sistor, a 0.35 pF capacitor, and a 0.7 nH inductor.

(a) What is the admittance of the transistor output at 8, 10, and 12 GHz?
(b) How does the susceptance vary with frequency?
(c) What is the shunt reactive element required to resonate the output admittance of the transistor at 8, 10, and 12 GHz?
(d) What are the equivalent inductances required to resonate the output admittance of the transistor at 8, 10, and 12 GHz?
(e) How does the inductance calculated in (d) vary with frequency?
(f) Describe a two-element circuit that has the characteristic identified in (e). (Note that this circuit would only be able to achieve the required characteristic over a smaller bandwidth than that required for a match from 8 GHz to 12 GHz.)

8. The output of a transistor is modeled as the shunt connection of a current source, a $68\ \Omega$ resistor, a 0.35 pF capacitor, and a 0.7 nH inductor.
   (a) What is the output admittance of the transistor at 8, 10 and 12 GHz?
   (b) How does the admittance vary with frequency?
   (c) Design a lumped-element matching network with two elements to match the transistor output at 10 GHz to a $50\ \Omega$ source.
   (d) Calculate the input admittance of the matching network, looking from the transistor, at 8, 10, and 12 GHz.
   (e) What is the input admittance of an ideal matching network, looking from the transistor, at 8, 10, and 12 GHz? Plot the actual and ideal admittance loci on a Smith chart using markers at 8, 10, and 12 GHz and indicating the direction of increasing frequency with arrows.

9. Consider a transistor having the S parameters shown in Table 3-1 and Figure 3-6(a). Ignore feedback effects and consider that the reflection coefficient looking into the output of the transistor is S_{22}^*.
   (a) Draw and describe the two-port input matching network problem with Port 1 at the output of the transistor and a $50\ \Omega$ termination at Port 2.
   (b) What is the ideal S_{11} of the input matching two-port at 8 GHz?
   (c) What is the ideal S_{11} of the input matching two-port at 10 GHz?
   (d) What is the ideal S_{11} of the input matching two-port at 12 GHz?
   (e) Plot the locus from 8 GHz to 12 GHz of S_{11} of the input matching two-port on a Smith chart.
   (f) Assume that the locus plotted in (e) from 8 GHz to 12 GHz can be realized using a lumped-element network. Comment on the difficulty of the design and the design approach.

10. At 10 GHz a capacitor, C_1, has a reactance of $-50\ \Omega$.
    (a) What is the impedance of C_1 at 8, 10, and 12 GHz?
    (b) How does the impedance of C_1 vary with frequency?
    (c) What is the inductance required to resonate the capacitance at 8, 10, and 12 GHz?
    (d) How does the inductance calculated in (b) vary with frequency?

11. The input of a transistor is modeled as a $20\ \Omega$ resistor in series with a 0.3 pF capacitor.
    (a) What is the impedance of the transistor input at 8, 10, and 12 GHz?
    (b) How does the impedance vary with frequency?
    (c) What is the series inductance required to resonate out the transistor capacitance at 8, 10, and 12 GHz?
    (d) Comment on whether a wideband match of a resistive source to the input of a transistor can be achieved using a frequency-independent inductor.

12. The input of a transistor is modeled as a $20\ \Omega$ resistor in series with a 0.3 pF capacitor. The transistor is part of an amplifier operating in a $50\ \Omega$ system.
    (a) Design a lumped-element matching network with two elements (inductors and/or capacitors) to match the transistor input at 10 GHz to a $50\ \Omega$ source.
    (b) Calculate the return loss (looking into the matching network from the source) at 8, 9, 10, 11, and 12 GHz.
    (c) Calculate the fraction of the available input power, expressed in decibels, delivered to the transistor at 8, 9, 10, 11, and 12 GHz and indicate the direction of increasing frequency with arrows.
    (d) Comment on the variation in amplifier gain solely due to mismatch at the transistor input.

13. Consider the input of a transistor having the S parameters shown in Table 3-1 and Figure 3-6(a). Ignore feedback effects so that for the active device $\Gamma_{\text{in}} = S_{11}$. Also an input matching

network terminated in 50 Ω at Port 1 and the active device at Port 2.

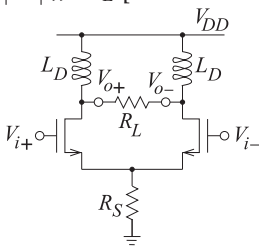
(a) What is the ideal 50 Ω S_{22} of the input matching network (i.e., seen from the transistor input) at 8 GHz?
(b) What is the ideal 50 Ω S_{22} of the input matching network (i.e., seen from the transistor input) at 10 GHz?
(c) What is the ideal 50 Ω S_{22} of the input matching network (i.e., seen from the transistor input) at 12 GHz?

14. Consider matching the input of a transistor having the S parameters shown in Table 3-1 and Figure 3-6(a). Ignore feedback effects and consider that the input reflection coefficient of the transistor $\Gamma_{\text{in}} = S_{11}$. Curve B in Figure 3-6(a) is the locus of the impedance looking into the matching network from the transistor. What two-element network has this locus? (One of the elements may be a resistor).

15. Consider synthesizing a two-port matching network terminated in a 50 Ω load and with an input reflection coefficient Γ_1 shown as Curve B in Figure 3-6(a). Draw and describe the two-port matching network problem.

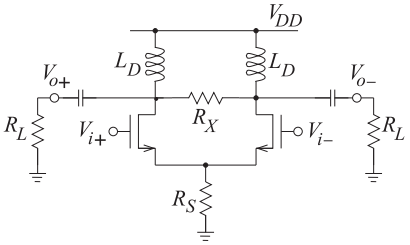
16. Consider synthesizing a two-port matching network terminated in a 50 Ω load and with an input reflection coefficient Γ_1 shown as Curve B in Figure 3-6(a). Can a broadband match be obtained using a two-element matching network? Explain your answer in terms of rotations on a Smith chart.

17. Consider the inductively biased differential Class A amplifier shown below. L_D is a choke inductor so $|sL| \gg R_L$.[Parallels Example 3.1.]



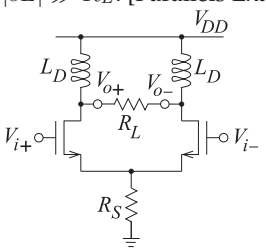
What is the CMRR when R_S = 20 kΩ, R_L = 10 kΩ, the transistor transconductance, g_m = 50 mS, and the drain-source resistance, r_d is 100 kΩ?

18. Consider the inductively biased differential Class A amplifier shown below. The capacitors can be treated as RF short circuits. L_D is a choke inductor so $|sL| \gg R_L$. [Parallels Example 3.1]



(a) Derive a symbolic expression for the CMRR of the amplifier assuming that the drain-source resistance of the transistors, r_0 or r_d, is much greater than both R_L and R_X, and so can be ignored.
(b) What is the CMRR when R_S = 10 kΩ, R_X = 30 kΩ, R_L = 10 kΩ, and the transistor transconductance, g_m is 10 mS.

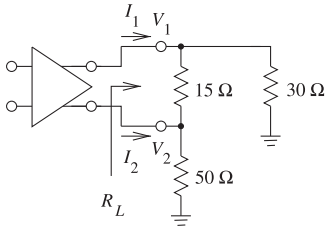
19. Consider the inductively-biased differential Class A amplifier shown below. L_D is a choke inductor so $|sL| \gg R_L$. [Parallels Example 3.1]



(a) Derive a symbolic expression for the differential mode gain of the amplifier.
(b) Derive a symbolic expression for the CMRR of the amplifier.
(c) What is CMRR when R_S = 10 kΩ, R_L = 10 kΩ, the transistor transconductance, g_m is 15 mS, and the transistors' drain-source resistance, r_d, is 100 kΩ?

20. A differential amplifier has a differential-mode gain of 20 dB and a common-mode gain of −3 dB.
(a) What is the the odd-mode gain?
(b) What is the the even-mode gain?

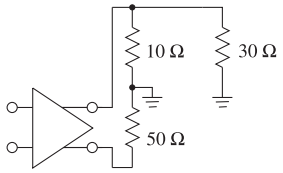
21. Consider the differential amplifier below. [Parallels Example 3.0]



(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?

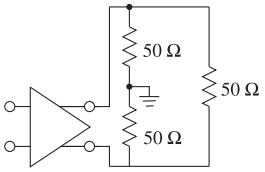
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?

22. Consider the differential amplifier below. [Parallels Example 3.0]



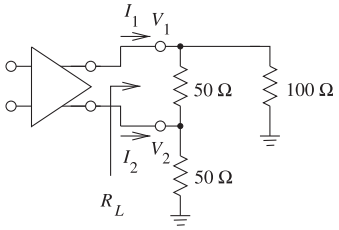
(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?

23. Consider the differential amplifier below. [Parallels Example 3.0]



(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?
(e) If the differential-mode gain of the amplifier is 20 dB and the common-mode gain is 2 dB, what is the odd-mode gain?

24. Consider the differential amplifier below. L_D is a choke inductor so $|sL| \gg R_L$. [Parallels Example 3.0]



(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?

25. A pseudo-differential amplifier is shown in Figure 3-28. Distributed biasing of this amplifier (replacing the inductors and R_{DD}), presents a common-mode impedance of $5\ \Omega$ and a differential-mode impedance of $1\ \text{k}\Omega$ to the drain terminals of the transistors in the middle of the amplifier band. The transconductance of each transistor is $g_m = 1$ S, and the internal parasitics of the transistors can be ignored.

(a) Draw the common-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(b) Draw the odd-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(c) Draw the even-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(d) Draw the differential-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(e) What is the common-mode gain?
(f) What is the differential-mode gain?
(g) What is the common-mode rejection ratio in decibels?

26. A pseudo-differential amplifier is shown in Figure 3-28. Distributed biasing of this amplifier (replacing the inductors and R_{DD}), presents a common-mode impedance of $5\ \Omega$ and an odd-mode impedance of $1\ \text{k}\Omega$ to the drain terminals of the transistors in the middle of the amplifier band. The transconductance of each transistor is $g_m = 100$ mS and the internal parasitics of the transistors can be ignored.

(a) Draw the common-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(b) Draw the odd-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(c) Draw the even-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(d) Draw the differential-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(e) What is the even-mode impedance presented to the amplifier?
(f) What is the differential-mode impedance presented to the amplifier?
(g) What is the even-mode voltage gain?
(h) What is the differential-mode voltage gain?
(i) What is the common-mode voltage gain?
(j) What is the odd-mode voltage gain?
(k) What is the common-mode rejection ratio?

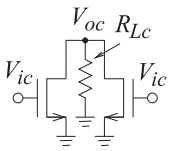
3.11.1 Exercises By Section

†challenging, ‡very challenging

§3.2 $1^{\dagger}, 2, 3$
§3.5 $4^{\ddagger}, 5^{\ddagger}, 6^{\dagger}, 7^{\dagger}, 8^{\dagger}, 9^{\dagger}$
§3.6 $10, 11, 12, 13, 14, 15, 16^{\dagger}, 17^{\dagger}, 18^{\dagger}, 19^{\dagger}, 20, 21, 22, 23^{\dagger}, 24^{\dagger}$
§3.7 $25^{\dagger}, 26^{\dagger}$

3.11.2 Answers to Selected Exercises

1(d) $34.7\ \Omega$
5(b) 121
10(d) $1/f^2$
12(c) 8 GHz, -2.77 dB
9 GHz, -0.57 dB
10 GHz, 0 dB
11 GHz, -0.33 dB
12 GHz, -1.03 dB
18(h) -100
20(b) -3 dB
21 $37.5\ \Omega$
25(a)



CHAPTER 4

Power Amplifiers

4.1 Introduction

This chapter considers amplifiers that are designed to produce large RF powers and achieve high efficiency. The design of such amplifiers is more complicated than the design of small signal amplifiers, which can be designed using linear techniques. It is now necessary to use nonlinear simulation tools and laboratory optimization to develop designs that achieve high efficiency with acceptable distortion [1]. The design of power amplifiers involves the trade-off of distortion and amplifier efficiency. Before beginning this chapter the reader should be familiar with the metrics for nonlinear distortion described in Section 4.5 of [2].

Transistors used in microwave amplifiers are large, as can be seen in Figure 4-1. This transistor is targeted for use in WCDMA amplifiers operating between 2.11 and 2.17 GHz with an output power of 140 W at 1 dB gain compression. The transistor includes input and output matching networks in the package. This provision by the transistor vendor reduces the complexity of amplifier design.

4.2 Simulation of Nonlinear Microwave Circuits

The circuit simulation tools used to model RF circuits are linear circuit simulators, transient circuit simulators (e.g., Spice), and nonlinear steady-state simulators. Transient circuit simulators can model large signals in RF circuits, but those available to RF designers cannot be used to model circuits requiring high dynamic range simulation or to model circuits with sharp frequency responses such as circuits with high-order filters. These simulators can be very slow, or perhaps impossibly slow, in modeling circuits with narrowband signals such as a digitally modulated carrier. When it is important to capture nonlinear and frequency-response effects precisely, nonlinear steady-state simulators are preferred. The two main types of nonlinear steady-state simulators available to designers are harmonic balance (HB) simulators [3, 4] and Spice-like transient simulators modified to efficiently find the response of a circuit to periodic excitation [5] (so-called periodic steady-state (PSS) analysis).

Nonlinear steady-state simulators exploit the narrowband nature of most radio systems and the circuit waveforms that are essentially steady-state, although not necessarily periodic. Such waveforms are called quasi-periodic waveforms. As an example of the waveforms to be determined, consider the

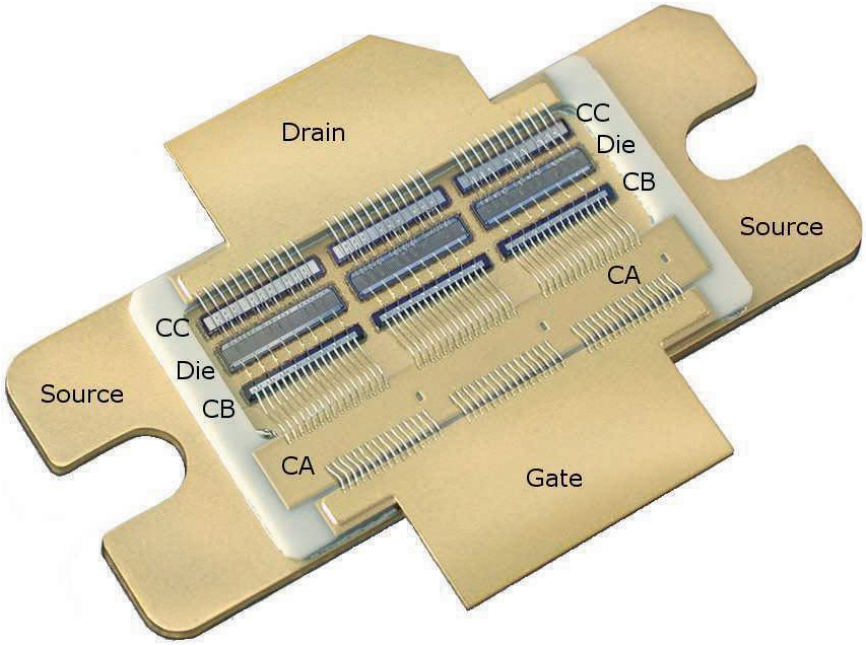


Figure 4-1: A 2.1 GHz silicon LDMOS transistor. The gate and drain tabs are 12.5 mm across. The three dies operate in parallel. The input matching network comprises a series inductance provided by bond wires, a shunt capacitor (Capacitor A, C_A), another series inductance from bond wires, and another shunt capacitor (C_B). The network is then connected to each gate finger of the transistor dies using short bond wires. The output matching network consists of a shunt capacitor (C_C) and series inductance. There are 189 bond wires. Used courtesy of Freescale Semiconductor Inc. Also see [1].

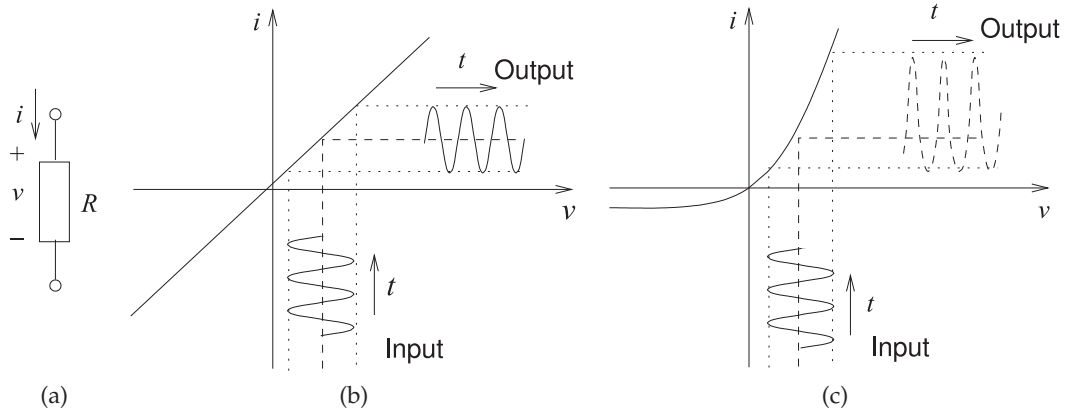


Figure 4-2: Current responses of a resistor: (a) resistor with passive convention defining voltage and current; (b) i-v characteristic of a linear resistor; and (c) i-v characteristic of a diode (a nonlinear resistor).

responses, shown in Figure 4-2, of linear and nonlinear resistors to an applied sinusoidal voltage.

Figure 4-2(b) shows the i-v characteristic of a linear resistor. With an applied sinusoidal voltage, the output current waveform of the linear resistor is also a sinusoid. If the applied voltage signal is a sum of sinusoids, then the output current is also a sum of sinusoids. The component of output current at each frequency only depends on the applied voltage component at that frequency. With a nonlinear resistor having the i-v characteristic in Figure 4-2(c), a large applied sinusoidal signal results in an output that is distorted and is a steady-state signal that has harmonics of the original signal. If the applied signal is the sum of two sinusoids of frequencies f_1 and f_2, then the output will be a steady-state signal with components having frequencies $mf_1 + nf_2$, where m and n are integers. The key feature here is that if a steady-state signal, a sum of sinusoids, is applied to a nonlinear circuit, then the output will also be a steady-state signal, a sum of sinusoids, but now each frequency component at the output is affected by every frequency component of the input signal. However, the simulation problem simplifies to finding the amplitudes and phases of the sinusoidal components rather than trying to determine the output waveform at a very large number of time points as done in a transient simulator. This is the basis of nonlinear steady-state simulation. In a narrowband radio, signals are very close to being sinusoidal with very slowly varying amplitude and phase [6].

4.2.1 Harmonic Balance Analysis of RF Circuits

With the harmonic balance method, the steady-state response of a nonlinear circuit is assumed to be a sum of sinusoids [3, 4]. This assumed form of the solution then allows simplification of the circuit equations, and simulation is used to determine the unknown coefficients: the magnitudes and phases of the sinusoids.

A harmonic-balance simulator can be many orders of magnitude more

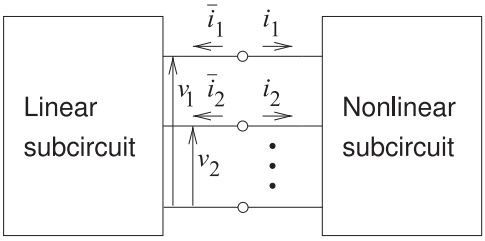


Figure 4-3: Analysis of a nonlinear circuit by the harmonic balance method partitions the circuit into linear and nonlinear subcircuits.

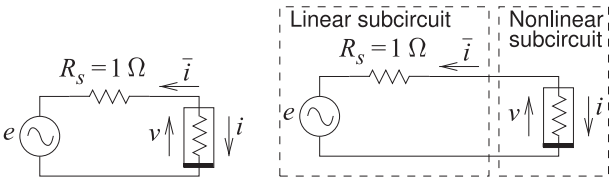


Figure 4-4: Example harmonic balance circuit: (a) circuit with nonlinear resistor; and (b) partitioned circuit. $e(t) = E\cos(\omega y)$.

efficient than a time-domain simulator, and lends itself well to analysis of narrowband circuits and to optimization. Another major advantage of the harmonic balance method is that linear circuits can be of practically any size, with no significant increase in overall simulation time.

4.2.2 Example: Harmonic Balance Analysis of a Simple Circuit

Internally a harmonic balance simulator partitions a circuit into two subcircuits, as shown in Figure 4-3. During analysis, the set of frequencies being considered is fixed and the linear block need only be calculated initially, and its admittance parameters stored and reused at every stage of the analysis. Newton's method (or similar iterative derivative-based minimization method) is used with harmonic balance to solve for the state of the circuit, that is, the amplitudes and phases of the voltage phasors at the interface.

As an example, consider the circuit in Figure 4-4(a) where the nonlinear resistor is described by

$$i(t) = v(t) + [v(t)]^2 . \tag{4.1}$$

The first step in analysis partitions the circuit into linear and nonlinear subcircuits, as shown in Figure 4-4(b). Harmonic balance simulation minimizes the frequency domain Kirchoff's current law error at the linear circuit-nonlinear circuit interface. Next the number of sinusoids (or tones) to be considered must be chosen. The choice in this example is to consider only the DC, fundamental at radian frequency ω, and second-harmonic tones. Then the voltage at the interface is

$$v(t) = V_0 + V_1\cos(\omega t) + V_2\cos(2\omega t). \tag{4.2}$$

Phase has been dropped, as this is a resistive circuit and all currents and voltages will have the same phase, nominally zero. Thus the unknowns are the amplitudes V_0, V_1, and V_2. With values of V_0, V_1, and V_2 assumed (and updated through iteration), the current flowing into the linear subcircuit can be calculated using the nodal admittance matrix of the linear subcircuit

yielding

$$\bar{i}(t) = \bar{I}_0 + \bar{I}_1 \cos(\omega t) + \bar{I}_2 \cos(2\omega t). \tag{4.3}$$

Similarly the nonlinear model of the element in the nonlinear subcircuit can be used to calculate the nonlinear currents:

$$i(t) = I_0 + I_1 \cos(\omega t) + I_2 \cos(2\omega t). \tag{4.4}$$

The linear subcircuit, shown in Figure 4-4(a), and the nonlinear subcircuit described by the quadratic model in Equation (4.1) result in the following circuit equations:

$$\begin{array}{lll} I_0 = V_0 + \frac{1}{2}V_2^2, & I_1 = V_1, & I_2 = \frac{1}{2}V_1^2 \\ \bar{I}_0 = V_0, & \bar{I}_1 = V_1 - E, & \text{and } \bar{I}_2 = V_2. \end{array} \tag{4.5}$$

Since only the DC, fundamental, and second-harmonic components are being considered, the KCL error, F, is

$$F = |f_0| + |f_1| + |f_2| \tag{4.6}$$

where the Kirchoff current error at DC, the fundamental, and the second harmonic are

$$f_0 = I_0 + \bar{I}_0, \quad f_1 = I_1 + \bar{I}_1, \quad \text{and} \quad f_2 = I_2 + \bar{I}_2, \tag{4.7}$$

respectively. Thus F is minimized, or alternatively the zeros of each sub-error, the f_ns, are found. The zeros can be found using a Newton–Raphson iterative technique to determine the voltages (V_0, V_1, and V_2) that yield the zeros of f_0, f_1, and f_2. Thus the $(i+1)$th analysis iteration is

$$^{i+1}\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} = {}^{i}\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} - \left[\mathbf{J}\left({}^{i}\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix}\right)\right]^{-1} \times \begin{bmatrix} f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}) \\ f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}) \\ f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}) \end{bmatrix}. \tag{4.8}$$

The Jacobian, $\mathbf{J}$, is a matrix of derivatives of the f_ns with respect to the V_ns. Thus

$$\begin{aligned} &\mathbf{J}\left({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}\right) \\ &= \begin{bmatrix} \dfrac{\partial f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} & \dfrac{\partial f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} & \dfrac{\partial f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} \\ \dfrac{\partial f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} & \dfrac{\partial f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} & \dfrac{\partial f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} \\ \dfrac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} & \dfrac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} & \dfrac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} \end{bmatrix}. \end{aligned} \tag{4.9}$$

Now each element of the Jacobian is affected by both the linear and nonlinear subcircuits, so, for example,

$$\frac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = \frac{\partial I_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} + \frac{\partial \bar{I}_2({}^{i}V_2)}{\partial V_1}. \tag{4.10}$$

Since the linear current $\bar{I}_2$ is dependent only on V_2 (see Equation (4.5),

$$\frac{\partial \bar{I}_2({}^iV_2)}{\partial V_1} = 0, \tag{4.11}$$

and following trigonometric expansion of Equation (4.1),

$$\frac{\partial I_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = \frac{\partial}{\partial V_1}\left(\tfrac{1}{2}{}^iV_1^2\right) = {}^iV_1. \tag{4.12}$$

Similarly

$$\left.\begin{array}{ll}
\dfrac{\partial f_0({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} = 1 + 2{}^iV_0 + 1 & \dfrac{\partial f_0({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = {}^iV_1 \\
\dfrac{\partial f_0({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} = {}^iV_2 & \dfrac{\partial f_1({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} = 2{}^iV_1 \\
\dfrac{\partial f_1({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = 1 + 2{}^iV_0 + {}^iV_2 + 1 & \dfrac{\partial f_1({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} = {}^iV_1 \\
\dfrac{\partial f_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} = 2{}^iV_2 & \dfrac{\partial f_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = {}^iV_1 \\
\dfrac{\partial f_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} = 2{}^iV_0 + 2 &
\end{array}\right\}. \tag{4.13}$$

Thus the equations to be solved by the harmonic balance simulator are

$$\left.\begin{array}{lll}
{}^iI_0 = {}^iV_0 + {}^iV_0^2 + V_1^2/2 + V_2^2/2 & {}^i\bar{I}_0 = {}^iV_0 & f_0 = {}^iI_0 + {}^i\bar{I}_0 \\
{}^iI_1 = {}^iV_1 + 2{}^iV_0\,{}^iV_1 + {}^iV_1\,{}^iV_2 & {}^i\bar{I}_1 = {}^iV_1 - E_1 & f_1 = {}^iI_1 + {}^i\bar{I}_1 \\
{}^iI_2 = {}^iV_2 + 2{}^iV_0\,{}^iV_2 + {}^iV_1^2/2 & {}^i\bar{I}_2 = {}^iV_2 & f_2 = {}^iI_2 + {}^i\bar{I}_2
\end{array}\right\}. \tag{4.14}$$

This is solved iteratively using the Newton–Raphson algorithm described by Equation (4.8) to provide an updated estimate of the voltages (${}^{(i+1)}V_0$, ${}^{(i+1)}V_1$, and ${}^{(i+1)}V_2$). The output of a program implementing this algorithm is shown in Table 4-1. Note how quickly the iterations arrive at the steady-state solution. This quick convergence to a steady-state solution is typical of harmonic balance simulation of nonlinear RF circuits.

The harmonic balance approach to nonlinear circuit analysis can be extended to consider excitation by sums of nonharmonically related sinusoids, which, for example, enables distortion to be determined using a two-tone test.

Many nonlinear RF circuits can be solved with just a few harmonics considered and only a few iterations are required to obtain convergence. The harmonic balance analysis is many orders of magnitude faster than simulation using transient circuit simulation and the dynamic range of the simulation can exceed 150 dB while a transient simulator is often limited to dynamic ranges of 80 dB and often much less is achieved. Consider a cell phone that can have transmit signals of up to 30 dBm, yet receive signals as small as -100 dBm. Circuit simulation of such a system requires that the simulator have a dynamic range 20 dB greater than the 130 dB dynamic range of the cellphone signals. This is easily met by harmonic balance simulators. However, there are limitations to the use of the harmonic balance method.

ITERATION 0					
$\bar{I}_0$	= 0	$\bar{I}_1$	= 0	$\bar{I}_2$	= 0
V_0	= 0	V_1	= 0.5	V_2	= 0
I_0	= 0.5	I_1	= 1	I_2	= 0.5
ITERATION 1					
$\bar{I}_0$	= 0	$\bar{I}_1$	= −0.5	$\bar{I}_2$	= 0
V_0	= −0.0769231	V_1	= 0.557692	V_2	= −0.0769231
I_0	= 0.125	I_1	= 0.5	I_2	= 0.125
ITERATION 2					
$\bar{I}_0$	= −0.0769231	$\bar{I}_1$	= −0.442308	$\bar{I}_2$	= −0.0769231
V_0	= −0.0892028	V_1	= 0.57747	V_2	= −0.0912325
I_0	= 0.087463	I_1	= 0.428994	I_2	= 0.0904216
ITERATION 3					
$\bar{I}_0$	= −0.0892028	$\bar{I}_1$	= −0.42253	$\bar{I}_2$	= −0.0912325
V_0	= −0.089835	V_1	= 0.578574	V_2	= −0.0919462
I_0	= 0.0896515	I_1	= 0.421762	I_2	= 0.0917795
ITERATION 4					
$\bar{I}_0$	= −0.089835	$\bar{I}_1$	= −0.421426	$\bar{I}_2$	= −0.0919462
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898363	I_1	= 0.421424	I_2	= 0.0919477
ITERATION 5					
$\bar{I}_0$	= −0.0898368	$\bar{I}_1$	= −0.421423	$\bar{I}_2$	= −0.0919482
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898368	I_1	= 0.421423	I_2	= 0.0919482
ITERATION 6					
$\bar{I}_0$	= −0.0898368	$\bar{I}_1$	= −0.421423	$\bar{I}_2$	= −0.0919482
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898368	I_1	= 0.421423	I_2	= 0.0919482
ITERATION 7					
$\bar{I}_0$	= −0.0898368	$\bar{I}_1$	= −0.421423	$\bar{I}_2$	= −0.0919482
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898368	I_1	= 0.421423	I_2	= 0.0919482

stop

Table 4-1: Circuit quantities at each iteration in the harmonic balance analysis of the circuit in Figure 4-4(a).

A major drawback is that harmonic balance does not work well with circuits containing large numbers of transistors, say a few tens of transistors. This is largely because the Jacobian becomes very large and analysis becomes unwieldy.

4.2.3 User's Guide to Using Harmonic Balance Analysis

Three major factors limit the accuracy of harmonic balance circuit simulation:

(i) The number of tones included in the analysis. If the number of tones is too small, there will be truncation error. Truncation error arises because, theoretically, an infinite number of harmonics can be generated by the interaction of large signals with even simple nonlinear circuits. The error can be reduced, of course, by specifying additional frequency components.

(ii) The aliasing errors due to a finite transform spectrum. This error can be reduced by considering many tones. The aliasing error is a numerically introduced error. This sets an upper limit on resolution.

(iii) The final value of the harmonic balance error. The major limiting factor here is how closely the Jacobian describes the actual error function.

Both the error function and the Jacobian have truncation error so ideally the Jacobian evaluation reflects the same truncation errors as the error function evaluation. In the end this comes down to the accuracy of the models. That is, whether the derivatives calculated in the model reflect the actual nonlinear relationship.

Naturally errors also arise due to the quality of the models of the active devices and of the linear circuit elements. Also, as the number of tones included in a harmonic balance analysis increases, the simulation time rapidly increases.

4.2.4 Periodic Steady-State Simulation of RF Circuits

Periodic steady-state (PSS) analysis is used with a Spice simulator to establish the response of a circuit to a periodic excitation signal [5]. Typically there would be one large sinusoid, such as a local oscillator or a carrier signal. The idea is that the dynamic state of the circuit is established by a transient simulation with a single sinusoidal excitation. Then a linear (or perhaps quadratic) model of the dynamic circuit is used with smaller signals. If the excitation signal is large, then the circuit is a time-varying linear circuit as far as smaller signals are concerned.

The PSS technique uses what is called the shooting method in which the simulator guesses the initial values of voltages at all terminals, charges on all capacitors, and currents through all inductors (i.e., the state-variables of the circuit) [7–13]. Then the circuit is simulated using transient analysis for one period of the exciting waveform. The state variables after one period are compared to the assumed state variables at the beginning of the period. If there is a difference, then the initial guess is changed and the process repeated. Convergence is usually achieved after a few iterations. Then the Fourier components of the state variables are found and a time-varying circuit calculated. From this, a model akin to a conversion matrix describing the dynamic circuit is established. There are many similarities to harmonic balance, as the frequencies of all signals in the circuit must be specified by the user. An advantage of the PSS technique is that a conventional Spice simulator, and the all important transistor models, can be used.

4.3 Switching Amplifiers, Classes D, E, and F

Switching amplifiers are the most efficient RF amplifiers, but are also the hardest to design. With linear amplifiers, such as Class A, B, AB, and C amplifiers, there is appreciable simultaneous voltage across the transistor and current flowing through it. Thus power is dissipated in the transistor in such an amplifier. The DC and AC **loadlines** at the transistor output of a linear amplifier essentially coincide, as is seen in the transistor output characteristic shown in Figure 4-5(a). The DC loadline is a straight line and the AC loadline closely follows the linear DC loadline; this is where the linear in linear amplifier comes from. Even when reactive effects cause the AC loadline to loop (and a small loop is seen in Figure 4-5(a)), the term linear amplifier is still used.

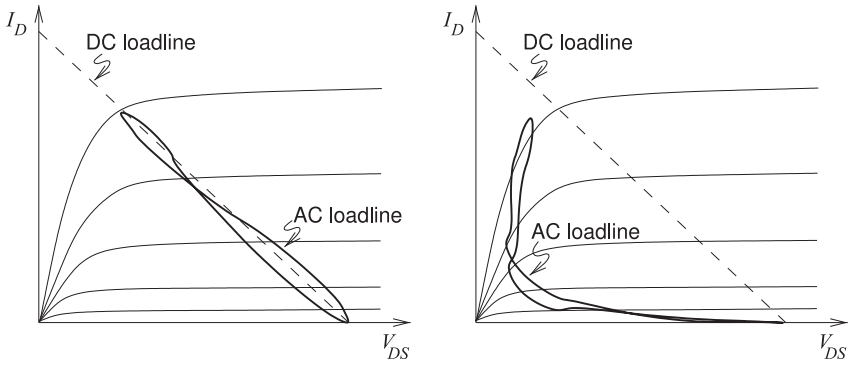
4.3.1 Dynamic Waveforms

In a switching amplifier there is little voltage across the output of the transistor when there is current flowing through it, and little voltage when there is current. This is seen in the **AC loadline**, also called **dynamic loadline**, of a switching amplifier, shown in Figure 4-5(b). This loadline is obtained through careful attention to the loading of the transistor at the harmonics.

The dynamic loadline of a switching amplifier is obtained by presenting the appropriate harmonic impedances to the transistor output. The particular scheme of harmonic termination (e.g., short or open circuits at the even and odd harmonics) leads to the designation of a switching amplifier as Classes D, E, F, etc. The key characteristic of all switching amplifiers is that when there is current through the transistor, there is negligible voltage across the output [14–20]. Also, when there is voltage across the transistor, there is little current through it (see Figure 4-5(b)).

Switching amplifiers are a conceptual departure from Class A, AB, B, and C linear amplifiers. The transistor output waveforms of a switching amplifier and those of Class A, AB, B, and C amplifiers are shown in Figure 4-6. Again it is seen, in Figure 4-6, that for a switching amplifier the voltages and current waveforms are shifted and voltage across the transistor and current through it do not occur simultaneously. The power dissipated by the transistor is the average of the product of the current through it and the voltage across the output. Thus the ideal switching amplifier consumes very little DC power, transferring nearly all of the DC power to the output RF signal. Bandpass filtering of the output of the amplifier results in a final RF output with little distortion. Switching amplifiers are the preferred amplifier in both handsets and basestations of cellular systems.

The theoretical maximum power-added efficiencies achieved by the various amplifier classes with a sinusoidal input signal are given in Table 4-2. With modulated signals, the maximum efficiencies cannot be achieved, as typically the average input power of the amplifier must be backed off by the peak-to-mean envelope power ratio (PMEPR) of the signal so that the peak carrier portion of the signal has limited distortion. Generally the acceptable distortion of the peak signal occurs at the 1 dB compression point of the amplifier. This is only an approximate guide, but useful. The PMEPRs of several digitally modulated signals are given in Table 4-3, together with their impact on efficiency. If there are two carriers, then the PMEPR of the



(a) Class A, AB, B, C linear amplifiers (b) Switching amplifier

Figure 4-5: DC and RF loadlines. An AC loadline is also called a dynamic loadline.

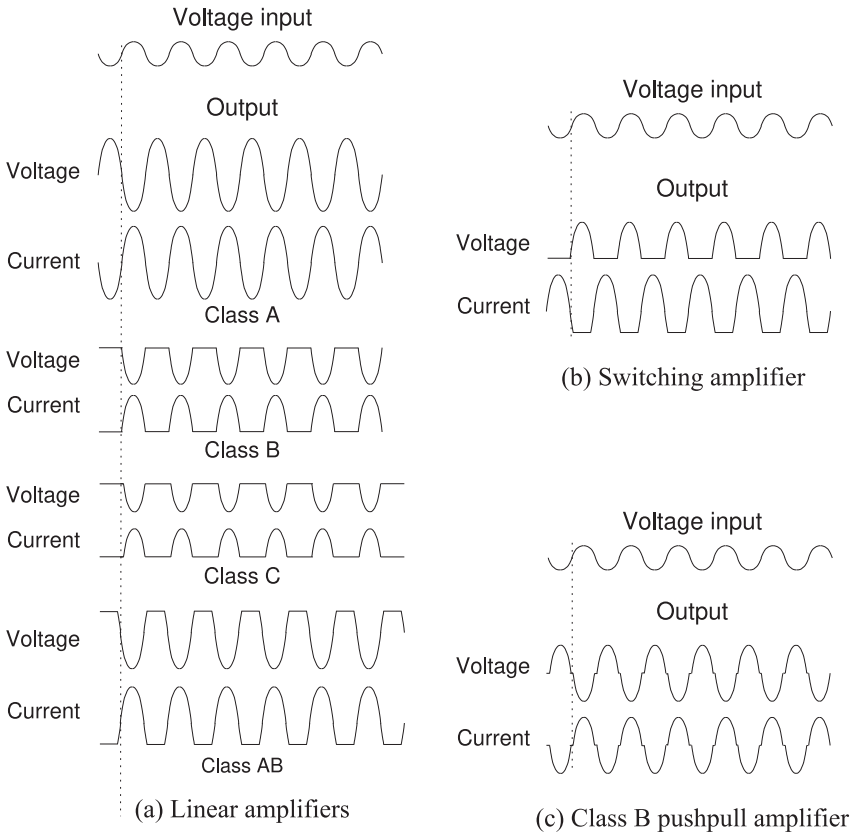


Figure 4-6: Input and output waveforms of a transistor for various classes of amplifier. The output voltage is measured across the transistor(s) and the current is through the transistor(s).

Table 4-2: Theoretical maximum efficiencies of amplifier classes with a sinusoidal excitation.

Amplifier class	Maximum efficiency
Class A (resistive bias)	25%
Class A (inductive bias)	50%
Class B	78.53%
Class C	100%
Class D	100%, but typically 75%
Class E	96%
Class F	88.36%

combined signal will be higher, requiring even greater amplifier backoff [21]. In practice, the efficiencies achieved will differ from these theoretical values. This is because the PMEPR does not fully capture the statistical nature of signals and because of coding and other technologies that can be used to reduce the PMEPR of a digital modulation scheme.

4.3.2 Conduction Angle

The conduction angle indicates the proportion of time an amplifying device, typically a transistor, is conducting current with a conduction angle of $360°$ indicating that the amplifying device is always on. The view is that a sinusoidal signal is being amplified which is appropriate since the majority of communication schemes produce an RF signal that looks like a sinewave with a very slowly-varying amplitude and phase. Class A amplifiers conduct current throughout the RF cycle and so have a conduction angle of $360°$. A Class B amplifier is biased so that only half of a sinewave is produced at the

Table 4-3: Efficiency reductions due to modulation type for a single modulated carrier. The Class A amplifier uses inductive drain biasing. The PMEPR increase is for multiple carriers. E.g. for GMSK PMEPR = 3.01 dB, 6.02 dB, 9.01 dB, 11.40 dB, 14.26 dB, and 17.39 dB for 1, 2, 4, 8, 16, and 32 carriers respectively. The PMEPR does not increase by 3 dB every time the number of carriers is doubled because statistically the envelope peaks of the individual carriers are less likely to align for more carriers.

Signal	PMEPR (dB)	Efficiency reduction factor	Class A (L bias) PAE	Class E PAE
FSK (ideal)	0	1.0	50%	96%
GMSK	3.0	0.501	25.1%	48.1%
QPSK	3.6	0.437	21.9%	42%
$\pi/4$DQPSK	3.0	0.501	25.1%	48.1%
OQPSK	3.3	0.467	23.4%	44.8%
8-PSK	3.3	0.467	23.4%	44.8%
64-QAM	7.8	0.166	8.3%	15.9%

output, so a Class B amplifier has a conduction angle is $180°$. A Class AB amplifier has a conduction angle between $180°$ and $360°$ with the bias, or one could say conduction angle, adjusted so that the distortion produced is acceptable. A Class C amplifier produces an output for less than half of the input sinewave and so it has a conduction angle of less than $180°$.

Switching amplifiers have high efficiency by ensuring that there is little current flowing through a transistor when there is voltage across it. So a switching amplifier is ideally either fully on or fully off. Ideally current flows for half the time and so the conduction angle is $180°$. However, the transistor must transition between these regions and the conduction angle indicates the degree of overlap. To minimize overlap, the conduction angle of an actual switching amplifier is less than $180°$.

4.3.3 Class D

The Class D amplifier was the first type of switching amplifier developed. The main concept of the Class D amplifier is using the transistor as a switch so that there is negligible current flowing through the transistor when there is voltage across it. The audio form of the Class D amplifier is shown in Figure 4-7(a) [14]. The two transistor inputs have the same RF signal applied but are level-shifted (but the circuit to do this has been omitted as well as the appropriate bias circuitry). Each transistor approximates Class C operation so only one transistor is switched on at a time. The current and voltage waveforms are shown in Figure 4-7(c). The transistors drive a resonant circuit with L_1 and C_1 acting as a bandpass filter. The filter reduces the distortion of the voltage and current waveforms presented to the output load, but results in the amplifier being narrowband. This Class D amplifier works very well at frequencies up to a few megahertz. Above that the transistors, having opposite polarity, are not well matched. At RF a more appropriate Class D amplifier is as shown in Figure 4-7(b) [14, 15, 19, 22] where, in this case, two nMOS transistors are used as they have higher mobility than pMOS transistors. Parasitic reactances result in substantial overlap of the current and voltage transition regions and hence there is loss of RF power in the

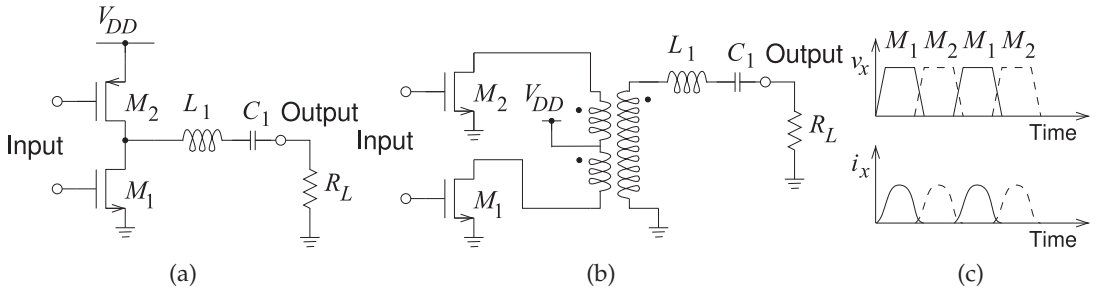


Figure 4-7: Class D amplifier: (a) low-frequency form; (b) microwave form; and (c) current and voltage waveforms (where v_x is the drain-source voltage and i_x is the drain current) indicating which transistor is turned on.

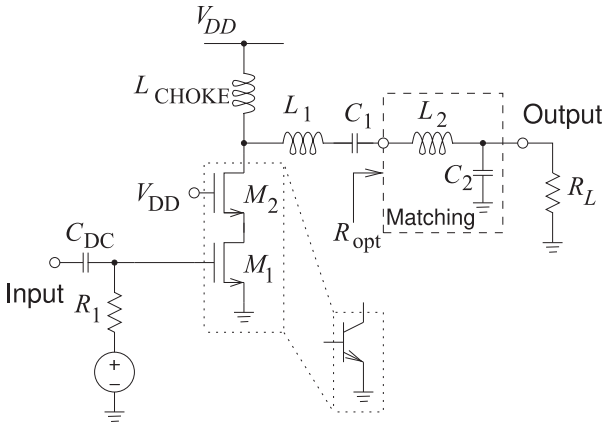


Figure 4-8: Class E amplifier.

transistors. This loss can be reduced using an alternative form of the Class D amplifier, called a current-mode Class D amplifier, which switches current rather than voltage [23]. Efficiencies of around 75% are achieved.

4.3.4 Class E

The Class E [15, 17] amplifier builds on the Class D amplifier concept of using a transistor as a switch rather than as a linear amplifier. An RF Class E amplifier is shown in Figure 4-8. The circuit shown uses two MOSFETS with M_1 being the switching transistor and M_2 acting in part as an active load. The main function of M_2 is to translate the drain current of M_1 into a voltage at the output. Bias is provided through L_{CHOKE}, which presents a high RF impedance. L_1 and C_1 provide a bandpass filtering function, while L_2 and C_2 provide matching to the load so that the impedance looking into the L_2, C_2 and R_L network is an optimum resistance, R_{opt}. The circuit is designed so that L_{CHOKE}, the parasitic output capacitance of the transistors, L_1, C_1, and R_{opt} form a damped oscillating circuit. The two MOSFETs can also be replaced by a single transistor, typically an HBT transistor, as shown by the insert in Figure 4-8, and sometimes a capacitor is added from the top of the transistor to ground if the parasitic transistor capacitance is insufficient. So at the output of the transistors there is a parallel LC circuit to ground

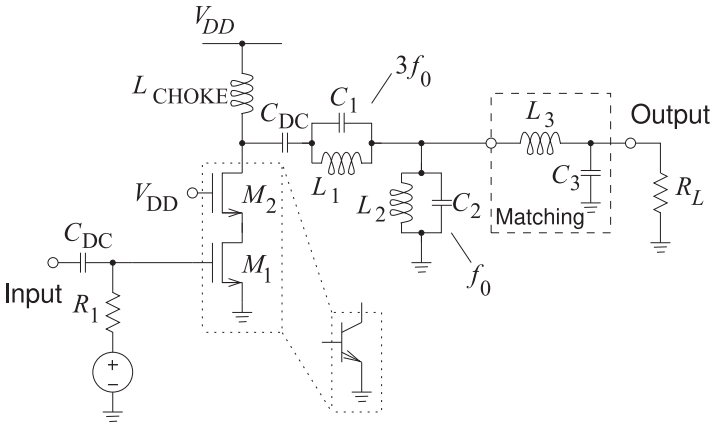


Figure 4-9: Class F amplifier.

(L_{CHOKE} and the parasitic capacitance of M_2) and a series LC circuit (L_1 and C_1). When the transistors are switched off (and have a high voltage across them), the series LC circuit provides current to the parallel LC circuit as well as to R_{opt}, instead of current being drawn through the transistors. When the transistors are switched on (and have little voltage across them), the oscillation proceeds in the opposite direction and current is delivered to R_{opt} through the transistors. This is the oscillation mechanism with the resistance R_{opt} providing damping. Design matches the natural oscillation frequency with the frequency of the RF signal.

4.3.5 Class F

In the Class E amplifier the voltage at the drain/collector of the transistor is approximately a square wave and the transistor current approximates a half sinusoid. The Class F amplifier takes this one step further and realizes an approximate square current wave through the transistors as well as an out-of-phase square voltage wave [14, 15, 19]. This is achieved using harmonic resonance.

A Class F amplifier is shown in Figure 4-9, where the narrowband RF signal has a center frequency f_0. Again M_1 operates as a switch producing a square voltage wave at the output of the transistors. The parallel L_1C_1 circuit is tuned to the third harmonic of the RF signal (i.e., $3f_0$), and this parallel circuit provides third-harmonic current when the transistors are turned off. The L_1C_1 circuit is an approximate short circuit at f_0, and the parallel L_2C_2 circuit, tuned to be resonant at f_0, presents an open circuit at f_0 and short circuits at the harmonics. The net result is that the current waveform passing through the transistors is a reasonable square wave with first and third harmonics and no second harmonic (note that a square wave consists of only odd frequency components). Also note that third-harmonic current does not pass through the load. This concept could be continued to provide similar behavior at the fifth harmonic as well, but it then becomes increasingly difficult to adjust the design to work as intended. With both the voltage and current waves being square, the lower the overlap, and the lower the power dissipated in the transistors.

4.3.6 Inverted Amplifiers

The Class D and F switching amplifiers described above are designed to switch the voltage between two states. The dual of these are the **inverted Class D amplifier** [20, 23–25] and the **inverted Class F amplifier** [26–29]. The design intent is that the transistors in these amplifiers switch the current rather than the voltage. They are also called **current-mode amplifiers**.

4.3.7 Summary

The efficiency advantages of switching amplifiers are significant and often justifies the higher design cost. They are used in power amplifiers in most basestations and are starting to be used in cellphones. At high microwave and millimeter-wave frequencies stability concerns and the high cost of design means that many amplifiers will continue to be linear amplifiers for some time.

4.4 Distortion and Digitally Modulated Signals

Digitally-modulated signals do not have constant envelopes so that the short-term peak power of the RF signal is higher than the average RF power. Efficient amplifiers must introduce minimal distortion and be efficient for peak signal levels as well as for the average signal level. There is a trade-off between distortion and efficiency and various amplifier architectures have been developed to implement good trade-offs. These architectures will be introduced in a later section. This section introduces approaches for describing distortion.

4.4.1 PMEPR and Probability Density Function

Amplitude variations occur with most digital modulation schemes. For example, a QPSK signal consists of two digital data streams, equal in amplitude, modulated in quadrature onto a carrier signal. If the data streams are not filtered prior to modulation, then the resulting modulated carrier signal has a constant envelope with instantaneous transitions from one constellation point to another. However, the occupied bandwidth of this signal is quite large, as the spectrum is that of a pulse train, $\sin(x)/x$, the sinc function. The first **sidelobe** of the sinc spectrum is only 13 dB down from the carrier level and typically is in the middle of the adjacent channel. To reduce the bandwidth of the modulated signal, a low-pass filter is applied to each digital data stream to minimize the out-of-band spectrum of the modulated signal. This comes with a drawback: the filters cause a finite memory effect resulting in amplitude variations as the ringing energy from a previous data pulse adds to the current filtered data pulse. As well, the transitions from one constellation point to another are slow and the amplitude of the modulated carrier varies significantly during the transition.

Amplitude variations of the modulated signal are characterized by waveform statistics such as the PMEPR. A signal with a high PMEPR requires that the RF system have high linearity to handle both the average power requirements and the peak amplitude excursions without generating excessive out-of-band distortion. A simple design technique is to reduce the average output power of an amplifier to a level that is below the 1 dB

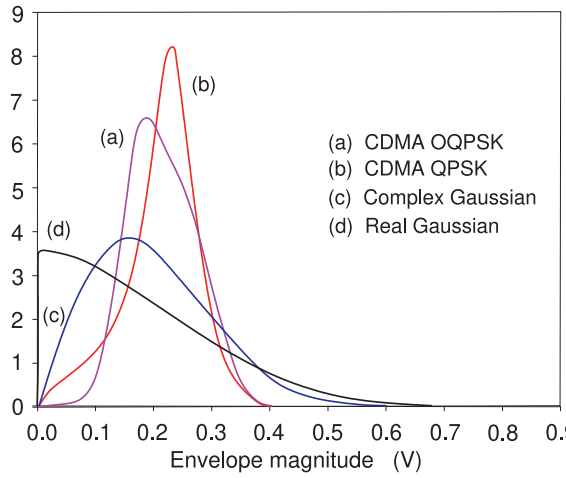


Figure 4-10: Amplitude PDF for CDMA and Gaussian modulated signals with the same power of 0 dBm. After [31].

Signal Modulation	PMEPR (dB)
CDMA OQPSK	5.4
CDMA QPSK	6.6
Real Gaussian	13.5
Complex Gaussian	11.8

Table 4-4: Peak-to-average ratios in decibels for OQPSK and QPSK used in CDMA compared to the PMEPRs of Gaussian signals.

gain compression power by the PMEPR, this is called **amplifier back-off**. However, it is possible for a signal with a high PMEPR to exhibit less nonlinear distortion than a signal with lower PMEPR [30]. The reason for this apparent inconsistency is because the signal peak may have a low probability of occurrence. Thus PMEPR by itself is an incomplete statistic for determining the linearity requirements of an amplifier.

The **amplitude probability density function** (APDF) is a more complete statistical description of the amplitude variations of a modulated signal. The APDF defines the maximum and minimum variation along with the relative probability of occurrence of amplitudes within the variation. The APDF is typically estimated from a histogram of envelope amplitudes, with a uniform bin size, by

$$f(A) = \frac{N}{\Delta A \times N_c}, \tag{4.15}$$

where N is the number of counts in the bin having amplitude A, ΔA is the bin width (i.e., the range of amplitudes in the bin), and N_c is the total number of samples. The shape of the amplitude density determines the sensitivity of a particular signal to spectral regrowth due to nonlinear gain compression or expansion. This is a more precise determinant of distortion than the single PMEPR metric.

As an example, consider Figure 4-10 which shows the APDF of a CDMA signal using OQPSK modulation, of the same signal but now using QPSK modulation of a real **Gaussian signal** and of a complex Gaussian QPSK signal (with I and Q each having Gaussian distributions) where the average power of each signal is set to 0 dBm. Gaussian signals are of particular interest because their simple statistics lend them, and their interaction with nonlinear circuits, to quasi-analytic treatment [32]. The PMEPR of each signal is given in Table 4-4. From the shape of the APDF it is possible to estimate which signal will be more sensitive to nonlinear gain compression.

For example, even though QPSK has a higher PMEPR than OQPSK, the probability that the OQPSK envelope is near the peak is higher than for the QPSK signal. This is seen in Figure 4-10. It is not surprising then that the measured spectral regrowth of an OQPSK signal is higher than that for a QPSK signal. So PMEPR is only a rough guide to the distortion that is produced.

4.4.2 Design Guidelines

Generally an amplifier is not operated in saturation nor near the **third-order intercept** (**TOI** or IP3) point. An exception is with constant or near-constant envelope modulation schemes such as FM, GMSK, SOQPSK, FOQPSK, and SBPSK. With little variation in the amplitude of the RF signal, saturating amplifiers can be used with the near-constant envelope schemes. Also some advanced amplifier technologies can operate with nonlinear transistor loadlines and still have low distortion. For example, with some switching amplifiers produce little intermodulation distortion but a lot of harmonic distortion which is easily filtered out. With nonconstant envelope modulation schemes (i.e., the signal PMEPR is more than 0 dB), the variation in the amplitude of the modulated RF carrier results in signal distortion and spectral regrowth (i.e., power will be transferred into neighboring communication channels). One general rule of design is to ensure that the peak of the RF pseudo-carrier is at or below the 1 dB gain compression point. The amplifier is backed off by an amount approximately PMEPR below the 1 dB gain compression point. Generally, however, third-order intermodulation is a much greater concern, as this relates more closely to the amount of power dumped into neighboring channels.

Modern communication schemes can require that the adjacent channel spectral regrowth be as much as 80 dB below the power of the main channel. As a result, the signal must be backed off considerably from the IP3 point. The input or output power at IP3 (IIP3 or OIP3) is obtained from two-tone characterization and thus is a weakly accurate indicator of distortion with a digitally modulated signal. However, since it is a simple measurement to make and understand, it is widely used. Experience provides a rule of thumb for the backoff needed to ensure a maximum level of spectral regrowth for a particular modulation format and transistor technology [21].

Since the complexity of designing high-power and highly efficient amplifiers is considerable, it is necessary to use measurements following design to optimize the system for high efficiency while ensuring acceptable distortion. The most common measurement approach is to use loadpull, described in the next section. Design of the baseband circuit can also affect intermodulation distortion and spectral regrowth [33–35]. It is very important that the transistor ground be the system ground and that there be good heat-sinking so that the thermal circuit does not contribute to spectral regrowth through the variation of thermally dependent transistor parameters as the signal envelope varies (and thus the heat generated varies).

While IP3 and the 1 dB compression point only refer to amplitude distortion, there will also be phase distortion. While amplitude distortion has no effect on constant amplitude signals, phase distortion does. Phase distortion affects the performance of constant envelope modulation schemes

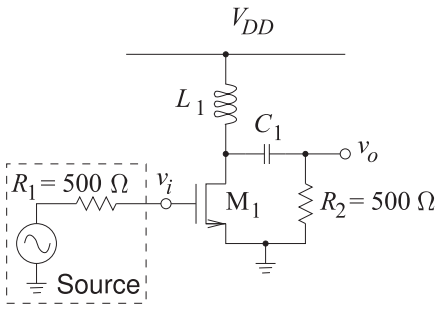


Figure 4-11: Class A inductively-biased MOSFET amplifier.

and the 1 dB compression point and IP3 provide some indication of possible phase distortion. Although EVM and BER are much better metrics of distortion for all digital modulation schemes, these are difficult to use in guiding initial design of RF hardware.

EXAMPLE 4.1 Intercept Point

The dynamic range derivations in Section 4.6 of [2] should be reviewed before working through this example.

The amplifier shown in Figure 4-11 is modeled, after expansion around the operating point, by a nonlinear transconductance with $i_{DS} = a_1 v_{GS} + a_3 v_{GS}^3$ with $a_1 = 0.01$ A/V and $a_3 = -0.1$ A/V^3. L_1 is an RF choke and can be assumed to present an open circuit at the operating frequency. C_1 is a biasing capacitor and can be treated as a short circuit at the operating frequency. R_2 is the load resistance.

(a) What is the small signal voltage gain of the amplifier?
(b) If the input signal is a two-tone signal consisting of two equal-amplitude sinewaves at 900 MHz and 901 MHz, what are the frequencies in the spectrum of v_o? Determine the output voltage across R_2?
(c) If a single sinusoid of amplitude 100 mV is applied to the gate of the MOSFET, what is the level of the fundamental tone at the output of the amplifier? What is the level of the third harmonic?
(d) What is the input-referred third-order intercept point, IIP3, of the amplifier?
(e) What is the output-referred third-order intercept point, OIP3, of the amplifier?

Solution:

(a) When the input signal v_i is small, $i_{DS} = a_1 v_i$, so the small signal output voltage is

$$v_o = -i_{DS} R_2 = -a_1 R_2 v_i$$

and the small signal voltage gain is

$$A = \frac{v_o}{v_i} = -a_1 R_2 = -0.01 \times 500 = -5.$$

(b) Eventually the symbolic amplitudes of the mixing terms will be required, so a trigonometric expansion is undertaken now. To minimize complexity, consider two cosinsoids, $A\cos(x)$ and $B\cos(y)$. So the input to the amplifier is

$$v_{GS} = A\cos(x) + B\cos(y).$$

The output of the amplifier is

$$
\begin{aligned}
v_o &= -i_{DS}R_2 = -R_2\left(a_1 v_{GS} + a_3 v_{GS}^3\right) \\
&= -R_2\left\{a_1\left[A\cos(x) + B\cos(y)\right] + a_3\left[A\cos(x) + B\cos(y)\right]^3\right\} \\
&= -R_2\left\{a_1\left[A\cos(x) + a_1 B\cos(y)\right] + a_3\left[A\cos(x) + B\cos(y)\right]\left[A\cos(x) + B\cos(y)\right]^2\right\} \\
&= -R_2\left\{a_1\left[A\cos(x) + a_1 B\cos(y)\right]\right. \\
&\qquad \left. + a_3\left[A\cos(x) + B\cos(y)\right]\left[A^2\cos^2(x) + 2AB\cos(x)\cos(y) + B^2\cos^2(y)\right]\right\} \\
&= -R_2\left\{a_1 A\cos(x) + a_1 B\cos(y)\right. \\
&\qquad + a_3\left[A\cos(x) + B\cos(y)\right] \\
&\qquad \left. \times \tfrac{1}{2}\left[A^2 + B^2 + A^2\cos(2x) + B^2\cos(2y) + 2AB\cos(x-y) + 2AB\cos(x+y)\right]\right\} \\
&= -R_2\left\{a_1\left[A\cos(x) + B\cos(y)\right]\right. \\
&\qquad + a_3\tfrac{1}{2}\left[\left(A^3 + AB^2\right)\cos(x) + A^3\cos(x)\cos(2x) + AB^2\cos(x)\cos(2y)\right. \\
&\qquad + 2A^2 B\cos(x)\cos(x-y) + 2AB^2\cos(x)\cos(x+y) + \left(A^2 B + B^3\right)\cos(y) \\
&\qquad + B^3\cos(y)\cos(2y) + A^2 B\cos(y)\cos(2x) + 2AB^2\cos(y)\cos(x-y) \\
&\qquad \left.\left. + \ 2AB^2\cos(y)\cos(x+y)\right]\right\} \\
&= -R_2\left(a_1\left[A\cos(x) + a_1 B\cos(y)\right]\right. \\
&\quad + a_3\tfrac{1}{2}\left\{\left(A^3 + AB^2\right)\cos(x) + \tfrac{1}{2}A^3\left[\cos(x) + \cos(3x)\right]\right. \\
&\qquad + \tfrac{1}{2}AB^2\left[\cos(2y-x) + \cos(x+2y)\right] + A^2 B\left[\cos(y) + \cos(2x-y)\right] \\
&\qquad + A^2 B\left[\cos(y) + \cos(2x+y)\right] + \left(A^2 B + B^3\right)\cos(y) \\
&\qquad + \tfrac{1}{2}B^3\left[\cos(y) + \cos(3y)\right] + \tfrac{1}{2}A^2 B\left[\cos(2x-y) + \cos(2x+y)\right] \\
&\qquad \left.\left. + AB^2\left[\cos(x) + \cos(2y-x)\right] + AB^2\left[\cos(x) + \cos(2y+x)\right]\right\}\right) \\
&= -R_2\left\{a_1 A\cos(x) + a_1 B\cos(y) \quad + a_3\tfrac{1}{4}\left[\left(3A^3 + 6AB^2\right)\cos(x)\right.\right. \\
&\qquad + \left(3B^3 + 6AB^2\right)\cos(y) + B^3\cos(3y) + 3A^2 B\cos(2x-y) \\
&\qquad \left.\left. + 3A^2 B\cos(2x+y) + A^3\cos(3x) + 3AB^2\cos(2y-x) + 3AB^2\cos(2y+x)\right]\right\}.
\end{aligned}
\tag{4.16}
$$

So with x representing 900 MHz and y representing 901 MHz, the frequencies in the output spectrum are 899, 900, 901, 902, 2700, 2701, 2702, and 2703 MHz.

(c) From Equation (4.16) and considering only one tone (i.e., $B = 0$), the output signal is

$$
v_o = -R_2\left[a_1 A\cos(x) + \frac{3a_3 A^3}{4}\cos(x) + \frac{a_3 A^3}{4}\cos(3x)\right]. \tag{4.17}
$$

So the coefficient of the fundamental at the output is

$$
v_o = -R_2\left[a_1 A\cos(x) + \frac{3a_3 A^3}{2}\cos(x))\right]. \tag{4.18}
$$

Here $A = 100$ mV, so the

$$
\begin{aligned}
\text{fundamental output} \quad &= -(500\ \Omega)\cdot\left[(0.01\ \text{A/V})\cdot(0.5\ \text{V}) + (-0.1\ \text{A/V}^3)\cdot(0.1\ \text{V})^3\right] \\
&= 0.05 - 0.0125\ \text{V} = 0.0375\ \text{V} = 37.5\ \text{mV}
\end{aligned}
\tag{4.19}
$$

$$
\text{third harmonic output} = \frac{a_3 R_2}{4}v_i^3 = (-500)\cdot(-0.1)\times(0.1)^3/4 = 0.0125\ \text{V}. \tag{4.20}
$$

(d) To answer this, determine the level of the fundamental and IM3 outputs for small v_{GS} and for two input tones having the same amplitude, $A = B = v_{GS}$. For a resistive nonlinearity, such as the transconductance here, the level of the lower and upper third-order intermods are the same. So, after examining Equation (4.16), consider $\cos(x)$ and $\cos(2x - y)$. The fundamental (at $\cos(x)$) is

$$v_o(900\text{ MHz}) = -R_2 a_1 v_i$$

and the level of the lower third-order intermod at $(2x - y)$ is

$$v_o(899\text{ MHz}) = -R_2 a_3 \tfrac{3}{4} (v_i)^3 .$$

IIP3 is the value of v_i when $v_o(900\text{ MHz}) = v_o(899\text{ MHz})$, that is, when

$$-R_2 a_1 v_i = -R_2 a_3 \tfrac{3}{4} (v_i)^3 .$$

That is, when $v_i = \sqrt{\left|\dfrac{4a_1}{3a_3}\right|} = \sqrt{\dfrac{4 \cdot 0.01}{3 \cdot 0.1}} = 0.3651\text{ V} = 365.1\text{ mV}.$

Thus $A_{\text{IIP3}} = 365.1\text{ mV}.$

Normally IIP3 is expressed in terms of power. Considering Figure 4-11, $E = 2v_i$, and so the available input power is

$$P_{\text{av}} = \frac{1}{2}\frac{v_i^2}{R_1} = \frac{v_i^2}{2R_1}.$$

Thus $\text{IIP3} = \frac{1}{2R_1} A_{\text{IIP3}}^2 = \dfrac{0.3651^2}{2 \cdot 500} = 0.0001333\text{ W} = 0.1333\text{ mW} = -0.875\text{ dBm}.$

(e) The voltage output-referred intercept point is

$$A_{\text{OIP3}} = |A|\, A_{\text{IIP3}} = 5 \cdot 0.3651\text{ V} = 1.8255\text{ V}$$

and $\text{OIP3} = (\text{Power gain}) \cdot \text{IIP3} = \dfrac{R_1}{R_2} A^2 \cdot 0.1333\text{ mW} = 3.333\text{ mW} = 5.23\text{ dBm}.$

4.5 Loadpull

Power amplifiers are designed beginning with an initial concept, detail designed using nonlinear circuit simulation tools, and finally optimized in the laboratory. However, computer-assisted design relies on models of components that can never capture all effects including, for example, thermal effects and parasitic coupling of components. A final experimental optimization for gain and efficiency while limiting distortion is nearly always required [36–38]. Once the design has been optimized in the laboratory, it is found that a design can be fixed for manufacturing. Minimal manual tuning of individual production units to peak performance is then usually all that is required. The most useful experimentally based design optimization approach uses the load-pull method. This technique uses computer-controlled variable input and output matching networks to search for the optimum input and output conditions. As well, performance with digitally modulated signals can be investigated, and this is very difficult to do in simulation.

A load-pull system is shown in Figure 4-12. The automated tuners realize variable input and output networks and are generally based on automated double-stub slabline tuners where the position of the stubs is also variable. The output impedance of the active device in a power amplifier is typically around a few ohms or less and it is often necessary to use an impedance transformer to scale the output impedance of the device up to the typically $50\ \Omega$ system impedance of the automated tuners.

The load-pull system shown in Figure 4-12 is configured to measure the reflected power at the input of the amplifier (through the input

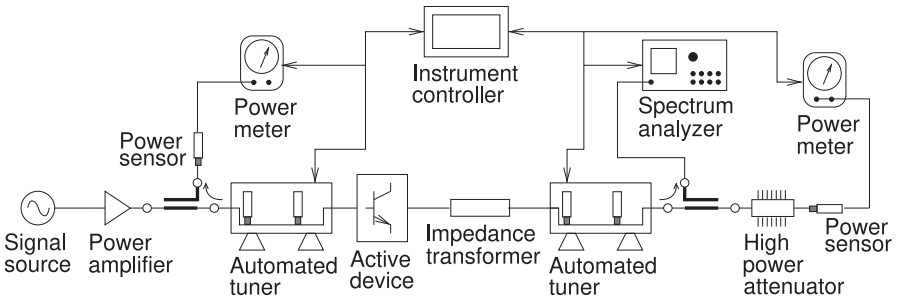


Figure 4-12: Load-pull system.

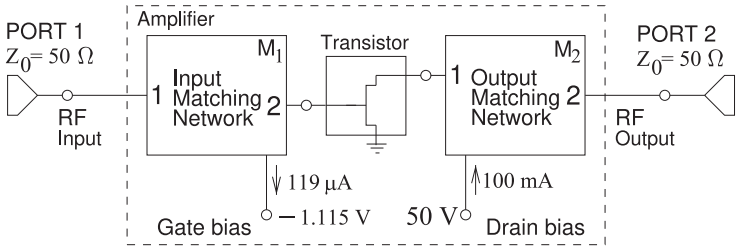


Figure 4-13: WiMAX amplifier block diagram showing the DC gate and drain currents.

directional coupler), the spectrum at the output of the amplifier (again through a directional coupler), and the output power (following a high-power attenuator). From the spectrum and power measurements, the gain, efficiency, output power, EVM, and spectral regrowth metrics are found. Many of these factors can be optimized [36] by systematically presenting impedances to the active device from a grid of possible input and output impedances. The tuners shown in Figure 4-12 provide controlled matching networks at the fundamental frequency. More elaborate systems can provide separate tuning at several harmonics. Microwave computer-aided design tools also support load-pull calculations for amplifiers.

The input automated tuner, the active device, the impedance transformer, and the output automated tuner become the amplifier. Once the optimum settings of the automated tuners are found for a range of frequencies, the design task then becomes realizing matching network with S parameters that correspond to the tuner settings.

4.6 Case Study: Design of a WiMAX Power Amplifier

In this section the design and simulation of a 10 W power amplifier module using a GaN HEMT transistor, here a Eudyna EGN010MK device, is examined.[1] A GaN (gallium nitride) transistor can have a very large drain-source voltage and can produce substantial output power. The amplifier schematic is shown in Figure 4-13. The amplifier is targeted for use in a WiMAX data communication system operating from 3.4 to 3.8 GHz.

4.6.1 Input and Output Matching Networks

The harmonic balance simulation employed in design uses a large-signal transistor model supplied by the manufacturer of the transistor and, as is usual, it is proprietary and cannot be viewed in the simulator. A

[1] AWR Design Environment Project File: WiMAX_amp.emp.

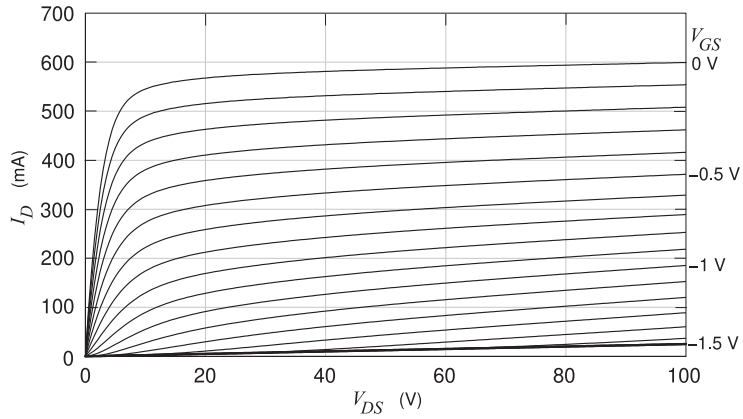
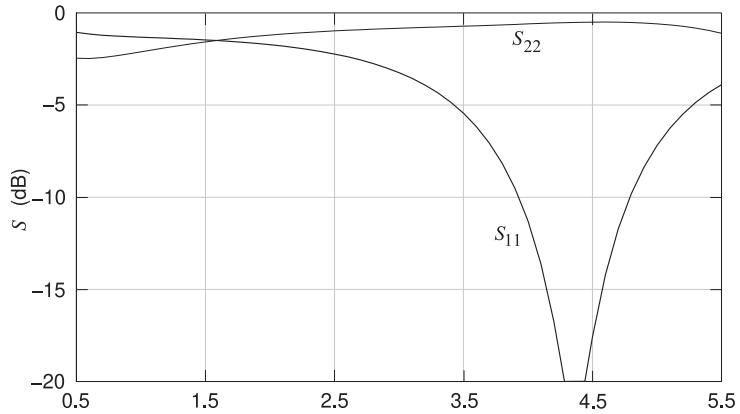
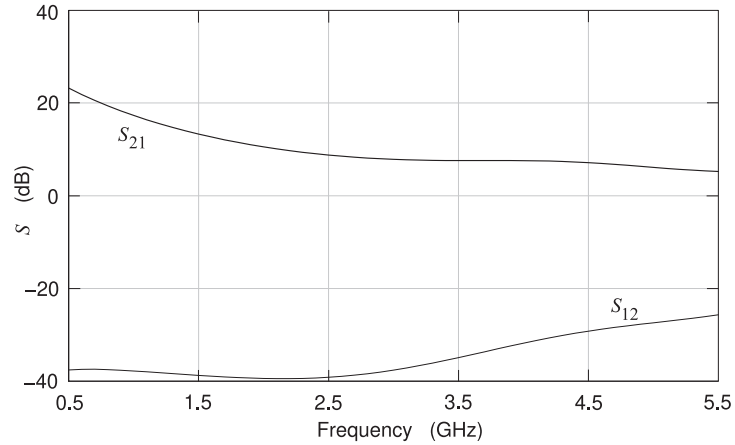


Figure 4-14: Output current-voltage characteristics of the active device in the WiMAX amplifier.



(a) S_{11} and S_{22}



(b) S_{21} and S_{12}

Figure 4-15: Small signal scattering parameters of the active device of the WiMAX amplifier.

designer must initially work with the transistor's output current-voltage characteristic, as shown in Figure 4-14. The transistor is biased with a gate-source voltage of -1.115 V, a drain-source voltage of 50 V, and a drain current of 100 mA. At this bias point the input and output 50 Ω S parameters of the transistor are as shown in Figure 4-15. Examination of S_{11} in Figure 4-15(a) indicates that the input impedance of the GaN transistor is approximately

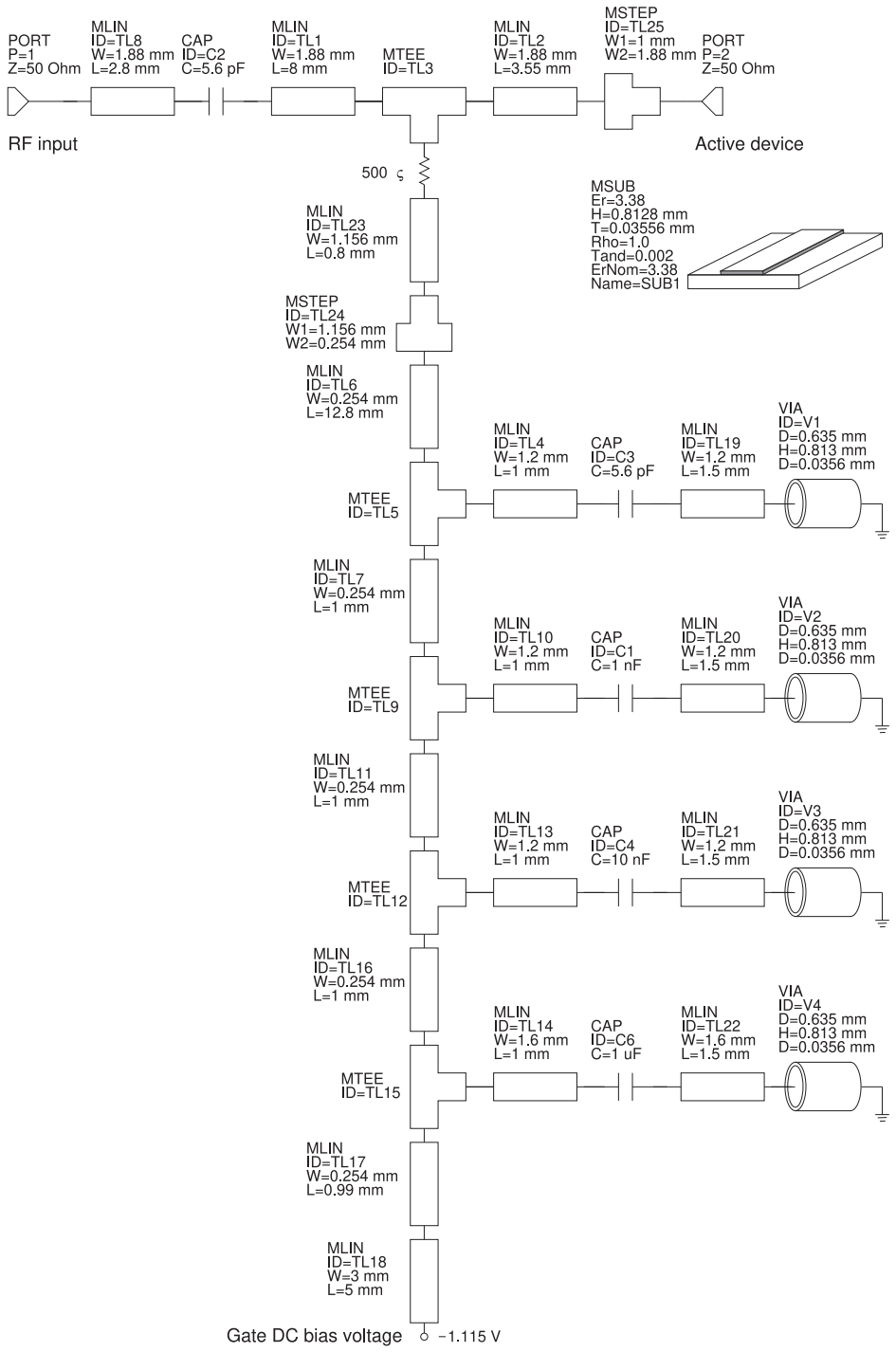


Figure 4-16: Input matching network of the WiMAX amplifier.

50 Ω over the WiMAX frequency range of 3.4–3.8 GHz. As a result, very little input matching is required. The circuit model of the input network is shown in Figure 4-16. The input network is shown with 50 Ω ports at Ports 1 and 2. These are used in evaluating the S parameters of the network but are removed when the network is incorporated in the complete amplifier shown

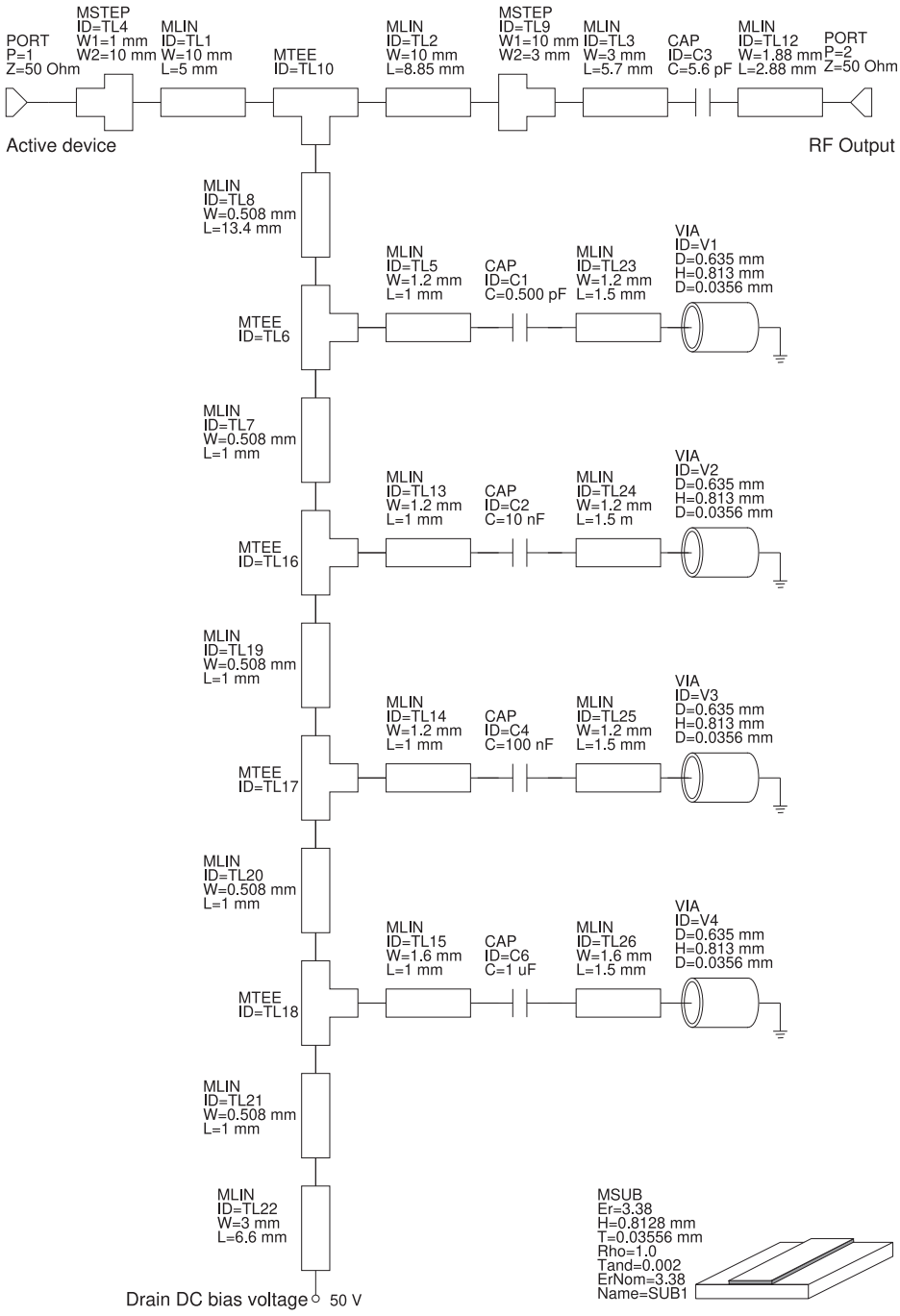


Figure 4-17: Output matching network of the WiMAX amplifier.

in Figure 4-13.

Examination of the output reflection coefficient of the transistor indicates that it can be approximately modeled by a current source in parallel with a 33 Ω resistor and a 4.4 pF capacitor (an impedance of $2.5 - j11\ \Omega$ at 3.6 GHz) and so output matching is required to transition to 50 Ω. The circuit model of the output network is shown in Figure 4-17. The main component of the output matching network is an approximate one-quarter wavelength long

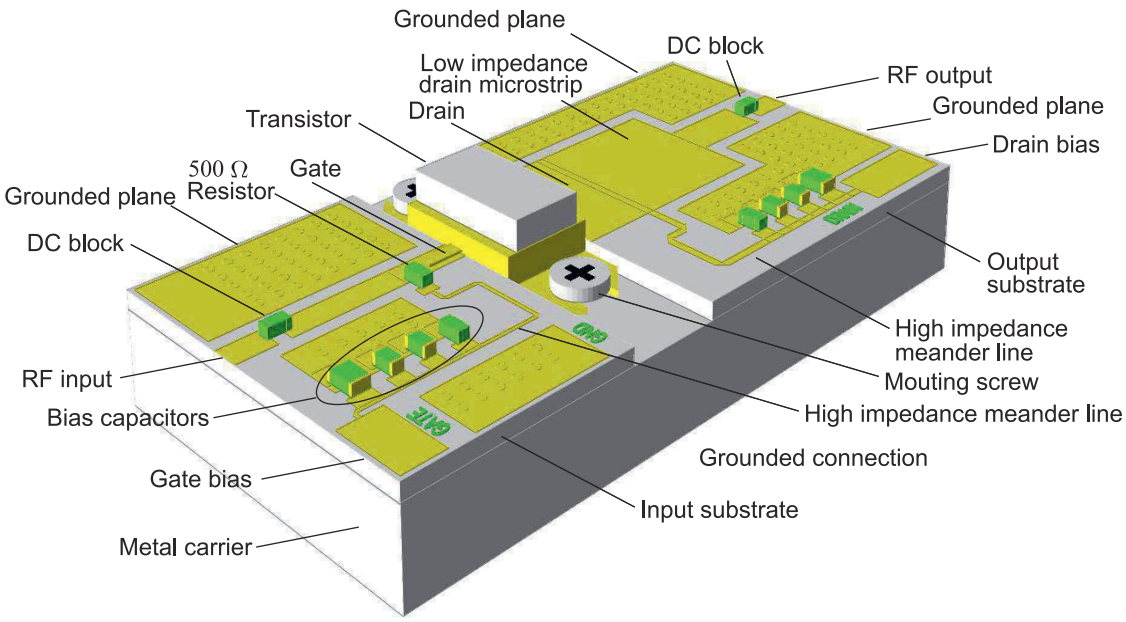


Figure 4-18: WiMAX amplifier operating from 3.4 to 3.8 GHz.

impedance transformer implemented by the microstrip transmission lines (MLIN) TL1 and TL2. This is the basis of the output network design and other components complete the network.

The complete power amplifier module is shown in Figure 4-18. A major consideration in the mechanical assembly is management of the heat produced by the active device. The semiconductor is valence-bonded by the transistor foundry to a metal plate which is secured to the metal carrier using mounting screws and thermal paste between the metal carrier and the metal backing of the transistor. The metal carrier is typically mounted on a heat sink.

The input and output networks are on separate substrates, as this enables the transistor to be mounted directly on the metal carrier. The input substrate supports a microstrip circuit with a 50 Ω RF input connected through a DC blocking capacitor to a microstrip line that drives the gate. Near the transistor, the gate microstrip line is connected through a 500 Ω resistor to a DC bias network. The 500 Ω resistor provides part of the RF isolation between the gate line and the gate bias network. The gate current drawn by the transistor is 11 μA so that there is a 0.06 V DC voltage drop across the 500 Ω resistor. From the 500 Ω isolation resistor, a high-impedance (i.e., narrow) microstrip meander line continues to a sequence of RF shorting capacitors of increasing capacitance ranging from 5.6 pF next to the meander line to 1 μF at the gate bias pad from which wirebonds lead to an external source. This arrangement minimizes the impact of the increasing parasitic series inductances of the capacitors as their capacitance values increase. Together, the 500 Ω resistor and the high-impedance line provide significant RF isolation from the external bias network. Note that

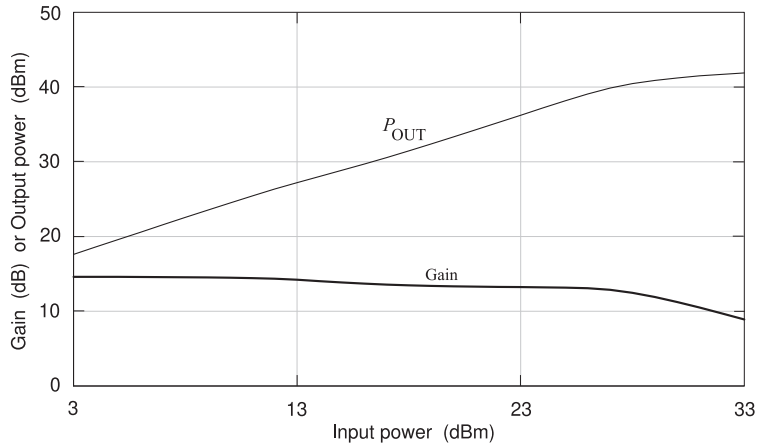


Figure 4-19: WiMAX amplifier simulated output power, P_{OUT}, and gain at 3.5 GHz.

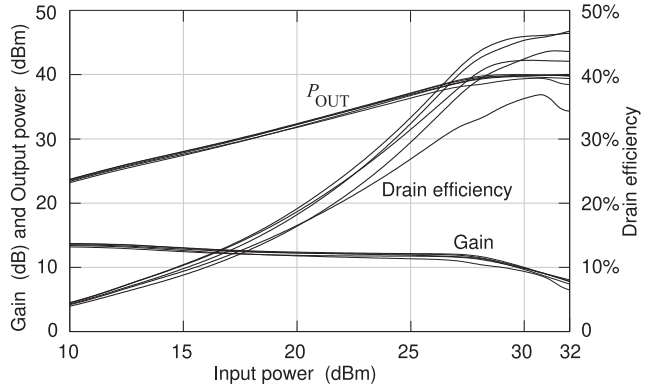


Figure 4-20: WiMAX amplifier simulated output power, gain, and drain efficiency at 3.4, 3.5, 3.6, 3.7, and 3.8 GHz.

areas of the input network without microstrip elements are grounded planes. These are regularly shorted by vias to the metal backing of the microstrip substrate, which is in turn connected to the metal carrier. This arrangement suppresses substrate modes.

The output network is similarly arranged, but now the bias resistance is omitted, as the voltage drop across it would be too great since the DC drain bias current is 100 mA. The output impedance of the active device two-port is very low, so the first microstrip line is wide and so has low characteristic impedance. Now RF isolation from the drain bias network is provided by the high-impedance meander line and there is a high-impedance contrast to the low-impedance drain line.

The nonlinear performance of the amplifier is characterized using one-tone and two-tone tests. The one-tone test results are shown in Figure 4-19, where the input power of a single input tone, a sinewave, at 3.5 GHz is swept from 3 dBm to 33 dBm. The output power at first increases linearly but then begins to saturate and the gain of the device drops. The small-signal gain is 13.3 dB and the output power at 1 dB gain compression is 28.2 dBm while the gain of the amplifier is 12.3 dB. While the gain and output power vary slightly over the WiMAX frequency range of 3.4 to 3.8 GHz (see Figure 4-20), the drain efficiency is more sensitive.

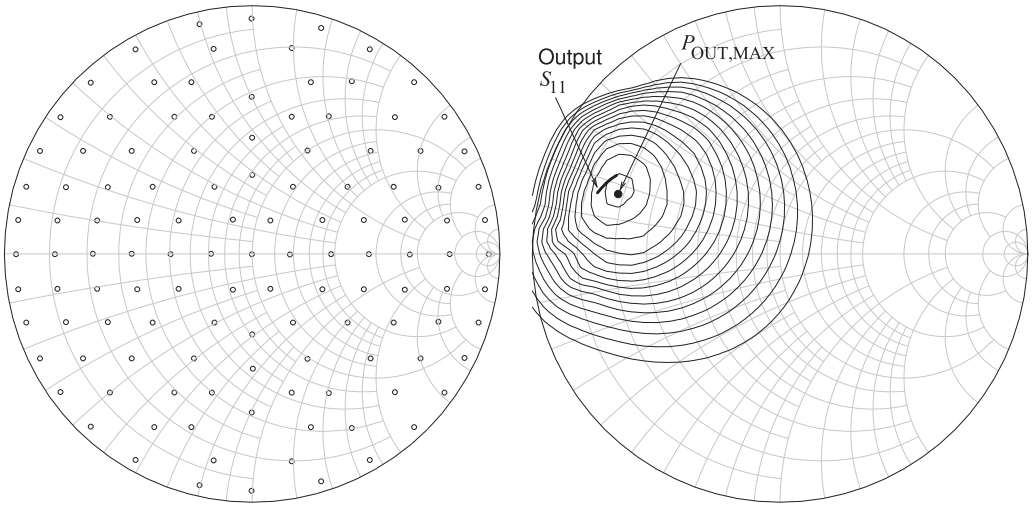


Figure 4-21: Simulated output load-pull characterization of the WiMAX amplifier: (a) load-pull points used; and (b) load-pull 0.1 dB contours of the amplifier at 3.5 GHz showing the locus of the reflection coefficient looking from the transistor into the output matching network as contours of constant output power. The family of contours begins with the point of maximum output power, $P_{\mathrm{OUT,MAX}}$ = 42.6 dBm. The output power for the first contour surrounding $P_{\mathrm{OUT,MAX}}$ is 42.5 dBm and the powers of the contours then reduce in 0.5 dB steps. Also shown is the S_{11} of the final output matching network design from 3.4 to 3.8 GHz. The 'Output S_{11}' is the reflection coefficient from 3.4 to 3.8 GHz of the output matching network looking from the active device indicating the simulated output power is between 42.4 dB and 42.5 dB (approximately). (A complete determination would require load-pull contours at several frequencies.)

4.6.2 Load-Pull

Optimum design of a power amplifier requires load-pull characterization in which, often, the input matching network is already designed, and in simulation a special load-pull element is used for the output matching network. This element can present a range of matching transformations between the output of the active device and the output load. An array of transformations is specified in terms of reflection coefficients such as that shown in Figure 4-21(a). At each of the points, the output power, P_{OUT}, is calculated and contours of constant output power are interpolated from these data, yielding the load-pull contours shown in Figure 4-21(b). The frequency locus (from 3.4 to 3.8 GHz) of S_{11} of the output matching network (looking from the active device) is also shown in Figure 4-21(b). The design choice made for the output matching network is for its S_{11} to be tangential to one of the load-pull contours so that the output power is constant across the frequency band. So rather than designing for maximum power transfer, which practically could only be achieved at one frequency, the design choice is to achieve flat gain across the band. Other quantities can also be plotted in the load-pull test. These include efficiency, the level of harmonics, and distortion terms.

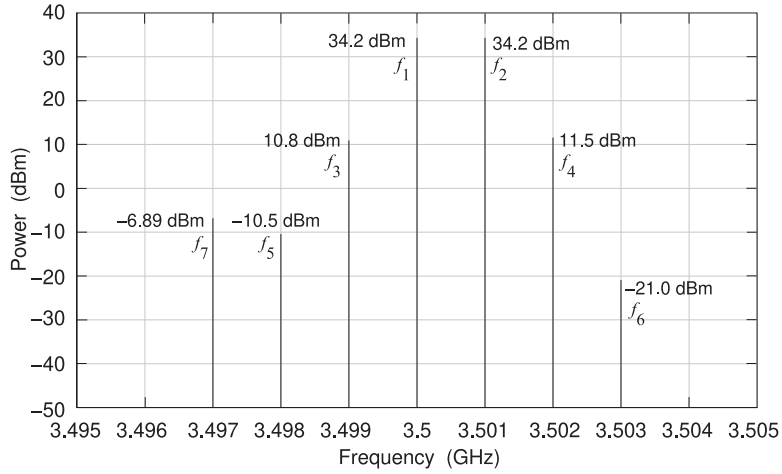


Figure 4-22: Simulated output spectrum with a two-tone input signal comprising tones at 3.500 GHz and 3.501 GHz each being 23 dBm.

4.6.3 Two-Tone Characterization

The second test used to quantify the nonlinear performance of a power amplifier is the two-tone test in which two closely spaced tones of equal power are applied at the input of the amplifier. The output spectrum is then calculated and tabulated for a range of input powers. The output spectrum under one condition is shown in Figure 4-22. The output spectrum consists of amplified versions of the two input tones at $f_1 = 3.500$ GHz and $f_2 = 3.501$ GHz as well as intermodulation products separated by multiples of the input tone spacing. The tones at $f_3 = 3.499$ GHz and $f_4 = 3.502$ GHz are called third-order intermodulation components, abbreviated as IM3, where $f_3 = 2f_1 - f_2$ and $f_4 = 2f_2 - f_1$. More specifically, the tone at f_3 is called the lower IM3 tone ($\mathrm{IM3_L}$) and the tone at f_4 is called the upper IM3 tone ($\mathrm{IM3_U}$). The next tones spaced one further interval out, the f_5 and f_6 tones, are called the fifth-order intermodulation distortion tones, the IM5 tones. Only one of the IM7 tones, f_7, has sufficient amplitude to be seen on the scale in the figure. The amplitudes of the output tones are not symmetrical around f_1 and f_2 and this effect is attributed to the impedance of the baseband circuit, as some of the input signal power is converted to baseband before being converted back up again [39–43]. The thermal modulation of the transistor characteristics, especially mobility, due to temperature variations at or slower than the rate at which the envelope varies, can also produce this effect. Minimizing temperature variations is an additional reason for very good heat-sinking of power devices.

In Figure 4-23 the output power at both the fundamental and $\mathrm{IM3_U}$ are plotted versus the input power in a two-tone test. The output power at the fundamental increases linearly with the input power until saturation is reached. The $\mathrm{IM3_U}$ signal initially increases with the third power of the input signal level, which corresponds to the model of the active device around the bias point at low signal levels being a cubic power series. As the power increases, high-order nonlinearities become significant and the IM3 behavior is less predictable.

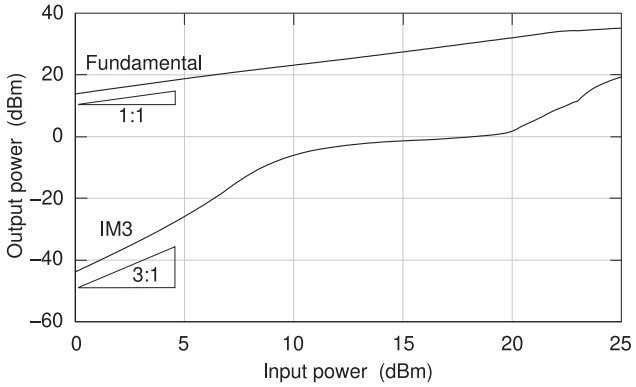
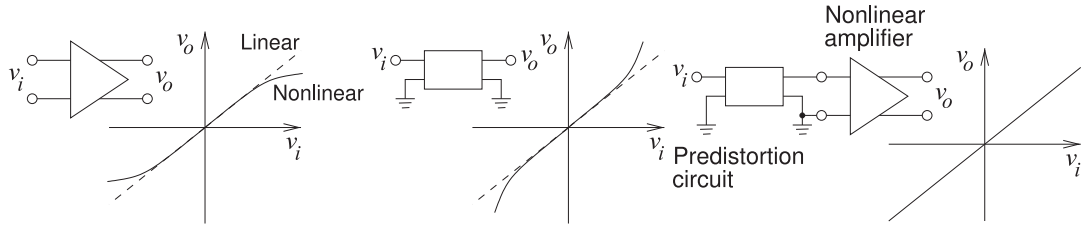


Figure 4-23: Response to a two-tone input signal: output power at the fundamental and at $\mathrm{IM3_U}$ versus input power. The two input tones are at 3.500 GHz and 3.501 GHz and the fundamental plotted is at 3.501 GHz.



(a) Nonlinear amplifier (b) Nonlinear compensator (c) Linearized amplifier

Figure 4-24: Predistortion linearizer concept.

4.6.4 Summary

Design of power amplifiers considers discrete tones and in particular uses two tones to characterize distortion. Discrete tones are required in simulation and two-tone characterization is convenient in a laboratory. Of course, digitally modulated signals are not discrete tones, but designers have found that the nonlinear responses to one tone and to two tones are good indications of the behavior with digitally modulated signals which are not convenient to use in simulation [6, 44].

4.7 Linearization

Linearization refers to strategies that compensate for the inherent phase and amplitude nonlinearities of RF amplifiers. There are three basic strategies that introduce a correction mechanism [45–50]:

(a) predistortion linearization,
(b) feed-forward linearization, and
(c) linearization by design.

Reducing distortion using the above techniques enables an amplifier to operate at higher efficiency while producing the maximum permissible amount of distortion.

4.7.1 Predistortion

The predistortion concept is illustrated in Figure 4-24. The essential nonlinearity of an RF amplifier is a tanh-like response, as shown in

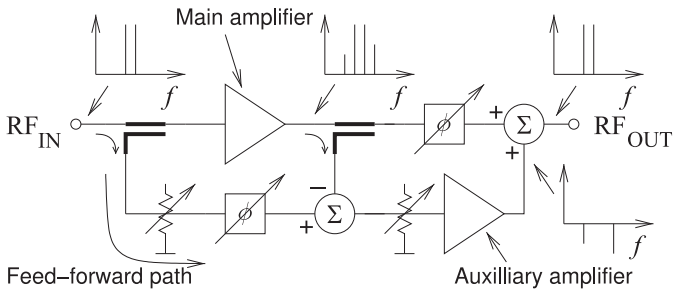


Figure 4-25: Feed-forward linearizer concept with the plots showing the spectra in different parts of the circuit.

Figure 4-24(a). With predistortion, a compensation network that has a complementary response, such as that shown in Figure 4-24(b), is developed [49, 50]. When the compensation network precedes the amplifier, ideally the output is linearized with little final distortion (see Figure 4-24(c)). Predistortion can be analog predistortion, as shown here, or the compensating function can be performed at baseband in a DSP unit. Predistortion requires a good behavioral analytic model of the amplifier from which distortion can be calculated. Then this model is reverted to synthesize the compensation network. Variations of the predistortion strategy include adaptive predistortion in which the distortion at the output of the amplifier is regularly sampled during operation and the coefficients of the compensation network updated. The linearization achievable is limited by the accuracy of the reversion synthesis, the stability of the amplifier circuit, and the stability of the adaptive process.

4.7.2 Feed-Forward Linearization

A linearization strategy that does not require characterization of the nonlinear process is feed-forward linearization, the concept of which is shown in Figure 4-25 [48]. In Figure 4-25 a two-tone input signal is considered for illustration. The two-tone signal is sampled prior to amplification by the main amplifier, which produces intermodulation distortion where third-order intermodulation distortion is shown in the spectrum at the main amplifier output. The original undistorted input signal is sampled by the directional coupler at the input and, following amplitude and phase adjustment, is compared to the sampled distorted signal at the output of the main amplifier. The difference is then the distorted waveform, which is amplified by an auxiliary amplifier and then added into the main signal path (after appropriate phase adjustment). This addition ideally cancels distortion in the high-power signal. Feed-forward linearization requires a very linear auxiliary amplifier, but this is easier to achieve than for the main amplifier since the auxiliary amplifier operates at much lower power levels.

4.7.3 Summary

The third category of linearization is linearization by design, achieved by choosing amplifier topologies that inherently compensate for distortion. Many of these topologies will be considered in the next sections.

The various linearization strategies can be combined, but linear by design topologies are very important, as predistortion and feed-forward linearization can consume significant DC power, especially for wideband signals.

4.8 Advanced Power Amplifier Architectures

A central challenge of power amplifier design is trading off power efficiency and distortion. Up to now discussion has focused on efficient design of single transistor amplifiers. In this section several architectures are described that enable the efficiency-distortion trade-off to go to a higher level. The three architectures discussed are the Doherty amplifier, the envelope tracking amplifier, the LITMUS amplifier, and the LINC amplifier.

4.8.1 Doherty Amplifier

A Doherty amplifier improves linearity and achieves high efficiency by using two amplifiers, one called a carrier amplifier and the other a peaking amplifier [45, 46, 51–55]. The basic Doherty amplifier configuration is shown in Figure 4-26(a). Here the input signal is split and half is applied to the carrier amplifier, typically a low-distortion Class AB amplifier, which amplifies small signals. However, as the input signal increases this amplifier becomes nonlinear and saturates, as seen in Figure 4-26(b). The other amplifier is typically a Class C amplifier and does not amplify small signals, but as the input becomes larger, the amplifier turns on. The Doherty amplifier has high linearity, as at high input powers the turn-on characteristic of the peaking amplifier complements the saturation of the carrier amplifier. The essential operation is that the input signal is split and half is shifted by $90°$. At the output this phase shift is matched by the $90°$ electrical length of the transmission line at the output of the carrier amplifier. This transmission line efficiently routes the output of the carrier amplifier to the load whether or not the peaking amplifier is turned on. When it is not turned on, the peaking amplifier presents an open circuit, but when it is on, it presents a Thevenin impedance Z_0. Then, the output power of both the carrier and peaking amplifiers are efficiently combined.

A more complete circuit design of an HBT Doherty amplifier is shown in Figure 4-27. The $90°$ hybrid at the input splits the signals and introduces the required phase shift. Base biasing voltages of the HBT transistors (i.e. V_{bc} and V_{bp}) are adjusted so that the carrier amplifier will amplify small signals, but the peaking amplifier will have a Class C-like characteristic and only turn on for large signals.

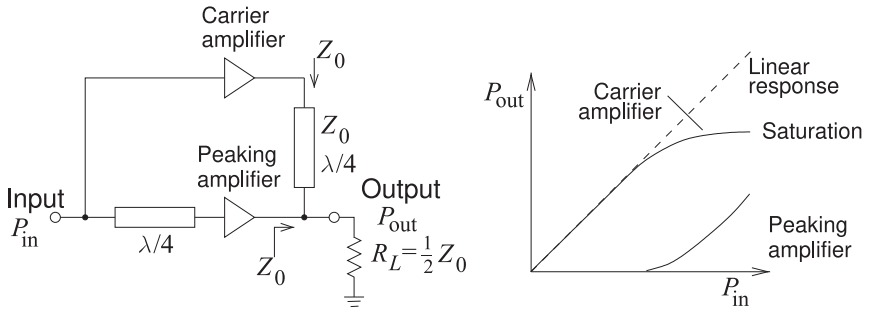


Figure 4-26: Doherty amplifier. (a) Doherty amplifier concept (b) Amplifier output characteristic

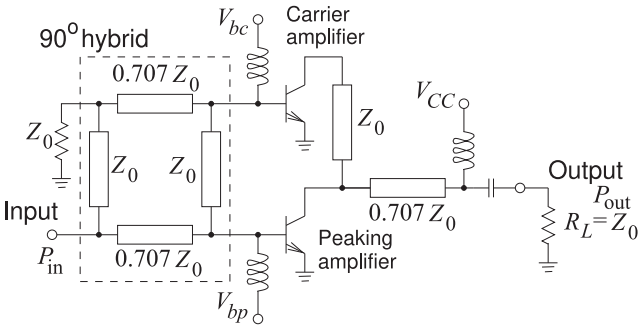


Figure 4-27: HBT Doherty amplifier with a 90° hybrid between P_{in} and the input of the transistors. All lines are $\lambda/4$ (i.e. 90°) long.

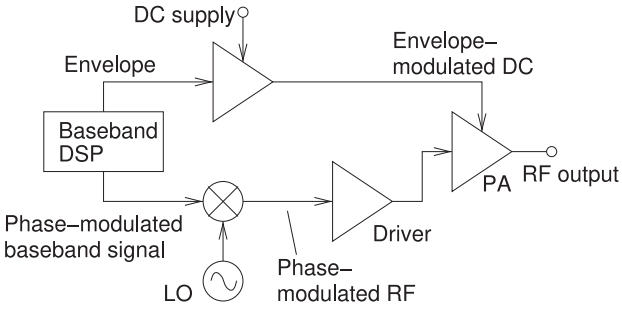


Figure 4-28: Khan transmitter using envelope elimination and restoration.

4.8.2 Envelope Tracking Amplifier

The concept of the envelope tracking amplifier was introduced with the **Khan transmitter** [56–59]. This technique is also called the **envelope-elimination and restoration** (**EER**) technique. The original amplifier concept [56], but updated to use modern DSP capabilities, is shown in Figure 4-28. The DSP operates at baseband and produces the digitally modulated signal to be transmitted. The DSP separately produces an amplitude-varying envelope signal and a constant-envelope phase-modulated baseband signal. These are amplified separately. This is in contrast to the usual approach of producing a baseband signal, or even a low-power RF signal, which has the complete modulation with both phase and amplitude variations. The phase-modulated signal is amplified by a driver before it is input to an efficient power amplifier that always operates in a highly linear mode since the DC source to the PA is adjusted to match the envelope of the final signal. Thus the PA is always operating at peak efficiency. These amplifiers have excellent linearity over a broad frequency range.

4.8.3 LINC Amplifier

The linear amplification with nonlinear component (LINC) amplifier separates a digitally modulated signal into two constant envelope components [18, 60–62]. This amplifier is also known as an **outphasing amplifier**. Each of the components is efficiently amplified by individual amplifiers and then recombined. The amplifier concept is shown in Figure 4-29.

The LINC amplifier operates as follows. The digitally modulated RF input signal can be expressed as

$$s(t) = a(t)\cos[\omega t + \theta(t)]. \tag{4.21}$$

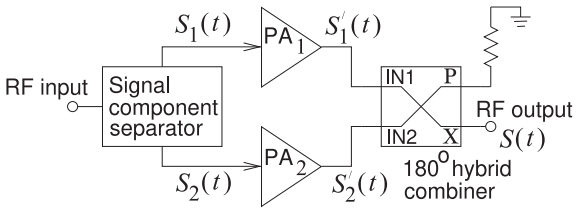


Figure 4-29:
LINC amplifier.

In complex exponential form this becomes

$$s(t) = a(t)\mathrm{e}^{\jmath\theta(t)}, \tag{4.22}$$

where $a(t)$ describes the amplitude modulation of the signal and $\theta(t)$ describes the phase modulation of the signal. This signal is split by the signal component separator to create two components, $S_1'(t)$ and $S_2'(t)$, with the same constant amplitudes and with complementary time-varying phase modulation:

$$S_1(t) = \tfrac{1}{2}\left[s(t) - e(t)\right]_{\mathrm{I}} = \tfrac{1}{2}V_m \cos\left[\omega t + \theta(t) - \psi(t)\right] \tag{4.23}$$

$$S_2(t) = \tfrac{1}{2}\left[s(t) + e(t)\right]_{\mathrm{I}} = \tfrac{1}{2}V_m \cos\left[\omega t + \theta(t) + \psi(t)\right], \tag{4.24}$$

where $\psi(t) = \cos^{-1}\left[a(t)/V_m\right]$, the subscript I denotes the in-phase component and $\frac{1}{2}V_m$ is the magnitude of the voltage signal applied to each of the PAs. The $e(t)$ term is called the quadrature signal and

$$e(t) = \jmath s(t)\sqrt{\frac{V_m^2}{a^2(t)} - 1}. \tag{4.25}$$

$S_1(t)$ and $S_2(t)$ are efficiently amplified separately to produce the amplified signals $S_1'(t)$ and $S_2'(t)$ and these are combined at the output by a $180°$ hybrid. Over time the amplitudes of $S_1'(t)$ and $S_2'(t)$ do vary but only when the average power of the RF signal changes. However in one data packet their amplitudes stay the same.

Ideally $S_1'(t)$ and $S_2'(t)$ are linearly scaled versions of $S_1(t)$ and $S_2(t)$ so that

$$S_1'(t) = kS_1(t) \quad \text{and} \quad S_2'(t) = kS_2(t), \tag{4.26}$$

and k may be complex. The output of the combiner is

$$S(t) = S_1'(t) + S_2'(t) = \tfrac{1}{\sqrt{2}}ka(t)\cos\left[\omega t + \theta(t) + \phi\right], \tag{4.27}$$

where k is the gain of each PA, ϕ is the phase shift introduced in the amplifier path, and the factor of $\frac{1}{\sqrt{2}}$ is introduced by the combiner. Thus $S(t)$ is a linearly amplified version of $s(t)$.

The LINC amplifier relies on the two PAs being well balanced and this must be maintained for a wide range of conditions and this can sometimes be difficult to achieve.

4.8.4 LITMUS Amplifier

The linear amplification by time-multiplexed spectrum (LITMUS) amplifier [63, 64], shown in Figure 4-30, utilizes the ability of a bandpass filter to

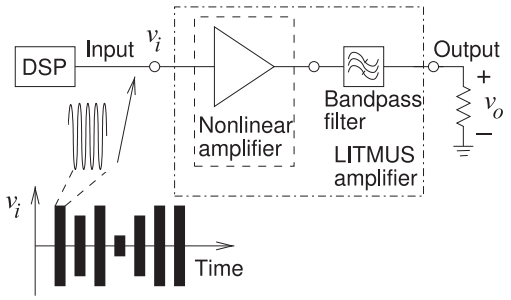


Figure 4-30: LITMUS amplifier. Each pulse of the input signal, v_i, has many sinusoids. For each pulse the sinusoids have different frequency, amplitude, and relative phase.

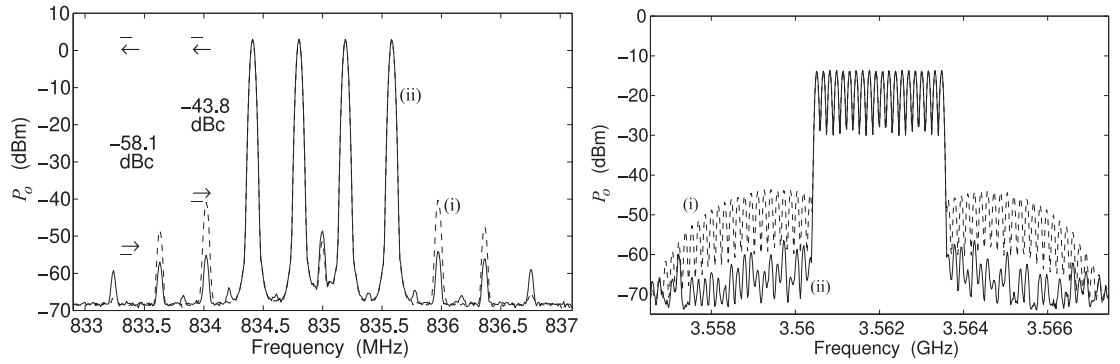
combine signals of different frequencies spaced in time. This is a result of the time-frequency behavior of filters, which is discussed in Section 2.18 of [2].

The most common way of reducing distortion of an amplifier is by setting the average output power of a modulated signal at a level that is below the 1 dB gain compression level by PMEPR. This is called backing off. So while an amplifier, such as a switching amplifier or a Class C amplifier could have close to 100% efficiency for a signal with a constant envelope, the back-off required for modulated signals with nonconstant envelopes significantly reduces the achievable efficiency.

A digitally modulated signal in a time interval Δt (coinciding with one or a few symbol interval times) can be decomposed into a sum of sinusoids that differ in amplitude, frequency, and relative phase [65]. (Generalizing this, the signal could be represented as a sum of constant envelope signals.) Now each sinusoid can be represented by a CW pulse of duration $\Delta\tau \ll \Delta t$ so that each pulse contains $\Delta\tau/T$ cycles of its sinusoidal constituent, where T is the period of the sinewave. These pulses can be separated in time, as shown for the input signal, v_i, in Figure 4-30. This input CW pulse train, generated in the DSP unit, is then applied to the input of the highly efficient nonlinear amplifier, which produces CW pulses at the output of the nonlinear amplifier. The output CW pulses are then combined in the bandpass filter. The bandpass filter spreads the pulses in time and reconstitutes the original, but now amplified, digitally modulated signal. During the time $\Delta\tau$ of each input pulse, the nonlinear amplifier is amplifying a constant envelope signal so the amplifier can operate in high-efficiency mode and saturation of the nonlinear circuit is not of concern. Also, since the individual frequency components are separated in time in the nonlinear part of the circuit, there can be no intermodulation distortion.

Theoretically a LITMUS amplifier completely suppresses intermodulation distortion, but the suppression achieved is limited by the bandwidth of the pulse train and also limited by reactive and memory effects in the nonlinear amplifier, especially baseband effects (including thermal effects), which provide long-term memory of the signals in the nonlinear circuit.

The output spectrum following amplification of a signal decomposed into four tones is shown in Figure 4-31(a). Curve (i) is the spectrum of the output of a conventional amplifier and Curve (ii) is the output of a LITMUS amplifier. Both amplifiers produce the same average RF power at the output. The output power of each of the four tones is 0 dBm. In Figure 4-31(a) six intermodulation tones are seen, three below the four fundamental tones and three above. These intermodulation tones are IM3 tones, as their frequencies can be numerically calculated as twice the frequency of one of the fundamental tones minus the frequency of one of the other tones. For



(a) Four tones (b) Twenty tones

Figure 4-31: Measured spectra at the output of the LITMUS amplifier for a different number of input tones. Curve (i) in both figures is the spectrum of the output of a conventional amplifier with a multitone input signal. Curve (ii) is the spectrum of the output of a LITMUS amplifier after the bandpass filter. The spike seen at 835 MHz is an artifact of the arbitrary waveform generator used to produce the signal. After [64].

the conventional amplifier, the largest intermodulation tone has an output power of -43.8 dBm. For the LITMUS amplifier the largest intermodulation tone has a power of -58.1 dBm. The intermodulation distortion is reduced by 14.3 dB.

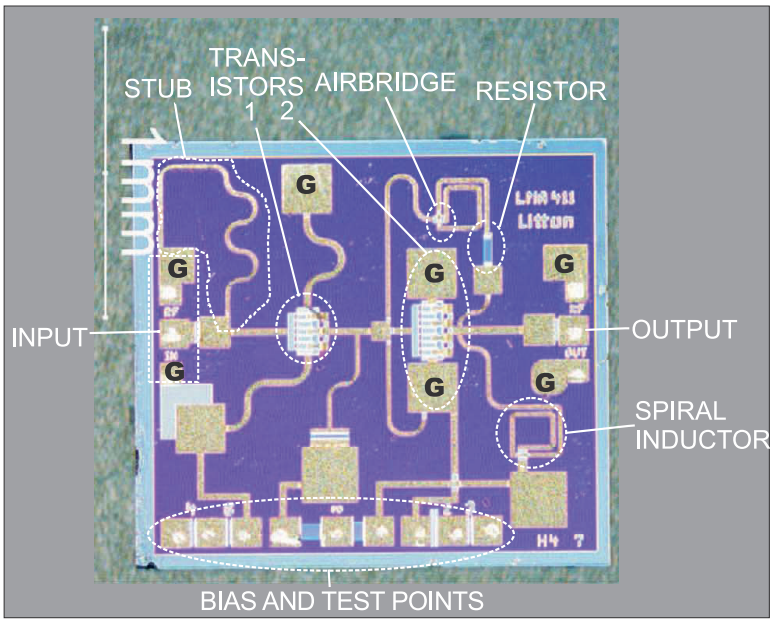
The amplification of a signal decomposed into 20 tones is shown in Figure 4-31(b). Again Curve (i) is the spectrum of the output of a conventional amplifier and Curve (ii) is the output of a LITMUS amplifier for the same output power. Now spectral regrowth is reduced by 13.9 dB.

4.8.5 Summary

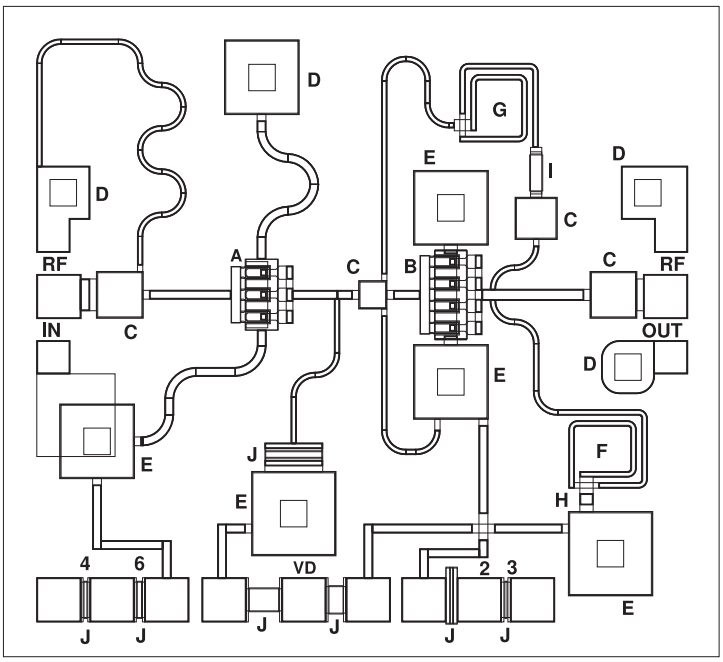
There are many variations on the advanced power amplifier architectures discussed in this section. These architectural concepts can be combined with each other and can also be combined with switching amplifier concepts. So the number of different power amplifier conceptual combinations is enormous and many papers have been written about them. Some amplifier topologies require high-performance DSP and high-performance transistors, while others require extensive manual tuning of each amplifier. These affect the cost of each amplifier. Also, the design cost of some amplifiers can be high and require the most experienced designers. These factors are considered in choosing the best amplifier architecture for a particular application.

4.9 MMIC Power Amplifiers

A linear Class A X-band amplifier implemented in GaAs MESFET **microwave monolithic integrated circuit** (**MMIC**) technology is shown in Figure 4-32. This amplifier operates from 8 to 12 GHz producing 100 mW of total power. This is a two-stage amplifier, as this results in the required gain, power, and bandwidth. The input and output pads are in a ground-signal-ground (GSG) configuration and can be used for both microwave



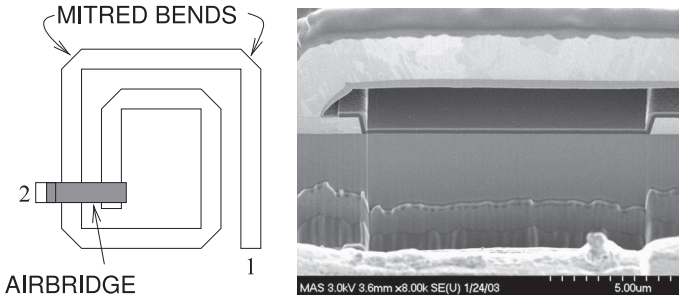
(a)



(b)

Figure 4-32: A two-stage X-band (8–12 GHz) MMIC amplifier producing 100 mW of power: (a) photomicrograph with key networks identified, **G** indicates ground; (b) annotated layout:
A, pHEMT in first stage.
B, pHEMT in second stage.
C, coupling capacitors.
D, via to backside metal.
E, via capacitor.
F, RF choke (bias) inductor.
G, RF feedback inductor.
H, airbridge.
I, feedback resistor.
J, bias (TaN) resistor.

probe testing and for wire-bonding. The first transistor has two source connections (at the top and bottom): a gate connection on the left and a drain connection on the right. The layout of the second transistor is the same. The larger second transistor has higher drain current. The source connections of the two transistors are grounded (indicated by the G connection). The matching network is implemented using stubs and capacitors. The second



(a) (b)

Figure 4-33: Details of the top spiral inductor in Figure 4-32(a): (a) layout of the top spiral inductor; and (b) scanning electron microscope image of the cross section of the **airbridge**.

transistor has a feedback network with a spiral inductor and series resistor. Drain bias to the second transistor is through a spiral inductor. Each of the spiral inductors use an airbridge (shown in Figure 4-33) to reduce parasitic capacitance by eliminating dielectric between the bridge metal and that of the spiral.

A second, higher-power X-band MMIC using pHEMT transistors is shown in Figure 4-34. This amplifier produces 1 W over the 8–12 GHz band. There are two stages, but the most striking contrast with the previous MMIC is the use of multiple transistors in each of the stages. The first stage has four transistors in parallel. The input matching network is integrated with a four-way power divider that drives the input of each transistor. An interstage matching network with four two-way power dividers drives the gates of 16 pHEMTS. A 16-way power combiner in the output matching network combines the signals at drains of each second-stage transistor, bringing them to a single output. The output combining network has four levels of two-way combiners and effectively enables transistors to be put in parallel. Putting transistors in parallel requires close matching of the transistors. Practically, the limit to the number of transistors that can be combined this way is 16 as there is loss with each level of combining.

4.10 RFIC Power Amplifiers

One of the distinguishing features of RFIC design is the synthesis of a circuit that intrinsically has the desired attributes. For an RFIC power amplifier the synthesized circuit must produce low levels of distortion while achieving high efficiency. The fundamental distortion mechanism in an RFIC is the near quadratic i-v characteristic of a MOS transistor and the tanh-like transfer characteristic of a MOS amplifier. This section has three examples of analytic and circuit techniques for calculating and managing distortion in MOS circuits.

4.10.1 Distortion in a MOSFET Enhancement-Depletion Amplifier Stage

In this section, power series analysis is applied to the MOS enhancement-depletion amplifier stage shown in Figure 4-35(a). This circuit is a Class A amplifier with a common-source enhancement gain stage and a depletion transistor as an active load. Recall that with a MOSFET, the voltage of the substrate has an important effect on operation of the transistor. With enhancement-mode transistors the substrate or body is typically connected

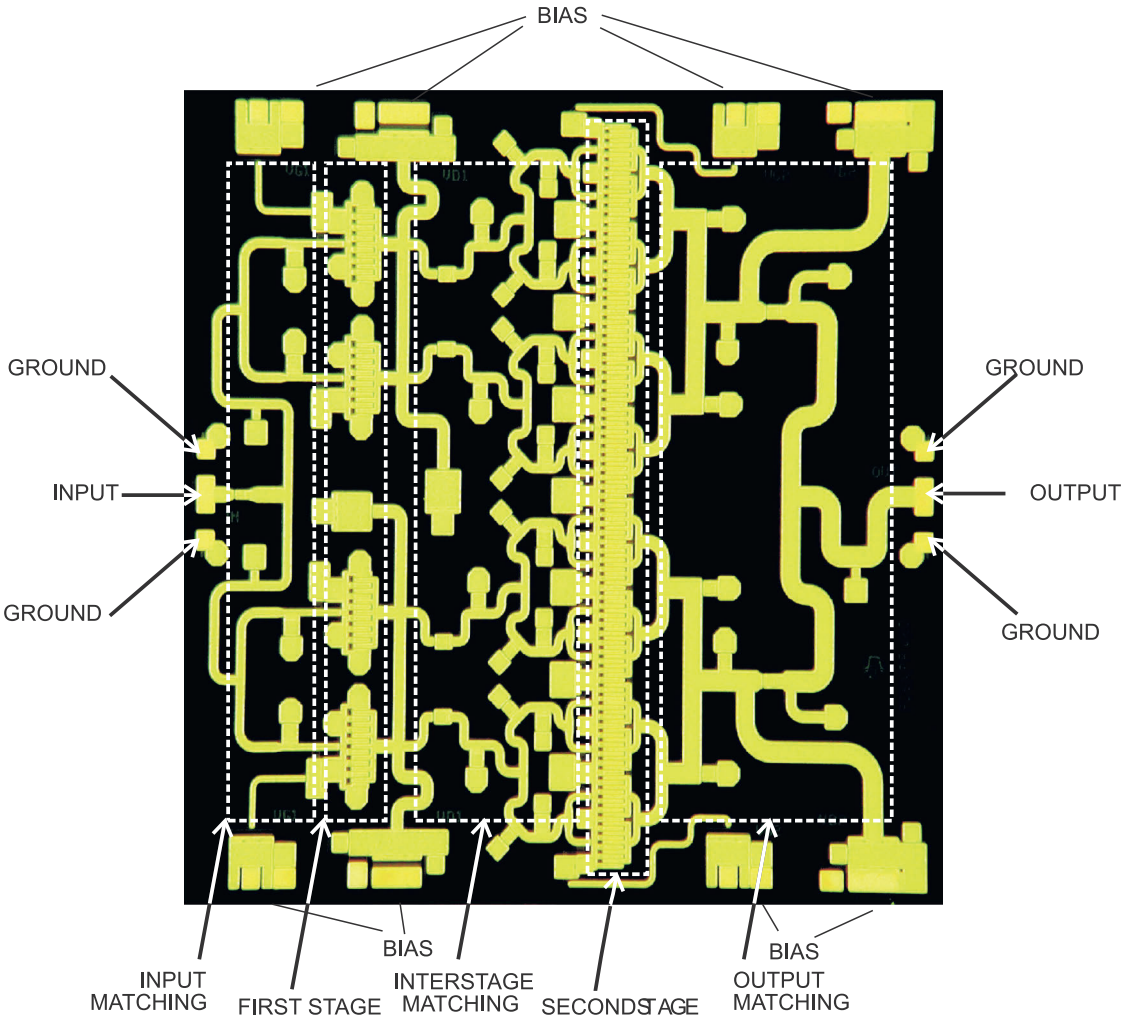


Figure 4-34: An 8–12 GHz MMIC amplifier producing approximately 1 W of output power with key networks identified. (Courtesy Filtronic, PLC, used with permission.)

to the most negative voltage in the circuit, in this case ground. So, since M_1 has the body connected to the source, there is no body effect for M_1. However, for M_2 the body effect must be considered. Using a simple MOSFET quadratic input-output relationship,

$$I_{D1} = \frac{K_1}{2}\left(V_A + V_X - V_{T0E}\right)^2 = I_{D2} = \frac{K_2}{2}\left(-V_{TD}^2\right), \tag{4.28}$$

where the threshold voltage of M_2, with the body effect, is

$$\begin{aligned} V_{TD} &= V_{T0D} + \gamma\left(\sqrt{2\phi_F + V_B + V_Y} - \sqrt{2\phi_Y}\right) \\ &= V_1 + \gamma\left(\sqrt{2\phi_F + V_B + V_Y}\right). \end{aligned} \tag{4.29}$$

(These equations are given in slightly different form in Section 1.A.1. The quantities ϕ_F and ϕ_Y are built-in inversion potentials.) Also K is

proportional to W/L, where W is the width of the transistor's channel and L is its length. The variable $V_1 = V_{T0D} - \gamma\sqrt{2\phi_Y}$ has been introduced, but as will be seen it will be canceled during the derivation.

The aim here is to develop a relationship between the input signal V_X and the output signal V_Y. The first step in developing a simple relationship that can be used in initial design is to relate the operating point voltage levels. Rearranging Equation (4.28),

$$\sqrt{\frac{K_1}{K_2}}\left(V_A + V_X - V_{TOE}\right) = -V_{TD}, \tag{4.30}$$

with the appropriate sign choice made. Combining Equations (4.28) and (4.29) yields

$$\sqrt{\frac{K_1}{K_2}}\left(V_A + V_X - V_{TOE}\right) = -\left(V_1 + \gamma\sqrt{2\phi_F + V_B + V_Y}\right) \tag{4.31}$$

$$\text{and } \sqrt{\frac{K_1}{K_2}}\left(V_X\right) + \sqrt{\frac{K_1}{K_2}}\left(V_A - V_{TOE}\right) = -\left(V_1 + \gamma\sqrt{2\phi_F + V_B + V_Y}\right). \tag{4.32}$$

Now when $V_X = 0 = V_Y$, that is, when there is no AC signal, Equation (4.32) becomes

$$\sqrt{\frac{K_1}{K_2}}\left(V_A - V_{TOE}\right) = -\left(V_1 + \gamma\sqrt{2\phi_F + V_B}\right). \tag{4.33}$$

Substituting Equation (4.33) into Equation (4.32) yields

$$\sqrt{\frac{K_1}{K_2}}V_X = \gamma\sqrt{2\phi_F + V_B} - \gamma\sqrt{2\phi_F + V_B + V_Y}, \tag{4.34}$$

$$\left(\sqrt{\frac{K_1}{K_2}}V_X - \gamma\sqrt{2\phi_F + V_B}\right)^2 = \gamma^2\left(2\phi_F + V_B + V_Y\right), \tag{4.35}$$

$$\frac{K_1}{K_2}V_X^2 - \left(2\gamma\sqrt{\frac{K_1}{K_2}}\sqrt{2\phi_F + V_B}\right)V_X + \gamma^2(2\phi_F + V_B) = \gamma^2 V_Y + \gamma^2(2\phi_F + V_B), \tag{4.36}$$

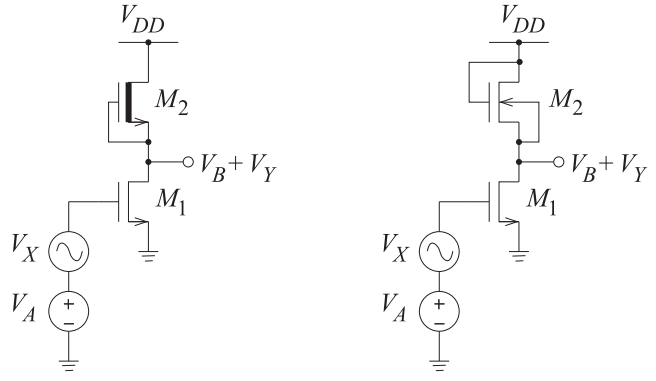


Figure 4-35: MOS amplifier stages.

(a) Enhancement-depletion stage (b) Ultralinear stage

and finally $$\left(\frac{K_1}{K_2}\right) V_X^2 - \left(2\gamma\sqrt{\frac{K_1}{K_2}}\sqrt{2\phi_F + V_B}\right) V_X = \gamma^2 V_Y. \tag{4.37}$$

Therefore the component of the output that differs from the quiescent point is

$$V_Y = \frac{-2}{\gamma}\sqrt{\frac{K_1}{K_2}}\sqrt{2\phi_F + V_B}V_X + \frac{1}{\gamma^2}\frac{K_1}{K_2}V_X^2. \tag{4.38}$$

For a sinusoidal input $V_X = |V_X|\cos(\omega t)$, $V_X^2 = \frac{1}{2}|V_X|^2 + \frac{1}{2}|V_X|^2\cos(2\omega t)$, which contains the second harmonic of the input signal.

So the second-harmonic distortion level, HD_2, the ratio of the second-harmonic component of V_Y to the fundamental component, is

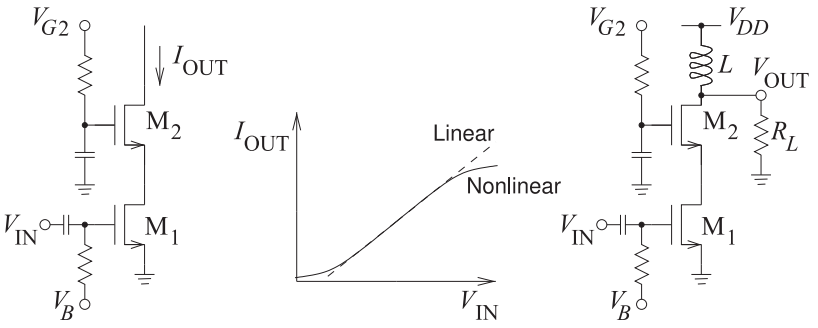
$$\begin{aligned}\mathrm{HD}_2 &= \left(\left|\tfrac{1}{2}\frac{K_1}{K_2}\frac{1}{\gamma^2}V_X^2\right|\right)\left(\left|\frac{-2\sqrt{2\phi_F + V_B}}{\gamma}\sqrt{\frac{K_1}{K_2}}V_X\right|\right)^{-1} \\ &= \frac{1}{4\gamma\sqrt{2\phi_F + V_B}}\sqrt{\frac{K_1}{K_2}}|V_X|. \end{aligned}\tag{4.39}$$

Examination of Equations (4.38) and (4.39) leads to the following conclusions. If there is a lot of voltage gain ($\propto\sqrt{K_1/K_2}$), then the input signal level must be small to keep the harmonic distortion down. Other design considerations that have the same effect are to use a wider device, which increases the transconductance (since $g_m \approx K \propto W/L$) of the transistor so that the drain current will be maintained for the lower input voltage levels. Of course, making the device wider increases capacitive parasitics, which will reduce the maximum operating frequency. Making the channel of the device shorter also increases the transconductance while not affecting the frequency performance. The important point is that there are trade-offs in modifying the performance of the circuit and these are only made apparent using the type of analysis presented here. The use of analysis and the optimization of design through synthesis is one of the cornerstones of RFIC design. In large part this is possible because of the (soft) quadratic-like current-voltage characteristics of MOSFETs.

4.10.2 Distortion in the Ultralinear MOS Connection

The circuit in Figure 4-35(b) is an enhancement-mode gain stage with an enhancement-mode load. Transistor M_2 is the n-channel enhancement load with its well tied to its source. That is, in the fabrication of M_2, a well is formed and the transistor is constructed in the well. The well now serves as the substrate for M_2. Since the well and the source of M_2 are connected together, M_2 is not subject to the body effect. One important property of this circuit is that the characteristics of M_1 and M_2 are matched. Using a similar approach to that used in the previous section, the input/output relationship of the circuit can be developed. Equating drain currents,

$$I_{D1} = \frac{K_1}{2}\left(V_A + V_X - V_{T0E}\right)^2 = I_{D2} = \frac{K_2}{2}\left(V_{DD} - V_B - V_Y - V_{T0E}\right)^2. \tag{4.40}$$



(a) Cascode transconductance cell

(b) Transconductance characteristic

(c) Cascode amplifier

Figure 4-36: A MOS cascode amplifier. L is a choke inductor and is an RF open circuit. V_B and V_{G2} are bias voltages. The cascode transconductance cell is used in this section as a tanh cascode cell (TCC).

With $V_X = V_Y = 0$, $$\sqrt{\frac{K_1}{K_2}}(V_A - V_{T0E}) = V_{DD} - V_B - V_{T0E}. \tag{4.41}$$

Therefore $$\sqrt{\frac{K_1}{K_2}}V_X + (V_{DD} - V_B - V_{T0E}) = (V_{DD} - V_B - V_Y - V_{T0E}), \tag{4.42}$$

and so $$V_Y = -\sqrt{\frac{K_1}{K_2}}V_X. \tag{4.43}$$

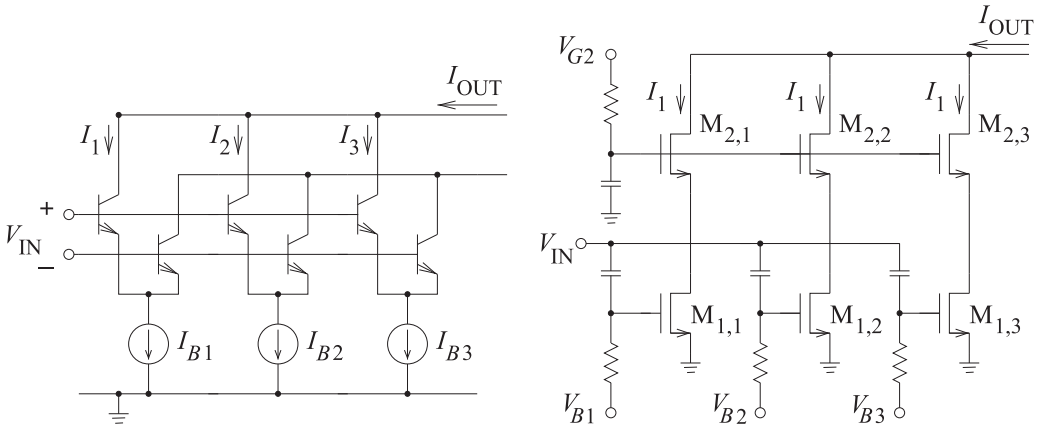
The result is that, provided the transistors are matched, the amplifier is inherently linear. So within the approximations of the drain current expressions there is no distortion, and this includes no third-order intermodulation distortion or spectral regrowth.

4.10.3 RFIC Power Amplifiers with Minimal Distortion

The most widely used linear-by-design strategy for linear RFIC amplifier design is to use what is called the multi-tanh method [66]. The idea is to scale several parallel amplifier stages to obtain an overall linear characteristic. Consider the MOS cascode amplifier shown in Figure 4-36(c), which is based on the transconductance cell shown in Figure 4-36(a). Applying an input voltage, $V_{\rm IN}$, produces an output current, $I_{\rm OUT}$, with the tanh-like characteristic shown in Figure 4-36(b). This characteristic is the dominant cause of amplitude distortion in a FET amplifier. The design strategy for linearizing the output characteristic of a FET amplifier is to put multiple amplifier stages in parallel so that the overall current-voltage characteristic optimally combines the tanh-like characteristic of each stage.

Figure 4-37 shows two circuits that exploit the multi-tanh strategy. Figure 4-37(a) is a three-stage multi-tanh differential amplifier transconductance cell in which the bias currents I_{B1}, I_{B2}, and I_{B3} are adjusted so that the overall input-output characteristic, $I_{\rm OUT}$ versus $V_{\rm IN}$, has greater linearity than that of the individual stages [66]. A similar concept is used to combine the output of multiple cascode stages (see Figure 4-37(b)) and this is a topology more suited to the development of RFIC power amplifiers [67].

As an example of the use of the multi-tanh design approach, consider the multiple tanh cascode amplifier in Figure 4-37(b) with three tanh cascode cells (TCCs). This amplifier has both amplitude and phase distortion. The amplitude distortion largely results from the tanh-like current-voltage



(a) Multi-tanh differential amplifier (b) Three tanh cascode cells (TCC)

Figure 4-37: Combining the output from multiple stages to obtain a highly linear overall transconductance characteristic.

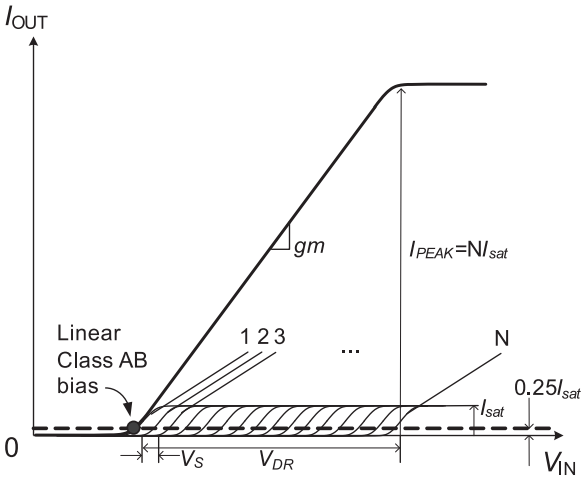


Figure 4-38: Combining the output from N Class AB cascode stages to obtain a highly linear overall transconductance characteristic. Shown are the individual $I_{\rm OUT}$ versus $V_{\rm IN}$ characteristics of each of N transistors each biased differently. The combined $I_{\rm OUT}$ versus $V_{\rm IN}$ characteristic is the sum of the individual transistor's $I_{\rm OUT}$ After [67].

characteristic of the individual cascode stages. Phase distortion largely results from the nonlinearity of the gate capacitance of the FETs. The tanh characteristic of each tanh cascode cell in Figure 4-37(b) is adjusted by scaling the transistors $\mathrm{M}_{1,1}$, $\mathrm{M}_{1,2}$, and $\mathrm{M}_{1,3}$, and changing their bias voltages V_{B1}, V_{B2}, and V_{B3}. The design approach is illustrated in Figure 4-38 and a synthesis approach is presented in references [68] and [67]. Each tanh cascode stage operates in high Class AB mode and with appropriate biasing the tanh characteristics are staggered. Note that for small input signals only one or a few stages are active. The currents from each stage are summed to yield an overall linearized transconductance response.

Phase distortion is reduced by using an analog predistortion circuit that realizes a nonlinear capacitance network that cancels the combined nonlinear gate capacitance of the cascode stages. The complete CMOS RF power amplifier topology is shown in Figure 4-39, where NCN is the nonlinear capacitor network. The die micrograph of this amplifier is shown in Figure

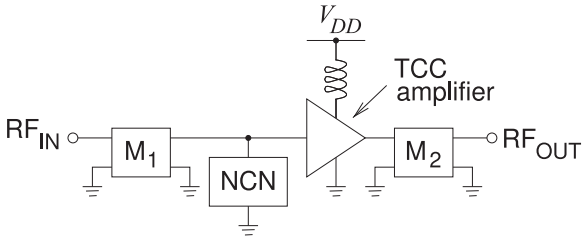
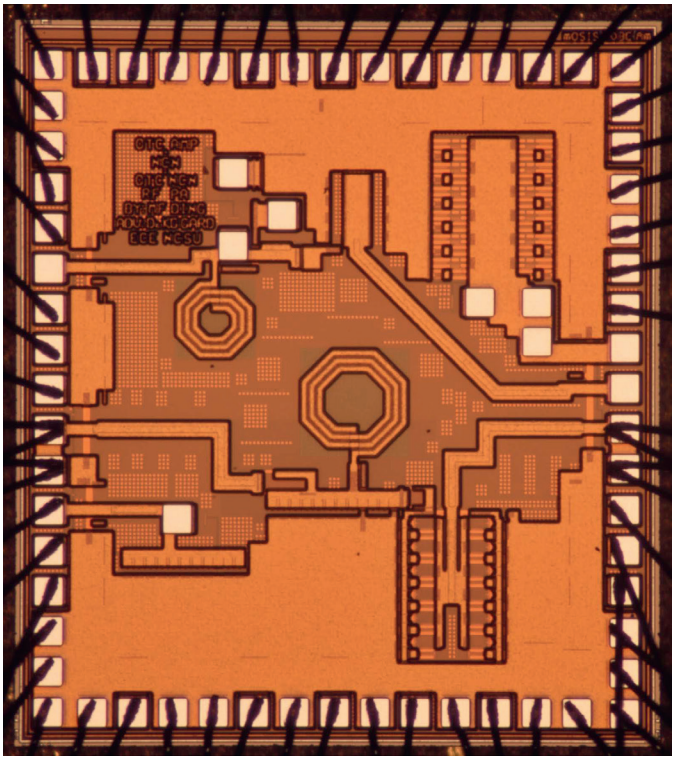
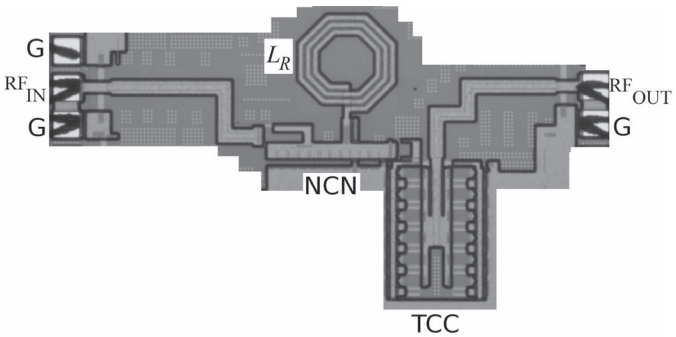


Figure 4-39: Complete RFIC power amplifier with a TCC amplifier with 12 Class AB TCC cells and a nonlinear capacitor network (NCN) and input and output matching networks (M_1 and M_2). After [67].



(a) Die



(b) 2D synthesized amplifier (NCN+TCC)

Figure 4-40: Die micrograph of the 2D synthesized amplifier with tanh cascode cell (TCC) amplifier stage and nonlinear capacitor network (NCN) stage. (There are two operating at different frequencies, one at the top of the die and one at the bottom. The bottom one is being referred to here.) The die also contains a TCC amplifier, an NCN network, and a conventional RF CMOS Class AB amplifier. The supply is 3.6 V and the ground connection is identified by G. The die is shown in (a) and the 2D synthesized amplifier in (b). The inductor L_R resonates out the linear component of the NCN capacitance and the linear component of the gate capacitance of the TCC amplifier. The matching networks and the coke inductor from V_{DD} are off-chip.

4-40(a) together with test circuits. The TCC amplifier with the nonlinear capacitor network is broken out in Figure 4-40(b). The performance of the RFIC power amplifier has excellent performance, as shown in Figure 4-41, and achieves an output power of 25 dBm with an efficiency of 42% and an ACPR of -22 dBc.

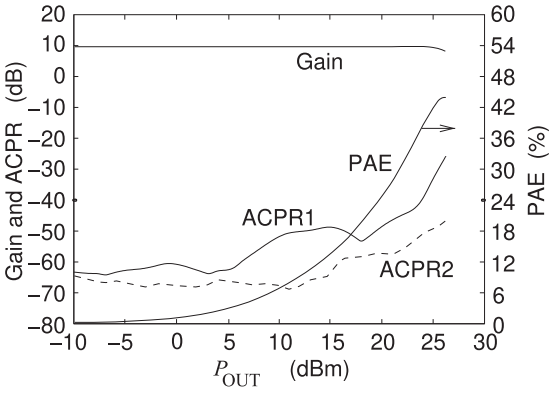


Figure 4-41: Performance of the TCC amplifier with the nonlinear capacitor predistortion circuit at 960 MHz with a WCDMA test signal. A gain of 9.4 dB with power-added efficiency of 41.6% is achieved with an output power of 24.9 dBm and meeting the 3G ACPR1 specifications of -33 dBc and an ACPR2 of -43 dBc. After [67].

4.11 Summary

The realization of efficient power amplifiers is one of the more competitive aspects of RF and microwave design. Power amplifiers must operate at high efficiency and hence operation is strongly nonlinear. Linear amplifier design is only an approximate guide to power amplifier design. Power amplifiers significantly affect the cost and reliability of RF front ends. This is true of base station amplifiers producing hundreds of watts, and equally true of handset power amplifiers producing hundreds of milliwatts. The cost of a power transistor is significant, and so it must operate close to its power output capability. It is important that the power amplifier designer extract every bit of performance from the power transistors, as every tenth of a decibel is significant.

In a basestation the cost of electricity for the RF front end, largely determined by the power amplifier, and by the required air-conditioning can be a significant part of the operating expense of the basestation. With handsets, the efficiency of the RF power amplifier impacts battery life. While it is known how to build amplifiers with high efficiency by combining the concepts presented in this chapter, combining too many concepts can result in a lengthy and costly design effort. The design cost and the possible need to manually tune individual amplifiers can appreciably raise the unit cost of each amplifier. The cost of design must be managed and designers should consider the requirement for individual amplifier tuning. For example, it is not reasonable to tune individual handset power amplifiers but it is for basestation power amplifiers. The trade-off for basestation amplifiers will be different from the trade-off for handset amplifiers, which have unit volumes that can be many orders of magnitude larger. Handset amplifiers are powered directly from the battery supply without voltage regulation. Thus handset power amplifier design must contend with limited supply voltages that can drop as the battery discharges. The limited supply voltage also restricts the choice of transistor technologies.

As with most aspects of RF design, intuition and experience is important in guiding initial topology selection. This is typically followed by computer-aided design and then optimization at the bench. All power amplifiers require manual, at-the-bench, optimization during design, as there are many effects that cannot be fully accounted for in the models used in microwave circuit simulators. Power amplifiers must contend with signals whose average power can vary significantly from packet to packet, and which can have a very large ratio of peak envelope power to mean envelope

power (i.e., large PMEPR). The standard design procedure is to arrive at a successful topology and initial layout using computer-aided design. Then load-pull is used in the laboratory to optimize the loading conditions for actual digitally modulated signals. It is possible that the topology may need to be changed to accommodate necessary changes identified during load-pull.

One of the significant costs and sources of reliability degradation in a handset is using multiple technologies for different parts of the RF front end. This requires heterogeneous integration rather than the more reliable monolithic integration. For example, in a handset it would be desirable to implement a power amplifier in silicon so that it can be integrated with the silicon driver and other circuitry. However, it is currently more feasible to obtain high efficiencies with compound semiconductor devices such as HBT and pHEMT transistors.

Ideally base station amplifiers would use silicon transistors, be able to efficiently amplifier many carriers (with PMEPRs exceeding 20 dB) simultaneously, be able to operate from 500 MHz to 5 GHz, and be reconfigurable for future unforeseen applications. Similarly, cellular handsets must support multiple bands and an important consideration is whether separate power amplifiers are used for each band. Ideally a handset would use a single silicon power amplifier as part of an RFIC that contains the rest of the RF front end, achieve efficiencies of 60% or greater, and operate from supply voltages that vary between 2 and 3 volts.

The RF spectrum is now being exploited beyond 200 GHz and power amplifiers are required from 100s of megahertz to more than 200 gigahertz. Currently the amplifiers with highest output power in this range merge semiconductor-based power amplifiers with vacuum-tube-based amplifier technologies.

Power amplifier design is a trade-off of available technologies that enable design with manageable design and operating costs while achieving requisite powers and efficiencies. There is a large trade-off space and there is a significant opportunity to provide competitive solutions.

4.12 References

[1] P. Aaen, J. Pla, and C. Balanis, "Modeling techniques suitable for CAD-based design of internal matching networks of high-power RF/microwave transistors," *IEEE Trans. on Microwave Theory and Techniques*, vol. 54, no. 7, pp. 3052–3059, Jul. 2006.

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

[3] R. Gilmore and M. Steer, "Nonlinear circuit analysis using the method of harmonic balance—a review of the art: part i, introductory concepts," *Int. J. on Microwave and Millimeter Wave Computer Aided Engineering*, vol. 1, pp. 22–37, Jan. 1991.

[4] ——, "Nonlinear circuit analysis using the method of harmonic balance—a review of the art: part ii, advanced concepts," *Int. Journal on Microwave and Millimeter Wave Computer Aided Engineering*, vol. 1, pp. 159–180, Apr. 1991.

[5] K. Kundert, "Introduction to RF simulation and its application," *IEEE J. of Solid-State Circuits*, vol. 34, no. 9, pp. 1298–1319, Sep. 1999.

[6] J. Sevic, M. Steer, and A. Pavio, "Nonlinear analysis methods for the simulation of digital wireless communication system," *Int. Journal on Microwave and Millimeter Wave Computer Aided Engineering*, pp. 197–216, May 1996.

[7] F. Colon and T. Trick, "Fast periodic steady-state analysis for large-signal electronic circuits," *IEEE J. of Solid-State Circuits*, vol. 8, no. 4, pp. 260–269, Aug. 1973.

[8] T. Aprille and T. Trick, "A computer algorithm to determine the steady-state response of nonlinear oscillators," *IEEE Trans. on Cir-*

cuit Theory, vol. 19, no. 4, pp. 354–360, Jul. 1972.

[9] S. Director and K. Current, "Optimization of forced nonlinear periodic circuits," *IEEE Trans. on Circuits and Systems*, vol. 23, no. 6, pp. 329–335, Jun. 1976.

[10] M. Nakhla and F. Branin, "Determining the periodic response of nonlinear systems by a gradient method," *Circuit Theory Appl.*, vol. 4, no. 3, pp. 255–273, Jul. 1977.

[11] I. Norenkov, Y. Yevstifeyev, and V. Manichev, "A steady-state analysis method for multiperiod electronic circuits," *Radiotekhnika*, no. 11, pp. 86–89, 1987.

[12] J. Aprille, T.J. and T. Trick, "Steady-state analysis of nonlinear circuits with periodic inputs," *Proc. of the IEEE*, vol. 60, no. 1, pp. 108–114, Jan. 1972.

[13] S. Director, "A method for quick determination of the periodic steady-state in nonlinear networks," in *Allerton Conf. Circuit System Theory*, Oct. 1971, pp. 131–139.

[14] H. Chireix, "High power outphasing modulation," *Proc. of the Institute of Radio Engineers*, vol. 23, no. 11, pp. 1370–1392, Nov. 1935.

[15] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Pothecary, J. Sevic, and N. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 814–826, Mar. 2002.

[16] E. Hegazi and A. Abidi, "A 17-mw transmitter and frequency synthesizer for 900-mhz gsm fully integrated in 0.35-μm CMOS," *IEEE J. of Solid-State Circuits*, vol. 38, no. 5, pp. 782–792, May 2003.

[17] N. Sokal and A. Sokal, "Class e-a new class of high-efficiency tuned single-ended switching power amplifiers," *IEEE J. of Solid-State Circuits*, vol. 10, no. 3, pp. 168–176, Jun. 1975.

[18] D. Cox, "Linear amplification with nonlinear components," *IEEE Trans. on Communications*, vol. 22, no. 12, pp. 1942–1945, Dec. 1974.

[19] F. Raab, "Class-F power amplifiers with maximally flat waveforms," *IEEE Trans. on Microwave Theory and Techniques*, vol. 45, no. 11, pp. 2007–2012, Nov. 1997.

[20] M. Eron, B. Kim, F. Raab, R. Caverly, and J. Staudinger, "The head of the class," *IEEE Microwave Magazine*, vol. 12, no. 7, pp. S16–S33, Dec. 2011.

[21] J. Sevic and M. Steer, "On the significance of envelope peak-to-average ratio for estimating the spectral regrowth of an RF/microwave power amplifier," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 6, pp. 1068–1071, Jun. 2000.

[22] S.-A. El-Hamamsy, "Design of high-efficiency RF class-D power amplifier," *IEEE Trans. on Power Electronics*, vol. 9, no. 3, pp. 297–308, May 1994.

[23] H. Kobayashi, J. Hinrichs, and P. Asbeck, "Current-mode class-d power amplifiers for high-efficiency RF applications," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, no. 12, pp. 2480–2485, Dec. 2001.

[24] J.-Y. Kim, D.-H. Han, J.-H. Kim, and S. Stapleton, "A 50 w ldmos current mode 1800 mhz class-d power amplifier," in *2005 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2005, p. 4 pp.

[25] H. Nemati, C. Fager, and H. Zirath, "High efficiency LDMOS current mode class-D power amplifier at 1 ghz," in *36th European Microwave Conf.*, Sep. 2006, pp. 176–179.

[26] D. Schmelzer and S. Long, "A GaN HEMT class F amplifier at 2 GHz with $> 80\%$ PAE," *IEEE J. of Solid-State Circuits*, vol. 42, no. 10, pp. 2130–2136, Oct. 2007.

[27] P. Saad, H. Nemati, M. Thorsell, K. Andersson, and C. Fager, "An inverse class-F GaN HEMT power amplifier with 78% PAE at 3.5 ghz," in *39th European Microwave Conf.*,, 2009, pp. 496–499.

[28] J. H. Kim, G. D. Jo, J. H. Oh, Y. H. Kim, K. C. Lee, and J. H. Jung, "Modeling and design methodology of high-efficiency class-F and class-F^{-1} power amplifiers," *IEEE Trans. on Microwave Theory and Techniques*, vol. 59, no. 1, pp. 153–165, Jan. 2011.

[29] E. Cipriani, P. Colantonio, F. Giannini, and R. Giofre, "Theoretical and experimental comparison of class F vs. class F^{-1} PAs," in *40th European Microwave Conf.*, Sep. 2010, pp. 1670–1673.

[30] J. Sevic and M. Steer, "Analysis of GaAs MESFET spectrum regeneration driven by a $\pi/4$-DQPSK modulated source," in *1995 IEEE MTT-S Int. Microwave Symp. Dig.*, May 1995, pp. 1375–1378.

[31] K. Gard, "Autocorrelation analysis of spectral regrowth generated by nonlinear circuits in wireless communication systems," Ph.D. dissertation, University of California at San Diego, 2003.

[32] H. Gutierrez, K. Gard, and M. Steer, "Nonlinear gain compression in microwave amplifiers using generalized power-series analysis and transformation of input statistics," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 10, pp. 1774–1777, Oct. 2000.

[33] J. Brinkhoff and A. Parker, "Effect of baseband impedance on FET intermodulation," *IEEE Trans. on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 1045–1051, Mar. 2003.

[34] F. Giannini, *Nonlinear Microwave Circuit Design*. Wiley, 2004.

[35] K. Lu, P. McIntosh, C. Snowden, and R. Pollard, "Low-frequency dispersion and its influence on the intermodulation performance of AlGaAs/GaAs HBTs," in *1996 IEEE MTT-S International Microwave Symposium Digest*, 1996, pp. 1373–1376.

[36] J. Sevic, K. Burger, and M. Steer, "A novel envelope-termination load-pull method for ACPR optimization of RF/microwave power amplifiers," in *1998 IEEE MTT-S Int. Microwave Symp. Dig.*, 1998, pp. 723–726.

[37] J. Sevic, M. Steer, and A. Pavio, "Large-signal automated load-pull of adjacent-channel power for digital wireless communication systems," in *1996 IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 2, 1996, pp. 763–766.

[38] J. Sevic, R. Baeten, G. Simpson, and M. Steer, "Automated large-signal load-pull characterization of adjacent-channel power ratio for digital wireless communication systems," in *ARFTG Conf. Dig.-Fall, 46th*, vol. 28, 1995, pp. 64–70.

[39] W. Jang, A. Walker, K. Gard, and M. Steer, "Capturing asymmetrical spectral regrowth in RF systems using a multislice behavioral model and enhanced envelop transient analysis," *Int. Journal of RF and Microwave Computer-Aided Engineering*, vol. 16, no. 4, pp. 400–407, 2006.

[40] A. Walker, M. Steer, and K. Gard, "Capturing asymmetry in distortion of an RF system using a multislice behavioral model," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 4, pp. 212–214, 2006.

[41] J. Hu, K. Gard, N. Carvalho, and M. Steer, "Dynamic time-frequency waveforms for VSA characterization of PA long-term memory effects," in *69th ARFTG Conf., 2007*, 2007, pp. 1–5.

[42] N. Borges de Carvalho and J. Pedro, "A comprehensive explanation of distortion sideband asymmetries," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 9, pp. 2090–2101, Sep. 2002.

[43] J. Pedro and N. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Artech House, 2003.

[44] J. Sevic and M. Steer, "On the significance of envelope peak-to-average ratio for estimating the spectral regrowth of an RF/microwave power amplifier," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 6, pp. 1068–1071, 2000.

[45] S. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, 1999.

[46] P. Kenington, *High-Linearity RF Amplifier Design*. Artech House, 2000.

[47] J. Pothecary, *High-power GaAs FET Amplifiers*. Artech House, 1999.

[48] N. Pothecary, *Feedforward Linear Power Amplifiers*. Artech House, 1999.

[49] J. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Trans. on Vehicular Technology*, vol. 39, no. 4, pp. 374–382, Nov. 1990.

[50] A. D'Andrea, V. Lottici, and R. Reggiannini, "RF power amplifier linearization through amplitude and phase predistortion," *IEEE Trans. on Communications*, vol. 44, no. 11, pp. 1477–1484, Nov. 1996.

[51] C. Campbell, "A fully integrated ku-band doherty amplifier mmic," *IEEE Microwave and Guided Wave Letters*, vol. 9, no. 3, pp. 114–116, Mar. 1999.

[52] M. Iwamoto, A. Williams, P.-F. Chen, A. Metzger, L. Larson, and P. Asbeck, "An extended doherty amplifier with high efficiency over a wide power range," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, no. 12, pp. 2472–2479, Dec. 2001.

[53] W. Doherty, "A new high efficiency power amplifier for modulated waves," *Proc. of the Institute of Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, Sep. 1936.

[54] B. Kim, J. Kim, I. Kim, and J. Cha, "The doherty power amplifier," *IEEE Microwave Magazine*, vol. 7, no. 5, pp. 42–50, Oct. 2006.

[55] F. Raab, "Efficiency of doherty RF power amplifier system," *IEEE Trans. Broadcast.*, vol. 33, no. 3, pp. 77–83, Sep. 1987.

[56] L. Kahn, "Single-sideband transmission by envelope elimination and restoration," *Proc. of the IRE*, vol. 40, no. 7, pp. 803–806, Jul. 1952.

[57] F. Raab and M. Poppe, "Kahn-technique transmitter for L-band communication/radar," in *2010 IEEE Radio and Wireless Symp.*, Jan. 2010, pp. 100–103.

[58] F. Raab, "Intermodulation distortion in Kahn-technique transmitters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 44, no. 12, pp. 2273–2278, Dec. 1996.

[59] ——, "Drive modulation in Kahn-technique transmitters," in *1999 IEEE MTT-S Int. Microwave Symp. Dig.*,, Jun. 1999, pp. 811–814.

[60] B. Stengel and W. Eisenstadt, "LINC power amplifier combiner method efficiency optimization," *IEEE Trans. on Vehicular Technology*, vol. 49, no. 1, pp. 229–234, Jan. 2000.

[61] A. Birafane and A. Kouki, "An analytical approach to LINC power combining efficiency estimation and optimization," in *33rd European Microwave Conf.*, Oct. 2003, pp. 1227–1229.

[62] X. Zhang, L. Larson, and P. Asbeck, *Design of Linear RF Outphasing Power Amplifiers*. Artech House, 2003.
[63] G. Mazzaro, K. Gard, and M. Steer, "Low distortion amplification of multisine signals using a time-frequency technique," in *2009 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2009, pp. 901–904.
[64] ——, "Linear amplification by time-multiplexed spectrum," *IET Circuits, Devices Systems*, vol. 4, no. 5, pp. 392–402, Sep. 2010.
[65] N. Carvalho, K. Remley, D. Schreurs, and K. Gard, "Multisine signals for wireless system test and design [application notes]," *IEEE Microwave Magazine*, vol. 9, no. 3, pp. 122–138, Jun. 2008.
[66] B. Gilbert, "The multi-tanh principle: a tutorial overview," *IEEE J. of Solid-State Circuits*, vol. 33, no. 1, pp. 2–17, Jan. 1998.
[67] M. Ding, K. Gard, and M. Steer, "A highly linear and efficient CMOS RF power amplifier with a new circuit synthesis technique," *IEEE Trans. on Microwave Theory and Techniques*, vol. 60, no. 8, pp. 1–2, Nov. 2012.
[68] D. M. and G. K.G., "Tanh cascode cell amplifier: an arbitrary transfer characteristics amplifier," *Electronics Letters*, vol. 46, pp. 1495–1497, Oct. 2010.

4.13 Exercises

1. Section 4.2.1 presented the harmonic balance equations for the analysis of a nonlinear resistor. DC, the fundamental, and the second harmonic were retained. Derive the harmonic balance equations if only the DC and fundamental tones are retained in the analysis. That is, derive the analysis in Section 4.2.1. You do not need to implement the computer program to perform the analysis.
2. A new digital modulation scheme produces a signal with a PMEPR of 5.6 dB. What is the maximum power-added efficiency of an inductively biased Class A amplifier if the amplifier gain is 40 dB? Explain your reasoning. [Hint: Consider Table 4-3.]
3. A digitally modulated signal has a PMEPR of 10 dB. What is the maximum power-added efficiency of a resistively biased Class A amplifier if the amplifier gain is 20 dB? Explain your reasoning.
4. An amplifier is driven by a digitally modulated signal with a PMEPR of 20 dB. What is the efficiency reduction factor?
5. A Class C amplifier is used to amplify a GSM signal.
   (a) What is the PMEPR of the GSM signal?
   (b) The 1 dB gain compression metric is not relevant when a Class C amplifier amplifies a GSM signal. Explain.
6. An amplifier with a single-tone output 1 dB gain compression power of 26 dBm is used to amplifier a digitally modulated signal with a PMEPR of 6 dB. What is the average power in dBm of the output signal if the 1 dB gain compression level is taken as the maximum acceptable distortion?
7. The RF input to a 10 dB power amplifier is an FM signal.
   (a) What is the PMEPR of the input RF signal?
   (b) What is the maximum power-added efficiency of an inductively biased Class A amplifier if the amplifier gain is 20 dB? Explain your reasoning.
   (c) What is the RF power at the output of the amplifier?
   (d) What is the DC power consumed by the amplifier if the 1 dB gain compression point sets the peak RF output power? That is, the maximum RF output power is when the amplifier is operating at the gain compression point.
8. The RF input to a 20 dB power amplifier is the combination of ten 10 W FM signals.
   (a) What is the PMEPR of one FM signal?
   (b) What is the PMEPR of the combined RF signal? (Consider that the FM signals are narrowband and that they are uncorrelated.)
   (c) What is the gain of the amplifier when it is outputting the largest distorted signal with acceptable distortion? (Assume that the maximum distorted signal is defined by the output power at the 1-dB gain compression.)
   (d) What is the total RF power at the output of the amplifier?
   (e) What is the DC power consumed by the amplifier if the 1 dB gain compression point sets the peak RF output power? That is, the maximum RF output power is when the amplifier is operating at the gain compression point.
   (f) What is the maximum undistorted power-added efficiency of an inductively biased Class A amplifier if the amplifier gain is 20 dB? Explain your reasoning.
9. The input of a 10 dB power amplifier consists of 10 GSM signals.

(a) What is the PMEPR of one GSM signal
(b) What is the PMEPR of the 10-GSM signal? (Consider that the GSM signals are narrowband and that they are uncorrelated.)
(c) What is the maximum power-added efficiency of an inductively biased Class A amplifier? Explain your reasoning.

10. What is the conduction angle of an ideal inductively biased Class A amplifier?

11. What is the conduction angle of an ideal Class E amplifier?

12. What is the conduction angle of an ideal Class F amplifier?

13. A matched amplifier with a $50\ \Omega$ source and a $50\ \Omega$ load has an RF input with peak voltage $V_{\rm IN}$ and input power $P_{\rm IN}$, and an output with peak voltage $V_{\rm OUT}$ and output power $P_{\rm OUT}$. The transfer function of the amplifier (including the input and output matching networks and the active device) is described by $V_{\rm OUT} = 5 \tanh V_{\rm IN}$.
(a) Sketch the transfer characteristic as $V_{\rm OUT}$ versus $V_{\rm IN}$.
(b) Derive an expression for $P_{\rm OUT}$ versus $P_{\rm IN}$.
(c) What is the linear voltage gain of the amplifier?
(d) What is the linear power gain of the amplifier in decibels?
(e) What is the saturated output power of the amplifier in dBm if the RF signal is a single sinusoid?
(f) What is the RF output power in dBm at the 1 dB gain compression point if the RF signal is a single sinusoid?

14. A single-stage amplifier has a linear gain of 16 dB, an output 1 dB gain compression point of 10 dBm, and an OIP3 of 30 dBm.
(a) What is the maximum sinusoidal input signal when the output of the amplifier is compressed by 1 dB?
(b) What is the input-referred third-order intercept point (IIP3)?

15. A single-stage amplifier has a linear gain of 16 dB, an output 1 dB gain compression point of 10 dBm, and an OIP3 of 30 dBm. A communication signal with a PAR of 6 dB is used. What is the maximum average power of the input signal before the output suffers significant compression? This is defined at the point at which the peak signal is compressed by 1 dB.

16. A matched amplifier with a $50\ \Omega$ source and a $25\ \Omega$ load has a sinusoidal RF input with peak voltage $V_{\rm IN}$ and input power $P_{\rm IN}$, and an output with peak voltage $V_{\rm OUT}$ and output power $P_{\rm OUT}$. The transfer function of the amplifier is described by $v_{\rm out}(t) = 5v_{\rm in}(t) - 0.5v_{\rm in}^3(t)$, where $v_{\rm in}(t)$ and $v_{\rm out}(t)$ are the instantaneous values of the input and output voltage. The maximum value of $v_{\rm in}(t)$ is 2 V.
(a) Sketch the instantaneous voltage transfer function of the amplifier. That is, plot $v_{\rm out}$ versus $v_{\rm in}$.
(b) Derive an expression for $V_{\rm OUT}$ versus $V_{\rm IN}$. That is, plot $V_{\rm OUT}$ versus $V_{\rm IN}$.
(c) On your previous sketch, overlay the magnitude voltage transfer function of the amplifier.
(d) Derive an expression for $P_{\rm OUT}$ versus $P_{\rm IN}$.
(e) What is the linear voltage gain of the amplifier?
(f) What is the linear power gain in dB of the amplifier?

17. The RF output of a cell phone has a driver amplifier followed by a power amplifier. The driver amplifier has a linear gain of 30 dB and an output 1 dB compression point of 20 dBm. The power amplifier has a linear gain of 12 dB and an output 1 dB gain compression power of 39 dBm.
(a) What is the linear gain of the driver-power amplifier cascade?
(b) What is the output 1 dB gain compression power in dBm of the cascade?

18. A matched single-stage amplifier in a $50\ \Omega$ system has a linear gain of 16 dB and an output 1 dB gain compression power of 10 dBm. What is the amplitude of the maximum sinusoidal input signal when the gain of the amplifier is compressed by 1 dB?

19. An inductively biased Class A HBT amplifier is biased with a collector-emitter quiescent voltage of 5 V and a quiescent collector-emitter current of 100 mA. When operated at the 1 dB compression point, the input RF power is 20 mW and the output power is 300 mW. Consider that the RF signal is a sinewave, and note that the quiescent collector-emitter voltage will be the supply rail voltage.
(a) What is the quiescent DC power consumed by the transistor? Express your answer in milliwatts?
(b) What is the output power in dBm?
(c) What is the efficiency of the amplifier? Note that the efficiency of an inductively biased Class A amplifier can be more than 50% if distortion can be tolerated.
(d) What is the power-added efficiency of the amplifier?
(e) If the input power is reduced by 10 dB so

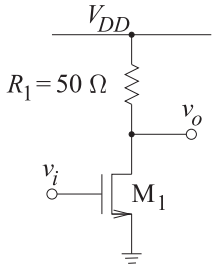
that the amplifier is no longer in compression, will the DC quiescent point change? Explain your answer.

(f) With 1 mW input power, what is the power-added efficiency of the amplifier if the quiescent point does not change? (The small-signal gain of the amplifier is 13 dB.)

20. An amplifier has a gain of 10 dB and an output power of 1 W. The amplifier has a power-added efficiency of 25%.
    (a) What is the total efficiency of the amplifier?
    (b) What is the efficiency of the amplifier?

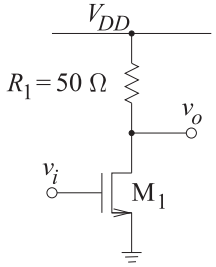
21. An amplifier with a gain of 20 dB and with a single-tone input-referred 1 dB gain compression power of 0 dBm is used to amplify a digitally modulated signal with a PMEPR of 8 dB. What is the average power in dBm of the output signal if the peak RF power is set equal to to the 1 dB gain compression level?

22. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS1} = a_1 V_{GS1} + a_3 V_{GS1}^3$, where $a_1 = 0.01\text{A/V}$ and $a_3 = -0.1\text{A/V}^3$. [You can use interim results from Example 4.1.]



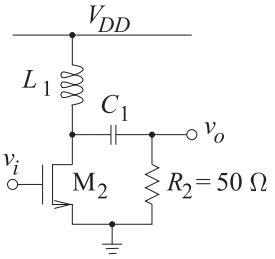
    (a) What is IIP3 in terms of voltage?
    (b) What is OIP3 in terms of voltage?

23. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS1} = 0.02 V_{GS1} - 0.5 V_{GS1}^3$. [You can use interim results from Example 4.1.]



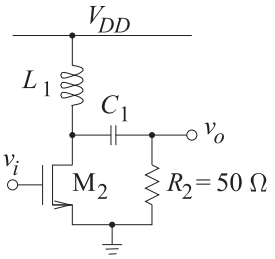
    (a) What is IIP3 in terms of voltage?
    (b) What is OIP3 in terms of voltage?

24. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS2} = b_1 V_{GS1} + b_3 V_{GS2}^3$, where $b_1 = 0.05\text{A/V}$ and $b_3 = -0.2\text{A/V}^3$. [You can use interim results from Example 4.1.]



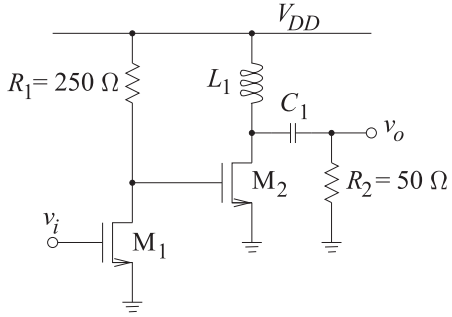
    (a) What is the IIP3 in terms of voltage?
    (b) What is the OIP3 in terms of voltage?

25. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS2} = 0.1 V_{GS1} - 0.4 V_{GS2}^3$. [You can use interim results from Example 4.1.]



    (a) What is the IIP3 in terms of voltage?
    (b) What is the OIP3 in terms of voltage?

26. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equations $i_{DS1} = a_1 V_{GS1} + a_3 V_{GS1}^3$ and $i_{DS2} = b_1 V_{GS1} + b_3 V_{GS2}^3$, where $a_1 = 0.01\text{A/V}$, $a_3 = -0.1\text{A/V}^3$, $b_1 = 0.05\text{A/V}$, and $b_3 = -0.2\text{A/V}^3$. [You can use interim results from Example 4.1.]



    (a) What is the IIP3 in terms of voltage?
    (b) What is the OIP3 in terms of voltage?

4.13.1 Exercises By Section

†challenging, ‡very challenging

§4.2 1†
§4.3 2, 3, 4, 5, 6, 7†, 8†, 9†, 10, 11, 12, 13‡
§4.4 14†, 15†, 16†, 17†, 18†, 19†, 20, 21
§4.5 22†, 23†, 24†, 25†, 26†

4.13.2 Answers to Selected Exercises

2 13.75%
5 0 dB
7(d) 796 mW
8(d) 10 kW
11 180°
13(d) 13.98 dB
16(b) $5V_{\mathrm{IN}} - 0.375V_{\mathrm{IN}}^3$
16(f) 17.0 dB
19(f) 3.8%
24 4.33 V

CHAPTER 5

Oscillators

5.1 Introduction

Most RF oscillators generate sinusoidal signals that are either used to drive mixers or, if modulated, to produce frequency modulated signals directly. In some designs the microwave oscillators drive flip-flop circuits that produce periodic square-wave signals with multiple phases as required for the LO drives of I/Q modulators. RF and microwave oscillators can be designed using either a two-port or a one-port approach. The classic treatment of oscillators is based on a two-port gain device with a feedback loop, but the oscillator can nearly always be viewed, and thus more conveniently designed at microwave frequencies, as a one-port in which a resonant circuit, called the **tank** circuit, is connected to an active circuit that presents a negative resistance. However, the stability, noise, and start-up analyses of an oscillator are based on a two-port with feedback.

All microwave oscillator designs are based on one of three basic oscillator configurations. Suitable configurations must have as few reactive elements as possible while enabling stable single-frequency oscillation. These configurations have three or four reactive element. Mapping a microwave oscillator design on to one of the standard oscillator designs is not simple mainly because active devices at microwave frequencies have substantial parasitic capacitances and also adjustments must be made to accommodate biasing. Stability of a microwave oscillator is of great concern

and the essence is that there should be as few significant energy storage elements as possible. If there are parasitic energy storage elements these should be quite small or absorbed into the capacitors of one of the basic oscillator configurations. With more energy storage elements than necessary the chances of unwanted resonances is much higher and hence instability of an oscillator is more likely.

At microwave frequencies the Q of lumped elements is limited. A lumped inductor in the tank circuit has a particularly low Q and if there is room it is replaced by a transmission line. Oscillators that are fixed in frequency can use a resonant circuit with high Q circuit elements. This contrasts with voltage-controlled oscillators (VCOs) that have a lossy variable element, nearly always a varactor, in the tank circuit in what is now a low Q resonant circuit. The voltage-controlled variation of (invariably) the capacitance of this element changes the frequency of the oscillator.

A microwave oscillator could be realized on-chip or realized as a hybrid circuit with a packaged active device and packaged lumped-elements and possibly transmission lines for the resonant circuit. A hybrid design has much more flexibility than an on-chip design, and if designed correctly has better performance than a monolithically integrated design. Hybrid design techniques are much more mature than chip-based designs but over time some of the techniques used with hybrid designs will migrate to on-chip designs. Also, on-chip designs are preferably differential with transistors in a push-pull, i.e. differential, configuration. Better performance of an on-chip oscillator can be obtained by using an off-chip resonator.

Oscillator theory derives from the analysis of a two-port with gain and feedback. This theory is described in Section 5.2. The following sections explore practical oscillator configurations. Section 5.3 presents the design technique for designing a fixed-frequency microwave oscillator and this is followed up with a design case study in Section 5.4. The distinguishing feature here is that the resonant circuit consists of high Q elements. Section 5.5 presents a design approach for a voltage-controlled microwave oscillator. A state-of-the-art case study of a 5 GHz VCO design is presented in Section 5.6. Section 5.7 describes the design of an on-chip differential oscillator. The last two sections, Sections 5.8 and 5.9 in this chapter describe oscillator phase noise, a characteristic that is the fundamental performance limiting parameter of an oscillator as, for example, it affects the sensitivity and performance of communication systems and the range of radar.

5.2 Oscillator Theory

Microwave oscillators are usually implemented as reflection oscillators with two connected one-port circuits with one being an active device configured as a one-port and presenting a negative conductance, and a second one-port network being the tank or resonator network which must be designed to have specific admittance versus frequency characteristics that ensure stability. However the great body of network theory available derives from control theory and general circuit theory, and is based on a two-port oscillator with linear frequency-selective feedback.

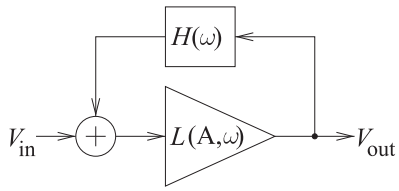


Figure 5-1: Representation of a feedback oscillator based on a two-port active device.

5.2.1 Theory of Oscillation

The two-port view of oscillators is based on the amplifier plus feedback loop shown in Figure 5-1. Here $L(A,\omega)$ describes the characteristics of an amplifier whose transfer function is dependent on both amplitude, A, and radian frequency, ω, and has a magnitude greater than 1. $H(\omega)$ describes the transfer characteristics of a linear feedback network that is dependent only on frequency. The output of the feedback system is described by

$$V_{\text{out}} = L(A,\omega)V_{\text{in}} + L(A,\omega)H(\omega)V_{\text{out}}, \tag{5.1}$$

and so

$$V_{\text{out}} = \frac{L(A,\omega)V_{\text{in}}}{1 - L(A,\omega)H(\omega)}. \tag{5.2}$$

The aim of most oscillator design is to use an active device whose characteristics are independent of frequency but whose transfer response depends on the output amplitude. Then Equation (5.2) becomes

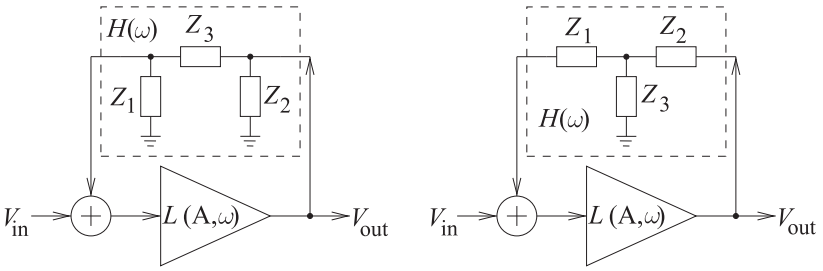
$$V_{\text{out}} = \frac{L(A)V_{\text{in}}}{1 - L(A)H(\omega)}. \tag{5.3}$$

Oscillation begins with input noise when the oscillator is powered on. If the denominator of Equation (5.3) is close to zero, oscillations build up at a frequency determined by the feedback network. As the amplitude of the oscillations builds, $L(A)$ compresses until the denominator is finite but close to zero[1], and there are stable oscillations. Stable oscillation is no accident and is the result of careful design, and much depends on the nature of the feedback network. When an amplifier becomes unstable, for example, the signal produced is often chaotic, with rapid variations in amplitude and frequency, and is not the single frequency sinusoid required of an oscillator.

5.2.2 Basic Oscillator Configurations

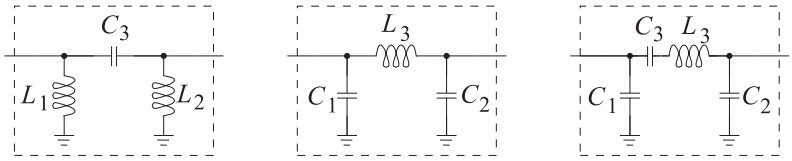
There are two dominant types of feedback networks, the Pi- and T-type networks, shown in Figure 5-2. Three Pi-type networks have proven to be particularly suited to the amplitude saturation characteristics of FET and BJT active devices and, ignoring parasitic capacitances, they result in stable oscillation. These are the Hartley, Colpitts, and Clapp circuits shown in Figure 5-3. The grounds shown can be considered to be common terminals instead.

[1] This description of oscillation, based on Equation (5.3), is behind the erroneous **Barkhausen stability criterion**, which is also known as the **Barkhausen oscillation criterion**. Barkhausen himself used the criterion, known as the **Barkhausen criterion**, to establish the frequency of oscillation as $L(A,\omega)H(\omega) = 1$. This was misinterpreted as a stability or oscillation criterion. It is a necessary criterion for two-port feedback oscillation, but not sufficient. It does not indicate whether a system is unstable. Instead, the Nyquist criterion is the necessary and sufficient criterion for oscillation in feedback oscillators [1–4]. The Barkhausen criterion should not be used in determining whether oscillation occurs.



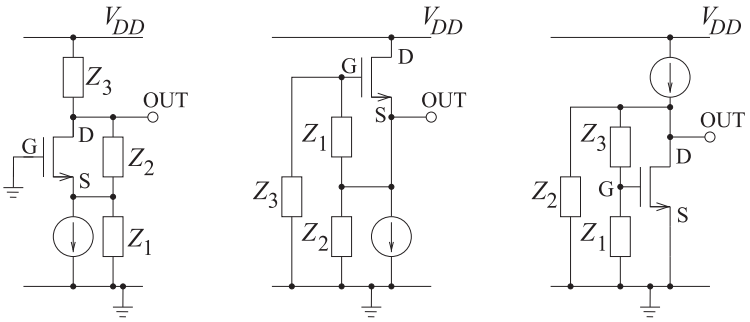
(a) Pi feedback network (b) T feedback network

Figure 5-2: Feedback oscillator with Pi- and T-type feedback networks.



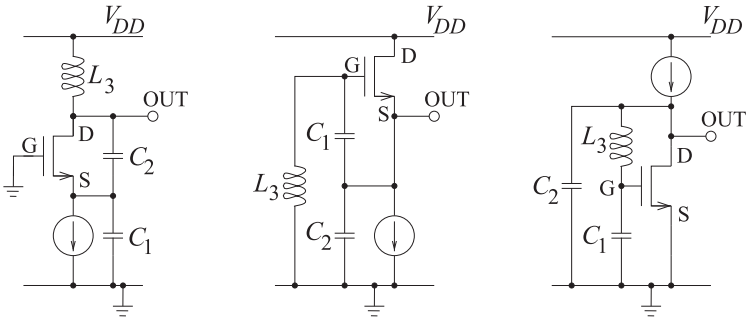
(a) Hartley (b) Colpitts (c) Clapp

Figure 5-3: Basic oscillator feedback networks.



(a) Common gate (b) Common drain (c) Common source

Figure 5-4: Circuit schematics for FET feedback oscillators using Pi-type feedback networks. The current sources provide bias.



(a) Common gate (b) Common drain (c) Common source

Figure 5-5: Circuit schematics for Colpitts FET oscillators. The current sources provide bias. (In (a) C_1 is across the inputof the gate-source transistor with connection through the ground.)

The Hartley and Colpitt's basic oscillator configurations have three large energy storage elements. The Clapp oscillator has four significant energy storage elements but it is important because sometimes it is desirable to use a very high Q resonator such as a piezoelectric crystal which is electrically modeled as a series LC circuit.

Single transistor oscillators using the feedback networks arranged in common-gate, common-source or common-drain configurations are shown in Figure 5-4. FET Colpitts oscillators are shown in Figure 5-5. The Colpitt's configuration is the most common configuration for a microwave oscillator

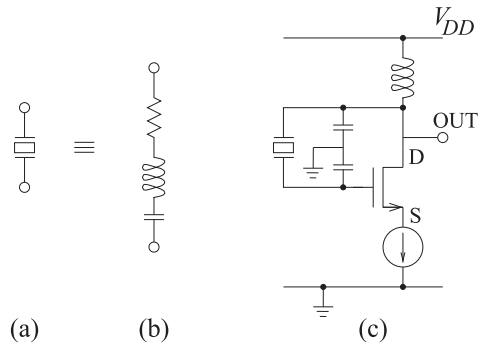


Figure 5-6: Crystal oscillator: (a) schematic symbol for a crystal; (b) its equivalent circuit; and (c) a FET crystal oscillator. The most commomly used form of a crystal oscillator is a Clapp oscillator because of the series LC electrical model of a piezolectric crystal.

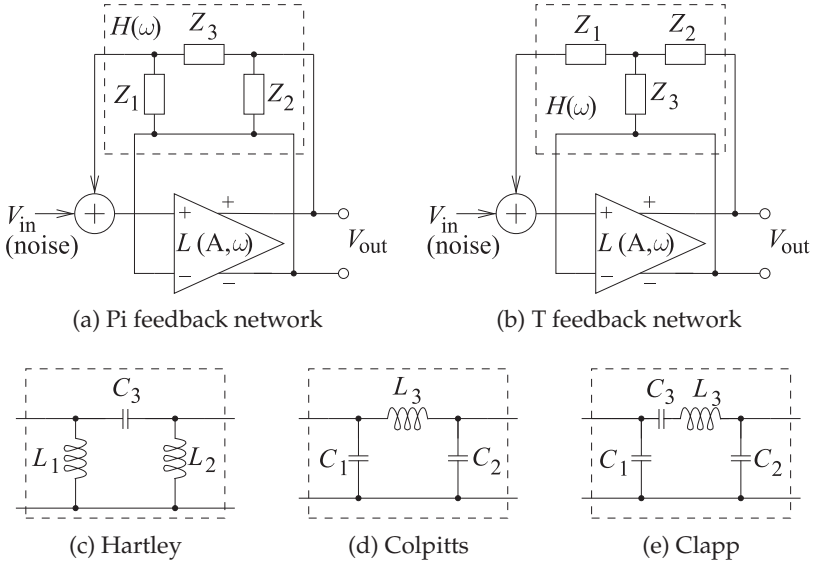


Figure 5-7: Differential feedback oscillator with Pi- and T-type feedback networks.

as it can absorb the parasitic capacitances of a transistor into the C_1 and C_2 capacitors, as in Figures 5-5(a and c). The circuits shown in Figures 5-2–5-5 are the bases of nearly every microwave oscillator. Many modifications lead to better stability, compensate for transistor parasitics, and accommodate differential signaling.

Crystal references, for example, are commonly used for fixed-frequency oscillators, as the piezoelectric (usually quartz) crystals used have a very high Q and the oscillation frequency is very stable, typically to a few parts in a billion, and better if the crystal is temperature stabilized. A circuit with a crystal-based feedback loop creating a Clapp FET oscillator is shown in Figure 5-6. Common quartz reference crystals available as off-the-shelf components are 10 MHz, 20 MHz, and 40 MHz but are available up to 300 MHz. So they are not microwave oscillators but they are used as reference signals in a phase-locked loop to precisely set the frequency of microwave VCOs. Fixed-frequency microwave oscillators do not have this method available for stabilizing the oscillation frequency.

The Hartley, Colpitts, and Clapp networks are also widely used as the bases of oscillators in CMOS RFICs, where the feedback network is

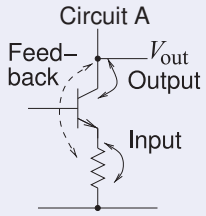
closely approximated using differential circuits [5–10]. The basic differential configurations are shown in Figure 5-7.

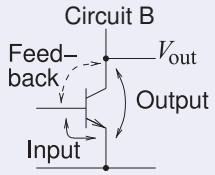
EXAMPLE 5.1 Common Emitter BJT Clapp Oscillator

Draw the schematic of a common-emitter BJT Clapp oscillator without biasing and then draw it with biasing.

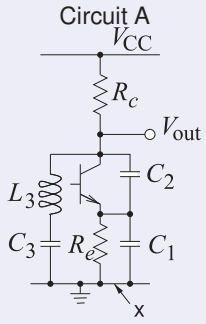
Solution:

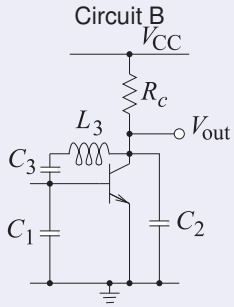
The first step is defining the input and the output connections to be used in the oscillator configuration. There are several ways to do this and the solution here considers two configurations, Circuit A and Circuit B.



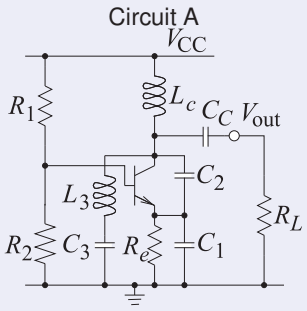


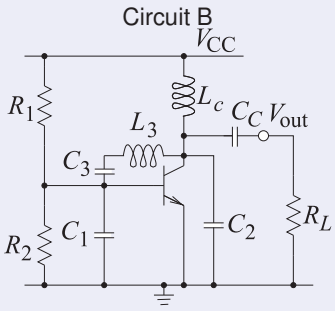
Thus the possible common-emitter Clapp oscillators are (as in Figure 5-4 but with a BJT)





In Circuit A the point x is the input terminal of the oscillator and the emitter terminal is the common point of the basic Clapp oscillator configuration. Appropriate bias must be applied to the base of the transistor. Also it would be good not to deliver power to R_c and instead deliver it to the load. In the circuits below a choke inductor L_c, an open circuit at RF, is used to apply bias. L_c also enables a large output voltage swing. C_c is a large coupling capacitor and has very low impedance at RF and blocks DC.





Of these Circuit B is the more attractive as the parasitic base-emitter capacitance will be absorbed into C_1, and the parasitic collector-emitter capacitance will be absorbed into C_2. Neither configuration can absorb the collector-base capacitance, C_{cb}, so extra care must be made that C_{cb} does not result in instability.

5.3 Reflection Oscillators

Design of stable microwave oscillators traditionally uses the one-port oscillator stability criterion outlined by Kurokawa [11, 12]. Most two-port oscillators can be designed by casting them in the form of a one-port oscillator. In applying the condition, each of the networks—the active device, the resonator load, and the device termination—are characterized as one-ports.

5.3.1 Kurokawa Oscillation Condition

The oscillation condition for a stable reflection oscillator can be expressed in terms of the admittances of the resonator network, $Y_r = G_r + \jmath B_r$, and of the active network, $Y_d = G_d + \jmath B_d$, see Figure 5-8(a). The Kurokawa oscillator condition establishes that for stable (single-frequency) oscillation [11]

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega} - \frac{\partial B_d}{\partial V}\frac{\partial G_r}{\partial \omega}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0, \tag{5.4}$$

where the subscript 0 refers to the operating point, r refers to the resonator, and d refers to the active device. The active device and resonator characteristics shown in Figures 5-8(b and c) satisfy the Kurokawa condition. In Equation (5.4) V_0 is the amplitude of the oscillation at the interface of the active and resonator networks. If the condition in Equation (5.4) is not met, then the oscillator may simultaneously oscillate at multiple frequencies. The Kurokawa oscillation condition must be met at all times which includes during start-up, including when bias is applied, of the oscillator.

With a fixed-frequency oscillator the resonator network is linear so G_r and B_r are independent of amplitude and G_r (by design) is independent of frequency. B_r varies with frequency. Ideally the active device has a frequency independent G_d and amplitude independent B_d Achieving this is a major design task. With these conditions $\partial B_d/\partial V \approx 0$ and $\partial G_r/\partial \omega_0 \approx 0$ and the

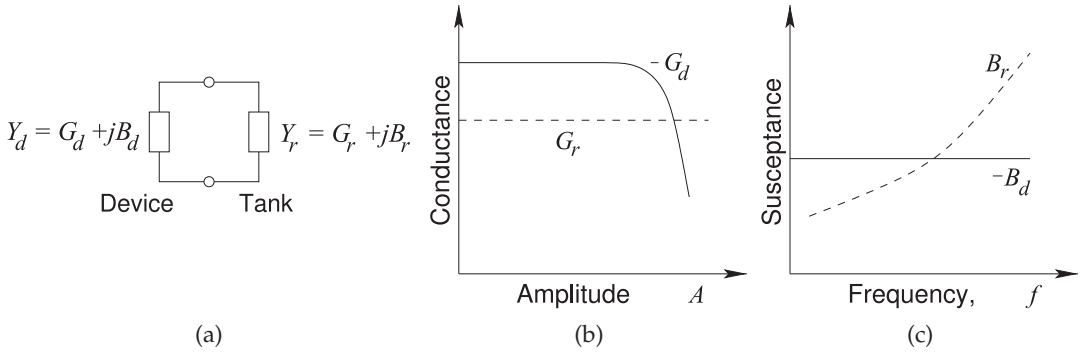


Figure 5-8: Reflection oscillator operation: (a) one-port oscillator; (b) as the amplitude of the oscillation increases, the magnitude of the device conductance, $|G_d|$, decreases while the conductance of the tank circuit, G_r, is constant; and (c) as the frequency of the oscillation increases, the susceptance of the tank circuit, B_r, changes while, B_d (ideally) does not change.

Kurokawa condition in Equation 5.4 simplifies to

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega_0}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0. \tag{5.5}$$

Thus design is greatly simplified for a fixed-frequency oscillator with a high-Q resonator.

5.3.2 Reflection Oscillator Design Approach

When a device with admittance $Y_d = G_d + \jmath B_d$ is connected in shunt to a resonator of admittance $Y_r = G_r + \jmath B_r$ (see Figure 5-8(a)), the voltage amplitude A and radian frequency ω of the resulting equilibrium oscillation are determined when $-G_d(A) = G_r(\omega)$ and $-B_d(A) = B_r(\omega)$. Here the assumption is that the device conductance only is a strong function of voltage amplitude, while the resonator admittance is a function only of angular frequency. This condition can be represented graphically by first denoting the locus of the negative of the device's complex admittance as $-Y_d(A) = -[G_d(A)+\jmath B_d(A)]$ (also referred to as the inverse device reflection coefficient, or $1/\Gamma$ locus (sometimes referred to as the $1/S$ locus)) and the locus of the resonator admittance as $Y_r(\omega) = G_r(\omega)+\jmath B_r(\omega)$. Then, for stable single-frequency oscillation, the intersection of these loci occurs at a single point (i.e., at a single amplitude and frequency combination).

In most oscillator design, the aim is to make the device admittance independent of frequency, and, of course, the admittance of the linear tank circuit is independent of the amplitude of the oscillating signal. These oscillation conditions are depicted in Figure 5-8(b and c). In Figure 5-8(b), as the amplitude of the oscillation increases, the magnitude of the device conductance, $|G_d|$, decreases while the conductance of the tank circuit, G_r, is constant. In Figure 5-8(c), as the frequency of the oscillation increases, the susceptance of the tank circuit, B_r, changes while B_d (ideally) is slow to change. The intersections define the amplitude and frequency of the oscillation.

If the device admittance is dependent on frequency, then it is difficult to avoid multiple intersections of the $-Y_d$ $(1/\Gamma_d)$ and Y_r (Γ_r) loci as viewed in the complex plane. The angle of intersection of the Y_r and $-Y_d$ loci is an important indicator of stability relating to multiple oscillations, oscillator start-up problems, and excess noise [13]. Thus the appropriate angular intersection of these loci is critical. It is difficult to achieve all of the objectives in design unless Y_d is frequency independent.

Ideally, resonator design requires that Q be maximized so that $G_r \approx 0$. This can be achieved with a fixed-frequency microwave oscillator as the resonator can be implemented with a capacitor and a transmission line segment both of which have very low loss. However the tunable capacitors in a VCO design are lossy and so the Q is not high. The two types of oscillators need a different design approach. Furthermore, for a VCO, voltage tuning of the resonator must satisfy the stability criteria, including a single point of intersection and appropriate angle of intersection, over the tuning range. With emphasis on these characteristics and since there are device capacitive parasitics, achieving a proper stable resonator-device interface can be troublesome. An alternative and equally viable approach to stability analysis of a broad class of oscillators, particularly for those using

three-terminal devices, is application of the two-port criteria developed for amplifier stability assessment. However, the one-port design approach is preferred by microwave designers because the one-port connection is closer to the intended operation. The one-port assessment of oscillator stability is not unlike the Bode criteria applied to two-port feedback systems [14, 15]. However, unlike the two-port open-loop assessment of stability, the one-port characterization technique is conveniently aligned with the measurements that can be made by a VNA [16, 17]. As well, the nonlinear limiting effect of the active device is readily measured.

5.3.3 Summary

Microwave oscillator design invariably uses the reflection oscillator approach in which a one-port active device network is connected to a one-port resonator network. Design is complicated enough and it is necessary to simplify design and limit the design space. The procedure almost universally followed is to design for the characteristics shown in Figure 5-8. Design of a fixed-frequency oscillator is further simplified because the conductance of the resonator is almost zero.

5.4 Case Study: Reflection Oscillator

In this case study[2] the design of the reflection oscillator shown in Figure 5-9(a) is examined. This is an 18 GHz common-gate oscillator with series inductive feedback provided by the transmission line TL_1. This circuit is derived from the Colpitts oscillator configuration. The resonator is resonant considerably below the oscillation frequency, and so presents a capacitance to the source of the transistor at the oscillation frequency.

5.4.1 Design Procedure

The circuit in Figure 5-9(a) is a modified common-base Colpitts oscillator and the RF equivalent circuit required to understand oscillator operation, and how the circuit corresponds to the standard Colpitts oscillator, are shown in Figure 5-10. This is the most common fixed-frequency microwave oscillator topology.

The Colpitts feedback network is shown in Figure 5-10(a) where an inductor provides feedback from the output to the input of the active device. Returning to the modified Colpitts oscillator of Figure 5-9, L_1 is a large choke inductor presenting an RF open circuit and C_1 is a large DC blocking capacitor presenting an RF short circuit. The gate transmission line, TL_1, presents a small inductance and most importantly provides inductive feedback from the output to the input of the active device and closely corresponds to L_3 in the Colpitts feedback network of Figure 5-10(a). The gate-source and drain-source parasitic capacitances of the transistor, C_{GS} and C_{DS}, correspond approximately to capacitors C_1 and C_2 in the Colpitts feedback network. This results in the RF equivalent circuit shown in Figure 5-10(b). In Figure 5-10(b) X_R is the reactance of the resonator and compensates for the nonideal nature of the modified Colpitts oscillator. Stable operation

[2] AWR Design Environment Project File: Reflection_Oscillator.emp.

of this oscillator requires that X_R be capacitive but have a variation with frequency substantially less than that of a capacitor. This is accomplished by having the oscillation frequency above the resonant frequency of the resonator.

In Figure 5-9 the reflection oscillator is composed of two networks, the active device network and the resonator network. The third network shown, the OscProbe network, is a probe used to control oscillator analysis using the harmonic balance method. The design approach here is to develop a topology incorporating the active device that presents a negative resistance at the interface, X, between the active device network and the resonator network. Since the active device and the resonator are each best modeled as parallel circuits, it is best to refer to admittance and so the active device network presents a negative conductance to the resonator. In addition, the small-signal admittances of the active network, $Y_d = G_d + jB_d$, and of the resonator network, $Y_r = G_r + jB_r$, are shown in Figure 5-11. Here the active network has a small-signal conductance that is negative and a small-signal

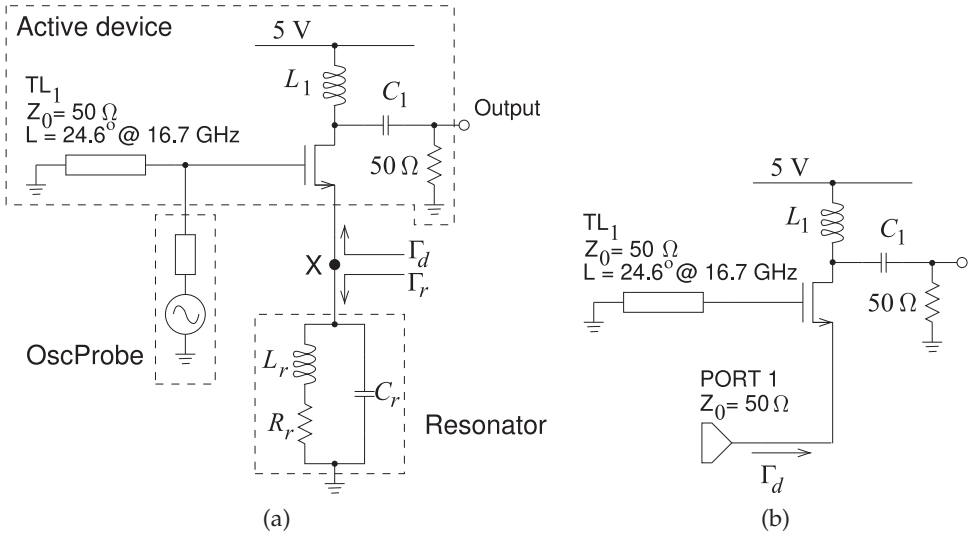


Figure 5-9: Reflection oscillator: (a) complete oscillator circuit used in simulation; and (b) configuration for measuring the large-signal reflection coefficient of the active device. $L_r = 5.6$ nH, $R_r = 10\ \Omega$, $C_r = 445$ fF, $L_1 = 15$ nH, and $C_1 = 10$ pF. The resonant frequency of the resonator is 3.19 GHz, but this is not the oscillation frequency. It presents the required slope of susceptance with respect to frequency at the oscillator frequency of 17.76 GHz.

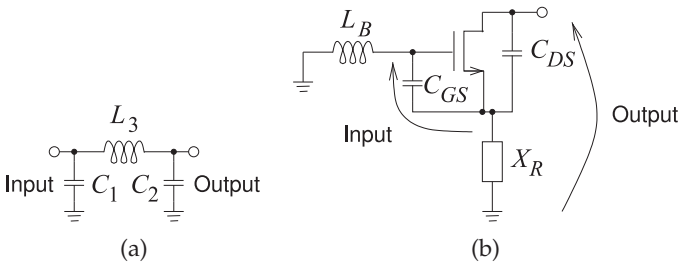


Figure 5-10: Operation of the modified Colpitts oscillator: (a) Colpitts feedback network; (b) the essential RF equivalent circuit of the Modified Colpitts oscillator of Figure 5-9.

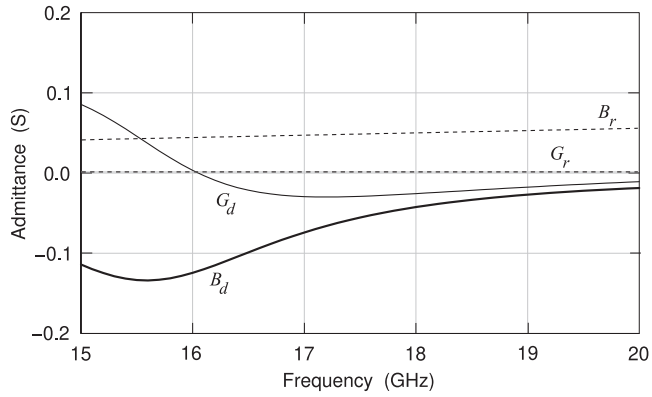


Figure 5-11: Small-signal admittance of the active device, $Y_d = G_d + \jmath B_d$, and admittance of the resonator, $Y_r = G_r + \jmath B_r$. B_d varies with frequency largely because of the frequency-dependent feedback provided by TL_1.

susceptance that is inductive. The resonator has negligible conductance and a capacitive susceptance.

The admittance looking into the source of the active network depends on the level of the signal at X in Figure 5-9. However, the admittance of the resonator is independent of the signal level since the resonator is a linear network. The design strategy is then to develop a feedback network, here the transmission line, TL_1, in the gate of the FET, that then presents a negative conductance to a frequency-selective structure, the resonator network.

The effect of the signal level is examined using the active network test circuit shown in Figure 5-9(b). Here a 50 Ω generator, at Port 1, drives the source of the active device. The power of the signal at the port is varied and it is found that the negative conductance varies from -0.0274 S at a small applied signal, to -0.0224 S at an applied signal level of -10 dBm, to -0.004 S at -7 dBm, and to 0.001 S at -6 dBm. Oscillation will occur when the conductance looking into the source terminal of the active device is approximately zero (since the conductance of the resonator is negligible) and the susceptances of the active network and the resonator network cancel.

Unfortunately the susceptance of the active network changes as the signal level varies since the active device capacitances are nonlinear. The effect of this will now be examined. It is instructive to view the effect of signal level on the oscillation condition by considering the reflection coefficients Γ_d for the active device network and Γ_r for the resonator network. This is shown in Figure 5-12, where a polar plot is used since $|\Gamma_d| > 1$ (except for a very large signal at X). Γ_r is approximately on the unit circle and is capacitive being in the lower half plane. Γ_d is shown for several signal levels, with a signal level of -20 dBm corresponding to the small-signal condition. As the signal level increases eventually to -6 dBm, both the conductance and susceptance of the active device change. The device conductance becomes positive as Γ_d crosses inside the unit circle. Resonance occurs when $\Gamma_d\Gamma_r = 1$. Since $G_r \approx 0$, this is when $\Gamma_d \approx 1/\Gamma_r$. That is, resonance will occur at the frequency where the locus of Γ_d intersects with $1/\Gamma_r$.

The results presented in Figure 5-12 for the reflection coefficient of the active network in Figure 5-9(b) are derived from a large-signal solution calculated using harmonic balance analysis. In most harmonic balance analyses it is only necessary to consider a few harmonics to obtain good results. In the simulation here, five harmonics are considered, with the fundamental frequency set by the frequency of the source at the driven port. The fundamental frequency was stepped from 15 GHz to 20 GHz. Oscillator

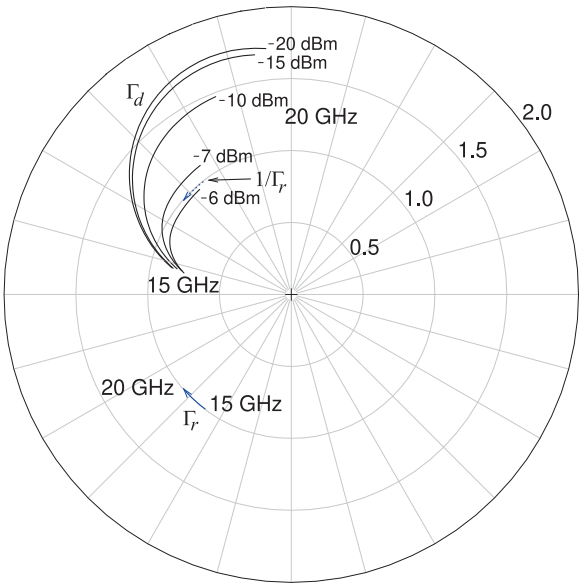


Figure 5-12: Reflection coefficient of the resonator, Γ_r, and of the active device, Γ_d, at various signal levels. Γ is plotted on a polar plot with the outer circle corresponding to $|\Gamma| = 2$. As desired, the locus of $1/\Gamma_r$ is parallel to Γ_d at a fixed signal level (here about -6 dB), and the direction of $1/\Gamma_r$ with increasing frequency is opposite that of Γ_d. The close parallel match is ensured by adjusting the susceptance-versus-frequency slope of the resonator. Considering the intersection of $1/\Gamma_r$ and Γ_d is equivalent to considering the intersection of $1/\Gamma_d$ and Γ_r.

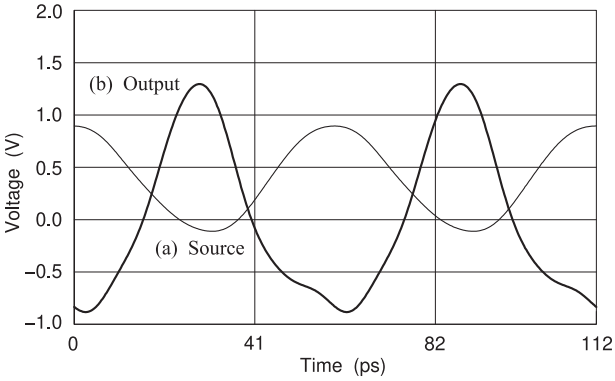


Figure 5-13: Voltage waveforms at the output of the oscillator (Curve (b)), and at the source terminal of the active device, Terminal x, which is the interface between the active device and the resonator (Curve (a)).

simulation also uses harmonic balance analysis, but now the frequency of oscillation is not known ahead of time. It is necessary to introduce another condition to enable the simulator to find the oscillation frequency. One of the techniques used in harmonic balance simulators is the introduction of an oscillator probe element such as the OscProbe element shown in Figure 5-9(a). The frequency of the source, F_{OSC}, in the OscProbe element is initially guessed by the simulator and the series impedance of the OscProbe element is a short circuit at the oscillation frequency and an open circuit at the harmonics. This extra condition is included in the harmonic balance equations and F_{OSC} is allowed to vary as well as the amplitude of the source. The oscillation solution is obtained when the current through the impedance in the OscProbe element is zero. The oscillation frequency is found to be 17.76 GHz. Note that this is not the resonant frequency of the resonator, as here the active device presents an inductance to the resonator. The resonator must present an effective capacitance that has the required susceptance-versus-frequency slope so that the frequency locus of $1/\Gamma_r(f)$ is parallel but oppositely directed to that of $\Gamma_d(f)$ (see Figure 5-12).

Figure 5-13 shows the waveforms at the output of the oscillator (Curve (b)) and at the interface (X) between the resonator and active device networks

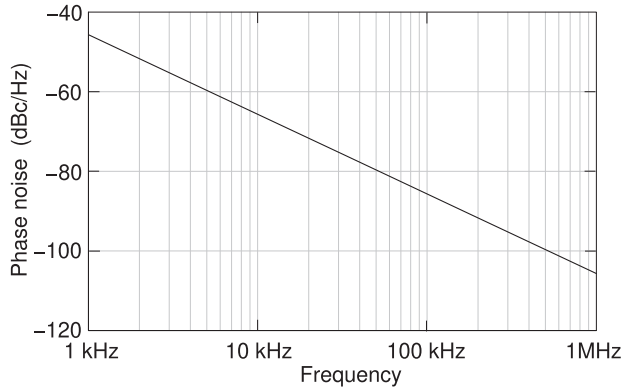


Figure 5-14: Phase noise of the reflection oscillator.

(Curve (a)). The output of the oscillator would be followed by a bandpass filter so that a single harmonic-free sinewave is presented to the external output terminal.

The phase noise of the oscillator is calculated from the oscillator steady-state conditions and is shown in Figure 5-14. The phase noise level is relatively high and this is principally due to the resistance R_r in the resonator, which models the loss in the lumped inductor. Performance would be improved if, instead, a distributed circuit (with transmission lines) was used to realize the required admittance and its derivative at the oscillation frequency.

5.4.2 Summary

The zero combined suspectance point in Figure 5-11 was at 17.76 GHz while the oscillation frequency shown in Figure 5-13 is at 17.29 GHz.The reason for this discrepancy is that the susceptance plot in Figure 5-11 is based on small-signal conditions whereas the simulation is for a large signal and includes harmonics. It can be expected that the oscillation frequency calculated using harmonic balance analysis will be different from that obtained from a small-signal analysis. Consider the Γ_d curves in Figure 5-12. (Generating this data took many simulation runs and did not consider harmonics.) The magnitude of Γ_d reduces as the applied signal level increases. This rotation is not desired and means that the susceptance of the active device is amplitude-dependent. This corresponds to the magnitude of the negative conductance of the active device getting smaller. Note that there is also a small rotation of Γ_d indicating that the susceptance of the active device also changes as the signal level increases. This means that the oscillation frequency will depend on the amplitude of the oscillation. In addition, the applied signal used to calculate the Γ_d at different power levels is not the actual oscillating signal.

5.5 Voltage-Controlled Oscillator (VCO)

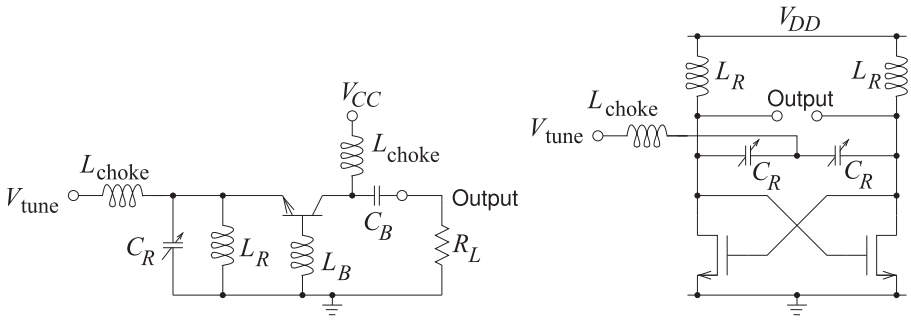
This section describes the special challenges in designing a microwave VCO. The main challenge derives from the resonator network now being lossy because of the use of a tuning element that unfortunately has low Q. In nearly all cases the tuning element is a varactor, a reverse-biased pn-junction or Schottky diode, which has substantial series resistance at microwave frequencies. This means that the full Kurokawa conditions for stable oscillation must be used and also these conditions must be met across the tuning range of the oscillator and during start-up.

5.5.1 Design Procedure

A broadband VCO is a key element in a communication system and is commonly used to drive a mixer to enable different frequency bands to be selected simply by changing the oscillation frequency. The design of a broadband VCO is relatively complex, with issues of simultaneous oscillation at multiple frequencies, phase noise, and power efficiency being of primary concern. In a battery-powered communication device the power drawn by the VCO is a substantial fraction of the total power consumed by the RF front end.

The design of a VCO builds on the standard oscillator design procedure which matches the inverse reflection coefficient of the active device, $1/\Gamma_d$, to the reflection coefficient of a tank or resonator circuit, Γ_r, so that $\Gamma_r = 1/\Gamma_d$ only at oscillation. This is complicated at microwave frequencies as parasitic capacitances are significant and special design strategies are required to mitigate their effect. The required stable oscillation conditions must be achieved over the entire tuning range while also maintaining constant output power and low phase noise.

Two VCO circuits are shown in Figure 5-15. The circuit in Figure 5-15(a) is a common-base oscillator with the feedback provided by the base inductor, L_B, presenting a negative conductance and a small shunt inductance at the emitter. The resonant circuit comprising C_R and L_R presents either a capacitance or an inductance at the oscillation frequency, and cancels the susceptance presented by the active device. The resonant circuit has a non-negligible conductance due to loss in the resonator. Ideally the active device's susceptance is independent of amplitude, however, the transistor's parasitics complicate the situation so that this susceptance also has amplitude dependence. The resonator is tunable because of the variable



(a) Common-base BJT VCO

(b) Differential VCO

Figure 5-15: VCO with tunable resonator comprising C_R and L_R. L_{choke} is an RF open circuit and C_B is an RF short circuit.

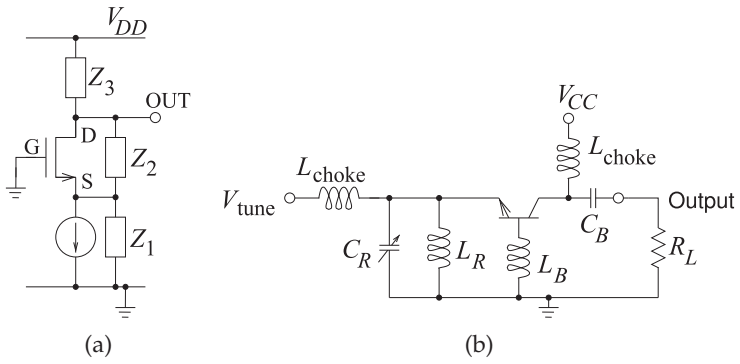


Figure 5-16: Common-gate/base feedback oscillator using a Pi-type feedback networks: (a) common-gate FET oscillator; and (b) common-base BJT oscillator.

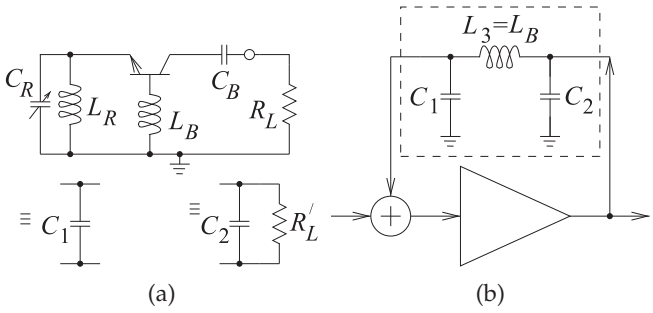


Figure 5-17: Transformation of the oscillator of Figure 5-16 into the form of a standard Colpitts feedback oscillator: (a) first stage of transformation replacing the L_{choke}s with RF open circuits, replacing C_B with an RF short circuit; and (b) final feedback form (compare with the circuits in Figures 5-2(a) and 5-3(b)). C_1 combines the inductance of the L_r-C_r resonator and the capacitive susceptance introduced by the device parasitics. C_2 derives mostly from the collector-emitter capacitance.

capacitance, usually a varactor.

The most common microwave single-transistor reflection oscillator topology is the common-base Colpitts BJT/HBT oscillator, or the similar common-gate FET oscillator, as shown in Figure 5-16. Consider the BJT oscillator in Figure 5-16(b). This is called a common-base configuration, although it is only approximate as will be seen when the RF signal flow path is traced. The parallel L_R-C_R combination is called a resonator, but it is resonant below (or above) the oscillation frequency, so at the oscillation frequency it is effectively a capacitor (or an inductor) but with an admittance-versus-frequency slope magnitude that is less than that of a lumped capacitor (or inductor).[3] Design of the susceptance slope is important to obtaining successful operation as a VCO. Thus one of the purposes of L_R is to enable $\partial B_r/\partial\omega$ to be selected in design. The base inductance L_B couples the output of the active device at the collector to the input at the emitter. So this circuit is a common-base Colpitt's oscillator (the equivalence is illustrated in Figure 5-17).

Parasitic and design elements (especially L_B and compensating capacitors) in the active network may require that the resonant network present an inductance. As long as the Kurokawa condition is met oscillation will be stable. VCO design requires considerable flexibility.

Figure 5-15(b) is the schematic of a differential oscillator realized using

[3] The base inductor L_B can be adjusted so that either a capacitive or inductive susceptance is presented to the resonator. As will be seen stable oscillation can be achieved either way.

MOSFETs, where the resonator comprises a pair of variable capacitances, the C_Rs, and a pair of inductors, the L_Rs. The chokes, the L_{choke}s, are short circuits at DC but block RF currents. On-chip design has many restrictions and it cannot be expected that the same performance can be obtained as with a hybrid design.

5.5.2 Managing Multioscillation and Phase Noise

The oscillation condition for a stable reflection oscillator needs to be revisited and discussed with respect to a VCO. Referring to the reflection oscillator in Figure 5-8, the Kurokawa oscillator condition for stable (single-frequency) oscillation is [11]

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega_0} - \frac{\partial B_d}{\partial V}\frac{\partial G_r}{\partial \omega_0}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0, \tag{5.6}$$

where the the subscript 0 refers to the operating point and V_0 is the amplitude of the oscillation at the interface of the active and resonator networks. The active network includes elements added to the active device mainly to compensate for parasitics. If the condition in Equation (5.6) is not met, then the oscillator may simultaneously oscillate at multiple frequencies or be chaotic.

Achieving single-frequency oscillation can be a challenge for fixed-frequency oscillator design but is especially challenging for a VCO as the condition must be met across the tuning range. Complicating design further is that with a VCO the resonator network is lossy so that the Kurokawa condition is not simplified as it is for a fixed-frequency oscillator, see Section 5.3.1. The condition is simplified then because of of the high Q resonator of a fixed-frequency oscillator where terms in Kurokawa's oscillation condition disappear because $G_r \approx 0$.

In VCO design the design procedure must be kept simple, and this opens up the design space to enable optimization of other characteristics such as minimizing phase noise and DC power consumption. One aspect of the preferred oscillator design procedure is to choose a topology that results in an effective device susceptance that as much as possible is independent of signal amplitude (i.e., $\partial B_d/\partial V|_{V=V_0} \approx 0$) and a loaded resonator conductance that is independent of frequency (i.e., $\partial G_r/\partial \omega_0|_{\omega=\omega_0} \approx 0$). Then the criterion for stable oscillation is the much simpler

$$\left(\frac{\partial G_d}{\partial V}\right)\left(\frac{\partial B_r}{\partial \omega_0}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0, \tag{5.7}$$

something that is much easier to satisfy in design. This equation is the same as the simplified criteria developed for the fixed-frequency oscillator but now there are two additional design requirements, i.e. $\partial B_d/\partial V|_{V=V_0} \approx 0$ and $\partial G_r/\partial \omega|_{\omega=\omega_0} \approx 0$.

Referring to the common-base circuit of Figure 5-17(a), the element in the base, L_B, induces a negative conductance at the interface with the L_R-C_R resonator [18]. It also induces an inductive susceptance and with the active device capacitance and other capacitors the total susceptance presented by the active network is either inductive or capacitive. A feature of the circuit in Figure 5-17(a) is that the L_R-C_R resonator is isolated from the load

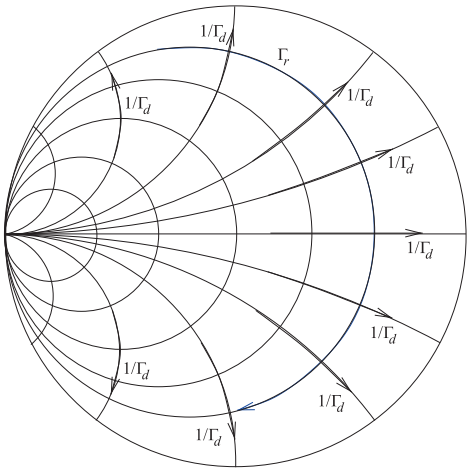


Figure 5-18: Requirements on Γ_r and Γ_d for stable oscillation as derived from Equation (5.7) with assumptions behind the derivation of the equation (i.e., $\partial B_d/\partial V|_{V=V_0} \approx 0$ and $\partial G_r/\partial \omega_0|_{\omega=\omega_0} \approx 0$). The locus of Γ_r is for increasing frequency. Several loci are shown for $1/\Gamma_d$ are shown at different frequencies. The arrowed loci of $1/\Gamma_d$ are for increasing frequency.

and so the load has little effect on the oscillation frequency. The output of the transistor is modeled as a shunt connection of a current source, a conductance, and a susceptance. Thus it is natural to treat the oscillator as a negative conductance oscillator. It is found that, as desired, this topology leads to a negative conductance that reduces in magnitude as the oscillation level increases. Another way of stating the same issue is saying that the L_r-C_r resonator network does not contain all of the reactance that should be assigned to C_1 in the simple Colpitts oscillator. Now $\partial G_d/\partial V|_{V=V_0}$ in Equation (5.7) is positive. (That is, G_d becomes less negative as the magnitude of the oscillating signal increases.) Thus for stable oscillation, $\partial B_r/\partial \omega|_{\omega=\omega_0}$ must be positive. The conditions for stable oscillation, using the design simplifications that led to Equation (5.7), are then as shown in Figure 5-18. What this figure indicates is that if Γ_d is capacitive ($1/\Gamma_d$ is in the top half plane) then Γ_r needs to be inductive, while if Γ_d is inductive ($1/\Gamma_d$ is in the bottom half plane) then Γ_r needs to be capacitive.

At microwave frequencies, parasitic reactances of a transistor are significant and so the active device does not have a signal-level independent susceptance (i.e., $\partial B_d/\partial V|_{V=V_0}$ is finite). In part this is because the parasitic capacitances of the transistor are voltage dependent. It is also due to the reactive feedback element used to produce the negative conductance. Parasitic elements further isolate the negative output conductance of a transistor from the port of the active network that interfaces with the resonator. As a result it may be necessary to introduce compensatory elements to translate the behavior of the active device into the correct form. This will be presented in the case study of the next section. The bottom line is that designing for $\partial B_d/\partial V|_{V=V_0} \approx 0$ is a major challenge. It is a less significant a problem to design the resonator to have the required conductance property (i.e., $\partial G_r/\partial \omega|_{\omega=\omega_0} \approx 0$).

5.5.3 Negative Resistance Oscillator

The above discussion concerns a negative conductance oscillator in which the magnitude of the device's negative conductance reduces with signal amplitude. Alternatively, a negative resistance oscillator could be realized using a capacitor in series with the emitter. This leads to an oscillator model in which the negative resistance is in series with a load resistance and

a series resonant circuit. Stable oscillation of this oscillator requires that the magnitude of the negative resistance reduce as the oscillation signal increases in size. However, it is difficult to realize a negative resistance that reduces in magnitude as the signal level increases and simultaneously achieve a reactance that is unchanged by the signal level. Thus it is better to use feedback to create a negative conductance shunting a susceptance. Design of a stable negative resistance oscillator is quite difficult and is nearly always avoided.

5.5.4 Summary

With the low Q resonator of a VCO the design procedure is quite different to that for a fixed-frequency oscillator which has a high Q resonator. The central reflection-based VCO design concept is the development of an active network presenting a near-voltage-independent susceptance (i.e., $\partial B_d/\partial V|_{V=V_o} \approx 0$) and a frequency-independent negative conductance ($\partial G_d/\partial \omega|_{\omega=\omega_0} \approx 0$). Once this is achieved, the conventional reflection oscillator design approach can be used and the resonator designed so that $\partial G_r/\partial \omega|_{\omega=\omega_0} \approx 0$. Design is both an art and a science. Sometimes the problem must be simplified for the designer to be able to conceptualize and synthesize the required circuit. This is usually the best approach to achieving microwave circuits that have near optimum performance. Simply optimizing a given topology is not enough. The correct topology must be synthesized first.

The requirements for oscillation start-up have not been considered yet, and these will be addressed in the following VCO design case study.

5.6 Case Study: Design of a C-Band VCO

This section presents the design of a high-performance microwave VCO operating from 4.5 to 5.3 GHz reported in [19]. The design objective is the generation of a frequency-independent negative conductance, $G_d(A)$, with a prescribed reflection coefficient shape, Γ_d, using a three-terminal active device in a common-base configuration with series-inductive feedback. The oscillator schematic is shown in Figure 5-19(a). This is a common-base oscillator in which the 2.2 nH base inductor L_3 provides negative feedback between the input and output of the transistor and the resonator presents an inductance at the oscillation frequency. So this circuit is a common-base Colpitts oscillator; the equivalence is illustrated in Figure 5-20.

5.6.1 Design Philosophy and Topology

Referring to Figure 5-19(a), the oscillator is partitioned into an active network to the right of the line x-x and a resonator network to the left of the dividing line. The negative conductance presented to the resonator is principally because of the feedback provided by the base inductor L_3. The choke, $L_{\text{CHOKE}} = 8.2$ nH, and associated elements provide DC bias to the transistor. The choke has an impedance magnitude of $258\ \Omega$ at 5 GHz and is effectively an RF open circuit. The output of the oscillator is taken from the collector of the transistor through a 2.2 pF capacitor that drives a $50\ \Omega$ load, Z_L. The emitter is connected to the resonator network through the parallel $47.5\ \Omega$ resistor and 10 pF capacitor. The 10 pF capacitors have an impedance

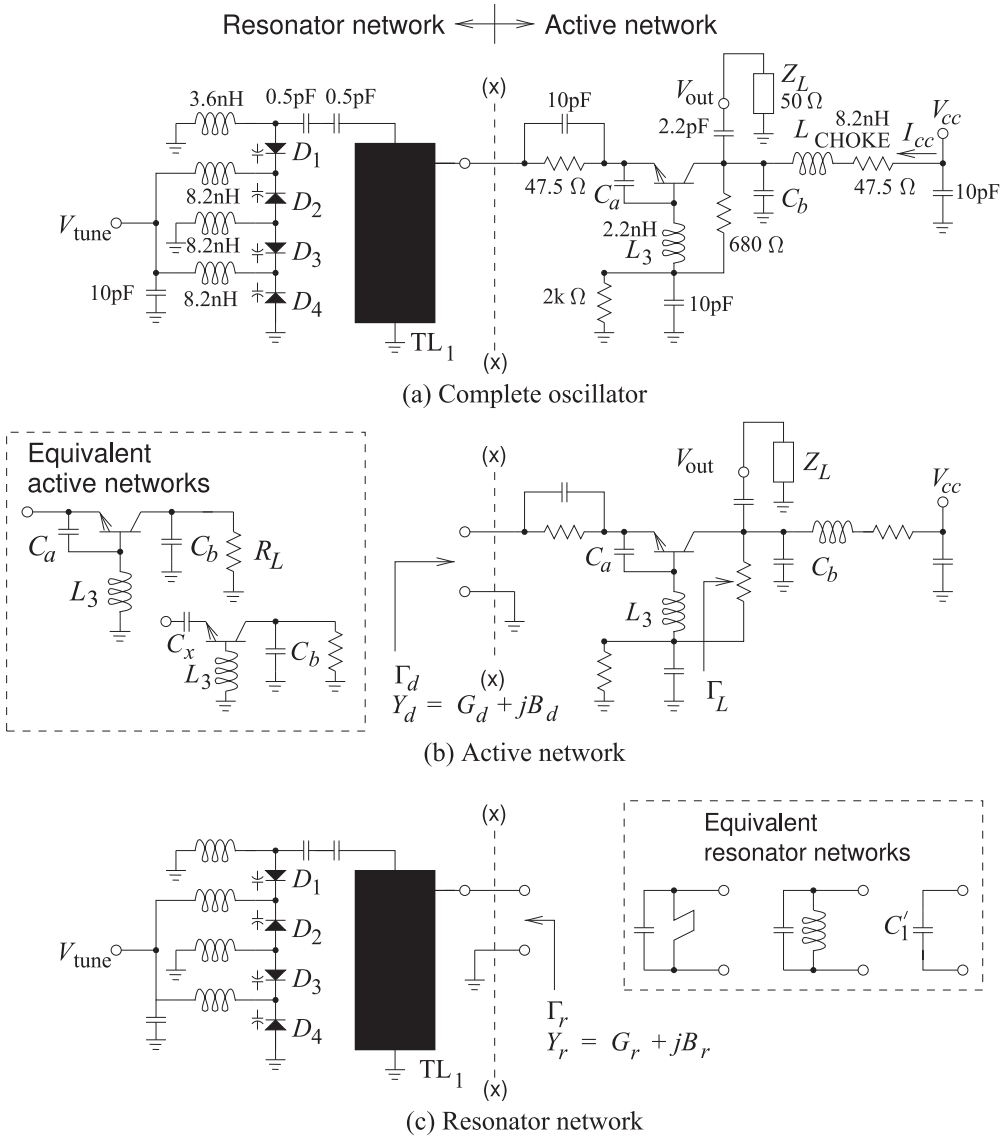


Figure 5-19: A 5 GHz common-base SiGe BJT oscillator: (a) oscillator showing the interface (x-x) between the resonator network (also called a tank circuit) and the device; (b) active network; and (c) resonator network. The element labeled TL_1 is a low impedance microstrip line. Capacitors C_a and C_b compensate for the frequency-dependent feedback provided by the base inductance. The choke inductor, $L_{CHOKE} = 8.2$ nH, presents an RF open circuit and is part of the bias circuit. $V_{CC} = 30$ V and $I_{CC} = 30$ mA. Each varactor diode (D_1–D_4) is model JDS2S71E. The transistor is a Si BJT model NE894M13, which is designed for oscillator applications above 3 GHz.

magnitude of 3 Ω at 5 GHz and are effectively RF short circuits.

The base inductor, L_3, provides feedback that results in negative conductance from the emitter to ground. Since the feedback is frequency dependent, this induced negative conductance is also frequency dependent.

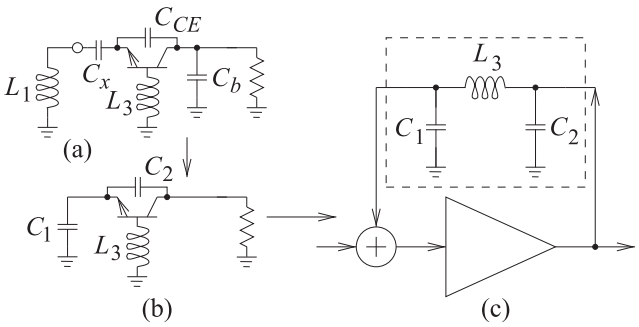


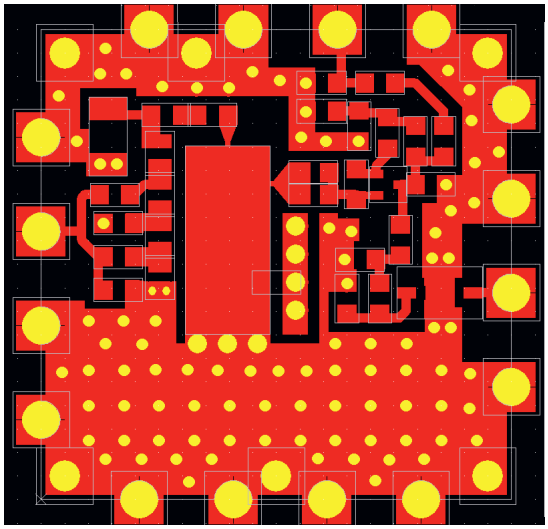
Figure 5-20: Transformation of the oscillator of Figure 5-19 into the form of a standard Colpitts feedback oscillator: (a) first stage of transformation combining the equivalent active and resonator networks in Figure 5-19; (b) combining L_1 and C_x (due to C_a and parasitics) to obtain an equivalent capacitance C_1; and (c) final feedback form (compare with the circuits in Figures 5-2(a) and 5-3(b)).

However, the reactive loading by C_a and C_b modifies the effective device conductance so that it becomes frequency independent. The design of C_a and C_b will be considered in depth later. The capacitors C_a and C_b are key to presenting an admittance to the resonator network that has the required characteristics for stable oscillation.

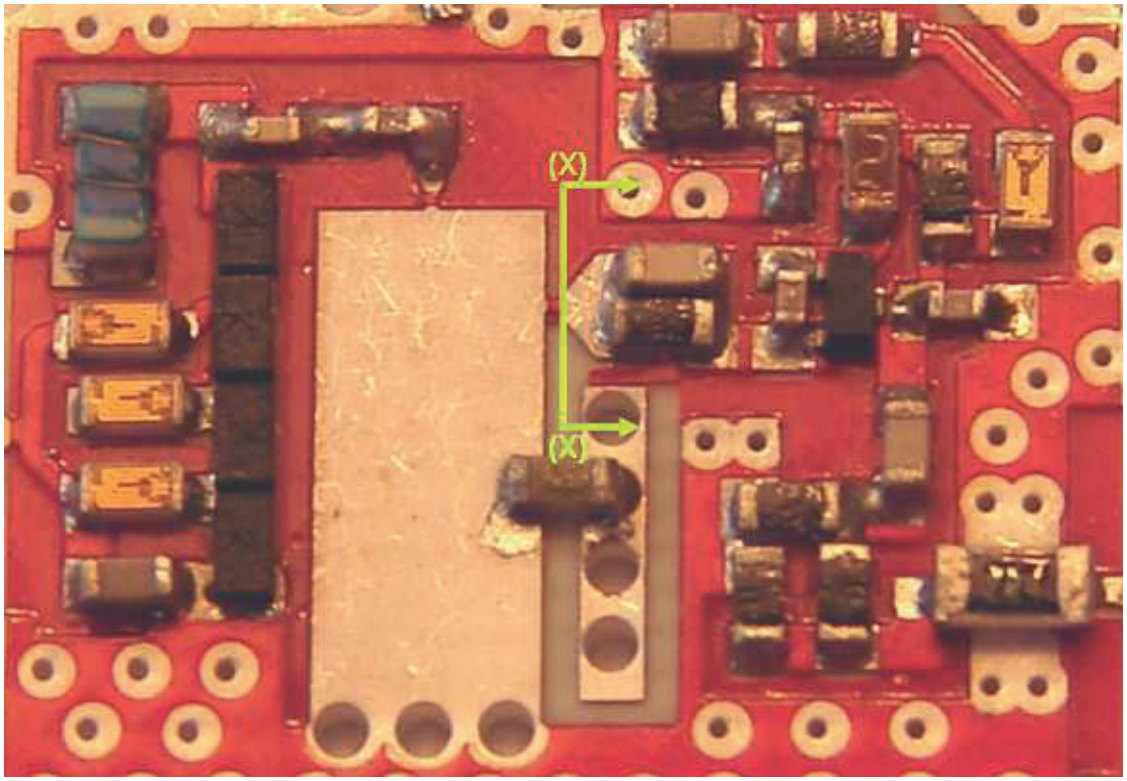
The resonator network, to the left of (x-x) in Figure 5-19(a), consists of a transmission line TL_1 that is coupled to a variable capacitance provided by the stack of four varactors. A single varactor would provide a voltage-tunable capacitance, but the stack of four varactors allows a four-times-larger RF voltage swing [20]. The 3.6 nH and 8.2 nH inductors provide DC shorts while presenting RF open circuits. The tapped transmission line, referring to the connection between the active and resonator networks not being at the top of the transmission line, improves the loaded Q of the resonator network. The resonator network is resonant at a frequency below the oscillation frequency, with the transmission line being inductive at the oscillation frequency. So at the oscillation frequency the resonator network presents an inductance with the required slope of admittance with respect to frequency[4]. The design of the varactor stack is discussed further in Section 5.3 of [21].

The layout and populated microstrip circuit board are shown in Figure 5-21(a and b). The resonator network is shown in Figure 5-22(a) and the active network is shown in Figure 5-22(b). Note the many vias to the backing ground plane. This is typical of microwave designs, as the vias eliminate substrate modes and the large grounded regions minimize parasitic coupling. The wide microstrip resonator, TL_1, is seen in Figure 5-22(a) and there is a shorting bar (available in chip form as a $0\ \Omega$ resistor) across it to a ground strip. The bar can be repositioned and used to tune the resonator network. While this is necessary during manual oscillator design optimization, it is retained in the final design (as is usual). The shorting bar effectively reduces the length of TL_1. At the output of the oscillator (see Figure 5-22(b)) a Pi attenuator (with $294\ \Omega$ resistors in the shunt legs and a $17.4\ \Omega$ series resistor) is between V_{out} and the $50\ \Omega$ bandpass filter. The filter blocks harmonics from the final output of the circuit. The impedance seen from the V_{out} terminal looking toward the filter is $50\ \Omega$.

[4] The resonator could be adjusted to present a capacitive susceptance.



(a)



(b)

Figure 5-21: C-band VCO circuit: (a) layout showing metalization and vias to ground planes (in yellow); and (b) populated circuit board with the resonator network to the left of the cutaway line (x-x) separated from the active circuit to the right.

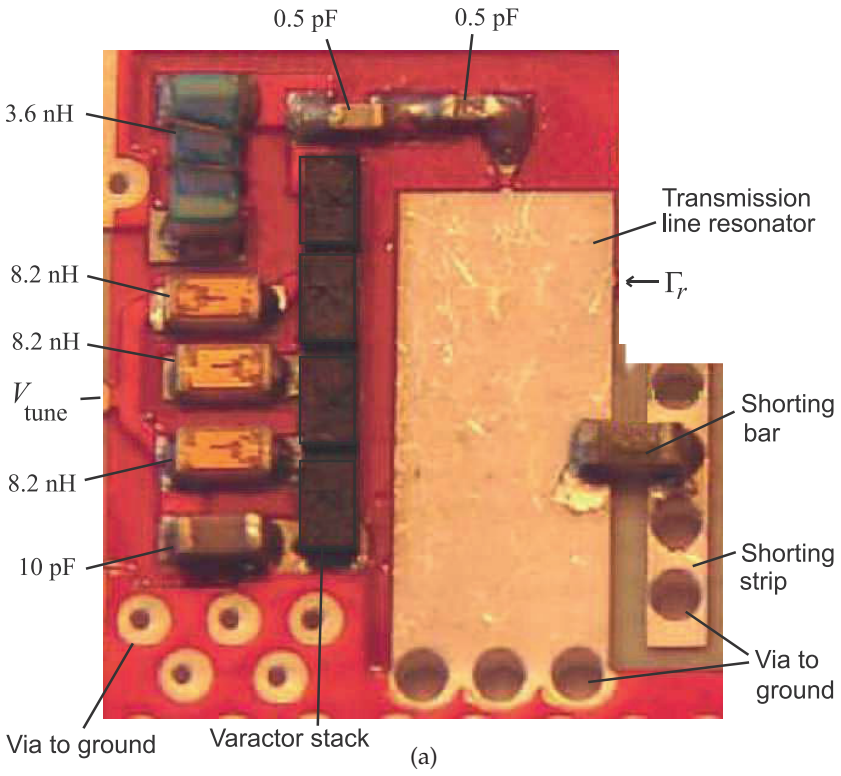


Figure 5-22: C-band VCO circuit: (a) annotated resonator network; and (b) annotated active network. The Pi attenuator (with 294 Ω resistors in the shunt legs and a 17.4 Ω series resistor) is between V_{out} and the 50 Ω bandpass filter.

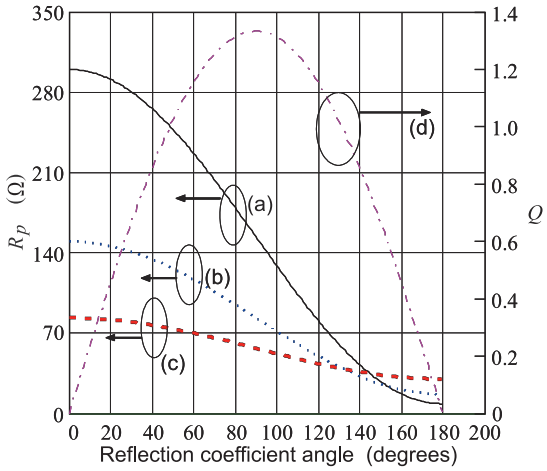


Figure 5-23: The resistance, R_P, of a parallel (or shunt-tuned) resonator required to satisfy the condition of oscillation for (a) $|\Gamma_d| = 1.4$, (b) $|\Gamma_d| = 2$, and (c) $|\Gamma_d| = 4$ versus the magnitude of the device reflection coefficient angle. Curve (d) is the active device $Q = Q_d = |B_d/G_d|$ for $|\Gamma_d| = 2$.

5.6.2 Design Strategy

Small-signal S parameters are generally good indicators of oscillator operation, particularly for the frequency of oscillation [22]. However, they do not provide sufficient information to determine if stable oscillation will occur. The design strategy here is to use simulations and measurements of the individual resonator and active networks to select the required modifications, particularly the selection of C_a and C_b and the position of the shorting bar. Measurements are needed as the characteristics are very sensitive to coupling parasitics, principally between the sides of the lumped elements, which cannot be captured in simulation.

5.6.3 Oscillator Start-Up Considerations

Another consideration in oscillator design is that the right conditions must be provided for oscillation start-up, which initially begins with the amplification of noise. Simply stated, the condition for oscillator start-up of a shunt-tuned oscillator is that for small signals (small A) $|G_d(A,\omega)| > G_r(\omega)$. That is, for small signals, the active network must present a negative conductance that is greater in magnitude than the positive conductance of the resonator network. Since $B_d(\omega) \approx -B_r(\omega)$, the condition for start-up of oscillation can also be expressed in terms of the Qs of the active and resonator networks. That is, for oscillation start-up, $Q_r > Q_d$ for small signals where the Q of the active network is $Q_d = |B_d/G_d|$ and the Q of the resonator network is $Q_r = |B_r/G_r|$. Since G_d is negative, $|\Gamma_d| > 1$, however, it is not sufficient to simply have a large value of $|\Gamma_d|$.

There is a specific angular range of Γ_d that provides the right condition for oscillator start-up. Now $1/\Gamma_d \approx \Gamma_r$ at steady-state oscillation and so designing for a particular Γ_d also determines Γ_r. It is found that the angle of Γ_d, $\angle\Gamma_d$, must be constrained so that losses in the resonator network are accommodated. The appropriate range is selected using Figure 5-23, which plots the required minimum parallel resistance, R_p $(= 1/G_r)$, of the resonator network as a function of $|\angle\Gamma_d|$ for several values of $|\Gamma_d|$. If $|\angle\Gamma_d|$ is less than $50°$, then the resonator network would need to have a higher Q, Q_r, in order to satisfy the oscillator start-up requirement. Now if a large tuning range of the VCO is required, then a large range of $\angle\Gamma_d$ must be considered. A

reasonable range derived from Figure 5-23 is that $\angle\Gamma_d$ between $30°$ and $70°$ should be supported. Thus, in the case of a shunt-tuned oscillator, the design of the interface of the resonator and active networks is a methodical process to provide an appropriate admittance to enable oscillator start-up over the tuning bandwidth of the VCO.

What the above means is that if the active network is designed to present a negative conductance and little susceptance (so that $\angle\Gamma_d \approx 180°$) to the resonator network, then the equivalent parallel resistance, R_p, of the resonator network needs to be very high to ensure start-up. A high R_p means that the Q of the resonator network, Q_r, must be high. However such a high Q_r would be difficult to achieve with a tunable resonator. To provide for greater likelihood of oscillator start-up as well as achievable Q_r, it is better for the active network to present $\angle\Gamma_d \approx 50°$. The wide tuning range requirement extends this range to something like $30° \leq \angle\Gamma_d \leq 70°$. It is thus not reasonable to simply embed reactances in the resonator network to compensate for parasitics in the active network and then to expect that the required start-up conditions be achieved. It may be possible to do this, but this approach would limit the ability to make other trade-offs that would optimize oscillator performance. So the point is that better oscillator performance can be achieved by going beyond the simple form of the Kurokawa oscillation condition embodied in Equation (5.7).

With a fixed-frequency common-base Colpitts oscillator the oscillation operating point (the intersection of Γ_r and $1/\Gamma_d$) can be in the left-half plane of the Smith chart as the loss of the resonator is negligible. However with a microwave VCO the resonator loss, due to the varactors, is appreciable and the location of the intersection of Γ_r and $1/\Gamma_d$ is important.

In summary, designing for a slightly reactive Γ_d is a subtle point that leads to a VCO design of superior performance. It is not necessary to understand this issue to follow the design procedure that follows. It is sufficient to follow the VCO design procedure by noting that one of the requirements is that $30° \leq |\angle\Gamma_d| \leq 70°$, that is, the input admittance of the active network at the resonator-active network interface should be slightly capacitive (or slightly inductive). In matching $1/\Gamma_d$ to Γ_r, this corresponds to a slightly inductive (or slightly capacitive) resonator network with $70° \geq \angle\Gamma_r \geq 30°$ (or $-70° \leq \angle\Gamma_r \leq -30°$).

5.6.4 Initial Design

The initial design of the oscillator with $C_a = 0.5$ pF and with $C_b = 0$ resulted in the simulated resonator and active network characteristics shown in Figure 5-24, where the locus of Γ_r at two bias voltages, 1 V and 3 V, are shown. Also shown are the small-signal characteristics of the active network, plotted as $1/\Gamma_d$, and determined (in harmonic balance simulation) using a $50\ \Omega$ port. The port impedance was high enough to suppress oscillation. The $1/\Gamma_d$ curve intersects with each of the resonator curves at multiple places so that multiple oscillations are possible.

The similar measured characteristics of the actual oscillator are shown in Figure 5-25. The resonator network is shown to the left of the (x-x) line in the oscillator schematic of Figure 5-19(a) and again in Figure 5-19(c). It is also shown in Figure 5-22(a). Measurement of the tank circuit yielded the set of resonator curves (Curves a–g) in Figure 5-25. Curve a is the locus of the

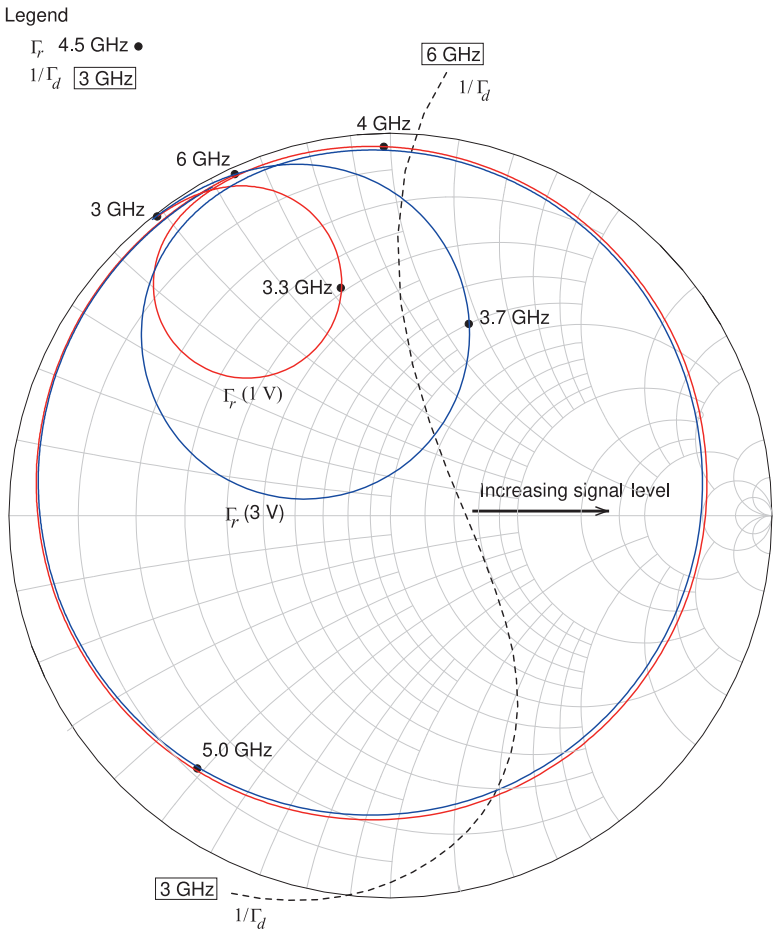


Figure 5-24: Simulated 50 Ω reflection coefficient of the resonator, Γ_r, and the inverse of the small-signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO with $C_a = 0.5$ pF and without C_b. This is a 50 Ω Smith chart.

reflection coefficient of the resonator network when the tuning voltage is 0 V, that is, Curve a is the locus with respect to frequency of Γ_r(0 V). The seven resonator responses shown in Figure 5-25, Curves a–g, are the resonator reflection coefficients for equally spaced tuning voltages from 0 V through 9 V. Comparing Figures 5-24 and 5-25 it is seen that there is reasonable agreement between the simulated and measured results. The difference can be attributed to the difficulty of performing the measurements at the correct location, as well as coupling between the components themselves not being captured in simulation.

The possibility of multiple oscillations is seen in Figures 5-24 and 5-25, as there can be multiple intersections of a Γ_r curve (for a particular bias) and the $1/\Gamma_d$ curve. (Note that the $1/\Gamma_d$ locus will shift to the right as the level of the oscillation signal increases.) Multiple oscillations are observed as seen in the spectrum at the output, V_{out}, of the oscillator (see Figure 5-26). An oscillator that oscillates at multiple frequencies is not desirable, of course, so the design must be altered to avoid the multiple intersections of the Γ_r (at a particular tuning voltage) and $1/\Gamma_d$ curves.

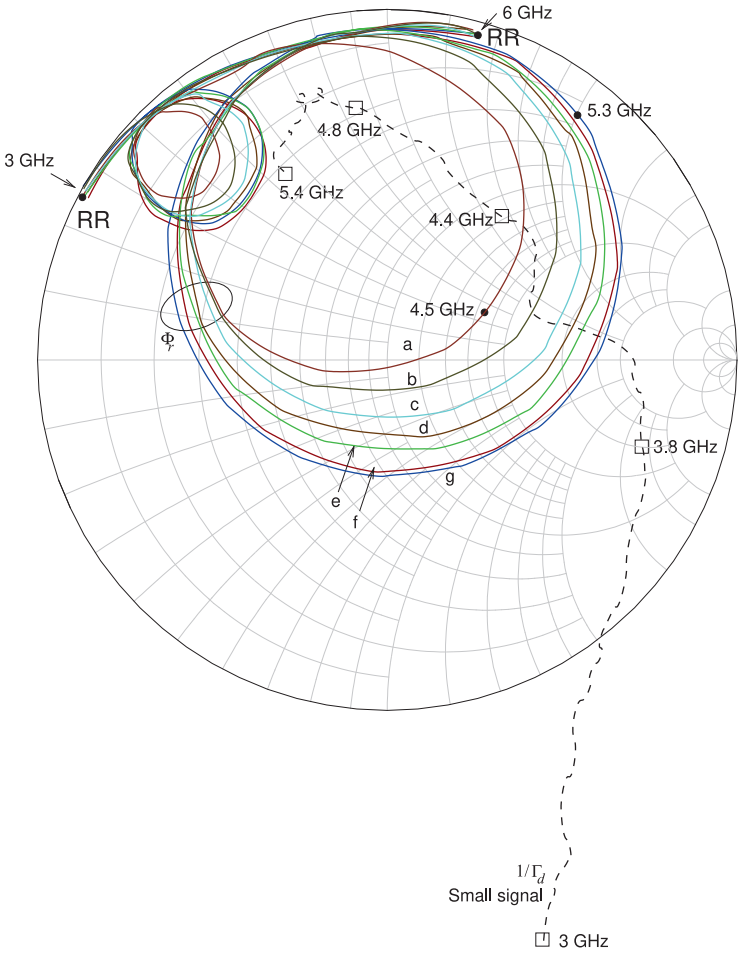


Figure 5-25: Measured reflection coefficient of the resonator network, Γ_r, and the inverse of the small-signal reflection coefficient of the active device network, $1/\Gamma_d$, of the C-band VCO with $C_a = 0.5$ pF and without C_b. Γ_r, Curve a, is the characteristic of the resonator network for a tuning voltage, V_{tune}, of 0 V. Curve g is for $V_{\text{tune}} = 9$ V. V_{tune} is equally spaced for Curves a–g. The locus of $1/\Gamma_d$ moves to the right as the signal level increases and the intersection of $1/\Gamma_d$ and Γ_r determines the oscillation frequency and the signal level.

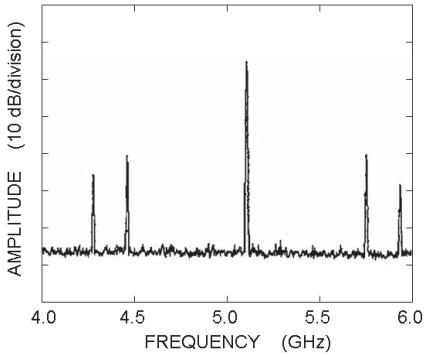


Figure 5-26: Multiple oscillations observed prior to reflection coefficient shaping by C_a and C_b. The spectrum was measured using a resolution bandwidth of 3 MHz and a video bandwidth of 1 MHz.

5.6.5 Avoiding Multiple Oscillations Through Reflection Coefficient Shaping

At an early stage in design (with $C_a = 0.5$ pF and $C_b = 0$) multiple oscillations were observed in Figure 5-26. The cause of these multiple oscillations is multiple intersections of the $1/\Gamma_d$ locus of the active circuit and Γ_r at a particular bias voltage. This is seen in both the simulated results in Figure 5-24 and the measured results in Figure 5-25. The next step in design is to shape $1/\Gamma_d$ of the active network so that there is a unique intersection

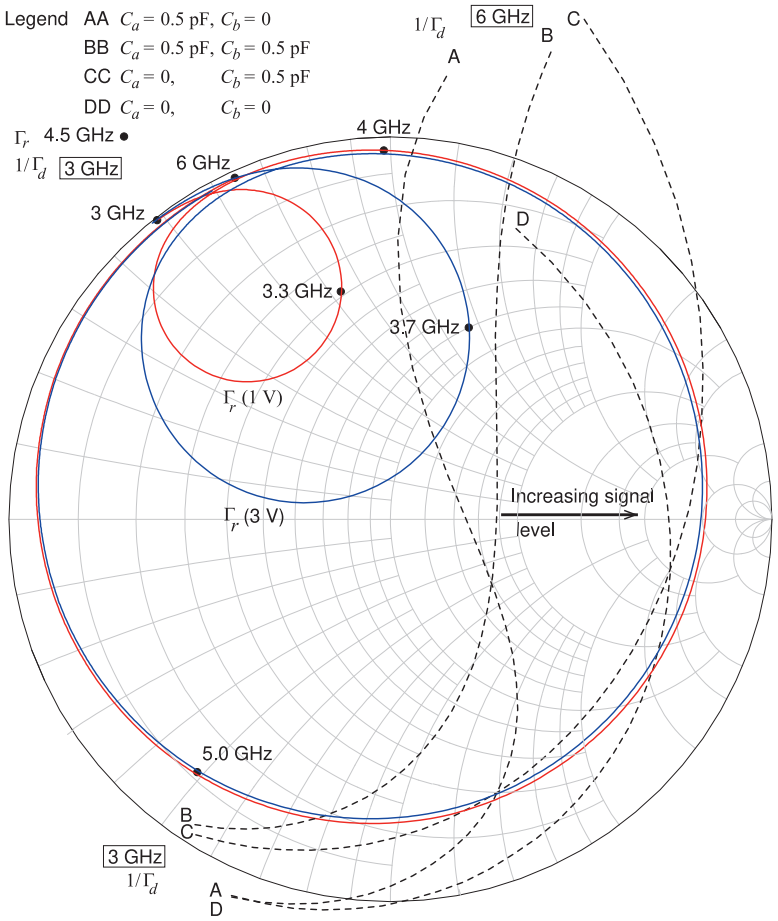


Figure 5-27: Simulated reflection coefficient of the resonator, Γ_r, and the inverse of the small-signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO for various values of the compensation capacitors C_a and C_b.

of $1/\Gamma_d$ and Γ_r at each bias voltage. The elements used to shape $1/\Gamma_d$ are C_a and C_b. At the same time that these are adjusted the design must achieve $(\partial B_d/\partial V|_{V=V_0}) \approx 0$. Also, the discussion in Section 5.6.3 indicated that the preferred angle of Γ_d to ensure start-up of oscillations is around $50°$. However, for a wide tuning range of the VCO it is only possible to achieve a specific angle of Γ_d at a single oscillation frequency. The trade-off is to choose $-70° \leq \angle\Gamma_d \leq -30°$, indicating that the active network should be slightly capacitive, which corresponds to a slightly inductive resonator network with $70° \geq \angle\Gamma_r \geq 30°$. Thus the intersection of Γ_r and $1/\Gamma_d$ should be in the top right quadrant of the Smith chart. (An alternative design choice which still would have resulted in a successful start-up of the oscillator is $70° \geq \angle\Gamma_d \geq 30°$.)

The simulated characteristics of the resonator and active networks are shown in Figure 5-27. The loci of Γ_r for two tuning voltages are shown, and the small-signal $1/\Gamma_d$ is shown for various values of C_a and C_b. Curve AA (of $1/\Gamma_d$) is for $C_a = 0.5$ pF and $C_b = 0$, as considered before, and will resulted in multiple oscillations. Curve DD is for $C_a = 0$ and $C_b = 0$ and is very close to Γ_r and it may be difficult for oscillation to begin. Another way of describing this is that Q_r is very close to the small-signal Q_d. So even if oscillation did start, it would reach steady state at a low signal level. The active networks represented by Curves CC and DD do not provide sufficient

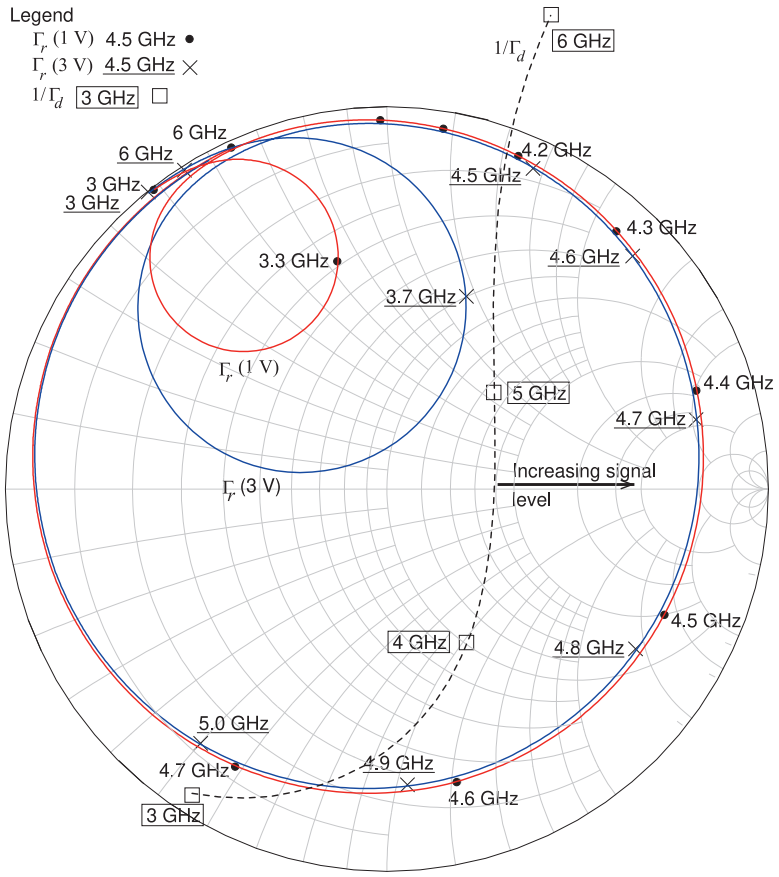


Figure 5-28: Simulated reflection coefficient of the resonator, Γ_r, and the inverse of the small-signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO.

negative conductance and only enable oscillation over a narrow frequency range. The best characteristic here is Curve BB for which $C_a = 0.5$ pF and $C_b = 0.5$ pF. This response provides a single intersection of the resonator curve for each tuning voltage and the active network curve. Also it indicates a fairly large magnitude of negative conductance so that the output oscillator power will be high. That is, the magnitude of the negative conductance reduces as the signal level increases and the large magnitude of the negative conductance at small signals means the signal level can grow significantly before the conductances of the resonator and active networks match.

The simulated characteristics of the resonator and active networks are repeated in Figure 5-28 for $C_a = 0.5$ pF and $C_b = 0.5$ pF (i.e. Curve BB in Figure 5-27). The response of the circuit now has the desired properties. First consider the locus of the reflection coefficient of the resonator network with 1 V across the varactor diodes (this is the Γ_r(1V) curve). There are two resonances between 3 GHz and 6 GHz, but it is the resonance between 4.0 GHz and 5.3 GHz that is close to the $1/\Gamma_d$ locus and so (for this varactor bias) is the only resonance that will affect oscillation. The locus of Γ_r(1V) approximately follows a constant conductance curve so that $(\partial G_r/\partial\omega|_{\omega=\omega_0}) \approx 0$ as required. As the signal level across the active network increases, the $1/\Gamma_d$ locus shifts to the right in the direction of constant susceptance so that $(\partial B_d/\partial V|_{V=V_0}) \approx 0$ as desired. Provided that the frequencies match, the point at which the Γ_r(1V) locus intersects the $1/\Gamma_d$ locus determines both the oscillation frequency and the oscillation level as

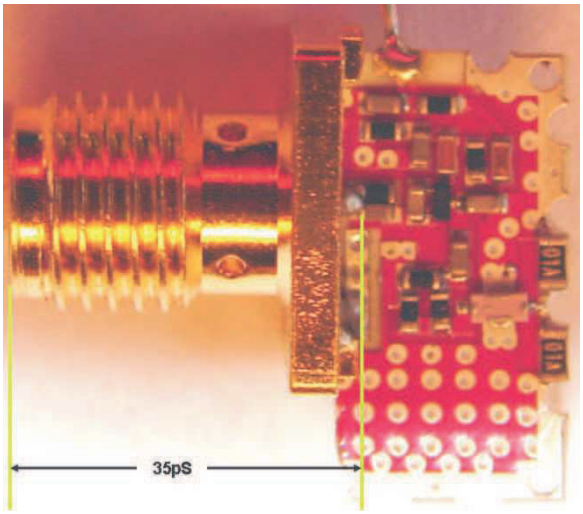


Figure 5-29: Measurement of the active circuit with a 50 Ω test fixture at the interface of the resonator and active networks. The card was cut at the interface to make the connection. A 35 ps delay due to the length of the SMA connector must be subtracted from measurements to reference measurements to the circuit card edge.

the locus of $1/\Gamma_d$ shifts to the right.

The simulated results discussed in the previous paragraph indicate that the oscillator will work as desired. However, there are many parasitics and coupling effects that are not fully captured in simulation. Final design optimization requires experimental investigation of the reflection coefficients looking into the resonator circuit and into the active circuit.

The next step in design of the VCO is using a VNA to measure the reflection coefficient of the active network in Figures 5-19(b) and 5-22(b) (shown again in its measurement configuration in Figure 5-29). The large signal locus in Figure 5-30 was measured with a 10 dBm signal applied to the active network at the 50 Ω measurement port. This curve is an indication of what the active network presents to the resonator under large signal conditions and it is used as a guide since it does not capture the full loading complexity, for example, the harmonic terminations are not correct.

Oscillation occurs when the characteristics of the active network (see the $1/\Gamma_d$ curve in Figure 5-30) match the characteristic of the resonator network, Curves a–g in Figure 5-30. First, for small signals the active network should provide

$$|1/\Gamma_d(A,\omega)| < |\Gamma(\omega)| \tag{5.8}$$

at all desired frequencies of operation. This is the requirement for oscillation start-up. Second, the rotation of $\Gamma_r(\omega)$ with respect to ω near the oscillation point (i.e., ω_0) should be positive (i.e., $(\partial B_r/\partial\omega|_{\omega=\omega_0}) > 0$ as developed in Section 5.5.2) and in the opposite direction to the $1/\Gamma_d(A,\omega)$ locus with respect to ω (i.e., in the same direction as the rotation of $\Gamma_d(A,\omega)$). This is indeed what happens and can be seen by closer examination of Curves a–g. In addition, the locus of $\Gamma_r(\omega)$ should approximately follow a line of constant conductance so that $(\partial G_r/\partial\omega|_{\omega=\omega_0}) \approx 0$. Now, device self-limiting stabilizes oscillation when the angles of $\Gamma_d(A,\omega)$ and $\Gamma_r(\omega)$ sum to zero. For single-frequency oscillation this must be obtained at each tuning voltage. Finally, the trajectory of the limiting $1/\Gamma_d(A,\omega)$ locus (i.e as the amplitude of oscillation, A or V, increases) should intersect the $\Gamma_r(\omega)$ locus, ideally at right angles to minimize phase noise [11, 17]. These requirements are referred to as a complementary relationship between the active and resonator networks.

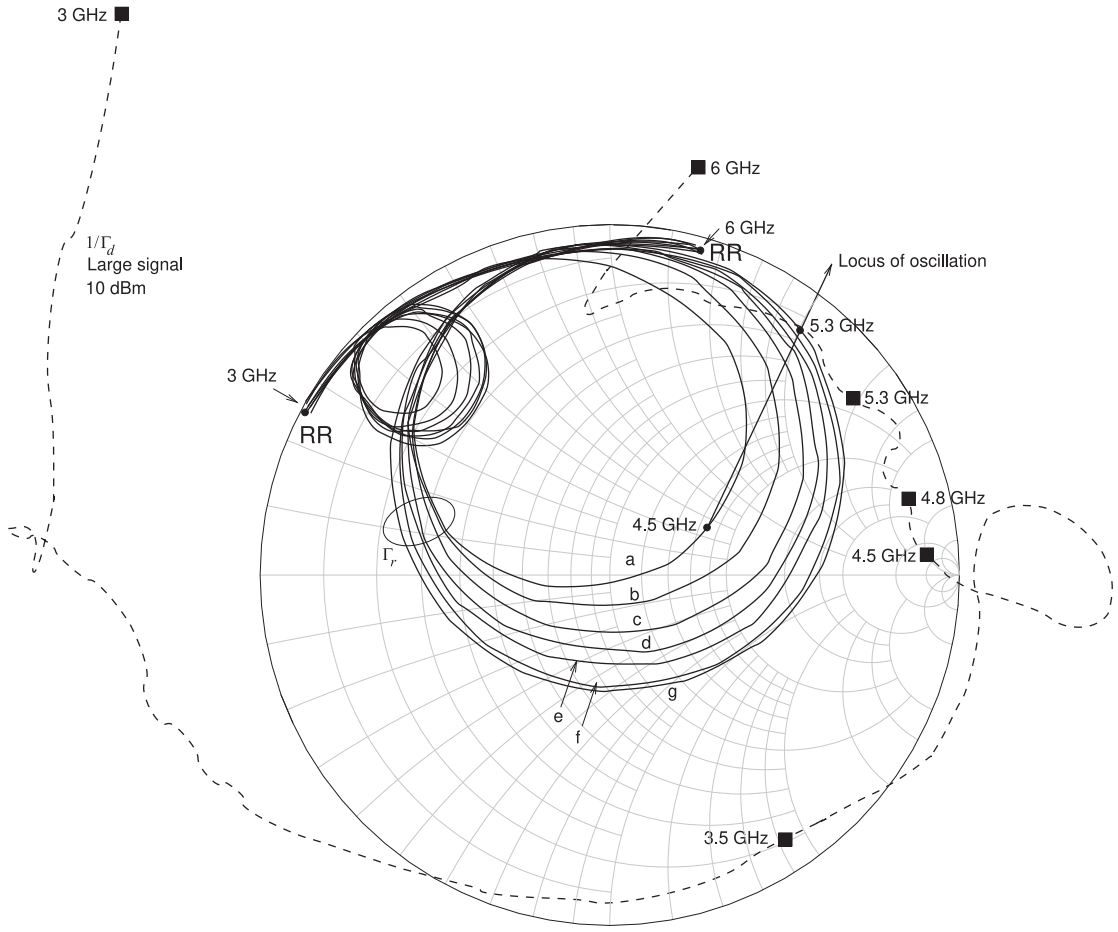


Figure 5-30: Measured reflection coefficient of the resonator network, Γ_r, and the inverse of the large signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO. For the large signal Γ_d measurement, the active network is driven at 10 dBm from a $50\ \Omega$ port. The Γ_r curves are identical to those in Figure 5-25. The curves with end points labelled RR identify the resonator curves.

While under small-signal conditions the loci may not coincide, the important point is that they do when limiting occurs, as well as providing for the start-up of oscillation. A slight counterclockwise rotation of the modified active device $1/\Gamma_d$ locus as the signal level increases (as well as a general right shift) ensures stable, single-frequency oscillation. Put another way, the trajectory of the negative conductance as device limiting occurs must be such that $1/\Gamma_d$ just intersects the Γ_r locus, and $\angle\Gamma_d$ must complement $\angle\Gamma_r$. This is the situation shown in Figure 5-30, where the modified device network characteristic is achieved by adding capacitive terminations to the collector and the emitter base terminals. Here, unlike the conventional common-base series feedback oscillator situation (as considered in Section 5.4), the input of the active network is now capacitive above 4.5 GHz (see Figure 5-30). Consequently the inductance of the resonator is successfully absorbed. Thus

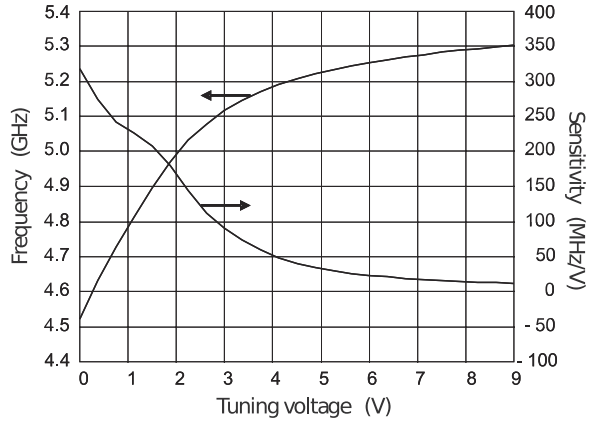


Figure 5-31: Measured tuning characteristic showing oscillation frequency and VCO sensitivity as a function of tuning voltage.

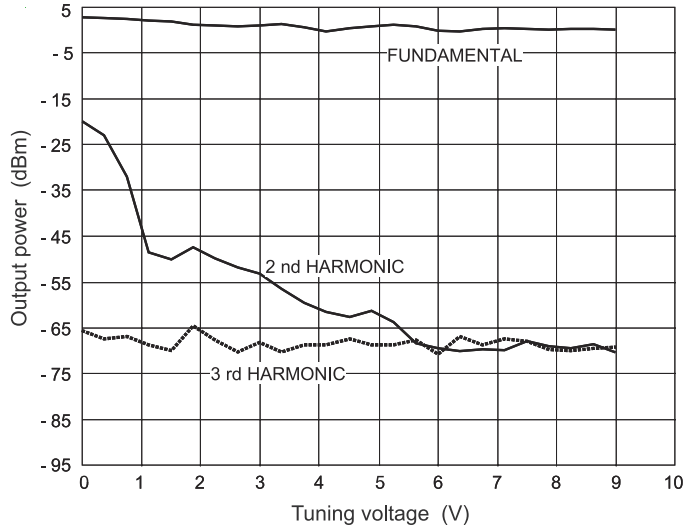


Figure 5-32: Measured output power and harmonics at V_{out} (before the bandpass filter) indicating low-level harmonic content.

the small-signal one-port reflection coefficient of the resonator is inductive initially. In effect the resonator is operated as a tunable shunt inductance rather than a tunable capacitive reactance.

5.6.6 VCO Performance

The most important metrics that describe the performance of the VCO are the phase noise, tuning bandwidth, output power, tuning gain or sensitivity, the output harmonic content, and the DC power consumption. The VCO characterized here, shown in Figure 5-19, includes the compensating capacitors C_a and C_b, both 0.5 pF. Figures 5-31 and 5-32 document the bandwidth and output powers of the VCO. As the varactor tuning voltage, V_{tune}, goes from 0 V to 9 V the filter tunes from 4.5 GHz to 5.3 GHz, producing a minimum output power of 0 dBm and varies by no more than 2 dB over the range. The DC power consumed is 150 mW. The tuning bandwidth is adjusted by varying the coupling (provided by the two 0.5 pF capacitors) between the varactor stack and the microstrip line, TL_1. Figure 5-32 also shows the power at the harmonics. At the final output these are further reduced by the bandpass filter.

The measured phase noise is shown in Figure 5-33 at 4.5 GHz

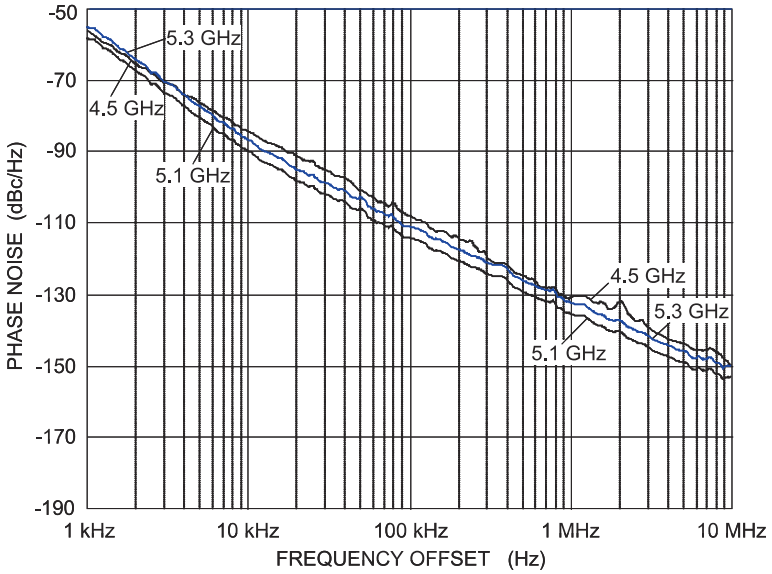


Figure 5-33: Phase noise measured at the top and bottom of the tuning range as well as at 5.1 GHz, where phase noise is optimum. Minimum phase noise floor -116 dBc/Hz at 1 kHz offset, -160 dBc/Hz at 10 MHz offset.

(corresponding to a tuning voltage of 0 V), and at 5.3 GHz (a tuning voltage of 9 V), as well as at 5.1 GHz where the best phase noise is obtained. The phase noise is approximately the same across the tuning range with a $1/f^2$ noise corner frequency, $f_{c,-2}$ (the transition from an f^{-1} dependency to an f^{-2} dependency) of 30 kHz. The phase noise at 10 kHz offset, $\mathcal{L}(10\text{ kHz})$, is better than -85 dBc/Hz, while $\mathcal{L}(1\text{ MHz})$ is better than -130 dBc/Hz. The best measured phase noise, $\mathcal{L}(1\text{ MHz})$, near band center (5.1 GHz) is -135 dBc/Hz.

The performance of a VCO should be quoted as the worst performance over the tuning bandwidth. For a tuning bandwidth of 770 MHz and center frequency of 4.92 GHz, the maximum phase noise of this VCO is -128 dBc/Hz. This improves to a maximum phase noise of -130 dBc/Hz for a bandwidth of 500 MHz centered at 5.05 GHz.

5.6.7 Summary

The VCO design here used a standard one-port reflection oscillator design approach with a technique that introduced compensation capacitors to manage the otherwise frequency-dependent susceptance of the active network. These compensation elements also resulted in the reflection coefficient of the augmented active device having the characteristics necessary to ensure stable oscillation and oscillator start-up.

5.7 Negative Transconductance Differential Oscillator

In RFICs it is common to use an oscillator with a tank circuit across a pair of matched transistors in a differential configuration. Such an oscillator is shown in Figure 5-34(a). So while this circuit is in a differential configuration, it is analyzed and designed as a reflection oscillator at RF.

The cross-connected differential common source pair creates a negative resistance while the fixed inductors (the Ls) and the voltage-tunable capacitors, the Cs, form the variable LC tank circuit. The tunable capacitors are typically implemented using semiconductor varactor diodes whose

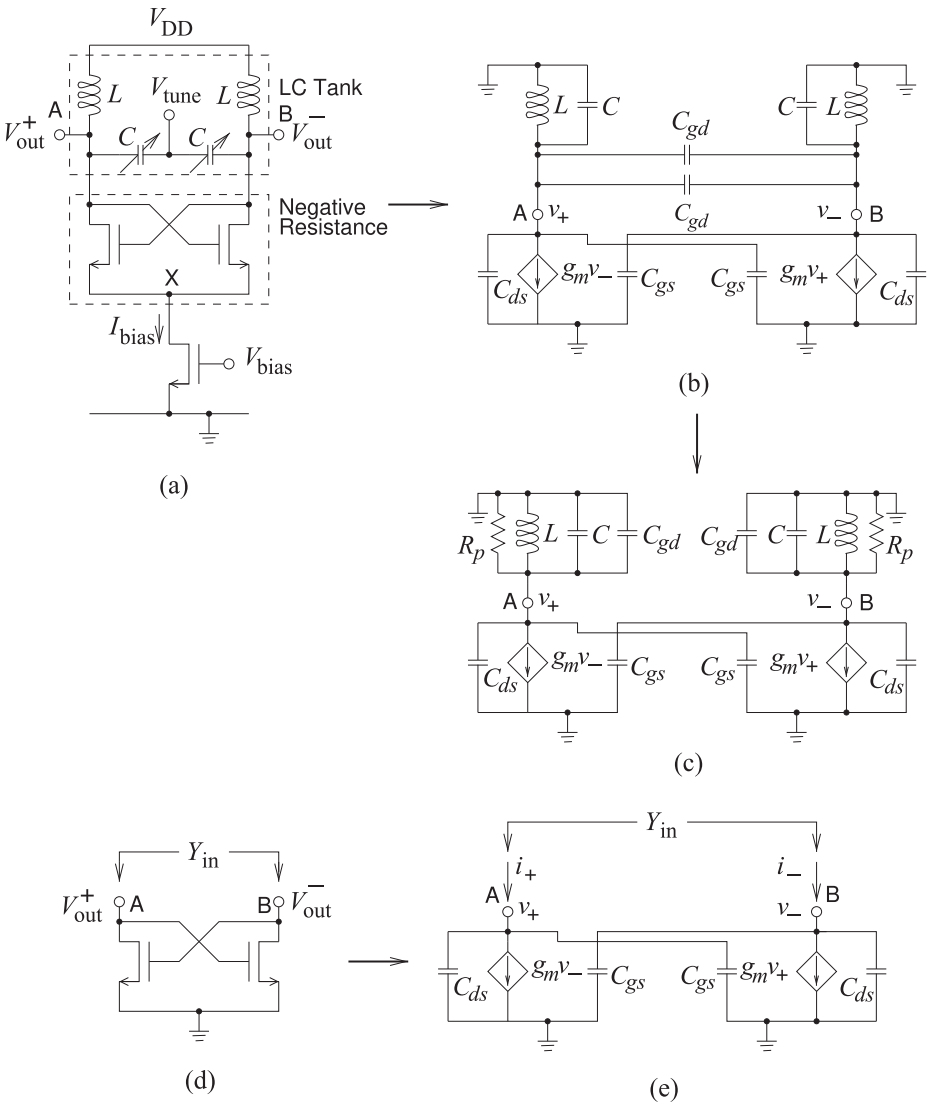


Figure 5-34: Negative-gm differential FET VCO: (a) schematic; and (b) small-signal model used in analyzing the oscillator; (c) small-signal model with C_{gd} incorporated in the tank circuit; (d) negative resistance network of the VCO; and (e) small-signal model used in deriving the input impedance of a negative resistance network. This is a modified form of a Colpitts oscillator. The L-C-C_{gd} resonant circuit operates below resonance and presents an effective inductance (a positive reactance) but with an admittance derivative with respect to frequency that is less than that of an actual inductor. This is essential for stability. The effective inductor, L_3 in Figure 5-7(b), connects the output of each of the transistors to its respective input. For each transistor C_{gs} is C_1, and C_{ds} is C_2, in 5-7(b)

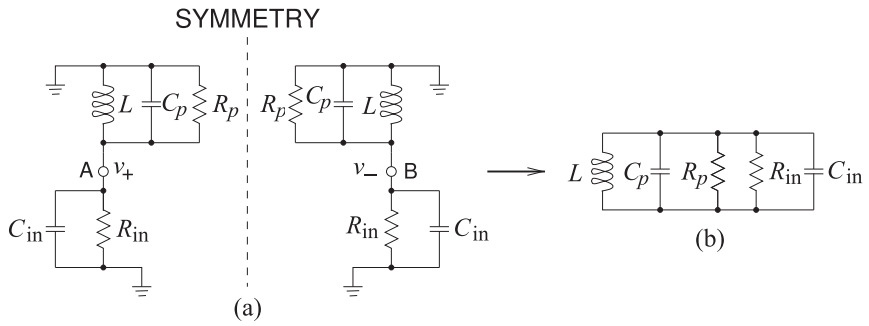


Figure 5-35: Reduced model of the differential FET VCO of Figure 5-34:(a) small-signal model with negative resistance FET network replaced by the equivalent resistance and capacitance; and (b) simplest parallel small-signal model combining the tank and the negative resistance network model.

capacitance can be adjusted by the tuning voltage V_{tune}. The bias transistor, the bottom FET, sets the current in the differential transistors and this current directly impacts the power consumption of the oscillator and the phase noise. The circuit is symmetrical so that the node between the two variable capacitors, the V_{tune} terminal, looks like an RF short as does the common source node of the differential pair, the node labeled X. This is key to developing the small-signal model shown in Figure 5-34(b), where the dominant parasitic capacitances of the transistors, the drain-source capacitance (C_{ds}), the gate-source capacitance (C_{gs}), and the gate-drain capacitance (C_{gd}) are seen. C_{gd} becomes part of the tank circuit. This leads to the simpler small-signal model shown in Figure 5-34(c). Removing the tank circuit leads to the small-signal active device models shown in Figure 5-34(d and e) which present a negative resistance to the tank circuit and load.

The input admittance of the negative resistance network (Figure 5-34(e)) can now be determined. Analysis begins by summing currents at the A and B nodes, respectively:

$$i_+ = (\jmath\omega C_{gs})v_+ + (\jmath\omega C_{ds})v_+ + g_m v_- \tag{5.9}$$

$$i_- = (\jmath\omega C_{gs})v_- + (\jmath\omega C_{ds})v_- + g_m v_+. \tag{5.10}$$

The differential input admittance is then

$$Y_{\text{in}} = \frac{i_+ - i_-}{v_+ - v_-} = -g_m + \jmath\omega(C_{gs} + C_{ds}). \tag{5.11}$$

Thus the negative resistance network is modeled as a negative resistance of value $R_{\text{in}} = (-1/g_m)$ in parallel with a capacitance $C_{\text{in}} = (C_{gs} + C_{ds})$. The dependence of R_{in} on g_m gives this oscillator its name "negative transconductance oscillator" or "negative-gm oscillator." The gate-drain capacitance C_{gd} is in parallel with the tank capacitance C and so a new equivalent capacitance $C_p = C + C_{gd}$ can be defined. Loss in the resonator circuit is modeled by a resistor R_p in parallel with C_p. The small-signal model of the oscillator is now as shown in Figure 5-35(a). This further reduces to the model shown in Figure 5-35(b). Oscillations will initiate if

$|1/R_{\text{in}}| = |g_m| > 1/R_p$. Also the oscillation frequency, f_0, is the frequency at which the shunt reactance is zero, that is,

$$f_0 = \frac{1}{2\pi}\frac{1}{\sqrt{L(C_p + C_{\text{in}})}} = \frac{1}{2\pi}\frac{1}{\sqrt{L(C + C_{gd} + C_{gs} + C_{ds})}}. \tag{5.12}$$

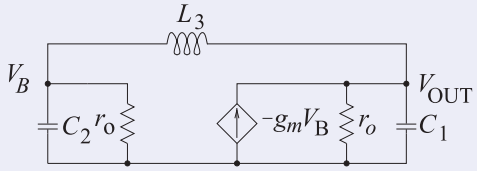
As oscillation builds up, $|g_m|$ reduces to the value of $1/R_p$ and stable oscillation is obtained. The negative-gm oscillator has the ideal characteristic if the negative conductance is the only element dependent on amplitude. Unfortunately the values of C_{gd}, C_{gs}, and C_{ds} also vary as the amplitude of the signal increases. This complicates design at microwave frequencies as these variations could lead to multiple simultaneous oscillations.

EXAMPLE 5.2 Oscillator Analysis

Determine the frequency of oscillation of a Colpitts common emitter BJT oscillator.

Solution:

Figure 5-36 shows two different implementations of a common emitter Colpitts BJT oscillator. The form in Figure 5-36(a) is the most direct implementation, with a clearly defined insertion of the Colpitts network in the collector-to-base feedback path. In Figure 5-36(a), the resistors R_1 and R_2 provide base biasing, and L_C is an RF choke. The oscillation frequency of this oscillator can be derived from the small-signal model of the oscillator. Since R_1 and R_2 will be relatively large resistances, and since L_C is an RF choke (it will look like an RF open circuit), the small-signal model of the oscillator is as shown below.



In this small-signal model, r_π is the base input resistance and r_o is the output resistance—both of these will be relatively large. The transconductance of the transistor is g_m. The network equations are obtained by summing the currents leaving the base node, with Y_1, Y_2, and Y_3 being the admittances of C_1, C_2, and L_3 respectively:

$$Y_2 V_B + G_\pi V_B + Y_3(V_B - V_{\text{OUT}}) = 0, \tag{5.13}$$

$$Y_1 V_{\text{OUT}} + g_m V_B + Y_3(V_{\text{OUT}} - V_B) + G_o V_{\text{OUT}} = 0, \tag{5.14}$$

and $G_\pi = 1/r_\pi$, $G_o = 1/r_o$. In matrix form

$$\begin{bmatrix} (Y_2 + Y_3 + G_\pi & -Y_3) \\ (g_m - Y_3) & (Y_1 + Y_3 + G_o) \end{bmatrix} \begin{bmatrix} V_B \\ V_{\text{OUT}} \end{bmatrix} = 0. \tag{5.15}$$

This can be simplified by noting that r_π and r_o will have admittances smaller than Y_1, Y_2, and Y_3. Thus Equation (5.15) becomes

$$\begin{bmatrix} (Y_2 + Y_3) & -Y_3) \\ (-Y_3) & (Y_1 + Y_3) \end{bmatrix} \begin{bmatrix} V_B \\ V_{\text{OUT}} \end{bmatrix} = 0. \tag{5.16}$$

Equation (5.16) has a solution only if the determinant of the matrix is zero. That is,

$$\begin{aligned}(Y_2 + Y_3)(Y_1 + Y_3) - Y_3 Y_3 &= Y_1 Y_2 + Y_2 Y_3 + Y_1 Y_3 + Y_3^2 - Y_3^2 \\ &= Y_1 Y_2 + Y_2 Y_3 + Y_1 Y_3 = 0.\end{aligned} \tag{5.17}$$

Now $Y_1 = j\omega C_1$, etc., where $\omega = 2\pi f$ is the radian oscillation frequency. Thus Equation (5.17) becomes

$$-\omega^2 C_1 C_2 + \frac{C_1}{L_3} + \frac{C_2}{L_3} = -\omega^2 C_1 C_2 + \frac{C_1 + C_2}{L_3} = 0. \tag{5.18}$$

Rearranging, the frequency of oscillation is

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{L_3}\frac{(C_1 + C_2)}{C_1 C_2}}. \tag{5.19}$$

The same result is obtained for the alternative form of the Colpitts oscillator shown in Figure 5-36(b).

5.8 Advanced Discussion of Oscillator Noise

This section presents a discussion of oscillator noise, and particularly the rapid increase of phase noise close to the carrier. Noise can be partitioned into amplitude and phase noise components. The nonlinear saturation of an oscillator suppresses amplitude noise so only phase noise is of concern. While usually associated with oscillators, phase noise is also added to a signal by an amplifier.

There is not a consensus as to the origins of close-to-carrier phase noise. This section begins with observations of oscillator noise in the frequency domain and in the time domain. Then three theories of excess oscillator noise, Leeson's theory, the linear time-invariant model, and the chaotic map model, are presented.

Not having a complete model of the physical origin of phase noise means that a simulator cannot reliably predict the phase noise of an oscillator. Also designing an oscillator with good phase noise performance currently relies heavily on experience and projections based on what has been achieved by a designer previously.

5.8.1 Observations of Oscillator Noise in the Frequency Domain

The most puzzling noise observed with oscillators is the noise observed at a small frequency offset from the carrier (i.e., the average oscillation signal). To develop an appreciation for the breadth of observations, the signals produced by several different oscillators will be considered. First, Figure 5-37

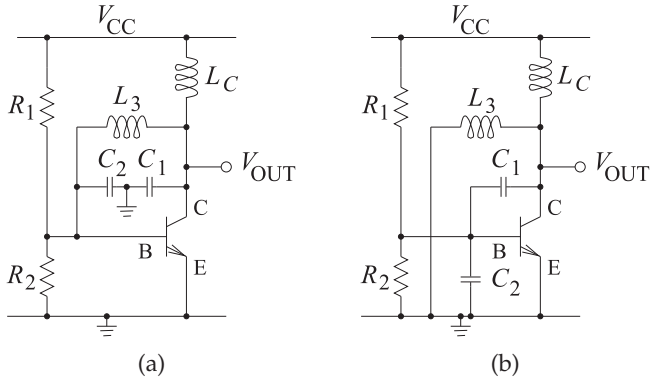


Figure 5-36: Common emitter BJT Colpitts oscillators: (a) configuration with a feedback network between the collector and base of the transistor; and (b) alternative configuration.

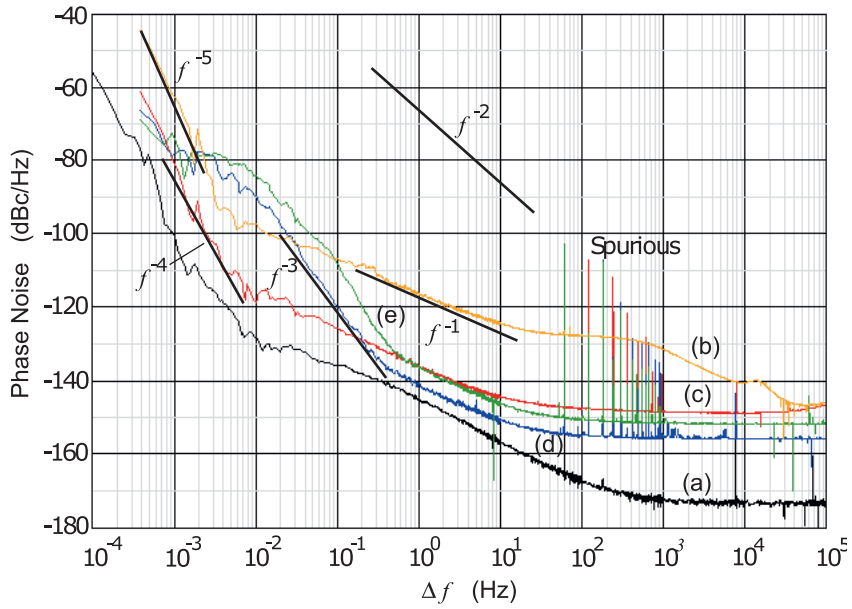


Figure 5-37: Measured phase noise of low-frequency oscillators: (a) instrument noise floor; (b) HP 5087A frequency distribution amplifier at 5 MHz (used to drive the external reference input of several test instruments using a single high-quality oscillator); (c) TADD-1 frequency distribution amplifier at 10 MHz; (d) TADD-1 frequency distribution amplifier at 5 MHz; (e) Spectracom 8140T frequency distribution amplifier at 10 MHz. Five phase noise regions are identified as f^{-5}, f^{-4}, f^{-3}, f^{-1}, and white noise. The spurious signals are related to injected harmonics of the 60 Hz power mains. Used with permission of John Ackermann [23].

is a plot of the phase noise observed at the output of several oscillators and amplifiers operating at 5 MHz and 10 MHz. Curve (a) is the noise floor of the noise measurement instrument and spurious tones are observed at multiples of 60 Hz, the power mains frequency. Curves (b), (c), (d), and (e) show phase noise varying in straight-line segments. Being a log-log plot, these curves show phase noise varying as f^{-5}, f^{-4}, f^{-3}, f^{-1}, and f^{0}. None of the phase noise plots here show a region with an f^{-2} dependence, although this is observed with other oscillators.

Another oscillator to consider is the VCO circuit shown in Figure 5-38 [24]. This is a 50 MHz VCO with a semiconductor varactor being the variable element with a zero-bias capacitance of 100 pF. The capacitance of the varactor is controlled by the voltage, V_b. With $V_b = 0$ V, the phase noise shown in Figure 5-39 was observed. The distinct phase noise regions have frequency dependencies of f^{0}, f^{-1}, f^{-2}, and f^{-3}. The phase noise of this oscillator is plotted again in Figure 5-40 for three different varactor bias voltages. The phase noise characteristics of the oscillator change even though the underlying physical sources of noise do not change (of course). Curve (a), with $V_b = 6$ V, and Curve (b), with $V_b = 0$ V, have an f^{-1} region around 20 kHz (see Figure 5-39 for more details), but the f^{-1} region is not observed in Curve (c) where $V_b = 18$ V. One interpretation is that the crossover frequencies have shifted. So what is particularly interesting here is

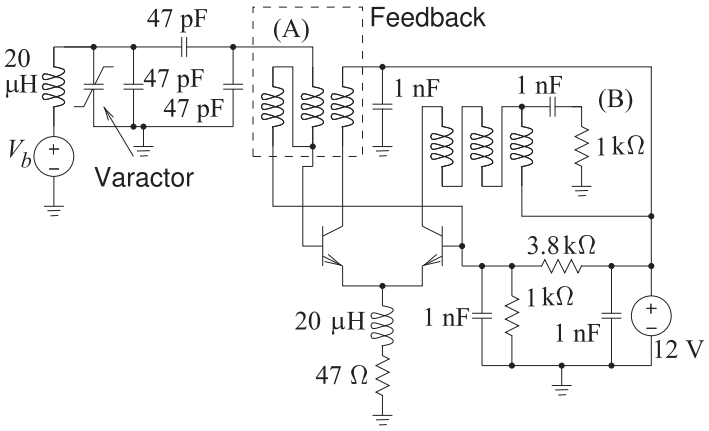


Figure 5-38: Varactor-tuned VCO schematic, from [24].

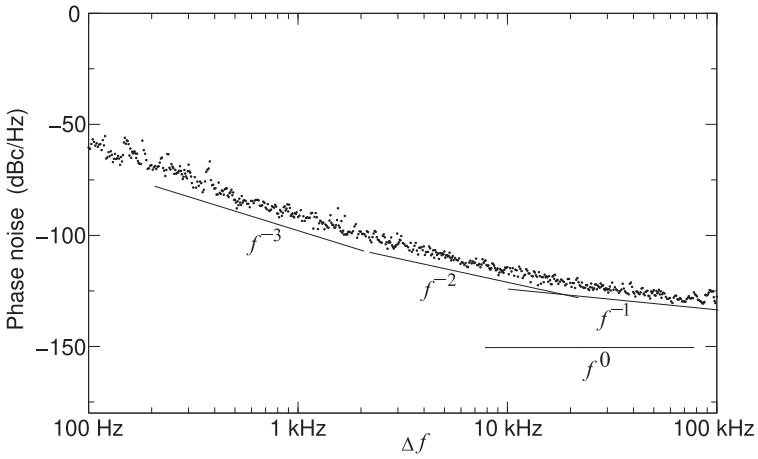


Figure 5-39: Measured phase noise of a 50 MHz BJT varactor-based VCO with the varactor biased at 0 V [25, 26]. Three phase noise regions are identified as f^{-3} (having a slope of -9 dB/octave), f^{-2} (having a slope of -6 dB/octave), and f^{-1} (having a slope of -3 dB/octave).

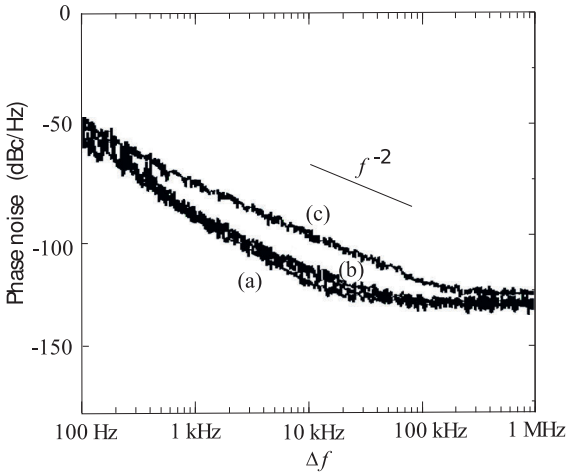


Figure 5-40: Measured phase noise of a 50 MHz BJT varactor-based VCO at three varactor bias voltages: (a) 6 V; (b) 0 V; and (c) 18 V [25]. The varactor breakdown voltage is 30 V Curve (b) was also plotted in Figure 5-39.

that the same physical source of noise can be manifested quite differently at the output of an oscillator when the circuit bias is changed.

The third phase noise example is for a 2.4 GHz power oscillator that has the output spectrum shown in Figure 5-41 with regions having dependencies of f^{-3} and f^{-0}, but nothing in between. (The slight increase in noise power spectral density at 40 kHz offset is due to dynamics of the oscillator's

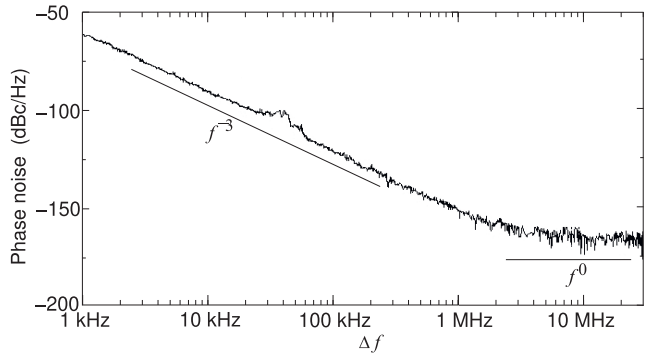


Figure 5-41: Phase noise of a 2.4 GHz power oscillator with an output power of 34.5 dBm [27, p. 323]. Two phase noise regions are identified as f^{-3} (having a slope of -9 dB/octave) and white noise (with an f^0 dependency).

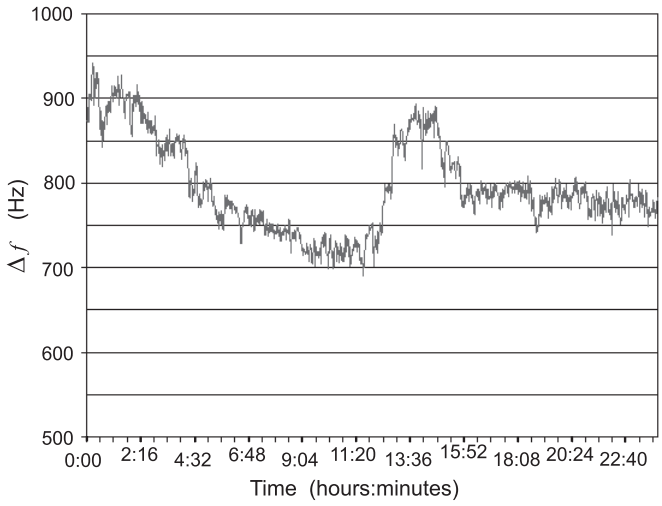


Figure 5-42: Long-term stability of a 10 GHz oscillator measured over a 24 hour interval after being on for 3 weeks. Used with permission of John Ackermann [23].

feedback loop.) Finally, the 5 GHz oscillator considered in Section 5.6 has f^{-3} and f^{-2} phase noise regions (see Figure 5-33).

So the whole range of phase noise dependencies on frequency offset are observed, but the universal observation is that the dependency of the noise power spectral density is to a non-positive integer power of frequency (i.e., $f^{-n}, n = 0, 1, \ldots$).

5.8.2 Observations of Oscillator Noise in the Time Domain

An important time-domain characterization of noise is referred to as random walk noise. An example of this is the variation of the oscillation frequency over a long period of time. The long-term stability of a 10 GHz oscillator is shown in Figure 5-42. This noise cannot be characterized in the frequency domain and instead is described by its **Allan variance**, $\sigma_y^2(\tau)$, or **Allan deviation**, $\sigma_y(\tau) = \sqrt{\sigma_y^2(\tau)}$, defined as follows.

If the frequency measured at time t is $f(t)$ and the nominal oscillation frequency is f_n, then the fractional frequency at time t is defined as

$$y(t) = \frac{f(t) - f_n}{f_n}. \tag{5.20}$$

Then the average fractional frequency over an observation time interval τ is

defined as

$$\bar{y}(t,\tau) = \frac{1}{\tau}\int_0^{\tau} y(t+t_v)dt_v. \tag{5.21}$$

This leads to the definition of the Allan variance as

$$\sigma_y^2(\tau) = \tfrac{1}{2}\left\langle \bar{y}_{n+1} - \bar{y}_n \right\rangle, \tag{5.22}$$

where τ is the observation period and $\bar{y}_n$ is the nth fractional frequency average over the time interval τ. Note that there is no dead-time between the nth and $(n+1)$th measurement time intervals.

The random walk shown in Figure 5-42 is an important clue to unraveling flicker noise. Figure 5-42 shows long-term memory and here it is shown that there is memory over several hours. Even on smaller time scales, random walk is apparent and there is a self-similar property—the hallmark of chaotic behavior. Could this random walk effect and $1/f$ noise arise from the same physical process? Most likely, but there is no accepted theory.

5.8.3 Excess Oscillator Noise: The Leeson Effect and Flicker Noise

As seen in Figures 5-37 to 5-41, oscillators have noise that increases as the offset Δf, from the mean oscillation frequency decreases. This noise has separable regions where noise varies as Δf^n, where n is an integer ranging from 0 to -5. There are transition regions between these discrete states, but there is not a region where n is a fractional number. Not all of the discrete states are observed because, presumably, either the crossover frequencies have changed order, or the frequency offset, Δf, was not low enough.

In 1966 Leeson [28] examined the effect of feedback on noise in oscillators (see Figure 5-1). The phase noise mechanism treated by this analysis is now called the Leeson effect. Leeson showed that white phase noise and white flicker noise (white here meaning independent of frequency) of the amplifier in the feedback loop translate to noise on the oscillation signal with power law dependencies of f^{-2}, called white frequency noise, and f^{-3}, called flicker frequency noise, respectively. These were the dominant "nonwhite" forms of noise observed in his time. However, his analysis did not predict the level of the noise accurately and sometimes was off by an order of magnitude.

The Leeson effect is briefly summarized here. First, it was observed that nearly every physical system has fluctuations that vary as $1/f$ at low frequencies. This includes electrical devices such as the amplifier in an oscillator feedback loop. This leads to equal amplitude phase and amplitude noise superimposed on the oscillation. Since noise is small, the amplitude fluctuations are suppressed by the saturation of the active device so that the only noise observed in good designs is phase noise. Leeson determined that the oscillator phase noise has a region with Δf^{-3} dependence that is due to low-frequency f^{-1} noise (i.e. around DC), a Δf^{-2} region due to white noise in the bandwidth of the oscillator's tank circuit, and also a white noise region outside the bandwidth of the tank circuit. The basis for the development of Leeson's oscillator phase noise model is shown in Figure 5-43. Mathematically [28],

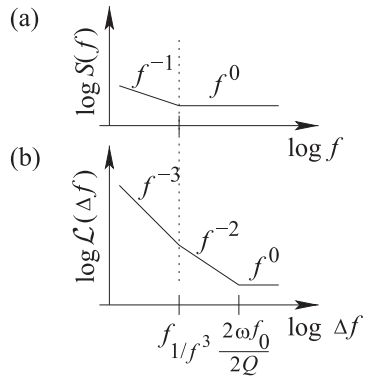


Figure 5-43: Derivation of oscillator noise spectra: (a) the noise spectra of an electronic material with noise increasing as the frequency reduces; and (b) the noise spectra close to the oscillation frequency of an oscillator.

$$\mathcal{L}(\Delta f) = \mathcal{L}(\Delta\omega) = \frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right], \tag{5.23}$$

where Q is the loaded Q factor of the oscillator's tank circuit and F is an empirical factor. $\mathcal{L}$ has the units of $\text{radians}^2/\text{Hz}$, or in decibels,

$$\mathcal{L}|_{\text{dB}}(\Delta f) = 10\log\left\{\frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right]\right\}, \tag{5.24}$$

which has the units of dB/Hz or more usually expressed as "decibels below the carrier" of power P_0, or dBc/Hz. This is the power at a specified offset such as the phase noise of a 5.05 GHz oscillator at an offset of 1 MHz and with an output power of 0 dBm being -130 dBc/Hz [19].

The derivation of the oscillator noise characteristics from first principles, which led to Equations (5.23) and (5.24), predicts noise levels that are much lower than those observed in practice [29, 30]. As well, the prediction inherent in Equation (5.23) is that by increasing the Q of the tank circuit the noise level will be reduced. However, this is not always obtained in practice. A further complication is that Equation (5.23) provides no mechanism for the generation of $1/(\Delta f)$ and $1/(\Delta f)^3$ noise in the oscillator phase noise spectrum. An adhoc modification of Equation (5.23) accounts for this:

$$\mathcal{L}(\Delta f) = \frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right]\left(1 + \frac{f_{c,-3}}{|\Delta f|}\right). \tag{5.25}$$

Given the inadequacy of this model it is just as well to use

$$\mathcal{L}(\Delta f) = \sum_{i=0}^{-5} b_i f_i^n \tag{5.26}$$

where the b_i coefficients are extracted from measurements.

The Leeson effect can be stated as oscillator phase noise being up-converted white noise around DC.

5.8.4 Excess Oscillator Noise: Linear Time-Variant Model

The Leeson effect model described in the previous subsection uses a linear time-invariant model of the oscillator and does not consider down-conversion of noise from frequencies near harmonics. The linear time-variant

model, also called the Hajimiri and Lee model, incorporates these higher-order conversion mechanisms [31].

Noise injected into an oscillator has a different impact depending on whether it is injected at the peak of the oscillating signal or at the zero-crossings. The noise injected at the peaks of the oscillating signal are quenched by the saturating effect of the active device in the oscillator. However, noise at or near the zero-crossings of the waveform introduces jitter and phase noise. This effect on phase noise can be described by an impulse sensitivity function [31]. Consider an impulse injected at phase $x = \omega_0 t$, then the time-domain impulse response is

$$h_\phi(t,\tau) = \frac{\Gamma(\omega_0 t)}{q_{\max}} u(t-\tau), \tag{5.27}$$

where $\Gamma(\ )$ is the impulse sensitivity function, $q_{\max}$ is the maximum charge displacement on the capacitor forming the tank circuit, t is the observation time, and τ is the time of the excitation. The excess phase of the oscillator (the additional phase induced onto the phase of the carrier) is

$$\phi(t) = \frac{1}{q_{\max}} \int_{-\infty}^{t} \Gamma(\omega_0 t) i(\tau) d\tau, \tag{5.28}$$

where $i(\tau)$ is the noise current injected in the oscillator.

$\Gamma(\ )$ can be derived approximately for some oscillators such as the CMOS LC oscillator in [32], where it was shown that the impulse sensitivity function can be expressed as a Fourier series with a fundamental component corresponding to the frequency of oscillation. The excess phase at the zero-crossings of the oscillator is

$$\phi(t) = \frac{1}{q_{\max}} \left[c_0 \int_{-\infty}^{t} i(\tau) d\tau + \sum_{m=1}^{\infty} c_m \int_{-\infty}^{t} i(\tau) \cos(n\omega_0 \tau) d\tau \right], \tag{5.29}$$

where c_m are coefficients of the Fourier series. The first term with the c_0 coefficient indicates noise that is up-converted from baseband, while the term in the summation is the contribution to the oscillator phase noise due to down-conversion of noise near the harmonic frequencies. With the assumption that the noise at the harmonics is white noise with mean-square current $\bar{i_n^2}$, then the noise spectral density is [32, 33]

$$\mathcal{L}(\Delta\omega) = 10 \log \left(\frac{\bar{i_n^2}}{\Delta f} \frac{\sum_{m=0}^{\infty} c_m^2}{4 q_{\max}^2 \Delta\omega^2} \right). \tag{5.30}$$

Here $\Delta f = 1$ Hz for noise in a 1 Hz bandwidth and $\Delta\omega$ is the radian offset frequency. Thus white noise at the harmonics is down-converted to f^{-2} noise at the oscillation frequency, and f^{-1} noise at baseband is up-converted to f^{-3} noise at the oscillation frequency.

The time-variant model provides a richer description of phase noise on an oscillating signal than does the Leeson model, but it does not describe f^{-1}, or $f^{-n}, n > 3$, noise that is observed with oscillators.

Thus the Hajimiri and Lee model relates to down-conversion of white noise at harmonics of the oscillation frequency to the near carrier noise which is in addition to Leeson's model of up-converted near-DC white noise. The Hajimiri and Lee, and Leeson models of phase noise led to designers developing microwave oscillators with significantly lower phase design yet there remains appreciable near-carrier phase noise.

5.8.5 Excess Oscillator Noise: Chaotic Maps and Flicker Noise

While not firmly established, it is possible that flicker noise originates from nonlinear dynamics and chaos [25, 26, 34–41]. In this model flicker noise derives from a nonlinear process with delayed feedback. The mathematical foundation is well established [42], describing a phenomenon called intermittency [43] that occurs when a physical process transitions between stable periodic states and chaotic states. Inherent to some forms of intermittency is long-term memory with a $1/f$ spectrum [44]. This has been established for many physical and biological systems.

Logistics Map

A classic example of **intermittency**, and the first widely accepted, is the following model of population dynamics. If t_n denotes discrete time and (the real number) x_n denotes the ratio of the existing population to the maximum possible population at t_n (so x_n is between 0 and 1), then what is called the logistics map provides the population ratio, $x_{(n+1)}$, at time t_{n+1}. The logistics map is [45]

$$F_\lambda(x) = \lambda x_n(1 - x_n), \tag{5.31}$$

and so

$$x_{n+1} = F_\lambda(x). \tag{5.32}$$

Here λ is a positive number representing the combined rate of reproduction and starvation. So environmental conditions determine λ, which is constrained so that $0 < \lambda \leq 4$. Depending on λ, the logistics map (i.e., Equation (5.31)) will produce a stable population or a random population depending on the value of λ, with $\lambda = 4$ producing white noise.

Thermal noise produces random fluctuations in the amplitude and phase of a sinusoidal signal that is being processed in a nonlinear electronic system such as an amplifier or an oscillator. Denoting the thermal amplitude fluctuations by $a_{t,I}(t)$ and the thermal phase fluctuations by $\phi_{t,I}(t)$, a sinusoidal signal with mean amplitude A and an initial phase of zero is

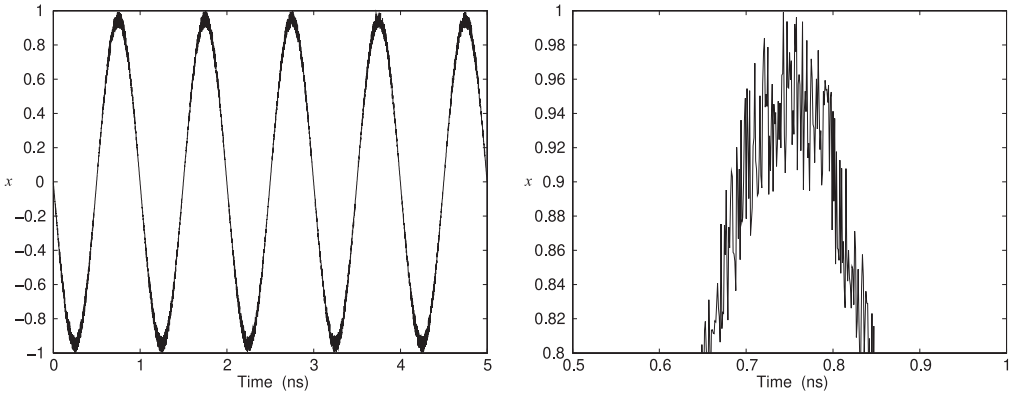
$$x(t) = A[1 + a_{t,I}(t)]\cos[\omega t + \phi_{t,I}(t)]. \tag{5.33}$$

Using the logistics map with $\lambda = 4$ (which produces white noise) to determine $a_{t,I}(t)$ and $\phi_{t,I}(t)$, a sinusoidal signal with thermal (white) noise is as shown in Figure 5-44. Using $a_{t,I}(t)$ and $\phi_{t,I}(t)$ determined from a Gaussian distribution would yield the same qualitative result. Of course most of this noise would be easy to remove by bandpass filtering but thermal noise will still appear within the finite bandwidth of the signal. It is just easier to visualize the effect of noise by plotting it on this scale.

Equation (5.31) is a simple nonlinear equation with delayed feedback that mixes x over time. What is called the rate of mixing describes the extent of correlation to past events and can be thought of as an exponential decay rate. However, with the logistics map, $F_\lambda(x)$ in Equation (5.31), the rate of this mixing is not controllable.

Logarithmic Map

There are many maps that will lead to $1/(\Delta f)$ effects and one of the most convenient to use in modeling flicker noise in electronics is called the loga-



(a) Waveform (b) Detail

Figure 5-44: Sinusoidal signal with superimposed white noise calculated using the logistics map rather than calculating noise as a Gaussian process.

rithmic map [46, 47]:

$$F_\beta(x) = \begin{cases} x(1 + Y(\beta)x\,|\log(x)|^{1+\beta}) & \text{if} \quad 0 \le x \le 1/2 \\ 2x - 1 & \text{if} \quad 1/2 < x \le 1 \end{cases} \tag{5.34}$$

and so $$x_{n+1} = F_\beta(x). \tag{5.35}$$

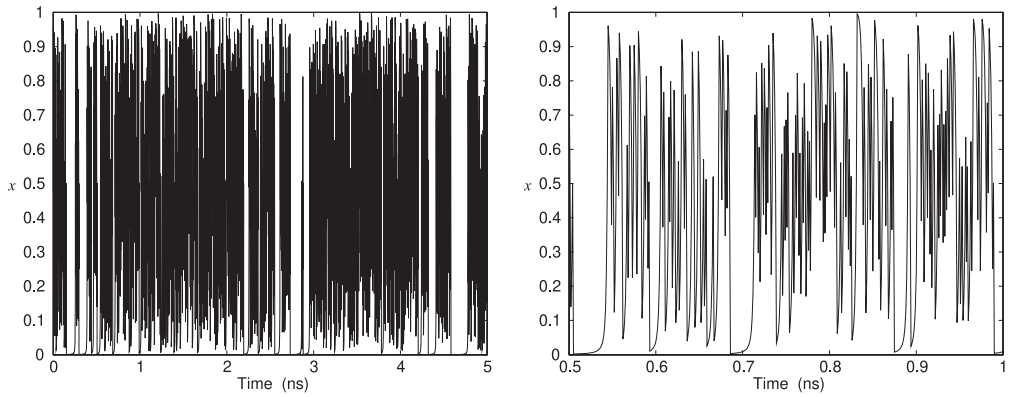
$F_\beta(x)$ is defined on the interval $0 < x \le 1$ and $Y(\beta) = 2(\log 2)^{-(1+\beta)}$ is chosen to ensure that $\lim_{x\to 1/2^-} f_\beta(x) = 1$. (Note that the map is discontinuous at $x = 1/2$.) If Δt is a fixed time interval, then $x(t + \Delta t) = F_\beta(x(t))$.

The logarithmic map uses only a single parameter, β, that controls the rate of mixing and thus the long-term memory property. The solution of the logarithmic map, Equation (5.34), with $\beta = 0.000005$ is shown in Figure 5-45.

The Fourier transform of the sequence plotted in Figure 5-45 is shown in Figure 5-46 and its autocorrelation is shown in Figure 5-47. The spectrum in Figure 5-46 (above 1 Hz) has an $f^{-0.5}$ dependence, and since the sequence corresponds to voltage, squaring this yields a $1/f$ power characteristic. The autocorrelation plot in Figure 5-47 shows the slow long-term decay in the correlation with respect to the discrete interval, i, between the sequence points. That is, the logarithmic map describes a process with slowly decaying correlations. The extended correlation is a measure of the rate of mixing. This interpretation corresponds very well to the understanding of physical systems, and in particular to electronic systems. It can be shown [47] that the rate of decay of correlation of this map is bounded as

$$R(n) \le B(\log n)^{-\beta}, \tag{5.36}$$

where n is the nth time interval. Thus the map is said to describe a logarithmic mixing rate that can be made as slow as desired by varying the value of β. It is this long-range dependence that produces f^{-1} (and f^{-2}, f^{-3}, etc.) noise. In an oscillator these result in phase noise with $1/(\Delta f)$, $1/(\Delta f)^2$, $1/(\Delta f)^3$, etc. characteristics. In semiconductors, for example, the



(a) Solution (b) Detail

Figure 5-45: Solution of the logarithmic map of Equation (5.34) with (a random initial seed) $x_0 = 0.477347$, $\beta = 0.000005$,and $t_{n+1} - t_n$ = 1 ps.

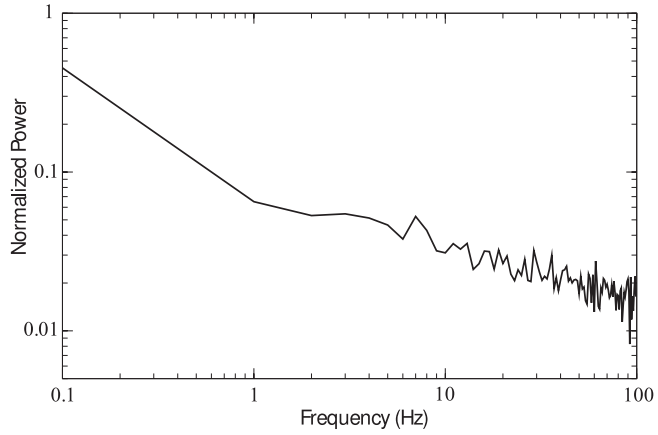


Figure 5-46: Spectrum of the logarithmic map with $\beta = 0.000005$.

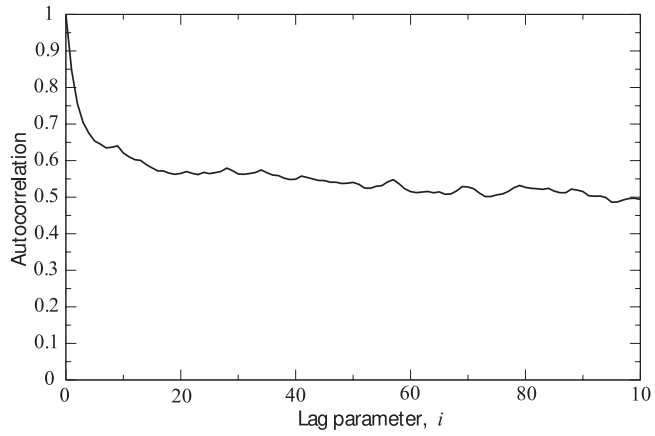


Figure 5-47: Correlation plot of the logarithmic map with $\beta = 0.000005$ for a million-point sequence.

rate of mixing is influenced by the density of traps [48] and lattice scattering [49] with the lower the density of traps (i.e. better quality semiconductor material) and the lower the amount of scattering, the lower the rate of mixing and hence the lower the level of flicker noise. The delayed feedback in the chaotic-map model is consistent with trapping and the observation that reducing traps improves phase noise performance.

The long-range mixing indicated by the slow decay of the correlation function (see Figure 5-47) is key to the f^{-1} response. Another function that yields a long-term correlation is the Ornstein–Uhlenbeck process [50–52] developed to describe Brownian motion. The autocorrelation function of this process decays exponentially and predicts f^{-2} noise but not f^{-1} noise [52]. The Ornstein–Uhlenbeck process decays too rapidly to predict the f^{-1} response. This is discussed further in [52].

So the solution of Equation (5.35) (the logarithmic map), shown in Figure 5-45, has complicated dynamics with long periods of stability with rapid transitions between stable and rapidly varying levels. The sequence of x_ns depends on the starting condition (i.e., x_0), but no matter how it starts, the power spectrum of the solution has an inverse frequency dependence (i.e., it is exactly f^{-1}). The logarithmic map, as with all chaotic maps, describes a nonlinear process with delayed feedback. This matches the situation in physical, biological, chemical, and financial systems. Since nearly every physical process can be described as a (perhaps weak) nonlinear process with delayed feedback, the widespread observation of $1/f$ fluctuations is not surprising. So the basis of $1/f$ noise is the most basic of physical processes.

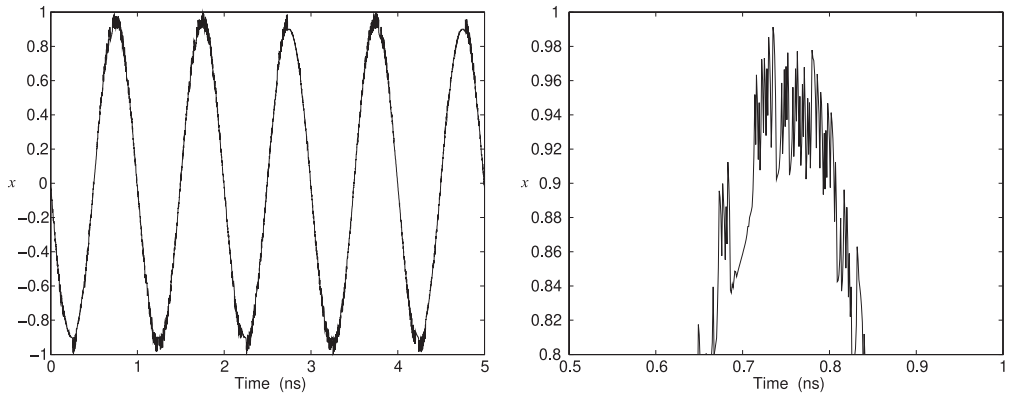
Intermittency (described by chaotic maps) results in random fluctuations in the amplitude and phase of a sinusoidal signal processed by a nonlinear electronic system such as an amplifier or oscillator. Denoting the amplitude intermittency by $a_I(t)$ and the phase intermittency by $\phi_I(t)$, a sinusoidal signal with mean amplitude A and an initial phase of zero is

$$x(t) = A[1 + a_I(t)]\cos[\omega t + \phi_I(t)]. \tag{5.37}$$

This signal is shown in Figure 5-48 with the logarithmic intermittency fluctuations $a_I(t)$ and $\phi_I(t)$ as shown in Figure 5-45, calculated using different seeds. The effect of intermittency fluctuations is greatly exaggerated here for visualization purposes. In practice, the fluctuations at the scale shown in Figure 5-48 could be eliminated using a bandpass filter. However, the fluctuations are self-similar (another property of chaotic processes) and is repeated at all scales. In a bandpass electronic system the in-band amplitude fluctuations are suppressed by device nonlinearity, but the phase fluctuations appear as phase noise on an oscillator.

5.8.6 Summary

Three oscillator phase noise models were presented. The Leeson model is based on up-conversion of white noise from baseband producing noise around the oscillator carrier with an f^{-2} dependency. The Hajimiri and Lee model is based on a linear time-varying model of the oscillator with f^{-1} noise at baseband resulting in oscillator phase noise with a $1/(\Delta f)^3$ dependency, and white noise at the baseband and harmonics resulting in noise around the oscillating frequency with an $1/(\Delta f)^2$ dependency. Up-conversion of noise has been demonstrated as a mechanism that describes



(a) Waveform (b) Detail

Figure 5-48: Sinusoidal signal with superimposed intermittency noise. Using Equation (5.37) with $A = 0.9$, $a_I(t)$ scaled to the interval $[0, 0.09]$, $a_I(0) = 0.477347$ (before scaling), $\phi_I(t)$ scaled to the interval $[0, 5]$ radians, and $\phi_I(0) = 0.00915926$ (before scaling).

some of the observed oscillator noise. Neither the Leeson nor the linear time-varying models describe the full set of observations of phase noise with $1/(\Delta f)^5, \ldots 1/(\Delta f)$ dependencies.

The chaotic map model is physically appealing and describes the origin of flicker noise as the time-delayed feedback of the output of a nonlinear process. This can produce a chaotic response called intermittency that embodies long-term memory. It has been shown through simulation that this model predicts the $1/(\Delta f)^3$, $1/(\Delta f)^2$, $1/(\Delta f)^1$, and $1/(\Delta f)^0$ dependencies of phase noise. It is also consistent with the random walk seen in time-domain observations of oscillator noise. However, the chaotic map-based model has not yet led to a compact formula for phase noise similar to Leeson's formula. The development of a compact phase noise model (e.g., like Leeson's model) will not be simple as integer calculus and transfer function-based analyses cannot be directly used with a chaotic map. However, it is clear that the description of phase noise is getting close to a satisfying physical explanation.

Phase noise can also be induced by vibrations [53, 54], and spurious signals from the environment (such as those coupled from the power mains) can also appear as phase noise.

5.9 Case Study: Oscillator Phase Noise Analysis

In this section the oscillator shown in Figure 5-38 is modeled and simulated in the time-domain [25, 26, 36]. The circuit includes three nonlinear devices, two identical BJTs and one varactor, which must be modeled. The process of developing a device model involves fitting the coefficients of a physically based model to measured characteristics. The development of models of the noise sources also requires fitting some noise parameters to measurements.

Device Modeling

The physical model used with each of the BJTs is shown in Figure 5-49 along with thermal, shot, and flicker noise sources. The Gummel-Poon BJT model was used with parameters provided by the manufacturer. The thermal noise sources, $i_{t,rb}$, $i_{t,rc}$, and $i_{t,re}$, are associated with the parasitic resistors at the collector, base, and emitter, respectively. The thermal noise current source models are

$$i_{t,rc} = \sqrt{\frac{2kT}{R_c}}\xi_c, \quad i_{t,rb} = \sqrt{\frac{2kT}{R_{bb}}}\xi_b, \quad \text{and} \quad i_{t,re} = \sqrt{\frac{2kT}{R_e}}\xi_e, \tag{5.38}$$

where ξ_c, ξ_b, and ξ_e are sequences of white noise generated by the logistic map of Equation (5.31) with $\lambda = 4$. The ξ_cs have values between 0 and 1. In this case the model is based on physical mechanisms and fitting of the noise model to measurements is not required.

Based on the development in [55] and [56], when the transistor is in the forward active region, minority carriers diffuse and drift across the base into the base-collector region. Since there is an electric field, the charges undergo acceleration when they enter the collector-base depletion region and are swept to the collector. This is a random process and is the source of shot noise in the collector. Recombination effects in the base-emitter region and carrier injection from the base into the emitter are also random processes contributing to shot noise in the base and emitter, respectively. Thus there are three shot noise current sources, $i_{\mathrm{s,ce}}$, $i_{\mathrm{s,be}}$, and $i_{\mathrm{s,bc}}$, which are proportional to the instantaneous collector-emitter, base-emitter, and base-collector currents, respectively. They are modeled as follows:

$$i_{s,ce} = \sqrt{e|i_{ce}|}\xi_{ce}, \quad i_{s,be} = \sqrt{e|i_{be}|}\xi_{be}, \quad \text{and} \quad i_{s,bc} = \sqrt{e|i_{bc}|}\xi_{bc}, \tag{5.39}$$

where ξ_{ce}, ξ_{be}, and ξ_{bc} are white noise sequences between 0 and 1 and generated by the logistics map with $\lambda = 4$, and e is the elementary charge.

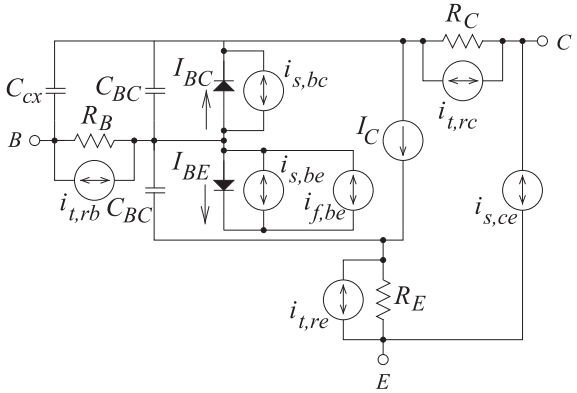


Figure 5-49: The Gummel-Poon BJT model, along with noise sources.

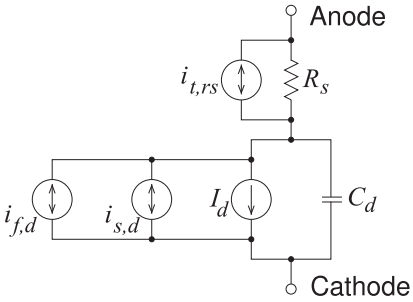


Figure 5-50: The p-n junction diode model with noise source.

Although there is more than one known source of flicker noise in a BJT [57], it has been shown that the dominant source of flicker noise can be modeled as a single noise current between the base and emitter terminals. It is a function of the instantaneous base-emitter recombination current [58] and so the flicker noise source is

$$i_{f,be} = k_f \sqrt{|i_{be}|^\alpha} \xi_f. \tag{5.40}$$

Here ξ_f is between 0 and 1 and is a sequence generated by the logarithmic map in Equation (5.34) and α controls the dependence of the flicker noise component on the non-ideal base current; here α is set to 2. The logarithmic map sequence, ξ_f, is parameterized by β, which was fit to noise measurements yielding $\beta = 0.000005$. The other parameter, k_f, sets the amplitude of the flicker noise and fitting to noise data yielded $k_f = 0.001$. The same flicker noise parameters were used with both BJTs. All of the noise sources were uncorrelated and this was achieved by randomly choosing initial seeds of each sequence, the ξs.

The diode model is shown in Figure 5-50. The noisy model of the diode includes a thermal current source for the parasitic resistance, a shot noise current source, and a flicker noise current source that is dependent on the current flowing through the diode [56]. The diode's thermal, shot, and flicker noise current sources are

$$i_{t,rs} = \sqrt{\frac{2kT}{R}} \xi_t, \quad i_{s,d} = \sqrt{e i_d} \xi_s, \quad \text{and} \quad i_{f,d} = k_{fd} \sqrt{i_d^\alpha} \xi_f, \tag{5.41}$$

respectively. Here the parameter k_{fd} is the scaling coefficient for the diode flicker noise and α controls the dependence of the flicker noise component on the current in the diode. As with the BJT, $\alpha = 2$ in the simulations reported here. The parameter β controls the slope of the autocorrelation characteristic. As before, the random variables ξ_t and ξ_s, describing the thermal and shot noise processes, were generated using the logistic map, and ξ_f, describing the flicker noise process, was generated using the logarithmic map. The amplitude of the flicker noise source was fit to measurements and the same $\beta = 0.00005$ parameter was used as was determined for the BJT model.

The noise model of the oscillator is completed by modeling the thermal noise current of each resistor as in Equation (5.38).

Oscillator Simulation

The VCO circuit of Figure 5-38 was simulated in the time domain using the transient simulator described in [25] and [26]. In the circuit model there are a total of three flicker noise parameters fit to measurements, k_f for the BJTs, k_{fd}

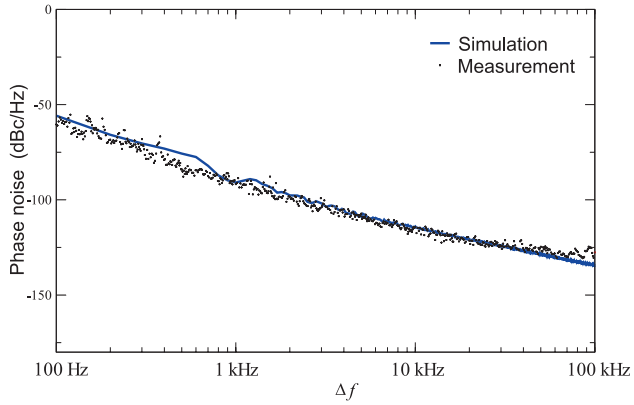


Figure 5-51: Phase noise comparison between data and experiment with bias voltage at 0 V.

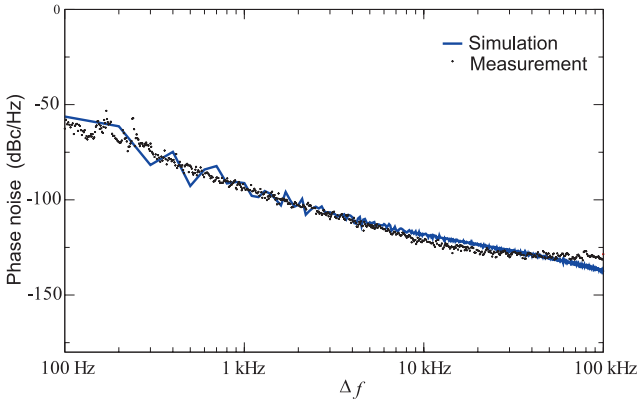


Figure 5-52: Phase noise comparison between data and experiment with bias voltage at 6 V.

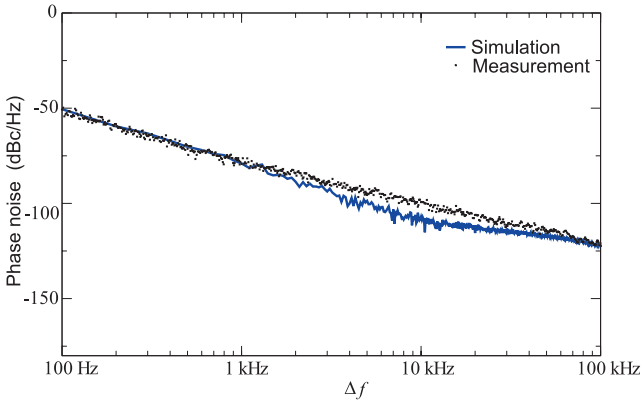


Figure 5-53: Phase noise comparison between data and experiment with bias voltage at 12 V.

for the diode, and the same value of β was used for the BJTs and the diode. Thus there are three noise parameters to be set in the VCO circuit. These were unchanged in simulations of the VCO with different varactor bias voltages. These are the only noise parameters that are not fixed by the device currents and parasitic resistance values. The phase noise output following simulation was Fourier analyzed yielding the phase noise results shown in Figures 5-51 to 5-53. These measured results were also presented in Figures 5-39 and 5-40, where the phase noise slopes were identified as f^0, f^{-1}, f^{-2}, and f^{-3}. So the simulated phase noise results in Figures 5-51 to 5-53 correctly calculate the various phase noise slopes and crossover frequencies for diverse varactor tuning conditions.

5.10 Summary

This chapter focused on the design of two categories of microwave oscillators: fixed-frequency oscillators and VCOs. All microwave oscillators can be considered to be an amplifier with a feedback network with the amplifier establishing the amplitude of oscillation and the feedback network setting the frequency of oscillation. As well, nearly all microwave oscillators can be considered as reflection oscillators in which the active device, with appropriate feedback, presents a negative resistance, or equivalently, a negative conductance to a linear frequency-selective circuit. Whether design based on negative conductance or negative resistance is used depends on which, resistance or conductance, reduces in magnitude as the signal level increases. Since transistors are essentially voltage-controlled current sources, and the current tends to saturate at high output levels, the natural view is to consider transistor-based oscillators as having a negative conductance that reduces in magnitude as the signal level increases. With negative-conductance reflection oscillators part or all of the feedback network appears as a linear two-terminal, that is one-port, circuit in parallel with the negative conductance from the active device. This two-terminal circuit is often called the resonator or tank circuit.

The chapter presented two case studies of oscillator design. This enabled design decisions to be illustrated that could not be presented in an algorithmic way. While the design case studies considered common-base Colpitts designs, the principles apply to other types of oscillators. Still the Colpitts common-base/common-drain oscillator core has proved to yield microwave VCOs with stellar performance. RFICs necessarily use differential circuits, but even then they are conveniently designed as negative conductance oscillators. Treating the design problem as that of interconnected one-port circuits considerably simplifies making design trade-offs, and so makes it easier to achieve an optimal VCO.

VCO design is the most complicated of microwave designs with significant trade-offs of low phase noise, stability, broad tuning range, rapid turn-on transient, high output power, and high efficiency. Several decades ago fixed-frequency oscillators were most common. With these the feedback network incorporates a high-Q resonant element that results in the phase noise on the oscillating signal being insignificant in nearly all microwave applications. The high-Q resonator of the fixed-frequency oscillators largely assures single frequency of operation, and the desired constant amplitude output is assured by ensuring the active device enters saturation. With a VCO, stability is not as easy to achieve. Stability here refers to the oscillator producing a sinewave of a single frequency and constant amplitude.

With VCOs the resonator is nearly always of relatively low Q, as it must be variable. Then a major concern in oscillator design is phase noise appearing on the output signal. Managing phase noise is complicated because the origins of phase noise are not well known and so physically based device models, that would enable the accurate determination of phase noise in computer-based RF circuit simulation, are not available. For noncompetitive VCOs there are many noise sources other than the intrinsic phase noise of an active device that dominate the phase noise of the output signal. For example, it is known that up-conversion of low-frequency noise, and down-conversion of harmonic noise, produce oscillator phase noise. However,

once these sources of phase noise are minimized in design, there is a remaining intrinsic phase noise component. It is this phase noise, intrinsic to active devices, that is the concern of competitive VCO design. While the origins of intrinsic phase noise in well-designed oscillators is not completely understood, there are best practices to follow that minimize the coupling of external noise to the oscillating signal. External noise at low frequencies will be up-converted unless care is taken. Good design practice is to eliminate low-frequency noise from the power supply and from the tuning voltage source for VCOs. Some more specific guidelines:

1. Good grounding is required with decoupling capacitors between the supply and ground.
2. Attention must be given to the signal return path for the supply and tuning voltage source to avoid common impedance coupling. Any noise on the tuning voltage in a VCO will result in phase noise on the VCO output. In a microstrip circuit, only minimum metal on the microstrip layer should be removed with floating metal connected to the ground through multiple vias. This suppresses substrate modes, minimizes parasitic coupling, and minimizes electro-thermal passive intermodulation distortion.
3. The oscillator output should drive a resistive load and it is common to use a resistive pad (i.e., attenuator) followed by a bandpass filter. This reduces the effect of the load on the oscillation frequency. Also it is common to isolate the tank circuit from the load as was seen in the case studies presented in this chapter.
4. Internal oscillator paths should be as small as possible to minimize the coupling of noise from the environment.

A VCO can be incorporated in a phase-locked loop which further reduces phase-noise and locks the oscillator to a low-frequency highly stable frequency reference. The oscillation frequency is then controlled by the fractional frequency division in the phase-locked loop.

5.11 References

[1] M. Odyniec, Ed., *RF and Microwave Oscillator Design*. Artech House, 2002.

[2] A. Rohde, U.L. Poddar and G. Bock, *The Design of Modern Microwave Oscillators for Wireless Applications*. Wiley, 2005.

[3] E. Faulkner, *Introduction to the Theory of Linear Systems*. Chapman & Hall, 1969.

[4] A. Pippard, *Response and Stability*. Cambridge University Press, 1985.

[5] L. Dai and R. Harjani, *Design of High Performance CMOS Voltage-Controlled Oscillators*. Kluwer Academic Publishers, 2003.

[6] M. Tiebout, *Low Power VCO Design in CMOS*. Springer, 2006.

[7] A. Aktas and M. Ismail, *CMOS PLLs and VCOs for 4G wireless*. Kluwer, 2004.

[8] D. Pederson and K. Mayaram, *Analog Integrated Circuits for Communication: Principles, Simulation and Design*. Springer, 2008.

[9] B. Razavi, *Design of Analog CMOS Integrated Circuits*. McGraw-Hill, 2001.

[10] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2004.

[11] K. Kurokawa, "Some basic characteristics of broadband negative resistance oscillator circuits," *The Bell System Technical J.*, vol. 48, no. 69, pp. 1937–1955, Jul.-Aug. 1969.

[12] G. Gonzalez, *Foundations of Oscillator Circuit Design*. Artech House, 2007.

[13] C. Rauscher, "Large-signal technique for designing single-frequency and voltage-controlled GaAs FET oscillators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 29, no. 4, pp. 293–304, Apr. 1981.

[14] R. Jackson, "Criteria for the onset of oscillation in microwave circuits," *IEEE Trans. On Microwave Theory and Techniques*, vol. 40, no. 3, pp. 566–569, Mar. 1992.

[15] D. Maclean, *Evaluating Feedback in Amplifiers*

and Oscillators: Theory, Design and Analogue Applications. Research Studies Press LTD., 2004.

[16] J. Boyles, "The oscillator as a reflection amplifier: an intuitive approach to oscillator design," *Microwave Journal*, vol. 29, no. 6, pp. 83–98, Jun. 1986.

[17] D. Esdale and M. Howes, "A reflection coefficient approach to the design of one-port negative impedance oscillators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 29, no. 8, pp. 770–776, Aug. 1981.

[18] G. Gonzalez and O. Sosa, "On the design of a series-feedback network in a transistor negative-resistance oscillator," *IEEE Trans. on Microwave Theory and Techniques*, vol. 47, no. 1, pp. 42–47, Jan. 1999.

[19] A. Victor and M. Steer, "Reflection coefficient shaping of a 5-GHz voltage-tuned oscillator for improved tuning," *IEEE Trans. on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2488–2494, Dec. 2007.

[20] A. Knights and M. Kelly, "Laterally stacked varactor formed by ion implantation," *Electronics Letters*, vol. 35, no. 10, pp. 846–847, May 1999.

[21] M. Steer, *Microwave and RF Design, Modules*, 3rd ed. North Carolina State University, 2019.

[22] P. J. Topham, A. Dearn, and G. Parkinson, "GaAs bipolar broadband oscillators," in *21st European Microwave Conf.*, Sep. 1991, pp. 178–183.

[23] http //www.febo.com.

[24] A. Victor, J. Nath, D. Ghosh, B. Boyette, J. Maria, M. Steer, A. Kingon, and G. Stauf, "Noise characteristics of an oscillator with a barium strontium titanate (BST) varactor," *IEE Proc., Part H, Microwaves, Antennas and Propagation*, vol. 153, no. 1, pp. 96–102, 2006.

[25] N. Kriplani, A. Victor, and M. Steer, "Time-domain modelling of phase noise in an oscillator," in *36th European Microwave Conf.*, 2006, pp. 514–517.

[26] N. Kriplani, S. Luniya, and M. Steer, "Integrated deterministic and stochastic simulation of electronic circuits: Application to large signal–noise analysis," *Int. Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 21, no. 6, pp. 381–394, 2008.

[27] A. Victor, "Microwave power oscillator utilizing thin-film ferroelectic varactors," Ph.D. dissertation, North Carolina State University, 2010.

[28] D. Leeson, "A simple model of feedback oscillator noise spectrum," *Proc. of the IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.

[29] T. Lee and A. Hajimiri, "Oscillator phase noise: a tutorial," *IEEE J. of Solid-State Circuits*, vol. 35, no. 3, pp. 326–336, Mar. 2000.

[30] S. Kogan, *Electronic Noise and Fluctuations in Solids*. Cambridge University Press, 1996.

[31] A. Hajimiri and T. Lee, "A general theory of phase noise in electrical oscillators," *IEEE J. of Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, Feb. 1998.

[32] ——, "Design issues in cmos differential lc oscillators," *IEEE J. of Solid-State Circuits*, vol. 34, no. 5, pp. 717–724, May 1999.

[33] ——, "Corrections to "a general theory of phase noise in electrical oscillators"," *IEEE J. of Solid-State Circuits*, vol. 33, no. 6, p. 928, Jun. 1998.

[34] G. Wornell and A. Oppenheim, *Signal Processing with Fractals: A Wavelet-Based Approach*. Prentice Hall, 1996.

[35] G. Wornell, "Wavelet-based representations for the 1/f family of fractal processes," *Proc. of the IEEE*, vol. 81, no. 10, pp. 1428–1450, Oct. 1993.

[36] N. Kriplani, "Modelling colored noise under large-signal conditions," Ph.D. dissertation, North Carolina State University, 2005.

[37] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysiological (fmri) time series analysis: resampling methods in time and wavelet domains," *Human Brain Mapping*, vol. 12, no. 2, pp. 61–78, 2000.

[38] B. Mandelbrot, *The Fractal Geometry of Nature*. Times Books, 1982.

[39] B. Mandelbrot and J. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[40] M. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. Dover Publications, 2009.

[41] R. Voss, *Fractals in Nature: From Characterization to Simulation*. Springer-Verlag, 1988.

[42] R. Devaney, *A First Course in Chaotic Dynamical Systems*. Perseus Book, 1992.

[43] Y. Pomeau and P. Manneville, "Intermittent transition to turbulence in dissipative dynamical systems," *Commun. Math. Phys*, vol. 74, no. 2, pp. 189–197, 1980.

[44] R. Bhansali, M. Holland, and P. Kokoszka, "Chaotic maps with slowly decaying correlations and intermittency," in *Fields Institute Communications: Asymptotic Methods in Stochastics*, L. Horváth and B. Szyszkowicz, Eds., 2004, p. ●.

[45] R. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol.

261, no. 5560, pp. 459–467, June 1976.

[46] T. Kawabe and Y. Kondo, "Intermittent chaos generated by logarithmic map," *Progress of Theoretical Physics*, vol. 86, no. 3, pp. 581–586, 1991.

[47] M. Holland, "Slowly mixing systems and intermittency maps," *Ergodic Theory and Dynamical System*, vol. 25, no. 1, pp. 133–159, Feb. 2005.

[48] E. Simoen, A. Mercha, L. Pantisano, C. Claeys, and E. Young, "Low-frequency noise behavior of SiO2–HfO2 dual-layer gate dielectric nMOSFETs with different interfacial oxide thickness," *IEEE Trans. on Electron Devices*, vol. 51, no. 5, pp. 780–784, May 2004.

[49] F. Hooge, T. Kleinpenning, and L. Vandamme, "•experimental studies on 1/f noise," *Reports on Progress in Physics*, 1981.

[50] G. Uhlenbeck and L. Ornstein, "On the theory of brownian motion," *Physical Review*, vol. 36, pp. 823–841, Sep. 1930.

[51] E. Bibbona, G. Panfilo, and P. Tavella, "The ornstein-uhlenbeck process as a model of a low pass filtered white noise," *Metrologia*, vol. 45, no. 6, pp. S117–S126, Dec. 2008.

[52] A. Suárez, *Analysis and Design of Autonomous Microwave Circuits*. John Wiley & Sons, Inc., 2009.

[53] A. Hati, C. Nelson, and D. Howe, "Effect of vibration on P and AM noise of oscillatory and non-oscillatory components at 10 GHz," in *2009 IEEE Int. Frequency Control Symp., Joint with the 22nd European Frequency and Time Forum*, Apr. 2009, pp. 524–529.

[54] W. Robins, *Phase Noise in Signal Sources*. Peter Peregrinus Ltd., 1982.

[55] M. Buckingham, *Noise in Electronic Devices and Systems*. Ellis Horwood, 1983.

[56] P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE*. McGraw-Hill, 1988.

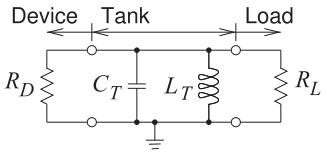
[57] A. van der Ziel, X. Zhang, and A. Pawlikiewicz, "Location of 1/f noise sources in BJT's and HBJT's—i. theory," *IEEE Trans. on Electron Devices*, vol. 33, no. 9, pp. 1371–1376, Sep. 1986.

[58] C. Green and B. Jones, "1/f noise in bipolar transistors," *J. Phys. D: Appl. Phys.*, vol. 18, pp. 77–91, 1985.

5.12 Exercises

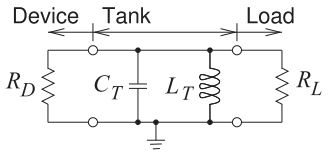
1. Draw the schematic of an opamp-based Hartley oscillator circuit.

2. Consider the circuit below. This is the equivalent circuit of an active device with a negative conductance $= 1/R_D$ connected to a tank circuit comprising capacitor C_T and inductor L_T, and then a load R_L. The oscillation amplitude will adjust so that $R_D = R_L$.

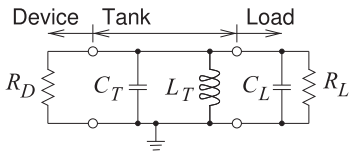


   (a) Write down a formula for the oscillation frequency f_0.
   (b) What is f_0 if $C_T = 1$ pF and $L_T = 1$ nH?

3. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit $C_T = 0.1$ pF and $L_T = 0.5$ nH. What is the oscillation frequency assuming that there is sufficient negative conductance for oscillation to occur?



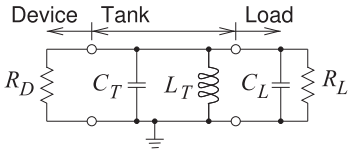
4. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor C_L in parallel with a resistor R_L. The oscillation amplitude will adjust so that $R_D = R_L$.



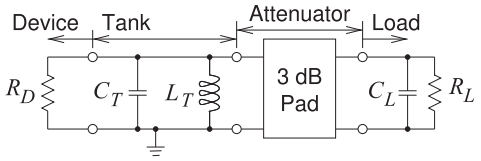
   (a) Write down a symbolic formula for the oscillation frequency f_0.
   (b) What is f_0 if $C_L = 0.2$ pF and $R_L = 50\ \Omega$?

5. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit $C_T = 0.1$ pF and $L_T = 0.5$ nH. The capacitance of the load is $C_L = 0.05$ pF and the load resistance is $50\ \Omega$. What is the oscillation frequency assuming that

there is sufficient negative conductance for oscillation to occur?

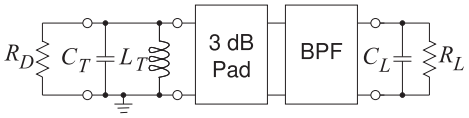


6. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor $C_L = 0.2$ pF in parallel with a resistor $R_L = 50\ \Omega$. Between the tank circuit is a 3 dB 50 Ω attenuator (i.e., the system is designed for system impedance of 50 Ω). The oscillation amplitude will adjust so that $R_D = R_L$.



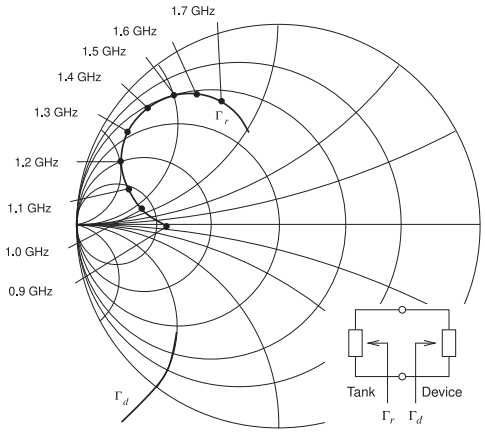
(a) Ignore the load capacitor C_L, what is the frequency of oscillation, f_0, of the oscillator?
(b) From now on consider the load capacitance. What is the oscillation frequency without the attenuator?
(c) At the frequency calculated in (a), what is the admittance looking into the attenuator from the tank circuit with the load comprising C_L and R_L? [Hint use the resistive PI network equivalent of the attenuator.]
(d) What is the equivalent shunt resistor and capacitor circuit looking into the attenuator from the oscillator? This is the effective load seen by the active device and tank circuit.
(e) What is the oscillation frequency with the attenuator and the RC load?
(f) Using your results above, discuss the effect of the attenuator on reducing the sensitivity of oscillation on the loading conditions.

7. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor $C_L = 0.2$ pF in parallel with a resistor $R_L = 50\ \Omega$. Between the tank circuit is a 3 dB 50 Ω attenuator (ti.e., the system is designed for system impedance of 50 Ω) and a bandpass filter with an insertion loss at the oscillation frequency of 2 dB. The oscillation amplitude will adjust so that $R_D = R_L$.



(a) Ignore the load capacitor C_L, what is the frequency of oscillation, f_0, of the oscillator?
(b) From now on consider the load capacitance. What is the oscillation frequency without the attenuator and bandpass filter?
(c) At the frequency calculated in (a), what is the admittance looking into the attenuator from the tank circuit with the bandpass filter and the load comprising C_L and R_L? [Hint: Consider that the effect of the insertion loss of the filter is an attenuation. Consider using a resistive PI network equivalent of the attenuator plus bandpass filter.]
(d) What is the equivalent shunt resistor and capacitor circuit looking into the attenuator for the oscillator? This is the effective load seen by the active device and tank circuit.
(e) What is the oscillation frequency with the attenuator and the RC load?

8. A reflection oscillator is shown below along with the Γ_r and Γ_d loci plotted on a Smith chart. Γ_r, the reflection coefficient looking into the resonator, rotates clockwise as frequency increases. Γ_d, the reflection coefficient of the active device, is amplitude dependent but is frequency independent. The arrowed Γ_d curve plots the loci of Γ_d as amplitude increases. [Hint: Consider Figure 5-18.]



(a) On the Γ_r plane show the region of the Smith Chart identifying oscillation.
(b) On the Γ_d plane show the region of the Smith Chart identifying oscillation.
(c) What is the frequency of oscillation?

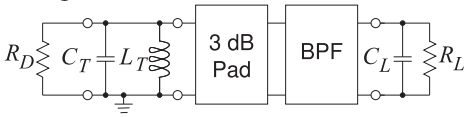
9. Consider an oscillator that can be modeled as shunt connections of a linear conductance, G_r,

a linear susceptance, B_r, a nonlinear or device conductance, G_d, and a nonlinear or device susceptance, B_d.

(a) Draw the circuit.
(b) Using G_r, B_r, G_d, and B_d, write down the Kurokawa oscillator condition that establishes stable single frequency oscillation.
(c) What are common design choices G_r, B_r, G_d, and B_d made that simplify the Kurokawa oscillator condition? Write down the resulting simplified Kurokawa oscillation condition.

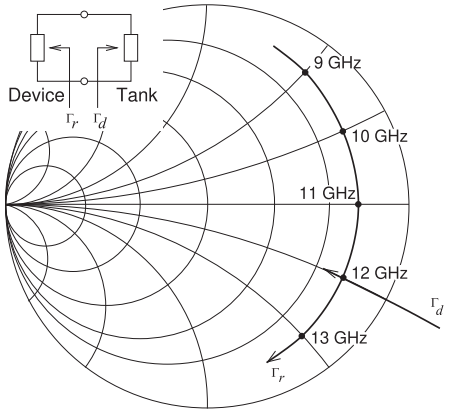
10. An oscillator has a linear conductance, G_r, a linear susceptance, B_r, a nonlinear or device conductance, G_d, and a nonlinear or device susceptance, B_d, all in shunt. The reflection coefficient looking into the linear network is Γ_r and the reflection coefficient looking into the device is Γ_d.
    (a) Use a Smith chart sketch to show the loci of the linear and nonlinear admittances (or equivalently their reflection coefficients Γ_r and Γ_d) for the case when loss in the linear network is low (i.e. $G_r \approx 0$). Do this when the oscillating signal is small and when it has reached steady state. Indicate the steady-state oscillation point. (You can use the negative or the inverse of either Γ_r or Γ_d as appropriate.)
    (b) Use a Smith chart sketch to show the loci of the linear and nonlinear admittances (or equivalently their reflection coefficients for the case when loss in the linear network is high (i.e. G_r cannot be ignored). Do this when the oscillating signal is small and when it has reached steady state. (You can use the negative of the inverse of either Γ_r or Γ_d as appropriate.)

11. The equivalent circuit of an oscillator is shown below with $R_L = 50\ \Omega$, $C_T = 1$ pF and inductance $L_T = 0.1$ nH.



    (a) What is the oscillation frequency assuming that there is sufficient negative conductance for oscillation to occur?
    (a) What is R_D when there is oscillation?

12. A reflection oscillator is shown below along with the Γ_r and Γ_d loci plotted on a Smith chart. Γ_r, the reflection coefficient looking into the resonator, rotates clockwise as frequency increases. Γ_d, the reflection coefficient of the active device, is amplitude dependent but is frequency independent. The arrowed curve plots the loci of Γ_d as amplitude increases. [Hint: Consider Figure 5-18.]



    (a) On the Γ_r plane show the region of the Smith Chart identifying oscillation.
    (b) On the Γ_d plane show the region of the Smith Chart identifying oscillation.
    (c) What is the frequency of oscillation?

13. The case study presented in Section 5.6 described the design of a 5 GHz VCO. The output of the oscillator was followed by a resistive Pi attenuator and then a bandpass filter. What is the attenuation (in decibels) and the system impedance of the attenuator. [Hint: You will need to carefully read Section 5.6.]

14. Consider a common emitter BJT Clapp oscillator.
    (a) Draw the schematic of the oscillator without the bias circuit.
    (b) Redraw the oscillator circuit including the bias circuit.

15. Consider a BJT Hartley oscillator.
    (a) Draw the schematic of the oscillator circuit in the common-base configuration. Do not show the bias circuit.
    (b) Redraw the schematic of the oscillator circuit in the common-base configuration, this time showing the bias circuit.

16. Consider a FET Clapp oscillator.
    (a) Draw the schematic of the oscillator circuit in the common drain configuration. Do not show the bias circuit.
    (b) Redraw the schematic of the oscillator circuit, this time showing the bias circuit.

17. Consider a FET Hartley oscillator.
    (a) Draw the schematic of the oscillator circuit in the common source configuration. Do not show the bias circuit.

(b) Redraw the schematic of the oscillator circuit, this time showing the bias circuit.

18 Consider a common source FET Clapp oscillator.
(a) Draw the schematic of the oscillator without biasing.
(b) Redraw the oscillator circuit including bias current sources.

19. Derive an expression for the oscillation frequency of the Colpitts BJT oscillator in the common emitter configuration shown in Figure 5-36(b).

20. A two-port feedback oscillator is shown in Figure 5-1.
(a) Draw the schematic of a feedback Colpitts oscillator.
(b) Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are C_1 = C_2 = 2 pF and L_3 = 5 nH. Ignore any phase shift introduced by the amplifier.

21. A two-port feedback oscillator is shown in Figure 5-1.
(a) Draw the schematic of a feedback Colpitts oscillator.
(b) Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are C_1 = 1 pF, C_2 = 3 pF, and L_3 = 1 nH. Ignore any phase shift introduced by the amplifier.

22. A two-port feedback oscillator is shown in Figure 5-1.
(a) Draw the schematic of a feedback Colpitts oscillator.
(b) Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are C_1 = 5 pF, C_2 = 1 pF, and L_3 = 10 nH? Ignore any phase shift introduced by the amplifier.

23. A negative-gm differential FET VCO, as shown in Figure 5-34, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a 1 kΩ load. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate. What is the oscillation frequency of the oscillator?

24. A negative-gm differential FET VCO, as shown in Figure 5-34, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a $50\ \Omega$ differential load that is in parallel with a 0.5 pF capacitor. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate.
(a) Draw the schematic of the oscillator with the load.
(b) Draw the equivalent tank circuit of the oscillator. This will need to include the effect of the load.
(c) What is the oscillation frequency of oscillator?

25. A negative-gm differential FET VCO, as shown in Figure 5-34, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a $50\ \Omega$ differential load that is in parallel with a 1 pF capacitor. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate. What is the oscillation frequency of the oscillator?

26. A common emitter Colpitts oscillator, as shown in Figure 5-36(a), has $C_1 = 1$ pF, $C_2 = 2$ pF, and $L_3 = 2$ nH. L_C is a choke inductor. $V_{CC} = 5$ V and the circuit is biased so that $g_m = 1$ mS. Ignore the internal parasitics of the transistor. What is the oscillation frequency of the oscillator? [Parallels Example 5.2]

27. A common emitter Colpitts oscillator, as shown in Figure 5-36(a), has $C_1 = 0.1$ pF, $C_2 = 0.2$ pF, and $L_3 = 0.5$ nH. L_C is a choke inductor. $V_{CC} = 5$ V and the circuit is biased so that $g_m = 1$ mS. Ignore the internal parasitics of the transistor. What is the oscillation frequency of the oscillator? [Parallels Example 5.2]

28. A digital communication system has a symbol rate of 1 MS/s.
(a) What is the highest frequency phase noise that will affect the bit error rate of the communication system? (It could be 1 MHz, 2 MHZ, no limit, 0.5 MHz, etc.)
(b) What changes would you make to the system (e.g., added components such as an attenuator, filter, amplifier, etc.) that will reduce the impact of high-offset phase noise?

29. A QPSK communication system has a transmitted bit rate of 100 kbit/s. Consider that the QPSK modulation scheme is ideal.
(a) What is the symbol rate?

(b) What would you do to the system (e.g., added components) to reduce the impact of high-offset phase noise?

30. A QPSK communication system has a transmitted bit rate of 1 Mbit/s. The QPSK modulation scheme is ideal (2 bit/s/Hz. DSP processing is such that phase noise slower than the duration of 5 symbols has no affect on the communication system throughput.
    (a) What is the symbol rate?
    (b) What is the lowest frequency phase noise that will affect the communication system?

31. An oscillator has phase noise that reduces away from the oscillation center frequency, f_{osc} at the rate of $1/\Delta f^2$ where Δf is the frequency offset from f_{osc}. If the phase noise at 100 MHz frequency offset from the oscillation frequency is -100 dBc/Hz, what is the phase noise at 10 MHz?

32. The phase noise of an oscillator measured at 1 MHz is -136 dBc/Hz. If the phase noise varies as $1/(\Delta f)$, where Δf is the offset from the center frequency of oscillation, what is the phase noise at 100 kHz offset?

33. A phase-locked microwave oscillator typically utilizes a low-Q oscillator. For such an oscillator the phase noise at the frequency that affects microwave systems has an inverse square relationship to frequency. The phase noise measured at 100 kHz is -106 dBc/Hz, what is the phase noise referred to 1 MHz?

34. A phase-locked microwave oscillator typically utilizes a low-Q oscillator. For such an oscillator the phase noise at the frequency that affects microwave systems often has an inverse square relationship to frequency. The phase noise measured at 1 MHz is -125 dBc/Hz, what is the phase noise at 100 kHz?

5.12.1 Exercises by Section

†challenging, ‡very challenging

§5.2 1
§5.3 $2^{\dagger}, 3^{\dagger}, 4^{\dagger}, 5^{\dagger}, 6^{\ddagger}, 7^{\ddagger}, 8, 9, 10, 11$
§5.5 12
§5.6 $13^{\dagger}$
§5.7 $14^{\dagger}, 15^{\dagger}, 16^{\dagger}, 17^{\dagger}, 18^{\dagger}, 19^{\dagger}, 20^{\dagger}, 21^{\dagger}, 22^{\dagger}, 23^{\dagger}, 24^{\dagger}, 25^{\dagger}, 26^{\dagger}, 27^{\dagger}$
§5.8 $28^{\dagger}, 29^{\dagger}, 30^{\dagger}, 31, 32^{\dagger}, 33^{\dagger}, 34^{\dagger}$

5.12.2 Answers to Selected Exercises

4 4.594 GHz
6(d) 49.8 Ω, 99.6 fF
7(e) 4.883 MHz
13 49.8 Ω, 2.97 dB
19 $\sqrt{\frac{C_1 + C_2}{L_3 C_1 C_2}}$
20 1.592 GHz
23 25.16 GHz
24(c) 10.27 GHz
30 100 kHz
34 -105 dBc/Hz

Index

MICROWAVE AND RF DESIGN: AMPLIFIERS AND OSCILLATORS Third Edition

Microwave and RF Design: Amplifiers and Oscillators presents the design of amplifiers and oscillators in a way that enables state-of-the-art designs to be realized. Detailed strategies and case studies are presented. Design of competitive microwave amplifiers and oscillators is particularly challenging as many trade-offs are required in design, and the design decisions cannot be reduced to a formulaic flow. The emphasis is on developing design skills. This book is suitable as both an undergraduate and graduate textbook, as well as a career-long reference book.

KEY FEATURES

- The fifth volume of a comprehensive series on microwave and RF design
- Open access ebook editions are hosted by NC State University Libraries at: https://repository.lib.ncsu.edu/handle/1840.20/36776
- 9 worked examples
- An average of 23 exercises per chapter
- Answers to selected exercises
- 6 extensive case studies following the design of competitive amplifiers and oscillators with world leading performance
- A companion book, *Fundamentals of Microwave and RF Design*, is suitable as a comprehensive undergraduate textbook on microwave engineering

ABOUT THE AUTHOR

Michael Steer is the Lampe Distinguished Professor of Electrical and Computer Engineering at North Carolina State University. He received his B.E. and Ph.D. degrees in Electrical Engineering from the University of Queensland. He is a Fellow of the IEEE and is a former editor-in-chief of *IEEE Transactions on Microwave Theory and Techniques*. He has authored more than 500 publications including twelve books. In 2009 he received a US Army Medal, "The Commander's Award for Public Service." He received the 2010 Microwave Prize and the 2011 Distinguished Educator Award, both from the IEEE Microwave Theory and Techniques Society.

OTHER VOLUMES

Microwave and RF Design
Radio Systems
Volume 1
ISBN 978-1-4696-5690-8

Microwave and RF Design
Transmission Lines
Volume 2
ISBN 978-1-4696-5692-2

Microwave and RF Design
Networks
Volume 3
ISBN 978-1-4696-5694-6

Microwave and RF Design
Modules
Volume 4
ISBN 978-1-4696-5696-0

ALSO BY THE AUTHOR

Fundamentals of Microwave and RF Design
ISBN 978-1-4696-5688-5

Published by NC State University

NC STATE UNIVERSITY

Distributed by UNC Press