial PCB ground layer to the larger main groundplane layer through multiple vias. (B) Have we prevented edge-coupled EMI from escaping through the periphery of the board? This can be done by not allowing the power plane copper to reach to the farthest edges of the PCB, and by placing a tightly spaced, grounded via fence around the edge of the board (which the power plane must not physically contact). (C) Have we prevented the power plane layer from functioning as a patch antenna, radiating EMI throughout the PCB? This can be achieved by extensively RF decoupling of the power plane layer to the groundplane layer with a sufficient number of suitable capacitors, which are placed on multiple locations throughout the board. (D) Have we prevented the top RF layer, even at high harmonic frequencies, from conducting local RFI/EMI throughout the system? We can do this by the use of a sufficient number of suitable decoupling capacitors on the top RF layer, terminated through vias to the main bottom groundplane, and inserted locally at all active stages and the power supplies, as well as by employing full, solid power and groundplanes that cover an entire dedicated board layer, which will increase the PCB's own power-to-ground capacitance, especially when the separation is 10 mils or less.

Rules for Minimizing Electromagnetic Interference Issues

Since most of today's modern RF circuits have, at their center, an integrated circuit of some sort, we normally find that most of the work in EMI suppression of radiated and conducted emissions is based around this fundamental radio frequency integrated circuit (RFIC) circuit, or its discrete or monolithic microwave integrated circuit (MMIC) power amplifier stage. We must also increase the RFIC's immunity to other radiated and conducted emissions generated within other circuits.

NOTE: *Since the wireless design may actually be earmarked for placement in a communications system's metal, or some other heavily shielded enclosure, we do not necessarily need our circuit to pass our own initial precompliance FCC testing (unless our design has actually been placed within that same enclosure during this testing), since it should be remembered that only the final, completed radio—with full production enclosure, antenna, and all support circuits—will be what is truly tested for FCC compliance, and not just our single, exposed RF circuit board.*

To minimize EMI in any system or board design:

1. Minimize the loop area of the PCB trace nets.
2. Do not place high-speed digital or unshielded RF circuits too close to unshielded cables (due to the cable's relatively long length, which can act as an efficient radiating antenna, causing emissions).
3. Never route high-speed traces near sensitive circuits or components, nor between pads.
4. Do not place unshielded transformers or inductors near sensitive components.
5. Keep all high-speed I/O traces short.
6. Filter high-speed digital signals to decrease their rise and fall time using a damping resistor, ferrite beads, or a shunt capacitor.
7. Maintain all high-speed circuits that communicate with each other together in the same close grouping.
8. Each separate PCB in a communication's system should have but a single common high-speed control cable from board to board.
9. Due to transmission line effects, use only matched terminations and controlled impedance lines when sending high-speed signals on a trace longer than one-tenth of a wavelength long.
10. For the most direct, low-impedance connection use a via or vias sent directly from the active circuit to its V_{CC} power plane.
11. If located on the same PCB power plane layer, separate different power plane sections (i.e., digital and analog), by at least the thickness to the nearest groundplane.
12. Do not place high-speed or RF traces near the edge of a PCB board, since such placement can cause edge emissions to increase by up to 20 dB.
13. Grounded guard traces may not be as useful between a high-speed and a low-speed trace as is simply maximizing the open-area spacing between the two lines.
14. Power traces should be as wide as possible to maximize capacitance and minimize inductance. (A trace over a conductive plane has an inductance of approximately 7 nH/in, while over just the dielectric, with no conductive plane, the inductance may be around 20 nH/in).
15. Never permit any traces of any appreciable length to end in an open or high impedance, since they will act as a monopole antenna and radiate or receive EMI. The amplitude of the EMI will be governed by the trace length and the EMI's frequency, field-strength, and polarization.
16. Placing high-speed control lines or RF traces in stripline (i.e., between two solid reference planes) will shield them from receiving and transmitting EMI by a level of almost 50 dB.
17. Digital clock generation circuits can be the primary source of generated noise on any PCB board, so using grounded and self-shielded clock ICs can assist in decreasing the EMI.

18. Modern surface-mount PCB layouts are very dense, which can aggravate crosstalk between components or traces. Therefore, make sure to maintain the separation of all sensitive components from on-board EMI generating circuits/components/traces.
19. Heavy PCB transient current demand, which can be commonly caused, for example, by a rapid RF power amplifier turn-on or multiple simultaneous digital devices switching, can produce EMI and a temporary drooping of the DC voltage supply amplitude as it attempts to provide enough immediate current to all on-board devices. It can be lessened by using low inductance V_{CC} and V_{dd} traces, sufficient low equivalent series resistance (ESR)/high value decoupling capacitors, and even localized power supplies. This will prevent EMI from the long power supply traces with their high, rapid current changes.
20. RF decoupling capacitors should be placed as close to active devices as possible, and then shunted to ground using a path of least impedance, which will always be multiple dedicated ground vias located at each decoupling component's PCB pad.
21. Due to the added inductance, never use *via thermals* or component pad thermals in an RF design.
22. So that digital EMI does not couple directly up from the ground and into the RF section, the digital circuits may need to have a separate groundplane area, instead of sharing a mutual groundplane. This is usually accomplished by simply splitting a common groundplane, with only a single thin copper trace connecting the two.
23. Large heat sinks can supply significant RF shielding to any equipped IC.
24. If some non-microstrip traces must run over each other on different layers, lay them at right angles to each other.
25. With exceedingly fast control signals (> 150 MHz) do not use tight, 45° bends when changing trace direction, but only 90° bends.
26. Temperature-reducing perforated RF shielding should be utilized only with extreme caution, and typically not when a design is to operate much above 2 GHz, as harmonics of the fundamental signal may then escape through the enclosure's holes.
27. Employ suitable shielding, since a complete lack of shielding or improper layout can cause a filter to become virtually useless if the signal to be filtered is able to propagate around the filter itself, thus allowing EMI to pass almost unfiltered. Surface acoustic wave (SAW) filters, due to their high insertion losses, are especially sensitive in this way.
28. It is prudent to ground all cases, shields, and septums straight to the PCB's groundplane.
29. Try to completely isolate AC-to-DC and DC-to-DC power supplies by using heavy decoupling and self-shielding packages. Because these supplies will be controlling large currents, they can produce considerable EM fields from the shear magnitude and fast rise/fall times of the current. (Even at low switching frequencies, these power supplies may generate significant interference out to their 50th harmonic).

30. At 400 MHz and above, tough EMI radiation problems can sometimes be mitigated by attaching RF absorbing foam or rubber to the top of an offending circuit.
31. Since EMI can easily be generated at any frequency by an amplifier that has become unstable, confirm stability of all active circuits.

13.2 Printed Circuit Board Design

13.2.1 Introduction

In the world of microwaves and RF, the PCB's layout, construction, and materials are almost as important as the circuit design itself. In fact, a perfect circuit design of even the most basic RF oscillator or filter can be ruined by improper board layout.

13.2.2 Printed Circuit Board Materials

At microwave frequencies, PCB insertion losses become a major concern when choosing a substrate material. When distributed filters and circuits or long runs of microstrip are used, common FR-4 substrate begins to exhibit extreme insertion losses at approximately 1.5 GHz and up. In fact, such lower cost generic FR-4 has a loss tangent of roughly 0.03 at microwave frequencies, while the higher quality FR-4 substrates, such as Isola's FR408, have a much superior loss tangent of 0.01—along with superior physical and electrical board tolerance parameters. Quality FR-4 substrates like FR408 make it possible, and even highly desirable from a cost *versus* performance perspective, to empl oy inexpensive FR-4 for small, consumer-grade wireless RF modules up to frequencies of 6 GHz. So, while generic FR-4 may have loss tangent figures that are too modest for an RF board designed at 5.8 GHz, using high-quality FR-4, along with a small, tight microwave board layout exploiting surface-mount technology (SMT) lumped passive elements, permits us to supply acceptable RF performance levels for most mass-produced, low-power radio applications up to extreme frequencies. Nonetheless, when designing larger PCBs, it should be kept in mind that even such quality FR-4s can still mean a loss of 0.5 dB/in at 6 GHz.

The case for using *only* branded FR-4 materials at up to 5.8 GHz is very compelling in any low-cost market, since all other higher frequency board materials are not only too expensive to be realistically or competitively used in consumer wireless applications, but may also be disturbingly soft structures, deforming quite easily. Still, for certain industrial and military applications, these softer microwave board substrates can be used when backed by a metal, or even generic FR-4, board stiffener. This will add the required rigidity, while permitting the selection of different dielectric constants and superior loss tangents, along with improved environmental performance and dielectric/dimensional tolerances. Unfortunately, most of these microwave board materials also demand very special processing techniques to fabricate a complete printed circuit board, so fabrication companies that process FR-4 at such a low price will frequently be unable to work with the more specialized substrates, further increasing our costs. Alternatively, there is a superior PCB material from Rogers that will allow *any* board fabrication house to easily manufacture a complete, and relatively low-loss, microwave PC board. This particular advanced substrate is the rigid, high-frequency Roger's RO4000 series, with a very good loss tangent of 0.0027, and an E_r of 3.38. When either large or high-powered microwave radios are to be designed, or those that employ

distributed filters, then this would be the most economical substrate in order to obtain reasonable microwave performance levels. But while the cost of the RO4000's processing at the fabrication facility is comparable or equal to FR-4, the board material itself is somewhat more expensive, therefore demanding that only quality FR-4 should be exploited whenever possible, even up to 6 GHz, for any significant production run of consumer wireless devices.

Since insertion loss becomes the dominant issue when selecting a PCB board material at microwave frequencies, we should quickly examine the reasons behind these types of losses. A PCB's fundamental losses are shared between the board's conductor and its dielectric. When utilizing low loss-tangent microwave substrate materials at frequencies under 2 GHz, the losses in the copper conductors can overwhelm any losses within the substrate itself. At higher frequencies, this may not be the case, and as the frequency reaches 20 GHz a board material with a dielectric constant of 3.5 will begin to display an almost equal loss within both the dielectric and in the conductor. Further, higher dielectric constant materials will increase both the dielectric and the conductor losses over that of lower dielectric constant materials. Indeed, it can be demonstrated that both the dielectric and conductor losses, which are essentially the PCB's overall insertion losses, increase with higher loss-tangent *and* dielectric constant, especially at elevated microwave frequencies. The following values will be the highest recommended dielectric constants to use on microwave PCBs to minimize insertion losses: $E_r = 10$ for 4 to 5 GHz operation, $E_r = 6$ for 6 to 7 GHz, $E_r = 4$ for 13 to 14 GHz, and $E_r = 3$ for up to 30 GHz.

Dielectric constants of smaller values can be used to control the size of the distributed circuit elements, such as microstrip and equivalent components. At microwave frequencies high dielectric constant materials may force the microstrip circuit topologies to become too small to be realistically realized, both economically and physically. Increasing the substrate's thickness, to a certain extent, will help to fight this effect (as well as lowering board losses), but at the cost of possibly confronting different undesired modes of signal propagation, along with increased board via inductances.

Dimensional and electrical tolerances of the PCB is another extremely critical parameter in microwave circuit board design, especially if there are distributed structures used within the circuit. This becomes readily apparent when viewing the results of several Rogers Corporation tests and simulations used to demonstrate the interactions that will occur when either the dielectric constant, loss tangent, or dielectric thickness of a PCB are varied, while the other two properties are left unchanged. Figure 13.1 presents a distributed edge-coupled bandpass filter centered at 1 GHz with a bandwidth of 10% and exploiting a 50-mil-thick Duroid material. With all electrical and dimensional substrate parameters ideal, Fig. 13.2 shows the filter's passband response in the frequency domain. Figure 13.3 illustrates what occurs to the filter's center frequency and passband when the dielectric constant is varied slightly both upward and downward. A small increase in the E_r shifts the passband lower in frequency, while a slight decrease in E_r shifts the passband higher. Variations in a board material's E_r will affect the passband frequency of a filter, because any change in the E_r will alter the velocity of propagation through the dielectric, and thus change the electrical length of each of its elements. Unfortunately, both manufacturing and temperature variations will influence the dielectric constant, necessitating a board material with a tight initial E_r tolerance, as well as a low *temperature coefficient of dielectric constant* (*TCK*). Now, by glancing over at Fig. 13.4, we see another board-related

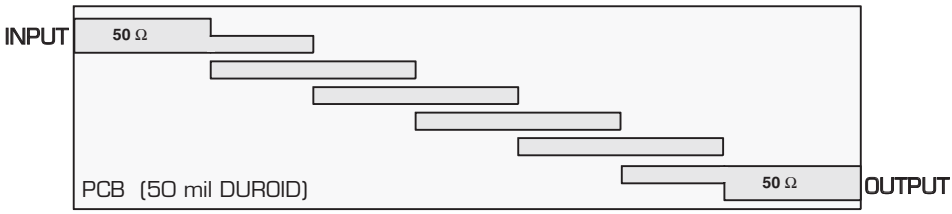


FIGURE 13.1 An edge-coupled distributed BP filter used for dielectric tolerance tests.

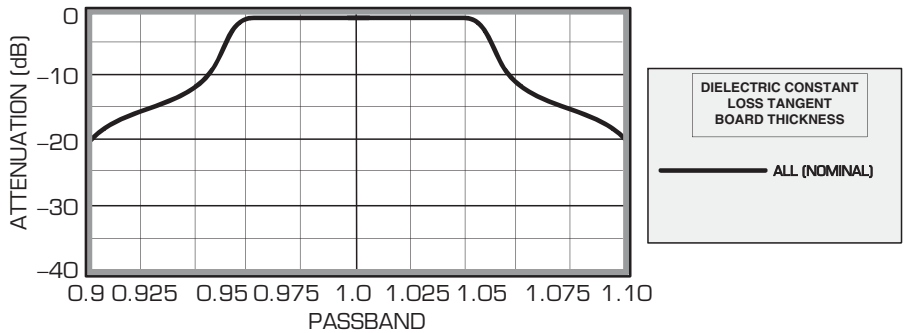


FIGURE 13.2 Dielectric constant, loss tangent, substrate thickness at nominal values; distributed filter's center frequency and losses are as expected.

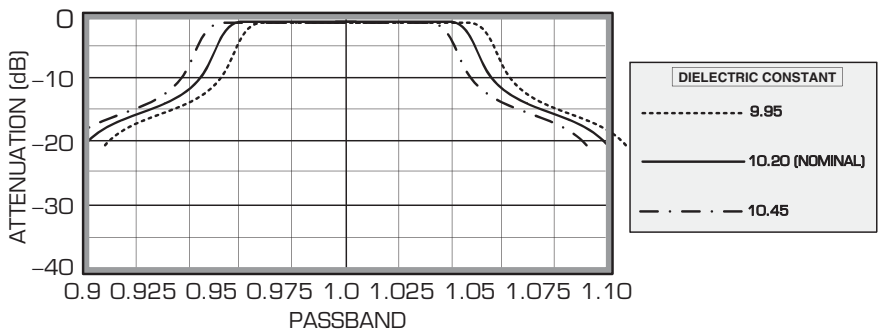


FIGURE 13.3 Variations in center frequency of a distributed filter as the dielectric constant is varied.

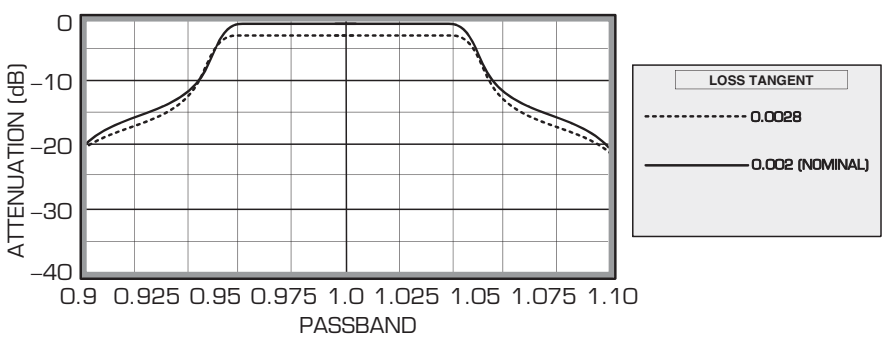


FIGURE 13.4 Increase in insertion loss of distributed filter as loss tangent is increased.

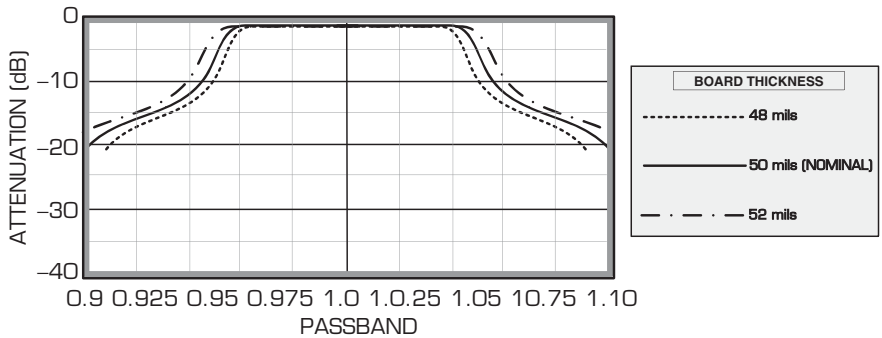


Figure 13.5 Variations in the bandwidth of a distributed filter as the board thickness is varied.

tolerance issue, this time for the loss tangent. The graph clearly demonstrates that as the PCB's loss tangent is increased, so does the signal attenuation, causing a slightly lower filter output than expected. Additionally, a board's dimensional tolerances will have large influences at high frequencies over the bandwidth of the distributed filter, with thicker than expected substrates increasing bandwidth, thinner substrates decreasing the bandwidth, as revealed in Fig. 13.5.

Distributed filters are very sensitive to dielectric constant, loss tangent, and thickness variations of the PCB's substrate, but lumped filters will demonstrate a much more subtle outcome, both because the distributed elements of such a structure are limited to the board's traces/pads and the component's own parasitics, and simply because lumped filters are normally used at much lower frequencies.

13.2.3 Printed Circuit Board Construction

A surface-mount printed circuit board, at its most basic level, normally comprises most or all of the following elements (Fig. 13.6) and, to save cost in consumer applications, will generally have only two (Fig. 13.7) to four layers of FR-4, employing through-hole vias only (no blind or buried vias):

1. The *solder mask*, which is a material, usually green in color, that covers the outer PCB layer and prevents the paste solder from sticking to any undesired metal surfaces.
2. The *metal layer*(*s*) (or *conductive layer*) is any layer on or within the PCB that is covered with a metal, such as copper, and is specified in various metal thicknesses, such as 1/2 oz (0.7 mils), 1 oz (1.4 mils), and so on. The metal may be plated with other metals for higher conductivity, or to protect exposed surfaces against rapid oxidation to preserve solderability. The outer metal layers usually consist of not only traces and ground pours, but also footprint pads that permit lumped components to be electrically and mechanically soldered to the PCB, while any inner metal layers are for extra AC, DC, or RF connections that were not convenient, or viable, to be routed on the PCB's top and bottom external layers. Almost all metal layers will have these metallic connections of traces and pads running from circuit to circuit, but some layers may employ only a single, fully poured and solid copper *plane* (a plane is a metal surface covering an entire PCB layer, and is generally used as a dedicated

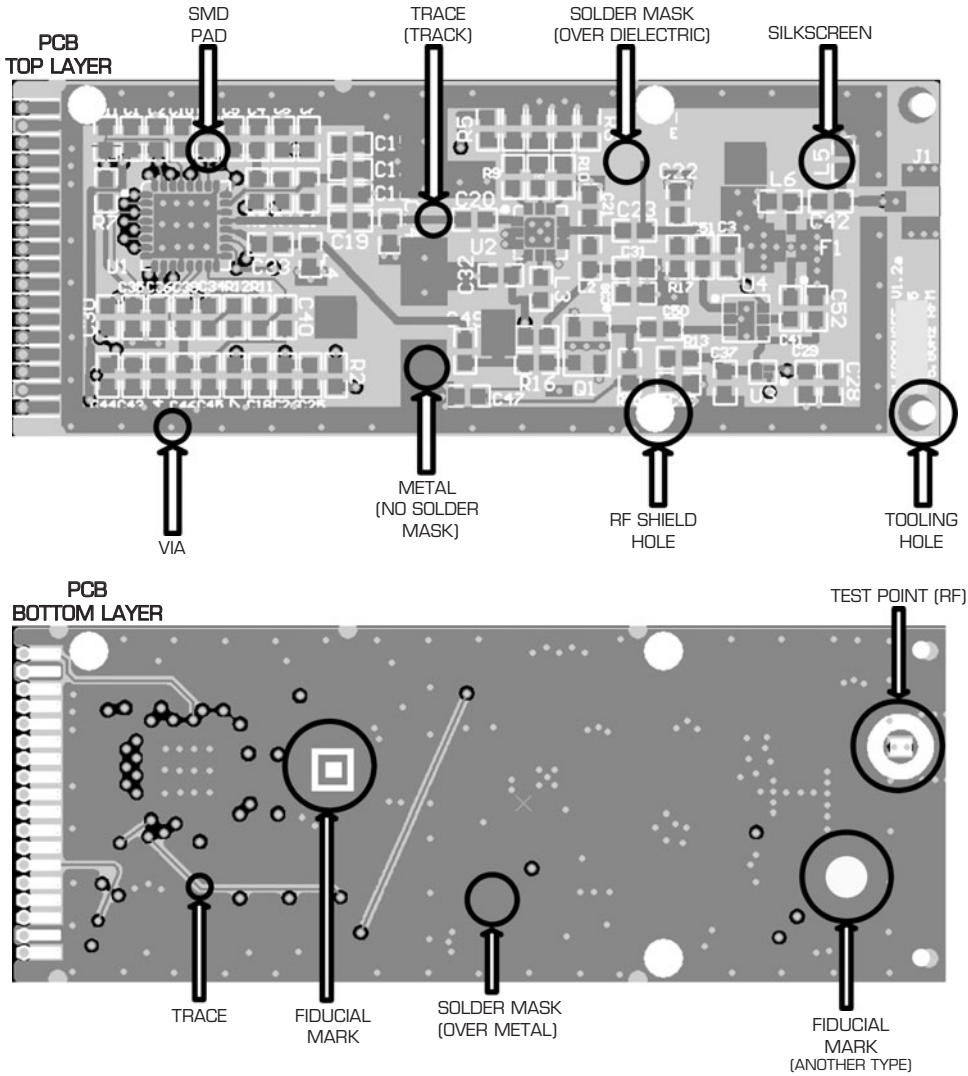


Figure 13.6 A production ready consumer-grade microwave PCB, top and bottom layers.

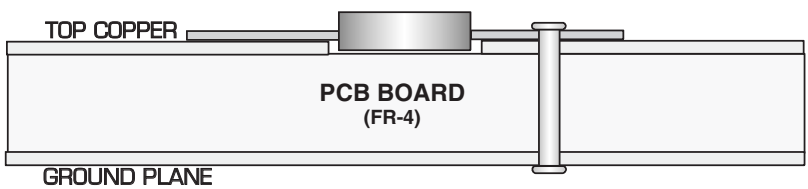


Figure 13.7 Side view of a through-hole via traveling from the top to the bottom groundplane.

ground or power layer, an internal EMI shield, or to assist in the dissipation of heat). The number of these metal layers, whether the metal layer consists of only traces or simply a single groundplane pour, signifies the quantity of board layers that a PCB is said to have, such as a *2-layer board* with its two metal layers, or a *4-layer board* with its four metal layers, and so on.

3. The *silkscreen* is a board's outer layer that is comprised solely of ink. The ink is applied by a silkscreen process, is typically yellow in color, and is used to display text or demarcation lines on the surface of the PCB. Each component should have its own silk-screened *reference designator* (a schematic-generated number that is unique to each lumped part), *component orientation marks* (used to indicate a part's proper polarization, orientation, and Pin 1 placement), and *part outlines* (squares or rectangles that display where components should be placed). All such markings must be visible, even after assembly, to aid in any later troubleshooting, rework, or field circuit repair. Additionally, placement of a *PCA assembly part number* (the part number of the *entire* assembled printed circuit board along with components, with PCA standing for *printed circuit assembly*), a *PCA REV* solid yellow blank area (for adding a later PCA revision number after any changes to the board or the components), and a *PCA assembly name* (the project's name or number) should also be placed on the board's outer layers via this silkscreen layer. Further, the name of your company, with its logo, will be inserted in a prominent location, along with, optionally, the date of manufacture. (There should also be a *PCB part number* [indicating the PCB's own exclusive part number, and not that of the printed circuit assembly (PCA)], which will be etched in one of the board's outer copper layers, along with an etched *PCB REV ID* letter or number (identifying the revision of the bare PCB itself), also in copper, to the right of the PCB's part number).
4. The *vias*, which are drilled and conductively plated holes located either completely through the entire printed circuit board, or only partially through certain select layers, and are used to connect metal layers together both electrically and thermally. However, if a specific ground or power plane layer does not want to be connected to this a via, then the layout engineer can easily specify that the copper on this layer should pour around the intruding via itself, and not contact any part of it (the small void without copper around the via is referred to as an a *antipad*). This antipad is required because a typical via must be drilled all the way from the top of the PCB, through all the internal layers, and out to the very bottom of the board.

   However, *blind vias* can also be used in a PCB design, which are plated holes that travel to a specific internal layer, and no farther. Or *buried vias*, which are placed within the PCB itself and connect only two or more internal layers. *Laser drilled vias* are also possible in order to achieve extremely small via diameters, thus saving board space. Both the blind, buried, and laser via types are significantly more expensive to fabricate than the through-hole vias discussed above, so are rarely used in most consumer wireless equipment (unless extreme board density is required due to PCB size constraints, as in some cell phone layouts).
5. The *substrate*, or *dielectric*, is the base insulating material that forms the bulk of the PCB's structure. In consumer products, fiberglass-based FR-4 is now by far the most common such substrate material used, even into the microwave frequencies. Military and high-end industrial requirements may exploit more

exotic and expensive dielectric materials, depending on the application and frequency. Each of these different substrate materials will have different dielectric constants, loss tangents, thermal properties, electrical and dimensional tolerances, strengths, rigidity, and so on.

6. *Fiducial marks* are 40-mil-wide copper-filled circles, devoid of solder mask, with a ring 100 mils in diameter that is clear of copper, masks, parts, and traces. One fiducial should be placed on an upper corner of a PCB, and another at the most distant opposing corner. These marks help in quickly and accurately verifying, through machine vision, that the PCB itself is oriented so as to properly receive the pick-and-placed parts, which are inserted by preprogrammed robotic assembly equipment. For precision placement operations, such as that for fine pitched parts of less than 20 mils, may also require their own *local fiducial* mark.
7. *Tooling holes*, used by the assembly and test equipment for proper PCB alignment by metal pins, should be placed in opposite far corners of the PCB, and are 125 mils in diameter and unplated.
8. *Test points* are required for most high-volume production runs, and are merely cleared and accessible locations on the PCB that are used by automatic test probes for rapid testing of both the DC and RF parameters of the unassembled and the assembled board. Therefore, these test points will usually involve a small area on the board that is devoid of solder mask, and that a miniature probe on a test platform can reach and make electrical contact. Any such test points on the PCB for RF testing must generally be a 50-Ω node, as well as having a nearby accessible groundplane area that is only a few 10s of mils away.

13.2.4 Printed Circuit Board Design Considerations

Auto-routing

Any successful board layout for RF circuits, especially at microwave frequencies, is a specialized art form that mixes experience, RF and EM knowledge, technical intuition, the desire to solve puzzles, and pure luck to obtain a first-pass success of a complex wireless board design. Indeed, when designing RF boards, as well as many modern high-speed digital PCBs, you cannot simply tell your PCB layout package to run its built-in *auto-placer* (for components) and its *auto-router* (for trace routing) to design most of your board. The results would be hopelessly deficient and completely inadequate. Still, both the auto placer and the auto router do have some limited utility, especially in DC, AC, or low-speed digital work.

Decoupling

Decoupling capacitors for the DC power supply of the entire system, or for an individual PCB, should have a very low capacitive reactance to ground for *all* frequencies and for *all* devices being fed by the supply. When this is so, any RF/EMI signals from the radio itself will be blocked from entering and contaminating the DC power supply (which could spread the noise to other circuits and systems), as well as preventing any internal power supply noise from reaching the radio's active stages. The decoupling function is accomplished by using capacitors and inductors which are optimized for diverse frequencies, such as electrolytics for low frequencies and ceramics/porcelains for higher frequencies.

To further supply maximum microwave decoupling, use 10-mil spacing from the top signal layer to the first groundplane, if possible. In this way, extra RF harmonic decoupling is supplied by the close proximity of these two planes, which are now acting as plates of a small value, microwave capacitor. This also provides lower inductive reactance to the lumped decoupling capacitor's ground vias, as well as a higher component body and trace pad capacitance to ground.

When designing common RF/microwave circuit boards, all RF decoupling components should be of the smallest practical size due to their superior high frequency operational characteristics (less parasitics). Currently this would be the 0402 package. For even better RF performance we could drop down to yet smaller sizes, such as the 0201 package. However, these extremely tiny parts currently have a higher initial price per component, as well as a higher final assembly cost, along with the unavoidable lower maximum possible capacitance and inductance values.

Microstrip

Transmission lines, usually microstrip, must always be utilized to maintain 50-ohm constant impedances when using PCB trace runs longer than a one-tenth of a wavelength. To lessen mismatch losses and reflections caused by trace impedance discontinuities, this practice is standard on any high-speed board. Some lower frequency circuits may exploit simple traces of any convenient width, or even short sections of bare wire, but these conductors should be kept very short to prevent any transmission line effects. (Even a 400-mil-long piece of wire can have an inductance of approximately 10 nH, or 63 Ω at 1 GHz, forming an almost pure inductor).

For the best microstrip performance, always try to use a true groundplane as the microstrip's reference layer, rather than a power plane (called an *image plane*), since coupling between the power plane and the groundplane is through the limited bandwidth of the decoupling capacitors, rather than the very wide bandwidth of through-hole vias.

All bends in microstrip should be mitered or rounded to prevent radiation into adjoining circuits, and with no bend any sharper than 90°. And if a component is narrower than the microstrip itself, then the microstrip should be tapered for a decreased impedance bump into the part (Fig. 13.8). A similar concept to microstrip tapering is shown in Fig. 13.9, in which the component is actually selected to be of the same width as the trace itself.

To decrease impedance variations and lower VSWR when an RF signal encounters a passive component, such as a coupling capacitor, the component should not only

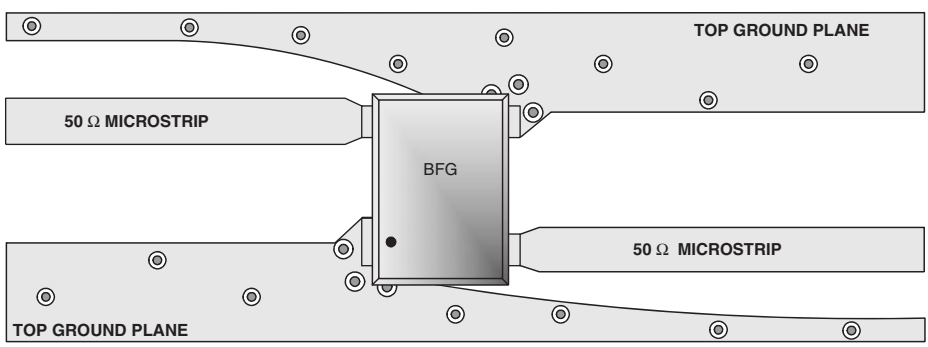


Figure 13.8 Proper RF board layout for a transistor as seen from the top of the PCB.

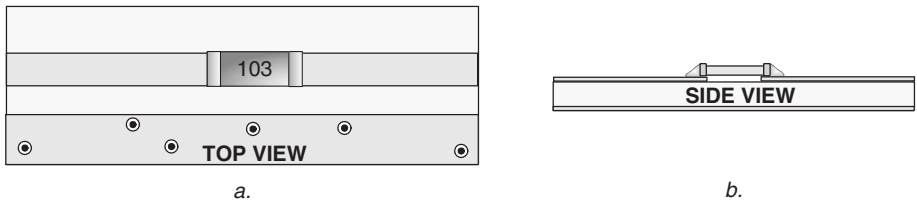


Figure 13.9 (*a*) Proper component width and (*b*) component soldering for decreased reflections at high frequencies.

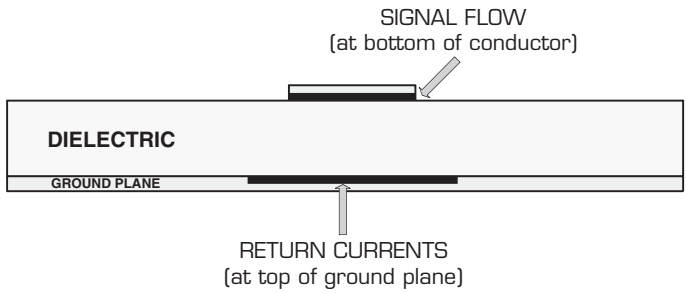


Figure 13.10 Currents in PCB microstrip.

ideally be of the same width as the microstrip, as just mentioned, but its solder fillet should be as smooth as possible so as not to disturb the RF signal flow. Another issue that creates impedance discontinuities is any break in the reference plane located directly underneath this microstrip, since the return currents for the board's transmission lines flow directly under the trace itself (Fig. 13.10).

Even on signal traces that employ 50-Ω microstrip, losses are a fact of life. Depending on the substrate in use and the frequency of operation, losses of between 0.1 and 1 dB/in can be caused by dielectric constant variations, high loss tangents, dielectric heating, copper losses, and undesired radiation. In many applications this may not be of much concern, but where every dB matters, such as in the front end of a low-noise receiver, we may want to work with a board material with less losses.

When 50-Ω microstrip must change PCB layers, a via transition must be used. This via transition will connect the 50-Ω microstrip on one layer to the 50-Ω microstrip (or stripline) on another layer. However, since vias have an impedance of approximately 30 Ω (the exact value dependent on multiple factors), the layer-to-layer transition will cause an undesired via-induced impedance bump. Therefore, we would typically endeavor to maintain all such RF carrying traces on a single PCB layer. Nevertheless, due to obstructions created by other traces and components, EMI suppression, and board size constraints, there are times when we simply must change layers. This can be done on most consumer boards by the above mentioned method, which is by the use of a single signal via to transition the microstrip to another layer. However, in more sensitive applications, we can actually form a virtual 50-Ω coax cable around this single via, as shown in Fig. 13.11, thus converting it into an almost perfect 50-Ω structure. To realize the 50-Ω transition via, we will surround it with four or five ground vias, which simulates a transmission line at our desired impedance. We may quickly *approximate* the required distance these ground vias must be spaced away from the center signal via with the enclosed Agilent AppCAD software:

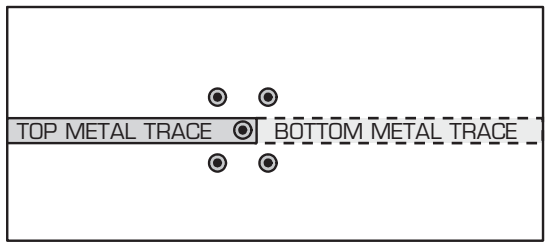


Figure 13.11 Microstrip transition to another layer using 50 Ω via to avoid any impedance discontinuities.

Select Passive Circuits, then COAX (ROUND). Enter the PCB substrate's dielectric constant, the signal via's diameter, and enter 60 in the D1 box. Select CALCULATE Z0. Modify the value in the D1 box until the Z0 box reads 50 Ω. You will now have a ballpark diameter spacing value to be used between the edges of the four or five ground vias. The center signal via, of course, will be in the exact center of these multiple ground vias, with the board's surrounding copper pour on each layer coming no closer to the center RF via than do the edges of these encircling ground vias.

Grounding

All ground returns for discrete active and passive devices and RFICs must be sent to the PCB's solid groundplane by the shortest route possible, which will be through multiple vias. This main groundplane must also always be located directly underneath the top component layer. The multiple vias will be part of the top groundplane pour, directly connecting the two planes, and improving electrical and thermal grounding and EMI mitigation, while also decreasing the inductance for the RF's return path to ground. This is especially important for proper stability and gain for a discrete transistor's emitter circuit, and for the correct frequency response of a lumped filter circuit.

Since through-hole vias will have a certain inductance (just as any conductor will), the path to ground can be further improved by using thinner substrates, with the via inductance value at different lengths being roughly approximated by:

$$L = 5.08h\left[\ln\left(\frac{4h}{d}\right)+1\right]$$

where L = inductance of the via, nH
h = length of the via, in
d = diameter of the via, in

This formula infers that varying the diameter of a via will not help much to change the via's inductance, while altering the via's length will have a profound effect; which is why all vias to ground should be kept short, especially at microwave frequencies. Vias must also be placed at regular intervals of an eighth of a wavelength, or less, through the top ground pour down to the bottom groundplane of any RF circuit.

All conductors will have a certain amount of resistance, and a printed circuit board's traces and groundplanes are no exception to this rule. At RF frequencies, this resistance to current flow is further exacerbated by the *skin effect*, as well as by the ever-increasing inductive reactance that occurs as frequencies rise. This overall nonzero impedance of

the ground circuit can adversely influence an active stage's operation and stability. And since copper will have this innate resistance and inductive reactance to all AC/RF signals, then any two points along a real ground trace or groundplane cannot possibly be at the same voltage. This will cause an ambiguity as to what the actual signal reference voltage truly is to any circuit that is attached to this ground. In other words, as all of the radio's active stages draw current or generate or amplify the RF or AC signal, the groundplane will vary its own voltage from 0 V to some other, non-zero value. So, as high-current stages switch on and off, or are being operated at a specific frequency, they will essentially modulate the ground's own voltage level by this same switching time or frequency. This can create stage instability in active circuits, as well as strong PCB EMI. The board's groundplane will only be returned to its ideal 0 V when all of the high-powered devices are switched off, and therefore not drawing current.

We can minimize this high resistance/reactance ground circuit that is not only not at 0 V, but is at various voltages levels along all parts of its entire surface, by utilizing a solid groundplane pour with few, if any, slots or voids in its surface; using multiple vias from the top component layer ground to the main lower groundplane; making direct connections of all top component ground pins directly to the top ground pour; employing a sufficient number and types of decoupling components; and exploiting a thin PCB substrate layer between our top component layer and our reference ground.

Undesired Coupling

Another issue in microwave PCB design is undesired coupling, or *crosstalk*. Since we will normally want the entire wireless printed circuit board layout to be as physically condensed as possible, not only to decrease our board costs, but also to minimize RF losses and to fit within the appropriate (small) enclosure, we must be careful that it is not so condensed as to increase undesired capacitive or inductive coupling between tightly spaced components or traces. This coupling can directly cause crosstalk, instability, and decreased filter stopband performance.

The capacitive or mutual inductance coupling of two traces, components, or wires must be generally accounted for in any design. This undesired coupling of RF or switching energy can be alleviated by:

1. Keeping RF or switching traces and components separated by distance.
2. Employing metallic shielding.
3. Reducing the area of the current carrying loops.
4. Using right angles between traces on different board layers.

PCB RF Connectors

Unless an RF connector is placed on the opposite side of the microstrip trace it is interfacing with, only PCB *end-launched* RF connectors should be used above 1 GHz. Employing any vertically launched connectors at the higher frequencies can create a stub effect, since the SMA's center conductor will be traveling through the substrate material to reach the other side (Fig. 13.12). This issue becomes severe with thicker substrates, causing the 50-Ω SMA connector and the 50-Ω microstrip to have an impedance bump at the connector/trace transition. This stub effect can be minimized by tuning it out by soldering a lumped capacitor in shunt to ground (a narrow-banded solution), or by milling-out or cutting-off most of the RF connector's center pin (only possible if the SMA does not need to connect

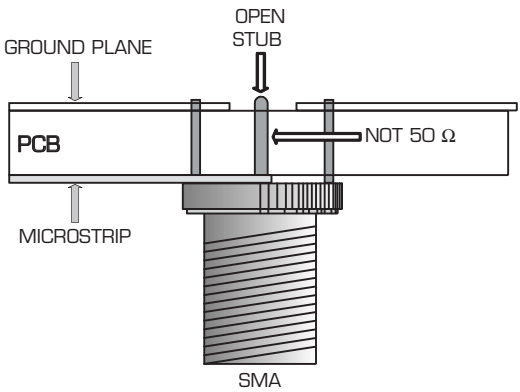


Figure 13.12 Vertical RF connector's stub effect when used at high frequencies.

to a bottom trace). We may also employ a fully surface-mount vertical SMA, but in comparison to the through-hole units, this solution will be mechanically weak.

Further, an end-launched RF connector should typically never use the pad size as recommended by the manufacturer, as this will generally add far too much capacitance, and will perturb the 50-Ω interface with the PCB's 50-Ω microstrip, drastically degrading S_{11}. The solution is to make the connector's center pin footprint pad the exact same width as the 50-Ω microstrip itself. This is especially essential when utilizing thin and/or high dielectric constant substrates.

PCB Components

Surface-mount components, as opposed to through-hole types, are all that should be considered at frequencies above a few 100 MHz. Nonetheless, not all *surface mount devices* (SMDs) are created equal when it comes to high-frequency operation. Only resistors, inductors, and capacitors that have been specified by the manufacturer to dependably operate above your desired design frequency—without the component coming close to any of its series or parallel resonances (except in certain coupling or bypass applications)—should be adopted for microwave operation.

However, since many RF resistors are not rated nor specified for a maximum recommended frequency of operation, they must sometimes be selected based solely on the type of internal architecture they employ. Usually, thin- and thick-film resistors can be depended on to reach to microwave frequencies. Still, as discussed in a previous chapter, as its resistance value increase, its ability to operate at its rated value is decreased due to parallel parasitic capacitances. In fact, for sensitive or very high-frequency circuit operation, candidate RF resistors should always be bench-tested to inspect their resistance changes versus frequency, and selected accordingly. For instance, it would be impossible to employ a 10,000-Ω resistor at 5.8-GHz, and then expect that this resistor will actually be anywhere near 10,000 Ω in value at this frequency. Its true resistance value, as seen at 5.8 GHz, will be more like 5000 Ω; and then only if we had selected a high quality 0402 thin-film resistor, *and* used the proper PCB layout.

While all active and passive components that are subjected to RF must be able to properly operate at our desired frequencies, any component that is out of the actual path of the RF does not. Consequently, in order to save cost, almost any generic low frequency component may be used in most AC and DC circuits (except when ESR or

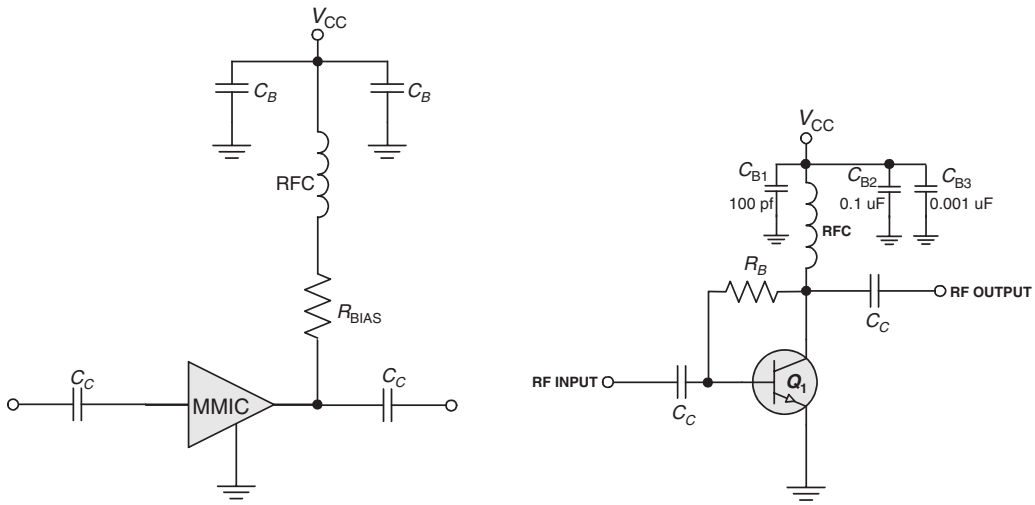


Figure 13.13 Proper decoupling for a MMIC, and the proper DC bias RF capacitor decoupling for a transistor.

equivalent series inductance [ESL] is an issue). This becomes especially meaningful in DC bias circuits (Fig. 13.13) where the RF choke (RFC) must obstruct most of the RF from entering the DC bias supply, with any RF that does get through being bypassed by the high-frequency capacitor of C_{B1} to ground, while C_{B2} and C_{B3} need not, and indeed cannot, operate at microwave frequencies due to their high capacitance values, and consequently high parasitic series inductance. However, all other components in this circuit, such as C_C, R_B, and RFC, must be capable of operating effectively up to the circuit's highest RF frequency (sometimes even up to its harmonics).

A major issue that can be overlooked by the uninitiated, is that the exact same manufacturer and model of passive and active component must always be employed throughout the entire production run for any RF/microwave wireless device. Indeed, we must also retain the same substrate material, thickness, and layout all the way from the product's simulation stage to the very last manufactured device rolls off the assembly line. The reasons for this are quite basic to all of RF design: if any component is changed or modified, then instability and a major alteration in RF performance of the microwave circuit can be expected (again, this does not apply to most AC and DC circuits). This is because of the variation in the particular component model's innate parasitics when changing from one part's model to another, or from one manufacturer to another. Since even the same model's component tolerance can prove a concern at microwave frequencies from part to part and lot to lot, such an extreme change as switching manufacturers or part numbers in midstream could prove catastrophic to any wireless microwave design. Therefore, a component change to another manufacturer's part, or to a similar model, will require retesting and possibly retuning of the RF circuit. An even more major change, such as multiple RF passive component substitutions, or the swapping of an active RF device for another model, or the switching of substrate materials or thicknesses, should automatically invoke a full run of bench tuning and performance retesting, not to mention the possibly significant EMC and FCC testing implications.

When laying out or assembling a complete microwave printed circuit board, no detail must be overlooked: The correct placement of RF inductors will minimize yield issues that may be experienced with such high-frequency RF designs. Therefore, the inductor's polarity mark, sometimes located on the inductor's package to decrease the possibility of magnetic coupling effects between neighboring close-in (unshielded) ceramic RF inductors (or the top of each inductor if no polarity mark is present), should always be facing upward when inserted onto each and every PCB. The reason for this becomes clear when you consider that not all RF ceramic chip inductors have their metallization layers spread evenly throughout the entire SMD package. In many instances, the inductor's metal layer is only on the top of the ceramic package itself. So, by varying whether the top of the inductor is facing the air above, or the substrate below, its inductance value may change by up to 20% depending on this single, often overlooked, factor.

Microwave Matching

Keep any open stubs short as compared to the wavelength of our signal of interest. These stubs may have been unintentionally formed by a trace and pad that is connected to the main microstrip, but is devoid of its component. If we do not minimize or eliminate this stub length, this will create a different complex impedance on the line than we had originally predicted. In fact, the open stub effect can even be formed if the trace itself does not *appear* to end in an open, simply by the use of an RFC (or other high-value inductor) too far up the trace in a matching or decoupling circuit (Fig. 13.14). Depending on the stub's exact length and width, it will perturb the expected match to varying degrees. When using an RFC, at its worst, this stub may be exactly 90° long, thereby removing our signal of interest almost entirely.

For a repeatable microwave design, it is important to maintain an RF-matching component's exact physical location on the PCB during its entire production run. This is again due to the correspondingly short wavelengths that we must contend with at microwave frequencies, which demand precise placement of all lumped components during PCB assembly. Ignoring this requirement at these elevated frequencies will create an alteration in the RF circuit, even with small changes of only a few 10s of mils shift from a part's original location. This is clearly demonstrated in Fig. 13.15, where an L network has been moved laterally along the microstrip by only 100 mils from its original and as-designed matching location, degrading the once excellent match. (Separating L from C by distance will also degrade the match.) In fact, a common microwave technique used to optimize an impedance match is by actually shifting a component along a *tuning rail*, which is simply a bare copper section of clear microstrip that allows us to move the component a certain distance forward and backward in order to optimize a match for an ideal S_{11}/S_{22} or P_{OUT}.

Substrates

Most low-cost, high-volume manufactured wireless products will consist of a PCB with a dielectric manufactured completely of FR-4 material. This is a pragmatic fact of life, and most wireless consumer goods, even up to the ISM band of 5.8 GHz, will only employ FR-4 material. To be competitive in today's market, we are usually prevented from using any other board material, as the price disparities are simply too great. Even when using FR-4, we may be limited to a maximum of two, but not more than four, PCB layers due to these same cost pressures. (See Sec. 13.2.2.)

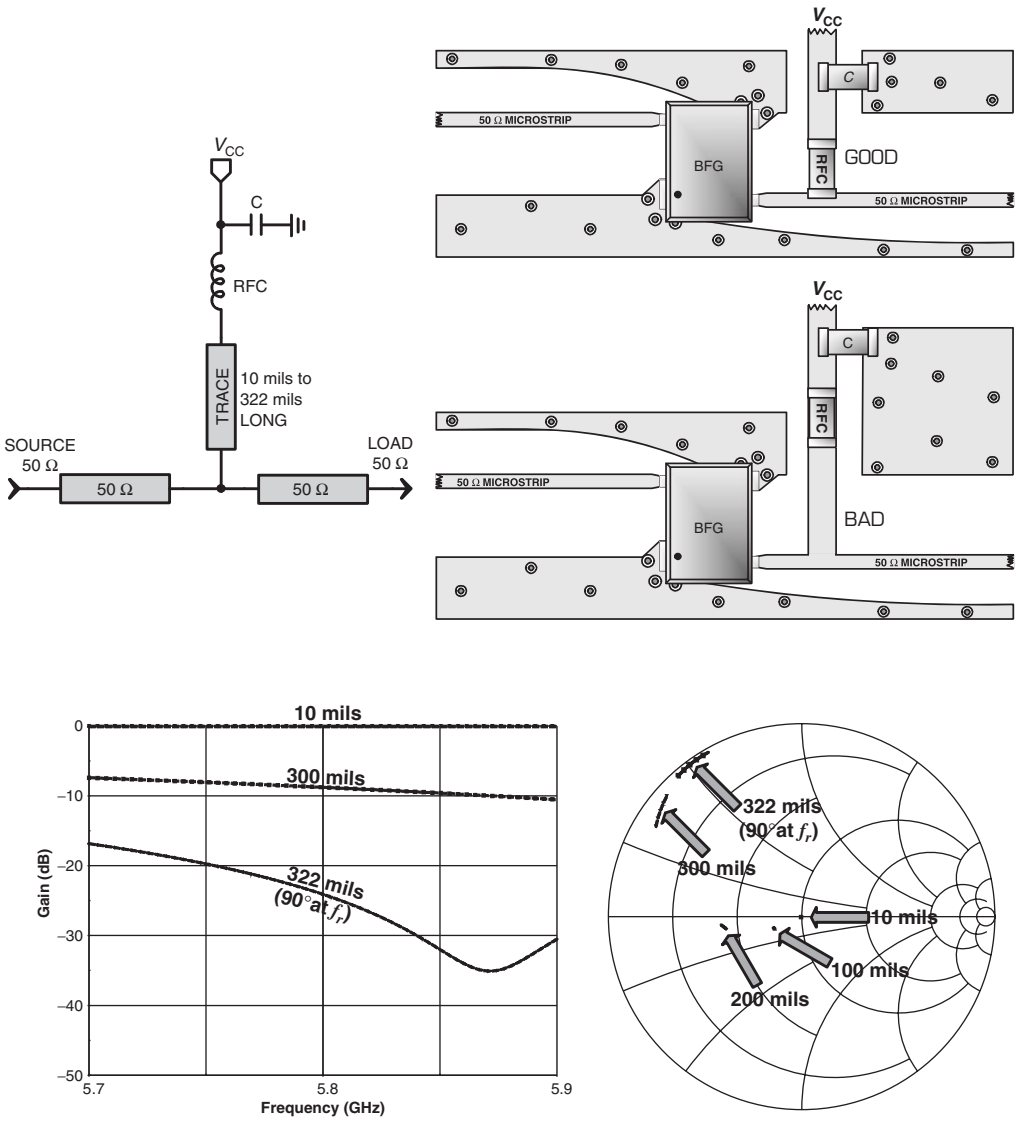


FIGURE 13.14 An RFC located at five different distances from its 50-Ω microstrip on a microwave PCB layout, and the effects on the circuit. If an RFC (or matching inductor) is to be positioned any significant fraction of a wavelength away in a high-frequency layout, it must be so simulated while using the proper trace lengths.

The reason for exploiting something other than FR-4 for our high-frequency signal carrying layers is because superior materials, such as Roger's RO4003, have lower RF losses, better dimensional tolerances, more closely controlled dielectric constant variations, and possesses better thermal properties. And in the case of RO4003, this Roger's material is completely compatible with basic FR-4 processing methods and procedures, and can also be safely and reliably laminated together with FR-4 at almost any PCB fabrication house. In fact, this method of using mixed dielectric materials on high-frequency microwave

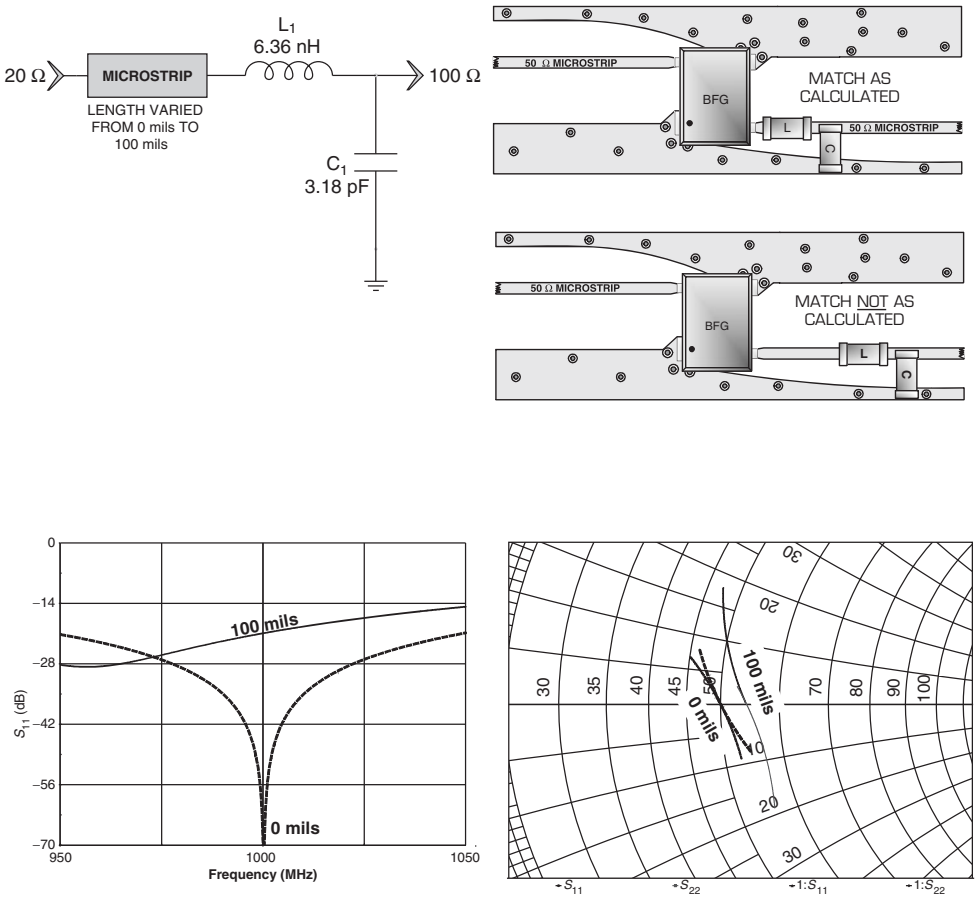


Figure 13.15 An L-matching network located at two different distances away from its 20-Ω load on 50-Ω microstrip to match a 100-Ω load, and the effect this distance has on the quality of the calculated match.

and RF-printed circuit boards will increase costs somewhat over pure, name-brand FR-4 dielectrics, but performance and repeatability from board to board will be much enhanced, and impedance control and RF losses improved (RO4003 has a decibel/inch insertion loss at 2.5 GHz of approximately 0.05 dB, while generic FR-4's is 0.2 dB). We can also specify our high-frequency RO4003 layers to be quite thin, while using the generic FR-4 as the thickest and stiffest part of the entire PCB stackup, which will decrease costs over employing a pure RO4003 board (less high-frequency material is expended).

In high-end industrial, exotic consumer, or military RF communication devices, better substrates may be the order of the day, but only if they are truly needed to meet specifications, or if they are implicitly called out by the customer or the particular market segment. So even in demanding military and industrial markets we must forever attempt to keep costs down, yet without substantially sacrificing quality, performance, or reliability. Consequently, there are many such applications that we can and must utilize nongeneric, name-brand FR-4; only when we absolutely cannot meet the designated RF design goals can we then go on to more advanced, and much more expensive, board materials.

Trace Widths and Vias

Using 0.5-oz copper in a 25°C ambient environment, a PCB can safely carry up to 1 A of current on the outer PCB layers per each 15-mil-wide trace. Due to the insulating properties of dielectrics, this width must be increased to 40 mils for the inner layers. Increasing the copper's thickness to 1 oz will permit us to decrease this trace width to 8 mils for the outer, and 20 mils for the inner, layers. To safely tap this much current off of a power plane, there should be at least five vias used (less if the vias have exceptionally large diameters). Therefore, lay down enough vias and wide-enough copper traces to guarantee that there will always be a low-resistance connection, with safe thermal characteristics, during any high-current applications.

Board Thermals

Only select *thermal pads* (Fig. 13.16) for a part's footprint when using larger, lower frequency parts that will be soldered directly to the groundplane, such as electrolytic filter capacitors. Due to the impedance added by these wagon-wheel/spoked-shaped structures, never employ thermal pads for any RF/microwave component's footprints, in which case the pad should only be attached directly to the ground pour. If this is not possible, and soldering has become difficult and unreliable due to the heat dissipation inherent in a large groundplane, then select the thickest possible thermal spoke pattern. This will at least minimize the impedance to ground over that of a thinner spoked pattern. The above also applies equally well to vias that connect to the inner and outer copper planes of PCBs; at no time should any RF via contact these planes through spokes (called *via thermals*), but only through a solid, low-impedance connection to the metal itself (Fig. 13.17).

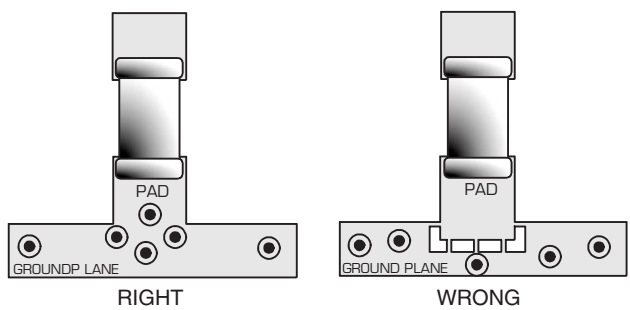


Figure 13.16 Component pads both with and without thermals. Thermals of any type should never be used anywhere in the path of RF frequencies. (Use in the AC or DC sections of an RF circuit is acceptable.)

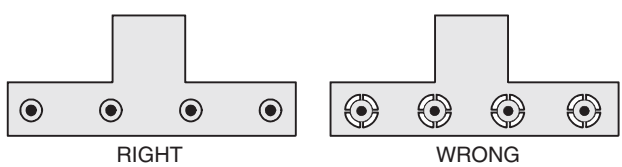


Figure 13.17 Direct via and thermal via connections to a copper plane. Thermals of any type should never be used anywhere in the path of RF frequencies.

Digital/Analog Isolation

Functional partitioning on the same PCB to separate the noisy digital sections from the more vulnerable analog sections can be accomplished by bisecting the two with RF shielding (around one or more sections), grounded fencing vias (at one-twentieth of a wavelength spacings and passing completely through the board), RC filtering (on all high-impedance digital lines), strong AC and RF decoupling (on all DC lines), separation by distance (of the digital and analog sections), or even using two completely different circuit boards (connected together via cable or special board-to-board connector).

Low-Cost PCB Design

The lowest cost consumer PCB will have components on only a single side; no more than two layers (four layers maximum); a minimum number of through-hole vias; no blind or buried vias; FR-4 substrate material; and is laid out within the standard trace/pad/part's tolerance rules of your particular PCB board house.

A modern board house's standard, no-premium fabrication rules and tolerances are shown in Table 13.1, along with the more advanced, but more expensive, cutting-edge fab capabilities—as well as capabilities that are just over the horizon.

13.2.5 Printed Circuit Board Design for Miscellaneous Circuits

This section addresses some challenging layout issues, such as that for digital, RFIC, SAW, and switched mode power supply (SMPS) circuits and devices.

High-speed Digital PCB Layout

A digital signal's rise and fall times, and not its frequency, govern the signal's speed as it relates to PCB design. The speeds of many of today's digital circuits will demand the same layout skills as that employed in high-frequency RF PCB design. Thus, if we need to integrate a high- or a low-speed digital control logic circuit onto our own analog wireless layout, we will be primarily concerned with three things: minimizing digital EMI far-field radiation into other analog systems/circuits; decreasing close-in capacitive and inductive coupling, called *crosstalk*, into other components, traces, or circuits within our own PCB; and preventing excessive distortion to the digital waveform. Therefore, to produce a trouble-free high-speed digital layout we must:

1. Force all digital signal traces to run near the digital groundplane, minimizing the current loop area, while employing a thin PCB substrate for a higher capacitance.
2. Use only 50-Ω microstrip for any digital tracks that are over 2 in/ns long, and that are properly terminated into 50 Ω. (As a general rule of thumb to avoid transmission line effects for digital signals, the maximum track length in mm should not be more than 46 times the fastest rise or fall time, in ns.)
3. Avoid any significant runs of trace or microstrip that are situated off the main signal track, and that terminates into a high impedance or an open circuit (such as into an RFC or an unstuffed PCB pad), as this forms an open stub which can act as an undesired bandstop filter at a frequency of $V_p(\lambda/4)$, or can cause severe impedance mismatching.
4. Minimized high-impedance nodes, since they will readily add noise from other radiated EMI sources to the digital circuit.

Streamline Circuits	Standard	Advanced	Emerging
Panel Size	18" × 24" & 19" × 25"	20" × 26" & 24" × 30"	
Layer Count	2 to 36	36+	50+
Laminate Material			
FR4 Tg 140	Yes	Yes	Yes
GETEK	Yes	Yes	Yes
Rogers	Yes	Yes	Yes
RoHs Materials	Yes	Yes	Yes
Exotic Material Types			
Finished Thickness Tolerances	0.005" to 0.250"	>0.250"	See Advanced
Finished Thickness (Multilayer)	0.008" to 0.220"	0.220"–0.350"	>0.300"
Minimum Core Thickness	0.003"	0.002"	0.002"
Finished Thickness Tolerance (+/–)	10%	7%	5%
Multiple Laminations	5	8	8+
Copper Foil Weights Internal	1/4 to 1/2	<0.02 Oz.	<0.05 Oz.
Copper Foil Weights External	1/4 to 3	4 to 5 Oz.	Up to 6 Oz.
Lines, Spaces & Pad Diameters			
Internal Line Width	0.0045"	0.002"	<0.003"
Internal Spacing	0.005"	0.002"	<0.003"
External Line Width	0.004"	0.002"	<0.003"
External Spacing	0.004"	0.002"	<0.003"
Int. Pad Size-A/R Per Side (Fin.-.001)	0.005"	0.004"	<0.004"
Ext. Pad Size-A/R Per Side (Fin.-.002)	0.003"	0.003"	<0.003"
SMT Pitch	0.10"	0.010"	<0.010"
Impedance	10%	2.50%	2–2.5%
Electroplating			
Tin Lead Plating Thickness	Copper Coverage	0.003–0.0015	<0.0015
Gold Plating Thickness	30 µ in	As specified	As specified
Minimum Drilled Hole Size	0.012	0.0098"	<0.0098"
Hole Aspect Ratio	10 to 1	15 to 1	20 to 1

(Streamline Circuits, Inc.)

Table 13.1 Modern PCB Fabrication Capabilities, Present and Future

Streamline Circuits	Standard	Advanced	Emerging
Conductor Finishes			
HASL	Yes	Yes	Yes
Solder with Reflow	Yes	Yes	Yes
White Tin	Yes	Yes	Yes
Lead Free			
Electroless Nickel	Yes	Yes	Yes
Immersion Gold	Yes	Yes	Yes
Immersion Silver	Yes	Yes	Yes
HASL	Yes	Yes	Yes
Tolerances			
Drilled Hole To Copper	0.008"	0.005"	<0.005"
Plated Hole Tolerances (+/–)	0.003"	0.002"	<0.002"
Non Plated Hole Tolerances (+/–)	0.001"	0.001"	<0.001"
Fabrication Tolerances (+/–)	0.005"	0.003"	<0.003"
Via Hole Finish			
Laser Micro Vias	0.004	0.003	<0.003
Blind/Buried Vias	0.004	0.003	<0.003
Laser Drill	0.004	0.003	<0.003
Micro Via Mechanical Vias	0.0059	0.0047	<0.0047
Soldermask and Legend			
Minimum Mask Clearance (LPI)	0.003"	0.0025"	<0.002"
Minimum Soldermask Thickness	0.0004	0.0004	0.0004
Legend Feature Size	0.008" Wide × 0.030" High Min.	0.006" Wide × 0.030" High Min.	N/A

(Streamline Circuits, Inc.)

Table 13.1 Modern PCB Fabrication Capabilities, Present and Future (*Continued*)

5. Insert extra groundplanes between the digital signal tracks of a multilayer board to effectively shield other digital or analog traces from any electromagnetic coupling, significantly reducing crosstalk. (It can be prudent to check all traces for crosstalk on a PCB design at the early layout phase. This can be accomplished by employing field solver or special crosstalk software.)
6. Keep in mind, and minimize, any openings made for connectors, LEDs, switches, or thermal vents within metal enclosures or groundplanes, since they may act as slot antennas, especially if the opening is greater than

one-tenth of a wavelength. (These slots will then radiate the fundamental and/or a harmonic frequency.)

7. Keep the system power supply separated from the digital and analog sections by very strong wideband decoupling on both sides.
8. Use self-shielded parts, if available, for any strongly radiating digital ICs, such as that for the clocks, as these harmonics can sometimes reach orders as high as 28*fr*, causing significant ingress emissions into sensitive RF circuits.
9. If required, employ local metallic RF shielding over and around any digital ICs or circuits that may radiate interference, as EMI can emanate from within a digital IC on even its low-frequency or DC pins. (The EMI itself is escaping the IC's package through internal crosstalk and ground bounce.)
10. Attempt to physically separate all high-speed digital circuits from the sensitive analog circuits. Located the DC and RF traces as far from all digital traces as possible.
11. Try to time the digital circuits to switch only when the analog circuitry will not be adversely affected.
12. Use local decoupling of digital circuits with sufficient values of capacitors to mitigate the effects of digital switching, since during this switching action there may be very high peak current requirements, which can cause rapid current pulses on the voltage supply lines, thereby producing sneak digital emissions into other circuits.

RFIC and MMIC PCB Layout

To prevent possible instability, decreased gain, and poor *P*1dB, an RFIC's own ground leads must minimize all undesired inductance to the PCB's top and bottom groundplanes. We can lessen the inductance by placing the RFIC's ground pins directly over the board's through-hole ground vias, but not so close that we create a *solder wicking* problem. (Solder wicking, caused through capillary action, can draw much of the joint's solder from the pad/pin, and into the via itself, forming a substandard electrical and mechanical contact.) Indeed, as a matter of basic, proper microwave and RF design, all grounds on the top side of a board must lead directly to the bottom groundplane by the shortest possible route. This shortest possible route is almost always through a direct through-hole via to ground. In fact, to lessen the inductive properties of multiple vias to ground even further, some high-frequency chips will demand that lower board layers not be any thicker than 32 mils from the component side down to the main groundplane. (As an example of why this is so, we may want to consider that a single via passing through a 32-mil board can have an inductance of approximately 0.15 nH, while a via passing through a 62-mil board may have an inductance of up to 0.5 nH. The 32-mil board thus demonstrates significantly less via inductance to any RFIC that is placed on the top of the PCB.) To decrease this inductance to ground to the lowest possible level, it is always advisable to use more than one ground via per component pin.

Most consumer microwave IC packages today can be found not only with conventional ground pins, but also with a single, large metal ground "slug" located under and inserted within the IC's own plastic base. This slug is used both as a low-impedance electrical and a low thermal resistance ground point that must be soldered to the PCB's metal layer. However, if such an RFIC possesses multiple internal amplifier stages, then extreme caution must be exercised during its layout, because if we make

the grounding too good for *some* (but not all) if these chips, we may actually be forcing them into instability. This can be due to the following cause: by creating an excessively good ground for all of the RFIC's ground pins, what we are actually doing is connecting all of the chip's internal stages directly together on the top groundplane, which has an ultralow impedance to RF. Therefore, each of the IC's internal stages are now effectively connected to each other through a mutual, and virtually zero impedance, connection. This can create positive feedback from one of the IC's active stages back to the first active stage, and without any really significant intervening mechanism to lessen this feedback. This may produce oscillations. However, by permitting each of the IC's separate ground pins and the ground slug itself to have their very own vias, which would now go straight down to the bottom groundplane, and if we do not have the 0-Ω top ground pour layer shorting all of the chip's ground pins together, then the now individual and dedicated vias to the PCB's ground will add just enough isolation (impedance) to the flow of this RF feedback that the loop will be attenuated between all of the chip's internal stages, eliminating any chance of feedback oscillations.

Even if an RFIC (or discrete amplifier) is designed to be unconditionally stable, poor PCB layout can cause severe stability problems. (Consider for a moment that all properly designed RF oscillators are utilizing only completely stable RF amplifiers as their core active elements, with only some (purposeful) external LC feedback elements used to force these unconditionally stable amplifiers into oscillation). Indeed, sub-band oscillations (below the amplifier's bandwidth), out-of-band oscillations (above the amplifier's bandwidth), and in-band oscillations may also occur unless the RFIC or its PCB follow these layout guidelines:

1. Grounding for a MMIC power amplifier must be highly effective, so utilize as much unbroken copper surface area as possible on the ground layer beneath the chip, along with a large number of vias from the MMIC's top layer to this main PCB ground. This is critical not only to PA stability, but since the strong circulating RF currents of the MMIC must see a virtually 0-Ω groundplane in order to prevent induced voltages from forming, this decreases EMI and noise radiation as well.
2. The V_{CC} for the RFIC and all other active components must be properly decoupled by using both low- and high-value capacitors for both high and low frequencies.
3. Certain RFICs with high-gain stages may need to be shielded to prevent them from bursting into oscillation.
4. To avoid bypassing the RFIC or creating feedback, the chip's input and output traces should be kept separated from each other.
5. To bias and match the chip, use only components designed for RF service, and that are rated to operate effectively over and above the maximum frequency of your circuit.

Switching Power Supply PCB Layout

Switching power supplies require very effective layout procedures to minimize EMI radiation, maintain regulation, and eliminate instability. These issues are especially acute when a switcher must provide high-current levels to a load.

The switcher's feedback traces are extremely prone to EMI, and should be run far from any noise sources, such as the switching inductor, with as small a loop and with as wide a trace as possible. Typically this trace will be run on a different layer to isolate it from noise sources generated by the switcher or other on-board circuits. Any other switcher traces carrying significant current should also be wide and very short to minimize the loop. To decrease input trace inductance, the input ceramic capacitor should be near to the switcher's input pin to supply it with a clean, noise-free voltage. If any compensation capacitor/resistor is required, it should be as close as possible to the input compensation pin of the switcher IC to maintain stability. The output capacitors, the switching inductor, and the output diode should all be tightly grouped together to decrease EMI, noise, and voltage errors. (Use only the switching inductor model as specified in the switch mode power supply's IC application note, or EMI and noise can become a severe issue.) To decrease EMI and noise, a solid groundplane should be on the bottom layer, as well as a flooded groundplane included on the top layer, with top ground copper directly under the switcher IC itself. The chip can also use this top ground pour, and the main groundplane directly below, to help in cooling the switcher through the thermal transfer of heat through the vias.

SAW Filter PCB Layout

In wireless design, SAW filters are becoming much more common than in the past, so a strong consideration for a board's proper layout must be accommodated for these devices. To prevent an incoming RF signal from simply bypassing the SAW due to its large insertions losses, very high PC board isolation must be maintained between the SAW's input and output ports. The precise isolation required is calculated by adding the ultimate rejection of the SAW with its own insertion loss. For example, if a SAW has an insertion loss of 20 dB, and its ultimate attenuation is 50 dBc, then the PC board will be required to supply a minimum isolation of 70 dB in order to maintain the filter's rated stopband performance levels. This high isolation, while extremely challenging, can be accomplished by combining multiple methods:

1. Separate the SAW's input and output ports by using a *plated through-slot* (or use a *via fence*), which is a solid piece of thin metal effectively running through the board's dielectric, all the way down to the main groundplane. This through-slot will decrease the RF leakage propagating through the substrate material.
2. Place a metal shield around the SAW, which separates the filter's input and output ports above the PCB itself.
3. Solder the SAW at multiple points, with multiple close-in and direct vias down to the main groundplane.
4. Position the SAW's input and output port-matching inductors at right angles to each other; or use shielded input/output inductors; or place one inductor on the top layer, and the other on the bottom layer.
5. After insertion of the SAW devices, make sure that each PCB on a production line will be completely cleaned of all solder paste residue.

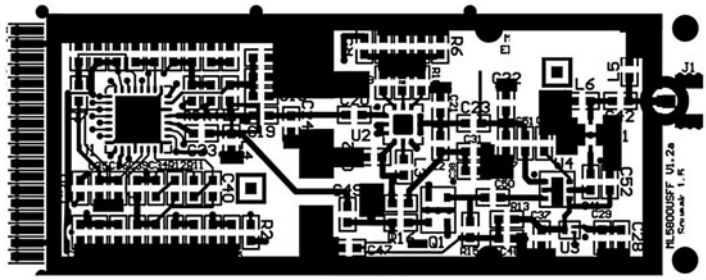
13.2.6 Printed Circuit Board Fabrication Files

After a PCB board layout has been completed and checked for all errors, it is time to create and collect the appropriate files to send on to the PCB fabrication and assembly house. Not providing all of the correct files required to complete the entire assembled board will lead to costly time delays and/or to misunderstandings; with these misunderstandings possibly leading to unintentional fabrication and/or assembly errors.

When submitting a printed circuit board design for manufacturing, the *fab house*, which is the facility that manufactures the PCB itself, and *assembly house*, which assembles the components onto this completed PCB (forming a PCA, or printed circuit assembly) will require:

1. *Gerber files* in 274X format, which contain the data that makes it possible to manufacture the PCB (such as copper geometry, size, and position). We can actually "read" the contents of any Gerber file by using a *Gerber viewer* [Fig. 13-18(*a*)]. (Gerber viewers are available free on the Web, and are invaluable for checking out your production ready PCB file for any errors.)
2. *Bill of materials*, or BOM, is a file in Excel or PDF format that contains all the schematic's part descriptions, values, reference designators, manufacturers, and so on [(Fig. 13.18(*b*)]. (All parts within the path of RF must be marked as permitting no substitutions. If not, then the assembly house may substitute generic, low-frequency components if they happen to run out of the exact model part you so carefully specified in your BOM.)
3. *Assembly drawings* file in PDF or DXF that clearly display the PCB's silkscreen, with its reference designators, Pin 1 designators, part's outlines, and so on, along with any special warning/instructions for the assemblers [Fig. 13-18(*c*)]. (Always specify that you do not permit any component substitutions without prior written permission.)
4. *Automatic placement data* (also referred to as *centroid data* or *pick-and-place data*) is an ASCII file generated within the PCB's layout package, and supplies the automated assembly equipment with each part's centroid (center of part) data for programming of the pick-and-place assembly robots used for the overall PCA.
5. *NC drill files* give all the PCB hole sizes and locations for the automatic via drilling machines.
6. *Netlist file*, required only if PCB bare board testing will be performed. (This testing will use either flying probe testers (low volume) or bed-of-nails testers (high volume), and will be able to inform the board house if the PCB has been properly fabricated, but only on a DC connectivity level.)
7. *Fabrication drawing* file contains the PCB's dimensions, stack-up, thickness, metal and dielectric composition, plating, size and location of tooling holes, cut outs, fiducial marks, along with any important notes [Fig. 13.18(*d*)].

a.

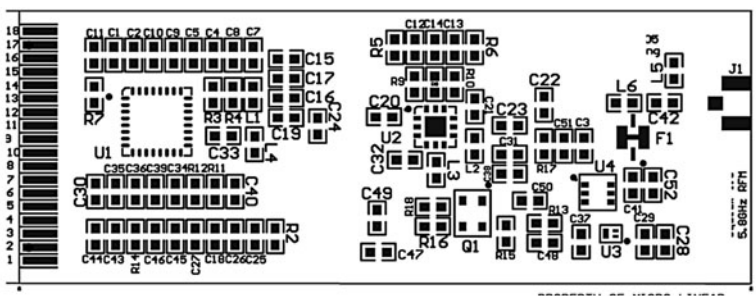


b.

V1.2a ML5800USFF BOM 3/30/05

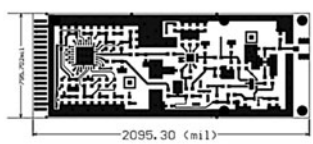
ITEM	QUANTITY	DESIGNATOR	DESCRIPTION	MANUFACTURER	PART NUMBER	VALUE
1	5	C1, C10, C15, C35, C37	CERAMIC 0402 SMT CAPACITOR	MURATA	GRM155R61A104KA01	0.1uF
2	7	C19, C33	CERAMIC 0402 SMT CAPACITOR	MURATA	GJM1555C1H1R2BB01	1.2pF
3	12	C2, C3, C8, C14, C16, C20, C29, C41, C42, C48, C51, C52	CERAMIC 0402 SMT CAPACITOR	MURATA	GJM1555C1H5R1CB01	5.1pF
4	3	C4, C28, C45	CERAMIC 0402 SMT CAPACITOR	MURATA	GRM155R60J105KE19	1uF
5	12	C5, C7, C9, C11, C12, C17, C26, C27, C32, C36, C44, C46	CERAMIC 0402 SMT CAPACITOR	MURATA	GRM1555C1H101JD01	100pF
6	3	C13, C34, C38	CERAMIC 0402 SMT CAPACITOR	MURATA	GRM155R71E682KA01	6.8nF
7	3	C18, C25, C43	CERAMIC 0402 SMT CAPACITOR	MURATA	GRM155R61A224KE18	0.22uF
8	6	C21, C22, C24, L4, L6, R17	CERAMIC 0402 SMT CAPACITOR, INDUCTOR, SMT RESISTOR 0402	MURATA, PANASONIC	N/F	N/F
9	5	C23, C31, C47, C49, C50	CERAMIC 0402 SMT CAPACITOR	MURATA	GJM1555C1HR75BB01	0.75pF
10	2	C30, C39	CERAMIC 0402 SMT CAPACITOR	MURATA	GRM1555C1H331JA01	330pF
11	1	C40	CERAMIC 0402 SMT CAPACITOR	MURATA	GRM1555C1H390JZ01	39pF
12	1	F1	CERAMIC BPF	MURATA	LFB215G78SG8A170	BPF
13	1	J1	UFL CONNECTOR	HRS	U.FL-LP(V)-040	UFL
14	2	L1, L2	INDUCTOR	MURATA	LQP15MN3N9B02	3.9nH
15	2	L3, L5	INDUCTOR	MURATA	LQP15MN2N2B02	2.2nH
16	1	Q1	SMALL SIGNAL RF TRANSISTOR	NEC	NESG2021M05	RF TRANSISTOR

c.

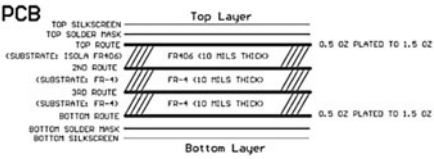


TOP LAYER SUBSTRATE TO BE ISOLA FR406 *ONLY*

THE VEGRANDIS 3 PROJECT: ML5800USFF PCB



LAYER STACK DETAIL FOR: ML5800USFF PCB



d.

REV 1.2
DRAWN BY:
DATE 2/4/08

NOTES:

1. FABRICATE PER IPC-6011 CLASS 2. (USING THIS DOCUMENT, GERBER DATA AND ASSOCIATED FILES).
2. UL RATING:94V-0 MIN.
3. SOLDER MASK MATERIAL SHALL MEET ALL REQUIREMENTS OF IPC-SM-840.
4. SOLDER MASK SHALL BE GREEN IN COLOR WITH A MINIMUM THICKNESS OF 2 MILS.
5. SILKSCREEN LEGEND SHALL BE APPLIED AS PER LAYER STACK-UP USING WHITE NON-CUNDUCTIVE EPOXY INK WITH A MAXIMUM THICKNESS OF 0.8 MILS.
6. FINISH IS TO BE ELECTROLESS NICKEL AND IMMERSION GOLD TO THE FOLLOWING SPECIFICATIONS:
   Ni -- 3 TO 8 MICRONS (120 TO 320 MICRO INCHES)
   Au -- 0.05 TO 0.15 MICRONS (2 TO 6 MICRO INCHES)
7. NO PARTS SUBSTITUTIONS PERMITTED: BOM MUST BE FOLLOWED EXACTLY.
8. NO CHANGES SHALL BE MADE TO THE GERBER FILES WITHOUT EXPRESS WRITTEN PERMISSION OF
9. THERE IS TO BE NO COPPER BALANCING.
10. PCB MATERIAL IS TO BE ISOLA FR406 FOR TOP LAYER, GENERIC FR-4 FOR OTHER LAYERS.
11. ALL LAYOUT DIMENSIONS PERFORMED IN MILS (ENGLISH).

FIGURE 13.18 Production files as would be sent to a board house for PCB fabrication; (*a*) Gerbers; (*b*) BOM; (*c*) assembly drawing; (*d*) fabrication drawing.

13.3 RF Shielding

13.3.1 Introduction

A solid RF shield (Fig. 13.19) will strongly attenuate any incoming or outgoing RF/microwave signals. Shields work by *reflecting* this high-frequency energy, as well as by turning the RF into heat, as it attempts to cross the metal boundary of the shield's structure. Shields are normally manufactured of highly conductive materials of either copper, steel, aluminum, or conductively coated plastic.

RF shields, at best, will only supply a real-world RF attenuation of about 40 dB, with the all too common perforated enclosures having lower performance. This attenuation level may be compromised even further if the shield box is not completely soldered down to the printed circuit board's groundplane, and with a continuous bead of solder around its entire periphery. The appropriate PCB ground needed for a RF shield consists of a top copper ground area, along with multiple vias that are spaced at one-twentieth of a wavelength, passing straight down through to the board's main groundplane. A further decrease in shield effectiveness may also occur if there are I/O control lines or RF traces entering or exiting the enclosure, since any trace, whether DC or AC, can easily transport EMI into or out of the shielded area. This is why it is fundamental to properly decouple all such traces with shunt capacitors, series ferrites of the appropriate frequency range or, if these traces are high-impedance digital signal lines, series 1kΩ resistors. After all of the I/O lines are properly decoupled, a shield's limited attenuation at high frequencies is caused by apertures of any kind, whether it be a hole in the enclosure's material, gaps between the soldering that attaches the shield to the PCB, or in the imperfections of the PCB's own groundplane coverage (since the PCB acts as the bottom surface for the shield box).

The aperture leakage from an RF enclosure is almost strictly governed by the aperture's longest dimension, no matter how thin that narrowest dimension may be. Therefore, any slit that is half of a wavelength long will act as a perfect slot antenna. Unfortunately, we cannot permit even this modest length of open slot in our enclosure, since apertures exhibit radiation behavior over and above that of standard slot antennas. This is due to the fluid currents being forced around the shield's open aperture, with a

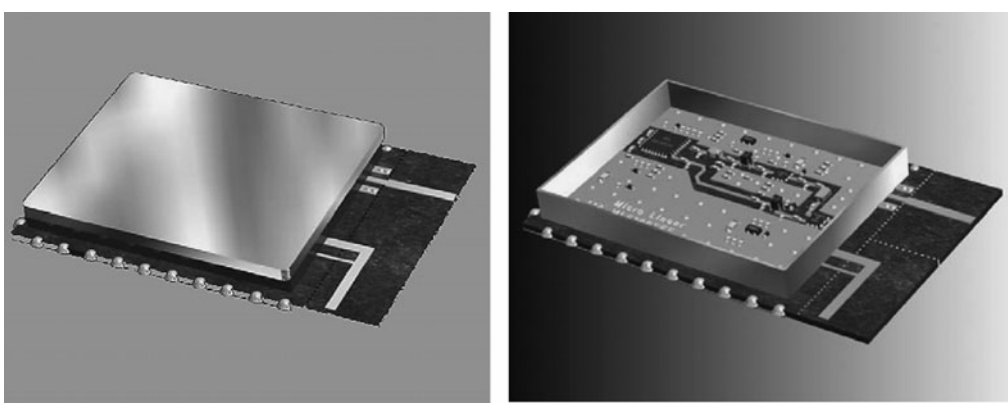


Figure 13.19 An RF PCB module with a shield's side walls and removable top cover, and with its top cover removed. (There are many single piece shields, without the separate removable top cover, available.)

voltage difference existing from one end of the aperture to the other. This dictates that, if possible, no slot or hole in the shield or the PCB metal should be larger than one-twentieth of a wavelength long.

Exploiting stripline to travel into or out of a PCB's shield enclosure is one important way to maximize shielding effectiveness, since not only is the RF line now almost completely shielded between two solid reference planes within the PCB, but the RF shield itself does not even need an aperture, as would be required with microstrip running across the top board layer. However, if we must use microstrip instead of stripline to carry the RF, we should make the shield's *mouse hole* (i.e., the actual aperture that permits the microstrip to enter or exit the enclosure) as small as possible; but not so small that the 50-Ω trace's impedance is modified by the shield's proximity, which would create an undesirable impedance bump.

Lower frequency (< 100 kHz) magnetic fields depend on metal thickness and permeability for an enclosure's shielding effectiveness. Thin copper or aluminum RF shields are not especially effective at mitigating such fields, but 40-mil-thick mild steel will be quite successful for these low-frequency magnetic fields, and functions by redirecting the undesired fields. *Mu-metal* is a costly, high-permeability magnetic shield material that is more effective than mild steel, and at even lower frequencies, but is magnetically fragile to physical impacts.

13.3.2 RF Shielding Resonances

When using RF enclosures, cavity resonances can become a real nuisance, especially at the high microwave frequencies we commonly design at today (> 2GHz). When box modes occur they cause amplifier instability, loss of filter performance, and other highly undesirable circuit symptoms. These modes arise because any conductive structure with two opposing sides will have a certain frequency that it is resonant to, which is the point when the largest dimension of the shield is one-half a wavelength, or some integer multiple. This will create, effectively, a cavity resonator waveguide effect, with little attenuation now occurring to this particular resonant frequency from one end of the shield to the other (and thus from one end of the shielded PCB to the other), along with causing varying and unpredictable impedances along the top of the board, produced by the box mode's standing waves at the specific resonances for that particular shield geometry.

When any amplifier within a shielded resonant cavity becomes unstable, or has poor gain flatness, or the RF filters on the PCB exhibit very suboptimal stopbands, a box mode can be inferred if proper circuit operation returns when we remove the RF shield from the PCB (or simply pull the shield's top cover off). Regrettably, none of this will be readily apparent when initially tuning the RF circuit itself, since we will normally accomplish this tuning without a shield in place.

To suppress such box modes (and still retain an RF shield) we can make the shield's cavity smaller, which may not be a viable option, or place metallic septums* inside the enclosure to break up the resonance. However, the easiest, quickest, and most pain-free

---

*Any septum that is placed in an RF shield cavity, and that is used to electromagnetically isolate certain stages from other circuits, can be extremely problematic. Unless the electrical contact between such a multi-cavity shield's top cover, sides, and internal septums are at much less than a quarter of a wavelength, then any small air-gap between these separate sections may cause a "slot antenna effect". This can force RF radiation to be leaked easily into the next chamber, or radiate outward into the air. Therefore, instead of employing a single stamped multi-cavity RF shield, it may be much safer to employ two (or more) completely separate shields.

Nevertheless, suppression of a box mode does not generally require the covering of the entire shield top, but rather only a small rectangular section.

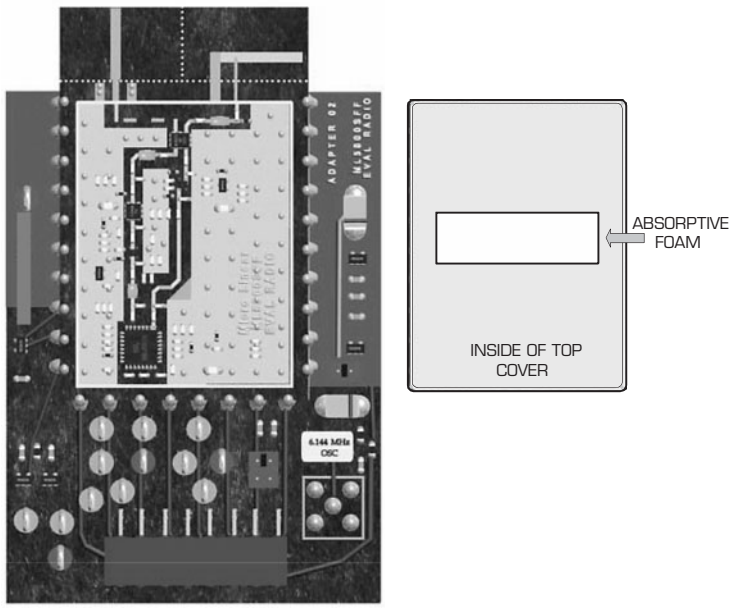


Figure 13.20 An RF PCB module showing the location of absorptive foam to break up box resonances. The optimum location of the foam will have to be found empirically.

method is in the use of *microwave absorber* materials, which are a rubber or sponge-like material that is especially manufactured to suppress a wide range of frequencies. The absorber can accomplish this feat through either the absorption or the attenuation of microwave energy. This material can be had in peel-and-stick form, thus making it effortless to apply to the bottom of a top shield cover (Fig. 13.20). The exact amount of absorber to place, and its precise location, must be arrived at empirically.

The RF/microwave absorber materials available are *magnetically loaded* and *dielectrically loaded* types. The exact type of absorber material to use is important, since the thickness of the absorber increases with a decrease in its effective frequency. This can mean that at lower microwave frequencies certain absorber materials may become too thick to be used within a particular cavity. Magnetically loaded rubber-like absorbers are made of an elastomer, and are perfect for box mode suppression since they are not only fairly thin, but are also nonconductive, and thus will not be at risk in shorting out any of the PCB's circuits. Dielectrically loaded foam-like absorbers, in contrast, are not only conductive, but are also relatively thick. They are, however, lower in cost than the magnetically loaded type.

13.4 Thermal Issues

13.4.1 Introduction

Most active devices will become damaged or destroyed if subjected to excessively high junction temperatures above that specified in their data sheets. This maximum temperature will usually range between 125 and 150°C. Therefore, we must run a component significantly below these maximum levels in order to obtain long-term circuit reliability, which will mean that we must derate all of our selected PCB

components. For instance, we should try to select an inductor or a resistor that is capable of surviving 50% more current than it is ever expected to see, and use a voltage regulator chip that can supply 25% more continuous current than it will ever need to.

The lifespan of any electronic component is directly linked to its operating temperature. We can actually quantify this generally and say that to double the lifetime of a part we must decrease its internal temperature by 10°C and, for every 10°C rise in internal temperature, the component's life will be correspondingly cut in half. In fact, most transistors in service fail due to strictly thermal issues, which can be caused by excessively high ambient temperatures, load mismatches, self-oscillations, and improper biases. (Indeed, if a silicon transistor is subjected even temporarily to excessive heat, it may either break down instantly, or may sustain a lasting but intermittent shifting of its electrical characteristics).

As mentioned elsewhere in this chapter, some discrete active devices and integrated circuits may have either a bottom metal slug imbedded in their package, or a row of pins (or both) that are dedicated not only as electrical grounds, but also to supply an efficient method for extracting internal heat from the semiconductor die. The more PCB top and bottom copper area that is connected to these slugs or pins, the more heat that can be removed, and the longer will be the life of the component. However, since the bottom groundplane normally comprises an entire solid layer of copper, it will have a much larger surface area of metal than the top layer, which means that the number of vias from the top copper ground slug attachment area, down into the bottom groundplane, is critical to proper thermal conductivity. This thermal conductivity will be further enhanced by filling up the vias with solder, which will dramatically increase their heat transfer capability. Increasing airflow with a fan or air vent is another method of cooling parts, as is the use of a heatsink. Certain critical, high-current devices, such as power amplifiers, may require all three heat mitigation methods, and may employ a large PCB copper area, forced fan cooling, and a heatsink attached directly to the PA's package.

For efficient thermal transfer, any effective heatsink requires a special grease-like compound be placed directly between the bottom of the heatsink and the top of the part's package. Since all we want to do is fill the air voids that are located between the chip and the heatsink, only a very thin coating should be applied, or thermal conductivity will actually suffer (such compounds are relatively poor thermal conductors). Advanced thermal materials have been developed that are more repeatable and effective than these old-style thermal grease compounds, and come in thin sheets of only 5-mils thick. Called *P725 Thermflow*, this product was developed by and is available from Chomerics, Inc.

Because of the readily available and popular surface-mount QFN packages for power devices (Fig. 13.21), which are rapidly replacing the much older, but more thermally conductive, metal flange type packaging, thermal problems have become a serious issue. These outdated flange-mount package styles could easily dissipate a large amount of heat from the die by means of the built-in flange that was firmly bolted, using thermal grease or gaskets, to a much larger metal structure. However, most of today's power device packages depend only on the printed circuit board itself to remove much of this self-generated heat; and the PCB's capability to do this is highly reliant on its layout geometry and the materials the board is composed of. But we can quickly approximate as to whether a particular integrated circuit is sufficiently robust to thermally survive within a specific design by means of the enclosed *AppCAD* design software (under the "*Device Thermal Calculator*" tab. Fig. 13.22).

A new component or circuit may also undergo accelerated thermal lifetime testing under high temperatures over a compressed timeframe, which will allow us to see the

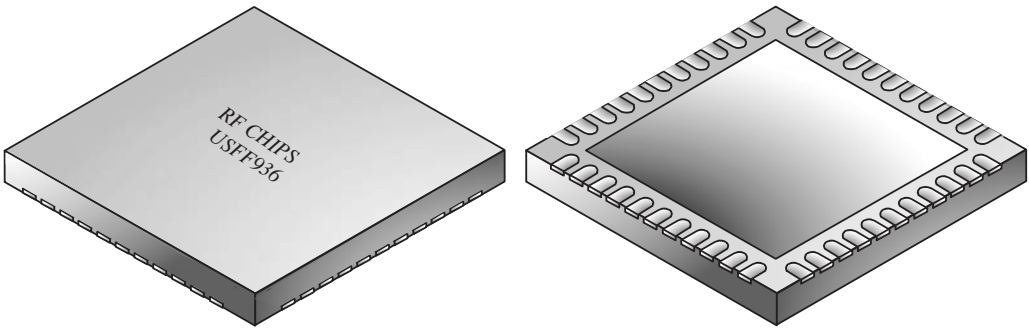


Figure 13.21 A common 44-pin QFN package with ground slug. QFNs may have almost any number of pins, depending on requirements. Top and bottom of package shown.

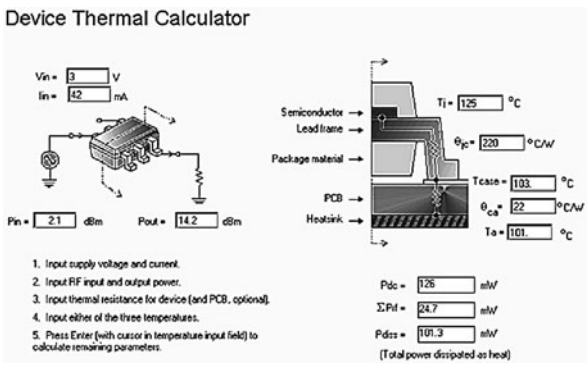


Figure 13.22 AppCAD's thermal calculator for active surface-mount devices.

part's statistical failure rate and reliability. As one of our highest and most important design goals is to produce a wireless device with as low a failure rate as possible within certain realistic cost constraints, these tests should not only include temperature cycling, but also humidity and shock. One way to perform an accelerated lifetime test is to place the DUT in a 125°C oven over a period of 1000 hours, using its maximum DC input voltage and maximum RF power. Before, during, and after this testing the DUT can be checked for any DC or RF changes, such as that for RF output power, gain, current draw, efficiency, stability, and harmonics.

13.5 Prototyping

13.5.1 Introduction

The construction of any electronic prototype encompasses many disciplines: engineer, technician, assembler, and mechanic. A properly constructed prototype can sometimes make or break a wireless project, as a prototype that will not function as expected may damage the confidence level in any new wireless concept or design.

13.5.2 Prototyping Considerations

When specifying and building a successful prototype for a wireless project, certain key issues should be kept in mind:

1. Derate components for a long, reliable, high-performance prototype lifetime.
2. Do not permit components to operate above their maximum rated thermal, voltage, current, or power.
3. Do not purchase unknown or suspect parts for the prototype simply because they are low in cost. Test and specify all parts from unknown manufacturers.
4. Confirm that components are of the appropriate value, and that they are correctly and neatly soldered to the PCB.
5. Verify a dependable supply for the selected components, as not all brand-new (or very old) lumped parts or RFICs are actually available.
6. Make sure that all DC power supplies are fully decoupled.
7. During project scheduling, allot enough time to construct not only the prototype, but also to debug the numerous operational, performance, and interface problems that will surface in any complex wireless project.
8. After initial tests and measurements, the redesign and/or reshuffling of circuits or components in a new prototype will almost certainly be needed, so only build one or two prototypes initially and then complete any further mandated units.
9. To reduce hash generation within the RF sections, confirm that all digital and analog power and grounds remain separated.
10. Establish that every trace on the PCB is correctly routed to the proper location, and without excessive undesired crosstalk or EMI risks.
11. Validate that all active devices are soldered to the correct terminals, especially diodes, and that all electrolytics are inserted with the proper polarity.
12. Check for any design or implementation flaws by placing the completed prototype under thermal, voltage, RF power, and input/output impedance stress testing.

This page intentionally left blank

CHAPTER 14
General Wireless Topics

Presented is a quick overview of software radio, direct conversion receivers, RF connectors, hybrid circuits, and FCC regulations.

14.1 Software Radio

14.1.1 Introduction

Practical digital radios, at least up to the first intermediate frequency (IF) stage, have become a reality for some advanced communication systems, typically military, as well as for lower frequency consumer radios. This has been possible thanks to advances in high-speed analog to digital converters (ADC), digital integrated mixers, and digital signal processors (DSP). However, complete software radios that can operate at high frequencies without analog down conversion are not yet feasible.

14.1.2 Software Radio Designs

The dream of a radio that can economically and quickly change from one type of modulation to another, over any desired bandwidth and frequency, by simply reprogramming it, is the promise of software radio. A complete front-to-back software radio is yet to be produced for the mass market, but the digital software back end is beginning to reach closer and closer to the RF front end of the receiver.

There are three basic types of software radios. The *software-defined radio* (SDR) type can modify a restricted set of distinct hardware functions by software, which is now already a reality in some dual-mode handset radios and base stations; *software radio* (SR) will permit some of the analog circuitry to be replaced by software, while the rest of the analog circuits will be reconfigurable by software; and *ideal software radio* (ISR), in which a limited amount of front-end analog circuitry is fixed and untunable, with the software completely controlling and changing the functions of the entire radio. Still, the ultimate software radio would have an analog-to-digital converter being fed directly by the antenna, without any intervening analog hardware whatsoever.

Since SDR-type radios are relatively common today, we will concentrate on this particular technology; but in order for SDR devices to become truly pervasive, these architectures will need to become cheaper, smaller, and more configurable (i.e., easily switch to different bands, protocols, bandwidths, and modulation schemes).

The majority of SDRs consist of a transmitter with an analog filter, solid-state power amplifier (SSPA), frequency conversion stage, and the digital stages (Fig. 14.1). The receiver section is constructed of an analog filter, low-noise amplifier (LNA), frequency

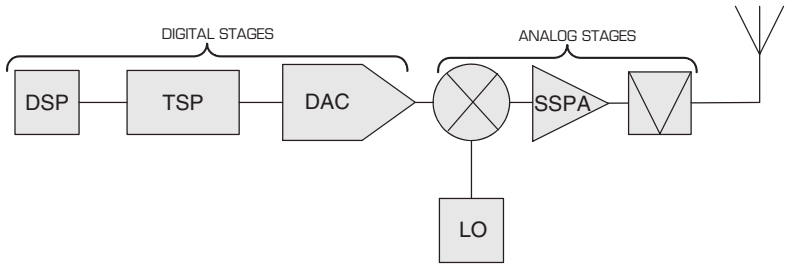


Figure 14.1 A software-defined radio transmitter.

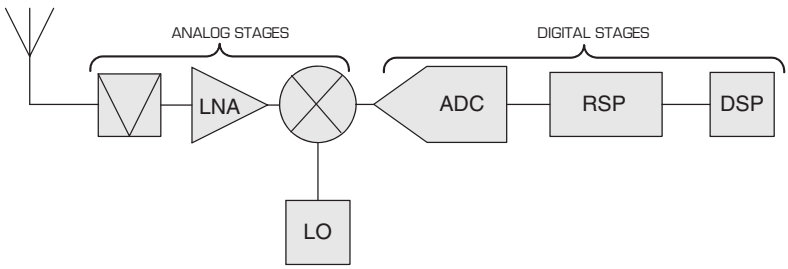


Figure 14.2 A software-defined radio receiver.

conversion stage, and digital IF and baseband functions incorporating signal processing hardware algorithms (Fig. 14.2).

There are two different kinds of SDR radios. The first type is called *heterodyne SDR*, as shown in Fig. 14.2: The antenna receives the RF signal, which is then filtered, amplified, and converted to IF by these analog stages. A very fast ADC then changes the analog IF into a digital signal, where the *receive signal processor* (RSP) filters and tunes the signal to the selected channel to provide baseband I and Q outputs into the DSP. This allows various channels, frequencies, and standards to be placed within a single radio. (In fact, the RSP is where the SDR radio has its tuning and selectivity, data rate, channel bandwidth, and even channel shaping abilities, and actually supplants the local oscillator (LO), channel select filter, quadrature mixer, and data decimation filter.) The RSP is essentially a special purpose DSP, and must be capable of the same speed as the ADC itself. The output of the RSP is then sent on to the DSP, which demodulates the baseband signal. The DSP, depending on its programming, can demodulate both digital and analog signals, and is able to receive FM, AM, QPSK, CDMA, and the like.

To really make quality software radios possible, a fast, low-noise ADC with wide dynamic range is a requirement. Indeed, this ADC must have an excellent signal-to-noise ratio (SNR) rating, which represents both quantization and thermal noise, as well as the analog-to-digital converter (A/D, also known as ADC) sample clock's wideband phase noise. The high dynamic range is also a must in order to decrease spurious responses, since within any multicarrier radio the responses of one channel can interfere with the weaker signals in another channel.

The transmitter of a heterodyne SDR system, as was shown in Fig. 14.1, functions so: The DSP places digital data to be modulated into the *transmit signal processor* (TSP). The TSP then modulates the carrier with this information. The digital modulated signal

is transformed to analog by the DAC, sent into the analog stages and frequency up-converted to RF by the mixer/LO, amplified by the SSPA, filtered, and sent out of the antenna.

The second type of SDR architecture is called *direct conversion software defined radio (DC SDR)*. DC SDR radios reduce much of the heterodyne system's disadvantages, such as the need for a front-end filter (used to set the radio's frequency, its maximum bandwidth, and to remove the image band), and converts the incoming RF directly into a zero IF. This removal of the front-end filter requirement is an important feature of a receiver that must be completely configurable over multiple bands, as RF filters that are frequency and bandwidth tunable over a wide range are extremely difficult and expensive to design and mass produce. As a result, for each wireless standard the heterodyne type SDR radio would need an expensive and large switchable filter bank, along with a fine-tunable LO, while the DC SDR completely removes this problem altogether, significantly lowering system costs.

14.2 Direct Conversion Receivers

14.2.1 Introduction

Direct conversion receivers (DCRs) (also called *zero-IF receivers*; Fig. 14.3), have seen limited use in the past due to implementation complexities. A DCR architecture comprises a receiver without an IF, since the input RF is directly mixed down baseband.

DCRs have a much lower parts count, and are thus cheaper to build, than the competing superheterodyne designs. DCRs do not require an image filter, since no image frequency is seen by the DC receiver, nor do they need high frequency IF filters and amplifiers, since there is no IF. However, direct conversion receivers have multiple problems that make a discrete DCR design almost impossible, while even RFIC implementations are rife with difficulties. Nonetheless, many of the DCR's limitations can, and have been, addressed relatively successfully within the domain of some of the newer RF-integrated circuits.

14.2.2 Direct Conversion Issues

Most design choices have tradeoffs, and a direct conversion receiver is no different. Superheterodyne receivers have more selectivity and sensitivity than an equivalent direct-conversion receiver, with most DCRs barely able to function up to 900 MHz, while attaining only −95-dBm sensitivity (at least −105 dBm is a requirement in many

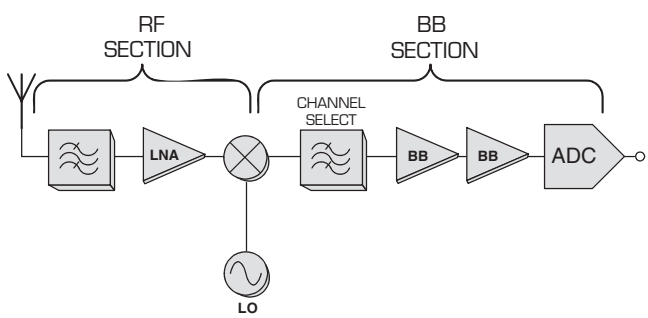


Figure 14.3 A direct conversion receiver block diagram.

systems). Selectivity naturally suffers due to the DCR's adoption of active lowpass filtering at baseband, giving direct conversion less interference rejection than the highly filtered superheterodyne receivers. Many direct conversion receiver RFIC's, as well, will have the following problems:

1. An incidental offset voltage is caused by the self-mixing in the direct conversion receiver; the LO leakage actually mixes with the original LO signal creating a DC voltage which can contaminate the signal of interest, lowering the SNR, and can even saturate the baseband amplifier stages.
2. LO leakage through the RF sections is a large consideration in design, since the LO is very close to the frequency of the incoming RF, and can be radiated by the receiver's antenna, causing in-band interference.
3. As the down-converted RF signal is usually of low amplitude and low frequency (near 0 Hz), flicker effect noise (*1/f*) from the mixer output decreases the SNR.
4. In very wide-band and high-frequency DCR receivers, the difficulty of maintaining equal amplitude and phase in both the I and Q legs (called *I & Q mismatch*) will cause an increase in the BER.
5. Any circuit that employs phase shifts of the incoming signal by 90°, as common zero-IF receivers do, undergo compromises with noise, linearity, and power.
6. Strong signals at the input mixer creates distortion, which reduces the DCR's sensitivity more quickly than with superheterodyne radios.

Up until recently the above problems have relegated DCRs to simple, undemanding use in devices such as frequency shift keying (FSK) pagers and a few amateur single-sideband (SSB) receivers. But RFIC companies have attempted to get around many of these difficulties by some ingenious design techniques. For instance, the above DC offset issue can be mitigated by transitioning from a direct conversion, zero-IF architecture to a very low, but not zero, baseband frequency by exploiting an LO that is not at the same frequency as the RF. This allows the operation of AC coupling into the mixer stage, which eliminates the DC offset. However, costs can rise somewhat due to the need for a higher frequency ADC. And the LO leakage problem from the direct conversion receiver's front end can be largely mitigated by a high reverse isolation LNA. Many other improvements in DCR design will be forthcoming in the next few years. But despite these implementation problems, DCR receivers are becoming popular in low-cost wireless designs simply because they require fewer components than the standard superheterodyne technology, and are therefore much cheaper to build.

14.3 Hybrid Circuits

14.3.1 Introduction

Hybrid design and construction may predate today's monolithic ICs, but they have not been completely supplanted by integrated circuits in all applications. However, hybrids are now used only in special applications where integrated circuit or discrete SMD PCB designs would be inferior in performance or reliability. Too, hybrid circuits can be an excellent choice for low-volume production runs when custom monolithic IC development would plainly suffer too high an initial design and production cost.

When packaged, an RF hybrid circuit may appear to a casual observer as a large, but ordinary, integrated circuit chip; or as a small, populated printed circuit board if the hybrid is left unpackaged. In most consumer applications the completed hybrid circuit will be placed in a special plastic package, with silicone or resin poured over it; or the circuit itself can be dipped within a liquid plastic and simply laid out to dry. Both of these techniques are used to protect the circuit, and to assist in displacing excess heat from the hybrid assembly.

A typical hybrid will take advantage of various circuit construction techniques, such as substrate-printed film resistors, surface mount and flip-chip ICs, SMD capacitors and inductors, distributed capacitors and inductors, microstrip transmission lines, and so on. In fact, hybrid manufacture is fundamentally a method of constructing a circuit on a specialized PCB by using any desired electronics technology that will decrease size and cost, while increasing reliability and performance.

14.3.2 Hybrid Circuit Assembly

Most of the integrated circuits that are placed on a hybrid's PCB (Fig. 14.4) are in their unpackaged form, even though many hybrids can and do use packaged ICs. However, the ability to exploit ICs in their bare die form means that the size and cost of the entire hybrid circuit can be significantly improved.

The bare IC die is attached to the hybrid's PCB with an epoxy that is optimized for thermal conductivity and/or electrical conductivity (or an epoxy may be selected that thermally and electrically insulates the die from the substrate, depending on the application). Then bare chip's bond wires are attached to the PCB's metallization layer by a bonding process, with gold wires requiring thermosonic bonding, and aluminum wires needing an ultrasonic procedure. Both methods demand a special machine that utilizes a chiseled point that contacts the wire, pressing it against the board's surface metallization while vibrating at ultrasonic frequencies. Thermosonic bonding also adds heat to this process.

For microwave frequencies, a far better bare die attachment technique is used that does not exploit bond wires at all, therefore providing a much decreased inductive effect. Instead, small solder balls are attached to the IC's I/O pad connections. The chip,

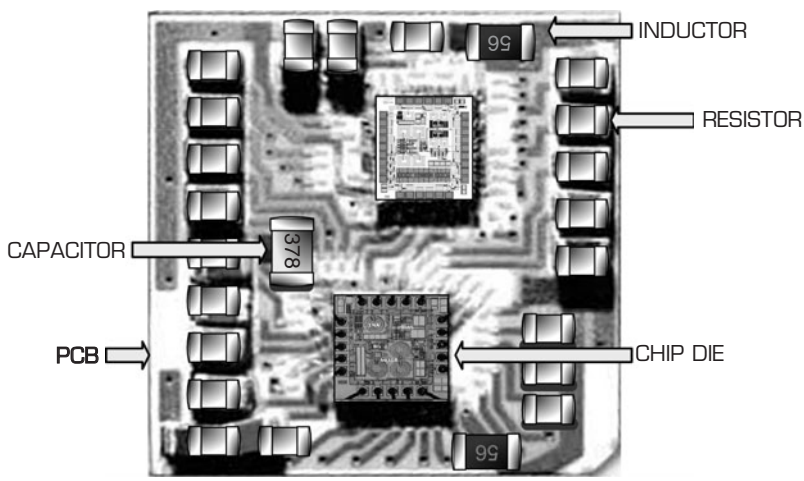


Figure 14.4 A hybrid RF circuit, showing SMD components.

now called a *flip-chip*, is affixed to the board by turning the bare die over so that these solder balls are in contact with the pads of the board's own metallization layer. The entire assembly is then heated, causing the flip-chip to become permanently attached, both mechanically and electrically, to the PCB.

There are other methods of attaching a bare die to a PCB, such as automatic tape bonding, adhesive *microbump* bonding, and so on. For more detailed information on these particular processes consult *Hybrid Microelectronics Handbook*, by Sergent and Harper.

All other surface-mount technology (SMT) components, such as discrete transistors, packaged RFICs, capacitors, resistors, and inductors, can be attached to the hybrid's substrate by the *reflow soldering process*. This common soldering method uses a solder paste that is printed onto the printed circuit board's (PCB) substrate and heated, causing the solder to reflow, and electrically and mechanically connecting the surface mount components to the board.

14.3.3 Hybrid Circuit PCB Materials

The composition of the PCB's substrate and conductor materials are an important consideration in any high-frequency design, and hybrids can use some highly specialized materials. *Metallization* is the conductive layer of the interconnecting traces and pads placed onto the substrate, and can be made of deposited copper, gold, or silver. (The entire PCB itself may be called a *metallized substrate* in these hybrid applications.) The board material is usually some type of ceramic, such as *alumina*, *aluminum nitride*, or *beryllia*. These ceramics are extremely rigid, very temperature stable, and possess very high material strength, which are all attributes that make ceramics perfect as a hybrid substrate material.

Alumina (aluminum oxide) is by far the lowest in cost and the most common of all the hybrid ceramic materials. It operates effectively up to 25 GHz, is mechanically unyielding—and so does not require a plate carrier—and has a very high dielectric constant of around 9.8 (which permits small circuit layouts). Alumina is used in applications that require stiffness, strength, temperature stability, and thermal conductivity.

The next substrate, *aluminum nitride*, is found only in specialized hybrid applications that require better thermal conductivity, but at a substantially increased cost, over alumina.

Beryllia, which is higher in cost than even aluminum nitride, is seen in applications where a somewhat lower dielectric constant is required, as well as improved thermal conductivity. However, Beryllia dust particles are highly toxic, and must not be inhaled when this hazardous substrate material is being machined.

14.4 RF Connectors

14.4.1 Introduction

An RF connector (Fig. 14.5) is a mechanical part that is used to supply a permanent or temporary connection for the low loss transfer of RF energy from one circuit, antenna, or system to another, and preferably at a constant impedance.

All connectors will have a finite lifetime. Indeed, some high-frequency, high-precision units should be replaced after as little as 100 connects/disconnects. This limitation is due to frictional wear between the two mating surfaces, causing a change in the connector's geometry, and thus increasing the insertion loss while decreasing the return loss.

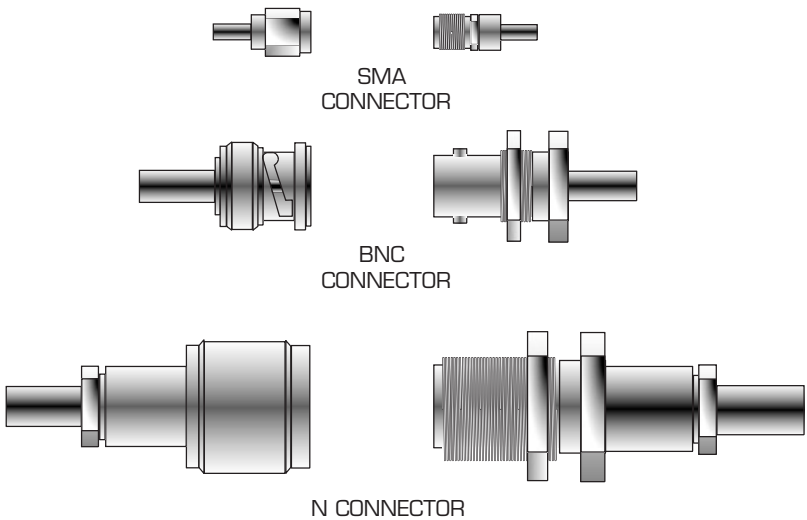


Figure 14.5 Common and popular RF connectors types, both for coaxial cables and for PCB and panel mounting.

Any RF connector that must function reliably out of doors in the wind, rain, ice, and snow should be specifically made for this type of abuse, or must be properly shielded. If not, corrosion will quickly cause damage, sometimes rapidly, decreasing the connector's rated specifications.

14.4.2 RF Connector Types

Over the years there have been many RF connectors invented and manufactured for different frequencies and applications, as well as to improve user-friendliness, performance, or price. The following RF connectors are the most common found today for coaxial connections from HF to SHF:

UHF: Developed by Amphenol in 1930s, UHF connectors are utilized in RF applications up to 300 MHz. This frequency limitation is due to their undesirable nonconstant impedance characteristics. Also called a *PL-259 connector*, they are found in low-cost, undemanding RF service.

N: Developed at Bell Labs, the N connector is a 50-Ω threaded unit, and is the foremost connector for test equipment and antennas up to 6 GHz. However, they are capable of operating up to 11 GHz.

BNC: A BNC can function up to 3 GHz, but are normally found only at lower frequencies and in lower cost test equipment, as well as on external whip antennas. Both 50- and 75-Ω versions are available.

TNC: A threaded version of the BNC connector, TNCs are designed for high vibration environments at up to 10 GHz.

SMA: The dominant 50-Ω microwave connector in use today, and quite capable of operation of up to 20 GHz. SMAs are small, low cost, and have threaded connections.

SMB: Non-threaded, push-on connectors operational up to 4 GHz. Available in both 50- and 75-Ω versions.

APC-7: An expensive, high-performance sexless threaded microwave connector that is qualified up to 18 GHz. Due to their cost they are almost exclusively found only on high-end (Agilent) test equipment.

3.5MM: A precision version of the SMA connector, and can function up to 34 GHz. They have an air dielectric for increased performance over that of the SMA type. 3.5MMs will maintain their operational characteristics for thousands of connects/disconnects.

Wiltron-K: A sturdy, high-frequency (40 GHz) connector that will mate with the SMA type, if required.

14.5 Federal Communications Commission Equipment Authorizations

14.5.1 Introduction

The Federal Communications Commission (FCC) is the governing body in the United States that controls all wireless communications. The FCC does this by allocating spectrum for the various wireless services, by licensing certain users of this spectrum, and by setting the power, stability, spurii, and bandwidth limits of all equipment capable of RF transmissions. Even receiving equipment is controlled by law, as it must not radiate excessive electromagnetic interference (EMI). Keeping this electromagnetic interference to a minimum is vital so that all operators of wireless equipment, as well as users of other electronic devices, will be able to enjoy relatively interference-free performance. The FCC accomplishes this task by mandating that any device, whether it be a transmitter or a receiver, must have some type of FCC *equipment authorization* if it is to be sold, leased, or advertised to any individual or company (unless such sale is to the federal government or to a foreign country).

14.5.2 FCC Equipment Law

As stated above, all electronic equipment that can emit electromagnetic radiation into the surrounding space is controlled under the FCC umbrella. The FCC breaks all such radiators down into three logical segments: *intentional radiators*, *unintentional radiators*, and *incidental radiators*.

Intentional radiators comprise all RF wireless transmitters within the field of communications. Unintentional radiators in RF communications would comprise receivers, which create high levels of radiation internally for frequency conversions, but do not purposely propagate this energy beyond their own cabinet. Both intentional and unintentional radiators must typically obtain FCC authorization to be marketed within the United States. The third category, incidental radiators, is not of interest to the average wireless designer, and comprise electronic devices that do not need nor desire to generate RF either internally or externally, but may do so anyway (such as light dimmers, neon signs, electric motors, and so on). While no FCC authorization is needed for this category, any such device must not produce electromagnetic radiation above similar devices on the market.

Both the intentional and unintentional radiators must have one of three FCC authorizations, depending on the product and its specific use: *Verification*, *Declaration of Conformity* (DoC), and *Certification*.

NOTE: *Due to the legal complexity, you may want to consult a law firm that specializes in FCC technical issues, such as* Fletcher, Heald & Hildreth *of Arlington, Virginia, before marketing a wireless device in volume, or if a device employs a novel modulation or technology. However, the following is a basic, general guide for the class of equipment and the type of authorization that normally must be obtained for most wireless communication's devices.*

Fixed microwave point-to-point and licensed broadcast transmitters (intentional radiators), as well as television and FM stereo and mono receivers (unintentional radiators), will usually obtain a *Verification Authorization*. This type of authorization simply requires the manufacturer of the device to test for FCC technical compliance. A Verification Authorization is by far the least complicated to acquire, as your company need not even file FCC compliance documents and, as soon as the device successfully passes, sales can actively begin. All test and design paperwork should be preserved over the manufacturing life of the product, plus 2 years. All that is demanded is that certain devices must be labeled as directed by the FCC rules.

Virtually all other consumer unintentional radiators are commonly authorized under the *Declaration of Conformity* (DoC). The DoC also allows your company to obtain its own equipment authorization and, as with the Verification Authorizations, the FCC is not even notified of the product's existence. However, the DoC is far more expensive and complex to obtain than the simple Verification Authorization; since all RF equipment tests must be performed by an accredited test facility. If passed, the equipment must be sold with the FCC logo attached—as well as with copies of paperwork called the *Compliance Information Statement* (which contains information on the product and its manufacturer). If the manufacturer so desires, he may also opt to obtain a *Certification*.

Mobile radios, licensed or unlicensed, as well as virtually all high-volume unlicensed transmitters of any kind, must obtain an FCC *Certification*. This covers the lion's share of wireless devices, such as cordless and cellular phones, WLANs, 802.11 devices, handi-talkies, citizen band radios, and so on. Certification is quite demanding and expensive, and involves complex conducted and radiated RF testing, substantial paperwork, and a large filing fee. Even then, the device cannot be sold until the FCC or a *Telecommunications Certification Body* (TCB) approves the application. The TCBs are third-party private certification (but not test) facilities that have been permitted by the FCC to quickly grant Certifications and to issue FCC ID codes—and within a much shorter time period than the FCC.

Whether a product has been authorized by Verification, DoC, or Certification, any future engineering changes to its circuits that may alter its RF compliance, such as a change in power, frequency, shielding, isolation, and so on, will demand a retest be conducted to confirm that these modifications have not dropped the device from the FCC compliance standards. If the compliance test on this modified device now fails the unit, then all sales must be halted and any problems fixed. However, if the engineering changes have either not affected the device, or have actually improved its compliance, then nothing need be done in the Verification or DoC cases (except continuing to maintain all new test and design paperwork). In the case of Certification it becomes a

bit more complicated, as almost any change to the radio's frequency-generating stages, or any increase in output power, demands a completely new FCC Certification. Still, if the modifications were to other circuits that are not involved in the normal compliance tests, and no degradation in test results occurred, then the device may be allowed to continue to operate as before, and with no new FCC filing necessary. But if the device has new test results that are slightly inferior to the original product, then the manufacturer may file a "Class 2" application, along with the current test data, and hope for FCC approval.

APPENDIX

Order of Operations

All mathematic operations must be performed in a specific sequence, or you can obtain the incorrect answer:

1. Complete any and all operations that are placed inside parentheses first, operating on any embedded parentheses and then working outward.
2. Calculate any numbers with exponents.
3. Now work on any operations needing multiplication and division, going from left to right.
4. Complete any operations needing addition and subtraction, going from left to right.

An easy way to remember all this is through PEMDAS (parenthesis, exponents, multiplication, division, addition, and subtraction.)

NOTE: *When a log is involved, we must solve for it first before performing any other operations. For instance, for the formula* gain = *10* log (P_{OUT}/P_{IN})*, we would immediately plug-in all the numbers for* P_{OUT} *and* P_{IN}*, then divide* P_{OUT} *by* P_{IN}*, so that we could then find their log, and then multiply this result by 10.*

During RF analysis and design, we may sometimes need an easy and quick way to convert from one unit to another without resorting to a calculator or a computer. The two tables below supply this conversion. Table A.1 converts from power (dBm and mW) to voltage (dBmV, mV_{RMS}) in a 50-ohm environment, while Table A.2 converts from return loss and VSWR into the resulting insertion loss, power reflected, and power transmitted.

dBm	mW	dBmV	mV_{RMS}
–50	0.000	–3.0	0.7
–45	0.000	2.0	1.3
–40	0.000	7.0	2.2
–35	0.000	12.0	4.0
–30	0.001	17.0	7.1
–25	0.003	22.0	12.6
–20	0.010	27.0	22.4
–15	0.032	32.0	39 .8
–10	0.100	37.0	70.7
–5	0.316	42.0	125.7
0	1.000	47.0	223.6
1	1.259	48.0	250.9
2	1.585	49.0	281.5
3	1.995	50.0	315.9
4	2.512	51.0	354.4
5	3.162	52.0	397.6
6	3.981	53.0	446.2
7	5.012	54.0	500.6
8	6.310	55.0	561.7
9	7.943	56.0	630.2
10	10.000	57.0	707.1
11	12.589	58.0	793.4
12	15.849	59.0	890.2
13	19.953	60.0	998.8
14	25.119	61.0	1120.7
15	31.623	62.0	1257.4
16	39.811	63.0	1410.9
17	50.119	64.0	1583.0
18	63.096	65.0	1776.2
19	79.433	66.0	1992.9
20	100.000	67.0	2236.1
21	125.893	68.0	2508.9
22	158.489	69.0	2815.0

Table A.1

dBm	mW	dBmV	mV_{RMS}
23	199.526	70.0	3158.5
24	251.189	71.0	3543.9
25	316.228	72.0	3976.4
26	398.107	73.0	4461.5
27	501.187	74.0	5005.9
28	630.957	75.0	5616.7
29	794.328	76.0	6302.1
30	1000.000	77.0	7071.1

TABLE A.1 (*Continued*)

Return Loss (dB)	$\|\Gamma O\|$	VSWR	Insertion Loss (dB)	Power Transmitted (%)	Power Reflected (%)
–3.0	0.708	5.848	–3.02	49.88	50.12
–3.5	0.668	5.030	–2.57	55.33	44.67
–4.0	0.631	4.419	–2.20	60.19	39.81
–4.5	0.596	3.946	–1.90	64.52	35.48
–5.0	0.562	3.570	–1.65	68.38	31.62
–6.0	0.501	3.010	–1.26	74.88	25.12
–7.0	0.447	2.615	–0.97	80.05	19.95
–8.0	0.398	2.323	–0.75	84.15	15.85
–9.0	0.355	2.100	–0.58	87.41	12.59
–10.0	0.136	1.925	–0.46	90.00	10.00
–15.0	0.178	1.433	–0.14	96.84	3.16
–20.0	0.100	1.222	–0.04	99.00	1.00
–25.0	0.056	1.119	–0.01	99.68	0.32
–30.0	0.032	1.065	0.00	99.90	0.10

TABLE A.2

This page intentionally left blank

Glossary

Active device Any component that can amplify a DC and/or AC signal, such as JFETs, BJTs, MOSFETs, and so on.

Active region The area of a bipolar junction transistor in which it is capable of amplifying an incoming signal. This area consists of a voltage swing of only about 0.2 V in amplitude, and is the region between *saturation* (0.8 V_{BE}) and *cutoff* (0.6 V_{BE}). Within these two V_{BE} values the I_B, and thus the I_C, is controlled.

AGC saturation point An area in which any further increase in a receiver's AGC voltage becomes nonlinear, or saturates, at the AGC knee.

Amplifier efficiency Important in power amplifier design, where it is referred to as *power added efficiency* (PAE), it quantifies the actual RF power output of the stage vs. the RF/DC input power. PAE can easily be calculated by:

$$\text{PAE} = (P_{\text{OUT}} - P_{\text{IN}}/\text{DC supply power}) \times 100$$

or

$$\text{Eff} = \frac{P_{\text{OUT}}}{P_{\text{IN}} + P_{\text{DC}}} \cdot 100$$

where PAE = power added efficiency of an amplifier stage, percent
Eff = efficiency of an amplifier, percent
P_{OUT} = RF output power, W
P_{IN} = RF input power, W
P_{DC} = power supplied to the amplifier by the DC bias, W.

AM to PM conversion The total of unwanted phase deviation (PM) that is created by any amplitude variations (AM) in a circuit or system. This can be caused by power supply ripple in an amplifier, by multipath fading in a receiver, or by anything that modifies the bias of an active device (which can then alter the device's output capacitance, causing PM). AM to PM conversion will cause a degradation of bit error rate in any phase modulated radio.

Application-specific integrated circuit (ASIC) A custom-designed and -built integrated circuit.

Automatic noise limiter (ANL) A diode circuit that is located after the baseband detector to cut off any noise spikes that would reach the audio section of an AM or SSB receiver.

Average power The energy of a signal averaged over a specific time period. Since *peak power* is the power contained only at a signal's highest peak, and most measurement devices would not only have difficulty evaluating such rapidly changing peak amplitudes in a complex signal, but these high amplitude peaks may actually contribute very little to the signal's total power. Indeed, some modulated signals may have unpredictable, momentary peaks that are ten times higher in amplitude than the average power over time. Therefore, the measurement of only peak amplitudes would not be very useful, so we will usually measure the average power of a signal over a short period of time.

Backplane A common connection point for multiple components, circuits, or systems.

Balanced amplifier A push-pull power amplifier.

Ballast Resistor Certain bipolar junction power transistors incorporate a very low value of emitter resistor within their internal structure to make them less prone to thermal runaway.

Balun A wideband transformer that is proficient at matching balanced lines or balanced stages to unbalanced line or stages, as well as modifying impedances (if desired).

Bandgap reference A temperature-stable voltage reference that is similar to a low voltage Zener.

Baseband signal The low-frequency modulation signal of a receiver or transmitter that has not traveled through any frequency conversion stages, and thus has not been transformed onto a higher frequency, such as music, voice, or digital data.

Baud rate Also referred to as the *symbol rate*, and defines the number of symbols generated by a communications system, per second. Each of these symbols comprises one or more binary bits of data.

Bilateral amplifier A single amplifier that amplifies in both directions, and is utilized in circuit-sharing transceivers to function in transmit and receive systems.

Bit error rate (BER) In a digital data communications system, the number of incorrectly received bits vs. the transmitted number of correct bits. For example, if a communications system has a BER of 10^{-6}, then it is stating that there is 1 bit error for every 1 million bits sent.

Bit rate The number of raw data bits transmitted over a communication's link, per second. Also referred to as the *data rate*.

Blanketing When a powerful conducted or radiated RF or EMI signal damages or destroys receiver reception over a wide band.

Blocking (Desensing) A strong undesired signal causes a receiver's front-end amplifier to saturate, lowering gain and, in some cases, completely blocking out the RF signal of interest.

Blocking dynamic range (BDR) The RF signal input range of a communication's receiver that causes a strong interferer to decrease the amplitude of a low-amplitude desired signal by 1 dB versus the receiver's no-signal noise floor. BDR is specified in dB at a particular frequency offset between the desired and the interfering signals. An excessively high-amplitude interferer above this BDR level can block, or desense, the desired signal completely. True measurements of a receiver's BDR may be hindered by the receiver's increase in its noise floor caused by the interferer's signal mixing with the receiver's local oscillators and other real-life noise impairments. This will create a noise-limited dynamic range measurement. Thus, in these cases, we must classify the receiver's BDR as the point that the receiver's noise floor rises by 1 dB, which is equivalent to the desired signal amplitude decreasing by 1 dB. In this way we will obtain the same SNR value as with a non-noise limited BDR test. (This test must be performed with AGC off).

Blocking interference A powerful undesired signal can overload a receiver, producing a reduction of the desired signal's amplitude that is either partial or complete.

Bluetooth (BT) A wireless standard that allows multiple radios to communicate on their own wireless LAN. BT functions within the 2.4-GHz ISM band, employs FHSS with a nominal hopping rate of 1600 hops/s, and has a total bandwidth of 83.5 MHz. Within this bandwidth BT can perform 79 hops, each with 1-MHz bandwidths, with a transmitted EIRP of 0 dBm (a high-power version has 20-dBm EIRP). The U.S. passband is 2.402 to 2.480 GHz or, including the guard-bands, 2.400 to 2.4835 GHz. Maximum transmit distance in open space for 0 dBm is about 10 m, while 20 dBm is 100 m. The 20 dBm BT device must have a power control built in, with a monotonic power control step size of not more than 8 dB and not less than 2 dB, all the way down to an EIRP of 4 dBm or less. Modulation is GFSK, with a binary 1 equaling a positive frequency deviation, and a binary 0 equaling a negative frequency deviation, with the minimum deviation always being greater than 115 kHz, but less than 175 kHz. Bluetooth can work point-to-point or point-to-multipoint. Point-to-multipoint uses a *piconet*, with one BT unit acting as a master unit, and the other BT units act as slaves. Multiple piconets can form a larger *scatternet* if any of their coverages overlap. Both master and slave units of one piconet can even operate in another piconet with a master in one piconet functioning as a slave in another piconet. BT units employ TDM so that they can function in more than one piconet at a time, with each piconet possessing its own hopping channel. This channel is one pseudorandom hopping sequence that hops through 79 RF channels, with the PN code unique to a particular piconet (the PN code is governed by the master unit). With the nominal rate of 1600 hops/s, the channel is divided into time slots of 625 μs, with each time slot numbered by the master. In this way, the master begins a transmission on even numbered slots, while the slave is allowed to send only during odd-numbered time slots. This is a duplex TDD method. However, the transmitted packets may last to five time slots, with the RF hop frequency frozen until completion of the packet.

Buffer amplifier A buffer amplifier is designed to isolate the load from the source, which makes a high S_{12} (isolation) and a high S_{11} and S_{22} (return loss) important for a good, nonreflective match. Typically a buffer will be placed between an LO and its mixer, preventing the LO frequency from being affected by a poor match at the mixer's port, as well as supplying some additional gain (many buffers, however, may have little or no gain). Some buffer amplifiers will have a high Z_{IN}, and are adopted mainly to block the loading of the output of an oscillator. An ordinary buffer may have an S_{12} of between 20 and 50 dB, and an S_{11} of 10 to 20 dB. High-isolation MMICs, instead of discretes, are sometimes appropriate in this isolation buffer role.

Butterworth characteristic A classification of a Butterworth filter's response, which has a very flat bandpass with no peak or ripple.

Bypass capacitor A capacitor used to shunt AC, preventing it from entering or exiting a circuit.

C/N ratio The ratio between a signal's carrier and the noise amplitude, in dB.

Carrier The frequency used to convey data or voice modulation from a wireless transmitter to a wireless receiver.

Cascaded amplifier One or more amplifiers connected in series to increase the gain over a single amplifier. This multiplies the voltage gain of each amplifier by the next amplifier, or adds the gain if using decibels.

Channel A specific section of the frequency spectrum that contains a carrier frequency, with sidebands; or a recognized passband frequency, standardized by common agreement.

Chebyshev characteristic A classification of a filter's response, which has a certain amount of passband ripple, but a sharp frequency cutoff.

Circuit *Q* The loaded *Q* of a circuit, with the narrower the circuit's bandwidth the higher the Q, or $f_C/(f_2 - f_1)$.

Circuit sharing A circuit design tactic that permits the joint use of certain stages for both the transmit and receive sections of a transceiver, such as the filters, oscillators, power supplies, frequency synthesizers, and/or amplifiers.

Circulators A device that can be employed as a duplexer or as an isolator. Circulators are important due to their unilateral nature, since they will allow the RF's signal power to flow in one direction only. In the duplexer role, one port is attached to an antenna, while the other two ports are attached to the transmitter and receiver. This will permit almost all the power to flow from the transmitter to the antenna port, while any signal at the antenna will only flow from the antenna to the receiver. Circulators can be mounted on waveguides or on microstrip.

Code division multiple access (CDMA) A multipoint technology that allows all client transceivers to occupy the same physical communications channel during the same time period. The modulated signal of each user is spread by an uncorrelated code sequence, which expands the data or voice signal to a wider bandwidth than that required by the information actually being carried. The code itself is unique to each client, which permits any receiver with the appropriate code to accept the transmitted signal, while rejecting all others. The receiver, which possesses the same code as the transmitter, amplifies the received signal, then multiplies it by a duplicate code, canceling out the original code sequence as sent by the transmitter. The device that removes the code is referred to as a *correlator*. After correlation, the now narrowband signal, which is a duplicate of the originally transmitted information before the added code sequence was inserted at the transmitter, is sent through a bandpass filter and demodulated in the normal manner.

Codec (coder/decoder) A digital piece of hardware that takes an analog voice and digitize it into digital 1s and 0s, which are then modulated onto the wireless system's carrier during transmit. After reception and demodulation by the receiver, the codec accepts the incoming serial digital bit stream and converts it back into an analog voice waveform that is then heard by the operator of the device. Many codecs today are actually *vocoders* (or *voice coder*), which also compresses the analog voice signal so that less than half the data bits need be sent to transmit the exact same signal intelligence as a typical codec.

Coherent reception The ability of a receiver to demodulate an incoming signal based on its phase, rather than its frequency. Since a coherent communications system's phase control must be very accurate, most receiver demodulators take advantage of recovering the transmitted carrier from the incoming input signal through a *Costas loop*, or by using the transmitter's pilot carrier. Coherent is 3 dB more sensitive than noncoherent demodulation since it reduces the received noise at the detector by half, thus supplying an improved BER. This 3-dB degradation when using noncoherent detection is due to its lack of any accurate phase information, so it must compare the previously received symbol to the currently received symbol, and then check as to whether there was a change in the current symbol's phase or not. However, this type of demodulation gives us two potential sources of error: a possibly corrupted new symbol, and a possibly corrupted prior reference symbol. Thus, we can see the huge improvement in sensitivity with the coherent receiver's local oscillator being locked in phase and frequency with that of the transmitted carrier, with the added phase information as supplied by the entire coherent system. However, coherent radios have an inherently higher cost and complexity over noncoherent systems.

Coherent transmission The capability of a modulator to manipulate the phase, as well as the frequency, of a signal. Typically generated with an I/Q modulator. (Directly modulating of a VCO will not permit the phase to be controlled, only the frequency, and would be referred to as *noncoherent transmission*.)

Common-mode conducted signals Wires will carry common-mode conducted signals if they function as a single conductor with all their signals in phase, and ordinarily having a return through ground, performing as a receiver or radiator for undesired signals. These conducted signals can be cured by using common-mode chokes and twin ferrite beads and twin capacitors to ground (see *differential-mode conducted signals*).

Conductive pad Transistors may employ an electrically and thermally conductive pad placed between their package and their heatsink to improve heat removal. These do not require thermal grease.

Conversion loss The loss in power between the input of a device at a specific input frequency, and the output of that same device at a specific frequency, in dB. Normally applies only to frequency mixers or multipliers, in which the input frequency is different from the device's output frequency.

Converter (DC to DC converter) A direct current power supply that changes one DC voltage into another DC voltage that can be higher or lower than the original.

Copper losses The losses in conductors produced by I^2R losses, skin effects, and crystallization.

Coupling capacitor A capacitor that freely admits AC on to other stages with low impedance, while blocking DC.

Cross modulation The undesired shift of baseband intelligence from an adjacent, strong channel to a desired, but weak, channel. The undesired signal modulates the desired signal in the receiver's front-end nonlinearities, creating internal amplitude modulated interference.

Crosstalk Electromagnetic or electrostatic coupling of signals between PCB traces or components.

CSMA/CA A multipoint protocol that controls when a wireless station may transmit, and when it may not. CSMA/CA stands for *collision sense multiple access with collision avoidance*, and is operated in WLANs and other multiuser environments. It is quite effective in low-data rate, bursty, wireless applications. CSMA/CA attempts to avoid data collisions by checking the RSSI (signal strength) level of its receiver section. If it senses that no other clients are transmitting within the LAN, it will then transmit its data packets. However, it will not attempt to transmit if a sufficiently strong signal is sensed, and will only try another transmission after a random time period has elapsed. Optional additions to the CSMA/CA protocol permits the radio to not only sense spectral energy from the RSSI, but also as to whether the signal is actually of the same type.

Cutoff A transistor in its off condition, conducting zero current, with its collector-emitter voltage now equaling its DC supply bias voltage.

Desensitization (Desensing) A powerful undesired signal that is within the bandpass of a receiver can create an overloading in the receiver's first RF amplifier or mixer. This produces decreased sensitivity to the desired signal, as well as distortion due to overdrive caused by the receiver's LNA or mixer being placed into a nonlinear area of its operation.

Die form An integrated circuit used without a package to conserve space.

Dielectric resonant oscillator (DRO) A microwave oscillator for very high-frequency operation. The heart of the DRO is a dielectric "puck," which functions as a resonant cavity.

Differential-mode conducted signals Signals that are out of phase with each other by 180° on two-conductor wire. The bulk of desired signals are in differential mode, and can be filtered by common LC filter circuits (see *Common-mode conducted signals*).

Digital signal processing (DSP) A method of altering an incoming signal digitally by converting it to digital pulses, manipulating it, then usually converting the signal back into an analog waveform. DSPs can dramatically improve the signal-to-noise ratio, as well as the fidelity, of a signal.

Dispersion An effect, especially important at high microwave frequencies, that is present in all transmission lines and PCBs in which a wideband (multifrequency) signal becomes distorted due to the different phase velocities of each of the frequencies that comprise the signal. This disrupts the phase relationships of the signal before it can reach the other end of the line. In a digital wideband signal this can cause the eye diagram to begin to close, which increases the BER.

DMOS A type of high-power, high-gain, and high-frequency RF power MOSFET.

Driver stage An amplifier that supplies enough power to drive a power amplifier. A *predriver* drives the driver.

Dummy load A 50-Ω resistive load for transmitters, used for testing without radiating.

Dynamic range The measurement between the lowest and the highest signals reproduced without significant noise or distortion in a system. High dynamic range in communication's receiver will allow the detecting of wanted signals with a difference in amplitude of 90 dB or more, with a usable BER, and within the hostile in-band and out-of-band jammer-infested real world. Designing these high dynamic range receivers not only involves increased system's cost and complexity caused by the need to employ double-balanced diode ring (or FET quad) mixers, along with high *P*1dB amplifiers and larger power supplies/batteries, but also much higher LO output power levels to drive the higher level mixers, also causing difficult-to-suppress LO radiation and LO harmonics. And since the LNA will need a higher current bias, which will not be optimal for low noise operation, increasing of the receiver's noise figure will also occur. Talk time with battery-operated radios will decrease due to all of these increased current needs.

E_b/N_0 Ratio of the energy per bit (E_b) over the spectral noise power density (N_0). In other words, it is the SNR for digital wireless systems but, more accurately, it is more or less the SNR divided by the amount of functional bits per symbol. Thus, it is the proportion of the average signal power to the average noise spectral density, and is equivalent to standard SNR normalized over every transmission frequency. E_b/N_0 can be presented as a table or graph versus bit-error-rate for different types of modulation. For instance, PSK may require an E_b/N_0 of approximately 10.5 dB for a bit-error-rate of 10^{-6}, while OOK (On/Off Keying) may require 14 dB for the very same BER.

EIRP (effective or equivalent isotropic radiated power) The power that an isotropic radiator would have to emit in order to create the output peak power density as seen in a real antenna's most favorable (i.e., maximum gain) radiation angle. EIRP includes not only the gain of the antenna itself added to the output power of the PA, but also losses within the feedline, matching circuits, and connectors.

ERP (effective radiated power) The true power delivered to the antenna multiplied by the power gain of the major lobe of the antenna. *EIRP*, or the *effective isotropic radiated power* (see above), is the product of the transmitted power multiplied by the gain over an *isotropic* antenna, while ERP employs a dipole instead of an isotropic source as its reference. The ERP of a transmitter system can be calculated by adding the output power, in dB, with the transmitter's antenna gain, in dB, and subtracting the feedline losses, or:

$$P\ (\text{dBm}) = P_{TX} + G_{TX} - L_{TX}$$

where P (dBm) = total amount of power transmitted from the transmitter's antenna, called ERP, dBm.
P_{TX} = output power of the transmitter before the antenna, dBm
G_{TX} = transmitter's antenna gain, dB (dBi for EIRP)
L_{TX} = feedline losses between the transmitter and the antenna, dB.

Electromagnetic compatibility (EMC) Tests of electronic devices are conducted to guarantee that they do not radiate excessive interfering electromagnetic signals. These devices should also not be unduly susceptible to another device's signals, or to small incoming levels of EMI. The EMC tests commonly span the spectrum from 60 Hz all the way up to 12 GHz, and above.

Electromagnetic interference (EMI) The interference caused to any electronic device, such as radios, televisions, telephones, and the like, due to an emitter of electromagnetic energy, especially that from electric motors, transmitters, relays, and computers.

Engineering change order (ECO) An explicit set of instructions to modify specific components, circuits, or specifications in an electronic device.

Error-vector-magnitude analyzer (EVM) Error-vector-magnitude analyzers, part of a test and measurement device called a *vector signal analyzer* (VSA), are a relatively new piece of RF equipment, and are capable of measuring any phase and amplitude impairments present in digital radio systems. Since advanced digital modulation schemes use both amplitude and phase to send information, and these phase and amplitude states can be crowded very close together (such as in QAM-16 and above), then noise and other impairments can severely degrade their BER. The VSA is perfectly suited to quantify these types of errors, and is important for checking all digital signal impairments, especially those caused by overdriven amplifiers, excessive group delay variations, amplitude ripple in the passband, LO feedthrough, phase noise, symbol timing, ISI, white noise, spurs, frequency offsets, and the like. Agilent is dominant in the field of EVM measurement analyzers.

Far field An antenna's field pattern at significant distances, comprising the electromagnetic field that is used for distance communications. Also referred to as the *radiation field*. The approximate distance from the antenna to the far field is approximate, and depends on the specific formula used:

$$(5\lambda) \times (2\pi) \qquad \text{or} \qquad 2D^2/\lambda \qquad \text{or} \qquad \lambda/(2\pi) \qquad \text{or} \qquad 3\lambda$$

where λ = signal's wavelength, m, D = antenna's longest dimension, m.

Faraday shield An electrostatic metallic shield located between the primary and secondary of an RF air-core transformer. This shield significantly decreases the capacitance that occurs between any two conductors that are separated by a dielectric, such as air. Therefore, harmonics,

noise, and spurious signal frequencies cannot easily pass through the now increased capacitive reactance of the two coils of the transformer. But the shield still permits the resonant frequency itself to be coupled to the secondary coil through normal transformer action.

Fast Fourier Transform (FFT) A technique employing algorithms to convert a time-domain signal into the frequency domain. Frequency domain signals are far superior for displaying harmonics than in the time domain.

Feedthrough capacitor An EMI filtering device that permits DC and low-frequency AC to penetrate an enclosure, while grounding noise and higher frequency AC/RF. It functions as a pass-through wire with a shunt capacitor to ground. Commonly used at the input to a metal cabinet or shield that contains a transmitter, receiver, or power supply.

Fencing Copper shielding or multiple vias placed within the substrate of a printed circuit board to block undesired signals from propagating throughout the PCB's dielectric.

Field strength A measurement of the electromagnetic energy radiating from an antenna, or other structure, in microvolt/meter.

Final power amplifier (FPA) The last amplifier that is located before the antenna of a transmitter.

Flat-fading A phenomenon that causes the entire transmitted signal to decrease in amplitude, significantly lowering SNR at the receiver. This is typically caused by atmospheric conditions in long, line-of-sight high-frequency links, as well as in sub-30MHz ionospheric communications. Flat fades of up to 20 dB are relatively common in both types of links.

Flywheel effect A tuned circuit has the capability to produce a sinusoidal wave even if only a simple pulse is received. It is able to perform this function due to the capacitor and inductor of the tuned tank circuit exchanging energy back and forth, creating an oscillation at its own resonant frequency. The sinusoid will not decay in amplitude as long as a pulse is received at the proper time to reenforce and restore the power lost within the tank's parasitic resistances.

Forward error correction (FEC) A technique of error correction to catch, locate, and correct transmission errors by sending redundant data to the receiver, along with the data payload. This prevents retransmission from being required when errors are found in the data, and improves BER. *Reed-Solomon* is one such type of FEC.

Frequency diversity Two separate frequency channels are transmitted in a communications link to provide better multipath performance, with the receiver selecting the frequency channel with the strongest signal.

Frequency tolerance The degree in which a wireless device's carrier frequency is permitted to deviate above and below its center frequency, expressed as a percentage.

Frequency translation Converting an input frequency to either a higher or a lower frequency, while still maintaining the original baseband modulation information.

Full duplex A communications method that allows both ends of a link to communicate with each other during the same time period.

Fundamental overload A type of interference, sometimes severe, to a communications receiver device. It is created by the powerful fundamental signal of an off-frequency interferer reaching the receiver, producing blocking or distortion products when the receiver is unable to reject these signals.

Gain The power or voltage increase or decrease from an electronic circuit's input to its output, in dB. The term *conversion gain* is sometimes used in mixers to specify gain due to the frequency translation that takes place.

Gain block A stable, high-gain, wideband 50-Ω amplifier usable from low frequency to at least 3 GHz and beyond.

Gain compression The point in an amplifier's or receiver's gain curve in which raising the input power of a signal will not linearly increase the power at the device's output dB for dB.

GASFET A field-effect transistor made of gallium arsenide that supplies gain at microwave frequencies. More expensive than silicon devices.

Ground The reference point for a circuit or system. It is not a "current sink," but is merely a conductive area that is presumed to be at 0 V. However, most grounding points on a PCB are only close to zero potential. Not being exactly at 0 V throughout its total surface is undesirable, and is caused by a real-life conductor's nonzero resistance and reactance value. Consequently, as AC current flows through the ground trace or a nonsolid, broken groundplane, a voltage will be formed across it, and will now no longer be a perfect 0-V reference point. Due to a conductor's nonzero resistance, flowing DC current will also produce this same issue. These unavoidable Ohm's law effects of $V = IR$ are, fortunately, minor in any well laid-out board, especially one employing a solid groundplane and multiple vias to the ground area.

Ground bounce The voltage from a transistor's emitter to ground of a high-power, high-frequency active device should ideally be 0 V, but is usually some fraction of a volt due to emitter lead inductance. This is referred to as ground bounce.

Group delay The comparative delay of specific frequencies over that of other frequencies, such as through a filter circuit.

Half duplex A communications link in which each end of the link can communicate with the other, but only at different time periods.

Harmonic distortion Distortion frequencies that are spaced at integer multiples from the fundamental frequency, and measured as dBc. Harmonic distortion is produced within the nonlinearities of any active device, and is very problematic in power amplifiers. The output harmonics created by the amplifier's nonlinearities can best be decreased by lowering the amplifier's input power, using a linearizer circuit, increasing the bias current, selecting a higher powered transistor, employing an active device with better linearity, modifying the transistor's output match, and/or filtering the output.

Harmonic suppression The degree to which harmonics of the fundamental frequency are attenuated before being output from a transmitter.

Heterojunction bipolar transistor (HBT) An active device that can operate up to 30 GHz and above in oscillators and power amplifiers. HBTs have poor noise figures, but good flicker noise characteristics.

Heat spreader A thin copper plate that assists in dissipating a device's heat over a wider area. The spreader is located between the device and the heat sink.

Helical resonators A filter that is constructed of a helically wound resonant transmission line surrounded by a conductive shield. Helical filters are used between 50 and 500 MHz, are highly selective, and have a flat response when placed in a series of four or more.

High-electron mobility transistor (HEMT) An active device that has the capability to operate up to 200 GHz, and possesses a very similar physical structure to that of a GaAs FET. HEMTs are mainly employed as the active element of an LNA for microwave receivers in DBS (direct broadcasting satellite television), radio telescopes, and in terrestrial and space telecommunications applications.

HEXFET A high-gain metal-oxide semiconductor field-effect transistor found in switching power supplies and audio amplifiers.

House wiring The 120-V color-coded mains wiring found in American residences or office buildings. The green wire is at ground (the round plug on the socket or plug); the white wire is the neutral (the wide slot); and the red or black wire is the hot lead (the narrow slot).

Hum A 60-, 120-, or 180-Hz low-frequency interference riding on the RF or baseband signal of interest, or on the DC supply voltage. Hum is normally created by an unshielded power transformer or poor filtering of the power supply.

Hysteresis loss AC transformer power losses caused by friction of the shifting magnetic domains within its core material, dissipated as heat. Air-core coils have virtually no hysteresis loss.

IEEE 802.11 wireless The most important wireless local area network (WLAN) specification today is referred to as the 802.11 standard. This specification was adopted by wireless manufacturers in order to build compliant radios that will interoperate, even when multiple vendor's radios are involved. Noncompliant wireless equipment is still common for specialized applications, such as low-cost WLAN systems, wireless video, or for higher bit-rate applications than 802.11 allows. The 802.11 standard details the *media access control* (MAC) and the *physical layer* (PHY) specifications only, and not how the specifications should be implemented. The specifications concentrate on the 2.4-GHz unlicensed band, with the older data rates of 1 and 2 Mbps using either FHSS or DSSS, the 11 Mbps utilizing DSSS, and the 54 Mbps using OFDM. The permitted total bandwidth comprises 83 MHz between 2.4 and 2.483 GHz. DSSS uses DQPSK (2 Mbps) and DBPSK (1 Mbps) modulations in three 20-MHz bandwidths, while FHSS employs two to four level GFSK in three hopping patterns with 1-MHz bandwidths in 79 frequency slots, with a hop occurring at a minimum of 2.5 hops/s. An antenna gain of no more than 6 dBi is permitted in the United States, with a maximum EIRP of 1 W. In a departure from CDMA, 802.11 DSSS radios all use the same code. The MAC layer of the 802.11 protocol specifies CSMA/CA (see *carrier-sense, multiple access/collision avoidance*) as the protocol to share the airwaves. The 802.11b radios of 11 Mbps have been largely replaced with 802.11g, which allows interoperability with 802.11b and the older 802.11, and also permits speeds of 6, 9, 12, 18, 24, 36, 48, and 54 Mbps using OFDM.

I/Q signals *I* is a signal with 0° phase shift, while *Q* is a signal with 90° phase shift. *Q* stands for *quadrature*, while *I* stands for *in phase*.

IF notch filter A tunable wave-trap utilized within a communication receiver's intermediate frequency stages to remove specific unwanted frequencies.

IMD dynamic range The difference between the noise floor and the power of two equal-in-amplitude input signals that cause third-order products to be 3 dB above the noise floor, measured in *dB*.

Incidental AM (IAM) Undesired amplitude modulation of an FM signal.

Incidental FM (IFM) Undesired frequency modulation of an AM signal.

Injection locking (*also called* injection pulling) A phenomenon in which an LC oscillator is shifted (or *pulled*) to a frequency that is exactly the same as a nearby, high-amplitude interfering frequency from another oscillator.

Insertion loss A loss instead of gain (i.e., negative gain) through a circuit, in *dB*.

Instability An amplifier that has instability may oscillate at some specific frequency, temperature, or input/output impedance. An unconditionally stable amplifier will not oscillate at any frequency or input/output impedance. However, in-phase external feedback can cause even an unconditionally stable amplifier to oscillate.

Isolator A device that permits a signal to pass in one direction, but not in the reverse direction. An isolator is used to prevent reflected waves or externally generated frequencies from entering the finals of some transmitters. It may be mounted in a waveguide or on microstrip, and can be constructed of a three-port circulator in which one of the circulator's ports is terminated.

Jitter The rapid fluctuation of a signal in frequency or amplitude, created by phase changes in a digital signal produced by phase noise in one or more of the system's local oscillators, causing inaccurate or ambiguous demodulation of a received symbol due to timing errors.

Linear mixing A mixing process that will permit one frequency to sum with and ride on another frequency, with no extra frequencies created, unlike nonlinear mixing.

Linearizer An electronic device adopted to reduce spectral regrowth in the power amplifiers of a transmitter. Most linearizers work by predistorting the signal before it actually reaches the PA stage, thus counteracting most of the nonlinear distortions introduced by the PA itself.

Loading coils An inductor will electrically increase the length of an antenna to make it appear longer than it really is, therefore permitting a short antenna to operate at a lower frequency. However, due to a higher *Q*, the antenna will have a narrower bandwidth than if it were at its own natural resonant frequency.

Local area network (LAN) A common collection of computers that are connected over a small area, either by wire or radio, to share files or peripheral equipment.

Loss resistance The undesired resistive part of an antenna's input impedance that contributes only to the heating of the antenna's structure, and does not assist in the actual RF radiation of the signal.

Mains voltage Voltage supplied by interior or exterior building power outlets. In the United States, this voltage is at 120 to 240 V; 117 to 234 V; or 110 to 220 V.

Mean time between failure (MTBF) The projected average time before a device is predicted to fail during normal operation, measured in hours.

Microphonics An undesired effect caused when the vibration of an electronic circuit or component creates noise amplitude modulations of a receiver or a transmitter. It is exacerbated by loose components in oscillators, such as crystals or capacitors.

Microprocessor One small computer on a single integrated chip.

Mil One thousandth of an inch.

Mismatch loss The decrease in power transferred to a load due to the lack of a true complex conjugate match. It is simply the loss, in *dB*, that occurs with any load mismatched to its

source, since only in a truly matched source-to-load condition could the maximum available power be transferred and absorbed by the load.

MMIC A high-frequency integrated circuit for microwave communications. A MMIC can either be an amplifier, a mixer, or a switch, but mainly refers to a Class A RF gain block with matched 50-Ω input and outputs. These MMIC amplifiers are available with low and high gains (8 to 33 dB), low noise (1.7 dB), low and high bandwidths (a few kHz up to 8 GHz), low and high *P*1dBs (1.5 to 22 dBm), and low and high bias currents (5 to 100 mA).

Modulus The number of counting states before a digital counter begins to repeat itself.

Monolithic component An electronic part that is formed from a single substrate material, with metal etching or deposition performed on the dielectric. A typical integrated circuit is a monolithic component.

Monolithic-crystal filter A miniature, low-cost multielectrode crystal filter that can operate at various frequencies between a minimum of 5 MHz and a maximum of 500 MHz.

Monotonic A specific attribute of a circuit's response to a certain stimulus that never changes its direction. For instance, if a filter had a steady, rising attenuation versus frequency characteristic, and that attenuation level never began to fall, then that would be considered monotonic. Another monotonic example would be when a VCO produced a constant decrease in frequency when supplied by a rising tuning voltage, and never displayed any tendency to change this frequency movement upward during the application of the rising input tuning voltage.

Monte Carlo analysis A computer simulation of a circuit run n times, with random component values, over a set range. For instance, if an LC circuit is simulated in a Spice program, and each component has a tolerance of ±10%, then Spice will choose random values within this ±10% from each component's ideal value, and run some operator chosen number of simulations (n) to discover how component tolerances will affect the circuit's operation.

Motorboating Audio amplifier self-oscillations that are caused by a low voltage supply, such as dying batteries.

Multiplexers A device with a single output, but many inputs, and which any desired input can be steered to this single output by control signals.

Near field The field pattern closest to a radiating antenna, and is comprises a reactive zone. Also referred to as the *induction field*, and extends to a very limited distance from the antenna structure itself of not further than $2D^2/\lambda$ (with D being the length of the largest radiating antenna element).

Near-field probes Probes to detect cable and printed circuit board emissions.

Netlist A text list of components that make up a circuit, along with how they are connected within the circuit. It is basically a text description of the circuit that is to be simulated or laid out, with each node given a number or name.

Noise factor (F) A figure of merit that is sometimes specified in mixer or amplifier data sheets. Noise factor compares the noise created by a device to the thermal noise created by a 50-Ω resistor at 290 K, or the decrease in SNR caused by the noise added by the device. Noise factor is also specified as the total equivalent input noise power divided by the input noise power due to the source only. It is a ratio, with no dB notation (see *Noise figure*).

Noise figure (NF) The noise figure is the amount of noise that is added by a receiver or individual stage. In a receiver, this noise is contributed mainly by the receiver's first filter and amplifier stage. Noise figure is the *noise factor* (*F*) specified in *dB*, or:

$$10 \log_{10}(F) \qquad \text{or} \qquad 10 \log_{10}(S/N_{(\text{OUTPUT})}/S/N_{(\text{INPUT})}).$$

Noise floor The amplitude, in dBm, of all the combined thermal noise of a wireless device. Below this level a modulated signal cannot easily be detected, with the noise floor limiting the maximum amplification of any such low-amplitude signal. The noise floor is a direct function of the bandwidth employed to take the noise measurement itself.

Notch filter A bandstop filter with a very sharp bandwidth of high attenuation. Prevents a specific undesired band of frequencies from passing through a circuit.

Nyquist theory States that the sampling rate must be at least double the highest frequency component of the received signal being processed in order to properly retrieve the information embodied within the signal.

Octave An octave is double a certain frequency, or half a certain frequency. For instance, if a 3-MHz signal is used as an example, then one octave above 3 MHz would be 6 MHz (or 2×3 MHz), while two octaves above 3 MHz would be 12 MHz (or $2 \times 2 \times 3$ MHz, or 2×6 MHz), three octaves would be 24 MHz (or $2 \times 2 \times 2 \times 3$ MHz, or 2×12 MHz).

Occupied bandwidth Generally the bandwidth of a digital signal, in Hz, that contains 99% of the signal's power. Other values besides 99% can, and are, utilized.

Original equipment manufacturer (OEM) The actual manufacturer of a specific circuit, system, or device.

Ortho-mode transducer (OMT) A device that functions as a waveguide diplexer at microwave frequencies. OMTs are composed of two output rectangular waveguides, and a single input circular waveguide. When both a horizontally and vertically polarized wave are sent into the circular waveguide input, which supports both polarizations with very little loss, these two polarizations will soon hit a junction. One offshoot of this junction has a horizontal rectangular waveguide, while the other has a vertical rectangular waveguide. Thus dual polarized signals are separated (or can be combined) into or out of a transceiver. Most OMTs have limited bandwidth capability.

***P*1dB compression point** The input or output location of an active device that specifies when the gain becomes 1 dB less than the gain at lower input power levels, measured in dBm. At this *P*1dB point harmonics and intermodulation products start to become a serious issue.

Pad attenuator A passive circuit that decreases the amplitude of a signal over a wide band of frequencies, specified in dB, while also maintaining the proper input and output impedance of the system.

Parametric instability Instability of an amplifier produced by the collector-to-base nonlinear capacitance causing low-frequency modulation of the transistor's output, and the outbreak of oscillations. Due to the small value of this nonlinear capacitance, parametric instability is normally only a problem at higher frequencies. Can be alleviated by high-value bypass capacitors.

Parts per million (PPM) Commonly used as a measurement of the frequency accuracy and stability of an oscillator. As an example, a particular frequency source of 155 MHz may be

specified as being frequency accurate to within ±0.1 PPM. However, since this oscillator is rated as accurate to within ±0.1 PPM, then we know that its frequency may vary by a maximum of ±15.50 Hz from the rated frequency, or $0.1/10^6$ 155 MHz = 15.5 Hz.

Peak envelope power (PEP) The average peak power of a linear signal with 100% modulation applied to the carrier, specified in watts or dBm.

Peak limiter A circuit used to confine a signal's maximum output amplitude.

Phy layer The Phy, or physical, layer defines the modulation and signaling attributes for wireless data transmission.

Poles A mathematical solution that is used in filter design to describe the points of a filter's transfer function that have infinite insertion loss, and that form a rolloff of 6 dB/octave. A series capacitor will thus act as a pole—an open circuit—at DC, while a series inductor acts as a pole at infinite frequency. Therefore, lowpass filters of the all-pole variety, such as Chebychev and Butterworth types, will have their poles all located at infinite frequency.

Port The location within a circuit or device that a signal can be inserted or extracted, either in a physical device or in a software simulation.

Power spectral density (PSD) The measure of the RF power within a given bandwidth, specified in dBm/Hz, V_{RMS}^2/Hz, or W/Hz.

Processing gain In direct-sequence spread spectrum (DSSS) communications, a modulated signal is sent into a correlator that spreads the signal over a much wider bandwidth than is actually required for the information being carried. The signal is then received, and the information is despread. This action creates *processing gain*, specified in *dB*. Processing gain in DSSS systems is obtained by the multiplication of data by the PN code, since it spreads the information out over a low spectral power density due to the much higher rate of the PN code over that of the information. The processing gain can simply be viewed as the difference in amplitude between the unspread and the spread signal. Because the transmitter spreads the intelligence out before it transmits it, and the receiver despreads it (i.e., converts it back to a narrowband signal), any interference at the receiver is converted to wider band, low-amplitude signals with a very low power spectral density. This means that the higher the processing gain, the more an interferer's ability to damage or destroy DSSS communications is reduced.

Quadrature phase Two sine waves that are 90° out of phase.

Radial stub A microstrip open stub on a PCB, flared like a trumpet bell at its open end, that permits a broad range of frequencies to pass as an RF short circuit.

Radials A simulated earth ground. It is a groundplane or counterpoise, normally consisting of four or more quarter-wavelength wires laid on or in the earth, that are used when insufficient conditions exist for grounding a quarter wavelength vertical antenna.

Radiation pattern The typical two-dimensional graphical polar representation of the antenna's far field and relative radiation amplitude versus a spatial direction angle, in degrees, in both the azimuth (0° to 360°) and the elevation (0° to 90°) planes, normally in dB.

Radiation resistance The resistive portion of the antenna's total input impedance that contributes to the RF signal's radiation, as opposed to the *loss resistance*, which is the resistive portion of the antenna's input impedance that only contributes to heating of the antenna's structure. In other words, the radiation resistance is the only part of an antenna's input impedance that actually performs useful work.

Radio frequency interference (RFI) Interference created by a generator of RF, such as a wireless transmitter, computer, or even the local oscillator of a radio receiver, into any other electronic device.

Received signal strength indicator (RSSI) meter A relative or real power indicator of the strength of the RF signal reaching a receiver. This signal strength is obtained from the AGC or RSSI circuits.

Reciprocal mixing A mechanism that degrades the SNR of a communication's receiver through the following action: The receiver's local oscillator, with its phase noise, and the receiver's mixer stage both downconvert a desired RF signal, as well as a close-in signal that is undesired. This mixing action will pass on the LO's phase noise to both the desired and the undesired signals, which effectively increases the desired and the undesired signal's bandwidths due to the new noise sidebands imparted by the noisy LO. This widened undesired signal can now interfere, or even swamp out, the desired signal, especially if the desired signal is of low amplitude and the undesired signal is of a much higher amplitude. Thus, reciprocal mixing in a receiver is created when the local oscillator's phase noise, which can be thought of as being composed of a huge multitude of miniature, low-powered LO outputs, mixes with any incoming RF signals or interference down to the IF frequency, and gets in the way of our desired signal over a broad IF bandwidth.

Reflected impedance A transformer has no true inherent impedance of its own, but merely reflects the impedance of the opposite winding. So, in transformer action, the impedance of the secondary load is reflected back into the primary winding. This impedance can be calculated by the turns ratio of the transformer times the impedance of the secondary load.

Regenerative feedback Any feedback that is at 0°, or in phase, from a device's output back into its input is called regenerative feedback, and can create increased amplification or oscillations. It is also referred to as *positive feedback*.

Repeater A device that is capable of retransmitting an RF signal as it is received from a transmitter. It is employed to extend the range of a low-powered transmitter, or to lengthen the line of sight of a low-altitude antenna.

Residual AM Undesired amplitude modulation created during frequency modulation of an oscillator. This can be mitigated during the design phase by using a limiter at the oscillator's output. The root cause may be the oscillator's active element or a frequency-determining LC component within the circuit that may be running out of bandwidth or gain, or even if the oscillator's output filter has too much amplitude ripple.

Residual FM A frequency modulation created during amplitude modulation. Can be lessened by buffering between the transmitter's oscillator and its modulator.

Residual PM Undesired phase modulation during amplitude modulation.

Resistor noise Due to random electron motion, all resistors will produce *white noise*. This noise power will increase in amplitude with the increased value of the resistance, and can be considered as an intrinsic in-series noise generator internal to every resistance, with its amplitude further escalating with a rise in temperature and bandwidth. If we want a truly low-noise circuit design, then we must decrease resistance, bandwidth, and/or temperature. This generated resistor thermal noise can be calculated by:

$$V_{\text{RMS}} = (4 \cdot \text{k} \cdot T \cdot R \cdot \Delta f)^{1/2}$$

where k = Boltzman's constant
V_{RMS} = voltage level of the RMS noise, V
T = temperature of the resistance R, K
R = value of the resistance, Ω
Δf = brickwall circuit bandwidth containing the resistance R, Hz.

Resonator A component that vibrates or oscillates only at its natural resonant frequency, and will not efficiently resonate at any other frequency, thus acting as a high Q resonant circuit. A quartz crystal is such a resonator.

Return loss (RL) The measurement of the difference between the RF power sent toward the input of a circuit, and the power of the returned (reflected) RF power, in *dB*. Most circuits, such as filters, amplifiers, attenuators, and so on, can easily be designed with a return loss of 10 dB or higher. With this 10-dB value of return loss, only one-tenth of the power incident to the circuit's input will *not* be passed on to the circuit's load, but will be reflected back toward the original source. Normally specified as S_{11} for input RL, and S_{22} for output RL:

$$\text{RL (dB)} = 10 \log_{10}(P_{REF}/P_{INC})$$

where P_{REF} = reflected power, in W
P_{INC} = incident power, W
RL = return loss, dB.

RF Stands for *radio frequency*, and is considered to be the section of the electromagnetic spectrum between 50 kHz and 3 GHz.

Radio frequency integrated circuit (RFIC) Integrated circuits designed, optimized, and constructed strictly for RF use.

RF leakage The amount of electromagnetic energy that escapes from a cable, connector, component, or circuit.

Ringing Damped sine wave oscillations that take place within a resonant tank circuit when hit by a pulse of energy.

Roofing filter A term for a first IF filter with a very tight passband, located after the output of the receiver's first mixer stage. Common to CW and SSB HF receivers, these filters are constructed, generally, as a crystal filter type in order to significantly reduce the IF bandwidth, and to attenuate out-of-band signals that would cause overloading in downstream amplifiers and mixer stages.

Saturation The transistor conducting as hard as it can, with $V_{CE} \approx 0$ V. (However, $V_{CE(SAT)}$ never truly reaches 0 V due to the transistor's own natural internal resistances, and thus V_{CE} may be up to 2 V during saturation.)

Selective calling Opening a communications channel for only a specific receiver or receivers by a distinct tone-coded signal sent to the receiver(s) from a transmitter. This will then break the squelch of only the desired radio(s).

Selective fading A multipath fading that only occurs at certain frequencies.

Selectivity The capability of a receiver to select a single signal out of many, while eliminating all others. Thus, a superheterodyne receiver's IF bandwidth must be narrow enough, and with sharp enough skirts, to filter out undesired frequencies, but wide enough to not attenuate the sidebands of the desired signals.

Selectivity factor A measurement of a filter's skirt steepness, or its bandwidth in Hz at the 3-dB down point divided by the bandwidth at the 60-dB down point. A perfect brickwall filter would be equal to one.

Self-resonance Due to a passive component's distributed capacitive and inductive reactances, a point is reached at a specific frequency when the component becomes parallel or series resonant, and this is called self-resonance.

Self-quieting If an FM receiver captures a harmonic or other spurious signal of one of its own local oscillators, this can cause the actual received signal of interest to be quieted, or muted. This undesired action can be reduced in the design phase by confirming that no harmonics or spurious mixer products fall within the receiver's IF, and that all stages are properly shielded and/or laid out.

Sensitivity The capability of a wireless receiver to successfully recover and demodulate low-amplitude signals. Sensitivity is measured as the RF signal level, in dBm (or μV), required at the receiver's input port in order to produce an output signal with a specific signal-to-noise ratio or bit error rate. The receiver's sensitivity value will be highly dependent on the internal noise and gain of the receiver itself, most specifically at its first stages. Indeed, if we can lower the NF of these particular stages and/or narrow the bandwidth of the IF, both of which will decrease the noise, then the receiver's output SNR, and thus its sensitivity, will be improved. However, we will normally be given a specific IF bandwidth to design with for a specific modulation, which will also require a specific SNR. Therefore, we will only have the receiver's NF to optimize. Since the receiver's front-end stages set the ratio between the incoming signal and the radio's own noise, the rest of the receiver is mainly present to amplify the signal and the noise, while maintaining the same ratio and adding relatively little noise of its own downstream (the incoming RF signal and the added noise will now be at a much higher amplitude after the front-end amplification, so any added noise further on down the receiver chain contributes less and less to the overall noise level). This is why the NF of a receiver's front end must be minimized and its gain maximized, so that we can optimize the SNR into the detector for maximum receiver sensitivity. Sensitivity can be calculated by:

$$-174 + \mathrm{NF} + 10\log(\mathrm{BW}) + \mathrm{SNR}_{\mathrm{REQUIRED}}$$

Shadowing Certain areas within a wireless coverage area that are blocked from receiving a broadcast or signal due to foliage or structures blocking the line-of-sight transmission.

Shielding A conductive structure that is normally shaped as a metal box of assorted sizes and shapes, and is used to protect a sensitive system, circuit, or component from stray electromagnetic fields. Copper makes a very good RF shield, while soft iron will function with decreased capability. At low-frequency AC, however, copper is a poor *magnetic* shield, while soft iron is much better in this particular application. The lower the frequency, the thicker the shield must be to be effective, while the higher the frequency, the thinner it can be and still function properly.

Sideband cutting If a receiver's IF filters are too selective they may actually attenuate the desired sidebands of a signal, thus distorting the higher detected audio frequencies or increasing the BER in a digital system.

Signal-to-noise ratio (SNR) The ratio between the signal voltage or power, to the noise voltage or power, expressed in dB.

Simplex operation In wireless communications, simplex refers to the successive communications between two transceivers on the same frequency, but not at the same time and without the use of a repeater.

Sinking current When an integrated circuit has current flowing *into* its I/O pin (see *Sourcing current*).

Skin effect At high frequencies a conductor's effective resistance increases, caused by the RF currents being forced to travel closer to the conductor's surface by the higher reactance deeper within the metal. The skin effect, sometimes referred to as "RF resistance," will create higher elevated losses to microwave signals than may have been expected during simple calculations for resistance, reactance, or impedance.

Smith chart Once the dominant graphical method employed to assist engineers in the rapid design of RF-matching networks between two dissimilar complex impedances. The Smith chart comprises constant resistance circles and constant reactance arcs. Indeed, before the PC and the handheld calculator, working with even the simplest RF-related complex math equation was a painfully slow and monotonous procedure. Philip H. Smith of Bell Labs solved this problem with the advent of his now well-known nomographic chart, which permitted engineers to perform calculations that normally took a slide-rule, paper, pencil, and a significant expenditure of brainpower and time to perform. These same calculations now took only moments, and with significantly less chance of mundane errors creeping into the final result. In fact, the Smith chart is still used today, mainly within software as opposed to on paper, to graphically view an RF circuit's impedances, scattering parameters, reflection coefficients, unconditional stability, noise figure, constant gain, return loss, and VSWR, as well as to still assist some engineers in the design of RF-matching networks and to work with transmission lines. Smith charts are a standard display option in almost any piece of microwave test equipment or RF simulation software that works with complex impedances.

Snubber A series resistor and capacitor network, or an RC parallel tank with a series diode, as well as other RCD combinations, that can be used to damp out or absorb voltage overshoots and ringing in an inductive circuit, thus preventing semiconductor damage and EMI. Snubbers work by controlling the transient voltage's maximum rate of change.

Sourcing current When a circuit has current flowing *out* of its I/O pin (see *Sinking current*).

Space diversity Two antennas separated by distance in order to mitigate multipath effects: If one antenna is receiving the faded signal, the other antenna will hopefully be receiving a higher amplitude signal, and the antenna with the strongest signal is then chosen for amplification by the digital receiver (there are additional antenna diversity methods).

Spectrum The complete range of electromagnetic waves from sound waves to x-rays. The section of the electromagnetic spectrum of interest in most wireless communications is between 20 kHz (VLF) and 100 GHz (EHF).

Spread The tolerance variation of a device's characteristics. For instance, a JFET may be rated as having a I_{DSS} of 12 mA, but may really have a I_{DSS} of anywhere between 9 and 15mA.

Spurious emissions Referred to as *spurs*, these are interfering and undesired frequency emissions that are inside or outside of the bandpass of a transmitter or receiver.

Spurious free dynamic range (SFDR) The dynamic range of a receiver from its smallest detectable signal to the highest amplitude signal in which, after the injection of a two-tone

input signal, there are no third-order or spurious products greater than 3 dB above the noise floor, measured in dB. SFDR can be increased by raising the receiver's IIP3 and lowering the receiver's noise figure or IF bandwidth, and can be calculated by:

$$\text{SFDR (dB)} = 0.66(\text{IIP3} - \text{NF} - 10\log(\text{BW}) + 171)$$

where IIP3 = receiver's third-order input intercept point, dBm
NF = receiver's input noise figure, dB
BW = bandwidth of the receiver's IF, Hz.

String A number of series connected amplifiers or frequency multipliers.

Stub tuner A device that permits the convenient manual adjustment of its input/output impedances. A stub tuner can be used to temporarily match, or to purposefully mismatch (to test for stability), a source to its load. The device consists of two or three (*double stub* or *triple stub*) movable short-circuited metallic stubs that transform its input/output impedances to almost any value. The entire instrument looks similar to a trumpet, and can be placed in any RF environment that possesses a standard RF connector, such as an SMA, N, or 7 mm. A variety of different model tuners are available that can cover specific frequency ranges between 0.3 and 18 GHz, with the most popular units being manufactured by Maury, Inc.

Subharmonics The lower frequency multiples of a frequency multiplier that are still present at the multiplier's output, along with the higher desired multiplied frequency.

Susceptibility How easily an electronic device is adversely affected by another device's electromagnetic fields.

Telemetry The remote wireless transmission of measurement signals, such as temperature, position, pressure, frequency, speed, and so on.

Television interference (TVI) Hampering of a television's received signal by a producer of RF, such as a transmitter, computer, or the LO of a receiver.

TEM mode propagation The propagation mode for freespace and coaxial transmission lines. In this mode, the electric field and the magnetic field are propagating at right angles to each other (mutual orthogonality) through space or in an unbalanced transmission line.

Termination A nonreactive load, at the system's own characteristic impedance (normally 50 Ω), utilized to prevent reflections from a cable or circuit's open end. This termination is used to prevent damage to or unstable operation of a circuit.

Thermal noise power RF random noise that is generated by any object that is heated. The higher the temperature and the wider the bandwidth the more noise power emitted. Noise power can be calculated by:

$$N = \text{kTB}$$

where k = Boltzmann's constant
N = noise power, W
B = channel bandwidth, Hz
T = system's temperature (normally 290K, or 17°C), K.

Thermal runaway A series of events that occurs to a bipolar junction transistor that results in its destruction. Thermal runaway is caused by poor biasing, which allows increasing internal heat within the BJT to create a rise in current through the device, producing further heat until, if unchecked, the transistor is destroyed.

Third-order intercept point (TOIP or IP3) Specifies the linearity of a device by supplying a theoretically calculated value that can be found by injecting two closely spaced signals into a circuit or system until the third-order intermodulation frequency products hypothetically reach the same output power as the fundamental's frequency amplitude, in dBm. The higher this value the better, since it will indicate that the system is able to receive a desired signal even while a strong close-in interferer is transmitting, and with less negative consequences on the BER. Thus, the system's self-generated third-order tones will be at a lower amplitude then as compared to a system with a lower specified TOIP, indicating that a high TOIP system has superior linearity and less distortion at high input signal levels. Indeed, a high value of TOIP is critical to a receiver, since the third-order products are far too close to the desired signal to be effectively filtered out. TOIP is graphed as the input power ($IP3_{IN}$) or the output power ($IP3_{OUT}$) level when this theoretical crossing of the fundamental and third-order gain curves occur.

Transformer An electronic component that is capable of magnetically coupling power in its primary winding into its secondary winding, and can transform the voltage, current, and impedance to different values. A transformer has no true inherent impedance of its own, but merely reflects the impedance of the opposite winding. This impedance can be calculated by the turns ratio of the transformer times the impedance of the secondary load.

Transient intermodulation Distortion in an amplifier created by its inability to react properly to an input signal's rapid amplitude variations.

Transient voltage suppressors (TVS) Devices designed to send a possibly damaging transient voltage to ground by tripping into a shorted state when a certain maximum amplitude is reached, such as *varistors* (MOVs) and Zener diodes and other such components.

Transition region The region between cutoff and saturation in an active device.

Transmitter noise Noise created and transmitted by a wireless transmitter.

Trimmer A small, adjustable, and fine-tunable capacitor, resistor, or inductor.

Tuning range The bandwidth of frequencies that a VCO can effectively tune and still remain within its design specifications.

Tuning sensitivity (tuning gain or VCO gain) The VCO's frequency change specification in MHz per DC tuning voltage. For instance, a specification of 48 MHz/V means that for every 1 V_{DC} at the VCO's tuning port, the VCO will alter its output frequency by 48 MHz (until the VCO runs out of tuning range or DC voltage).

Tuning voltage The recommended maximum range of DC voltage that can be applied to a VCO's input tuning port while remaining within its design specifications.

Twisted pair A transmission line with a characteristic impedance of 100 Ω, and constructed of two (normally) unshielded wires that are twisted together. Utilized for short runs in low-frequency applications, twisted pair is able to decrease low-frequency magnetic hum pickup and RF-producing radiated fields.

Varistor A component that protects circuits against rapid transients, and consists of the *MOV* (metal oxide varistor) and the *ZNR* (zinc oxide resistor) devices. These function by decreasing the varistor's resistance when a spike of voltage reaches a certain input level, shunting the transient to ground.

Vector network analyzer (VNA) An electronic instrument that is invaluable for testing filters, linear amplifiers, attenuators, passive components, and so on. VNAs can sweep a chosen

bandwidth and then output a display of all four *S*-parameters, as well as impedance, phase, delay, Smith chart representations, log and linear magnitudes, VSWR, and so on. Many models of VNA are capable of sweeping the input power of an amplifier to test for *P*1dB, as well as operate up to 12 GHz (more exotic vector network analyzer versions can reach 110 GHz). Vital for accurate VNA measurements is a complete two-port calibration of the unit with a high-quality short, open, and load.

Nonlinear *large-signal VNA*'s (LSNA) are being developed that possess the advanced capability to accurately measure, under large signal input conditions, both slightly and strongly nonlinear networks. This makes them capable of being used on amplifiers all the way from a Class A low-noise amplifier that is slightly overdriven, up to a completely nonlinear Class C power amplifier.

Vias Small holes drilled through a PCB's substrate and plated with a metal conductor, used to connect an external layer to any other layer of the board (called a *through-hole via*). *Blind vias* are used to connect one external layer of a multilayer board to an inner layer or layers, while a *buried via* connects two or more inner layers.

Voltage-controlled crystal oscillator (VCXO) A voltage-controlled oscillator with a crystal as the resonator instead of an LC circuit. VCXOs possess superior frequency stability, but a very small tuning range.

Wave-trap A bandstop filter.

Wireless local area network (WLAN) A group of devices connected wirelessly so as to permit mobile, or untethered, connections to each other, the Internet, or shared peripherals (see *IEEE802.11 wireless and Bluetooth*).

Zeros A mathematical solution that is used in filter design to describe the points of a filter's transfer function that has zero insertion loss, and that forms a roll-up of 6 dB/octave. An ideal series capacitor will therefore act as a zero, a short circuit, at infinite frequency, while an ideal series inductor will act as a zero at DC. A Chebychev lowpass filter will have its zeros distributed around its passband, while a Butterworth's zeros will all be located at DC (see *Poles*).

This page intentionally left blank

Bibliography

Books

Abrie, Pieter L.D., *Design of RF and Microwave Amplifiers and Oscillators*, Artech House, ISBN 0-89006-797-X, 1999.

Albulet, Mihai, *RF Power Amplifiers*, Nobel Publishing, ISBN 1884932126, 2001.

Banerjee, Dean, *PLL Performance, Simulation, and Design,* 3d ed., National Semiconductor, ISBN 0-9708207-1-2, 2003.

Besser, Les, Gilmore, Rowan, *Practical RF Circuit Design for Modern Wireless Systems, Volume 1, Passive Circuits and Systems*, Artech House, ISBN 1-58053-521-6, 2003.

Besser, Les, Gilmore, Rowan, *Practical RF Circuit Design for Modern Wireless Systems, Volume 1I, Active Circuits and Systems*, Artech House, ISBN 1-58053-522-4, 2003.

Bowick, Chris, *RF Circuit Design*, Newnes, ISBN 0-7506-9946-9, 1982.

Carr, Joseph J., *Microwave & Wireless Communications Technology*, Newnes, ISBN 0-7506-9707-5, 1997.

Carr, Joseph J., *Practical Radio Frequency Test & Measurement*, Newnes, ISBN 978-0-7506-7161-3, 2002.

Carr, Joseph J., *Secrets of RF Circuit Design*, 2d ed., McGraw-Hill, ISBN 0-07-011673-3, 1997.

Cripps, Steve, *RF Power Amplifiers for Wireless Communications,* 2d ed., Artech House, ISBN 1596930187, 2006.

DeMaw, Doug, *Practical RF Design Manual*, MFJ Publishing, ISBN 1-891237-00-4, 1997.

Dungan, Frank, *Electronic Communications Systems*, Breton Publishers, ISBN 0-534-07698-X, 1987.

Dye, Norm, Granberg, Helge, *Radio Frequency Transistors*, Butterworth-Heinemann, ISBN 0-7506-9059-3, 1993.

Dye, Norman, Granberg, Helge, *Radio Frequency Transistors* 2d ed., Newnes, ISBN 0750672881, 2001.

Evans, Alvis J, Jerry D. Mullen, Danny H. Smith, Gerald Luecke (Editor), Kenneth M. Krone (Editor). *Basic Electronics Technology*, The Learning Center, 1985.

Feher, Dr. Kamila, *Wireless Digital Communications*, Prentice Hall, ISBN 0-13-098617-8, 1995.

Frenzel, Louis, *Communication Electronics*, McGraw-Hill, ISBN 0-02-801842-7, 1995.

Fry, Jim, *Electronic Circuits*, Heath Publishing, ISBN 0-87119-010-9, 1985.

Ginsberg, Gerald L., *Printed Circuits Design*, McGraw-Hill, Inc., ISBN 0-07-023309-8, 1990.

Golio, Mike (Editor), *RF and Microwave Semiconductor Device Handbook*, CRC, 2002.

Gottlieb, Irving M., *Practical Oscillator Handbook*, Newnes, ISBN 0-7506-6312-3, 1997.

Gottlieb, Irving M., *Practical RF Power Design Techniques*, McGraw-Hill, ISBN 0-8306-4129-7, 1993.

Gottlieb, Irving M., *Simplified Practical Filter Design*, TAB Publishing, ISBN 0-8306-8355-0, 1990.
Grob, Bernard, *Basic Electronics*, McGraw-Hill, ISBN 0-07-024928-8, 1984.
Grob, Bernard, *Electronic Circuits and Applications*, McGraw-Hill, ISBN 0-07-024931-8, 1982.
Hagen, Jon B., *Radio Frequency Electronics*, Cambridge University Press, ISBN 0521-55356-3, 1996.
Hall, M.P.M., Barclay, L.W., Hewitt, M.T., *Propagation of Radiowaves,* Institution of Electrical Engineers, ISBN 0852968191, 1996.
Hardy, James, *High Frequency Circuit Design*, Preston Publishing Company, Inc., ISBN 0-8359-2824-1, 1979.
Hayward, Wes, *Radio Frequency Design*, ARRL, ISBN 0-87259-492-0, 1996.
Hayward, Wes, DeMaw, Doug, *Solid State Design*, ARRL, ISBN 0-87259-040-2, 1995.
Heathkit, *Electronic Communications*, Heath Publishing, ISBN 595-2443-01, 1984.
Heathkit, *Semiconductor Devices*, Heath Publishing, ISBN 595-2875-05, 1984.
Hickman, Ian, *Practical Radio-Frequency Handbook,* 2d ed., Newnes, ISBN: 0-7506-34472, 1997.
Howard W. Sams (Editors), *Electronic Tables and Formulas*, Howard W. Sams Publishers, ISBN 0-672-21532-2, 1983.
Jung, Walter G. (Editor), *Op Amp Applications*, Analog Devices, ISBN 0-916550-26-5, 2002.
Kaufman, Milton, *Radio Operator's License Q&A Manual*, Hayden Books, Div. of Howard W. Sams & Co., ISBN 0-8104-0666-7, 1987.
Lancaster, Don, *Active Filter Cookbook,* 2d ed., Newnes, 0-7506-2986-X, 1996.
Larson, L.E., *RF and Microwave Circuit Design for Wireless Communications*, Artech House, ISBN 0-89006-818-6, 1996.
Laverghetta, Thomas S., *Handbook of Microwave Testing*, Artech House, ISBN 0-89006-070-3, 1981.
Laverghetta, Thomas S., *Microwave Measurements and Techniques*, Artech House, ISBN 0-89006-053-3, 1985.
Lenk, John, *Lenk's RF Handbook*, McGraw-Hill, ISBN 0-8306-4560-8, 1993.
Maas, Stephen, *Nonlinear Microwave and RF Circuits,* 2d ed., Artech House, ISBN 1580534848, 2003.
Maas, Stephen A., *RF and Microwave Circuit Design Cookbook*, Artech House, ISBN 0-89006-973-S, 1998.
Madhu, Swaminathan, *Electronics: Circuits and Systems*, Howard W. Sams, ISBN 672-21984-0, 1986.
Margolis, Art, *Computer Technician's Handbook,* McGraw-Hill, ISBN 0-8306-3279-4, 1990.
Marston, R.M., *Passive and Discrete Circuits Pocket Book*, Newnes, ISBN 0-7506-0857-9, 1997.
Matthaei, G., Young, L., Jones, E.M.T., *Microwave Filters, Impedance-Matching Network, and Coupling Structures*, Artech House, ISBN 0-89006-099-1, 1980.
Matthys, Robert J., *Crystal Oscillator Circuits*, Kreiger Publishing, ISBN 0-89464-552-8, 1992.
Metzger, Daniel L., Electronics *Pocket Handbook*, Prentice Hall, ISBN 0-13-251835-X, 1982.
Miller, Gary M., Beasley, Jeffrey S., *Modern Electronic Communication*, Prentice Hall, ISBN 0-13-016762-2, 2002.
Montrose, Mark I., *Printed Circuit Board Design Techniques for EMC Compliance*, Wiley Interscience, ISBN 0-7803-5376-5, 2000.

Motorola Literature, Authors: Various, *Motorola RF Application Reports*, Distribution, (Catalogue number HB215/D), 1995.
Navy Education and Training Command, *Electrical Engineering Training Series*, U.S. Navy, 1988.
Orr, William, *Radio Handbook*, Howard W. Sams Publishing, ISBN 0-672-22424-0, 1995.
Pease, Robert A., *Troubleshooting Analog Circuits*, Butterworth-Heinemann, ISBN 0-7506-9499-8, 1993.
Pozar, David M., *Microwave Engineering*, 2d ed., John Wiley and Sons, ISBN 0-471-17096-8, 1998.
Rhea, Randall W., *HF Filter Design and Computer Simulation*, McGraw-Hill, ISBN 0-07-052055-0, 1995.
Rhea, Randall W., *Oscillator Design and Computer Simulation*, 2d ed., McGraw-Hill, ISBN: 0-07-052415-7, 1995.
Rohde, Ulrich, Whitaker, Jerry, *Communications Receivers*, 2d ed., McGraw-Hill, ISBN 0-07-053-608-2, 1997.
Sayre, Cotter W., *Complete RF Technician's Handbook*, 2d ed., Prompt Publications, ISBN 0-7906-1147-3, 1998.
Schetgen, Robert, *The ARRL Handbook for Radio Amateurs*, ARRL, ISBN 0-87259-172-7, 1995.
Schmitt, Ron, *Electromagnetics Explained,* Newnes, ISBN 0-7506-7403-2, 2002.
Sergent, Jerry E., Harper, Charles A., *Hybrid Microelectronics Handbook,* 2d ed., McGraw-Hill, ISBN 0-07-026691-3, 1995.
Shrader, Robert, *Electronic Communication*, Glencoe, ISBN 0-07-057157-0, 1991.
Smith, Bradford L., Carpentier, Michel-Henri, Reinhold, Van Nostrand, *Microwave Engineering Handbook*, ISBN 0-442-13588-0, 1992.
Straw, R. Dean (Editor), *The ARRL Handbook for Radio Communications*, ARRL, 2006.
Thomas, Jeffery L., Edgington, Francis M., *Digital Basics for Cable Television Systems*, Prentice Hall, ISBN 0-13-743915-6, 1999.
Tocci, Ronald, *Fundamentals of Electronic Devices*, Merril Publishing, ISBN 0-675-09887-4, 1982.
Turner, Rufus, *The Illustrated Dictionary of Electronics*, Tab Books, ISBN 0-8306-1366-8, 1982.
U.S. Navy, *Basic Electronics*, Dover Publications, ISBN 0-486-21076-6, 1973.
Vizmuller, Peter, *RF Design Guide, Systems, Circuits and Equations*, Artech House, ISBN 0-89006-754-6, 1995.
Ward, Al, Sayer, Will, Pocock, Emil, Banghart, Richard P., Troetschel, William, Krauss, Geoffrey, Rautio, Dr. James C., et al. *ARRL UHF/Microwave Experimenter's Manual*, ARRL, ISBN 0-87259-312-6, 1997.
Williams, Tim, *The Circuit Designer's Companion*, Newnes, ISBN 0-7506-1756-X, 1991.
Wong, Kin-Lu, *Planar Antennas for Wireless Communications*, Wiley-Interscience, ISBN 0-471-26611-6, 2003.

Application Notes and Data Sheets

1.5A, Low Noise, Fast Transient Response LDO Regulators, App. Note LT1963A, Linear Technology, 2001.
1800 to 1900 MHz Amplifiers Using the HBF-0405 and HBFP-0420 Low Noise Silicon Bipolar Transistors, App. Note 1160, Hewlett Packard, 1994.

1800 to 1900 MHz Amplifiers Using the HBFP-0405 and HBFP-0420 Low Noise Silicon Bipolar Transistors, Agilent, 1999.
3-Stage Amplifier, Murata, November 2004.
3-Terminal Adjustable Regulator, National Semiconductor, 1998.
3-Volt Low Noise Amplifier for 0.8-6 GHz Applications, Hewlett-Packard (Agilent), 1999.
500mA-Peak Output LDO Regulator, App. Note MIC5219, Micrel, Inc., January 2005.
500MHZ to 2500MHz VCO Buffer Amplifiers, App. Note MAX2472/2473, Maxim Integrated Products, June 1999.
A 1.9 GHz Low Noise Amplifier Optimized for High IP3 Using BFP540, App. Note 057, Infineon Technologies, 2004.
A 2-Watt, 2 GHz Power Amplifier, App. Note AN172, AMCOM Communications, Inc., January 2001.
A 400, 900 and 1800 MHz Buffer/Driver Amplifier using the Agilent HBFP-0450 Silicon Bipolar Transistor, App. Note 1206, 2003.
A 5.0 GHz Bipolar Active Mixer, App. Note S010, Hewlett Packard (Agilent), 2004.
A High Third-Order Intercept Low Noise Amplifier for 1900 MHz Applications Using the Infineon Silicon-Germanium BFP620 Transistor, App. Note 060, Infineon Technologies, Eagleware Corp., 2005.
A Low Distortion PIN Diode Switch using Surface Mount Devices, HP App. Notes #1049, Hewlett-Packard (Agilent), 1997.
A Low-Cost Surface Mount PIN Diode Pi Attenuator, App. Note 1048, Hewlett-Packard (Agilent), 1997.
A Temperature Compensated Linear Power Detector, Avago Technologies, 2006.
About Ferrite Beads, App. Note 101, Linear Technology, 1999.
Active GaAs FET Mixers, Using the ATF-10136, ATF-13736, and ATF-13484, Hewlett-Packard (Agilent), 1992.
Active Mixer Designing Using the NE25139 Dual Gat MESFET, App. Note AN1020, California Eastern Laboratories, 1997.
Amplifier Design, App. Note 282A, Freescale Semiconductor, Inc., 2004.
An Analysis and Performance Evaluation of a Passive Filter Design Technique for Charge Pump PLL's, App. Note AN-1001, National Semiconductor Corp., 1996.
An SPDT PIN Diode T/R Switch for PCN Applications, App. Note 1067, Agilent Technologies, 1999.
Application of Microwave GaAs FETs, CEL AU82901-1-982, California Eastern Labs, 2001.
Applications for the HSMP-3890 Surface Mount Switching PIN diode, HP App. Note 1072, Hewlett Packard, 1992.
Applications for the HSMP-3890 Surface Mount Switching PIN Diode, App. Note 1072, Hewlett Packard, 1996.
AV104-12, AV104-12LF: GaAs IC 25 dB Voltage Variable Attenuator Single Positive Control 0.45-2.5 GHz, Skyworks, 2001.
Base Station Antennas for Digital Cellular Systems, Decibel Products, Inc., 1998
Basic MODAMP MMIC Circuit Techniques, App. Note S001, Hewlett-Packard (Agilent), 1993.
Basics of Dual Fractional-N synthesizers/PLLs, Skyworks, May 2005.
BCR400, App. Note 025, Siemens Semiconductor Group, 1999.
Biasing ERA Amplifiers, Mini-Circuits, 1998.

Biasing MSA Series RF Integrated Circuits, App. Note 5003, Hewlett-Packard (Agilent), 1997.
Bridged T Attenuator, www.rfic.co.uk, 2007.
Buffer Amplifiers Solve VCO Problems, App. Note 2019, Maxim, May 2003.
Capacitive Tap Matching, TSV Engineering, 1999.
Capacitor PI Network for Impedance Matching, App. Note 026, Modelithics, 2005.
Cascaded Input IP3 Calculation, RF Micro Devices, 2000.
Ceramic Input Capacitors Can Cause Overvoltage Transients, App. Note 88, Linear Technology, March 2001.
Charge Pump Gain, National Semiconductors, 2006.
Charge Pump Saturation Effects in PLL Frequency Synthesizers, Applied Radio Labs, 2000.
Chip RF Crossovers Can Eliminate Need for Multi-Layer P.C. Boards, Applied Microwave & Wireless, 2003.
Choosing an EMPOWER Grid, App. Note PN11, Eagleware Corp., November 2001.
Coaxial Stub Tuners, Maury Microwave Corp. June 2002.
Components\Box Modes.WSP, Agilent Genesys Help Files, 2005.
Computer Aided Engineering Solutions for RF & Microwave, Applied Computational Sciences, October 2006.
Constellation Displays, LeCroy Corp., App. Note LAB303A, LeCroy Corperation, 1999.
Converting Field Strength to Power, App. Note AN0005, RF Micro Devices, 1997.
Crystal Oscillator Basics and Crystal Selection for rfPIC and PICmicro Devices, App. Note AN826, Microchip Technology, 2002.
CTS Crystal Application Notes, CTS Corp., 1998.
DECT (1.9 GHz) Transmit—Receive PIN-Diode Switch, App. Note 007, Siemens Semiconductor Group, 2001.
Design for Assembly—Tips and Tricks, Ditron Manufacturing, 2006.
Design of "PI" and "T" Network Attenuators for Inter-Stage Buffering, App. Note AN0015, RFMD, 2002.
Design Tools, Q Microwave, 2005.
Design with PIN Diodes, App. Note APN1002, Alpha Industries, June 1999.
Device Characterization with Harmonic Source and Load Pull, App, Note 5C-044, Maury Microwave , December 2000.
Differential and Common Mode Noise, App. Note TEoEA-1, Murata, 1998.
Difficult Amplifier Specifications and Tradeoffs, Miteq, Inc., 2002.
Digital Down Conversion (DDC) Theory, Hunt Engineering, 2007.
Digital Modulation in Communications Systems—an Introduction, App. Note 1298, Hewlett-Packard (Agilent), 1997.
Digital Radio Theory and Measurements, App. Note 355A, Hewlett-Packard (Agilent), 1992.
Digital Receivers Bring DSP to Radio Frequencies, Pentek, Inc., 2000.
Digital Wireless System EVM Measurement, Wireless Conference, 1996.
Dual Fractional-N Synthesizers/PLLs, Skyworks, August 2003.
Eagleware Help Files, Agilent, 2006.
Eagleware Help Files, Eagleware Corp. 1999.
Effective Capacitance vs. Frequency, American Technical Ceramics, 2003.
Electromagnetic Compatibility and Interference, www.parker.com, 2007.
Electro-Magnetic Interference and Electro-Magnetic Compatibility (EMI/EMC), App. Note 34097, Vishay, 2001.
Electronic Prototyping: Tips and Pitfalls, University of Nebraska-Lincoln Electronics Shop Web Page, 1994.

EMI Testing Fundamentals, Laird Technologies, 2006.
Equipment Authorization, FCC, 2007.
Field Strength vs. Radiated Power, Procom, 2007.
Filter Design in Thirty Seconds, App. Note SLOA093, Texas Instruments, December 2001.
Filter Networks and Impendence, www.bobsamerica.com, 2007.
Filters, Mini-Circuits, September 1999.
Fractional/Integer-N PLL Basics, Texas Instruments, August 1999.
Frequency Tripler, Genesys Help Files, 2006.
Fundamentals of RF and Microwave Noise Figure Measurements, App. Note 57-1, Agilent Technologies, 2002.
Glossary of Specifications and Terms, Modco, Inc., 1998.
HF/VHF Power Transistor, App. Note BLW86, Philips Semiconductors, August 1986.
High Dynamic Range Receiver Parameters, Watkins-Johnson Company Tech Notes, 1980.
High Frequency Transistor Primer, Part II, Noise and S-Parameter Characterization, Hewlett Packard, 1993.
High Speed Design Techniques, Analog Devices, ISBN 0-916550-17-6, 1996.
High-Frequency Transistor Primer, Part 1, Silicon Bipolar Electrical Characteristics, Hewlett Packard, 1998.
How to Choose a Quartz Crystal Oscillator for the MAX1470 Superheterodyne Receiver, App. Note 1017, Dallas Semiconductor, March 2002.
HSMS—286X Series, Surface Mount Microwave Schottky Detector Diodes, Avago Tech, 2006.
IAM-8 Series Active Mixers, App. Note S013, Hewlett Packard, 1993.
ICs Simplify the Design of Digital Communications Links, RFMD, 1994.
Impedance Matching: A Primer, Electus Distribution, 2001.
Importance of 2nd Harmonic Tuning for Power Amplifier Design, App. Note 5C-045, Maury, March 2000.
INA Series RFIC Amplifiers, App. Note S112, Hewlett-Packard (Agilent), 1997.
Intermodulation Distortion, App. Note SWRA030, Texas Instruments, 1999.
Introduction to Indoor Radio Propagation, Spread Spectrum Scene, 1998.
Introduction to Single Chip Microwave PLL's, App. Note AN-885, National Semiconductor Corp., 1995.
Is Bias Current Necessary?, App. Note 987, Agilent Technologies, 1999.
ISM-Band and Short Range Device Antennas, Texas Instruments, 2003.
Kinds of Diodes, Hewlett Packard, 1997.
L, S Band SPDT Switch, California Eastern Laboratories, January 2001.
Layout Guidelines for MMIC Components, Hittite Microwave Corporation, February 2001.
Layout Guidelines for Switching Power Supplies, App. Note 1149, National Semiconductor, October 1999.
Linear Regulators in Portable Applications, *App*. Note 751, Dallas Semiconductor, May 2001.
LMX2306/LMX2316/LMX2326, Data Sheet, National Semiconductor Corp., 1999.
LNA/Mixer ICs Ease Wireless Receiver Design, RFMD, 1993.
Low Cost Integrated Solution for Analog Cellular RF Blocks, RFMD, 1995.
Low Noise Amplifiers for 1600 MHz and 1900 MHz Low Current Self-Biased Applications Using the ATF-35143 Low Noise PHEMT, Hewlett-Packard (Agilent), 1999.

Low Power Active Mixer, App. Note LT5560, Linear Technologies, 2002.
LSSP Simulation Description, Agilent ADS Help Files, 2005.
M/Filter Types, Chapter 9, Eagleware, Inc., 1998.
Match Impedances in Microwave Amplifiers, App. Note AN1033/D, Motorola, Inc., 1993.
MAX2620 Data Sheet, Maxim, May 1998.
Maximum Power Transfer Theorem, www.Microwaves101.com, 2007.
Microwave Alternatives, Competition and Markets, M-Pulse, 1998.
Microwave Bipolar/FET Bias Circuits, www.rfic.co.uk, 2007.
Microwave Transistor Bias Considerations, App. Note 944-1, Hewlett Packard, 1993.
Mixer 2x2 Spurious Response and IP2 Relationship, Dallas Semiconductor, December 2002.
Mixer: Types, Agilent Technologies, Genesys Mixer Help Files, 2005.
Modulating the Alphabet Soup, Analog Devices, 1999.
Most Important EMC Design Guidelines, Clemson Vehicular Electronics Laboratory, 2002.
Network Analyzer Measurements: Filter and Amplifier Examples, App. Note 1287-4, Hewlett Packard, 1999.
OCXO's: Oven-Controlled Crystal Oscillators, Wenzel Associates, Inc., 1995.
Oscillator Fundamentals—Understanding VCO Concepts, Mini-Circuits, 2001.
Oscillators NegR VCO, Genesys Help File, 2005.
Other Crystal Oscillator Types, Wenzel Associates, Inc., 1996.
Output Return Loss of High Power Class AB Amplifiers, App. Note AN056, 2002, Sirenza Microdevices, 2002.
Part Value Tolerance in Surface Mount Components, App. Note 022, Modelithics, 2006.
Passive S-Parameters Issues, Johanson Technology, 1999.
PC Board Layout Tips, SAWTEK, Inc., 1998.
PCB Layout Tips, Triquint Company, 2005.
PCB Thermal Capability, App. Note AN-0003, Anadigics, 2004.
PCB Trace Antenna Design Considerations and Implementation Guidelines for the AT86RF401 RF Wireless MicroTransmitter, App. Note, Atmel, February 2003.
Phase Noise—Technical Details, Radio Labs, 2006.
Phase Noise Measurement of Z-COMM VCOs, App. Note AN-109, Z-Communications, Inc., 2002.
Phase Noise, www.rfic.co.uk, 2007.
PLL Basics, www.altera.com. 2007.
Poles and Zeros, K&L Microwave Inc., 2005.
Proper Output Loading of Z-Comm VCOs, App. Note AN-102, Z-Communications, 2003.
PTI Filter Guide, PTI, Inc., 1997.
Quadrature Modulator/Demodulator RF 2703, RFMD, 2006.
Quadrature Modulator/Demodulator, App. Note RF2713, RFMD, 2006.
Radial Stub, Electrical Engineering Software, 2001.
Radiated Emissions Measurement How To, A.H. Systems, Inc. 2006.
Representative Active Resonator Parameters, PTI, Inc., 1997.
RF Connector Interface Styles and Applications, Johnson Components, 1997.
RF Design for GPS Receivers, U-blox Ag., 2006.
RF Power Device Impedances: Practical Considerations, App. Note AN1526, Freescale, 1991.
RF Related Conversions, A. H. Systems, 2006.
RF Transmitting Transistor and Power Amplifier Fundamentals, Philips Semiconductors, March 1998.

RFIC Components for Cordless Phones, RFMD, 1994.
RO4003 High Frequency Material Insertion Loss Comparison with other Material Types, App. Note RO2.9.7, Rogers Corporation, 2000.
SA8025 Fractional-N Synthesizer for 2GHz Band Applications, Philips Semiconductors, August 1997.
Selecting a Wireless LAN Technology, Proxim, 1997.
Silicon Bipolar MMIC 56GHz Active Double Balanced Mixer/IF Amp, Hewlett-Packard (Agilent), 1997.
Silicon Variable Capacitance Diode, Siemens Semiconductor Group, October 2004.
Simulating the Effect for Mounting on SRF and S-Parameters for High Frequency Multi-Layer Ceramic Capacitors, Johanson Technology, May 2000.
Smith Chart, Agilent Technologies, 2002.
Software-Defined Radio, Wipro Technologies, Inc., 2002.
Solectek's Wireless Multipoint Media Access Control Protocol, Solectek, Inc., 2006.
Solid-State Phase-Locked Microwave Signal Sources, CTI, Inc., 1997.
Sources of Inaccuracy for S-Parameter Data, Johanson Technology, April 1999.
S-Parameter S2P File Format Guide, App. Note AN3009, Tyco Electronics, 1993.
S-Parameter Techniques, App. Note 95-1, Hewlett-Packard (Agilent), 1996.
SRF & PRF and Their Relation to RF Capacitor Applications, Johanson Technology, January 1999.
Streamline Circuits, Capabilities, Streamline Circuits, 2007.
Surface Mount PIN Diodes, Hewlett Packard, 1999.
Surface Mount, Multilayer Ferrite Beads, App. Note 34146, Vishay, 1999.
Switches, Electronics Club, 2003.
Switches, www.minicircuits.com, September 1999.
Quadrature Modulator/Demodulator IC, TA 0019, RFMD, 1999.
TCXO's – Temperature Compensated Crystal Oscillators, Wenzel Associates, 1999.
TechTip 3, Rogers Corp., 1998.
TechTip 5, Rogers Corp., 1998.
TechTip 8, Rogers Corp., 1998.
Testing and Troubleshooting Digital RF Communications Designs, Agilent Technologies, 2002.
Testing and Troubleshooting Digital RF Communications Receiver Designs, App. Note AN 1314, Agilent Technologies, 1995.
The Advantages of Constant Envelope Modulation, http://sss-mag.com/sstopics.html, 2008.
The Design of 2.4GHz Bipolar Oscillator by Using the Method of Negative Resistance, ADS App. Note, Microwave Laboratory, September 2001.
The Design Principles of Buffer Amplifiers, http://my.integritynet.com.au/purdic, 2007.
The IEEE 802.11 Wireless LAN Standard, WLANA (The Wireless LAN Alliance), 2007.
The Importance of 2nd Harmonic Tuning for Power Amplifier Design, Application Note 5C-045, Maury Microwave (Gary R. Simpson and Mark Vassar), 2003.
The Quest for the Ideal RF Amplifier, Amplifier Research, Inc., 1999.
Theory of Intermodulation Distortion Measurement (IMD), Maury Microwave Corp. July 1999.
Thermal Management for Surface Mount Components, Hittite Microwave Corporation, 2001.
Three Methods of Noise Figure Measurement, Dallas Semiconductor, November 2003.
Two-Diode Odd-Order Frequency Multipliers, Wenzel Associates, 1996.
UHF and Microwave Designer's Handbook, Watkins-Johnson Company, 1998.

UHF Power Transistor, App. Note BFG21W, Philips Semiconductors, 1998.
Ultra Low Noise Amplifiers for 900 and 2000 MHz with High IP3, App. Note KV96-157, Philips Semiconductors, 2002.
Understanding and Enhancing Sensitivity in Receivers for Wireless Applications, App. Note SWRA030, Texas Instruments, May 1999.
Understanding Dynamic Hardware Specifications, National Instruments Corp., 2006.
Understanding Ferrite Bead Inductors, App. Note TE04EA-1, Murata, 2003.
Understanding Model Parameters and Applications Limitations, App. Note 23, Zetex, March 1996.
Understanding Model Parameters and Applications Limitations, App. Note AN23, Zetex, 2001.
Understanding Radio for the Practical Engineer-Part One, Mike Wilson, C. Eng., 2006.
Understanding Radio for the Practical Engineer-Part Three, Mike Wilson, C. Eng., 2006.
Understanding Radio for the Practical Engineer-Part Two, Mike Wilson, C. Eng., 2006.
Understanding the FCC Regulations for Low-Power, Non-Licensed Transmitters, Office of Engineering and Technology Federal Communications Commission, October 1993.
Understanding VCO Concepts, Mini-Circuits, Inc. September 1999.
Using a Network Analyzer to Characterize High-Power Components, Agilent Technologies, 2003.
Using Ferrite Beads to Keep RF out of TV Sets, Telephones, VCR's, Burglar Alarms and Other Electronic Equipment, Palomar Engineers.
Using Genesys to Design Power Amplifiers, App. Note 21, Eagleware Corp., August 2003.
Using Peregrine Phase-Locked Loop Integrated Circuits in Reference and System Clock Applications, App. Note AN16, Peregrine Semiconductor, 2003.
Using Schottky Diode Mixers, SM 5 BSZ, June 2001.
Using the ATF-10236 in Low Noise Amplifier Applications in the UHF through 1.7 GHz Frequency Range, Agilent, 1999.
Using the BCR410W Bias Controller with BFP405 in Amplifier Circuits, App. Note 064, Infinion, February 2003.
Using the HMC199MS8 as a Low Cost 1-Bit Attenuator, Hittite Microwave Corporation, 2005.
Using the MSA-0520 and MSA-1023 Medium-Power MODAMP Silicon MMIC Amplifiers, App. Note S007, Hewlett-Packard (Agilent), 1993.
Using Vector Modulation Analysis in the Integration, Troubleshooting and Design of Digital RF Communications Systems, Product Note, HP 89400-8, Hewlett Packard, 1994.
VCXO's—Voltage Controlled Crystal Oscillators, Wenzel Associates, 2000.
VHF Quadrature Modulator, RFMD, 2002.
Voltage Controlled Oscillators, Synergy Microwave Corp., 2003.
Voltage Controlled Oscillators, www.Electronics-tutorials.com, 2007.
Wave Propagation Along a Transmission Line, Agilent Technologies, 2007.
Waveguides Below Cutoff-Part 4: Shielding, Nutwood UK LTD., 1999.
Which Simulator Should I Use?, Eagleware Help Files, 2003.
Wireless Analog Design Solutions Guide, 8th Edition, MAXIM, 1999.
XOs—Non-Compensated Crystal Oscillator, Wenzel Associates, Inc., 2000.

Papers and Magazines

Abidi, Asad A., "Direct-Conversion Radio Transceiver for Digital Communications," IEEE, Aug. 29, 1995.
Ackerly, Bryan, "Simulation Tools, Real Components and Limitations," http://users.bigpond.net.au, 2007.

Adams, Greg, "Designing Resistive Unequal Power Dividers," Moorestown Microwave Co., 2007.

Akamatsu, Sonoko, Baylis, Charles, Dunleavy, Larry, "Accurate Simulation Models Yield High-Efficiency Power Amplifier Design," *IEEE Microwave Magazine*, December 2005.

Andren, Carl A., "Comparison of Frequency Hopping and Direct Sequence Spread Spectrum Modulation for IEEE 802-11 Applications at 2.4 GHz," Harris Semiconductor, April 1997.

Andren, Carl, "ZIF Makes Dual-Band WLAN Radios Practical," Intersil Corp, 2003.

Armstrong, Eur Ing, Keith, "Design Techniques for EMC Part 4: Shielding," www.emc-journal.co.uk, 2007.

ARRL, "FCC Part-15 Rules: Unlicensed RF Devices," ARRL, 2006.

Aubin, John F., "A Brief Tutorial on Antenna Measurements," *Microwave Journal*, August 2005.

Baker, Thomas, "Successful LNA Design Involves Performance Trade-Offs," www.rfdesign.com/, November 2006.

Banerjee, Dean, "Changes in PLL Parameters Impact Receiver System Performance," www.wirelessdesignonline.com, April 2001.

Banerjee, Dean, Boyer, Paul, "Delta Sigma PLLs Raise the Standard for Performance," National Semiconductor, 2006.

Banerjee, Dean, "PLL Building Blocks," National Semiconductor, 2003.

Banerjee, Dean, "PLL Performance," National Semiconductor, 2003.

Barel, Alain, "Nonlinear Microwave Measurements in Belgium," *HF Journal*, April 2002.

Basraoui, Mahmoud, Prasan, S.N., "Wideband, Planar, Log-Periodic Balun," Bradley University, 2003.

Bateman, Andrew, "Theory and Implementation of 16-QAM Radio Modems," Wireless Systems International Limited, 1998.

Bensky, Alan, "Range Estimation for Short-Range Event Transmission Systems," *RF Design* Magazine, November 2002.

Bhatti, Saleem, "Design Limitations on Maximum Data-Rate and Channel Capacity," www.cs.ucl.ac.uk, 2007.

Bible, Steven, "Selecting Crystals for Stable Oscillators," *Microwaves & RF,* November 2002.

Blevins, Bruce A., "Small Satellite Antennas," PSL, 1998.

Bogatin, Eric, "What Really is Inductance?," Bogatin Enterprises, LLC, October 1999.

Brannon, Brad, "Basics of Designing a Digital Radio Receiver (Radio 101)," Analog Devices, Inc., 2005.

Brannon, Brad, "Basics of Designing a Digital Receiver," Analog Devices, 2001.

Breed, Gary A., "Transistor Biasing Issues for Linear Amplification of Complex Signals," *High Frequency Electronics*, March 2003.

Breed, Gary, "A Review of Key Oscillator Specifications and What They Mean," *High Frequency Electronics*, November 2003.

Breed, Gary, "Intermodulation Distortion Performance and Measurement Issues," *High Frequency Electronics*, May 2003.

Breitbarth, Jason, "Octave Tuning, High Frequency Varactor Oscillator Design," Thesis, Oregon State University, 1997.

Brink, Anton, "Resonators," Department of Electroscience, Lund Institute of Technology, 2001.

Brown, Jim, "Understanding and Solving RF Interference Problems," Audio Systems Group, 2007.
Browne, Jack, "Coaxial Tuners Control Impedances to 65 GHz," *Microwaves & RF*, January 2003.
Bryant, James M., *Practical RF Design*, 1994.
Bullinger, Rex, "How to Measure Carrier-to-Noise," *CED: Communications Engineering & Design*, November 1994.
Burgess, Larry, "Radiated power and Field Strength from UHF ISM Transmitters," *High Frequency Electronics*, November 2006.
Butler, Lloyd, "Negative Resistance Revived," *Amateur Radio Magazine*, November 1995.
Capwell, John, Clausen, Bill, Weller, Tom, Dunleavy, Larry, "Accurate Models Simplify Reference Designs for RFIC Amplifiers," High Frequency Electronics, 2005.
Carioca, Cezar A., DeCarvalho, Paulo H., Abdalla Jr., Umberto, "Computer-Aided Design of Diode Frequency Multipliers," *Applied Microwave & Wireless* Magazine, April 1999.
Carlini, Jim, "A 2.45 GHz Low Cost, High Performance VCO," *Microwave Journal*, 2001.
Chavers, Tracey, "Micro-Coax Pigtails for RF Measurements to and Beyond 5 GHz," *RF Design* Magazine, February 2006.
Chen, Yen-Ming, "RC Snubber Design Using Root-Loci Approach for Synchronous Buck SMPS," University of Waterloo, 2005.
Croston, Robin, "Effects of Parasitics in Circuit Simulations," Applied Microwave & Wireless, 2001.
Curtin, Mike, O'Brien, Paul, "Phase-Locked Loops for High-Frequency Receivers and Transmitters-Part 2," Analog Dialogue, 1999.
Cushing, Rick, "Replacing or Integrating PLL's with DDS Solutions," Analog Devices, 1999.
Davis, Jr. "Stabilization circuit for Transistor RF Power Amplifiers," USPO , September 1980.
Dearn, Andrew, "How to Design RF Circuits," Plextek Ltd.
Dearn, Andrew, "Oscillators," Plextek Ltd.
Delphy, Ulrich, "Stabilize Transistors in Low-Noise Amplifiers," *Microwaves & RF*, May 2006.
Dittmer, Tim W., "Advances in Digitally Modulated RF Systems," Harris Corp., 1997.
Dixon, Paul, "Dampening Cavity Resonance using Absorber Material," www.rfdesign.com, May 2004.
Doherty, Bill, "PIN Diode Fundamentals," MicroNote Review 701, Microsemi Corp., 1998.
Dr. Math, "Order of Operations," http://mathforum.org/dr.math/faq/faq.order.operations.html, 2007.
Dunleavy, Dr. Lawrence P., Liu, Dr. Jiang, "Understanding P2D Nonlinear Models," *Microwaves & RF*, July 2007.
Eriksson, Hans, Waugh, Raymond W., "A Temperature Compensated Linear Diode Detector," www.rfdesign.com, June 2000.
Evans, John, Microwave "Circuit Board Design and Manufacturing Considerations," Novacom Microwave, 1999.
Fakatselis, John, "Processing Gain in Spread Spectrum Signals," Harris Semiconductor, 2000.

Fiore, Richard, "Capacitors in Broadband Applications," *Applied Microwave & Wireless*, May 2001.
Fitton, Mike, "Principles of Digital Modulation," Toshiba-Trel, 1997.
Fox, Adrian, "What is a PLL Synthesizer?," Analog Dialogue, 2002.
Gear, Jack, T., "Microwave Absorbers Manage Military Electronics RF Interference," www.rfdesign.com, August 2004.
Gerke, Daryl, Kimmel, William, "EMI and Circuit Components: Where the Rubber Meets the Road," www.ednmag.com, September 2000.
Gilbert, Barrie, Nash, Eamon, "Controlling RF Power Transmission using a Demodulating Logarithmic Amplifier," Analog Devices, 1998.
Gorin, Joe, "Detector Selection for Spectrum Analyzer Measurements," *RF Design* Magazine, February 2003.
Graves, Danny R., "Introduction to Printed Circuit Board Prototyping (2PDH)," www.pdhengineer.com, 2005.
Green, Arnold H., "Shielding Magnetic Fields," *American Institute of Physics*, October/November 2001.
Handlon, Dennis, "Making EMI-Radiated Emissions Measurements," *Microwaves & RF* magazine, April 2000.
Harrison, Linden, "Introduction to depletion-mode MOSFETs," www.wirelessnetdesignline.com, August 2006.
Hartley, Rick, "Controlling radiated EMI through PCB Stack-Up," www.pcdmag.com, July 2000.
Hartley, Rick, "Designing PCBs with Mixed Materials," Printed Circuit Design & Manufacture , March 2006.
Hendriks, Paul, Schreier, Richard, DiPilato, Joe, "High Performance Narrowband Receiver Design Simplified by IF Digitizing Subsystem in LQFP," Analog Dialogue, 2001.
Holladay, Ken, "Design Loop Filters for PLL Frequency Synthesizers," Microwaves & RF, September 1999.
Howlad, Rob, "Impairments R Us," *Communications System Design* Magazine, 1999.
Huang, Shang Hsiu, "Modeling and Analysis of the Charge-Pump Phase-Locked Loop," National Chiao-Tung University, 2005.
Hunter, Mark, "The Basics of Radio System Design," Plextek Communications.
Hymowitz, Charles E., Sands, Bill, "Simulating Class C RF Amplifiers," Intusoft Spice, 1996.
Jacob, Gerald, "Power Meters Keep Up with Complex Digital Signal Demands," Nelson Publishing, 1998.
Jarvis, T.P., Beng, Marriott I.R., "Improving the Immunity of Sensitive Analogue Electronics," *EMC Journal*, 1997.
Jarvis, Tim, "Best Practice in Circuit Board Design," KTL-Hull, 1997.
Jeffries, D., "Transmission Lines," January 1999.
Jeganathan, K., "Design of a Simple Tunable/Switchable Bandpass Filter," *Applied Microwave and Wireless,* March 2000.
Johnson, Dr. Howard, "Bypass Capacitor Layout," *Signal Consulting, Inc.*, August 1997.
Kameche, Samir, Feham, Dr. Mohammed, Kameche, Dr. Mohamed, "PLL Synthesizer Tunes DCS11800 Band," *Microwaves & RF* magazine, June 2007.
Kang, Wei, Hong, Wei, Zhao, Jianing, "Optimized Transceiver Aids Broadband Wireless Access," *Microwaves & RF* magazine, June 2007.

Katz, Prof. Randy H., "Radio Propagation," University of California, Berkeley, 1996.
Kelly, Bob, "How to Debug a PLL Frequency Synthesizer," www.rfdesign.com, February 2004.
Koo, Raymond, "RF Switches," Microsemi Corp., 2002.
Koren, Victor, "Design of a Constant Insertion Loss Variable Frequency LC Band Pass Filter," *Applied Microwave & Wireless,* 2000.
Kraemer, Bruce, "Digital Modulation: Today and Tomorrow," *Portable Design* Magazine, December 1997.
Kraemer, Bruce, "Nonlinearity Effects in Wireless," *Portable Design* Magazine, May, 1997.
Lakshminaraynan, V., "Minimizing Failures in Electronic Systems by Design," www.ednmag.com, August 2000.
Lascari, Lance, "Accurate Phase Noise Prediction in PLL Synthesizers," Applied Microwave & Wireless, 2000.
Laursen, Soeren, "Output-Matching-Circuit Performance Critical for PA Design," *Portable Design*, October 2004.
Lazarus, Mitchell, "Keeping the FCC Happy—A Reference Article," *Wireless Design and Development* Magazine, November, 1999.
Lesea, Austin, "Error Rate: What is it? What Does it Mean?," Xilinx, March 2004.
Lesurf, Jim, "Negative Resistance Oscillators," University of St. Andrews, Scotland, 2005.
Lewallen, Roy W., "On Solid-State PA Matching Networks," *QST*, April 1983.
Lewallen, Roy, "RMS Power," EZNEC, 2004.
Loewen, Myron, "Designing an FCC Approved ASK rfPIC," Microchip Technology, Inc., 2002.
Lucek, Jarek, Damen, Robbin, "LNA Design for CDMA Front End," NXP, 2001.
Lundheim, Lars, "On Shannon and Shannon's Formula," Norwegian University of Science and Technology, 2002.
Lythall, Harry, "Synthesizers Basics," SMOVPO, March 2001.
Maloratsky, Leo, G., "Reviewing the Basics of Microstrip Lines," *Microwaves & RF*, March 2000.
Mannion, Patrick, "Direct Conversion Prepares for Cellular Prime Time," Electronic Design, November 1999.
Martell, Mike, "Using Bipolar Transistors As Switches," www.rason.org, September 2001.
McKeen, Trent, "Flexible Ferrite Gasket for Receiver Second Image Protection," (Patent, U.S.) 2003.
McLarnon, Barry, "VHF/UHF Microwave Radio Propagation: A Primer for Digital Experimenters," TAPR/ARRL Digitial Communications Conference, 1998.
McMorrow, Scott, "Handling Signal Return Current," *Printed Circuit Design*, September 2002.
MDS, "Troubleshooting Scada," www.sine-wave.com, 2002.
Mercer, Dr. Sean, "Minimizing RF PCB Electromagnetic Emissions," www.rfdesign.com, January 1999.
Mercer, Sean, "Exact Simulation of LNAs Reduces Design Cycle Time," Applied Microwave & Wireless, 1999.
Miceli, Andrew, "Digital Modulation," *Wireless Review*, December 1998.
Minsoek, Kim, "A Study of Implementation of Digital Signal Processing," Yokohama National University, Japan, 2002.

Mohammed Ali, Firas, "Practical Approach Yields Class C PA," *Microwaves & RF*, October 2006.

Mühlhaus, Dr. Ing Volker, "Circuit Simulator or Electromagnetic Field Solver?," www.muehlhaus.com, December 2001.

Neal, Lee, "Open and Closed Loop Characteristics of PLL," Micro Linear, 2003.

Neal, Lee, "Regulatory Considerations When Using ML2722, ML2724, and ML5800," Micro Linear, October 2004.

Neu, Thomas, "Designing Controlled Impedance Vias," www.edn.com, October 2003.

O'Hearn, P.E., Green, David, "A 2.3 GHz Transceiver Module Using Low Temperature Co-fired Ceramics," CiteSeer, 2001.

Olney, Barry, "EMC Design for High Speed PCB's," *VeriBest*, 1996.

Raab, Frederick, H., Asbeck, Peter, "RF and microwave Power Amplifier and Transmitter Technologies—Part 2," *High Frequency Electronics*, May 2003.

Rhea, Randall W., "Different Zin and Zout," Eagleware Corp., February 2004.

Rhea, Randall W., "Exploiting Filter Symmetry," *Microwave Journal*, March 2001.

Rhea, Randall W., "Transforms Aid the Design of Practical Filters," *Applied Microwave & Wireless,* 2002.

Rhea, Randall W., "Transmission Zeros in Filter Design," *Applied Microwave & Wireless,* 2001.

Rhea, Randall, W., "Filters with Complex-Impedance Terminations," Eagleware, 2002.

Rhea, Randy, "New Class of Oscillators," *IEEE Microwave Magazine*, June 2004.

Rhea, Randy, "The Yin-Yang of Matching: Part 1—Basic Matching Concepts, *High Frequency Electronics*, 2006.

Rhea, Randy, Clausen, Bill, "Recent Trends in Oscillator Design," *Microwave Journal*, January 2004.

Rivas-Torres, Wilfredo, "Guidelines for Successful Simulation Setup," High Frequency Electronics, December 2005.

Rivas-Torres, Wilfredo, "Using S-Parameter Data Effectively" *Planet Analog*, March 2006.

Rivas-Torres, Wilfredo, "Using S-Parameters Data Effectively," CommsDesign, March 2006.

Robertson, Dave, "Digital Communication Systems in Introduction," Analog Dialogue, 1996.

Robertson, Dave, "Selecting Mixed Signal Components for Digital Communication Systems—Sharing the Channel," Analog Devices, Inc., 1999.

Roscoe, Ron, "The Decibel Unit of Measurement," MIT Department of Electrical Engineering & Computer Science, 2003.

Sayre, Cotter, W., "Accurate RF Circuit Simulation," Modelithics, November 2006.

Sayre, Cotter W., "Combline BPF for ML5800 Reference Design," Micro Linear, July 2004.

Sayre, Cotter W., "Distributed LPF for ML5800 Reference Design," Micro Linear, July 2004.

Sayre, Cotter W., "Feasibility Report on a Tunable ISDB-T_{SB} Bandpass Filter," Micro Linear, July 2005.

Sayre, Cotter, W., "Lumped Discrete Bandpass Filter for TX Output of ML5800 RFIC," Micro Linear, January 2005.

Sayre, Cotter, W., Moore, Steve, "Minuteman RF Board Specification," Micro Linear, December 2002.

Scalpi, Anthony M., "Crystal Oscillator Design and Negative Resistance," Cypress Semiconductor, 2005.
Schmitt, Ron F., Allen, John W., "Designing and EMC-Compliant UHF Oscillator," www.rfdesign.com, October 2000.
Sevic, John, "Introduction to Tuner-Based Measurement and Characterization," Maury Microwave, August, 2004.
Sifri, Jack, Kanaglekar, Niranjan, "Simulation Tools Converge on Large RFICs," Communication Systems Design, June 2002.
Silver, H. Ward, "Hands-On Radio, Experiment # 22—Stubs," *AARL.org*, November 2004.
Silver, J.P., "Oscillator Basics Tutorial," www.rfic.co.uk, 2007.
Simpson, Chester, "Capacitors are Key to Voltage Regulator Design," National Semiconductor, 2005.
Simpson, Chester, "Linear Switching Voltage Regulator Fundamentals," National Semiconductor, 1995.
Skolnick, David, Levine, Noam, "Why Use DSP?," Analog Devices, 1999.
Smithson, Geoff, "Introduction to Digital Modulation Schemes," Plextek Ltd.
Smithson, Geoff, "Practical RF Printed Circuit Board Design," Plextek Ltd.
Srirattana, Nuttapong, "Power Amplifier Design Considerations," Georgia Institute of Technology, 2005.
Stanford Microdevices, "SXA/SXT-289 New Bias Scheme,": Stanford Microdevices, August 2000.
Stauffer, George H., "Finding the Lumped Element Varactor Diode Model," *High Frequency Electronics* magazine, November 2003.
Stein, John C., "Indoor Radio WLAN Performance, Part II: Range Performance in a Dense Office Environment," Harris Semiconductor, 1999.
Struzak, R., "Digital Modulation Basics," International Centre for Theoretical Physics, January 2003.
Subbiah, Palani, "Bit-Error Rate (BER) for High-Speed Serial Data Communication," Cypress Semiconductor, 2001.
Sztein, Eduardo, "RF Design of a TDMA Cellular/PCS Handset, Part I," www.uwcc.org, 2007.
Thornber, Martin, "Amplitude Equalizer Flattens Gain from 50 MHz to 7 GHz," Microwaves *& RF*, October 2007.
Troop, Nick, "Spur Analysis and Out of Band Transmit Mask for ML5800 Implications for Filter Requirements," Micro Linear, April 2003.
Tsaliovich, A., "The 20H Rule vs. Shielding," Besser Associates, 1999.
Umstattd, Ruth, "Operating and Evaluating Quadrature Modulators for Personal Communication Systems," National Semiconductor, 1993.
VanDyke, Rulon, "Relating Cascaded Noise Figures to Real-World Performance," *RF Design*, January 2003.
VanDyke, Rulon, "Using CAD for RF Architecture Engineering and Optimization," Eagleware Corp., 2006.
Vig, John R., "Introduction to Quartz Frequency Standards," Communications-Electronics Command, U.S. Army, 1979.
Vogel, Mark O., "Key RF Upconverter Parameters for Optimizing Your Cable Modem System," 3Com Corp, 2000.
Vogel, Mark O., "Measuring the Power of a Digital Signal," 3Com Corp., 1999.
Wagle, Kaustubh, Dharia, Anuj, "The Ups and Downs of Digital Data Conversion," Evaluation Engineering, May, 2006.

Waugh, Raymond W., "Choosing the Right Diode for your AGC Detector," Agilent Technologies, July 1999.
Waugh, Raymond W., "Large Signal Detectors for Cellular Handsets and Base Stations," Agilent Technologies, 2002.
Wenzel, Charles, "Switching Diode Frequency Doublers," Wenzel Associates, 2000.
White, Ian, "Transistor PA Bias Circuits," *Radcom* Magazine, December 1995.
White, Joseph F., "Achieving High Gain Over Broad Bandwidths," *Microwaves & RF*, January 2006.
White, Joseph F., "Cascading Amp Stages for Higher Gain," *Microwaves & RF*, January 2006.
Wight, Jim, "Delta-Sigma Fractional-n Synthesizers: Enabling Next-Generation Wireless Applications," Hearst Electronic Products, 2006.
Williamson, Steve, "How to Design RF Circuits—Synthesizers," Plextek Communications.
Wright, Helen, "Testing Digital Communications Transmitters and Receivers," *Microwave Journal*, July 2000.
Yeap, Y.W., Chua, L.H., Tan, S.H., "Design of a 3.5 W S-Band Power Amplifier Based on Small-Signal S-Parameter Analysis," *Microwave Journal*, July 2004.
Yu, Jong Won, "Receiver Nonlinear Performance," Kaist University, 2006.
Zhou, Jianyi, Zhao, Jianing, "A Low Cost RF Transceiver for 2.4 GHz DSSS Short-Range Communication Systems," *Microwave Journal*, June 2004.
Zyren, Jim, Petrick, Al, "Tutorial on Basic Link Budget Analysis," Intersil Corp. June 1998.

Index

This page intentionally left blank

A

B

C

D

E

F

G

H

I

M

P

Q

R

S

T

Y

Z

This page intentionally left blank

CD WARRANTY

This software is protected by both United States copyright law and international copyright treaty provision. You must treat this software just like a book. By saying "just like a book," McGraw-Hill means, for example, that this software may be used by any number of people and may be freely moved from one computer location to another, so long as there is no possibility of it being used at one location or on one computer while it also is being used at another. Just as a book cannot be read by two different people in two different places at the same time, neither can the software be used by two different people in two different places at the same time (unless, of course, McGraw-Hill's copyright is being violated).

LIMITED WARRANTY

Customers who have problems installing or running a McGraw-Hill CD should consult our online technical support site at http://books.mcgraw-hill.com/techsupport. McGraw-Hill takes great care to provide you with top-quality software, thoroughly checked to prevent virus infections. McGraw-Hill warrants the physical CD contained herein to be free of defects in materials and workmanship for a period of sixty days from the purchase date. If McGraw-Hill receives written notification within the warranty period of defects in materials or workmanship, and such notification is determined by McGraw-Hill to be correct, McGraw-Hill will replace the defective CD. Send requests to:

McGraw-Hill
Customer Services
P.O. Box 545
Blacklick, OH 43004-0545

The entire and exclusive liability and remedy for breach of this Limited Warranty shall be limited to replacement of a defective CD and shall not include or extend to any claim for or right to cover any other damages, including, but not limited to, loss of profit, data, or use of the software, or special, incidental, or consequential damages or other similar claims, even if McGraw-Hill has been specifically advised of the possibility of such damages. In no event will McGraw-Hill's liability for any damages to you or any other person ever exceed the lower of suggested list price or actual price paid for the license to use the software, regardless of any form of the claim.

McGRAW-HILL SPECIFICALLY DISCLAIMS ALL OTHER WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Specifically, McGraw-Hill makes no representation or warranty that the software is fit for any particular purpose and any implied warranty of merchantability is limited to the sixty-day duration of the Limited Warranty covering the physical CD only (and not the software) and is otherwise expressly and specifically disclaimed.

This limited warranty gives you specific legal rights; you may have others which may vary from state to state. Some states do not allow the exclusion of incidental or consequential damages, or the limitation on how long an implied warranty lasts, so some of the above may not apply to you.