

ELG 6369

NONLINEAR MICROWAVE DEVICES AND EFFECTS

CHAPTER IV

MICROWAVE AMPLIFIERS

Amplification is one of the basic functions in analog electronics. The task of designing a microwave amplifier consist first to ensure that the active device around which design is done satisfy the circuit requirements. Therefore, the active device as a designable part of the circuit can

- Meet performance specifications that might otherwise be unattainable,
- Afford improved margins of process tolerance so that yield increases and
- Potentially save on the number of elements or transistors required in matching networks, thus leading to lower circuit cost and small size.

In the microwave area, such circuits can be built around diodes or transistors. Following the power criteria, amplifiers can be classified into three groups: low-gain or low-noise amplifiers (equivalent to a linear circuit), medium-gain amplifiers (designed through a quasi-linear approach) and high-gain or power amplifiers (designed through a nonlinear approach).

In this chapter, we will emphasize first on one particular aspect of nonlinear microwave circuits, namely the parametric amplification. This point is important to understand the nonlinear effects present in microwave circuits since parametric amplifiers are practically the earliest nonlinear circuits used in the microwave range.

For more simplification, we will focus on one-port reactive components like varactors *that can be modeled by a single nonlinear capacitor* in the parametric amplification mode. Reactive device means no active losses will be considered. Parametric amplifiers with a variable-capacitance diode are used in radar tracking and communications Earth stations, Earth satellite stations, and deep-space stations. The noise temperature of parametric amplifiers cooled to the temperature of liquid helium, about 5 K, is in the range of 20 to 30 K. Usual parametric amplifiers gains are about 40 dB.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

A – DEFINITION OF PARAMETRIC AMPLIFICATION

In order to highlight the parametric effect, let us consider the circuit shown on Figure IV-1 where a capacitor C is excited by a current source $i(\omega_0 t)$ or a voltage source $e(\omega_0 t)$.

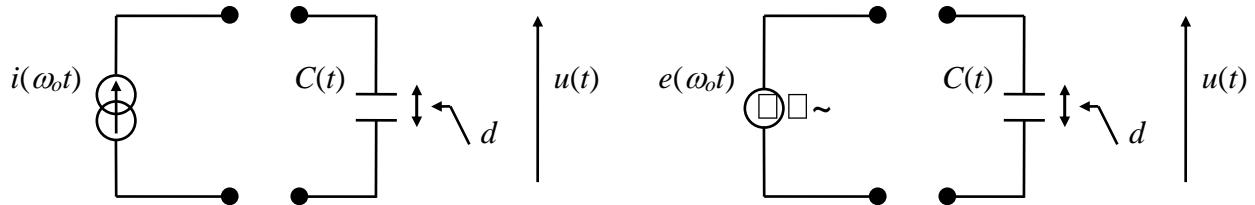


Fig. IV-1. Capacitor C excited by a current source $i(\omega_0 t)$ or a voltage source $e(\omega_0 t)$.

The capacitor consists of two conducting plates, separated by a thin insulating layer of thickness d . This thickness can be varied instantaneously with a pulsation ω_p , called the pump pulsation. Thus, the capacitance value will increase and decrease accordingly (Figure IV-2-a). If we fix the value of pulsation ω_p as twice of that of the source pulsation ω_0 , we have two possibilities:

- The two pulsations are in-phase. Thus, we can correspond the capacitance decrease (equivalent to a larger thickness d) to point M₁ where the applied voltage is at its maximum, and the capacitance increase (equivalent to a smaller thickness) to point M₂ where the voltage is equal to zero (turn back the capacitance plates to their initial position). Consequently, the corresponding voltage $u_1(\omega_p t)$ across the capacitor will always increase: This is called *parametric amplification* (Figure IV-2-b).
- The two pulsations ω_0 and ω_p are out-of-phase. We have then a “beating”, the amplitude increases and decreases periodically (Figure IV-2-c) and could reach the extreme limit where the *parametric attenuation* effect appears (Figure IV-2-d).

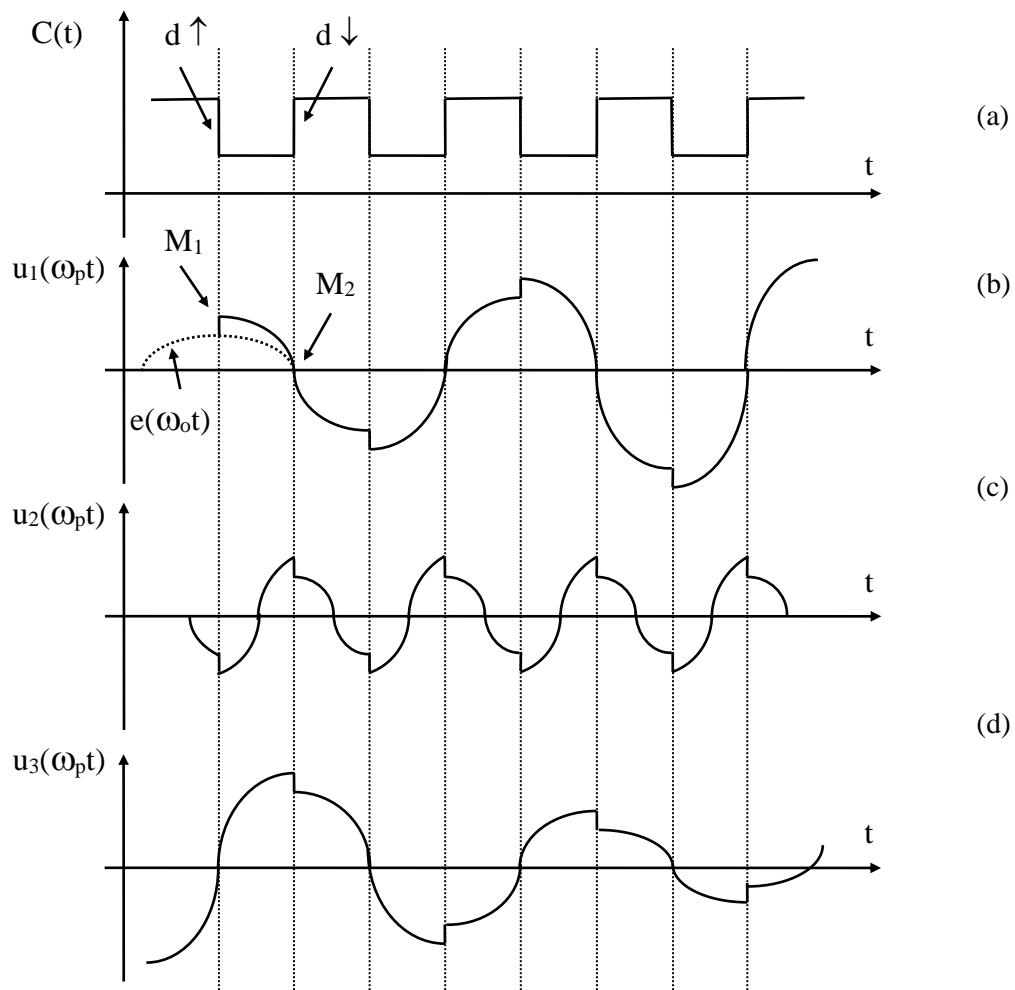


Fig. IV-2. Excitation of a capacitance $C(\omega_p t)$ by a signal ω_o :

- (a) Variation of the capacitance $C(\omega_p t)$ as function of the thickness d between the two planes
- (b) The two pulsations ω_p and ω_o are in phase \rightarrow Amplification
- (c) The two pulsations ω_p and ω_o are out-of-phase \rightarrow Interference (beating)
- (d) The two pulsations ω_p and ω_o are out-of-phase (limit case) \rightarrow Attenuation

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Since it is impossible to bound or predict neither the input phase nor its pulsation, the above phase and pulsation conditions { $\omega_p = 2 \omega_o$ } cannot be respected. Therefore, in addition to ω_p and ω_o { with $\omega_p \neq 2 \omega_o$ }, we introduce a new pulsation noted ω_i and defined by

$$\frac{\omega_p}{2} + \left\{ \frac{\omega_p}{2} - \omega_o \right\} = \omega_p - \omega_o = \omega_i \quad \text{with} \quad \frac{\omega_p}{2} - \left\{ \frac{\omega_p}{2} - \omega_o \right\} = \omega_o \quad (\text{IV-1})$$

Adding a "idler circuit" of pulsation ω_i to the original system will allow avoiding the above conditions. We can then suggest two possible configurations for a parametric amplifier in accordance with the source impedance value: excitation by a current source if the impedance is high (Figure IV-3) or by a voltage source if the impedance value is low (Figure IV-4). A parametric amplifier that is not phase-sensitive, referred to as a *non-degenerative parametric amplifier*, uses a pump circuit with a frequency higher than twice the input signal.

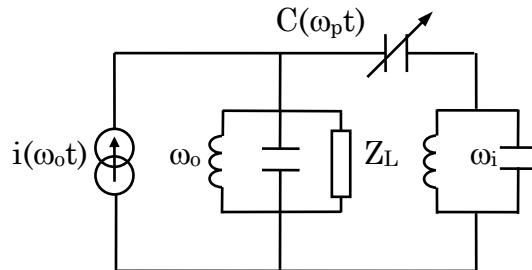


Fig. IV-3. Parametric amplifier excited by a current source.

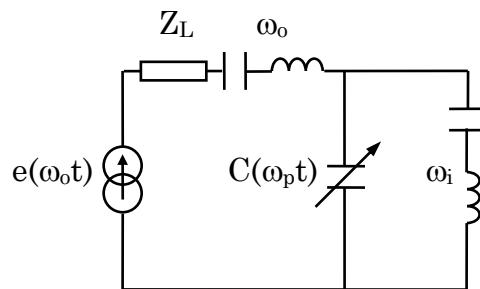


Fig. IV-4. Parametric amplifier excited by a voltage source.

B – PARAMETRIC DEVICES

Many microwave diodes exhibit a nonlinear capacitance: IMPATTs, Varactors, Schottky, ... but we will focus on the simplest one, namely the varactor diode. As shown in Figure IV-5, the dynamic excursion of the Q point varies from U_a (breakdown or avalanche voltage) to Φ (barrier voltage); the diode is electrically equivalent to a pure reactance.

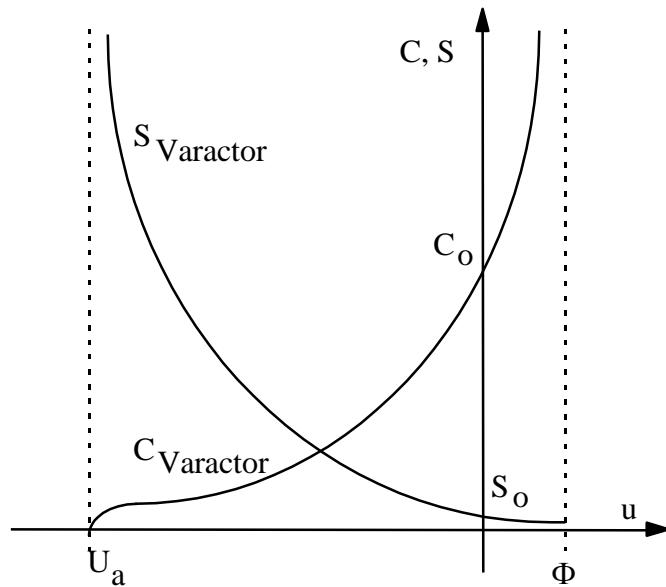


Fig. IV-5. Characteristics of the Capacitance C and elastance S of a varactor diode.

The capacitance variation $C(u(t))$ can be expressed by (chapter II, section D-V):

$$C_j(V_j) = C_{jo} \left\{ 1 - \frac{V_j(t)}{\Phi} \right\}^{-\gamma} \quad (\text{IV-2})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

but it is more convenient to use the elastance $S(u(t))$ defined as

$$S(u) = \frac{du(t)}{dq(t)} = S_o \left\{ 1 - \frac{u}{\Phi} \right\}^\gamma \quad (\text{IV-3})$$

where S_o is the value of the elastance at $\{ u = 0 \}$.

C – PARAMETRIC CIRCUIT ANALYSIS

The $C(u)$ characteristic is nonlinear, therefore the junction voltage is not a pure sinusoid. For each harmonic $k\omega$ of a pulsation ω , $u(t)$ or $q(t)$ can be written as

$$\begin{aligned} u(t) &= \sum_{k=-\infty}^{\infty} U_k e^{jk\omega t} & U_k = U_{-k}^* \\ q(t) &= \sum_{k=-\infty}^{\infty} Q_k e^{jk\omega t} & Q_k = Q_{-k}^* \end{aligned} \quad (\text{IV-4})$$

We can also express the Fourier coefficients of the elastance as

$$S(t) = \sum_{k=-\infty}^{\infty} S_k e^{jk\omega t} \quad S_k = S_{-k}^* \quad (\text{IV-5})$$

Note that

$$S_k \neq \frac{1}{C_k} \quad (\text{both in magnitude and phase})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

In such circuits, the amplitude of the pump (ω_p) is usually larger than that of the source (ω_o), so the large-signal / small-signal analysis can be applied. However, since the circuit topology is simple, it is worth to use analytical instead of numerical tools to highlight how a nonlinear circuit can be analyzed. Let us express the charge using equation (IV-3)

$$q(t) = \frac{\Phi}{S_o(1-\gamma)} \left\{ 1 - \left\{ 1 - \frac{u}{\Phi} \right\}^{1-\gamma} \right\} \quad (\text{IV-6})$$

By considering the particular values

$$q(\Phi) = Q_\Phi = \frac{\Phi}{S_o(1-\gamma)} \quad (\text{IV-7})$$

$$q(U_a) = Q_a = Q_\Phi \left\{ 1 - \left\{ 1 - \frac{U_a}{\Phi} \right\}^{1-\gamma} \right\} \quad (\text{IV-8})$$

we obtain

$$q(t) = Q_\Phi \left\{ 1 - \left\{ 1 - \frac{u}{\Phi} \right\}^{1-\gamma} \right\} \quad (\text{IV-9})$$

and, with the condition $\gamma \neq 1$,

$$\left\{ 1 - \frac{u}{\Phi} \right\}^\gamma = \left\{ 1 - \frac{q}{Q_\Phi} \right\}^{\frac{\gamma}{1-\gamma}} \quad (\text{IV-10})$$

Furthermore, using the maximum value of the elastance obtained at $u = U_a$

$$S_{\max} = S(U_a) = S_o \left\{ 1 - \frac{U_a}{\Phi} \right\}^\gamma \quad (\text{IV-11})$$

we can deduce a more convenient relationship between elastance and voltage

$$S(u) = S_o \left\{ 1 - \frac{u}{\Phi} \right\}^\gamma = S_o \frac{\left\{ 1 - \frac{U_a}{\Phi} \right\}^\gamma}{\left\{ 1 - \frac{U_a}{\Phi} \right\}^\gamma} \left\{ 1 - \frac{u}{\Phi} \right\}^\gamma = S_{\max} \left\{ \frac{u - \Phi}{U_a - \Phi} \right\}^\gamma \quad (\text{IV-12})$$

or elastance and charge

$$S(q) = S_{\max} \left\{ \frac{q - Q_\Phi}{Q_a - Q_\Phi} \right\}^{\frac{\gamma}{1-\gamma}} \quad (\text{IV-13})$$

For gradual junction ($\gamma = 1/3$):

$$S(u) = S_{\max} \sqrt[3]{\frac{u - \Phi}{U_a - \Phi}} \quad \rightarrow \quad \text{Nonlinear}$$

$$S(q) = S_{\max} \sqrt[2]{\frac{q - Q_\Phi}{Q_a - Q_\Phi}} \quad \rightarrow \quad \text{Nonlinear}$$

Both sources show nonlinear behaviors.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

For abrupt junction ($\gamma = 1/2$):

$$S(u) = S_{\max} \sqrt[2]{\frac{u - \Phi}{U_a - \Phi}} \quad \rightarrow \quad \text{Nonlinear}$$

$$S(q) = S_{\max} \frac{q - Q_\Phi}{Q_a - Q_\Phi} \quad \rightarrow \quad \text{Linear}$$

A voltage source (pump) response is nonlinear while the current one is linear: Therefore, a user can decide on linear or nonlinear pumping depending on the application.

Now, the final step will be to fully characterize the elastance, i.e., to determine the S_k coefficients of equation (IV-5). As mentioned above, since these coefficients are functions of the two pulsations ω_p and ω_o , we use the large-signal / small-signal analysis approach.

I - Large signal analysis: Excitation by the large signal only (the pump)

In the case of a current pumping, the charge could be developed in a series expansion as

$$q(t) = Q_{dc} + 2Q_{acp} \cos(\omega_p t) = Q_{dc} + Q_{acp} e^{j\omega_p t} + Q_{acp} e^{-j\omega_p t} \quad (\text{IV-14})$$

which gives

$$S(q) = S_{\max} \left\{ \frac{q - Q_\Phi}{Q_a - Q_\Phi} \right\}^{\frac{\gamma}{1-\gamma}} = S_{\max} \left\{ \frac{Q_{dc} - Q_\Phi}{Q_{dc} - Q_\Phi} \right\}^{\frac{\gamma}{1-\gamma}} \left\{ \frac{Q_{dc} + 2Q_{acp} \cos(\omega_p t) - Q_\Phi}{Q_a - Q_\Phi} \right\}^{\frac{\gamma}{1-\gamma}}$$

that can be rewritten as

$$S(q) = S_{\max} \left\{ \frac{Q_{dc} - Q_{\Phi}}{Q_a - Q_{\Phi}} \right\}^{\frac{\gamma}{1-\gamma}} \left\{ 1 + \frac{2Q_{acp}}{Q_{dc} - Q_{\Phi}} \cos(\omega_p t) \right\}^{\frac{\gamma}{1-\gamma}} \quad (\text{IV-15})$$

$$\rightarrow \quad S(q) = S(Q_{dc}) \left\{ 1 + 2\tau_q \cos(\omega_p t) \right\}^{\frac{\gamma}{1-\gamma}} \quad (\text{IV-16})$$

where τ_q is the current modulation factor (i.e., when the junction is excited by a current). A similar form can be also found using the voltage source

$$S(u) = S_{\max} \left\{ \frac{U_{dc} - \Phi}{U_a - \Phi} \right\}^{\gamma} \left\{ 1 + \frac{2U_{acp}}{U_{dc} - \Phi} \cos(\omega_p t) \right\}^{\gamma} = S(U_{dc}) \left\{ 1 + 2\tau_u \cos(\omega_p t) \right\}^{\gamma} \quad (\text{IV-17})$$

where τ_u is the voltage modulation factor (i.e., when the junction is excited by a voltage). We can notice that equations (IV-16) and (IV-17) have the same format

$$m(t) = \left\{ 1 + 2\tau \cos(\omega_p t) \right\}^{\alpha} = \sum_{k=-\infty}^{\infty} M_k e^{jk\omega_p t} \quad (\text{IV-18})$$

where, according to α { $\alpha = \gamma$ or $\alpha = (\gamma / 1 - \gamma)$ }, the M_k coefficients are equal to

$$M_k = \begin{cases} \frac{S_k}{S(Q_{dc})} \\ \text{or} \\ \frac{S_k}{S(U_{dc})} \end{cases} = \frac{1}{\pi} \int_0^{\pi} \left\{ 1 + 2\tau \cos(\omega_p t) \right\}^{\alpha} \cos(k\omega_p t) d(\omega_p t) \quad (\text{IV-19})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Setting

$$\omega_p t = \theta - \pi \quad (\text{IV-20})$$

$$x = \{1 + 4\tau\}^{-1/2} \quad (\text{IV-21})$$

the values of the integrals (deduced from the Legendre's polynomials) were calculated by Lowan (Appendix IV-1)

$$\int_0^{\pi} \left\{ x + \sqrt{x^2 - 1} \cos \theta \right\}^{\alpha} \cos(k\theta) d\theta = \frac{\pi P_{\alpha}^k(x)}{(\alpha + 1) \cdots (\alpha + k)} \quad (\text{IV-22})$$

These values allow obtaining the M_k (Fig. IV-6) since, for a given a , we can deduce all required S-parameter ratios (Table IV-1).

$$M_k = \frac{(-1)^k (1 - 4a^2)^{\frac{\alpha}{2}} P_{\alpha}^k(1 - 4a^2)^{-\frac{1}{2}}}{(\alpha + 1) \cdots (\alpha + k)} \quad (\text{IV-23})$$

Table IV-1. S-parameter ratios (in %) for $a = 0.5$

Junction	Pump	S_2 / S_1	S_3 / S_1	S_4 / S_1	S_5 / S_1	S_6 / S_1	S_7 / S_1	S_8 / S_1	S_9 / S_1	S_{10} / S_1
G	U	28.6	14.3	8.8	6.7	4.5	3.4	2.8	2.3	1.9
G	Q	20.0	8.6	4.8	3.1	2.0	1.5	1.2	0.9	0.8
A	U	20.0	8.6	4.8	3.1	2.0	1.5	1.2	0.9	0.8
A	Q	-	-	-	-	-	-	-	-	-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

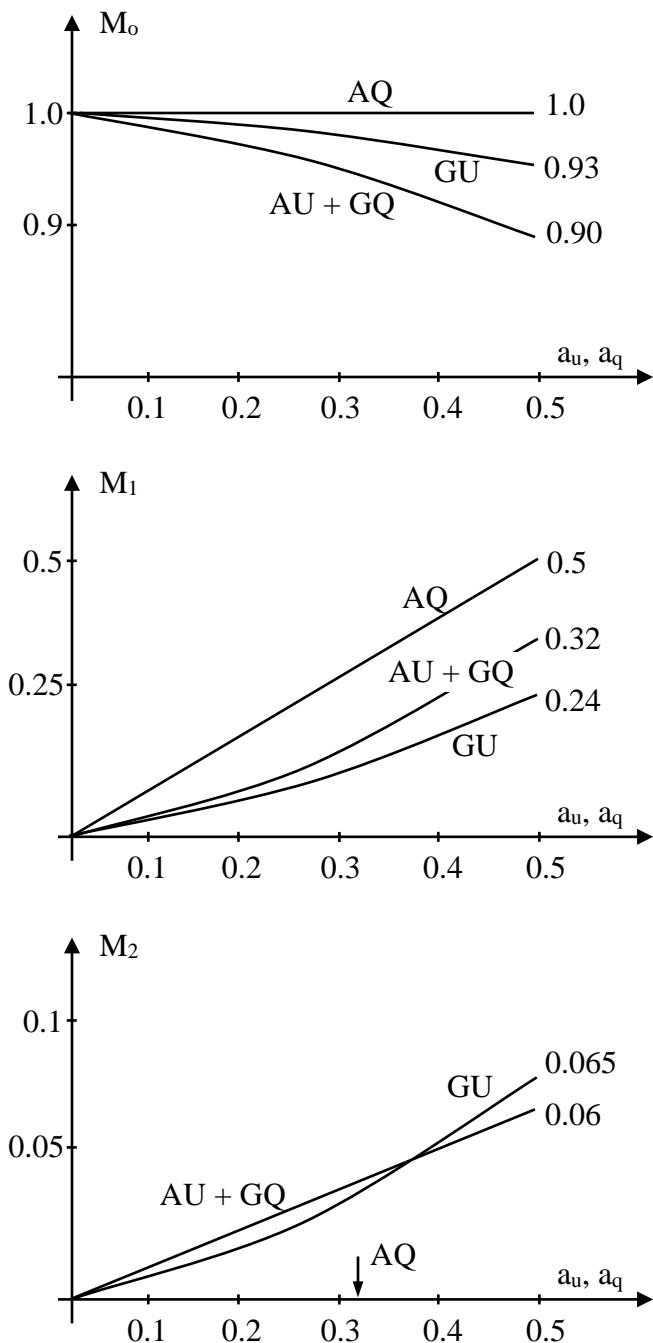


Fig. IV-6. M_k coefficients for different pump sources.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

II - Small signal analysis: Excitation by two signals (the pump and the signal)

With an excitation signal f_o , the output frequency $f_i = \omega_i/2\pi$ can be expressed as

$$f_i = f_{m,n} = |mf_o + nf_p| \quad \text{with} \quad m, n = \dots, -1, 0, 1, \dots \quad (\text{IV-24})$$

Knowing that the capacitance is a loss-less element, the energy conservation criteria is equivalent to

$$\sum_m \sum_n P_{m,n} = 0 \quad (\text{IV-25})$$

where $P_{m,n}$ is the average power at frequency $f_{m,n}$. It is positive if generated by a source and negative if dissipated in a circuit. After multiplying each power by a number equal to unity, this relation can be reformatted as:

$$\sum_m \sum_n P_{m,n} \left\{ \frac{m\omega_o + n\omega_p}{m\omega_o + n\omega_p} \right\} = 0 \rightarrow \omega_o \sum_m \sum_n \frac{m P_{m,n}}{\omega_{m,n}} + \omega_p \sum_m \sum_n \frac{n P_{m,n}}{\omega_{m,n}} = 0 \quad (\text{IV-26})$$

Therefore, with

$$u(t) = \sum_m \sum_n U_{m,n} e^{j\omega_{m,n} t} \quad \text{and} \quad q(t) = \sum_m \sum_n Q_{m,n} e^{j\omega_{m,n} t} \quad (\text{IV-27})$$

and using

$$I = \frac{dQ}{dt} = j\omega Q$$

the powers are equal to

$$P_{m,n} = 2 \operatorname{Re} \left\{ U_{m,n} I^*_{m,n} \right\} = -2 \omega_{m,n} \operatorname{Im} \left\{ U_{m,n} Q^*_{m,n} \right\} \quad (\text{IV-28})$$

Here, **Re** and **Im** represent the real part and the imaginary part, respectively. We can now introduce normalized powers:

$$\frac{P_{m,n}}{\omega_{m,n}} = -2 \operatorname{Im} \left\{ U_{m,n} Q^*_{m,n} \right\} \quad (\text{IV-29})$$

and note that they are not frequency dependent because the coefficients $U_{m,n}$ and $Q_{m,n}$ are related only to the diode curve $C(t)$.

Equation (IV-26) can be reformatted in order to show the fundamental relationships between all powers, also known as the “Manley-Rowe relations” (Appendix IV-2) :

$$\left. \begin{array}{l} \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{mf_o + nf_p} = 0 \\ \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \frac{nP_{m,n}}{mf_o + nf_p} = 0 \end{array} \right\} \text{or} \quad \left. \begin{array}{l} \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{\omega_{m,n}} = 0 \\ \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \frac{nP_{m,n}}{\omega_{m,n}} = 0 \end{array} \right\} \quad (\text{IV-30})$$

Summation boundaries are chosen so that each power $P_{m,n}$ is taken into account only once. These relations show that the source powers are converted into output powers at frequency $f_{m,n}$ but they did not show how the undesirable powers are dissipated. In practice, the diode is not ideal and this rule is played by the series resistances noted $R_{m,n}$ (see Figure IV-7 where the $F_{m,n}$ are ideal pass-band filters centered on frequency $f_{m,n}$).

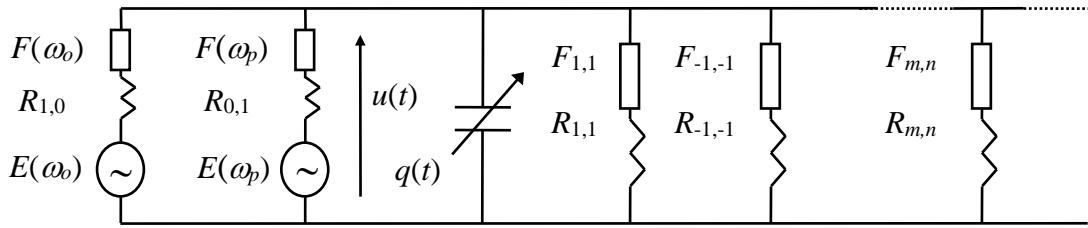


Fig. IV-7. Parametric amplifier configuration.

To illustrate this concept, let us consider the case of a frequency multiplier (Figure IV-8).

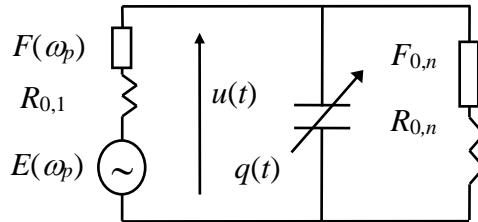


Fig. IV-8. Parametric amplifier configuration.

In this circuit, $m = 0$ and $n = 0, 1$. Thus,

$$\frac{P_{0,1}}{\omega_p} + \frac{n P_{0,n}}{n \omega_p} = 0 \quad \rightarrow \quad \frac{P_{0,n}}{P_{0,1}} = -1 = G_{\max} \quad (\text{IV-31})$$

With an ideal varactor, the maximum expected value of the gain for a frequency multiplier is unity due to the transfer of power between harmonics. A more complex example is the circuit shown in Figure IV-9. In this circuit, we have three frequencies, i.e., two input signals $\omega_{1,0}$ (i.e., ω_0) and $\omega_{0,1}$ (i.e., ω_p) and one output signal $\omega_{1,1}$: the circuit is a mixer (converter). Furthermore, since $\omega_{1,1} = \omega_{0,1} + \omega_{1,0}$, the mixer is an upper sideband up-converter.

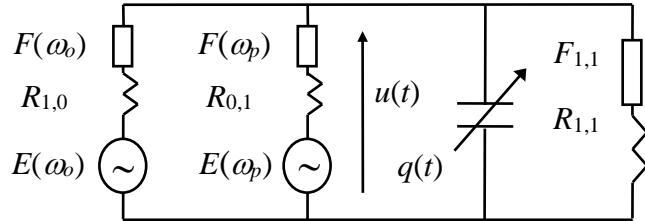


Fig. IV-9. Parametric upper sideband up-converter configuration.

In this circuit, $\{ m = 0, 1 \}$ and $\{ n = 0, 1 \}$. Then,

$$\frac{P_{1,0}}{\omega_o} + \frac{P_{1,1}}{\omega_o + \omega_p} = 0 \quad \rightarrow \quad \frac{P_{1,1}}{P_{1,0}} = -\frac{\omega_o + \omega_p}{\omega_o} = G_{\max} \quad (\text{IV-32})$$

In the ideal case, the gain of the upper sideband parametric up-converter is higher than unity but cannot exceed G_{\max} .

In general, in order to obtain the optimum efficiency, real powers due to undesirable frequencies must not be present. This condition is assured by loading the diode with reactive impedances at all frequencies except the desired one. But the series resistance R_s of the diode cannot allow this condition to be satisfied only if the loads are open circuited in order to eliminate the power dissipation in R_s . This implies a zero current in the resistance for all undesirable frequencies, which is practically impossible to realize.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

An alternative approach consists to put short-circuits and then to provide power dissipation in the output network. In practice, this is done by adding circuits, called "idle", for each undesirable frequency. But this will complicate significantly the final circuit, so that analyzing such circuits is relatively complex. Therefore, a more simplified but approximate approach can be utilized, namely the coupled circuit analysis.

D – COUPLED CIRCUIT REPRESENTATION

To analyze parametric amplifiers, the nonlinear element, namely the capacitance, is separated from the rest of the circuit represented by its Thevenin's equivalent impedance noted $Z(\omega)$ (Figure IV-10).

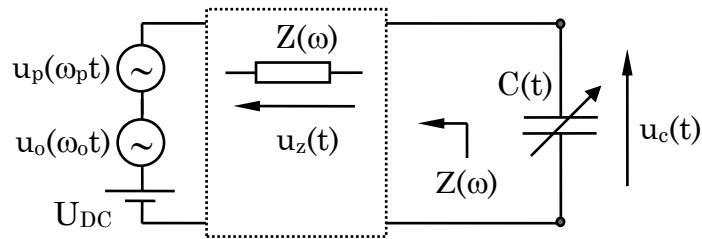


Fig. IV-10. Equivalent circuit of the parametric amplifier.

Thus, the source voltage $u(t)$ can be expressed in the form of an immittance (Appendix IV-3) :

$$\begin{aligned}
 u(t) &= u_c(q(t)) - \int_0^t Z(t-\tau) q(\tau) d\tau = U_{dc} + u_p(\omega_p t) + u_o(\omega_o t) \\
 &= U_{dc} + 2U_{acp} \cos(\omega_p t) + 2U_{aco} \cos(\omega_o t + \varphi) \\
 &= U_{dc} + U_{acp} e^{j\omega_p t} + U_{acp} e^{-j\omega_p t} + \{U_{aco} e^{j\varphi}\} e^{j\omega_p t} + \{U_{aco} e^{j\varphi}\} e^{-j\omega_p t}
 \end{aligned} \tag{IV-33}$$

or

$$u(t) = U_{dc} + 2 \operatorname{Re} \left\{ U_p e^{j\omega_p t} \right\} + 2 \operatorname{Re} \left\{ U_o e^{j\omega_p t} \right\} \quad (\text{IV-34})$$

Integration, with initial conditions equal to zero, gives

$$u_c(q(t)) + Z(0)q(t) - \int_0^t Z(t-\tau)q(\tau) d\tau = u_p(\omega_p t) + u_o(\omega_o t) + Z(t-0)q(0) \quad (\text{IV-35})$$

$$\rightarrow u_c(q(t)) - \int_0^t Z(t-\tau)q(\tau) d\tau = u_p(\omega_p t) + u_o(\omega_o t) \quad (\text{IV-36})$$

$$\rightarrow F(q(t)) = u_p(\omega_p t) + u_o(\omega_o t) \quad (\text{IV-37})$$

This nonlinear Frechet's operator F is calculated using the large-signal small-signal analysis (Appendix IV-4).

I - Large-signal analysis for the nonlinear operator: Excitation by the pump

We have:

$$F[q(\omega_p t)] = u(\omega_p t) \quad (\text{IV-38})$$

II - Small signal analysis for the operator: Excitation by the pump and the signal

We have

$$F[q_p(\omega_p t) + q_o(\omega_o t)] = u(\omega_p t) + u(\omega_o t) \quad (\text{IV-39})$$

Therefore:

$$F[q_p(\omega_p t) + q_o(\omega_o t)] - F[q_p(\omega_p t)] = u(\omega_o t) \quad (\text{IV-40})$$

The operator F can be expended using Frechet's series:

$$\begin{aligned} F[q_p(\omega_p t) + q_o(\omega_o t)] &= F[q_p(\omega_p t)] + dF \Big|_{q_p(\omega_p t)} q_o(\omega_o t) + \dots \\ \rightarrow \quad dF \Big|_{q_p(\omega_p t)} q_o(\omega_o t) &= u(\omega_o t) \end{aligned} \quad (\text{IV-41})$$

Thus, using equation (IV-36), we have

$$\frac{du_c}{dq} \Big|_{q_p(\omega_p t)} q_o(t) - \int_0^t Z(t-\tau) q_o(\tau) d\tau = u_o(\omega_o t) \quad (\text{IV-42})$$

III - Steady-state case

Knowing that q is the integral of i over time, the above equation can be expressed as

$$S(\omega_p t) \int_0^t i_o(\tau) d\tau + \int_0^t Z(t-\tau) i_o(\tau) d\tau = u_o(\omega_o t) \quad (\text{IV-43})$$

which gives in the steady-state case,

$$S(\omega_p t) \int_0^t i_o(\tau) d\tau + Z(\omega) i_o(\omega_o t) = u_o(\omega_o t) \quad (\text{IV-44})$$

This linear equation is similar to (IV-36). To solve it, let us set

$$u_o(t) = 2U_{aco}\cos(\omega_o t + \varphi) = 2\operatorname{Re}(U_{aco} e^{j\omega_p t}) \quad (\text{IV-45})$$

and

$$S(\omega_p t) = \sum_{n=-\infty}^{\infty} S_n e^{jn\omega_p t} \quad (\text{IV-46})$$

$$i_s(t) = 2\operatorname{Re}\left(\sum_{n=-\infty}^{\infty} I_n e^{j(\omega_o + n\omega_p)t}\right) = 2\operatorname{Re}\left(\sum_{n=-\infty}^{\infty} I_n e^{j\omega_n t}\right) \quad (\text{IV-47})$$

Note that

$$S_n = S_{-n}^* \quad \text{and} \quad I_n \neq I_{-n}^*$$

Thus, this allows bringing out an impedance matrix

$$[\mathbf{Z}] [\dots \quad I_{-1} \quad I_0 \quad I_1 \quad \dots]^t = [\dots \quad 0 \quad U_0 \quad 0 \quad \dots]^t \quad (\text{IV-48})$$

Line #m

$$\Rightarrow [\mathbf{Z}] = \begin{matrix} \vdots & & & & & \\ -1 & \left[\begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ \dots & Z_{-1,-1} & Z_{-1,0} & Z_{-1,1} & \dots \end{array} \right] \\ 0 & \left[\begin{array}{cccc} \dots & Z_{0,-1} & Z_{0,0} & Z_{0,1} & \dots \end{array} \right] \\ 1 & \left[\begin{array}{cccc} \dots & Z_{1,-1} & Z_{1,0} & Z_{1,1} & \dots \end{array} \right] \\ \vdots & \vdots & \vdots & \vdots & & \end{matrix} \quad (\text{IV-49})$$

Column #n

where the exponent t indicates the transpose matrix. Elements of the matrix $[Z]$ are equal to

$$Z_{m,n} = \begin{cases} \frac{S_o}{j\omega_m} + Z(\omega_m) & m = n \\ \frac{S_{m-n}}{j\omega_n} & m \neq n \end{cases} \quad (\text{IV-50})$$

where ω_n can be equivalent to $\omega_{m,n}$ upon the condition where the amplitude of the pump signal is much more greater than the one of the input signal:

$$A_o(\omega_o) \ll A_p(\omega_p) \quad (\text{IV-51})$$

$$\rightarrow \omega_{m,n} = m\omega_o + n\omega_p \approx \omega_o + n\omega_p = \omega_n \quad (\text{IV-52})$$

In our problem, the elastance is the time-dependent variable. $Z_{m,m}$ are the internal time-invariant impedances of the circuit and $Z_{m,n}$ (for $m \neq n$) the parametric coupling impedances between the powers at different frequencies with

$$-M \leq m \leq M \quad \text{and} \quad -N \leq n \leq N$$

The coupling is achieved by the elastance. Therefore, the parametric circuit is reduced to a set of coupled non-parametric circuits as shown on Figure IV-11. A similar procedure can bring out an admittance matrix which elements are

$$Y_{m,n} = \begin{cases} j\omega_m C_o + Y(\omega_m) & m = n \\ j\omega_m C_{m-n} & m \neq n \end{cases} \quad (\text{IV-53})$$

where the equivalent sources are current sources in a set of coupled circuits I_M to I_N .

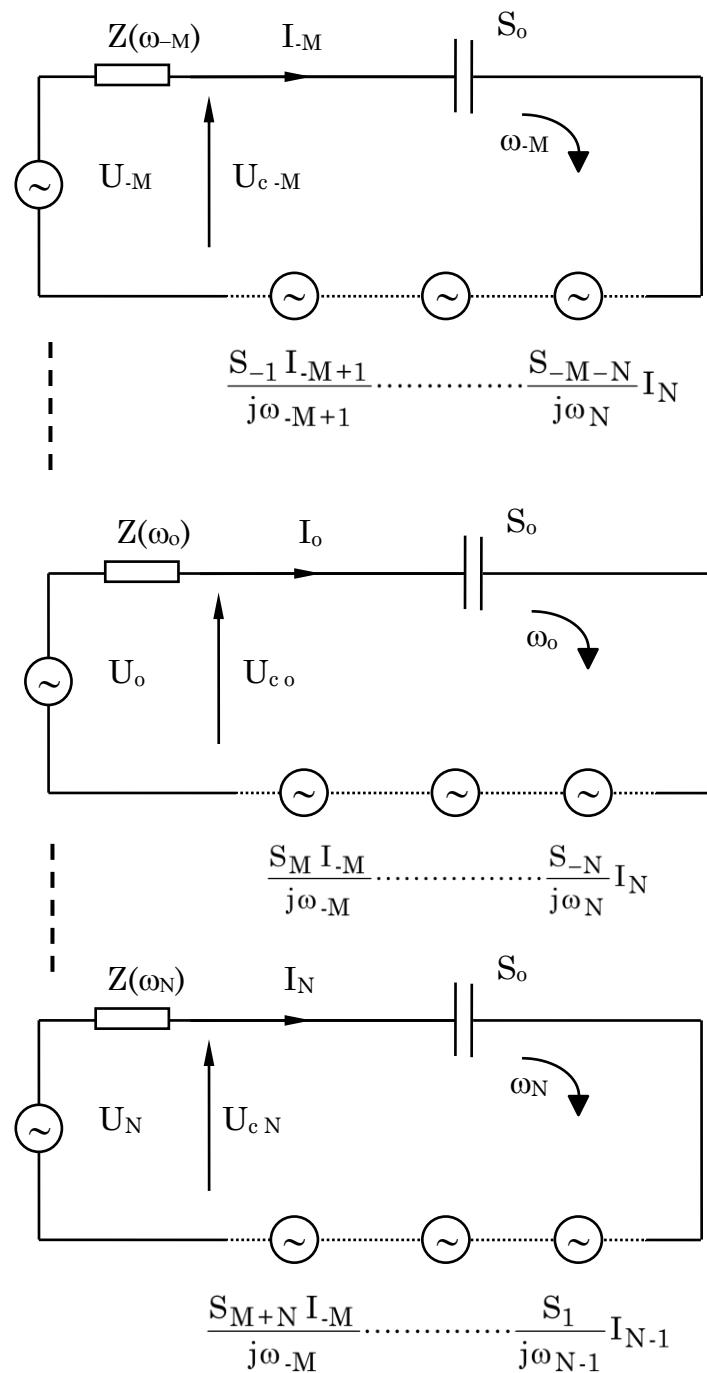


Fig. IV-11. Parametric circuit representation as a set of coupled non-parametric circuits.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

D – PARAMETRIC CIRCUIT PARAMETERS

In order to determine the circuit parameters, let us use the example of a parametric amplifier (lower sideband) excited by a voltage source (Figure IV-12).

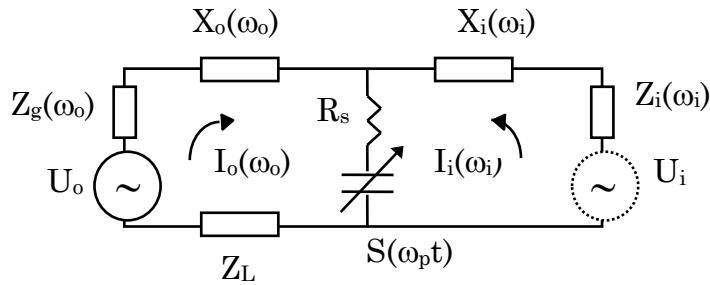


Fig. IV-12. Parametric amplifier with idler circuit.

In this circuit, with

$$\{ \omega_l = \omega_p - \omega_b = -\omega_1 \} \quad (\text{IV-54})$$

we have

$$U_i = U_{-1}^* \quad \text{and} \quad I_i = I_{-1}^* \quad (\text{IV-55})$$

so, the current-voltage relationship can be written as

$$\begin{bmatrix} U_i^* \\ U_o \end{bmatrix} = \begin{bmatrix} Z_{-1,-1} & Z_{-1,0} \\ Z_{0,-1} & Z_{0,0} \end{bmatrix} \begin{bmatrix} I_i^* \\ I_o \end{bmatrix} \quad (\text{IV-56})$$

where the internal impedances $Z_{m,m}$ are equal to

$$Z_{0,0} = R_{0,0} + jX_{0,0} = Z_g(\omega_o) + Z_L(\omega_o) + R_s + j \left[X_o(\omega_o) - \frac{S_o}{\omega_o} \right] \quad (\text{IV-57})$$

$$Z_{-1,-1} = R_{-1,-1} + jX_{-1,-1} = Z_i^*(\omega_i) + R_s - j \left[X_i(\omega_i) - \frac{S_o}{\omega_i} \right] \quad (\text{IV-58})$$

with

$$Z_{m,n} = \frac{S_{m-n}}{j\omega_n} = -j \frac{S_{m-n}}{\omega_n} = jX_{m,n} \quad \rightarrow \quad X_{m,n} = -\frac{S_{m-n}}{\omega_n} \quad (\text{IV-59})$$

Thus, the input impedance is equal to

$$Z_{in} = R_s - j \frac{S_o}{\omega_o} - \frac{\omega_o \omega_i}{Z_i^*(\omega_i) + R_s - j \left(X_i(\omega_i) - \frac{S_o}{\omega_i} \right)} \quad (\text{IV-60})$$

This circuit is based on parametric coupling. At the resonant frequency, i.e., at

$$X_{-1,-1} = 0 = X_i^* + \frac{S_o}{\omega_i} - X_i(\omega_i) \quad (\text{IV-61})$$

the input impedance is reduced to

$$R_{in}|_{res} = R_s - \frac{|S_i|^2}{\omega_o \omega_i (R_i + R_s)} \quad (\text{IV-62})$$

Let us introduce the quality factor of a parametric circuit:

$$q_{m,n} = \frac{X_{m,n}}{R_s} = -\frac{S_{m-n}}{\omega_n R_s} \quad (\text{IV-63})$$

$$\rightarrow q_m^n = q_n^m = q_{m,n} * q_{m,n} \quad (\text{IV-64})$$

Thus,

$$q_0^i = \frac{|S_i|^2}{\omega_o \omega_i R_s^2} \quad (\text{IV-65})$$

Let us also define the notion of normalized impedance based on the capacitance, i.e.,

$$z = \frac{Z}{R_s}$$

So equation (IV-62) can be reformulated as

$$r_{in}|_{res} = 1 - \frac{q_0^i}{1 + r_i} \quad (\text{IV-66})$$

The input resistance can be negative as shown in Fig. IV-13.b A negative resistance can reach to a gain $G > 1$. In fact, the load power is equal to

$$P_L = \frac{U_o^2 Z_L}{(Z_L + Z_g + Z_{in})} \quad (\text{IV-67})$$

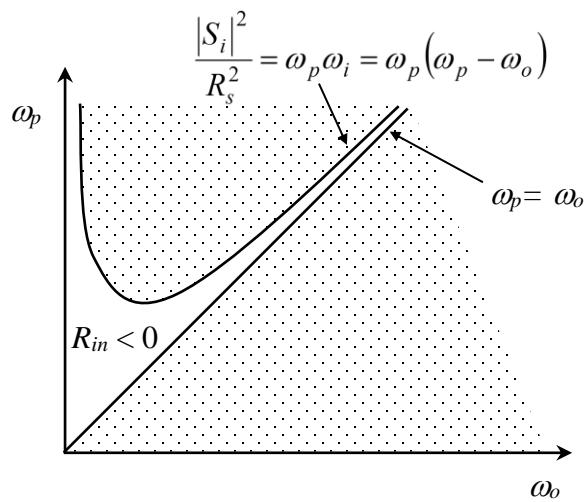


Fig. IV-13. Input impedance vs. frequency.

At resonant frequency,

$$P_L|_{R_{in} \neq 0} = \frac{U_o^2 Z_L}{(Z_L + Z_g + R_{in})^2} \quad (\text{IV-68})$$

$$P_L|_{R_{in} = 0} = \frac{U_o^2 Z_L}{(Z_L + Z_g)^2} \quad (\text{IV-69})$$

Thus, the power gain is equal to

$$G|_{res} = \frac{(Z_L + Z_g)^2}{(Z_L + Z_g + R_{in})^2} \quad (\text{IV-70})$$

A negative input impedance leads to $G > 1$: amplification!

Many configurations have been proposed but the coupled parametric amplifier via a circulator is the most used one (Figure IV-14).

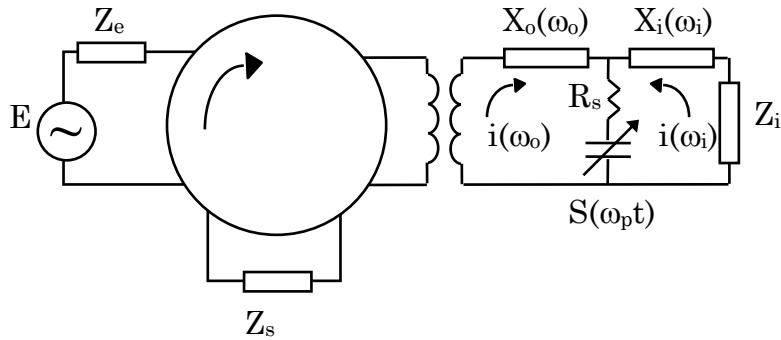


Fig. IV-14. Parametric amplifier: coupling by a circulator.

As expected, this kind of amplifier exhibits a negative resistance. In fact, the energy conservation theorem shows that the only powers that are present are $P_{1,0}(\omega_0)$, $P_{0,1}(\omega_p)$ and $P_{-1,1}(\omega_i)$ with the relation

$$P_{1,-1} = P_{-1,1} \quad (\text{IV-71})$$

Then:

$$\left. \begin{aligned} \frac{P_{1,0}}{\omega_o} + \frac{P_{1,-1}}{\omega_o - \omega_p} &= \frac{P_{1,0}}{\omega_o} - \frac{P_{-1,1}}{\omega_p - \omega_o} = 0 \\ \frac{P_{0,1}}{\omega_p} + \frac{P_{-1,1}}{\omega_p - \omega_o} &= \frac{P_{0,1}}{\omega_p} - \frac{P_{1,-1}}{\omega_o - \omega_p} = 0 \end{aligned} \right\} \quad (\text{IV-72})$$

By setting $P_{0,1}(\omega_p)$ as a positive power, $P_{-1,1}(\omega_i)$ and $P_{1,0}(\omega_o)$ are negative. We have then a possibility of oscillations in input or the existence of a negative resistance.

If we consider the resonant case, we have

$$R_{in} = R_s - \frac{|S_i|^2}{\omega_o \omega_i} \frac{1}{R_s + R_i} \quad (\text{IV-73})$$

Therefore, this resistance could have a negative value, Therefore, parametric amplifiers can exhibit an amplification gain greater than unity.

E – PARAMETRIC AMPLIFICATION: DIODES VERSUS TRANSISTORS

Transistors are also used for parametric amplification. In this case, the negative resistance is obtained by a feedback loop, which make the transistor unstable (Figure IV-15).

Compared to usual transistor amplifiers, these amplifiers are more stable, exhibit better *gain*Bandwidth* product, and present less sensitivity to transistor parameters variations.

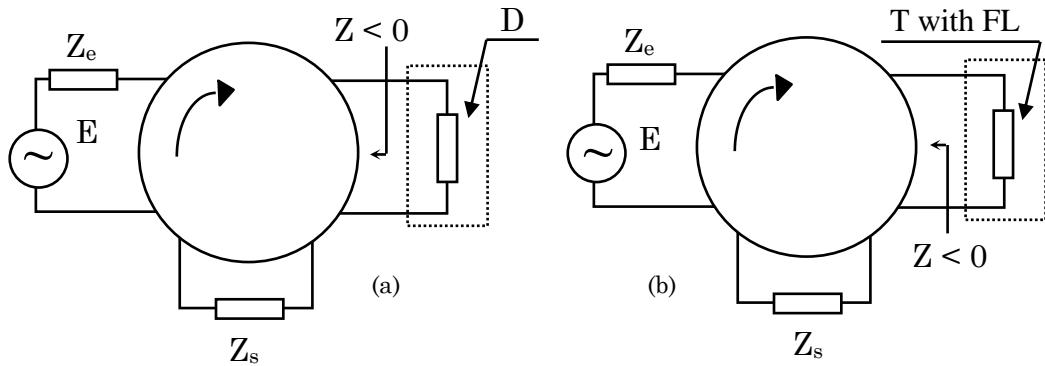


Fig. IV-15. Parametric amplifier with coupling via circulator

(a) : Diode amplifier (D).

(b) : Transistor amplifier (T) with a feedback loop (FL).

F – TRANSISTOR AMPLIFIERS

Most important design considerations in a microwave transistor amplifier are stability, power gain, bandwidth, noise and dc requirements. Usually, transistor amplifiers can be classified as

- reactively matched,
- lossy matched,
- feedback,
- distributed and
- balanced.

Reactively matched circuits represent the most common amplifier configuration. In fact, circuits like “wideband” or “broadband” amplifiers refer to amplifiers with a bandwidth of typically one or two octaves, i.e., that essentially have a pass-band characteristic. Such amplifiers are often of the

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

reactively matched type and find application in phased-array radars, electronic counter measure (ECM) systems, civil radars and satellite communication systems (SATCOMS). Amplifiers having bandwidths of a few tens of percent mostly employ reactive matching, and find application in communications such as direct broadcast from satellite (DBS) systems.

On the other hand, ultra-broadband amplifiers, as a class, have a bandwidth of several octaves, and often have a low-frequency response down to dc, or at most a few hundred megahertz. Those circuits, which respond at dc, can rightly be regarded as low-pass. They are required for electronic warfare (EW) systems, guided weapons (GW), instrumentation and pulse amplifiers.

As a guide, reactively-matched amplifiers allow the most gain to be derived from a given FET (and they also offer the lowest noise).

Designing an amplifier usually starts with a set of specifications followed by the selection of the proper transistor. Then, a systematic mathematical solution, supported by numerical/graphical methods, is developed to determine the transistor loading (source and load reflection coefficients) for user-defined stability and gain criteria.

But first, what is the meaning of “gain”?

I – Power definitions

We will base our definitions ***on the reactively matched transistor configuration***. The transistor can be defined using S-parameters

$$b_1 = S_{11}a_1 + S_{12}a_2 \quad b_2 = S_{21}a_1 + S_{22}a_2 \quad (\text{IV-74})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

where a_k and b_k are the incident and reflected waves of port k respectively. Let us consider the simple circuit shown on Figure IV-16. The transistor is excited by a voltage source of magnitude E and internal impedance Z_s and loaded by Z_L . The reflection coefficients of these impedances are noted Γ_S and Γ_L respectively.

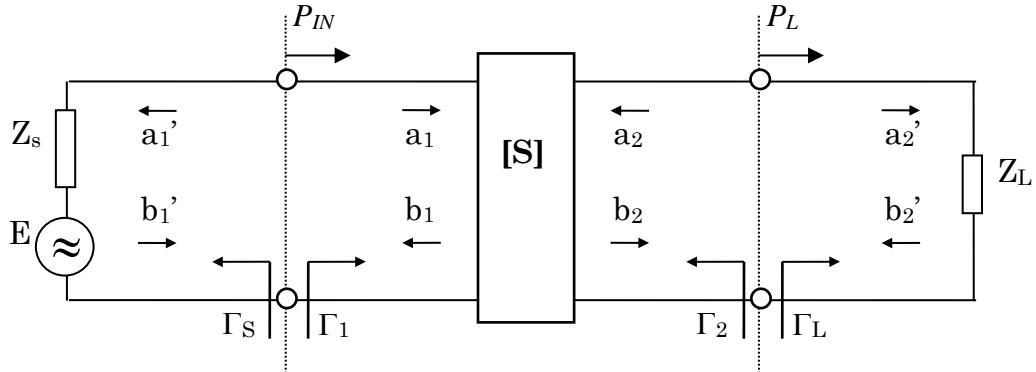


Fig. IV-16. Two-port transistor representation in terms of scattering parameters.

From this mathematical representation, we can introduce some definitions of powers used in amplifier design:

- *Input power delivered to the transistor:* the input power to the transistor is defined as

$$P_{IN} = P_1 = \frac{1}{2}|a_1|^2 - \frac{1}{2}|b_1|^2 = \frac{1}{2}|a_1|^2(1 - |\Gamma_1|^2) \quad (\text{IV-75})$$

- *Power delivered to the load:* The power delivered to the load is defined as the difference between the incident and reflected power, namely

$$P_L = \frac{1}{2}|b_2|^2 - \frac{1}{2}|a_2|^2 = \frac{1}{2}|b_2|^2(1 - |\Gamma_L|^2) \quad (\text{IV-76})$$

- *Power available from a source:* The power available from a source is defined as the power delivered by a source to the conjugately matched load, i.e.,

$$\Gamma_1 = \Gamma_s^* \quad (\text{IV-77})$$

Thus, the power available from the source is given by

$$P_{AVS} = P_{IN} \Big|_{\Gamma_1 = \Gamma_s^*} = \frac{1}{2}|b_1|^2 - \frac{1}{2}|a_1|^2 = \frac{1}{2}|a_1|^2 - \frac{1}{2}|b_1|^2 = \frac{\frac{1}{2}|b_1|^2}{1 - |\Gamma_s|^2} \quad (\text{IV-78})$$

- *Power available from the network (transistor):* The power available from the transistor is defined as the power delivered by the transistor when

$$\Gamma_2 = \Gamma_L^* \quad (\text{IV-79})$$

Thus, the power available from the source is given by

$$P_{AVN} = P_L \Big|_{\Gamma_2 = \Gamma_L^*} = \frac{1}{2}|b_2|^2 - \frac{1}{2}|a_2|^2 = \frac{1}{2}|b_2|^2 \left(1 - |\Gamma_2|^2\right) \quad (\text{IV-80})$$

II – Gain definitions

The transducer power gain G_T is defined as the ratio of the power delivered to the load to the power available from the source.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Thus,

$$G_T = \frac{P_L}{P_{AVS}} = \frac{|b_2|^2}{|b_1|^2} \left(1 - |\Gamma_L|^2\right) \left(1 - |\Gamma_S|^2\right) \quad (\text{IV-81})$$

It can also be expressed in terms of the transistor S-parameters and the reflection coefficients

$$G_T = \frac{1 - |\Gamma_S|^2}{|1 - \Gamma_S \Gamma_1|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_L \Gamma_2|^2} \quad (\text{IV-82})$$

Similarly, the operating power gain is defined as the ratio of the power delivered to the load to the input power delivered to the transistor

$$G_P = \frac{P_L}{P_{IN}} = \frac{|b_2|^2}{|a_1|^2} \frac{\left(1 - |\Gamma_L|^2\right)}{\left(1 - |\Gamma_1|^2\right)} \quad (\text{IV-83})$$

$$G_P = \frac{1}{1 - |\Gamma_1|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2} \quad (\text{IV-84})$$

Finally, the available power gain G_A is defined as the ratio of P_{AVN} over P_{AVS}

$$G_A = \frac{|b_2|^2}{|b_1|^2} \left(1 - |\Gamma_2|^2\right) \left(1 - |\Gamma_S|^2\right) \quad (\text{IV-85})$$

$$G_A = \frac{1 - |\Gamma_S|^2}{|1 - S_{11} \Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2} \quad (\text{IV-86})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

III – Amplifier design goals

Any amplifier design is mainly based on the right choice of source and load impedances Z_S and Z_L . This is essential since the amplifier performance and parameters (gain, input/output impedance, noise figure, etc.) are function of these impedance values. The purpose of any linear (small-signal) amplifier design is finally to determine the appropriate reflection coefficients Γ_S and Γ_L .

For linear and quasi-linear amplifiers, this design is achieved from the small-signal S-parameters of the active component, i.e., the transistor. For power (nonlinear) amplifiers, this design is achieved through load-pull transistor characterization. The load is determined to allow maximum output power.

IV – Definition of stability and matching

The stability of an amplifier, or its resistance to oscillate, is a fundamental consideration which can be determined from the S-parameters, the matching networks, and the terminations:

- Make an amplifier stable is to load it with source and load impedances exhibiting real positive parts ($|\Gamma_S| < 1$ and $|\Gamma_L| < 1$). In other words, the impedances should be passive.

The matching of an amplifier is its ability to transfer a maximum of power from the source to the transistor and from the transistor to the load:

- Match an amplifier means to built a maximum power transfer from source to network and from network to load.

V - Reactively matched transistor design

Referring to Figure IV-16, let Γ_1 be the reflection coefficient at the transistor input when its output port is matched ($a_2 = 0$) and Γ_2 the reflection coefficient at the transistor output when the input port is matched ($a_1 = 0$). These two matching conditions can be expressed as

$$\Gamma_1 = \frac{b_1}{a_1} = \Gamma_S^* \quad (\text{IV-87})$$

$$\Gamma_2 = \frac{b_2}{a_2} = \Gamma_L^* \quad (\text{IV-88})$$

where the (*) sign indicates the complex conjugate. Using equation (IV-74), a substitution between the two expressions leads to

$$\Gamma_1 = S_{11} + \frac{\Gamma_L S_{12} S_{21}}{1 - \Gamma_L S_{22}} = \Gamma_S^* \quad (\text{IV-89})$$

$$\Gamma_2 = S_{22} + \frac{\Gamma_S S_{12} S_{21}}{1 - \Gamma_S S_{11}} = \Gamma_L^* \quad (\text{IV-90})$$

By combining these two relations, we obtain second-order equations which solutions are given by

$$\Gamma_S = \frac{B_1 \pm \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \quad (\text{IV-91})$$

$$\Gamma_L = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \quad (\text{IV-92})$$

with

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2$$

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2$$

$$C_1 = S_{11} - \Delta S_{22}^*$$

$$C_2 = S_{22} - \Delta S_{11}^*$$

where Δ is the determinant of the S matrix

$$\Delta = S_{11} S_{22} - S_{12} S_{21} \quad (\text{IV-93})$$

We can note that a permutation of subscripts 1 and 2 did not affect the above expressions. Furthermore, since

$$B_1^2 - 4|C_1|^2 = B_2^2 - 4|C_2|^2 = 4(K^2 - 1)|S_{12} S_{21}|^2$$

we can introduce a new factor called the Rollet factor or the stability factor defined as

$$K = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{12}||S_{21}|} \quad (\text{IV-94})$$

This relation highlights that the input/output matching and stability are based on the K value:

- a) $K \equiv 1$ it is not a practical case since $|\Gamma_S| = |\Gamma_L| = 1$

b) $K \geq 1$ expanding the two solutions of Γ_S and Γ_L gives

$$\Gamma_S' = \frac{B_1 - \sqrt{B_1^2 - 4|C_1|^2}}{2C_1}; \quad \Gamma_S'' = \frac{B_1 + \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \quad (\text{IV-95})$$

$$\Gamma_L' = \frac{B_2 - \sqrt{B_2^2 - 4|C_2|^2}}{2C_2}; \quad \Gamma_L'' = \frac{B_2 + \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \quad (\text{IV-96})$$

Note that $|\Gamma_S'| \cdot |\Gamma_S''| = 1$ and $|\Gamma_L'| \cdot |\Gamma_L''| = 1$

Thus, one solution of Γ_S and Γ_L has a magnitude greater than unity and the other has a magnitude lower than unity. So, it is always possible to simultaneously match the input *AND* the output of the transistor, with no effect on stability.

To explicit which solution has to be retained, we have to consider Δ :

- $|\Delta| < 1$ The pair of solutions is $\{\Gamma_S', \Gamma_L'\}$.
- $|\Delta| > 1$ The pair of solutions is $\{\Gamma_S'', \Gamma_L''\}$.

c) $K \leq -1$ For this value of K , the two possible solutions are $\{\Gamma_S', \Gamma_L''\}$ or $\{\Gamma_S'', \Gamma_L'\}$. However, these solutions cannot be useful for amplifiers for two raisons. First, it is not possible to match simultaneously the input and the output. Second, the amplifier is unstable because we will associate a solution with a magnitude greater than unity with another one of magnitude less than unity.

d) $-1 \leq K \leq 1$ Simultaneous matching is impossible. The amplifier is conditionally stable (or potentially unstable), because only some values of Γ_S and Γ_L can be used to avoid oscillations. In this case, we need to determine the allowed regions in which we can select the source/load impedances that assure stability.

Reported into the reflection coefficient space (Smith chart), these regions are equivalent to circles, and thus are called "stability circles".

To summarize, when the amplifier is potentially unstable, there may be values of Γ_S and Γ_L for which the real parts of Z_S and Z_L are positive. These values of Γ_S and Γ_L (i.e., regions in the Smith chart) can be determined using the following procedure. First, find the limits of such regions where Γ_S and Γ_L produce $\{ |\Gamma_1| = 1 \}$ and $\{ |\Gamma_2| = 1 \}$, i.e.,

$$|\Gamma_S| = 1 \quad \rightarrow \quad |\Gamma_1| = \left| S_{11} + \frac{\Gamma_L S_{12} S_{21}}{1 - \Gamma_L S_{22}} \right| = 1 \quad (\text{IV-97})$$

$$|\Gamma_L| = 1 \quad \rightarrow \quad |\Gamma_2| = \left| S_{22} + \frac{\Gamma_S S_{21} S_{12}}{1 - \Gamma_S S_{11}} \right| = 1 \quad (\text{IV-98})$$

After substituting,

$$\left| \Gamma_L - \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12} S_{21}}{|S_{22}|^2 - |\Delta|^2} \right| \quad (\text{IV-99})$$

$$\left| \Gamma_S - \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \right| = \left| \frac{S_{21} S_{12}}{|S_{11}|^2 - |\Delta|^2} \right| \quad (\text{IV-100})$$

The derivations of the above equations show two families of circles:

Γ_L values for $|\Gamma_1|=1 \rightarrow$ output stability circle: Centre Ω_2 and radius R_2

$$\Omega_2 = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \quad R_2 = \frac{|S_{12} S_{21}|}{\left| |S_{22}|^2 - |\Delta|^2 \right|} \quad (\text{IV-101})$$

Γ_S values for $|\Gamma_2|=1 \rightarrow$ input stability circle: Centre Ω_1 and radius R_1

$$\Omega_1 = \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \quad R_1 = \frac{|S_{21} S_{12}|}{\left| |S_{11}|^2 - |\Delta|^2 \right|} \quad (\text{IV-102})$$

So, on one side of the circle boundary, in the Γ_L plane, we will have $|\Gamma_1| > 1$ and on the other side $|\Gamma_1| < 1$. Similarly, in the Γ_S plane, on one side of the circle boundary, we will have $|\Gamma_2| > 1$ and on the other side $|\Gamma_2| < 1$.

Next we need to determine which area in the Smith chart represents the stable region, i.e., the region where values of Γ_L ($|\Gamma_L| < 1$) produce $|\Gamma_1| < 1$ and where values of Γ_S ($|\Gamma_S| < 1$) produce $|\Gamma_2| < 1$. To this end, we observe that if $Z_L = Z_0$, then $\Gamma_L = 0$ and $|\Gamma_1| = |S_{11}|$. If the magnitude of S_{11} is less than 1, then $|\Gamma_1| < 1$ when $\Gamma_L = 0$. In other words, the centre of the Smith chart in Figure IV-17-a represents a stable operating region. On the other hand, if the magnitude of S_{11} is greater than 1, then the centre of the Smith chart in Figure IV-17-a is in an unstable operating region for the input. Choosing any value of source impedance in such region leads to oscillation.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

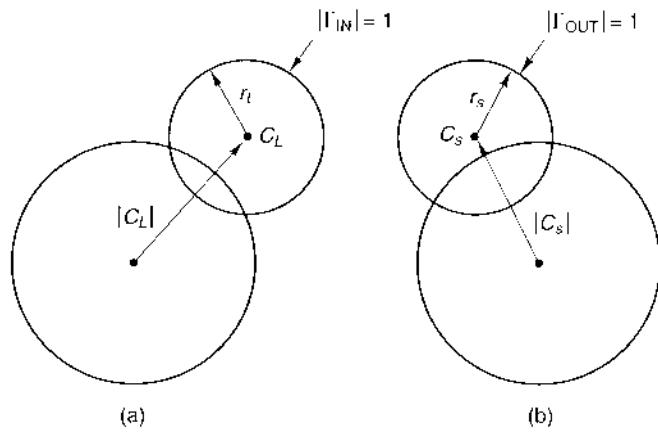


Fig. IV-17. Stability circles in the Smith chart (a) for Γ_1 (Γ_L plane), (b) for Γ_2 (Γ_S plane).

Most of the time, microwave amplifiers used for narrowband or wideband applications face stability problems at certain frequency ranges. Instability is primarily caused by three phenomena: internal feedback around the transistor, external feedback around the transistor caused by an external circuit, or excess of gain at frequencies outside of the band of operation. Figure IV-18 shows six passive feedback-networks that are used usually to stabilize amplifier circuits (from

<http://scholar.lib.vt.edu/theses/available/etd-07152001-172453/unrestricted/chap3.PDF>)

Similarly, the magnitude of S_{22} will decide upon the location of the stable region for Γ_2 (Figure IV-17-b). Figures IV-19-a and IV-19-b show stable and unstable regions versus the magnitude of S_{11} (for the Γ_L plane) and S_{22} (for the Γ_S plane), respectively.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

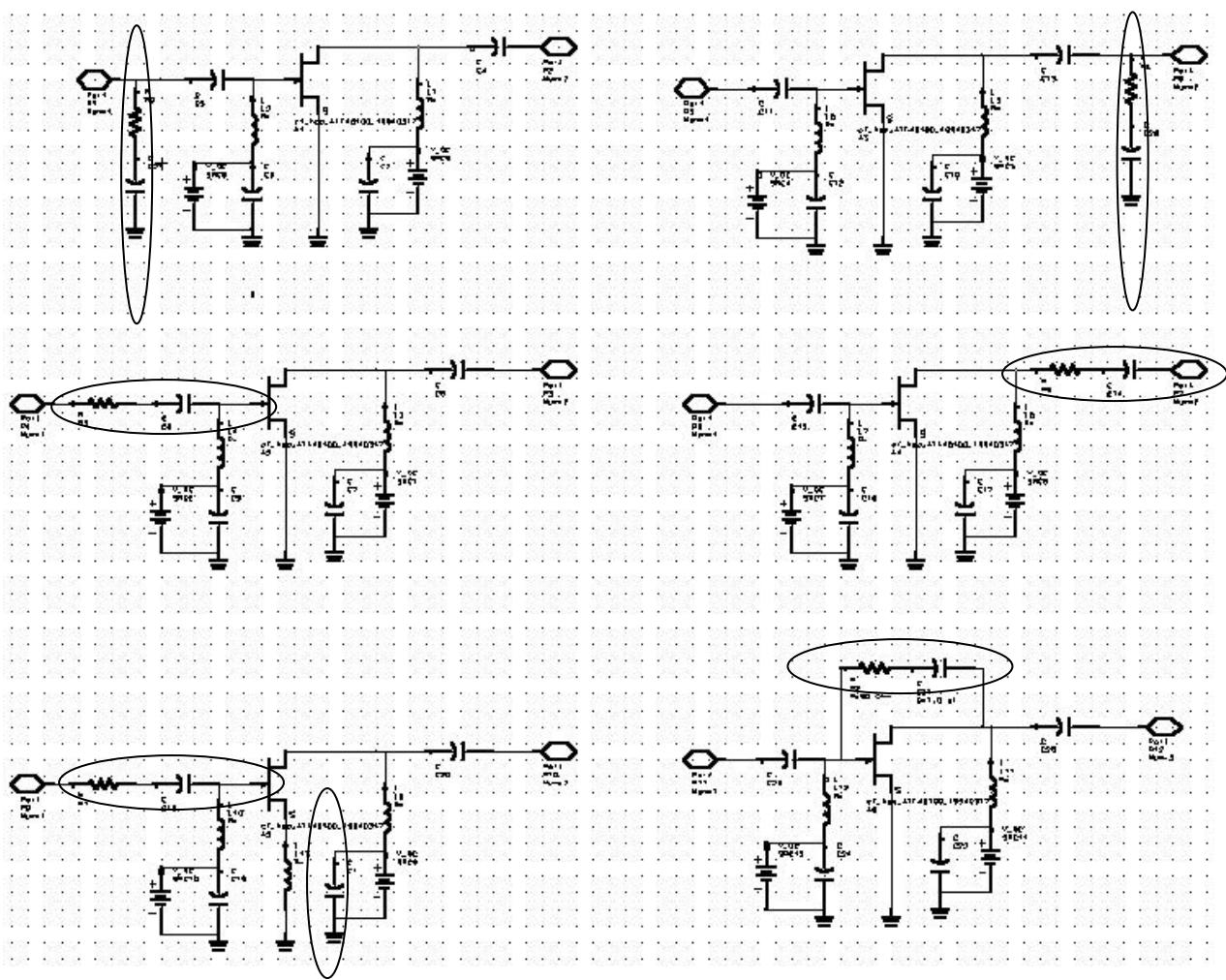


Fig. IV-18. Stable and unstable regions in the Γ_L plane (here Γ_{in} is Γ_1).

However, these circles give only an area on the smith chart. To find appropriate impedance values for a transistor amplifier, the designer needs to specify the values of the gain and the noise factor desired for the designed amplifier.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

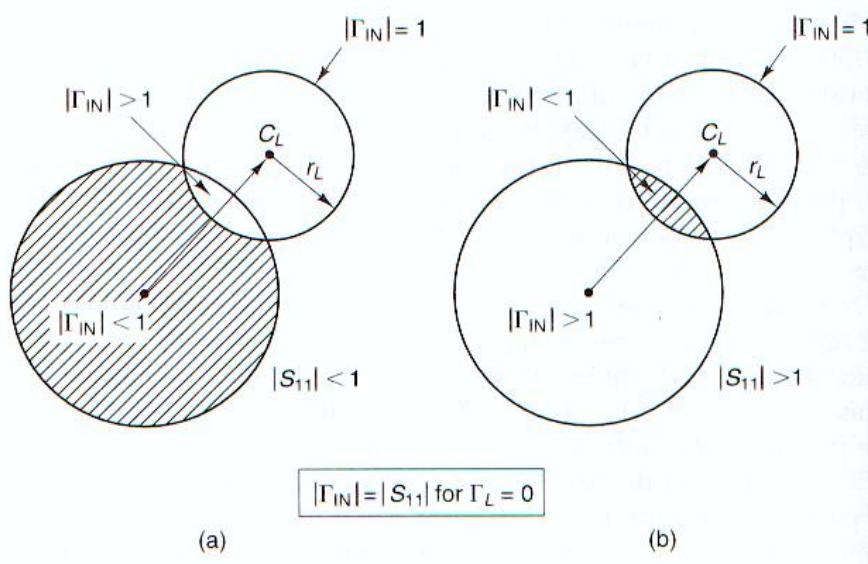


Fig. IV-19-a. Stable and unstable regions in the Γ_L plane (here Γ_{in} is Γ_1).

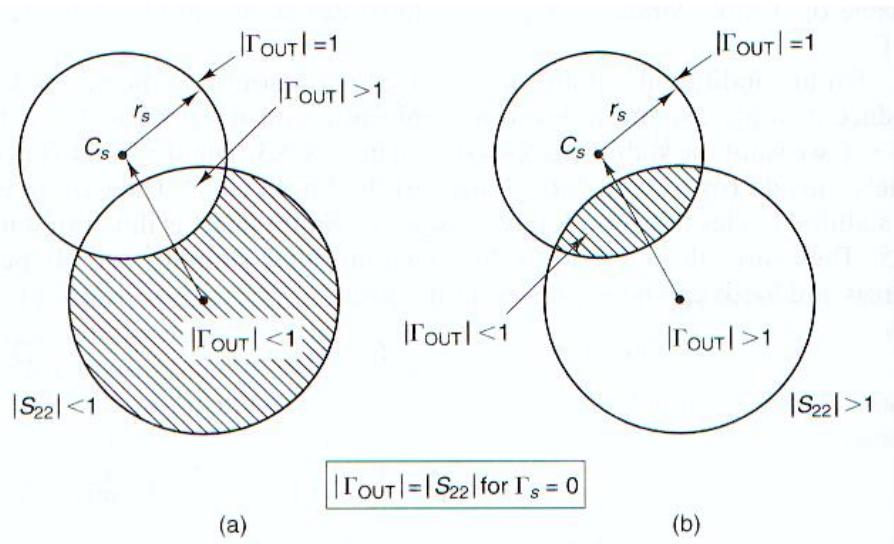


Fig. IV-19-b. Stable and unstable regions in the Γ_s plane (here Γ_{out} is Γ_2).

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

G – AMPLIFIER DESIGN

I – Power Gain

When the input is matched, the power gain becomes

$$G_p = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2)}{1 - |S_{11}|^2 + |\Gamma_L|^2 (|S_{22}|^2 - |\Delta|^2) - 2 \operatorname{Re}(C_2 \Gamma_L)} \quad (\text{IV-103})$$

By introducing the normalized gain

$$g_p = \frac{G_p}{|S_{21}|^2} = \frac{(1 - |\Gamma_L|^2)}{1 - |S_{11}|^2 + |\Gamma_L|^2 (|S_{22}|^2 - |\Delta|^2) - 2 \operatorname{Re}(C_2 \Gamma_L)} \quad (\text{IV-104})$$

we can demonstrate that the values of Γ_L that give a constant gain g_p are located on a circle, called "constant-gain circle" of radius R_p and centered at Ω_p

$$\Omega_p = \frac{g_p (S_{22}^* - \Delta * S_{11})}{1 + g_p (|S_{22}|^2 - |\Delta|^2)} \quad (\text{IV-105})$$

$$R_p = \frac{\sqrt{1 - 2 K |S_{12} \cdot S_{21}| \cdot g_p + |S_{12} \cdot S_{21}|^2 \cdot g_p^2}}{\left| 1 + g_p (|S_{22}|^2 - |\Delta|^2) \right|} \quad (\text{IV-106})$$

Figures IV-20 and IV-21 show a typical set of constant-gain circles on the Smith chart.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Note 1: The gain is related to the output impedance or load, i.e., the load reflection coefficient Γ_L . Choosing a specific value for the gain implies choosing a specific load value.

Note 2: Any value for the load has to be inside the stable region for potentially unstable cases. For stable transistors, any load value can be selected (with the restriction: $\text{Re}(Z) > 0$).

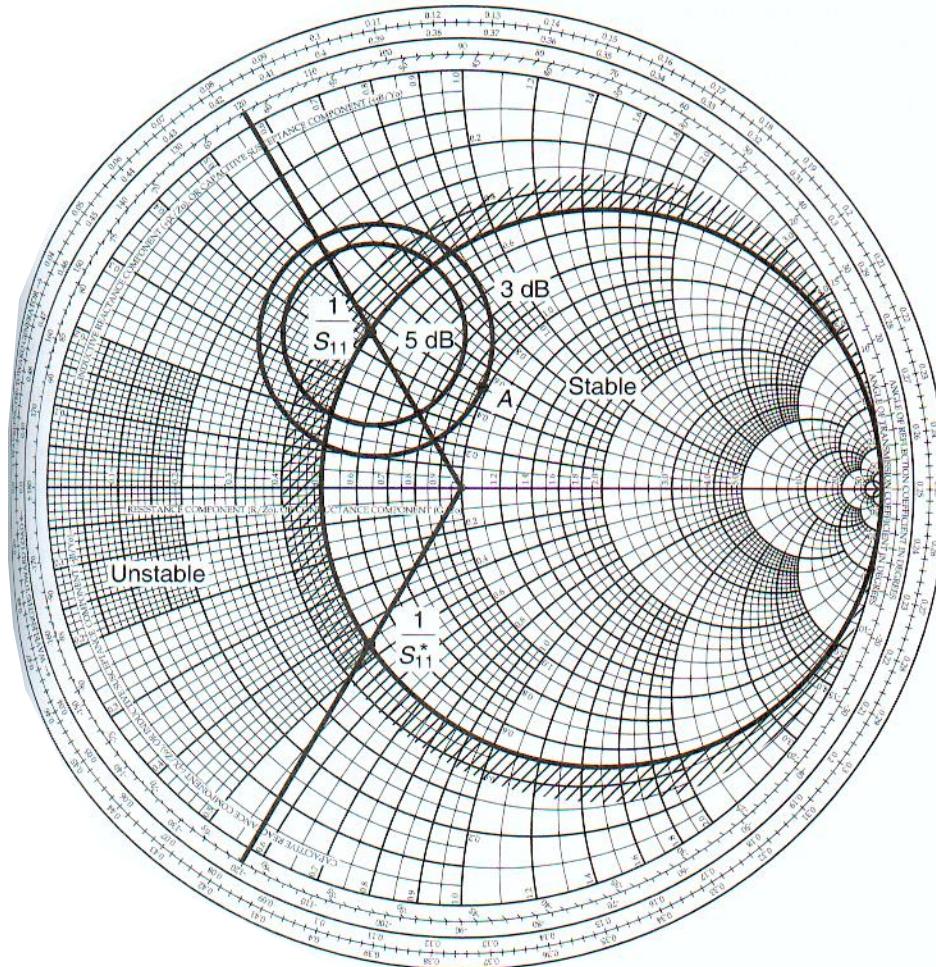
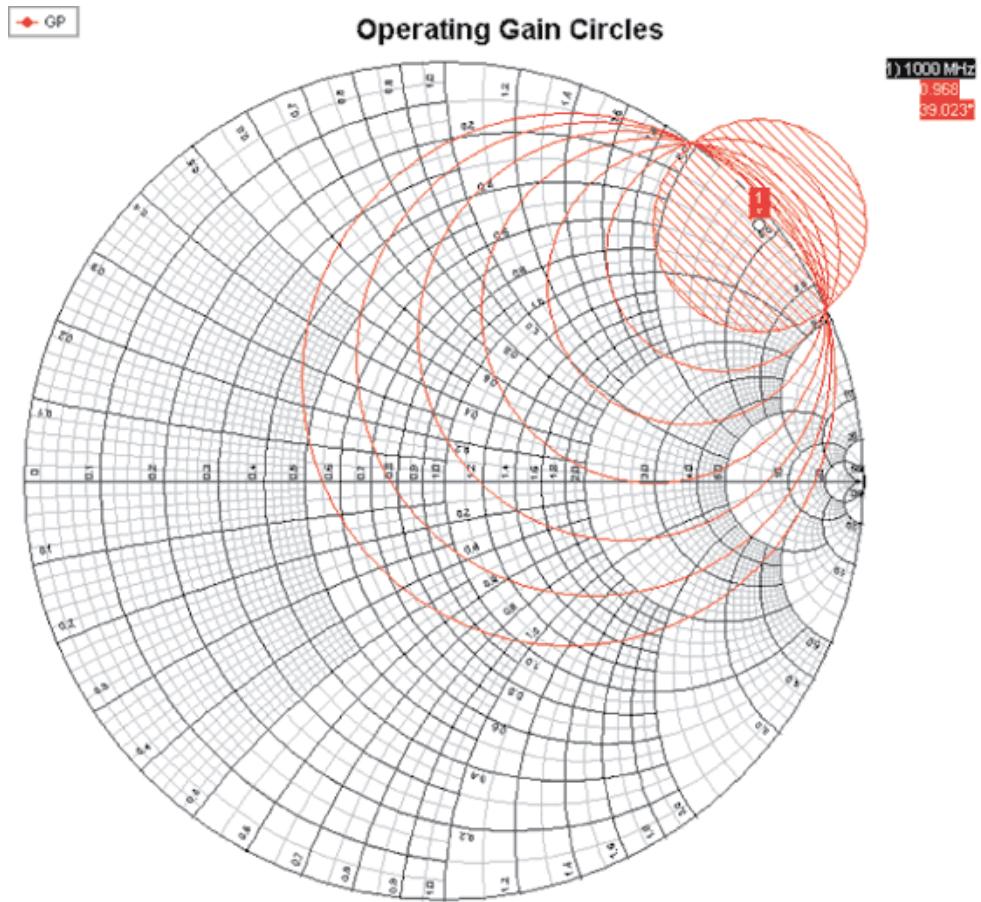


Fig. IV-20. Constant-gain circles for a given transistor, including stable and unstable regions of the load reflection coefficient.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.



2. These operating gain circles for the unstabilized model 2N6679A transistor are plotted for levels $-1, -2, -3, -4, -5$, and -6 dB below the simultaneous match gain. The (shaded) load instability circle is also shown.

Fig. IV-21. From Joseph F. White, *Microwave & RF Journal*

(<http://mwrf.com/Authors/AuthorID/1686/1686.html>)

Once Γ_L is selected, maximum output power is obtained with a conjugate match at the input, namely, with $\Gamma_S = \Gamma_1^*$

$$\Gamma_S = \left(\frac{S_{11} - \Delta \Gamma_L}{1 - \Gamma_L S_{22}} \right)^* \quad (\text{IV-107})$$

Note that the designer has to verify that the corresponding value of Z_S belongs to the input stable region, i.e., $|\Gamma_1| < 1$.

II – Noise factor

First, the designer has to know that gain is related to the load (output reflection coefficient Γ_L) while noise figure is related to the source impedance (i.e., Γ_S). So, selecting a given NF will imply selecting a specific source impedance. The noise figure of a two-port network is given by

$$F = F_{\min} + 4 \frac{R_n}{Z_o} \frac{|\Gamma_S - \Gamma_{opt}|^2}{|1 + \Gamma_{opt}|^2 (1 - |\Gamma_S|^2)} \quad (\text{IV-108})$$

$$F = F_{\min} + \frac{R_n}{Z_o} \frac{|\Gamma_S - \Gamma_{opt}|^2}{|1 + \Gamma_{opt}|^2 (1 - |\Gamma_S|^2)} \quad (\text{IV-109})$$

depending on the definition of R_n , the noise resistance of the transistor. It can be defined also in terms of source admittances

$$F = F_{\min} + \frac{r_n}{g_s} |y_s - y_{opt}|^2 \quad (\text{IV-110})$$

with

$r_n = R_n/Z_0$: equivalent normalized noise resistance of the two-port (e.g., transistor),

$y_s = g_s + j b_s$: normalized source admittance, with

$$y_s = \frac{1 - \Gamma_s}{1 + \Gamma_s} \quad (\text{IV-111})$$

$\Gamma_s = g_{opt} + j b_{opt}$: normalized source admittance that results in the minimum (or optimum) noise figure called F_{\min}

$$y_{opt} = \frac{1 - \Gamma_{opt}}{1 + \Gamma_{opt}} \quad (\text{IV-112})$$

F_{\min} , r_n and Γ_{opt} are called the transistor noise parameters. They are usually given by the manufacturer or have to be determined experimentally.

The value of F_{\min} occurs when $\Gamma_{opt} = \Gamma_s$. The noise resistance r_n can be measured by reading the noise figure $F = F_s$ when $\Gamma_s = 0$.

Then, using (IV-108), we obtain

$$F_S = F_{\min} + 4r_n \frac{|\Gamma_{opt}|^2}{|1 + \Gamma_{opt}|^2} \quad \rightarrow \quad r_n = (F_S - F_{\min}) + \frac{|1 + \Gamma_{opt}|^2}{4|\Gamma_{opt}|^2} \quad (\text{IV-113})$$

Equation (IV-108) can be used to design Γ_S for a given noise figure F . Thus, we can rearrange that equation as

$$N = \frac{|\Gamma_S - \Gamma_{opt}|^2}{1 - |\Gamma_S|^2} = \frac{F - F_{\min}}{4} \frac{Z_o}{R_n} |1 + \Gamma_{\min}|^2 \quad (\text{IV-114})$$

showing that the right-hand side is a constant N called the noise figure transistor parameter. Equation (IV-114) can be recognized as the equation of a circle in the Γ_S plane. For a given N , the center and radius of such circles called “constant noise figure circles”, are given by

$$R_n = \frac{\sqrt{N^2 + N(1 - |\Gamma_{\min}|^2)}}{1 + N} \quad C_n = \frac{\Gamma_{\min}}{1 + N} \quad (\text{IV-115})$$

As for the gain, the value of Γ_S allows to select the corresponding value of Γ_L

$$\Gamma_L = \left(\frac{S_{22} - \Delta \Gamma_S}{1 - \Gamma_S S_{11}} \right)^* \quad (\text{IV-116})$$

Figures IV-22 and IV-23 show a typical set of constant-noise figure circles on the Smith chart.

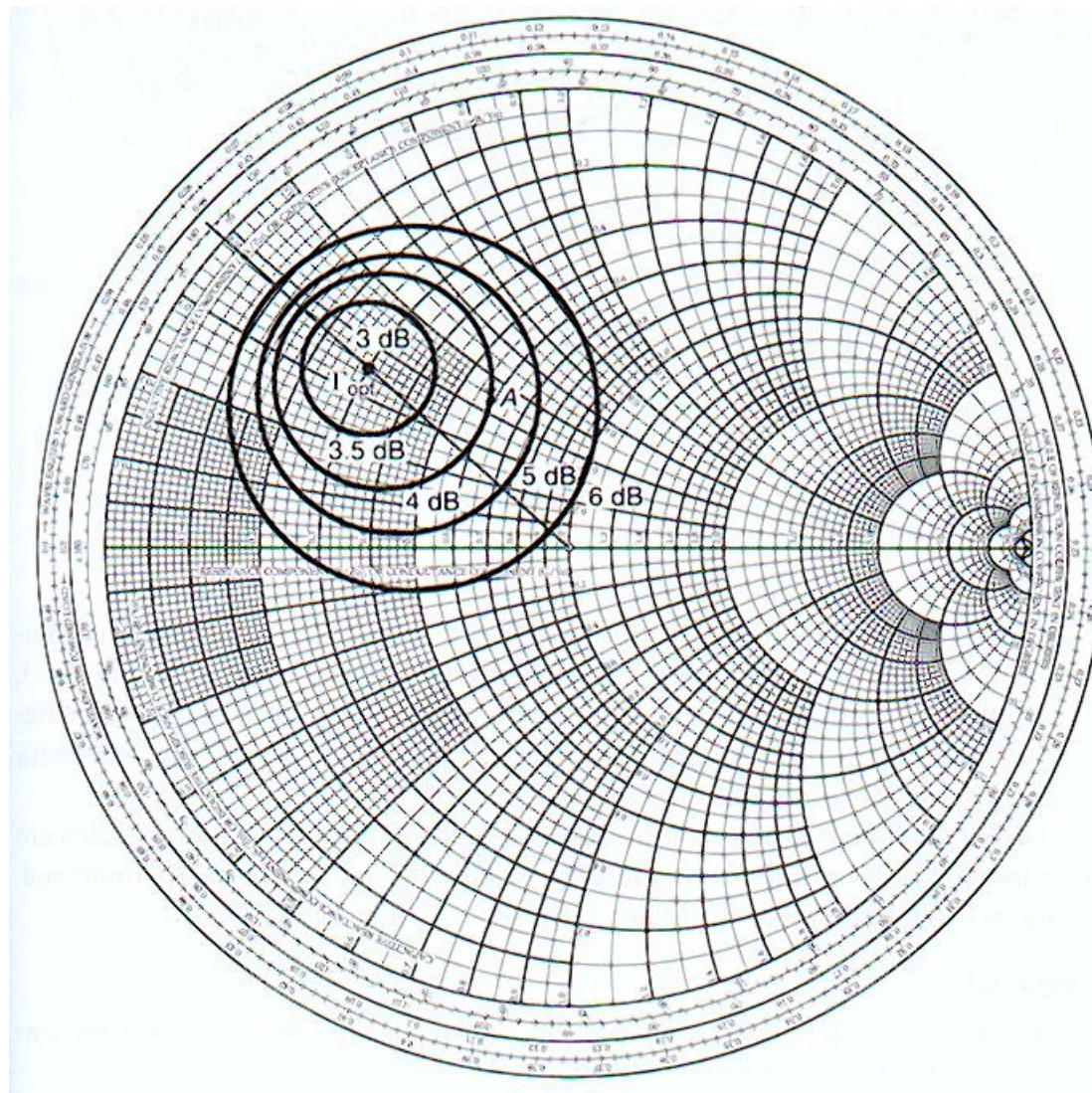


Fig. IV-22. Constant-noise figure circles for a given transistor.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

5. These constant noise-figure circles were generated for the amplifier of Fig. 4. The circles are for -0.25, -0.5, -1, -1.5, -2, -2.5, -3 and -6 dB below $NF_{MIN} = 0.5\text{dB}$

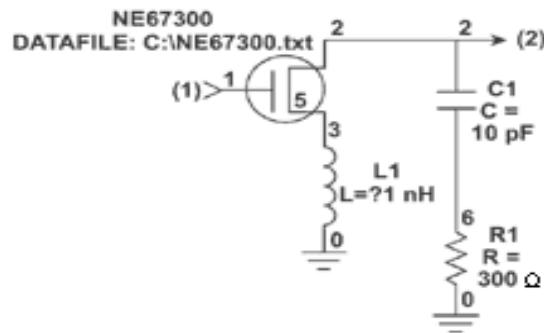
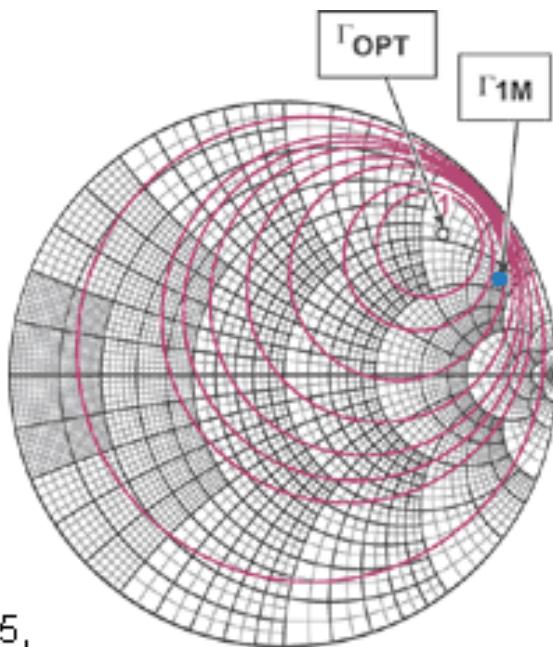


Fig. IV-23. From Joseph F. White, *Microwave & RF Journal*

(<http://mwrf.com/Authors/AuthorID/1686/1686.html>)

Note: Constant-gain circles:

when the radius increases, gain decreases.

Constant-noise figure circles:

when the radius increases, noise figure increases.

III – Compromise between power gain and noise figure

A typical design should link between all the above notions: constant-gain circles and constant-noise figure circles (Figures IV-24 and IV-25).

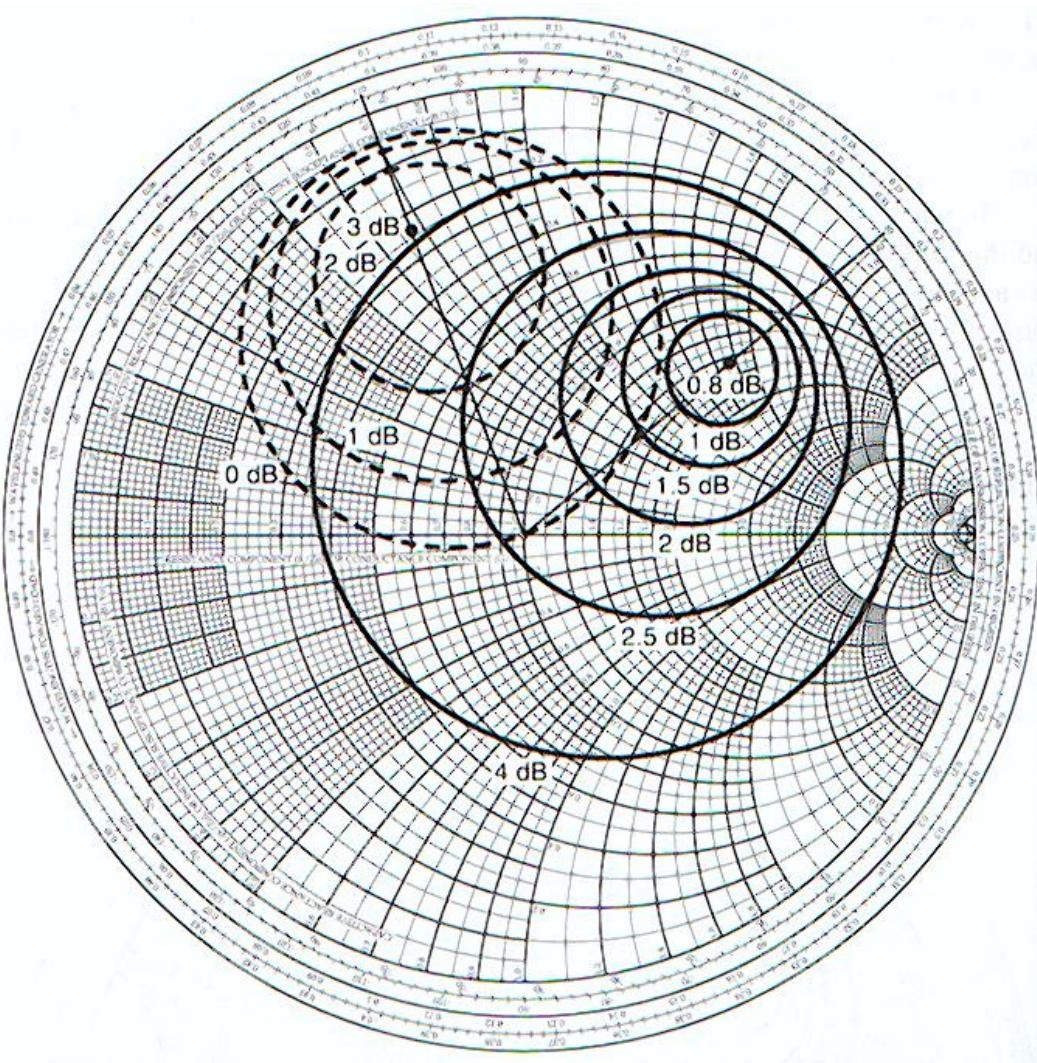


Fig. IV-24. Constant-gain circles (dashed curves) and constant -noise figure circles (solid curves) for a given transistor.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

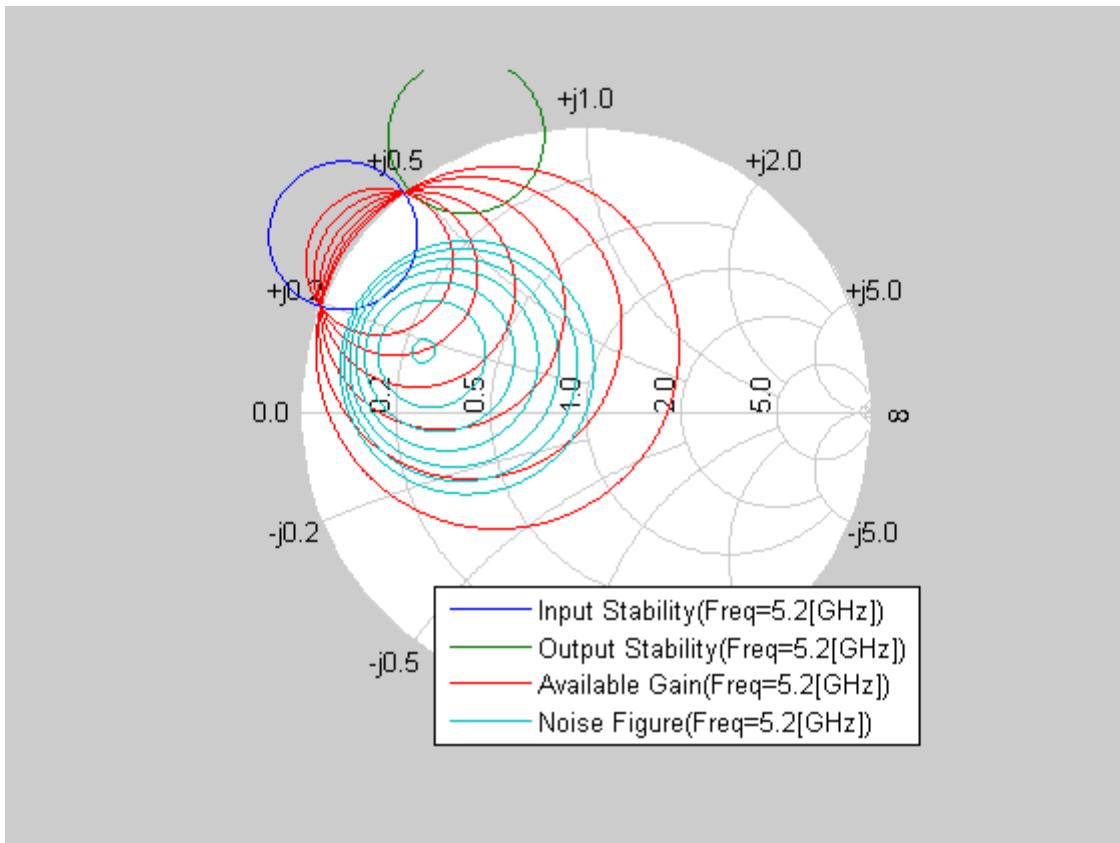


Fig. IV-25. From Mathworks notes

[http://www.mathworks.com/products/rftoolbox/demos.html?file=/products/demos/shipping/rf/
lna_match.html](http://www.mathworks.com/products/rftoolbox/demos.html?file=/products/demos/shipping/rf/lna_match.html)

To these circles, we should add the stability circles to fully design the amplifier. Here we have three cases:

- For a given gain and noise figure, we have two intersect points between the constant-gain circle and the constant-noise figure circle. Hence, we have two possible values for Γ_S . The designer has to assure that the selected point belongs to the input stability region, and that

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

the corresponding impedance value is realistic (in terms of technology limitations, substrate performance, matching network, etc.). Use equation (IV-116) to obtain Γ_L (same restrictions for the load impedance as for the source impedance).

- One intersection point, thus one possible value for Γ_S and Γ_L .
- No intersection: it is impossible to design an amplifier with the specified gain and noise figure specifications.

A typical design is shown in Figure IV-26 highlighting all the circles required for proper design.

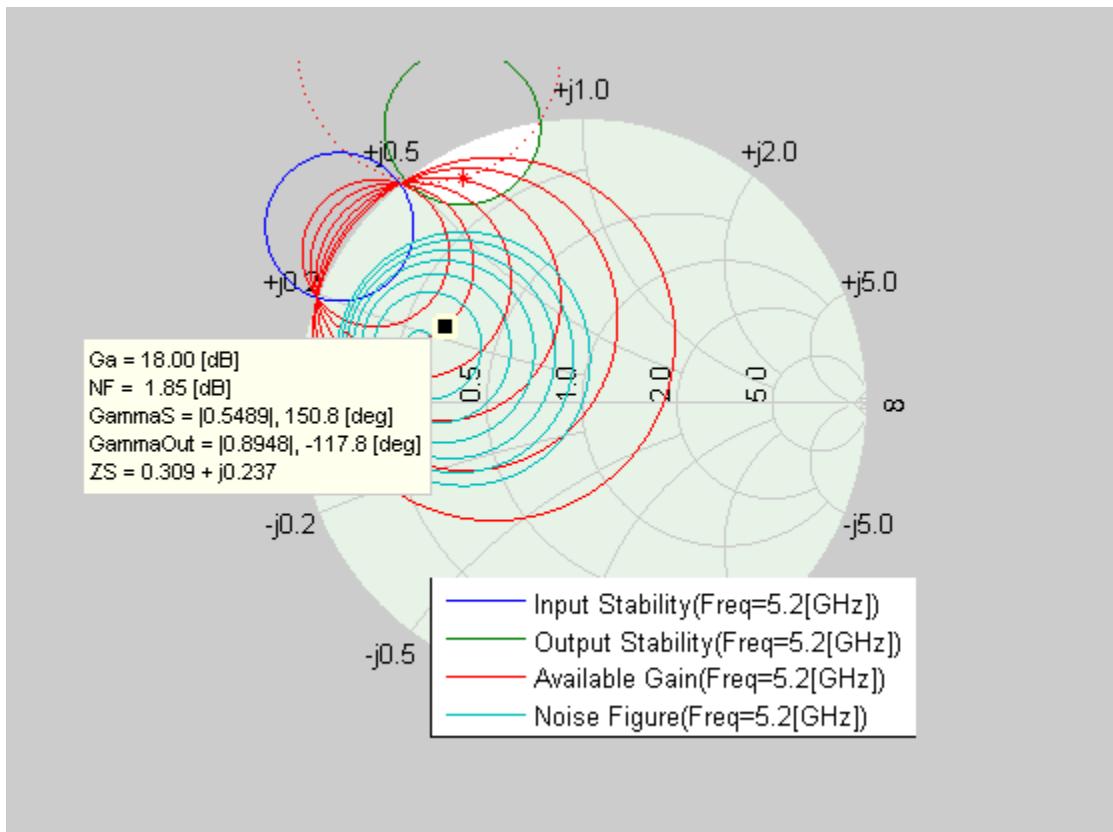


Fig. IV-26. From Mathworks notes

http://www.mathworks.com/products/rftoolbox/demos.html?file=/products/demos/shipping/rf/Lna_match.html

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

IV – Noise figure of multistage amplifiers

In case of a cascaded circuit (multi-stage amplifier), the well-known Friis equation, named after Harald T. Friis, allows us obtaining the total noise figure. Friis's formula is used to calculate the total noise figure of a cascade of N two-port networks, each with its own noise figure and gain. It is given as

$$F_{1\dots N} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 * G_2} + \dots + \frac{F_N - 1}{G_1 * G_2 * \dots * G_N}$$

where F_n and G_n are the noise factor and available power gain, respectively, of the n^{th} stage. Note that both magnitudes are expressed as ratios, not in decibels.

Note : The above equation assumes that all N stages exhibit the same bandwidth, i.e., $B_1 = \dots = B_N$.

H - TYPICAL CONFIGURATIONS OF MATCHING NETWORKS

In order to meet certain user specifications, matching networks should be inserted between source and transistor and between transistor and load (Figure IV-27). Their design is critical.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

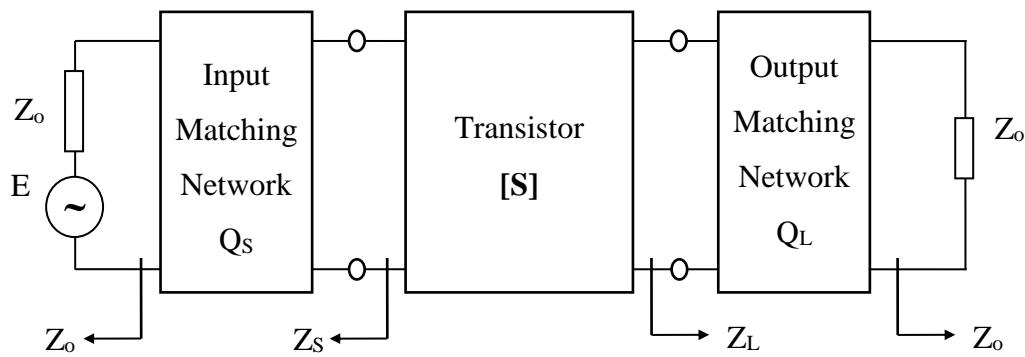


Fig. IV-27. Block diagram of a microwave amplifier with matching networks.

I – Matching networks for narrow band amplifiers

For narrow band amplifiers, the analysis of matching circuits at microwave frequencies can be cumbersome in analytical form. Figure IV-28 shows some basic narrow band matching networks. Note that the networks are based on reactive components (LC) to avoid losses.

II – Solved example

In this section, we present some basic derivations for matching networks. The goal is to match the impedance $Z = R + jX$ to $Z_o = R_o = 50 \Omega$

$$a - R < R_o, \text{ i.e., } R < 50 \Omega$$

The appropriate network is shown on Fig. IV-29.

We have

$$Z = R + jX = \frac{1}{\frac{1}{R_o} + jX_1} + jX_2 = \frac{1}{\frac{1}{50} + jX_1} + jX_2$$

$$\rightarrow R + jX = \frac{R_o}{1 + R_o^2 X_1^2} + j \left[X_2 - \frac{R_o^2 X_1^2}{1 + R_o^2 X_1^2} \right]$$

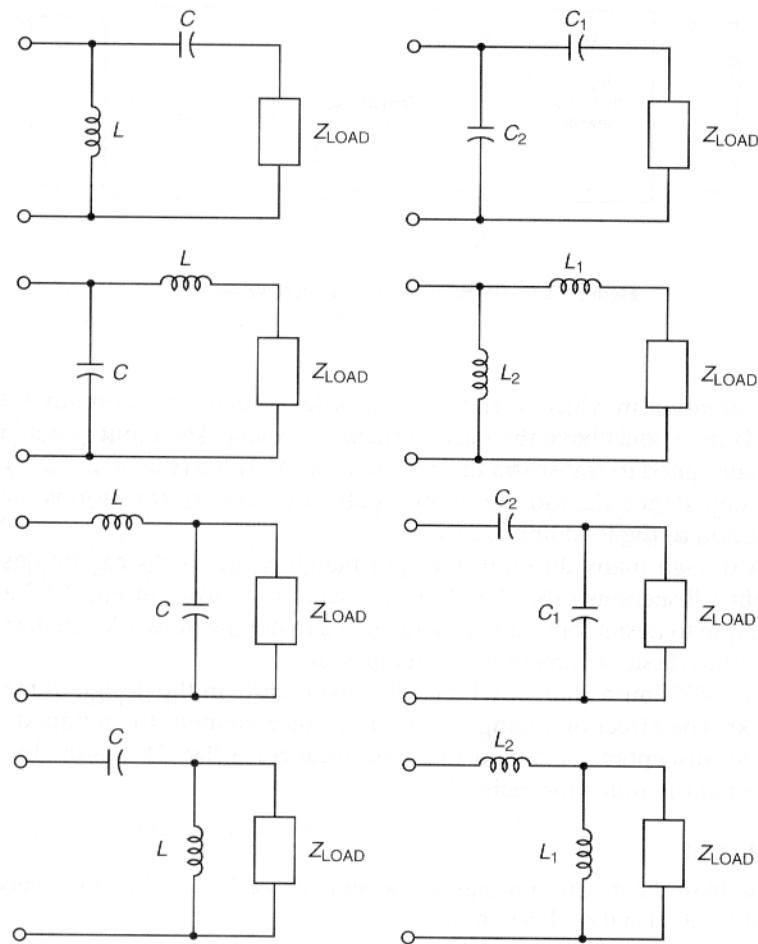


Fig. IV-28. Matching networks for narrow band amplifiers.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

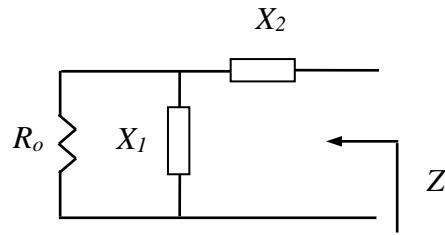


Fig. IV-29. Matching network for $R < 50 \Omega$.

Separating real and imaginary parts gives

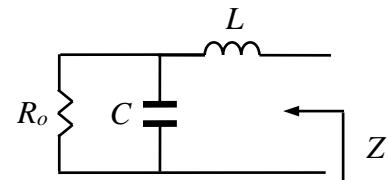
$$R = \frac{R_o}{1 + R_o^2 X_1^2} \quad X = X_2 - \frac{R_o^2 X_1^2}{1 + R_o^2 X_1^2}$$

Therefore,

$$\rightarrow \quad X_1 = \frac{\sqrt{R_o - R}}{R_o \sqrt{R}} \quad X_2 = X + \sqrt{R[R_o - R]}$$

Example: at 550 MHz, $Z = (8.1 + j 0) \Omega$

The corresponding matching network is



$$X_1 = \frac{\sqrt{50 - 8.1}}{50 \sqrt{8.1}}$$

X_1 is shunt admittance:

$$\rightarrow \quad X_1 = C \omega = \frac{\sqrt{50 - 8.1}}{50 \sqrt{8.1}} \quad \rightarrow \quad C = \frac{1}{2 * \pi * 550 * 10^6} \frac{\sqrt{50 - 8.1}}{50 \sqrt{8.1}} = 13.2 \text{ pF}$$

$$X_2 = 0 + \sqrt{8.1[50 - 8.1]} \quad X_2 \text{ is series impedance:}$$

$$\rightarrow \quad X_2 = L\omega = 0 + \sqrt{8.1[50 - 8.1]}$$

$$\rightarrow \quad L = \frac{1}{2 * \pi * 550 * 10^6} \sqrt{8.1[50 - 8.1]} = 5.3 \text{nH}$$

b - R > R_o, i.e., **R > 50 Ω**

The appropriate network is shown on Fig. IV-30.

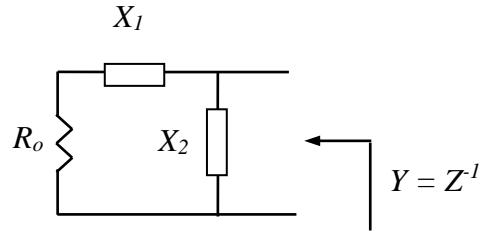


Fig. IV-30. Matching network for $R > 50 \Omega$.

We have

$$Y = G + jB = \frac{1}{R_o - jX_1} - jX_2 = \frac{1}{50 - jX_1} - jX_2$$

$$\rightarrow \quad G + jB = \frac{R_o}{R_o^2 + X_1^2} + j \left[\frac{X_1}{R_o^2 + X_1^2} - X_2 \right]$$

Separating real and imaginary parts gives

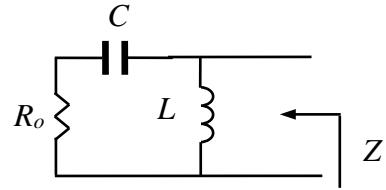
$$G = \frac{R_o}{R_o^2 + X_1^2} \quad B = \frac{X_1}{R_o^2 + X_1^2} - X_2$$

$$\rightarrow \quad X_1 = \sqrt{R_o \left(\frac{1}{G} - R_o \right)} \quad X_2 = \sqrt{G \left(\frac{1}{R_o} - G \right) - B}$$

Example: at 550 MHz, $Z = (122.3 + j 155.4) \Omega$ or $Y = (3.127 - j 3.974) \text{ mS}$

The matching network is then

$$X_1 = \sqrt{R_o \left(\frac{1}{G} - R_o \right)}$$



As we started with $(-jX_1)$ then:

$$\rightarrow \quad X_1 = \frac{1}{C\omega} = \sqrt{50 \left(\frac{1}{3.127 \cdot 10^{-3}} - 50 \right)}$$

$$\rightarrow \quad C = \frac{1}{2 * \pi * 550 * 10^6} \left[50 \left(\frac{1}{3.127 \cdot 10^{-3}} - 50 \right) \right]^{-\frac{1}{2}} = 2.49 \text{ pF}$$

$$X_2 = \sqrt{G \left(\frac{1}{R_o} - G \right) - B}$$

As we started with ($-jX_2$) then:

$$\rightarrow X_2 = \frac{1}{L\omega} = \sqrt{G\left(\frac{1}{R_o} - G\right)} - B$$

$$\rightarrow L = \frac{1}{2 * \pi * 550 * 10^6 \left[\sqrt{3.127 * 10^{-3} \left(\frac{1}{50} - 3.127 * 10^{-3} \right)} - (-3.974 * 10^{-3}) \right]}$$

$$L = 25.75 \text{ nH}$$

III – Matching networks for relatively narrow band amplifiers

For amplifiers with a relatively narrow band, designers can use T or Π matching networks as shown in Figure IV-31.



Fig. IV-31. T or Π matching networks.

IV – Matching networks for wide-band amplifiers

For wideband amplifiers, a designer would need an input/output matching over a large bandwidth. Accordingly, the impedance will be not constant but will vary with frequency in the form of a transfer function like { $Z(j\omega) = Z(s)$ }. We have then to extract the number of poles and zeros to determine the matching cell.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Let $[S(j\omega)]$ be the scattering matrix of the two-port network Q (Figure IV-32).

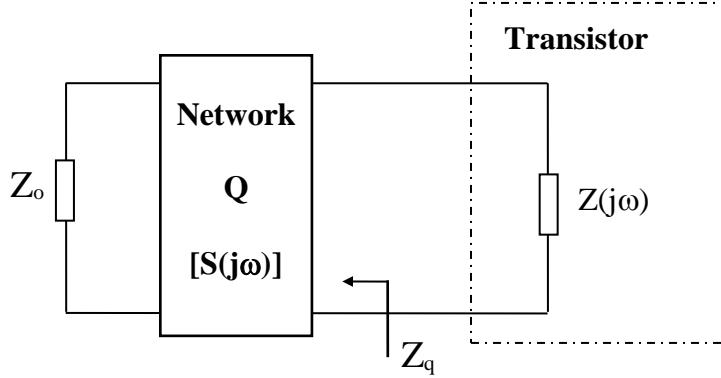


Fig. IV-32. Block diagram for wideband matching networks.

The power gain of the network is equal to

$$G(\omega^2) = |S_{21}(j\omega)|^2 = 1 - |S_{22}(j\omega)|^2 = 1 - \left| \frac{Z_q(s) - Z(-s)}{Z_q(s) + Z(-s)} \right|^2 \quad (\text{IV-117})$$

where $Z_q(s)$ is the impedance seen by the transistor. The network is now equivalent to a wideband network. Therefore, the procedure to obtain the configuration of the matching two-port network can be summarized in the following steps:

- * Verify that the function $G(\omega^2)$ is rational wrt ω^2 and satisfy to passive system criteria :

$$0 \leq G(\omega^2) \leq 1$$

- * Determine the reflection coefficient $\rho(s)$ with minimum phase using spectral factorization

$$\rho(-s) \rho(s) = 1 - G(-s^2) = \frac{N(s^2)}{M(s^2)}$$

- * Expand $N(s^2)$ and $M(s^2)$ as a product of two polynomials

$$\frac{N(s^2)}{M(s^2)} = \frac{n(-s)n(s)}{m(-s)m(s)} \quad (\text{IV-118})$$

where $n(s)$ and $m(s)$ are Hurwitz's polynomials built from the zeros of $N(s^2)$ and $M(s^2)$ (Appendix IV-4). $\rho(s)$ can be written relatively to an all-pass function $B(s)$:

$$\rho(s) = \pm \frac{n(s)}{m(s)} = \prod_{i=1}^m \frac{s - s_i}{s + s_i} S_{22}(s) = B(s) S_{22}(s) \quad (\text{IV-119})$$

s_i are the poles of $Z(-s)$ for $\{\operatorname{Re}(s) > 0\}$. Substituting S_{22} in equation (IV-117) gives

$$Z_q(s) = \frac{[Z(s) + Z(-s)]B(s)}{B(s) - \rho(s)} - Z(s) \quad (\text{IV-120})$$

- * Use the impedance given by equation (IV-120) as the input impedance of the matching network, as shown in Figure IV-32. Thus, synthesize the network.

I – DC BIAS NETWORKS FOR AMPLIFIERS

In some technical papers, it has been said that the least considered factor in microwave amplifier design is the bias network. While considerable effort is sent in designing the amplifier for a given specific gain and noise figure, little effort is spent on the dc bias network. But we should have in mind that the cost per decibel of power gain or noise figure is high in microwaves, so we cannot sacrifice the amplifier performance by a poor dc bias design.

The purpose of a good dc bias design is to select the proper quiescent point and hold the quiescent point constant over variations in transistor parameters and temperature. A resistor bias network can be used with good results over moderate temperature changes. However, an active bias network is usually preferred for large temperature changes.

I – BJT bias networks

At low frequencies, a bypassed emitter resistor is an important contribution to the Q point stability. At microwave frequencies, the by-pass capacitor (in parallel with the emitter resistor) can produce oscillations by making the input port unstable at some frequencies, mainly because of nonlinear behaviors in the microwave circuit. Furthermore, an emitter resistor will degrade the noise performance of the amplifier. Therefore, in most microwave transistor amplifiers, the emitter lead of the transistor is directly grounded.

At microwave frequencies, the transistor parameters that are affected most by temperature are I_{CBO} , h_{FE} (or β), and V_{BE} .

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Since the collector current source is the main nonlinear element in the transistor, we note such temperature-sensitive parameter as

$$I_C = f(I_{CB0}, h_{FE}, V_{BE}) \quad (\text{IV-121})$$

Thus, defining stability factors as

$$S_i = \left. \frac{\Delta I_C}{\Delta I_{CB0}} \right|_{\substack{\Delta h_{FE} = 0 \\ \Delta V_{BE} = 0}} \quad (\text{IV-122})$$

$$S_{h_{FE}} = \left. \frac{\Delta I_C}{\Delta h_{FE}} \right|_{\substack{\Delta I_{CB0} = 0 \\ \Delta V_{BE} = 0}} \quad (\text{IV-123})$$

$$S_{V_{BE}} = \left. \frac{\Delta I_C}{\Delta V_{BE}} \right|_{\substack{\Delta h_{FE} = 0 \\ \Delta I_{CB0} = 0}} \quad (\text{IV-124})$$

allows to reformulate equation (IV-121) as

$$\Delta I_C = S_i \Delta I_{CB0} + S_{h_{FE}} \Delta h_{FE} + S_{V_{BE}} \Delta V_{BE} \quad (\text{IV-125})$$

For a given dc bias network, the stability factors can be determined and thus, equation (IV-125) can be used to predict the variations of the collector current wrt temperature. Figure IV-33 shows two practical grounded-emitter bias networks that can be used at microwave frequencies.

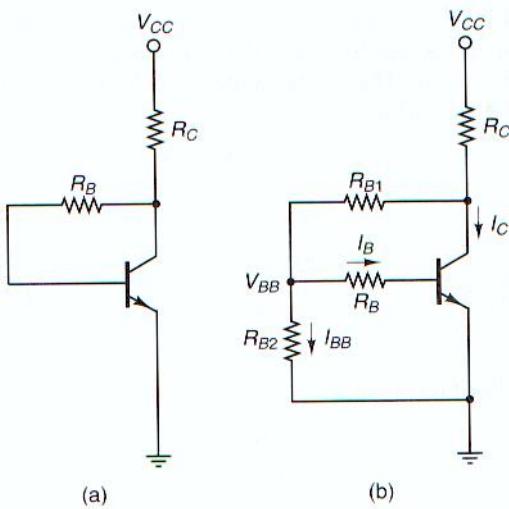


Fig. IV-33. Two dc bias networks for BJTs. (a) Voltage feedback bias network, (b) Voltage feedback bias network with constant-base current source.

II – FET bias networks

The FET can be biased in several ways. Five basic dc network configurations for FET amplifiers are shown in Figure IV-34.

IMPORTANT: The column “how” in Figure IV-34 indicates the polarity of the sources, as well as the sequence in which the voltages must be applied to prevent transient burnout of the device during turn-on.

Figure	How	Amplifier characteristics	Power supply used
(a)	Apply V_G , then V_D	Low noise High gain High power High efficiency	Bipolar, Minimum source inductance
(b)	Apply V_S , then V_D	[same as (a)]	Positive supply
(c)	Apply V_S , then V_G	[same as (a)]	Negative supply
(d)	Apply V_D	Low noise High gain High power Lower efficiency Gain easily adjusted by varying R_S	Unipolar, incorporating R_S automatic transient protection
(e)	Apply V_G	[same as (d)]	Negative unipolar, incorporating R_S

Fig. IV-34. Five basic dc network configurations for FET amplifiers.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

J - POWER AMPLIFIERS

So far, we have presented design techniques based on small-signal S-parameters for maximum or arbitrary user-defined gain and/or noise figure. The small-signal S-parameters are not useful for power amplifier design because such amplifiers operate in nonlinear regions. However, the small-signal approaches can be still used for large-signal amplifiers operating in class A (i.e., with an almost linear output power) but they are not suitable for class AB, B, C, D, E, F

A set of large-signal S parameters is required to properly characterize the transistor behavior for power applications. Unfortunately, measuring such parameters is difficult. Therefore, an alternative set of large-signal parameters is needed to characterize the component. This can be done by providing information about source and load reflection coefficients as functions of output power and gain. Figure IV-35 shows typical large-signal reflection coefficients.

One set of large-signal parameters provided by manufacturers is the measurement of the source and load reflection coefficients together with the output power, when the transistor is operated at its 1-dB compression point. The 1-dB gain compression point is defined as

$$G_{1dB}(dB) = G_o(dB) - 1 \quad (\text{IV-126})$$

where $G_o(dB)$ is the small signal linear power gain in decibels. Since the power gain G_p is defined as P_{out}/P_{in} , we can write the output power at the 1-dB compression point (Figure IV-36)

$$P_{out}(dBm) = G_p(dB) + P_{in}(dBm) \rightarrow P_{1dB}(dBm) = G_{1dB}(dB) + P_{in}(dBm) \quad (\text{IV-127})$$

$$\rightarrow G_{1dB} = \frac{P_{1dB}}{P_{in}} \rightarrow P_{1dB}(dBm) - P_{in}(dBm) = G_o(dB) - 1 \quad (\text{IV-128})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

In Figure IV-37, the dynamic range is that range where the amplifier has a linear power gain. This range is limited at low power levels by the noise level. An input signal $P_{i,mds}$ is detectable only if its output power level $P_{o,mds}$ is above the noise power level. Since the thermal noise power level of a two-port, with noise figure F , is given by

$$P_{N_o} = kT B G_A F \quad (\text{IV-129})$$

and observing that $kT = -174$ dBm (or $kTB = -174$ dBm/Hz at $T = 290^{\circ}\text{K}$), and assuming that the minimum detectable input signal is X decibels above thermal noise (a typical value of $X(\text{dB})$ is 3dB), we can write

$$P_{i,mds} = -174 \text{ dBm} + 10 \log B + F(\text{dB}) + X(\text{dB}) \quad (\text{IV-130})$$

$$P_{o,mds} = -174 \text{ dBm} + 10 \log B + F(\text{dB}) + X(\text{dB}) + G_A(\text{dB}) \quad (\text{IV-131})$$

I – Fundamental considerations in power amplifier design: Class-A amplifiers

Figure IV-38 shows a simplified circuit of a FET power amplifier.

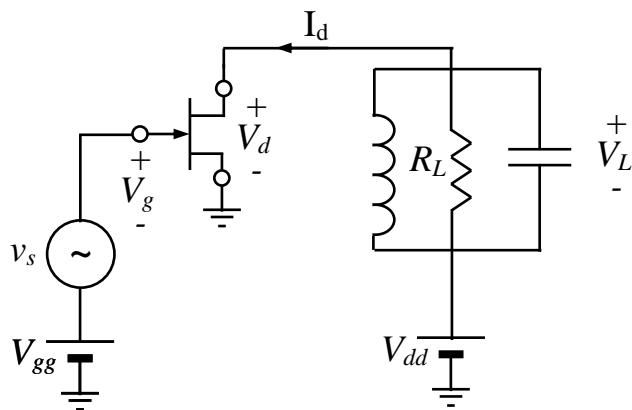


Fig. IV-38. Equivalent circuit of an ideal FET power amplifier.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

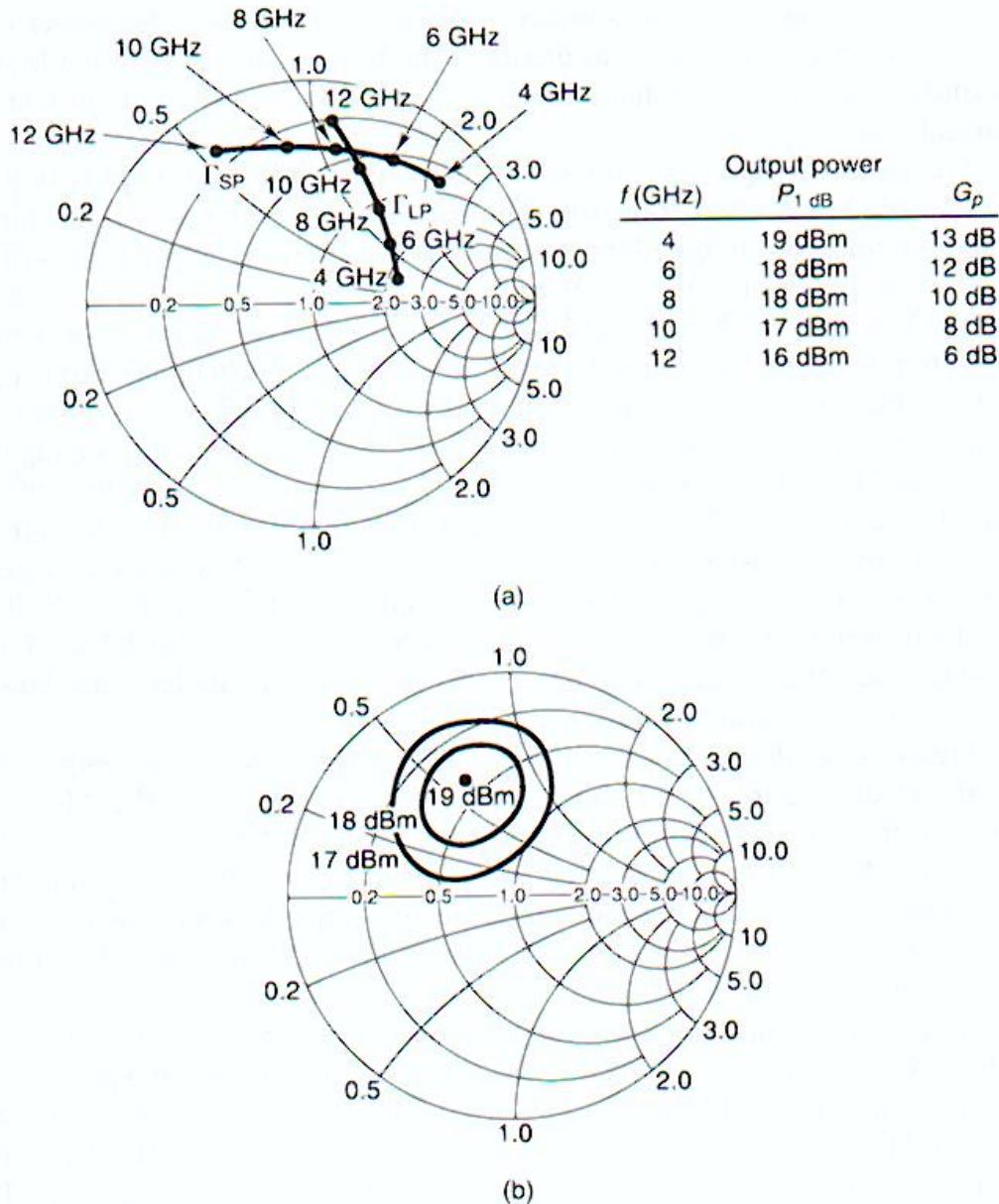


Fig. IV-35. (a) Typical large-signal reflection coefficients. (b) Typical output power contours as function of Γ_L .

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

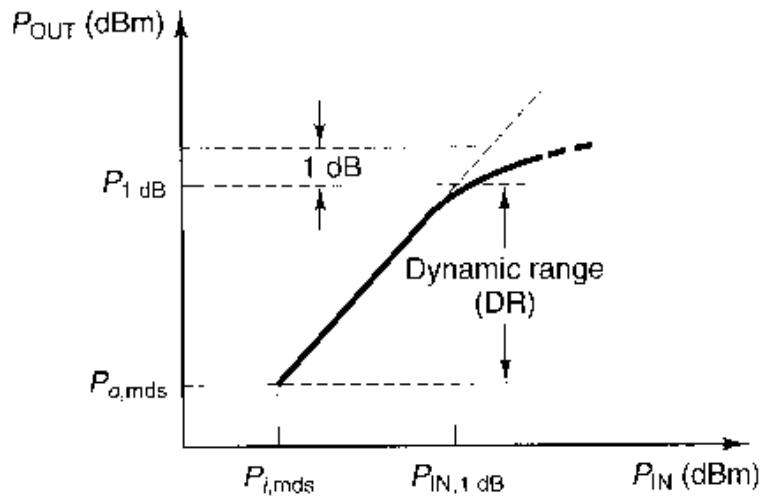


Fig. IV-36. 1-dB compression point and dynamic range of microwave amplifiers.

It consists of FET, excitation and bias sources, tuned circuit and load R_L . Ideally, we assume that the FET is equivalent to an ideal transconductance (since the most important nonlinear parameter is the controlled current source) without resistive or reactive parasitics ($V_g = V_{gg}$ and $V_d = V_{dd}$). The tuned circuit is resonant at the excitation frequency. The bias current is I_{dd} . The application of a sinusoidal excitation $v_s(t)$ to the gate generates an RF component of drain current, $\Delta i_d(t)$. The RF component of the drain voltage, $\Delta v_d(t)$, is equal to the voltage drop across R_L , i.e.,

$$v_L(t) = \Delta v_d(t) = -\Delta i_d(t) R_L \quad (\text{IV-132})$$

Each curve of the FET I-V characteristics (Fig. IV-39) represents a range of values for drain current and voltage. Since we wish to maximize the power delivered to the load, we can maximize the excursion of both $\Delta v_d(t)$ and $\Delta i_d(t)$, with

$$|v_L(t)| = V_{dd} \quad |i_L(t)| = I_{dd} \quad (\text{IV-133})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

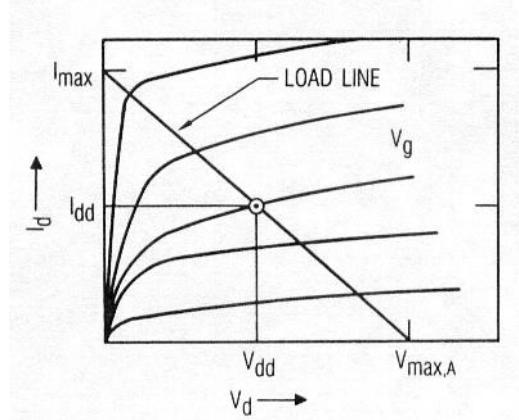


Fig. IV-39. Drain I-V characteristics.

The load line dictates that these conditions are met when

$$R_L = \frac{V_{\max, A}}{I_{\max}} = \frac{V_{dd}}{I_{dd}} \quad (\text{IV-134})$$

Then, the drain voltage will vary from 0 to $2V_{dd}$ (same for the current: from 0 to $2I_{dd}$). Under these conditions, the output power is equal to

$$P_L = 0.5 |v_L(t)| |i_L(t)| = 0.5 V_{dd} I_{dd} \quad (\text{IV-135})$$

$$\rightarrow P_{L\max} = \frac{1}{2} \left(\frac{1}{2} V_{\max, A} \right) \left(\frac{1}{2} I_{\max} \right) = \frac{1}{8} V_{\max, A} I_{\max} \quad (\text{IV-136})$$

The dc current remains constant at I_{dd} at all excitation levels; therefore, by defining the dc power as { $P_{dc} = V_{dd} I_{dd}$ }, the dc-RF conversion efficiency will be equal to

$$\eta_{dc} = \frac{P_L}{P_{dc}} = \frac{0.5 * V_{dd} * I_{dd}}{V_{dd} * I_{dd}} = 50\% \quad (\text{IV-137})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

An amplifier operating in this way is called a **Class-A amplifier**. In theory, the maximum efficiency of such amplifier is 50%. In practice, this reasoning is too simple. First, it is not possible to vary the drain voltage and current all the way to the peak of the load line, where $I_d = I_{max}$ and $V_d = 0$ because of the knee in the uppermost of the I-V curves. The result is that $|v_L(t)|$ cannot reach V_{dd} (similarly for the drain current). Second, the FET is non linear, so the current waveform is not sinusoidal. The tuned circuit constrains $i_L(t)$ to be sinusoidal, however, so in practice

$$|i_L(t)| < |\Delta i_d(t)| \quad (\text{IV-138})$$

Then, two undesirable characteristics of the Class-A amplifier are the relatively low efficiency and the dissipation of a large amount of power even when it is not excited.

II – Class-B amplifiers

Many of the disadvantages of Class-A are circumvented by **Class-B** operation. For such amplifiers, the transistor conducts during half of the input signal cycle. This is accomplished by biasing the transistor at cut-off. The gate-bias voltage is set at the turn-on voltage V_t . Thus, the quiescent drain current is zero and so, the FET dissipates no power in the absence of excitation.

Because the harmonic components of i_d , it is not possible to draw a true load line for Class-B amplifiers. The dc drain current is then the average of the half-cosine waveform. From Fourier analysis, we can find that under full excitation,

$$I_{dc} = \frac{I_{\max}}{\pi} \quad \rightarrow \quad P_{dc} = V_{dd} \frac{I_{\max}}{\pi} \quad (\text{IV-139})$$

The tuned circuit allows only the fundamental component of $i_d(t)$ to pass.

The power delivered to the load is consequently

$$P_L = \frac{1}{2} |i_L(t)| * |v_L(t)| = \frac{1}{2} I_1 * |v_L(t)| \quad (\text{IV-140})$$

where I_1 is the magnitude of the fundamental component of the load current and equals to $0.5 * I_{max}$. From Figure IV-40, we have

$$|v_L(t)| = |\Delta v_d(t)| = V_{dd} \quad (\text{IV-141})$$

$$\rightarrow P_L = \frac{1}{2} I_1 * |v_L(t)| = \frac{1}{2} \left(\frac{1}{2} I_{max} \right) V_{dd} = \frac{1}{4} I_{max} V_{dd} \quad (\text{IV-142})$$

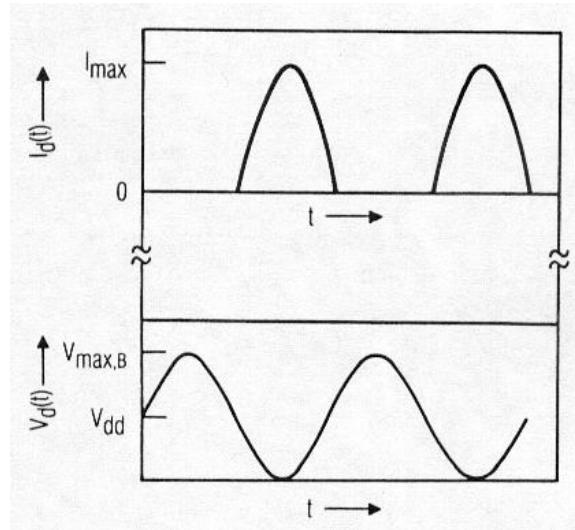


Fig. IV-40. Drain voltage and current waveforms in the ideal Class-B amplifier.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

and the dc-RF efficiency is

$$\eta_{dc} = \frac{P_L}{P_{dc}} = \frac{\pi}{4} = 0.78 = 78\% \quad (\text{IV-143})$$

This improvement (compared to Class-A) is achieved by allowing the channel to conduct during only half the period of the excitation; during time that the transistor is off, it dissipates no power. Since $V_{max,B} = 2V_{dd}$, the maximum output power for Class-B amplifiers is

$$P_{L\max} = \frac{1}{2} \left(\frac{1}{2} V_{max,B} \right) \left(\frac{1}{2} I_{\max} \right) = \frac{1}{8} V_{max,B} I_{\max} \quad (\text{IV-144})$$

which is the same as that of the Class-A amplifier if $V_{max,A} = V_{max,B}$. In order to achieve the maximum output power, the load resistance must be

$$R_L = \frac{V_{dd}}{I_1} = \frac{V_{dd}}{0.5 * I_{\max}} = \frac{2 V_{dd}}{I_{\max}} = \frac{V_{max,B}}{I_{\max}} \quad (\text{IV-145})$$

which is also the same as that of the Class-A amplifier. Furthermore, because in both amplifiers the load and the fundamental component of the load current are the same, the output power must also be the same.

a- Disadvantages of the Class-B amplifier

In practice, because the maximum drain voltage is limited by the gate-drain avalanche breakdown, $V_{max,A}$ is always greater than $V_{max,B}$. In Class-A, the maximum drain-gate voltage occurs when $V_d = V_{max,A}$ and $V_g = V_t$.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Thus, if V_a is the drain-gate avalanche breakdown voltage:

$$V_{\max,A} = V_a - |V_t| \quad (\text{IV-146})$$

The Class-B amplifier is biased at $V_{gg} = V_t$, so the maximum negative excursion of V_g is $2V_t$. Then,

$$V_{\max,B} = V_a - 2|V_t| \quad (\text{IV-147})$$

So $V_{\max,B}$ is less than $V_{\max,A}$. Accordingly, the maximum output power of a Class-B is slightly lower than that of a Class-A using the same device. Moreover, we can also note that the gate voltage of a Class-A amplifier varies between 0 and V_t while in a Class-B amplifier the gate voltage varies between zero and $2V_t$.

More input power is required to achieve the Class-B amplifier's wider gate-voltage variation but the output is nearly the same; thus Class-B amplifiers have inherently lower gain than Class-A amplifiers.

b- Effect of harmonic generation

Another disadvantage of the Class-B amplifier is that it generates a high level of harmonics in the drain current by switching the FET on and off during each excitation cycle. If the device is terminated in the same impedance at the fundamental and second-harmonic frequencies, the second-harmonic output of an ideal Class-B amplifier is a few dB below the fundamental output. One solution to the problem of harmonics is to use a “push-pull” configuration, in which the excitation is applied out of phase to the inputs of two Class-B amplifiers and the outputs are

combined out of phase. The phase shift of the output combiner must be 180° at the harmonic frequencies as well as the fundamental frequency. This configuration can reduce significantly the levels of even harmonics.

III – Class-AB amplifiers

For all the above reasons, we rarely operate microwave power FETs in a true Class-B mode. So called Class-B microwave amplifiers are usually biased near $0.1I_{dss}$ and are actually operated in a mode somewhere between Class-B and Class-A.

Conversely, Class-A amplifiers are often not operating in a classical Class-A mode. They are sometimes biased to a minimal current level and driven well into saturation. Both types of operation are called ***Class-AB***.

In practice, Class-AB amplifiers represent a relatively good compromise between the extremes of either class:

- Class-AB amplifiers usually have better efficiency than Class-A amplifiers and better gain than Class-B amplifiers.
- The linearity of Class-A operation with the efficiency of Class-B operation can be attained using two Class-B transistors connected in a complementary amplifier configuration.

When working at relatively low frequencies, i.e., below a few gigahertz, most Class-B amplifiers use bipolar transistors (Figure IV-41). However, MESFET and HEMT Class B amplifiers are now widely used in the centimeter- and millimeter-wave ranges (see next section).

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

The bipolar amplifier shown in Figure IV-41-a suffers from crossover distortion since the base-to-emitter junction of each transistor must be forward biased before the transistor operates in its linear region. A type of Class-AB is obtained if the transistors in Figure IV-41-a are biased at a Q point with a voltage $V_{BE} = 0.6V \sim 0.7V$ so that the crossover distortion is almost eliminated. A class-AB complementary amplifier is shown in Figure IV-41-b.

IV – Class-C amplifiers

As for Class-B, **Class-C** amplifiers are based mainly on bipolar transistors when working below the X band (Figure IV-42). In Class-C, the transistor is cut off until the ac signal between the base and emitter makes it conduct. The emitter is grounded and the input RFC sets the quiescent value of the base to emitter voltage at zero. For the input ac signal, the input RFC is an open circuit and the input signal applied between the base and emitter makes the transistor conduct. The conduction angle is less than 180° , about 140° for good efficiency and low harmonic content at the output.

The output-matching network must have a high Q value in order to suppress the harmonics and pass the amplified fundamental signal. Efficiency is certainly better in Class-C operation, but it is difficult to calculate.

Typical characterization of transistors fabricated to operate in Class-B and Class-C operation is at V_{CE} voltages of 12.5V, 28V, and 50V:

- The 12.5V amplifiers find applications in mobile transmitter equipment such as those used by police cars, trucks, and taxis.
- The 28V and 50V amplifiers find applications in high-power base-station transmitters and receivers.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

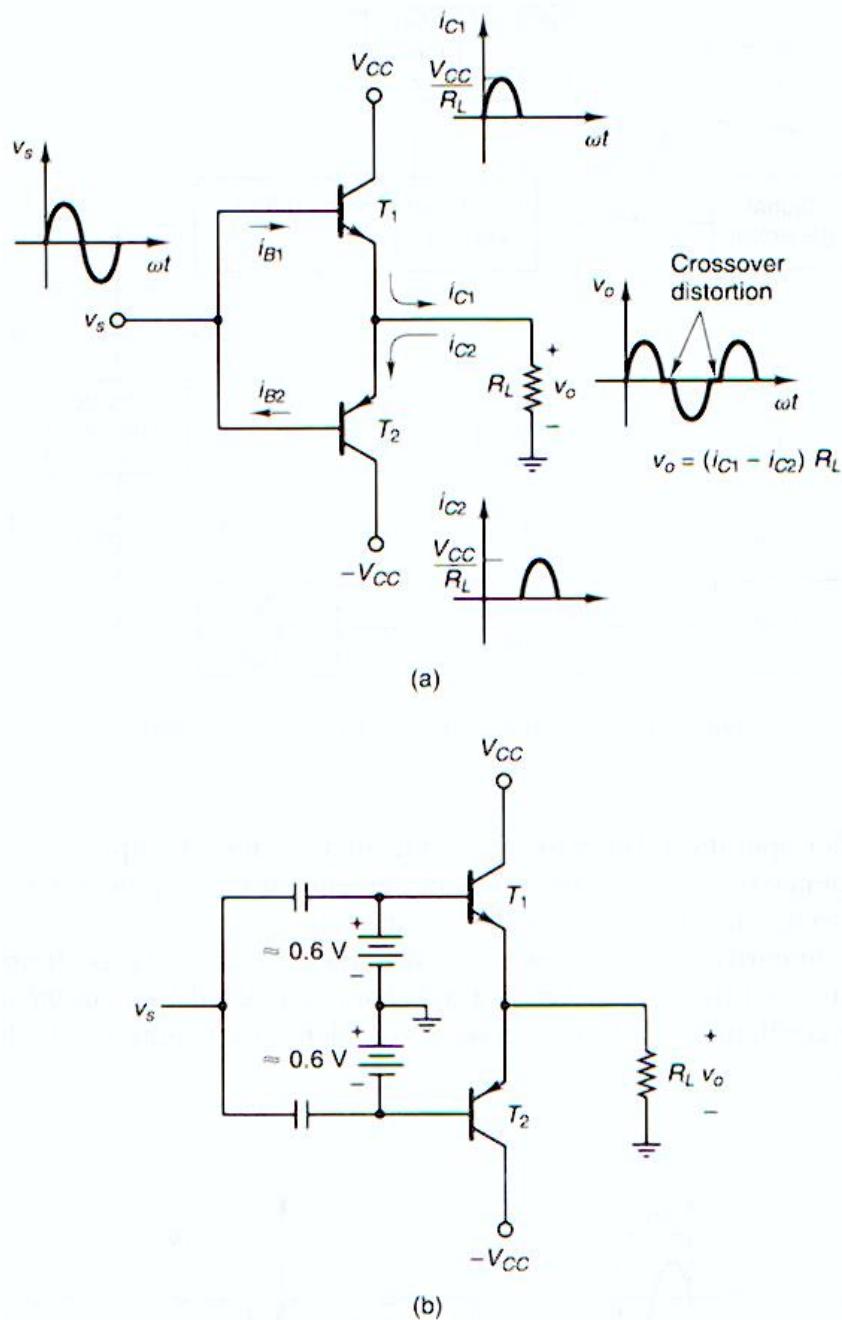


Fig. IV-41. (a) Basic Class-B complementary amplifier versus (b) Basic Class-AB complementary amplifier.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

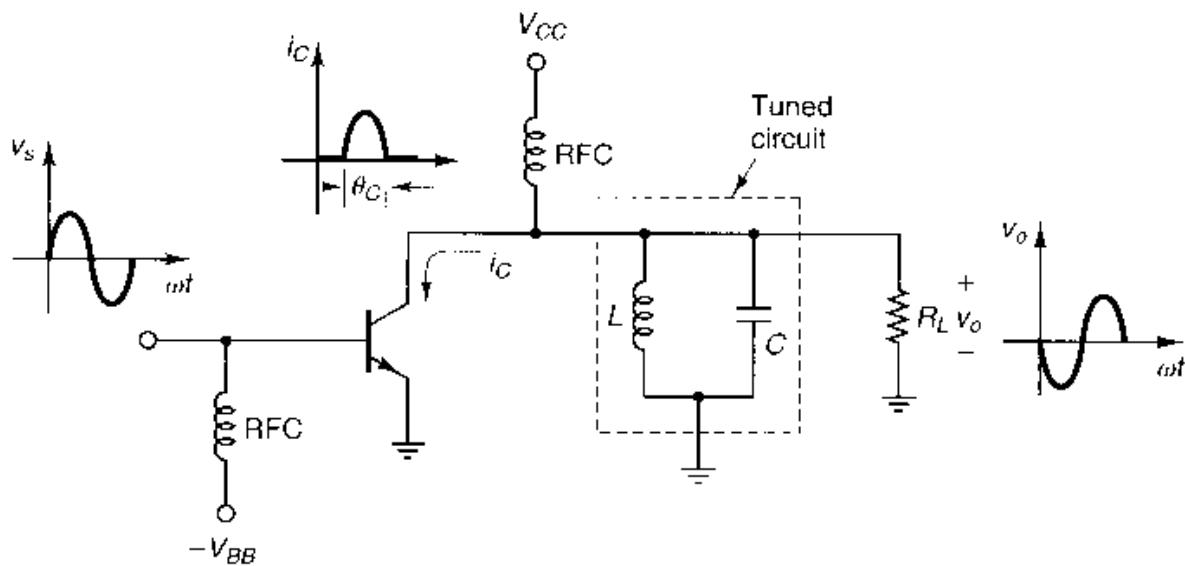


Fig. IV-42. A basic Class-C amplifier.

K – APPROXIMATE DESIGN OF CLASS-A FET AMPLIFIERS

The first step in the design of a Class-A FET amplifier is to select an appropriate transistor. As a general rule, at frequencies below 15 GHz, most high-quality devices can produce output powers, as a function of gate width, of 0.3 W/mm at 1-dB compression point and 0.5 W/mm when saturated (however, many experimental devices have produced more than 1 W/mm).

In designing the Class-A power amplifier, one have to remember that the transistor should exhibit an ideal linear transfer function. Therefore, although a practical Class-A amplifier is weakly nonlinear, we should be able to rely fairly on linear-amplifier theory we developed above, *in the initial approximate design step*.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

The fundamental problem in designing a Class-A amplifier, as in designing a small-signal amplifier, is to pick the appropriate source and load impedances:

- ⇒ ***In a power amplifier, the load impedance must be selected to achieve the desired output power and the source impedance must provide a conjugate input match.***
- ⇒ ***Additionally, we must select a bias point that results in both adequate power and good efficiency.***

We base our design on the terminal I-V characteristics of a FET (i.e., using the terminal voltages V_{gs} and V_{ds} instead of the internal voltages V_g and V_d) as shown in Fig. IV-43. It is preferable to use the internal I-V curves, but they are not accessible through measurements, and recognizing that this initial design is approximate, we shall accept a plot of the FET's terminal as an approximation of the internal ones.

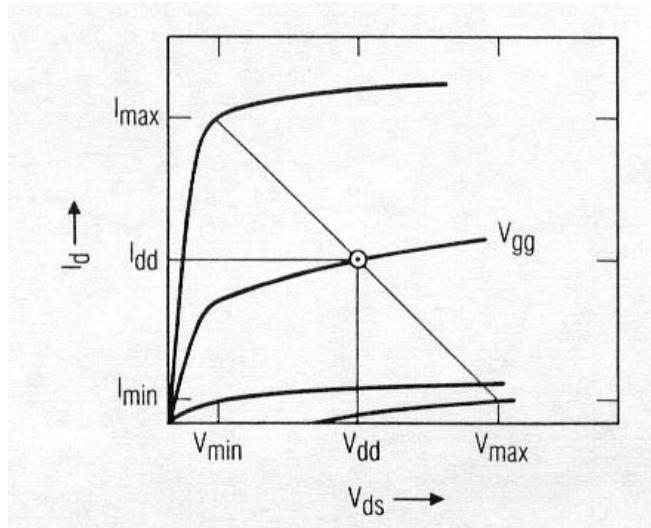


Fig. IV-43. Terminal I-V characteristics of a FET.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

We use the load line to select the load conductance. To achieve so, we must take into account the limits on the drain voltage and current:

- V_{min} : The minimum drain-source voltage is limited by the knee of the I-V characteristics at $V_g = 0.6V$ (i.e., approximately 1.5V).
- V_{max} : The maximum drain-source voltage is fixed by the physical limitations of the transistor (given in data sheets provided by the manufacturer).
- I_{min} : Because of the variation in V_t with V_d , and the gate-drain avalanche limitation, V_d cannot be driven to the point where $I_d = 0$. Thus there is a finite drain current I_{min} at V_{max} , the maximum drain-source voltage (given in data sheets provided by the manufacturer).
- I_{max} : The maximum drain current is limited accordingly to the given transistor performance (given in data sheets provided by the manufacturer).

Based on these values, the quiescent dc voltage V_{dd} is halfway between V_{max} and V_{min} . The quiescent dc current I_{dd} is halfway between I_{max} and I_{min} .

The load conductance is equal to the slope of the load line:

$$G_L = \frac{I_{max} - I_{min}}{V_{max} - V_{min}} \quad (\text{IV-148})$$

When an unpackaged FET is biased in its saturation region (for a packaged FET, the load admittance is more complex to determine due to the presence of parasitics, but the approach is still the same), the drain-source capacitance C_{ds} is the dominant component of the output admittance. Because we wish to present a real load of conductance G_L to the terminals of the controlled source

I_d , the susceptance of the load must resonate with C_{ds} . Thus, the initial good estimate of the load admittance is

$$Y_L = G_L - j\omega C_{ds} \quad (\text{IV-149})$$

Because the load at the current source is real, the ac part of $V_d(t)$, which equal the load voltage $V_L(t)$), and the load current ($I_L(t) = -I_d(t)$) are in phase. Therefore, the output power is

$$P_L = \frac{1}{2} \left[\frac{1}{2} (V_{\max} - V_{\min}) \right] \left[\frac{1}{2} (I_{\max} - I_{\min}) \right] \quad (\text{IV-150})$$

Note: Below the compression point, the dc power is $\{ P_{dc} = V_{dc} * I_{dc} \}$. Above the compression point, the dc power is greater but much of it is converted to RF output power. Thus, the *quiescent dc power can be considered an upper limit to the power dissipated by the device.*

The input of the power FET amplifier is designed to be conjugately matched. We can determine the corresponding input impedance of the FET by using the small-signal S-parameters and the approach detailed in the first part of this chapter.

After the source and load matching impedances are determined, the matching networks can be designed at the fundamental frequency. This latter requirement looks not very severe for Class-A amplifiers because the second and higher harmonic currents are not very significant. *But we have to remember that the source and load impedances are generally very low and a combination of low impedances and high current densities requires careful consideration because of power dissipation.* Capacitors, even those used for such purposes as dc blocking, must have high quality factors, and inductors should not be made from narrow microstrips or fine wire (gold ribbon is a good material

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

for inductors that must carry high currents). Thus, the topology of the matching circuit can often be selected to minimize the currents in components with relatively high loss.

Later, we can estimate the source and load impedance at higher harmonics by analyzing these matching circuits. A general circuit simulator can be used for that purpose. The final step is to check the final circuit via a harmonic-balance analysis to tune the approximate values.

L – APPROXIMATE DESIGN OF CLASS-B FET AMPLIFIERS

The design of Class-B amplifiers is similar to the one detailed for Class-A amplifiers. The load impedance for an ideal Class-B amplifier is the same as that of an ideal Class-A amplifier having the same output power. However, it is not possible to estimate the linear gain or input impedance of a Class-B amplifier for small-signal S-parameters. Instead, we have to use the equivalent circuit model of the transistor to estimate the input impedance and the harmonic-balance to determine the gain.

The maximum value of V_d is lower than that of a Class-A amplifier. In FET amplifiers that are limited by gate-drain avalanching, the output power in Class-B operation is lower than in Class-A operation. Otherwise, the output powers are nearly identical. Thus, the same procedure can be used to determine the load impedance as long as V_{max} is chosen to have its Class-B value.

The dc drain current of a Class-A amplifier under full excitation is estimated to I_{max}/π (at the 1-dB saturation point). Thus, the dc power dissipation is

$$P_d = V_{dd} \frac{I_{max}}{\pi} \quad (\text{IV-151})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

In an ideal Class-A amplifier, V_g varies between V_t and the threshold of gate conduction ($\approx 0.5V$). In a Class-B amplifier, the range is between $2V_t$ and the same maximum voltage. Therefore, in order to deliver the same output power, the Class-B amplifier, requires twice the voltage across C_{gs} as the Class-A; accordingly, the Class-B input power must be 6dB greater. In practice, the difference in gain between Class-B and Class-A amplifiers using the same FET is usually 3 to 5 dB.

An adequate initial estimate of the input impedance of the Class-B amplifier is

$$Z_{in} = R_g + R_i + R_s + \frac{\langle g_m \rangle L_s}{C_{gs}(V_t)} + j \left[\omega L_s - \frac{1}{\omega C_{gs}(V_t)} \right] \quad (\text{IV-152})$$

The term $\langle g_m \rangle$ is the transconductance averaged over the excitation cycle ($\approx 20\%$ of the peak transconductance).

M – DESIGN EXAMPLES

We shall use an X-band power MESFET (gate width 2.4mm and output power capability of 1W) which equivalent circuit is shown in Fig. IV-44. The operating frequency is 10GHz.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

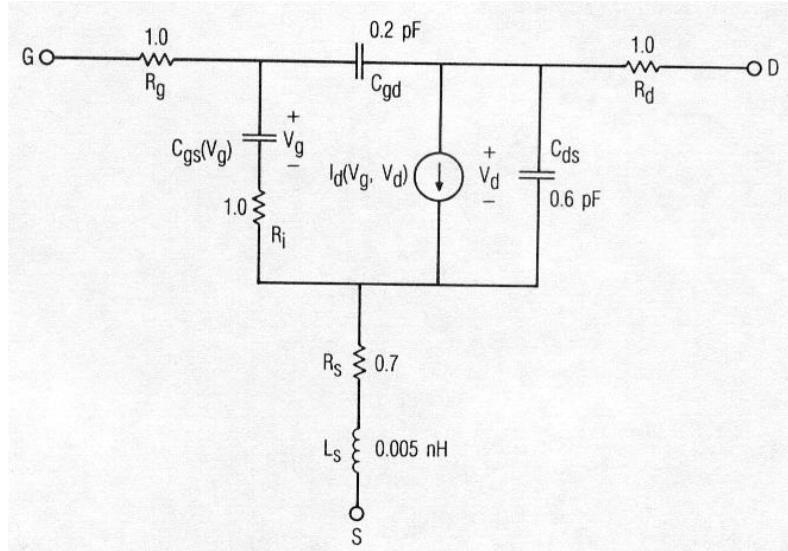


Fig. IV-44. Equivalent circuit of the 2.4mm power MESFET.

The I-V characteristics (Fig. IV-45) can be modeled using the drain-current Curtice cubic model:

$$I_d = \left(a_0 + a_1 V + a_2 V^2 + a_3 V^3 \right) \tanh(\alpha V_{ds}) \quad (\text{IV-153})$$

with:

$$V = V_{gs} \left(1 + \beta (V_{dso} - V_{ds}) \right)$$

and

$$\begin{aligned} a_0 &= 0.5304 & a_1 &= 0.2595 & a_2 &= -0.0542 \\ a_3 &= -0.0305 & \alpha &= 1.0 \end{aligned}$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

One other useful parameter is

$$V_t = -2.5V$$

The gate-source capacitance C_{gs} is modeled as a uniform doped Schottky barrier with $\phi = 0.7V$ and $C_{gso} = 4.0\text{pF}$

$$C_{gs} = C_{gso} \left\{ 1 - \frac{V_j(t)}{\Phi} \right\}^{-1/2} = \frac{C_{gso}}{\sqrt{1 - \frac{V_j(t)}{\Phi}}} \quad (\text{IV-154})$$

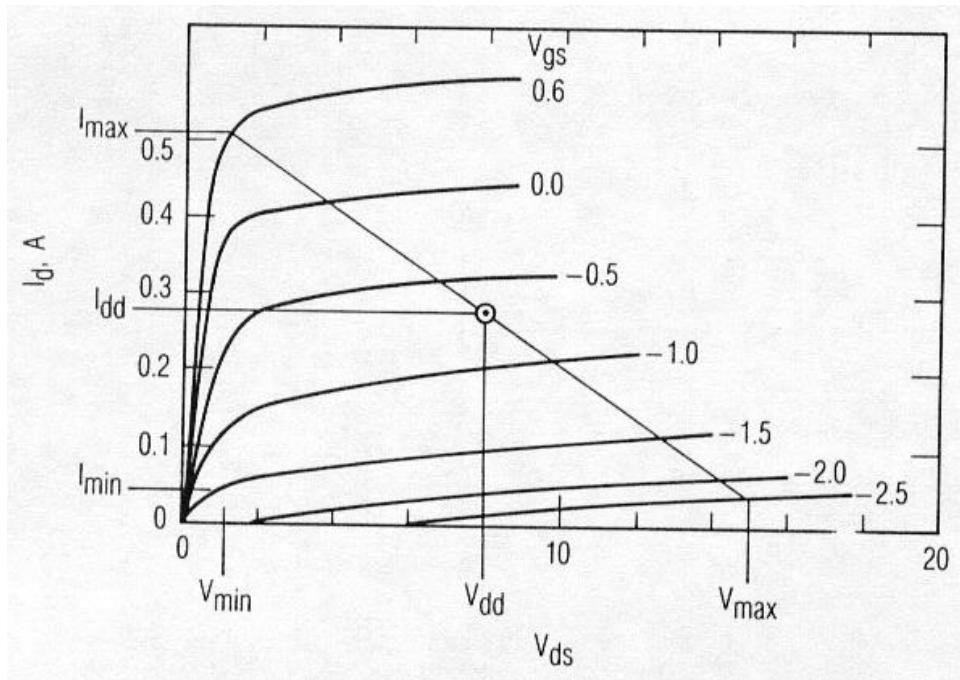


Fig. IV-45. I-V characteristics.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

I – Example 1 - Class-A amplifier

We determine first the load impedance. Using Fig. IV-45, the load line is constructed with the constraint $V_{dd} = 8.0\text{V}$ (and $I_{dd} = 270 \text{ mA}$). Thus,

$$V_{max} = 14.7 \text{ V} \quad V_{min} = 1.3 \text{ V} \quad I_{max} = 500 \text{ mA} \quad I_{min} = 40 \text{ mA}$$

we have

$$G_L = \frac{V_{max} - V_{min}}{I_{max} - I_{min}} = \frac{0.5 - 0.04}{14.7 - 1.3} = 0.034 \text{ S} \quad (\text{IV-155})$$

$$P_L = \frac{1}{2} \left[\frac{1}{2} (V_{max} - V_{min}) \right] \left[\frac{1}{2} (I_{max} - I_{min}) \right] = 0.771 \text{ W or } 28.9 \text{ dBm} \quad (\text{IV-156})$$

It is more reasonable to anticipate a more realistic 1-dB compression point of 28 dBm. Thus, from equation (IV-149) we obtain

$$Y_L = G_L - j\omega C_{ds} = (0.034 - j0.05) \text{ S} \quad (\text{IV-157})$$

$$\rightarrow Z_L = (9.3 + j13.5) \Omega \quad (\text{IV-158})$$

For the input impedance, we must find C_{gs} at the bias point ($V_g = -0.7\text{V}$):

$$C_{gs} = \frac{C_{gso}}{\sqrt{1 - \frac{V_j(t)}{\Phi}}} = \frac{4.0 \text{ pF}}{\sqrt{1 - \frac{-0.7}{0.7}}} = 2.8 \text{ pF} \quad (\text{IV-159})$$

The transconductance is found by differentiating I_d or from the I-V curves. Its value is 315 mS. Using these values in equation (IV-152) gives

$$Z_{in} = Z_s^* = (2.3 - j4.3)\Omega \quad (\text{IV-160})$$

and a gain $G_t = 10$ dB. Assuming that the matching networks present the optimum terminations (i.e., short-circuit at the gate and drain at all harmonics), an harmonic-balance analysis shows that the approximate design is very good: the output power at 1-dB compression point is 27.8 dBm, with an input impedance of $(2.5 - j4.5)\Omega$ and an output impedance of $(9.0 + j12.0)\Omega$. The dc current at saturation is 290 mA, and the dc power is 2.3 W. The output power and power added efficiency η_a , are shown in Fig. IV-46.

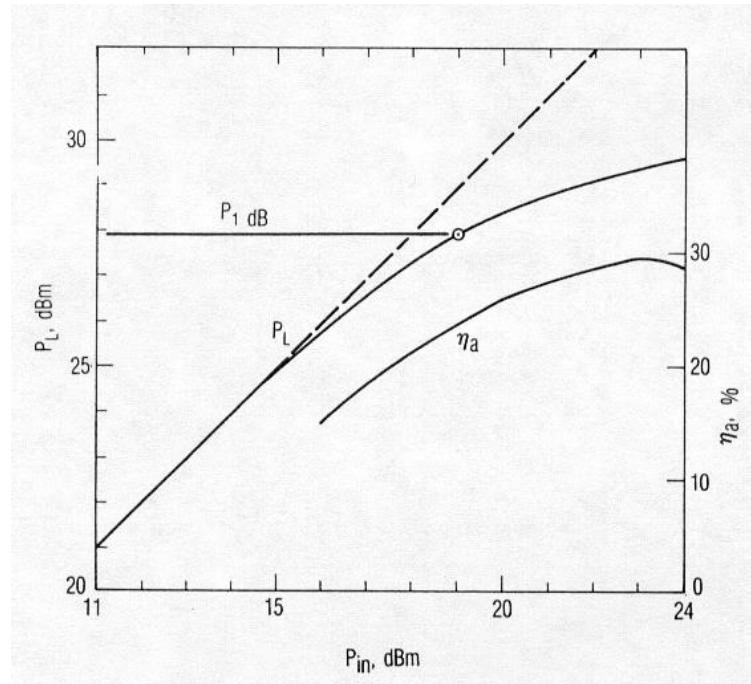


Fig. IV-46. Output power and power added efficiency, η_a , of the Class-A amplifier.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

II – Example 2 - Class-B amplifier

We now design a Class-B amplifier by modifying the previous design. The first modification is to reexamine the bias voltages: $V_{dd} = 8$ V and adjust the gate bias V_{gg} to -2.0 V, so that the quiescent current is $0.1 * I_{dss}$ or 40 mA.

Assuming the harmonic source and load impedances are short circuits, we can use the same load impedance for Class-B as for Class-A. The gate-source capacitance is lower because the voltage across C_{gs} is more negative. Consequently the imaginary part of the input impedance should be greater. In fact, using equation (IV-152) we obtain an initial estimate of

$$Z_s^* = Z_{in} = (2.7 - j 7.6) \Omega \quad (\text{IV-161})$$

Although the drain voltage and current ranges are slightly different ($I_{min} = 0$ in Class-B), we find, via the harmonic-balance analysis, that the Class-B load impedance is practically equal to the Class-A load and the input impedance is $(2.7 - j 5.2) \Omega$. P_L and η_a , as function of P_{in} , are shown in Fig. IV-47.

We see that the saturated output power is equal to 29.3 dBm, nearly identical to that of the Class-A amplifier. The small-signal gain of an ideal Class-B amplifier is undefined because the FET is turned off when quiescent. When the FET has a small quiescent drain current, the small-signal gain can be defined; however, that gain may be meaningless parameter because the transconductance at the low drain current may be very low and not accurately determined.

We can define the small-signal gain of the Class-B amplifier as the gain derived from the straight part of P_{in}/P_L characteristic, which, from Fig. IV-47 is 5.2 dB.

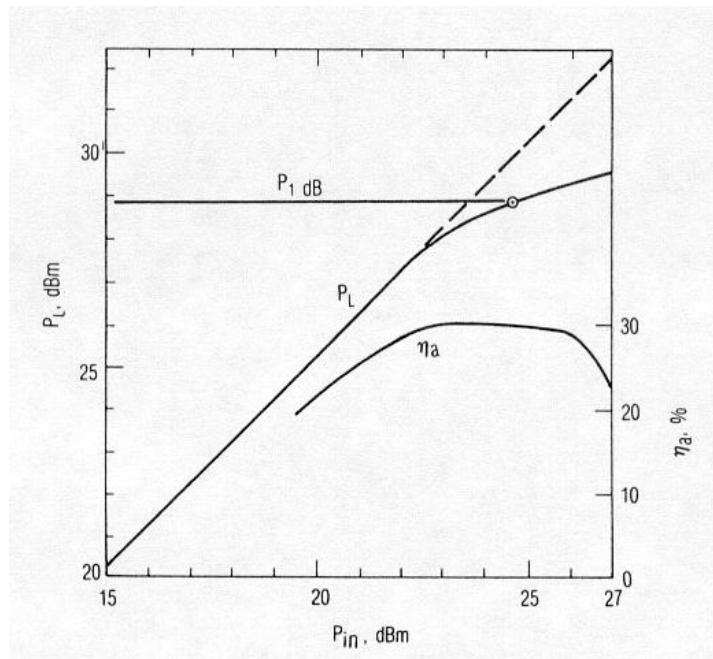


Fig. IV-47. Output power and power added efficiency, η_a , of the Class-B amplifier.

III – Effect of nonzero harmonic terminations

We assumed that all harmonics are short-circuited. The effect of deviations from these ideal conditions is that the Class-B amplifier is more sensitive to harmonic terminations than the Class-A amplifier.

All the harmonic components of the drain current circulate in the terminating impedance, $Z_L(\omega)$. So, if $Z_L(\omega)$ has a nonzero real part at any harmonic frequency, there can be output power at that harmonic. However, even if $Z_L(\omega)$ is purely reactive at some harmonic, V_d has a voltage component at that harmonic; this component can cause the fundamental output power to be reduced because output power is limited by the need to keep $\{ 0 < V_d(t) < V_{max} \}$. The addition of a harmonic

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

component to $V_d(t)$ may force the fundamental component of $V_d(t)$ to be reduced, so that $V_d(t)$ remains within the prescribed limits. In practice, harmonic-balance analysis is the only accurate means to assess these effects.

However, we can understand intuitively that a relatively large value of C_{ds} (here 0.36 pF) will effectively short-circuit the channel at harmonics of the excitation frequency, making the amplifier relatively insensitive to harmonic termination effects. Therefore, unless the second-harmonic termination resonates with C_{ds} , it has a minimal effect on fundamental output. However, if the load does resonate with C_{ds} , the results can have a serious effect on the performance of the amplifier. Fig. IV-48 shows the effect of a purely reactive second-harmonic termination on the output power of the Class-B amplifier.

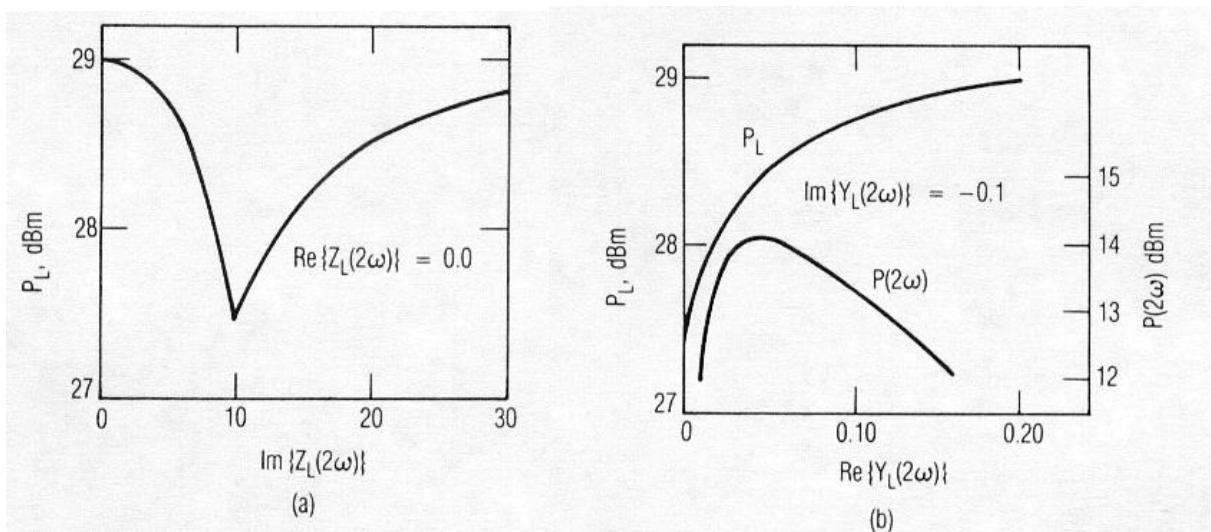


Fig. IV-48. Class-B amplifier: Effect of the second-harmonic load termination on the fundamental output power (a) effect of a purely reactive second-harmonic termination, (b) fundamental and second-harmonic output powers as function of $\text{Re}(Y_L(2\omega))$.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Here the output power decreases by more than 1.5 dB. Fig. IV-48-b shows the effect of adding a conductance in parallel with the susceptance (-0.1 S) that resonates C_{ds} . Although the fundamental output power rises monotonically with $\text{Re}(Y_L(2\omega))$, the harmonic output peaks near $\text{Re}(Y_L(2\omega)) = 0.05$. The worst case harmonic level is 14 dB below the fundamental output compared to 7 dB below the output in an ideal Class-B amplifier.

We must remember that this relative insensitivity to harmonic terminations resulted from the large value of C_{ds} in this particular device. Other MESFETs may have much lower values of C_{ds} and in these, the harmonic terminations may have much more significant effects.

N – HARMONIC-BALANCE ANALYSIS OF FET POWER AMPLIFIERS

We discussed earlier about the advantages of applying the harmonic-balance analysis to power amplifiers. In this section, we will examine the details involved in applying such technique to power amplifier circuits. We have considered I_d and C_{gs} as nonlinear elements. C_{gd} can be also treated as a nonlinear element if the user can plug an appropriate model to the circuit simulator or use an internal model available in the software library.

Fig. IV-49-a shows the equivalent circuit of the MESFET used for the harmonic balance analysis. It includes source and load terminations, bias sources, and the excitation source. Fig. IV-49-b shows the circuit divided into linear and nonlinear sub-circuits.

Ports 1 through 3 are the ports to which the nonlinear elements are connected (C_{gs} , C_{gd} , and I_d respectively). Ports 4 and 5 are the ports to which the excitation and bias voltages are applied. We note that $V_1 = V_g$ and $V_2 = V_d$.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

I - Linear sub-circuit

The Y parameters of the linear sub-circuit are found in the usual manner. Since the linear circuit has a reciprocal property, $Y_{i,j} = Y_{j,i}$. Therefore, the Y parameters are

$$Y_{1,1}(\omega) = -Y_{1,2}(\omega) = -Y_{2,1}(\omega) = -Y_{1,3}(\omega) = -Y_{3,1}(\omega) = \frac{1}{R_i} \quad (\text{IV-162})$$

$$Y_{1,4}(\omega) = Y_{4,1}(\omega) = Y_{1,5}(\omega) = Y_{5,1}(\omega) = 0 \quad (\text{IV-163})$$

and

$$Y_{2,2}(\omega) = \frac{1}{R_i} + \left[\left(\frac{1}{Z_d} + \frac{1}{Z_g} \right)^{-1} + R_s + j\omega L \right]^{-1} + j\omega C_{ds} \quad (\text{IV-164})$$

$$Y_{4,4}(\omega) = \left[\left(\frac{1}{Z_d} + \frac{1}{R_s + L_s} \right)^{-1} + Z_g \right]^{-1} \quad (\text{IV-165})$$

$$Y_{5,5}(\omega) = \left[\left(\frac{1}{Z_g} + \frac{1}{R_s + L_s} \right)^{-1} + Z_d \right]^{-1} \quad (\text{IV-166})$$

$$Y_{2,4}(\omega) = Y_{4,2}(\omega) = -Y_{4,4}(\omega) \frac{Z_d}{R_s + j\omega L_s + Z_d} \quad (\text{IV-167})$$

$$Y_{2,5}(\omega) = Y_{5,2}(\omega) = -Y_{5,5}(\omega) \frac{Z_g}{R_s + j\omega L_s + Z_g} \quad (\text{IV-168})$$

$$Y_{2,3}(\omega) = Y_{3,2}(\omega) = -Y_{2,4}(\omega) + Y_{1,1}(\omega) \quad (\text{IV-169})$$

$$Y_{3,3}(\omega) = Y_{4,4}(\omega) + Y_{1,1}(\omega) \quad (\text{IV-170})$$

$$Y_{3,4}(\omega) = Y_{4,3}(\omega) = -Y_{4,4}(\omega) \quad (\text{IV-171})$$

$$Y_{3,5}(\omega) = Y_{5,3}(\omega) = -Y_{5,5}(\omega) \frac{R_s + j\omega L_s}{R_s + j\omega L_s + Z_g} \quad (\text{IV-172})$$

$$Y_{4,5}(\omega) = Y_{5,4}(\omega) = -Y_{3,5}(\omega) \quad (\text{IV-173})$$

where

$$Z_d = Z_L(\omega) + R_d \quad (\text{IV-174})$$

$$Z_g = Z_s(\omega) + R_g \quad (\text{IV-175})$$

II - Nonlinear sub-circuit

In order to use the Newton's method, we need to formulate the Jacobian for the nonlinear elements,

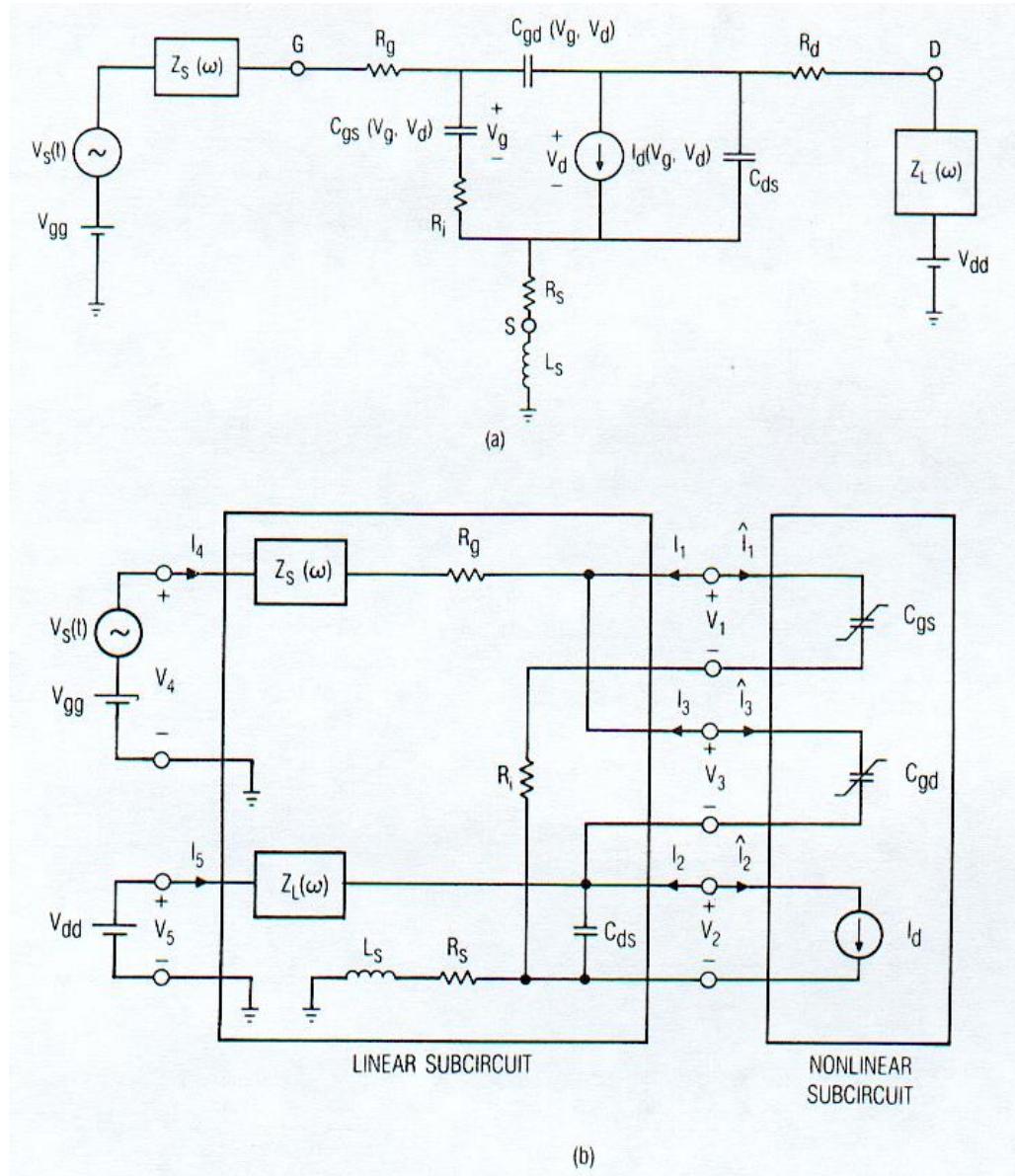


Fig. IV-49. (a) Equivalent circuit of the FET power amplifier; (b) the same circuit separated into linear and nonlinear sub-circuits.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

$$[\mathbf{J}_B] = \begin{bmatrix} \frac{\partial B_{1,0}}{\partial V_{1,0}} & \frac{\partial B_{1,0}}{\partial V_{1,1}} & \frac{\partial B_{1,0}}{\partial V_{1,2}} & \dots & \frac{\partial B_{1,0}}{\partial V_{1,M}} & \frac{\partial B_{1,0}}{\partial V_{2,0}} & \frac{\partial B_{1,0}}{\partial V_{2,1}} & \frac{\partial B_{1,0}}{\partial V_{2,2}} & \dots & \frac{\partial B_{1,0}}{\partial V_{N,M}} \\ \frac{\partial B_{1,1}}{\partial V_{1,0}} & \frac{\partial B_{1,1}}{\partial V_{1,1}} & \frac{\partial B_{1,1}}{\partial V_{1,2}} & \dots & \frac{\partial B_{1,1}}{\partial V_{1,M}} & \frac{\partial B_{1,1}}{\partial V_{2,0}} & \frac{\partial B_{1,1}}{\partial V_{2,1}} & \frac{\partial B_{1,1}}{\partial V_{2,2}} & \dots & \frac{\partial B_{1,1}}{\partial V_{N,M}} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{\partial B_{1,M}}{\partial V_{1,0}} & \frac{\partial B_{1,M}}{\partial V_{1,1}} & \frac{\partial B_{1,M}}{\partial V_{1,2}} & \dots & \frac{\partial B_{1,M}}{\partial V_{1,M}} & \frac{\partial B_{1,M}}{\partial V_{2,0}} & \frac{\partial B_{1,M}}{\partial V_{2,1}} & \frac{\partial B_{1,M}}{\partial V_{2,2}} & \dots & \frac{\partial B_{1,M}}{\partial V_{N,M}} \\ \frac{\partial B_{2,0}}{\partial V_{1,0}} & \frac{\partial B_{2,0}}{\partial V_{1,1}} & \frac{\partial B_{2,0}}{\partial V_{1,2}} & \dots & \frac{\partial B_{2,0}}{\partial V_{1,M}} & \frac{\partial B_{2,0}}{\partial V_{2,0}} & \frac{\partial B_{2,0}}{\partial V_{2,1}} & \frac{\partial B_{2,0}}{\partial V_{2,2}} & \dots & \frac{\partial B_{2,0}}{\partial V_{N,M}} \\ \frac{\partial B_{2,1}}{\partial V_{1,0}} & \frac{\partial B_{2,1}}{\partial V_{1,1}} & \frac{\partial B_{2,1}}{\partial V_{1,2}} & \dots & \frac{\partial B_{2,1}}{\partial V_{1,M}} & \frac{\partial B_{2,1}}{\partial V_{2,0}} & \frac{\partial B_{2,1}}{\partial V_{2,1}} & \frac{\partial B_{2,1}}{\partial V_{2,2}} & \dots & \frac{\partial B_{2,1}}{\partial V_{N,M}} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{\partial B_{2,M}}{\partial V_{1,0}} & \frac{\partial B_{2,M}}{\partial V_{1,1}} & \frac{\partial B_{2,M}}{\partial V_{1,2}} & \dots & \frac{\partial B_{2,M}}{\partial V_{1,M}} & \frac{\partial B_{2,M}}{\partial V_{2,0}} & \frac{\partial B_{2,M}}{\partial V_{2,1}} & \frac{\partial B_{2,M}}{\partial V_{2,2}} & \dots & \frac{\partial B_{2,M}}{\partial V_{N,M}} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{\partial B_{N,M}}{\partial V_{1,0}} & \frac{\partial B_{N,M}}{\partial V_{1,1}} & \frac{\partial B_{N,M}}{\partial V_{1,2}} & \dots & \frac{\partial B_{N,M}}{\partial V_{1,M}} & \frac{\partial B_{N,M}}{\partial V_{2,0}} & \frac{\partial B_{N,M}}{\partial V_{2,1}} & \frac{\partial B_{N,M}}{\partial V_{2,2}} & \dots & \frac{\partial B_{N,M}}{\partial V_{N,M}} \end{bmatrix}$$

or

$$\mathbf{J}_B = \mathbf{Y}_{3x3} + \frac{\partial \mathbf{I}}{\partial \mathbf{V}} + j \boldsymbol{\Omega} \frac{\partial \mathbf{Q}}{\partial \mathbf{V}} \quad (\text{IV-176})$$

with

$$\frac{\partial B_{n,k}}{\partial V_{m,l}} = Y_{m,n} (k = l) + \frac{\partial I_{Gn,k}}{\partial V_{m,l}} + j k \omega_p \frac{\partial Q_{n,k}}{\partial V_{m,l}} \quad (\text{IV-177})$$

In this example, the Jacobian has the following form

$$\mathbf{J}_B = \begin{bmatrix} \mathbf{J}_{1,1} & \mathbf{J}_{1,2} & \mathbf{J}_{1,3} \\ \mathbf{J}_{2,1} & \mathbf{J}_{2,2} & \mathbf{J}_{2,3} \\ \mathbf{J}_{3,1} & \mathbf{J}_{3,2} & \mathbf{J}_{3,3} \end{bmatrix} \quad (\text{IV-178})$$

where

$$\mathbf{J}_{m,n} = \mathbf{Y}_{m,n} + \mathbf{G}_{m,n} + j\Omega \mathbf{C}_{m,n} \quad (\text{IV-179})$$

The term $Y_{m,n}$ ($k = l$) is $Y_{m,n}$ ($k\omega_p$) when $\{ k = l \}$ and zero when $\{ k \neq l \}$. The admittance matrix $\mathbf{Y}_{m,n}$ (a 3x3 matrix) is defined by the Harmonic-Balance equations developed in chapter III, i.e.,

$$\begin{bmatrix} [\mathbf{I}_1] \\ \vdots \\ [\mathbf{I}_N] \end{bmatrix} = \begin{bmatrix} [Y_{1,N+1}] & [Y_{1,N+2}] \\ \vdots & \vdots \\ [Y_{N,N+1}] & [Y_{N,N+2}] \end{bmatrix} \begin{bmatrix} [\mathbf{V}_{N+1}] \\ [\mathbf{V}_{N+2}] \end{bmatrix} + \begin{bmatrix} [Y_{1,1}] & [Y_{1,N}] \\ \vdots & \vdots \\ [Y_{N,1}] & [Y_{N,N}] \end{bmatrix} \begin{bmatrix} [\mathbf{V}_1] \\ \vdots \\ [\mathbf{V}_N] \end{bmatrix} \quad (\text{IV-180})$$

$$[\mathbf{I}] = [\mathbf{I}_s] + [\mathbf{Y}_{N \times N}] [\mathbf{V}] \quad (\text{IV-181})$$

This matrix has the form of

$$[\mathbf{Y}_{m,n}] = \begin{bmatrix} Y_{m,n}(0) & 0 & 0 & \cdots & 0 \\ 0 & Y_{m,n}(\omega_p) & 0 & \cdots & 0 \\ 0 & 0 & Y_{m,n}(2\omega_p) & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & Y_{m,n}(M\omega_p) \end{bmatrix} \quad (\text{IV-182})$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

The terms $\mathbf{G}_{m,n}$ and $\mathbf{C}_{m,n}$ are the Fourier coefficients of the derivative waveforms. These matrices have the same form as the conversion matrices used in the small-signal analysis (Chapter III)

$$\frac{\partial I_{Gn,k}}{\partial V_{m,l}} = \frac{1}{T} \int_{-\frac{1}{2T}}^{\frac{1}{2T}} \frac{\partial i_{gn}(t)}{\partial V_m(t)} e^{-j(k-l)\omega_p t} dt \quad (\text{IV-183})$$

and

$$\frac{\partial Q_{n,k}}{\partial V_{m,l}} = \frac{1}{T} \int_{-\frac{1}{2T}}^{\frac{1}{2T}} \frac{\partial q_n(t)}{\partial V_m(t)} e^{-j(k-l)\omega_p t} dt \quad (\text{IV-184})$$

where n and m subscripts refer to the port numbers and k and l subscripts indicate the harmonic number. T is the fundamental excitation period.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

The Ω matrix is defined as

$$[\Omega] = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \omega_p & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 2\omega_p & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & \cdots & M\omega_p & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \omega_p & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 2\omega_p & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & M\omega_p \end{bmatrix}$$

O – OTHER AMPLIFIERS

In section F, we divided amplifiers between reactively matched, lossy matched, feedback, distributed and balanced amplifiers. We already developed in details the first group of amplifiers. We will examine briefly the characteristics of the other kinds of amplifiers used in the RF/microwave area.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

I – Lossy matched amplifiers

Because the gain of a FET decreases roughly at 6dB per octave, an amplifier gain characteristic, which is approximately independent of frequency, requires power to be reflected at the lower frequencies (Fig. IV-50).

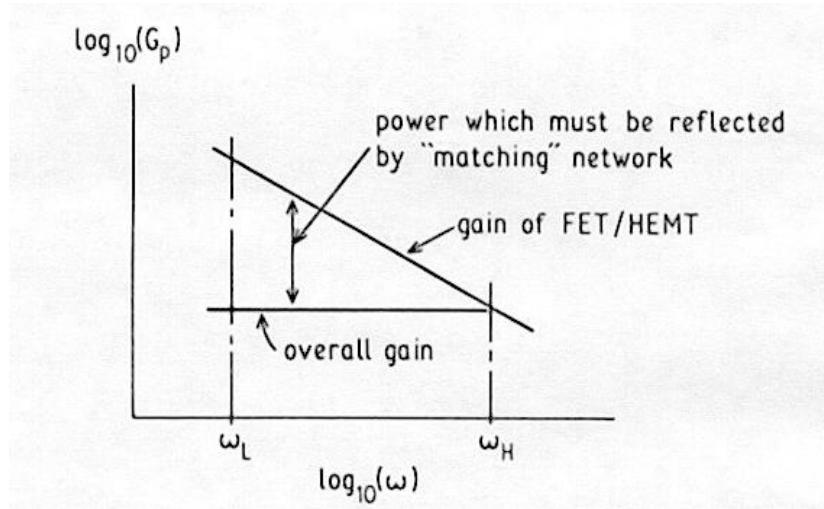


Fig. IV-50. Mechanism of achieving level gain using a lossless equalizer or matching network.

In other words, the input or inter-stage match must become worse at low frequencies. So instead of using reactive matching networks (Fig. IV-51), one means of improving the match at low frequencies is to provide resistive elements, which become more tightly coupled to the main signal path through the amplifier at low frequencies (Fig. IV-52). M_i are lossless impedance matching networks and F_i are low-pass lossless impedance transformers. At the higher range of the specified frequency band, the transformers transform the value of R_i into high impedances that do not load the signal path through the amplifier. In this limit the circuit behaves as a reactively matched amplifier.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

The gain of the amplifier is then dependent upon the resistor values. Therefore, a lossy matched amplifier consists of two circuits both having the same gain: a resistively-coupled amplifier providing gain and matching at the lower frequencies in the band to be covered, and a reactively-matched amplifier providing gain and matching at the higher frequencies. The circuit is thus applicable to flat-gain, ultra-broadband amplification.

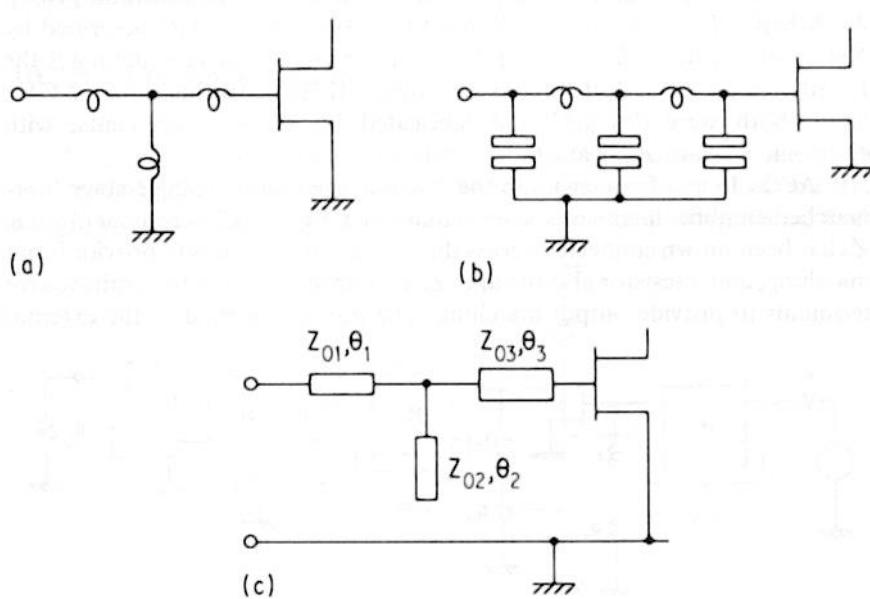


Fig. IV-51. Various reactive matching network configuration.

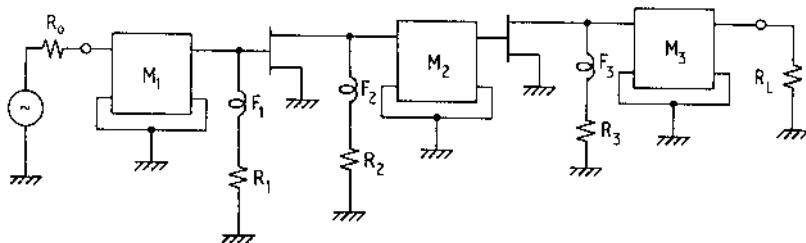


Fig. IV-52. Configuration of lossy matched amplifier.

II – Feedback amplifiers

Feedback amplifiers (Fig. IV-53) could be designed through a first-trial method, finding initial circuit values that can then be iterated by computer methods to achieve the desired performance. The connection has the following attributes:

- (a) It offers gain equalization over a fairly wide bandwidth,
- (b) It can improve the $|S_{21}|$ at the highest frequencies, compared with the bare FET,
- (c) The matches can be improved concurrently with (a) and (b), thus simplifying the task of matching network design,
- (d) It reduces the sensitivity of the overall circuit to device variations

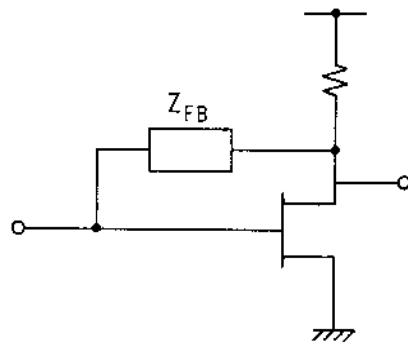


Fig. IV-53. Shunt feedback amplifier.

III – Distributed amplifiers

With a conventional amplifier, any attempt to increase the gain by increasing the FET transconductance usually results in an increase in input capacitance as well. The distributed amplifier, also called the travelling wave amplifier, offers a means of combining the transconductance of several FETs without combining their input capacitances (Fig. IV-54).

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

IV – Balanced amplifiers

Usually, the input reflection coefficient for a wideband reactively matched amplifier can easily be 0.8 or greater, i.e. a return loss as high as -2dB . To prevent the reflected energy giving rise to standing waves on any transmission line between the source and the amplifier, and to prevent the energy being dissipated in the source, a balanced amplifier configuration can be employed (Fig. IV-55). Two quadrature couplers are used: one at the input, to split the signal from the generator (source) into two equal parts to be fed to two amplifiers operating in parallel, the other to combine the outputs of the amplifiers into a single load. If the amplifier is truly balanced, $S_{11}^{(a)} = S_{11}^{(b)}$, and $S_{22}^{(a)} = S_{22}^{(b)}$. Regardless of the values of these reflection coefficients, the reflection coefficients at the input and the output of the balanced amplifier are zero. Furthermore, the following magnitude is unchanged.

$$S_{21}^T = \frac{j(S_{21}^{(a)} + S_{21}^{(b)})}{2} \quad (\text{IV-185})$$

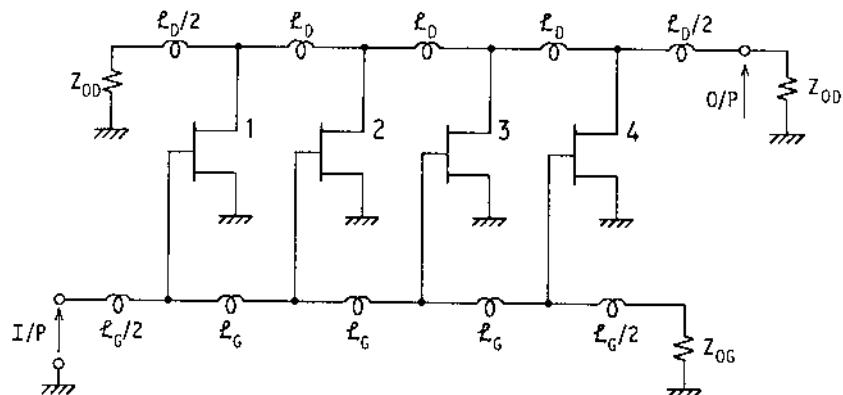


Fig. IV-54. Distributed amplifier configuration.

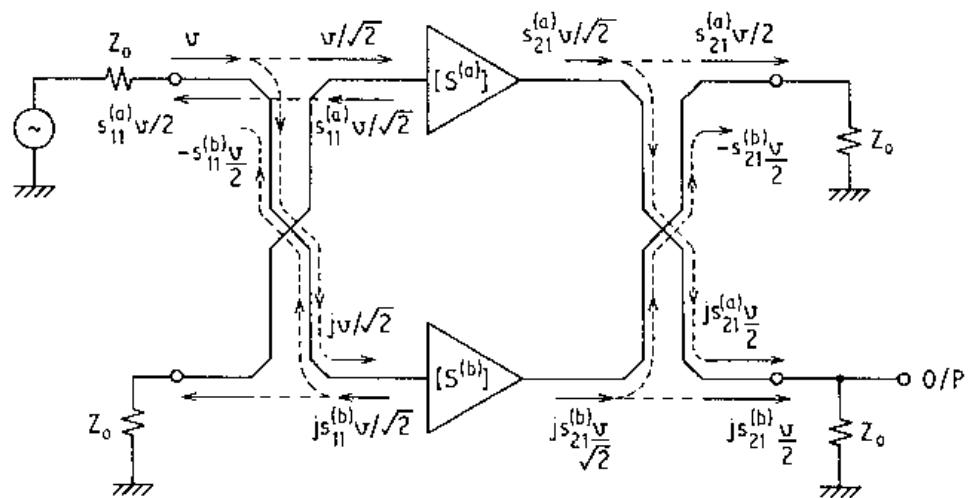


Fig. IV-55. Schematic of a balanced amplifier showing signal flow paths.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

APPENDIX IV-1: LOWAN'S SERIES

The following summary is from:

Heinz Unger, "Lagrange-Hermitesche Interpolation im Komplexen," *Zeitschrift für Angewandte Mathematik und Physik (ZAMP)*, Vol. 3, No. 1, 51-65, Jan. 1952.

Link:

<https://commerce.metapress.com/content/l717864036435j28/resource-secured/?target=fulltext.pdf&sid=vwwd2r45myakbw45ikrych45&sh=www.springerlink.com>

Summary Increased attention has been recently devoted to the interpolation of an analytic function $g(z)$ in the complex plane. Only if the increment in a given table is sufficiently small, $g(z)$ may be approximated by a quadratic or cubic polynomial. To get an interpolation polynomial $P(z)$ of higher degree we employ the extension of the Lagrange formula given by Hermite, based on the values of the function and some of its derivatives:

$$P(z) = \sum_{\mu=1}^n \sum_{p=0}^q L_{\mu,p}(z) g^{(p)}(z_\mu)$$

Choosing the points z_μ in the corners of a regular polygon the relation for the coefficients $L_{\mu,p}$ may be found as follows:

$$L_{\mu,p}(z) = e^{i(2\pi/n)p(\mu-1)} L_{1,p}[z^{(\mu)}],$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

where

$$z^{(\mu)} = z e^{i(2\pi/n)(1-\mu)}$$

It is convenient and advantageous to choose a square grid. Employing e.g. an interpolation polynomial of the degree 11, it is unnecessary to tabulate the twelve coefficients $L_{\mu,p}(z)$. You need only three, namely $L_{1,0}$, $L_{1,1}$ and $L_{1,2}$. The corner values together with the first and second derivatives of the special function $g(z)$ must be known. It should be noted, that in many cases the derivatives could be easily computed (i.e. by the differential equation).

Read also:

J.C.P. Miller, “Zeros of Legendre polynomials of orders 2-64 and weight coefficients of Gauss quadrature formulae by H. J. Gawlik”, *Mathematics of Computation*, Vol. 14, No. 69, Jan. 1960.

APPENDIX IV-2: MANLEY-ROWE RELATIONS

The two reference papers are

- [1] J.M. Manley, H.E. Rowe, "Some general properties of nonlinear elements. Part I : general energy relations," *Proc. IRE*, Vol 44, 904-913, July 1956.
- [2] H.E. Rowe, "Some general properties of nonlinear elements. Part II : small signal theory," *Proc. IRE*, Vol 46, May 1958.

Other useful references involving the application of Manley-Rowe relations in physics are

- [3] A.V. Petukhov, V.L. Brudny, W.L. Mochan, J.A. Maytorena, B.S. Mendoza, T. Rasing, "Energy conservation and the Manley-Rowe relations in surface nonlinear-optical spectroscopy," *Physical review letters*, vol. 81, pp. 566-569, 1998.
- [4] B. Saikia, M. Nambu, S. Bujarbarua, "Conservation relations and the generalized Manley-Rowe relations for the plasma-maser in magnetized plasma," *Physics of Plasmas*, Vol. 2, pp.1746-1754, May 1995.
- [5] A.K. Popov, V.M. Shalaev, "Negative-index metamaterials: second-harmonic generation, Manley-Rowe relations and parametric amplification," *Applied Physics B*, Vol. 84, , pp. 131-137, July 2006.
- [6] T. Tanaka, "Manley-Rowe relations in energy transports among plasma waves,", *Journal of the Physical Society of Japan*, Vol.60, pp. 2645-2654, 1991.
- [7] V. Belevitch, "On frequency-power relations," *Int. J. Circuit Theory and Applications*, Vol. 4, pp. 93-98, 2006.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

An interesting application in optics from

<http://electron9.phys.utk.edu/optics507/modules/m9/second.htm>

Frequency mixing represents one of the most general but important processes in nonlinear optics. Two or more waves interact in a nonlinear medium to produce an output at various sum or difference frequencies. Polarization components for an incident field containing two frequency components ω_1 and ω_2 , produced by the $\chi^{(2)}$ tensor are

$$P_i^{(2)}(\omega')/\epsilon_0 = \chi_{ijk}^{(2)}(E_j'(\omega_n)E_k'(\omega_m)\delta(\omega_n + \omega_m - \omega')).$$

where ω_n and ω_m can take on the values $\pm\omega_1$ and $\pm\omega_2$ and $E'(\omega_n)$ is the field strength associated with ω_n .

When $\omega_n = \omega_m = \omega$ and $\omega' = 2\omega$, then $P^{(2)}(\omega')$ is the source of second harmonic generation (SHG). Franken and his co-workers first observed SHG in 1961. They focused the 694.3 nm beam from a ruby laser onto a quartz crystal and obtained a very low-intensity output beam at a wavelength of 347.15 nm. This wavelength lies in the ultraviolet region of the spectrum. The beam has exactly half the wavelength and twice the frequency of the incident laser light.

When $\omega_n = \omega_m = \omega$ and $\omega' = 0$, then $P^{(2)}(\omega')$ is the source of optical rectification. When an intense, plane-polarized laser beam traverses an appropriate crystal, one of the effects produced is a constant electric polarization of the medium. A voltage difference proportional to the intensity of the beam will appear across the medium (Fig. AIV2-1).

The lack of power sources in the range of 100GHz-10THz is often called the ``tera hertz gap''. Optical rectification can be used to generate subpicosecond, terahertz (THz) bandwidth

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

electrical signals, which are important for a variety of applications, including material characterization and the development of high-speed optoelectronic devices.

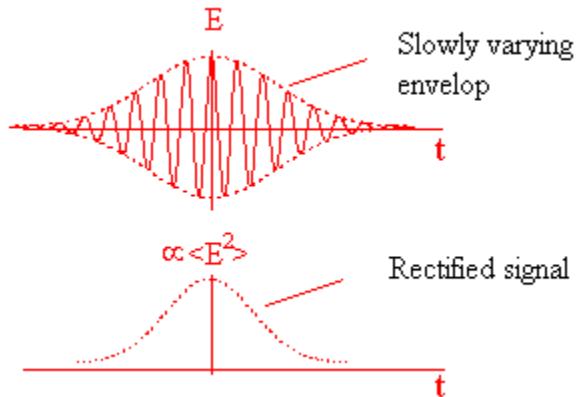


Figure AIV2-1: Rectified signal.

SHG and optical rectification are special cases of sum and difference frequency mixing processes. In general, the $\chi^{(2)}$ terms provide a coupling between three electromagnetic waves [$\delta(\omega_n + \omega_m - \omega')$]. Each wave has its own frequency ω_i , wave vector \mathbf{k}_i , state of polarization \mathbf{e}_i , as well as a complex amplitude $E_i = A_i \exp(i\phi_i)$. In the same manner the $\chi^{(3)}$ terms provide a coupling between four electromagnetic waves [$\delta(\omega_n + \omega_m + \omega_o - \omega')$]. This parametric coupling of light waves in a nonabsorbing medium may be considered as the scattering of photons between eigenmodes or waves of the electric field by the material nonlinearity.

In parametric, non-dissipative processes photon energy and momentum must be conserved. We need

$$\sum_i \hbar \omega_i = \sum_f \hbar \omega_f, \quad \sum_i \hbar \mathbf{k}_i = \sum_f \hbar \mathbf{k}_f,$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

where i denotes the incident and f the emerging photons. Color dispersion generally gives rise to a momentum mismatch $\Delta\mathbf{k}$. This limits the active volume of emission to a layer of thickness $\sim I\Delta\mathbf{k}l^1$ ¹. It is possible, however, to compensate the color dispersion by making use of optical birefringence in anisotropic crystals.

To understand the physical origin of the $\chi^{(2)}$ terms we use a simple model similar to the model we used for lih materials (linear, isotropic and homogeneous). In lih materials we assumed that the electrons are bound by a harmonic force to the atomic cores, and we included a damping force. The equation of motion for an electron in the presence of an electric field is given by

$$\mathbf{F}_e = -q_e \mathbf{E}_{\text{eff}} = m(\ddot{\mathbf{r}} + \gamma \dot{\mathbf{r}} + \omega_0^2 \mathbf{r})$$

If the electric field is given by $\mathbf{E}_{\text{eff}} = \mathbf{E}_0 \exp(-i\omega t)$, then $\mathbf{r} = \mathbf{r}_0 \exp(-i\omega t)$, with

$$\mathbf{r}_0 = \frac{-(q_e/m) \mathbf{E}_0}{(\omega_0^2 - \omega^2 - i\gamma\omega)}.$$

In a nonlinear material we assume that the electrons are in anharmonic potential wells and the restoring force (neglecting all directional aspects) is given by

$$\mathbf{F}_e = m(\ddot{\mathbf{r}} + \gamma \dot{\mathbf{r}} + \omega_0^2 \mathbf{r} + D \mathbf{r}^2)$$

where the nonlinear term ($D\mathbf{r}^2$) is assumed to be a small correction. If the electrons are driven by an electric field $\mathbf{E}(t) = E(t)\mathbf{i}$ containing two frequency components ω_1 and ω_2 ,

$$\mathbf{E}(t) = E_1 \exp(-i\omega_1 t) + E_2 \exp(-i\omega_2 t) + E_1^* \exp(i\omega_1 t) + E_2^* \exp(i\omega_2 t),$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

then $x(t)$ is approximately given by

$$x(t) = x(\omega_1)\exp(-i\omega_1 t) + x(\omega_2)\exp(-i\omega_2 t) + \\ x(\omega_1-\omega_2)\exp(-i(\omega_1-\omega_2)t) + x(\omega_1+\omega_2)\exp(-i(\omega_1+\omega_2)t) + \text{c.c.}$$

where

$$x(\omega_1) = -(q_e/m)E_1/(\omega_0^2 - \omega_1^2 - i\gamma\omega), \quad x(\omega_2) = -(q_e/m)E_2/(\omega_0^2 - \omega_2^2 - i\gamma\omega),$$

$$x(\omega_1+\omega_2) = -Dx(\omega_1)x(\omega_2)/(\omega_0^2 - (\omega_1+\omega_2)^2 - i\gamma\omega),$$

$$x(\omega_1-\omega_2) = -Dx(\omega_1)x^*(\omega_2)/(\omega_0^2 - (\omega_1-\omega_2)^2 - i\gamma\omega),$$

(To verify that $x(t)$ is an approximate solution for equation 1, insert $F(t) = q_e E(t)$ and $x(t)$ into equation 1 and equate coefficients of terms with the same time dependence on both sides. Neglect terms involving $Dx(\omega_1 \pm \omega_2)$ or their complex conjugates. Neglect the optical rectification produced by each of the components.) With

$$p = -q_e x \quad \text{and} \quad P = Np = \epsilon_0(\chi^{(1)}E + \chi^{(2)}E^2)$$

we have

$$\chi^{(2)}(\omega'=\omega_1+\omega_2) = [(ND/\epsilon_0)(q_e^3/m^2)]/[(\omega_0^2 - \omega_1^2 - i\gamma\omega)(\omega_0^2 - \omega_2^2 - i\gamma\omega)(\omega_0^2 - (\omega_1+\omega_2)^2 - i\gamma\omega)]$$

$$\chi^{(2)}(\omega'=\omega_1-\omega_2) = [(ND/\epsilon_0)(q_e^3/m^2)]/[(\omega_0^2 - \omega_1^2 - i\gamma\omega)(\omega_0^2 - \omega_2^2 + i\gamma\omega)(\omega_0^2 - (\omega_1-\omega_2)^2 - i\gamma\omega)]$$

This simple model verifies some fundamental restrictions on the elements of the susceptibility tensor $\chi^{(2)}$, the Manley-Rowe relations. The Manley-Rowe relations describe the balance in the photon fluxes of the beams participating in a parametric process. For example compare the processes $\omega' = \omega_1 + \omega_2$ and $\omega_2 = \omega' - \omega_1$. We have

$$\chi^{(2)}(\omega' = \omega_1 + \omega_2) = \chi^{*(2)}(\omega_2 = \omega' - \omega_1).$$

The Manley-Rowe relations relate the rates of photon production and photon destruction.

Parametric up-conversion in a crystal can be used to convert a signal from a low frequency ω_1 to a high frequency ω' by mixing it with a strong laser beam of frequency ω_2 . The signal (ω_1) and pump (ω_2) photons are annihilated while simultaneously a photon of frequency ω' is created. Conservation of energy requires $\omega' = \omega_1 + \omega_2$ and conservation of momentum requires $\mathbf{k}' = \mathbf{k}_1 + \mathbf{k}_2$. The number of output photons cannot exceed the number of input photons. Under normal dispersion n increases with frequency. Suppose the crystal is a positive uniaxial crystal. For a positive uniaxial crystal $n_o > n_e$. There probably exists a common direction for all three vectors such that $\mathbf{k}' = \mathbf{k}_1 + \mathbf{k}_2$, if we let \mathbf{k}_1 and \mathbf{k}_2 be ordinary rays and \mathbf{k}' be an extraordinary ray. Since $k = \omega n/c$, we need $\omega' n_e(\omega') = \omega_1 n_o(\omega_1) + \omega_2 n_o(\omega_2)$.

Let z be the direction of propagation of the waves, i.e. the direction of \mathbf{k}' , \mathbf{k}_1 , and \mathbf{k}_2 . Let x be the direction of polarization of the ordinary rays \mathbf{k}_1 and \mathbf{k}_2 . Since we have an anisotropic crystal, the nonlinear polarization produced by the $\chi^{(2)}$ terms will, in general, have a component along the y -direction, and therefore an electromagnetic wave at the sum frequency with the polarization of the extraordinary ray will be produced.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Parametric amplification involves an input signal at ω_1 together with an intense pump beam at ω_2 , $\omega_2 > \omega_1$. A photon of the pump beam with energy $\hbar\omega_2$ interacts with a photon of energy $\hbar\omega_1$ and splits into two photons, one with energy $\hbar(\omega_2 - \omega_1)$ and one with energy $\hbar\omega_1$. The coupling is provided by the non-linearity of the crystal. The wave at ω_1 is amplified. This is accompanied by the generation of an idler beam at $\omega' = \omega_2 - \omega_1$. The idler beam then interacts with the pump beam to produce additional amounts of signal and idler light. A photon of the pump beam with energy $\hbar\omega_2$ interacts with a photon of energy $\hbar\omega' = \hbar(\omega_2 - \omega_1)$ and splits into two photons, one with energy $\hbar(\omega_2 - \omega') = \hbar\omega_1$ and one with energy $\hbar\omega'$.

Parametric up-conversion and parametric amplification have the same input conditions. We use an input signal at ω_1 together with an intense pump beam at ω_2 , $\omega_2 > \omega_1$. The interaction that takes place is determined by the index matching condition, $\mathbf{k}' = \mathbf{k}_1 + \mathbf{k}_2$. If this condition is satisfied for $\omega' = \omega_1 + \omega_2$, then the sum frequency is generated. If the condition is satisfied for $\omega' = \omega_2 - \omega_1$, then the difference frequency is generated.

Consider the case of parametric amplification (Fig. AIV2-2), $\omega_3 = \omega_2 - \omega_1$.

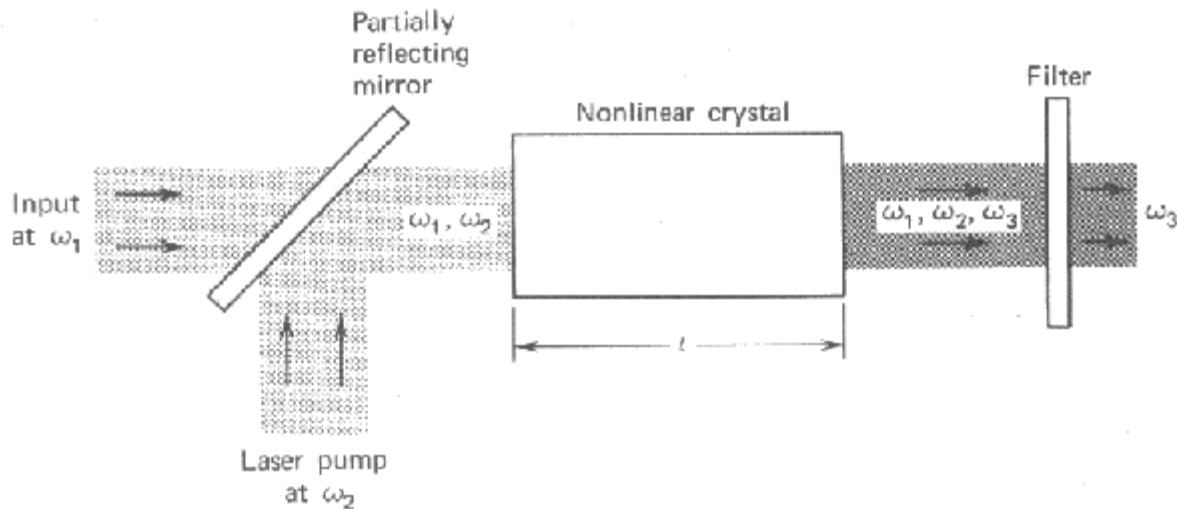


Figure AIV2-2: Parametric amplifier at ω_3 .

Beams with frequencies ω_1 , ω_2 , and ω_3 propagate in the medium and are coupled by the non-linearity of the medium. If we consider the intensity of the strong pump beam at ω_2 to be essentially unaffected, then we can use the coupling of modes formalism to describe time evolution of the intensities of the beams at ω_1 and ω_3 .

Let $x_i(t) = A_i(t) \exp(-i\omega_i t)$ describe the electric fields of beams 1 and 3 where $A_i(t)$ is assumed to be a slowly varying envelop. If we include a damping term, the first-order differential equation for $x_1(t)$ and $x_3(t)$ are

$$dx_1/dt = -i\omega_1 x_1 - x_1/\tau_1 + s_1 \quad \text{and} \quad dx_3/dt = -i\omega_3 x_3 - x_3/\tau_3 + s_3.$$

The source terms are due to the interactions with the pump field and can be evaluated from power considerations. The rate of change of the energy in mode 1 is given by

$$d|x_1|^2/dt + 2|x_1|^2/\tau_1 = x_1^* s_1 + x_1 s_1^* = \text{Power fed into the mode}$$

Note: $d|x_1|^2/dt = d(x_1 x_1^*)/dt = x_1^* dx_1/dt + x_1 dx_1^*/dt.$

This rate is proportional to $\omega_1 E(\omega_1) P^{(2)*}(\omega_1) + \omega_1 E^*(\omega_1) P^{(2)}(\omega_1)$, since the power (Power per unit volume = $\mathbf{j} \cdot \mathbf{E}$) is produced by polarization current flowing against the electric field. When the polarization varies with time, there are charges in motion, so there is a polarization current $\mathbf{j} = d\mathbf{P}/dt$. If P varies harmonically, i.e. $P \propto \exp(i\omega t)$, then $dP/dt \propto \omega P$. So

$$x_1^* s_1 + x_1 s_1^* \propto \omega_1 E(\omega_1) P^{(2)*}(\omega_1) + \omega_1 E^*(\omega_1) P^{(2)}(\omega_1).$$

But

$$P_i^{(2)}(\omega_1)/\epsilon_0 = \chi_{ijk}^{(2)}(E_j(\omega_2)E_k^*(\omega_3)\delta(\omega_2 - \omega_3 - \omega_1)) + \text{c.c.}$$

and therefore

$$x_1^* s_1 \propto \omega_1 E^*(\omega_1) P^{(2)}(\omega_1) \propto \omega_1 E^*(\omega_1) (E_j(\omega_2)E_k^*(\omega_3)f(\chi^{(2)})),$$

or

$$s_1 = A_2 A_3^* \omega_1 f(\chi^{(2)}).$$

Similarly

$$s_3 = A_2 A_3^* \omega_3 f(\chi^{(2)}).$$

We can therefore write

$$dx_1/dt = -i\omega_1 x_1 - x_1/\tau_1 + A_2 A_3^* \omega_1 f(\chi^{(2)}).$$

and

$$dx_3/dt = -i\omega_3 x_3 - x_3/\tau_3 + A_2 A_3^* \omega_3 f(\chi^{(2)})$$

$$dA_1/dt = -A_1/\tau_1 + \kappa_{13} A_3^*$$

and

$$dA_3/dt = -A_3/\tau_3 + \kappa_{31} A_1^*$$

with

$$\kappa_{13} = A_2 \omega_1 f(\chi^{(2)}) \text{ and } \kappa_{13}/\omega_1 = \kappa_{31}/\omega_3.$$

In the absence of loss ($1/\tau_1 = 1/\tau_3 = 0$) we find $(d/dt)(|A_1|^2/(\hbar\omega_1)) = (d/dt)(|A_3|^2/(\hbar\omega_3))$. $|A_i|^2/(\hbar\omega_i)$ represents the number of photons in mode i. Photons are generated in pairs at frequencies ω_3 and ω_1 .

If damping is included, then for $|\kappa_{13}|^2(\omega_3/\omega_1) > 1/(\tau_1\tau_3)$ the system has an exponentially growing solution. If the magnitude of the pump field is large enough so that the threshold $|\kappa_{13}|^2(\omega_3/\omega_1) = 1/(\tau_1\tau_2)$ is exceeded, then the system starts from unavoidable noise background, and fields build up at frequencies ω_3 and ω_1 . But as the fields grow exponentially, the depletion of the pump beam

Chapter IV: Microwave Amplifiers

can no longer be neglected, the coupling coefficients decrease in magnitude, and the system settles into a steady-state solution.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

APPENDIX IV-3: TRANSIENT IMMITANCE

Let us consider the step function shown in Fig. AIV3-1.

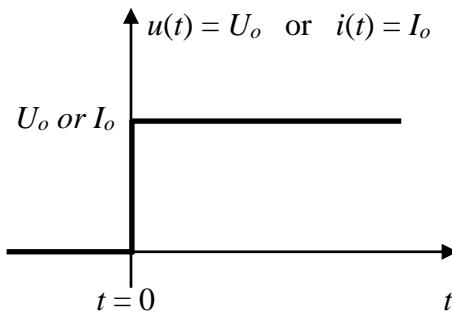


Fig. AIV3-1. Step function.

This voltage or current step function allows us to define a transient “immitance” (i.e., impedance Z , admittance Y or transmission function T). In fact, let us assume $U_o = 1$ or $I_o = 1$, therefore (Fig. AIV3-2),

$$\text{Step current : transient impedance } i(t) = I_o = 1 \quad @ \quad t > 0$$

$$u(t) = I_o Z(t) \quad \rightarrow \quad u(t) = Z(t)$$

$$\text{Step voltage : transient admittance } u(t) = U_o = 1 \quad @ \quad t > 0$$

$$i(t) = U_o Y(t) \quad \rightarrow \quad i(t) = Y(t)$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

and

Step input voltage : transient transmission $u_1(t) = U_o = 1$ @ $t > 0$

$$u_2(t) = U_o T(t) \rightarrow u_2(t) = T(t)$$

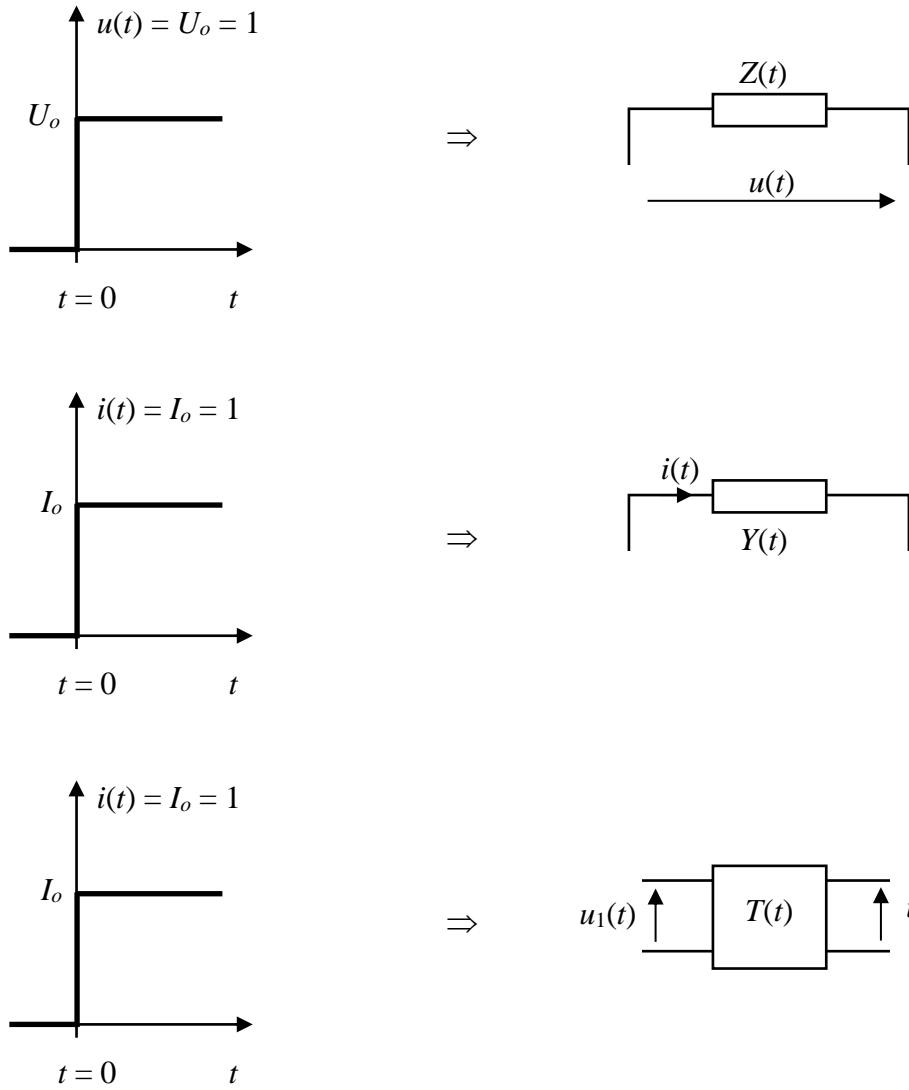


Fig. AIV3-2. Transient parameters.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Once the transient immittance known, we use the Duhamel integral to obtain

$$i(t) = \int_0^t Y(t-\tau) u(\tau) d\tau, \quad u(t) = \int_0^t Z(t-\tau) i(\tau) d\tau, \quad u_2(t) = \int_0^t T(t-\tau) u_1(\tau) d\tau$$

Duhamel integral : Background (Source: http://en.wikipedia.org/wiki/Duhamel's_integral)

The response of a linear, viscously damped single-degree of freedom (SDF) system to a time-varying mechanical excitation $p(t)$ is given by the following 2nd order ordinary differential equation

$$m \frac{d^2 x(t)}{dt^2} + c \frac{dx(t)}{dt} + k x(t) = p(t)$$

where m is the (equivalent) mass, x stands for the amplitude of vibration, t for time, c for the viscous damping coefficient, and k for the stiffness of the system or structure. If a system is initially rest at its equilibrium position, from where it is acted upon by a unit-impulse at the instance $t=0$, i.e., $p(t)$ in the equation above is a delta function $\delta(t)$,

$$x(0) = \left. \frac{dx(t)}{dt} \right|_{t=0} = 0,$$

then by solving the differential equation one can get a fundamental solution (known as a unit-impulse response function)

$$h(t) = \begin{cases} \frac{1}{m \omega_d} e^{-\zeta \omega_n t} \sin(\omega_d t) & t > 0 \\ 0 & t < 0 \end{cases}$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

where

$$\varsigma = \frac{c}{2m\omega_n}$$

is called the damping ratio of the system, ω_n is the natural circular frequency of the undamped system (when $c = 0$) and $\omega_d = \omega_n \sqrt{1 - \varsigma^2}$ is the circular frequency when damping effect is taken into account (when $c \neq 0$). If the impulse happens at $t = \tau$ instead of $t = 0$, i.e., $p(t) = \delta(t - \tau)$, the impulse response is

$$h(t - \tau) = \frac{1}{m\omega_d} e^{-\varsigma\omega_n(t-\tau)} \sin(\omega_d(t - \tau)) \quad t \geq \tau$$

Thus, regarding the arbitrarily varying excitation $p(t)$ as a superposition of a series of impulses:

$$p(t) = \sum p(\tau) * \Delta\tau * \delta(t - \tau)$$

then it is known from the linearity of system that the overall response can also be broken down into the superposition of a series of impulse-responses:

$$x(t) = \sum p(\tau) * \Delta\tau * h(t - \tau)$$

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

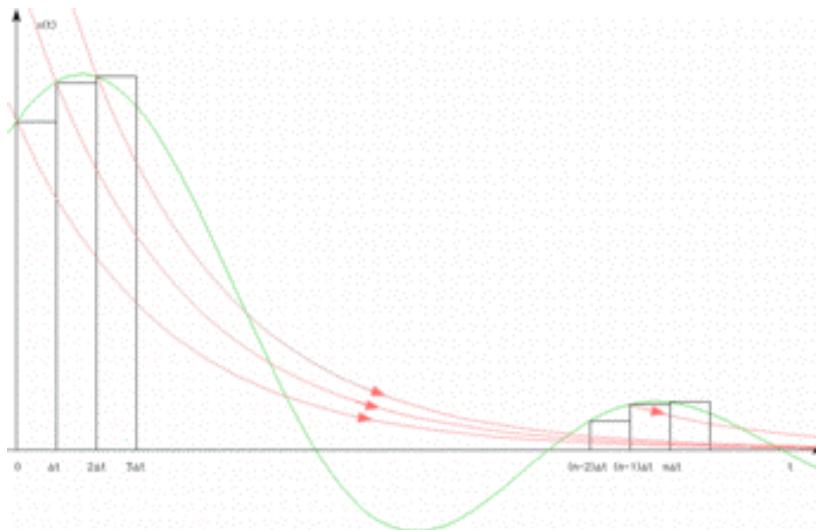
Letting $\Delta\tau \rightarrow 0$, and replacing the summation by integration, the above equation is strictly valid

$$x(t) = \int_0^t p(\tau)h(t-\tau)d\tau$$

Substituting the expression of $h(t - \tau)$ into the above equation leads to the general expression of Duhamel's integral

$$x(t) = \frac{1}{m\omega_d} \int_0^t p(\tau)e^{-\zeta\omega_n(t-\tau)} \sin(\omega_d(t-\tau))d\tau$$

General form of the Duhamel's convolution integral



The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Duhamel's convolution integral

$$y(\tau) = \int_0^{\tau} x(t)h(\tau-t)dt$$

describes the dependence of a system's output signal from its input with respect to the input signal's amplitude and its time course. Here, $y(\tau)$ is the system response, $x(t)$ the signal input, and $h(\tau - t)$ the weighting term for time dependent response. As described in the above equation, the reaction of the transfer element is corresponding to the sum of the previous input signals during infinitesimally short periods multiplied with a weighting term that is specific for the transfer system. The weighting term gives for each point of time the strength, with which the respective input signal influences the system's output.

(from <http://www.medical-cybernetics.de/methods/duhamel.html>)

APPENDIX IV-4 : FRECHET' S SERIES

The notions used in this part are derived from

Maurice Fréchet, *Transactions of the American Mathematical Society*, Vol. 5, No. 4, 493-499, Oct. 1904.

Link:

<http://matwbn.icm.edu.pl/ksiazki/sm/sm119/sm11935.pdf>

Read also

M. Buntinas, “Strong summability in Fréchet spaces with applications to Fourier series,” *J. approx. theory*, Vol. 68, n°1, 56-73, 1992.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

APPENDIX IV-5: HURWITZ'S POLYNOMIALS

Read

Ulrich L. Rohde, *Microwave and Wireless Synthesizers*, John Wiley & Sons, 1997, see Appendix A: Mathematical Review (p 507-549)

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

REFERENCES

(In this list, pioneer papers are privileged)

General references

- [1] S.A. Maas, *Nonlinear microwave circuits*, Norwood MA: Artech House, 1988.
- [2] M.E. Hines, "The virtues of nonlinearity - detection, frequency conversion, parametric amplification and harmonic generation," *IEEE Trans. Microwave Theory Tech.*, Vol. 32, 1097-1104, Sept. 1984.
- [3] B.L. Smith, M.H. Carpentier, *The microwave engineering handbook*, Microwave Technology Series, Vol I and II, London: Chapman & Hall, 1993.

Parametric Amplifiers

- [4] Z. Heshmati, I.C. Hunter, R.D. Pollard, "Microwave parametric frequency dividers with conversion gain," *IEEE Trans. Microwave Theory Tech.*, Vol. 55, 2059-2064, Oct. 2007.
- [5] J. Tignon, A. Huynh, P. Roussignol, C. Delalande, R. Ferreira, R. Andre, R. Romestain, L.S. Dang, "Polariton parametric amplifier coherent dynamics," *Physica-E-Low-Dimensional Systems and Nanostructures*, Vol. 21, 820-824, 2004.
- [6] R.E. Lipsey, S.H. Jones, "Accurate design equations for 50-600 GHz GaAs schottky diode varactor frequency doublers," *IEEE Trans. Microwave Theory Tech.*, Vol. 46, 1876-1882, Sept. 1998.
- [7] J. Watson, T. Lepine, P. Georges, A. Brun, "Tunable visible ultrashort pulses by parametric generation and amplification," *IEEE Conf. on Lasers and Electro-Optics Europe*, 16-17, Sept. 1994.
- [8] A. Khanifar, R.J. Green, "Photo-parametric amplifier/converter in subcarrier multiplexed lightwave communication systems," in *IEEE Int. Microwave Symp. Dig.*, Alburquerque, 761-764, 1992.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [9] D.F. Peterson, G.I. Haddad, "Read-type varactors for parametric amplifier applications," *IEEE Trans. Microwave Theory Tech.*, Vol. 28, 945-951, Sep. 1980.
 - [10] C.S. Aitchison, A. Wong, "A VHF hybrid parametric amplifier," *IEEE Trans. Microwave Theory Tech.*, Vol 28, 825-833, Aug. 1980.
 - [11] C.S. Aitchison, A. Wong, "A hybrid parametric amplifier," *IEEE Trans. Microwave Theory Tech.*, Vol 28, 833-839, Aug. 1980.
 - [12] S. Egami, "A design theory for wide-band parametric amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. 22, 119-124, Feb. 1974.
 - [13] R.W. Laton, G.I. Haddad, "Characteristics of IMPATT-diode reflection amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. 21, 668-680, Nov. 1973.
 - [14] H.C. Bowers, "IMPATT diodes start competing as microwave system amplifiers," *Electronics*, Vol. 45, 82-88, Aug. 1972.
 - [15] A.S. Bains, "High-power microwave amplifier using IMPATT diodes," *Electron. Lett.*, Vol. 8, 427-428, Aug. 10, 1972.
 - [16] K.S. Grabowski, *Wzmacniacze I mieszacze parametryczne z dioda pojemnosciowa*, Warszawa: Wydawnictwa N.T., 1968.
 - [17] A.I. Grayzel, "The overdriven varactor upper sideband upconverter," *IEEE Trans. Microwave Theory Tech.*, Vol. 15, 561-565, Oct. 1967.
 - [18] R.L. Ernst, "Multiple idler parametric amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. 15, 9-22, Jan. 1967.
 - [19] L.S. Napoli, R.J. Ikola, "An avalanching silicon diode microwave amplifier," *Proc. IEEE*, Vol. 53, 1231-1232, Sept. 1965.
 - [20] B.S. Perlman, "Current-pumped abrupt junction varactor power frequency converters," *IEEE Trans. Microwave Theory Tech.*, Vol. 13, 150-161, Mar. 1965.
 - [21] J. Chramiec, "On the gain and bandwidth of a parametric super regenerative amplifier," *Bull. Acad. Pol. Sci. Ser. Sci. Tech.*, Vol. 13, 27-33, Oct. 1964.
-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [22] W.J. Getsinger, G.L. Matthaei, "Some aspects of the design of wide-band up-converters and nondegenerate parametric amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. 12, 77-87, Jan. 1964.
- [23] K.K.N. Chang, *Parametric and Tunnel diodes*, Englewood: Prentice-Hall, 1964.
- [24] D. Sabih, "Broadband hybrid-coupled parametric amplifier," *Microwave J.*, Vol. 5, 87-94, May 1962.
- [25] P. Penfield, R.P. Rafuse, *Varactors applications*, Cambridge: MIT Press, 1962.
- [26] G.L. Matthaei, "A study of the optimum design of wide-band parametric amplifiers and up-converters," *IRE Trans. Microwave Theory Tech.*, Vol. 9, 23-38, Jan. 1961.
- [27] W. Louisell, *Coupled mode and the parametric electronics*, New-York: Wiley, 1960.
- [28] H.E. Rowe, "Some general properties of nonlinear elements. Part II : small signal theory," *Proc. IRE*, Vol. 46, May 1958.
- [29] J.M. Manley, H.E. Rowe, "Some general properties of nonlinear elements. Part I : general energy relations," *Proc. IRE*, Vol. 44, 904-913, July 1956.

Parametric Amplifiers

- [30] W. Louisell, *Coupled mode and the parametric electronics*, New-York: Wiley, 1960.
- [31] K.S. Grabowski, *Wzmacniacze I mieszaczce parametryczne z dioda pojemnosciowa*, Warszawa: Wydawnictwa N.T., 1968.
- [32] W.J. Getsinger, G.L. Matthaei, "Some aspects of the design of wide-band up-converters and nondegenerate parametric amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 12, 77-87, Jan. 1964.
- [33] P. Penfield, R.P. Rafuse, *Varactors applications*, Cambridge: MIT Press, 1962.
- [34] K.K.N. Chang, *Parametric and Tunnel diodes*, Englewood: Prentice-Hall, 1964.
- [35] L.S. Napoli, R.J. Ikola, "An avalanching silicon diode microwave amplifier," *Proc. IEEE*, Vol 53, 1231-1232, Sep. 1965.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [36] H.C. Bowers, "IMPATT diodes start competing as microwave system amplifiers," *Electronics*, Vol 45, 82-88, Aug. 1972.
 - [37] A.S. Bains, "High-power microwave amplifier using IMPATT diodes," *Electron. Lett.*, Vol 8, 427-428, Aug. 10, 1972.
 - [38] R.W. Laton, G.I. Haddad, "Characteristics of IMPATT-diode reflection amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 21, 668-680, Nov. 1973.
 - [39] D.F. Peterson, G.I. Haddad, "Read-type varactors for parametric amplifier applications," *IEEE Trans. Microwave Theory Tech.*, Vol 28, 945-951, Sep. 1980.
 - [40] B.S. Perlman, "Current-pumped abrupt junction varactor power frequency converters," *IEEE Trans. Microwave Theory Tech.*, Vol 13, 150-161, Mar. 1965.
 - [41] A.I. Grayzel, "The overdriven varactor upper sideband upconverter," *IEEE Trans. Microwave Theory Tech.*, Vol 15, 561-565, Oct. 1967.
 - [42] R.E. Lipsey, S.H. Jones, "Accurate design equations for 50-600 GHz GaAs schottky diode varactor frequency doublers," *IEEE Trans. Microwave Theory Tech.*, Vol 46, 1876-1882, Sep. 1998.
 - [43] J.M. Manley, H.E. Rowe, "Some general properties of nonlinear elements. Part I : general energy relations," *Proc. IRE*, Vol 44, 904-913, July 1956.
 - [44] H.E. Rowe, "Some general properties of nonlinear elements. Part II : small signal theory," *Proc. IRE*, Vol 46, May 1958.
 - [45] G.L. Matthaei, "A study of the optimum design of wide-band parametric amplifiers and up-converters," *IRE Trans. Microwave Theory Tech.*, Vol 9, 23-38, Jan. 1961.
 - [46] D. Sabih, "Broadband hybrid-coupled parametric amplifier," *Microwave J.*, Vol 5, 87-94, May 1962.
 - [47] J. Chramiec, "On the gain and bandwidth of a parametric super regenerative amplifier," *Bull. Acad. Pol. Sci. Ser. Sci. Tech.*, Vol 13, 27-33, Oct. 1964.
 - [48] R.L. Ernst, "Multiple idler parametric amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 15, 9-22, Jan. 1967.
-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [49] S. Egami, "A design theory for wide-band parametric amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 22, 119-124, Feb. 1974.
- [50] C.S. Aitchison, A. Wong, "A VHF hybrid parametric amplifier," *IEEE Trans. Microwave Theory Tech.*, Vol 28, 825-833, Aug. 1980.
- [51] C.S. Aitchison, A. Wong, "A hybrid parametric amplifier," *IEEE Trans. Microwave Theory Tech.*, Vol 28, 833-839, Aug. 1980.
- [52] A. Khanifar, R.J. Green, "Photo-parametric amplifier/converter in subcarrier multiplexed lightwave communication systems," in *IEEE Int. Microwave Symp. Dig.*, Alburquerque, 761-764, 1992.

Reflection amplifiers

- [53] M.T. Hickson, P. Gardner, D.K. Paul, "High gain millimetric negative resistance low-noise amplifiers," *IEEE Electron Device Lett.*, Vol. 29, 1408-1409, Aug. 1993.
- [54] P. Gardner, D.K. Paul, "Aspects of the design of low noise, negative resistance, reflection mode transistor amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, 1869-1875, Nov. 1991.
- [55] G.V. Rosenbaum, "Feedback mapping applied to the GOES RF amplifiers," *Microwave Syst. News & CT*, 145-155, May 1985.
- [56] R.W. Laton, G.I. Haddad, "Characteristics of IMPATT-diode reflection amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. 21, 668-680, Nov. 1973.
- [57] J.O. Scanlan, J.T. Lim, "A design theory for optimum broadband reflection amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. 12, 504-511, Sept. 1964.

Negative resistance amplifiers

- [58] J.O. Scanlan, J.T. Lim, "A design theory for optimum broadband reflection amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 12, 504-511, Sept. 1964.
-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [59] R.W. Laton, G.I. Haddad, "Characteristics of IMPATT-diode reflection amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 21, 668-680, Nov. 1973.
- [60] G.V. Rosenbaum, "Feedback mapping applied to the GOES RF amplifiers," *Microwave Syst. News & CT*, 145-155, May 1985.
- [61] P. Gardner, D.K. Paul, "Aspects of the design of low noise, negative resistance, reflection mode transistor amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 39, 1869-1875, Nov. 1991.
- [62] M.T. Hickson, P. Gardner, D.K. Paul, "High gain millimetric negative resistance low-noise amplifiers," *IEEE Electron Device Lett.*, Vol 29, 1408-1409, Aug. 1993.

Transistor Amplifiers

- [63] T.T. Ha, *Solid-state microwave amplifier design*, New-York: Wiley, 1981.
 - [64] C. Gentili, *Amplificateurs et oscillateurs micro-ondes*, Paris: Masson, 1984.
 - [65] Vendelin G.D., Pavio A.M., Rohde U. L., *Microwave circuit design using linear and nonlinear techniques*, New York: Wiley & Sons, 1990.
 - [66] G.D. Vendelin, *Design of amplifiers and oscillators by the S-parameter method*, New-York: Wiley, 1982.
 - [67] M.L. Edwards, J.H. Sinsky, "A new criterion for linear 2-port stability using a single geometrically derived parameter," *IEEE Trans. Microwave Theory Tech.*, Vol 40, 2303-2311, Dec. 1992.
 - [68] H. Fukui, "Available power gain, noise figure, and noise measure of two-ports and their graphical representation," *IEEE Trans. Circuit Theory*, Vol 13, 137-142, June 1966.
 - [69] L.A. Geis, L.P. Dunleavy, "Power contour plots using linear simulators," *Microwave J.*, Vol 39, 60-68, June 1996.
 - [70] M.L. Edwards, S. Cheng, "Conditionally stable amplifier design using constant μ -contours," *IEEE Trans. Microwave Theory Tech.*, Vol 44, 2634-2640, Dec. 1996.
 - [71] U.L. Rohde, "Designing a matched low noise amplifier using CAD tools", *Microwave J.*, Vol 29, 154-160, Oct. 1986.
-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [72] G.D. Vendelin, W.C. Mueller, "Noise parameters of microwave transistors," *Microwave J.*, Vol 30, 177-186, Nov. 1987.
 - [73] R.K. Froelich, "An improved model for noise characterization of microwave GaAs FET's," *IEEE Trans. Microwave Theory Tech.*, Vol 38, 703-706, June 1990.
 - [74] G. Capponi, B.D. Maio, P. Livreri, "HEMT for low-noise microwaves: CAD-oriented performance evaluation," *IEEE Trans. Microwave Theory Tech.*, Vol 43, 1226-1229, June 1995.
 - [75] S. Munoz, J.L. Sebastian, J.D. Gallego, "Modeling the bias and temperature dependence of a C-class MESFET amplifier," *IEEE Trans. Microwave Theory Tech.*, Vol 45, 527-233, April 1997.
 - [76] G.L. Matthaei, L. Young, E.M.T. Jones, *Microwave filters, impedance matching networks and coupling structures*, London: Artech House, 1980.
 - [77] Y. Tajima, P.D. Miller, "Design of broad-band power GaAs FET amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 32, 261-267, Mar. 1984.
 - [78] L.D. Abrie, *The design of impedance-matching networks for radio-frequency and microwave amplifiers*, Norwood: Artech House, 1985.
 - [79] M. Staloff, "Computer-aided design and analysis of microwave transistor amplifiers," *Microwave J.*, Vol 30, 253-268, May 1987.
 - [80] A.J. Parsons, R.G. Meadows, "Matching techniques for microwave integrated circuits," *Microwave Eng. Europe*, 47-57, Oct. 1994.
 - [81] H. Yamada, S. Ohara, T. Iwai, Y. Yamaguchi, K. Imanishi, K. Joshin, "Self-linearizing technique for L-band HBT power amplifier: effect of source impedance on phase distortion," *IEEE Trans. Microwave Theory Tech.*, Vol 44, 2398-2402, Dec. 1996.
 - [82] F.J.O. Gonzales, J.L.J. Martin, A.A. Lopez, "Effects of matching on RF power amplifier efficiency and output power," *Microwave J.*, Vol 41, 60-72, April 1998.
 - [83] R.M. Barrett, "Microwave printed circuits - A historical survey," *IRE Trans. Microwave Theory Tech.*, Vol 3 (2), 19, Mar. 1955.
-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [84] K.C. Gupta, R. Garg, R. Chadha, *Computer aided design of microwave circuits*, Dedham MA : Artech House, 1981.
 - [85] T.C. Edwards, *Foundations for microstrip circuit design*, 2nd Ed., Chichester, England : Wiley, 1992.
 - [86] R. Soares, J. Graffeuil, J.J. Obregon, *Application des transistors à effet de champ en arséniure de gallium*, Paris: Eyrolles, 1984.
 - [87] P.F. Combes, J. Graffeuil, J.F. Sautereau, *Composants, dispositifs et circuits actifs en micro-ondes*, Paris: Dunod Université, 1985.
 - [88] M.R. Weiss, D. Pavlidis, "An investigation of the power characteristics and saturation mechanisms in HEMT's and MESFET's," *IEEE Trans. Microwave Theory Tech.*, Vol 36, 1197-1206, Aug. 1988.
 - [89] S. Toyoda, "High efficiency amplifiers," in *IEEE Int. Microwave Symp. Dig.*, San Diego, 253-256, 1994.
 - [90] M. Maeda, H. Takehara, M. Nakamura, Y. Ota, O. Ishikawa, "A high power and high efficiency amplifier with controlled second-harmonic source impedance," in *IEEE Int. Microwave Symp. Dig.*, Orlando, 579-582, 1995.
 - [91] J. Staudinger, "Multiharmonic load terminaison effects on GaAs MESFET power amplifiers," *Microwave J.*, Vol 39, 60-77, June 1996.
 - [92] F.H. Raab, "Class-F power amplifiers with maximally flat waveforms," *IEEE Trans. Microwave Theory Tech.*, Vol 45, 2007-2012, Nov. 1997.
 - [93] R.N. Braithwaite, "Nonlinear amplification of CDMA waveforms: an analysis of power amplifier gain errors and spectral regrowth," in Proc. *Vehicular Technology Conf.*, 2160-2166, 1998.
 - [94] J.S. Kenney, A. Leke, "Design considerations for multicarrier CDMA base station power amplifiers," *Microwave J.*, Vol 42, 76-86, Feb. 1999.
 - [95] W.H. Leighton, "RF amplifier design with large signal S-parameters," *IEEE Trans. Microwave Theory Tech.*, Vol 21, 809-814, Dec. 1973.
-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [96] A. Platzker, Y. Tajima, "Large-signal GaAs FET amplifier CAD program," in *IEEE Int. Microwave Symp. Dig.*, Dallas, 450-452, 1982.
 - [97] W.R. Curtice, R.L. Camisa, "Self-consistent GaAs FET models for amplifier design and device diagnostics," *IEEE Trans. Microwave Theory Tech.*, Vol 32, 1573-1578, Dec. 1984.
 - [98] A. Materka, T. Kacprzak, "Computer calculation of large-signal GaAs FET amplifier characteristics," *IEEE Trans. Microwave Theory Tech.*, Vol 33, 129-135, Feb. 1985.
 - [99] W.R. Curtice, "Nonlinear analysis of GaAs MESFET amplifiers, mixers and distributed amplifiers using the harmonic balance technique," *IEEE Trans. Microwave Theory Tech.*, Vol 35, 441-447, April 1987.
 - [100] C. Guo, M. Camiade, D. Rousset, A. Cessey, J.J. Obregon, A. Bert, "Optimum design of nonlinear FET amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 35, 1348-1354, Dec. 1987.
 - [101] S.A. Maas, *Nonlinear microwave circuits*, Norwood MA: Artech House, 1988.
 - [102] R.J. Gilmore, M.B. Steer, "Nonlinear circuit analysis using the method of harmonic balance - a review of the art : Part I, Introductory concepts; Part II, Advanced concepts," *Int. J. on Microwave and Millimeter CAE*, Vol 1, Jan. / Apr. 1991.
 - [103] G.M Lambrianou., C.S. Aitchison, "Optimization of third-order intermodulation product and output power from an X-band MESFET amplifier using Volterra series analysis," *IEEE Trans. Microwave Theory Tech.*, Vol 33, 1395-1403, Dec. 1985.
 - [104] A.M. Crosmun, S.A. Maas, "Minimization of intermodulation distortion in GaAs MESFET small-signal amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol 37, 1411-1417, Sept. 1989.
 - [105] T. Larsen, "Determination of Volterra Transfer functions of non-linear multiport networks", *Int. Journal of Circuit Theory and Applications*, Vol 21 (2), 107-131, 1993.
 - [106] J.C. Pedro, J. Perez, "Accurate simulation of GaAs MESFET's intermodulation distortion using a new drain-source current model," *IEEE Trans. Microwave Theory Tech.*, Vol 42, 25-33, Jan. 1994.
-

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

- [107] M.C.E. Yagoub, H. Baudrand, J. Boucher, A. Duarte Doria Neto, "Acccurate method for the intermodulation analysis of nonlinear microwave circuits," in Proc. *Int. Microwave and Opto. Conf.*, Natal, Brazil, 1997.
- [108] R.H. Caverly, "Distortion in RF and microwave control devices," *Microwave J.*, Vol 40, 74-82, Dec. 1997.
- [109] T. Mader, J. Schoenberg, L. Harmon, Z.B. Popovic, "Planar MESFET transmission wave amplifier," *Inst. Elect. Eng. Electron. Lett.*, Vol 29, 1699-1701, Sept. 1993.
- [110] A.R. Perkons, T. Itoh, "A 10-element active lens amplifier," in *IEEE Int. Microwave Symp. Dig.*, San Francisco, 1119-1122, 1996.
- [111] A.H. Baree, I.D. Robertson, "Monolithic MESFET distributed baluns based on the distributed amplifier gate-line termination technique," *IEEE Trans. Microwave Theory Tech.*, Vol 45, 188-195, Feb. 1997.
- [112] Datum Inc., "An RF distribution amplifier," *Microwave J.*, Vol 41, 136-138, April 1998.

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to upload the course materials, including any lecture recordings you may have, to any website. If you require clarification, please consult your professor.

Mathematical-Tools

La comparaison des coefficients de x et y dans les premiers et les seconds membres de ces équations, en égard aux expressions de r_2 , r_4 , α_i , β_i , γ_i , donne après quelques transformations

$$(a c - b^2) B = -(r_4 - r_2) \sqrt{a c - b^2}, \quad (a c - b^2) D = -(r_4 - r_2) \sqrt{a c - b^2},$$

d'où $B = D$, ($L_{12} = L_{21}$), ce qu'il fallait démontrer.

Summary

Basing on the probability theory, ONSAGER and CASIMIR have established that the coefficients L_{ik} of the phenomenological relations between the "Flows" J_i and the "Forces" X_i , $J_i = \sum_k L_{ik} X_k$ near the steady states [i.e. taking the expression of variation of entropy as $\Delta S = -(\sum_{i,k} g_{ik} x_i x_k)/2$] satisfy the equalities $L_{ik} = L_{ki}$.

In the preceding it is shown that these are phenomenological consequences of NEWTON's Mechanics in the more general case where ΔS is assumed as $\Delta S = \sum_i g_i x_i - (\sum_{i,k} g_{ik} x_i x_k)/2$.

(Reçu le 14 juillet 1951.)

Lagrange-Hermitesche Interpolation im Komplexen¹⁾

Von HEINZ UNGER, Darmstadt²⁾)

1. Einleitung

Die Interpolation einer analytischen Funktion im Komplexen wird mit zunehmender Bedeutung nichtelementarer Funktionen ein immer dringlicheres Problem. Die dabei zu beachtenden Gesichtspunkte unterscheiden sich im allgemeinen wesentlich von denen der reellen Interpolation. Die Tafeln erfordern im Komplexen doppelten Eingang und können kaum so engmaschig angelegt werden, daß etwa kubische Interpolation [1]³⁾ noch ausreicht. Es sind von vornherein hochgradige Interpolationsformeln ins Auge zu fassen. Dabei muß man versuchen, mit möglichst wenig Angaben auszukommen, um die Tafelwerke nicht zu umfangreich werden zu lassen. Andererseits muß aber auch der Aufwand zur Berechnung eines Zwischenwertes mit hochgradigen Näherungs-

¹⁾ Es handelt sich um eine ausführliche Wiedergabe eines Vortrages, der auf der Tagung der Gesellschaft für angewandte Mathematik und Mechanik (Freiburg i. Br. vom 29. bis 31. März 1951) gehalten wurde. Vgl. auch Z. angew. Math. Mech. 31, 246–247 (1951).

²⁾ Institut für Praktische Mathematik der Technischen Hochschule Darmstadt.

³⁾ Die Ziffern in eckigen Klammern beziehen sich auf das Literaturverzeichnis auf S. 64.

polynomen in erträglichen Grenzen bleiben. Beide Forderungen richtig gegen einander abzuwagen, rechtfertigt besondere Untersuchungen. Es handelt sich hier vornehmlich um die Berechnung einzelner Zwischenwerte, nicht um die Aufgabe der Untertafelung, die wieder andere Methoden erfordert.

In einer Arbeit von SCHNELL [2] ist gezeigt worden, daß bei der Interpolation im Komplexen die Aufspaltung der Funktion in Real- und Imaginärteil ein Umweg ist. Die Verwendung der bekannten Interpolationsformeln, wie Besselsche, Stirlingsche, Everettsche Formel, mit komplexem Argument kommt gleichfalls kaum in Frage, da die Stützwerte bei hochgradiger Interpolation im allgemeinen zu weit vom interessierenden Bereich abliegen, was eine ungünstige Restgliedabschätzung zur Folge hat.

Als naheliegende Stützstellenanordnung erscheint im Komplexen das regelmäßige Polygon. Dabei gelangt man zu besonders vorteilhaften Formeln, wenn man nicht, wie bei der üblichen Lagrangeschen Interpolation, allein die Funktionswerte in den n Eckpunkten des Polygons heranzieht, sondern nach der auf HERMITE [3] zurückgehenden Erweiterung der Interpolationsformeln auch noch die Ableitungen bis zu einer bestimmten Ordnung benutzt. Man kommt auf diese Weise schon mit wenig Stützstellen zu Interpolationsformeln hohen Grades, wie sie für unsere Zwecke benötigt werden. Wir setzen für das Folgende voraus, daß in den Polygoneckpunkten die Funktionswerte und die q ersten Ableitungen (in allen Punkten bis zur gleichen Ordnung¹⁾) bekannt sind. Die Polygoneckpunkte bilden dann die «Gitterpunkte», für deren Argumente die Funktionswerte nebst Ableitungen zu tabellieren sind.

Es wird sich zeigen, daß man sich dabei unter Verwendung eines kartesischen Koordinatensystems im allgemeinen auf das Quadrat als besonders zweckmäßigen Sonderfall des regelmäßigen Polygons beschränken kann.

Für die praktische Durchführung der Interpolation müssen die Interpolationskoeffizienten in einer Hilfstafel vorliegen. Im Falle des Quadrats aber brauchen von den $n(q+1)$ benötigten Koeffizienten nur $q+1$ in einem Quadrat der Seitenlänge 1 mit einer bestimmten Maschenweite (zum Beispiel 1/100) vertafelt zu werden. Den gleichen Vorteil kann man bei anderen Polygonen (zum Beispiel Sechseck) nur unter Verwendung von Polarkoordinaten erreichen. Im allgemeinen ist aber das Arbeiten mit kartesischen Koordinaten bequemer.

Die Formeln werden zunächst alle so aufgestellt, daß der Polygongeschwungspunkt als Stützstelle unberücksichtigt bleibt, da dies den häufigeren Fall der Stützstellenanordnung darstellt. In dem ergänzend behandelten Fall mit Berücksichtigung des Mittelpunktes müssen dann außer den $q+1$ Koeffizienten der Eckpunkte noch die Koeffizienten für den Mittelpunkt bereitgestellt werden; doch braucht dies nur in einem Teile des Quadrats zu geschehen.

¹⁾ Der allgemeinere Fall, daß q_1 Ableitungen in z_1 , q_2 in z_2 usw. berücksichtigt werden, wird hier nicht betrachtet. Weiter bleibt die Verallgemeinerung, daß Ableitungen von beliebigem Grade herangezogen werden, einer besonderen Arbeit vorbehalten.

Den Vorteil der Kenntnis nur einer kleinen Zahl von Interpolationskoeffizienten kann man bei regelmäßigen Polygonen auch dadurch erreichen, daß man zum Beispiel auf der Mitte jeder Polygonseite eine weitere Stützstelle hinzunimmt und mit der gewöhnlichen Lagrange-Formel arbeitet. Es erscheint aber im allgemeinen wesentlich einfacher, an den Eckpunkten die Ableitungen einer speziellen Funktion mit heranzuziehen, da diese oft durch Rekursionsformeln oder Differentialgleichungen gebildet werden können. Schließlich ist auch der Aufbau der Interpolationskoeffizienten nach LAGRANGE-HERMITE besonders einfach, was für die Berechnung dieser Koeffizienten eine nicht unwe sentliche Rolle spielt.

Vorliegende Arbeit beschäftigt sich mit dem Aufbau der Interpolationskoeffizienten und ihrem Zusammenhang untereinander. Bezuglich der praktischen Durchführung sei auf [4] verwiesen.

2. Lagrange-Hermitesche Interpolation

Wir setzen voraus, daß an den durchweg als verschieden angenommenen Stützstellen

$$z_1, z_2, \dots, z_n \quad (1)$$

der komplexen Veränderlichen $z = x + i y$ die Funktionswerte und die q ersten Ableitungen der zu interpolierenden analytischen Funktion $g(z)$ gegeben sind:

$$g_1, g_2, \dots, g_n; \quad g'_1, g'_2, \dots, g'_n; \quad \dots; \quad g_1^{(q)}, \dots, g_n^{(q)}. \quad (2)$$

Damit liegen $n(q+1)$ Angaben vor, und wir suchen ein Polynom $P_\gamma(z)$ vom Grade γ mit

$$\gamma = n(q+1) - 1, \quad (3)$$

welches an den Stützstellen die vorgegebenen Funktionswerte und Ableitungen annimmt.

Wir setzen

$$P_\gamma(z) = \sum_{\mu=1}^n \sum_{p=0}^q L_{\mu,p}^{n,q}(z) g_\mu^{(p)} \quad (4)$$

und damit

$$g(z) = P_\gamma(z) + R_{\gamma+1}(z) \quad (5)$$

mit dem Restglied $R_{\gamma+1}(z)$. Zur Vereinfachung der Schreibweise lassen wir im folgenden bei den Interpolationskoeffizienten $L_{\mu,p}^{n,q}(z)$ die oberen Indizes n, q wieder fort, da Verwechslungen kaum zu befürchten sind, schreiben also einfach $L_{\mu,p}(z)$. Wir führen folgende Vektoren ein¹⁾:

¹⁾ Die geschweifte Klammer wird zur bequemeren Schreibweise einer einspaltigen Matrix verwendet.

$$\left. \begin{aligned} \mathfrak{g} &= \{g_1, g_2, \dots, g_n; \quad g'_1, \dots, g'_n; \quad \dots; \quad g_1^{(q)}, \dots, g_{n+1}^{(q)}\}, \\ \mathfrak{z} &= \{1, z, z^2, \dots, z^\gamma\}, \\ \mathfrak{z}_\mu &= \{1, z_\mu, z_\mu^2, \dots, z_\mu^\gamma\}. \end{aligned} \right\} \quad (6)$$

Weiter benützen wir die $[n(q+1)]$ -reihige Matrix

$$\mathfrak{B}_{n,q} = \left\{ \mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_n, \frac{d\mathfrak{z}_1}{dz_1}, \dots, \frac{d\mathfrak{z}_n}{dz_n}, \dots, \frac{d^q \mathfrak{z}_1}{dz_1^q}, \dots, \frac{d^q \mathfrak{z}_n}{dz_n^q} \right\}, \quad (7)$$

deren Determinante eine Verallgemeinerung der bekannten Vandermondeschen Determinante $\det \mathfrak{B}_n$ darstellt [5] und sich durch diese wie folgt ausdrücken läßt¹⁾:

$$\det \mathfrak{B}_{n,q} = (-1)^{\frac{q(q+1)}{2}} \cdot \frac{n(n-1)}{2} (1! 2! \dots q!)^n [\det \mathfrak{B}_n]^{(q+1)^2} \quad (8)$$

mit

$$\det \mathfrak{B}_n = \prod_{\mu > \nu}^{1,n} (z_\mu - z_\nu). \quad (9)$$

Sind also, wie vorausgesetzt, die z_ν voneinander verschieden, so muß mit $\det \mathfrak{B}_n$ auch $\det \mathfrak{B}_{n,q}$ von Null verschieden sein.

Setzen wir daher für das Näherungspolynom an:

$$P_\gamma(z) = \mathfrak{z}' \mathfrak{a} = \sum_{\nu=0}^{\gamma} a_\nu z^\nu, \quad (10)$$

so erhalten wir zur Bestimmung der $\gamma+1$ Koeffizienten a_ν das Gleichungssystem:

$$\mathfrak{B}_{n,q} \mathfrak{a} = \mathfrak{g}, \quad (11)$$

welches eine eindeutige Lösung besitzt.

In Erweiterung der Darstellung der Lagrangeschen Interpolation [6] läßt sich $P_\gamma(z)$ in der Form schreiben:

$$P_\gamma(z) = -\det \begin{pmatrix} \mathfrak{B}_{n,q} & \mathfrak{g} \\ \mathfrak{z}' & 0 \end{pmatrix} : \det \mathfrak{B}_{n,q}, \quad (12)$$

während das Restglied lautet:

$$R_{\gamma+1}(z) = \det \begin{pmatrix} \mathfrak{B}_{n,q} & \mathfrak{g} \\ \mathfrak{z}' & g \end{pmatrix} : \det \mathfrak{B}_{n,q}. \quad (13)$$

¹⁾ Auf die Literaturstelle [5] wurde ich von Herrn Prof. Dr. U. WEGNER in dankenswerter Weise hingewiesen. Dort wird ein rekurrentes Verfahren zur Berechnung der verallgemeinerten Vandermondeschen Determinante angegeben, die sich von (7) nur durch andere Zeilenanordnung und un wesentliche Faktoren unterscheidet. Der allgemeine explizite Ausdruck findet sich dort nicht. Das Ergebnis (8) wurde von mir auf andere Weise gefunden, indem eine gewöhnliche Vandermondesche Determinante von der Ordnung $n(q+1)$ differenziert wird. Auf diese Weise läßt sich auch der allgemeine Fall, vgl. Fußnote S. 52, mit q_μ Ableitungen in z_μ leicht erledigen.

Zur Durchführung der Interpolation werden die $n(q+1)$ Polynome $L_{\mu,p}(z)$ vom Grade γ

$$L_{\mu,p}(z) = \mathfrak{z}' \mathfrak{c}_{\mu,p} \quad (14)$$

mit

$$\mathfrak{c}_{\mu,p} = \{(c_{\mu,p})_0, (c_{\mu,p})_1, \dots, (c_{\mu,p})_\gamma\} \quad (15)$$

benötigt. Zur Bestimmung der Koeffizienten $(c_{\mu,p})_\nu$ sind $n(q+1)$ Gleichungen erforderlich, die mit dem Vektor

$$\mathfrak{e}_\nu = \left\{ 0, 0, \dots, 1, 0, \dots, 0 \right\}_{\nu\text{-te Zeile}} \quad (16)$$

folgendermaßen lauten:

$$\mathfrak{B}_{n,q} \mathfrak{c}_{\mu,p} = \mathfrak{e}_{n,p+\mu}. \quad (\mu = 1, 2, \dots, n; \quad p = 0, 1, \dots, q) \quad (17)$$

Die Bestimmung der Koeffizienten ist unter den gemachten Voraussetzungen eindeutig. Fassen wir die $n(q+1)$ Vektoren $\mathfrak{c}_{\mu,p}$ zur $(\gamma+1)$ -reihigen quadratischen Matrix

$$\mathfrak{C}_{n,q} = (\mathfrak{c}_{1,0}, \mathfrak{c}_{2,0}, \dots, \mathfrak{c}_{n,0}, \dots, \mathfrak{c}_{1,q}, \dots, \mathfrak{c}_{n,q}) \quad (18)$$

zusammen, so folgt aus (17):

$$\mathfrak{B}_{n,q} \mathfrak{C}_{n,q} = \mathfrak{E}. \quad (19)$$

Die Matrix der Koeffizienten ist also die Kehrmatrix¹⁾ zur verallgemeinerten Vandermondeschen Matrix (7) (vgl. auch [7]).

Da sich die Vandermondesche Determinante (9) nicht ändert, wenn z durch $z - z^*$ ersetzt wird, so ändert sich auch $\det \mathfrak{B}_{n,q}$ nicht. In den Spalten der Matrix $\mathfrak{C}_{n,q}$, deren ν -te Zeile die $(\nu-1)$ -te Ableitung von $L_{\mu,p}(z)$, dividiert durch $(\nu-1)!$, enthält, tritt dieser Ausdruck dann an der Stelle $z = z^*$ anstatt in $z = 0$ auf.

Für die zu interpolierende Funktion $g(z)$ müssen wir voraussetzen, daß $g(z)$ im Innern und auf der Berandung des kleinsten konvexen Polygons, das durch die Punkte z, z_1, z_2, \dots, z_n bestimmt wird, regulär ist. Allerdings muß für die praktische Durchführung zusätzlich gefordert werden, daß die nächste Singularität noch genügend weit entfernt liegt, um eine brauchbare Restgliedabschätzung zu erhalten.

Von dem Restglied, das in allgemeiner Form wie folgt lautet:

$$R_{\gamma+1}(z) = [(z - z_1) \dots (z - z_n)]^{q+1} \frac{1}{2\pi i} \oint_C \frac{g(\zeta)}{[(\zeta - z_1) \dots (\zeta - z_n)]^{q+1}} \cdot \frac{d\zeta}{\zeta - z}, \quad (20)$$

¹⁾ Einen Spezialfall dieser Beziehung mit einer etwas abgeänderten verallgemeinerten Vandermondeschen Determinante findet man bei KOWALEWSKI [7]. Die Elemente der Kehrmatrix stellen auch die Koeffizienten der Partialbruchzerlegungen gewisser leicht angebbarer rationaler Funktionen dar, worauf schon in [5] hingewiesen wird.

wobei C den Rand des oben erwähnten Polygons darstellt, interessieren uns in diesem Zusammenhang hauptsächlich Abschätzungen. Dazu wird die Formel (20) unter Anwendung bekannter Beziehungen (vgl. [6], S. 16, Darstellung des Restgliedes nach HERMITE, Darboux'scher Mittelwertsatz usw.) umgeformt. Hierbei sei auch auf die Arbeit von P. MONTEL [8] verwiesen.

3. Einheitspolygon in der komplexen Ebene

Wir legen als Stützstellenanordnung ein regelmäßiges Polygon von n Ecken und einer Seitenlänge 1 zugrunde, im folgenden kurz Einheitspolygon genannt, und wählen ein z -Koordinatensystem mit dem Nullpunkt im Polygonmittelpunkt. Die Numerierung der Eckpunkte z_μ sei so durchgeführt, daß

$$z_1 = \frac{1}{e^{\frac{i 2\pi}{n}} - 1} = \frac{e^{-i\pi}\left(\frac{1}{2} + \frac{1}{n}\right)}{2 \sin \frac{\pi}{n}} \quad (21)$$

und daß

$$z_\mu = z_1 e^{i \frac{2\pi}{n} (\mu-1)}, \quad (22)$$

so daß

$$z_2 - z_1 = 1 \quad (23)$$

wird (Figur 1). Die Wahl dieses Polygons bedeutet keine Einschränkung, da

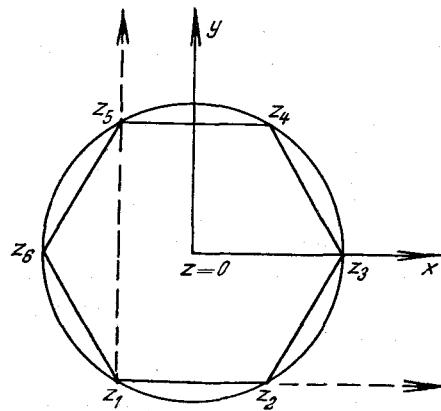


Fig. 1
Lage des Einheitspolygons.

jedes regelmäßige Polygon sich auf das gewünschte durch eine Lineartransformation zurückführen läßt. Laut Voraussetzung soll sich die zu interpolierende Funktion $g(z)$ in und auf dem Rande des Polygons regulär verhalten.

Die Vandermondesche Determinante $\det \mathfrak{B}_n$ hat den Wert:

$$\det \mathfrak{B}_n = \left(2 \sin \frac{\pi}{n} \right)^{n(n-1)/2} \sqrt{n^n} e^{i \frac{\pi}{2} (n-1)(n-2)}.$$

Bei der Quadratwurzel ist immer das positive Zeichen zu nehmen.

Neben der Veränderlichen z werden wir noch die Veränderliche

$$w = z - z_1 \quad (24)$$

benutzen, deren Achsenystem in Figur 1 gestrichelt eingetragen ist. In ihr sind die Koordinaten der laufenden Eckpunkte des Polygons dann:

$$w_\mu = \frac{e^{i \frac{2\pi}{n} (\mu-1)}}{e^{i \frac{2\pi}{n}} - 1} \quad \text{mit } \mu = 1, 2, \dots, n. \quad (25)$$

Die Veränderliche z bzw. w beschränke sich auf das Innere und den Rand des Einheitspolygons.

Das Einheitspolygon stellt die für die komplexe Ebene natürliche Verallgemeinerung gleichabständiger Stützstellen auf der reellen Achse dar.

4. Die Lagrange-Hermiteschen Koeffizienten für das Einheitspolygon

Setzt man die Interpolationsformel (5) für das Einheitspolygon an, so muß der zu berechnende Wert unverändert bleiben, wenn man eine Lineartransformation von z durchführt. Macht man insbesondere die Drehung

$$z^{(\mu)} = z e^{i \frac{2\pi}{n} (1-\mu)} \quad \text{für } \mu = 1, 2, 3, \dots, n, \quad (26)$$

bei der der Punkt $z = z_\mu$ im Uhrzeigersinn in den Punkt z_1 übergeführt wird (vgl. Figur 1), so zeigt sich, daß von den $n(q+1)$ benötigten Koeffizienten nur $q+1$ bereitzustellen sind. Denn es gilt nach Einsetzung von (26) in (5) die Beziehung:

$$L_{\mu,p}(z) = L_{1,p}[z^{(\mu)}] e^{i p \frac{2\pi}{n} (\mu-1)}. \quad (\mu = 1, 2, \dots, n) \quad (27)$$

Für $q=0$ erhält man für die gewöhnlichen Lagrange-Koeffizienten, die wir mit $L_\mu(z)$ bezeichnen, den Zusammenhang:

$$L_\mu(z) = L_{1,p}[z^{(\mu)}].$$

Für jeden p -Wert wird also nur die Kenntnis eines Koeffizienten, nämlich $L_{1,p}$, im Einheitspolygon an n Argumentstellen benötigt. Andererseits würde es auch genügen, jeden Koeffizienten in dem n -ten Teil des Polygons zu kennen.

Interpoliert man im Reellen mit äquidistanten Stützstellen, so benötigt man, wenn n gerade ist, $(q+1) n/2$, bei ungeradem n dagegen $(q+1) [(n+1)/2]$ Koeffizienten. Hierauf ist für die gewöhnlichen Lagrange-Koeffizienten zum Beispiel in [9], S. XVII, hingewiesen. Im Einheitspolygon werden die Verhältnisse also besonders einfach.

Übrigens erhält man durch Differenzieren von (27) für die k -te Ableitung¹⁾

$$L_{\mu, p}^{(k)}(z) = L_{1, p}^{(k)}[z^{(\mu)}] e^{i \frac{2\pi}{n} (\mu-1)(p-k)},$$

womit auch die Ableitungen auf diejenigen mit dem Index $\mu = 1$ zurückgeführt sind.

Die Interpolationsformel (5) lautet unter Berücksichtigung von (27):

$$g(z) = \sum_{\mu=1}^n \sum_{p=0}^q L_{1, p}[z^{(\mu)}] e^{i \frac{2\pi}{n} p(\mu-1)} g_{\mu}^{(p)} + R_{\gamma+1}(z), \quad (5a)$$

wobei $z^{(\mu)}$ nach (26) zu bilden ist.

Führen wir an Stelle von z die Veränderliche w nach (24) ein und bezeichnen wir die Lagrange-Hermiteschen Koeffizienten als Funktionen von w mit $M_{\mu, p}(w)$ und die zu interpolierende Funktion mit $G(w)$, so geht (5a) über in

$$G(w) = \sum_{\mu=1}^n \sum_{p=0}^q M_{1, p}[w^{(\mu)}] e^{i \frac{2\pi}{n} (\mu-1)} G_{\mu}^{(p)} + R_{\gamma+1}^*(w) \quad (5b)$$

mit

$$w^{(\mu)} = w e^{i \frac{2\pi}{n} (1-\mu)} + w_{n-\mu+2} \quad \text{für } \mu = 2, 3, \dots, n. \quad (28)$$

Für die praktische Durchführung der Interpolation wird im allgemeinen die Veränderliche w bevorzugt (vgl. [1]).

Aus (5a) folgt die für Kontrollen wichtige Beziehung

$$\frac{z^k}{k!} = \sum_{\mu=1}^n \sum_{p=0}^k L_{1, p}[z^{(\mu)}] e^{i \frac{2\pi}{n} p(\mu-1)} \frac{z_{\mu}^{k-p}}{(k-p)!},$$

die für die gewöhnlichen Lagrange-Koeffizienten mit $k = 0$ übergeht in

$$1 = \sum_{\mu=1}^n L_1[z^{(\mu)}].$$

Da die Lagrange-Hermiteschen Koeffizienten $L_{\mu, p}(z)$ an allen Stützstellen außer in $z = z_{\mu}$ eine $(q+1)$ -fache Nullstelle besitzen müssen, lässt sich ein wesentlicher Bestandteil durch die gewöhnlichen Lagrange-Koeffizienten aus-

¹⁾ Unter $L_{\mu, p}^{(k)}$ ist jedesmal die Ableitung nach dem Argument zu verstehen.

drücken. Es gilt die Beziehung:

$$L_{\mu, p}(z) = [L_{\mu}(z)^{q+1}] l_{\mu, p}(z). \quad (29)$$

Hierbei stellt $l_{\mu, p}(z)$ ein Polynom q -ten Grades dar. Von den $n(q+1)$ Gleichungen (17) sind durch $[L_{\mu}(z)]^{q+1}$ bereits $(n-1)(q+1)$ erfüllt. Die restlichen Gleichungen

$$[L_{\mu, p}^{(k)}(z)]_{z=z_{\mu}} = \begin{cases} 1 & \text{für } k = p \\ 0 & \text{für } k \neq p \end{cases} \quad \text{mit } k = 0, 1, 2, 3, \dots, q \quad (30)$$

dienen zur Bestimmung der $q+1$ Konstanten des Polynoms $l_{\mu, p}(z)$. Dieses Polynom muß daher noch durch $(z - z_{\mu})^p / p!$ teilbar sein, woraus der für $p = q$ besonders einfache Aufbau der Koeffizienten folgt.

Die Beziehung (29) gilt für jede Stützstellenanordnung. Beim Einheitspolygon wird sie infolge (27) nur für $\mu = 1$ benötigt.

Eine Restgliedabschätzung ist bei der komplexen Interpolation und insbesondere bei der hier benützten hochgradigen von noch größerer Bedeutung als im Reellen, da hier die Genauigkeit eines Zwischenwertes nur durch eine sorgfältige Restgliedabschätzung garantiert werden kann.

Bedeutet $K_{n(q+1)}$ eine obere Schranke für den Betrag der $[n(q+1)]$ -ten Ableitung der Funktion $g(z)$ auf dem Rande des Polygons, so läßt sich folgende Abschätzung des Restgliedes angeben:

$$|R_{\gamma+1}(z)| \leq \left[\frac{\left(\cos \frac{\pi}{n}\right)^n + 1}{\left(2 \sin \frac{\pi}{n}\right)^n} \right]^{q+1} \frac{K_{n(q+1)}}{[n(q+1)]!}. \quad (31)$$

Die Hauptaufgabe bei der praktischen Fehlerabschätzung ist die Abschätzung der $[n(q+1)]$ -ten Ableitung auf dem Rande. Für eine Orientierung genügt es oft, die Abschätzungen in den Eckpunkten zu kennen.

5. Mitnahme des Polygonmittelpunktes

Nimmt man den Mittelpunkt des Polygons als weitere Stützstelle hinzu und setzt dort wieder den Funktionswert $g_0^{(p)}$ als bekannt voraus, so erhält man an Stelle von (5) folgende Interpolationsformel:

$$g(z) = \sum_{\mu=1}^n \sum_{p=0}^q N_{\mu, p}(z) g_{\mu}^{(p)} + \sum_{p=0}^q N_{0, p}(z) g_0^{(p)} + R_{\gamma^*+1}(z), \quad (5c)$$

wobei $\gamma^* = (n+1)(q+1) - 1$ den Grad der Interpolationskoeffizienten darstellt¹⁾. Führen wir wieder die Transformation (26) durch, so erhalten wir für

¹⁾ Wir lassen auch hier die beiden Indizes $n+1$ und q weg und bezeichnen die Koeffizienten mit $N_{\mu, p}(z)$, weil keine Verwechslungen zu befürchten sind.

die Polynome $N_{\mu,p}(z)$ die zu (27) entsprechende Beziehung, wobei also für jeden p -Wert ein Polynom benötigt wird. Außerdem gilt:

$$N_{1,p}(z) = \left(\frac{z}{z_1}\right)^{q+1} [L_1(z)]^{q+1} l_{1,p}^*(z) = [N_1(z)]^{q+1} l_{1,p}^*(z).$$

Dabei sind auch die Polynome $l_{1,p}^*(z)$ wie $l_{1,p}(z)$ vom Grade q , und $N_1(z)$ ist der gewöhnliche Lagrange-Koeffizient.

Für die beim Mittelpunkt auftretenden Polynome gilt:

$$N_{0,p}(z) = N_{0,p}[z^{(\mu)}] e^{i \frac{2\pi}{n} p(\mu-1)} \quad \text{für } \mu = 2, 3, \dots, n.$$

Diese Koeffizienten brauchen also nur in einem Dreieck ($1/n$ des Polygons) bekannt zu sein. Unter Heranziehung des gewöhnlichen Lagrange-Koeffizienten $N_0(z)$ kann der Koeffizient wie folgt ausgedrückt werden:

$$N_{0,p}(z) = [N_0(z)]^{q+1} \frac{z^p}{p!} \quad \text{mit} \quad N_0(z) = e^{i\pi(n/2)} \left(2 \sin \frac{\pi}{n}\right)^n z^n + 1.$$

Bei Vorliegen einer Potenztafel im Komplexen ist dieser Koeffizient einfach zu bilden.

Bei Mitnahme des Mittelpunktes empfiehlt es sich, die Koordinate z auch für die praktische Rechnung beizubehalten.

Es sei noch darauf hingewiesen, daß auch die Formeln, bei denen das Polygon um π/n gedreht ist, sowie der Fall, daß die Stützstellen mit Funktionswerten und Ableitungen auf der Mitte der Polygonseite liegen, leicht auf dem hier beschriebenen Wege hergeleitet werden können.

Für das Restglied gilt mit entsprechenden Bezeichnungen wie bei (31) die Abschätzung:

$$|R_{\gamma^*+1}| \leq \left[\frac{\left(\cos \frac{\pi}{n}\right)^n + 1}{\left(2 \sin \frac{\pi}{n}\right)^{n+1}} \right]^{q+1} \frac{K_{(n+1)(q+1)}^*}{[(n+1)(q+1)]!}.$$

6. Das Quadrat als Einheitspolygon (ohne Mittelpunkt)

Für die praktische Durchführung der Interpolation erweist sich die quadratische Stützstellenanordnung als besonders vorteilhaft. Hier hat man die Koeffizienten $L_{1,p}(z)$ für $p = 0, 1, \dots, q$ zu z -Werten eines quadratischen Netzes im Einheitsquadrat, also einem Quadrat mit der Seitenlänge 1, zu vertafeln. Wählt man als Maschenweite des Netzes zum Beispiel 1/100, so sind für jeden der $q+1$ Interpolationskoeffizienten $L_{1,p}(z)$ 10000 Werte anzugeben. (Durch Spiegelung an der Diagonale von z_1 nach z_3 könnte diese Zahl noch auf etwa die Hälfte reduziert werden, vgl. hierzu [4].) Ist z dann ein Tafelargument, so

sind beim Quadrat auch die Werte $z^{(\mu)}$ nach (26) für $\mu = 2, 3, 4$ wieder Tafelargumente, die zu z in ganz einfacher Beziehung stehen, nämlich:

$$z^{(1)} = x + i y, \quad z^{(2)} = y - i x, \quad z^{(3)} = -x - i y, \quad z^{(4)} = -y + i x.$$

Damit können die Koeffizienten $L_{1,p}$ für alle vier Argumentwerte $z^{(\mu)}$ bequem aus der Tafel abgelesen werden. Die Berechnung der $L_{\mu,p}(z)$ erfordert nach (27)

$$L_{\mu,p}(z) = L_{1,p}[z^{(\mu)}] i^{(\mu-1)p} \quad \text{für } \mu = 2, 3, 4 \quad (27a)$$

dann nur noch die Multiplikation mit Potenzen von i .

Der Übergang auf die für die Vertafelung bequemere Veränderliche w bereitet keine Schwierigkeit.

Für $q = 0, 1$ und 2 ist die Vertafelung der Koeffizienten $L_{1,p}$ in der Veränderlichen w zur Zeit im Gange, und zwar mit einer Maschenweite von $1/100$, die sich aus Restgliedabschätzungen für die meisten Zwecke als die günstigste ergab [4].

Will man feiner als mit der in der Interpolationstafel vorgegebenen Maschenweite (zum Beispiel $1/100$) unterteilen, so kann man entweder zunächst die vier umliegenden in der Tafel enthaltenen Eckpunkte in der angegebenen Weise berechnen und dann mit einer Interpolationsformel niederen Grades nachinterpolieren, deren Genauigkeit dann stets ausreichen wird; oder aber die Interpolation wird nach anderen Gesichtspunkten durchgeführt, worauf teils in Abschnitt 7, teils in einer späteren Arbeit eingegangen wird.

Zur Berechnung der Funktionswerte in der Quadratmitte erhält man für $q = 0, 1$ und 2 die folgenden Formeln:

$$\left. \begin{aligned} g(0) &= \frac{1}{4} [g_1 + g_2 + g_3 + g_4] + R_4, \\ g(0) &= \frac{1}{4} [g_1 + g_2 + g_3 + g_4] + \frac{1+i}{32} [g'_1 + i g'_2 - g'_3 - i g'_4] + R_8, \\ g(0) &= \frac{1}{4} [g_1 + g_2 + g_3 + g_4] + \frac{11(1+i)}{256} [g'_1 + i g'_2 - g'_3 - i g'_4] \\ &\quad + \frac{i}{256} [g''_1 - g''_2 + g''_3 - g''_4] + R_{12}. \end{aligned} \right\} \quad (32)$$

Dabei sind die Funktionswerte und Ableitungen an den Stellen

$$z_1 = -\frac{1+i}{2}, \quad z_2 = \frac{1-i}{2}, \quad z_3 = \frac{1+i}{2}, \quad z_4 = -\frac{1-i}{2}$$

zu nehmen.

Aus (31) erhält man hierbei für R_{12} die Abschätzung

$$|R_{12}| \leq 0,7 \cdot 10^{-10} K_{12}.$$

Dies bedeutet, daß zum Beispiel die Zylinderfunktion $J_0(\zeta)$ für $\zeta = 5,5 + 1,5 i$ nach (32) aus den Werten für

$$\zeta_1 = 5 + i, \quad \zeta_2 = 6 + i, \quad \zeta_3 = 6 + 2i, \quad \zeta_4 = 5 + 2i$$

auf neun Dezimalen genau berechnet werden kann, wenn die Stützwerte sowie die ersten und zweiten Ableitungen auf neun Dezimalen genau bekannt sind.

Vergleicht man bei quadratischer Stützstellenanordnung die Interpolation nach LAGRANGE-HERMITE unter Benützung der ersten und zweiten Ableitungen mit der Interpolation nach der gewöhnlichen Lagrange-Formel, so wird letztere im Mittel bei gleicher Genauigkeit etwa 1/100 der Maschenweite erfordern, angenähert gleiche Beträge für die im Restglied auftretenden Ableitungen vorausgesetzt. Man spart also durch die Verwendung der Lagrange-Hermiteschen Formel in einem einzigen Quadrat die Angabe von etwa 10000 Funktionswerten bei einem Aufwand von 12 komplexen Multiplikationen zur Berechnung eines Zwischenwertes, der in Anbetracht der großen Ersparnis an Zahlenmaterial wohl vertretbar ist.

Auf besondere Eigenschaften der Koeffizienten $L_{1,p}(z)$ wird in [4] eingegangen, wo auch besondere Kontrollformeln aufgeführt sind.

7. Eine Umschreibung der Lagrange-Hermiteschen Formel

Zu einer anderen Vorgehensweise bei der Auswertung der Lagrange-Hermiteschen Interpolationsformel – hierbei erweist es sich als zweckmäßig mit der Variablen w nach (24) zu arbeiten – gelangt man, wenn man bei der Formel (5b) aus dem Koeffizienten $M_{1,p}(w)$ den von p unabhängigen Bestandteil $[M_1(w)]^{q+1}$ gemäß der Beziehung (29) herauszieht:

$$M_{1,p}(w) = [M_1(w)]^{q+1} \lambda_{1,p}(w), \quad (29a)$$

worin $\lambda_{1,p}(w)$ wie das entsprechende $L_{1,p}(z)$ ein Polynom q -ten Grades ist:

$$G(w) = \sum_{\mu=1}^n [M_1[w^{(\mu)}]]^{q+1} \sum_{p=0}^q \lambda_{1,p}[w^{(\mu)}] e^{i \frac{p}{n} \frac{2\pi}{n} (\mu-1)} G_{\mu}^{(p)} + R_{q+1}^{*}(w). \quad (5b^*)$$

Die Koeffizienten von $\lambda_{1,p}(w)$ ergeben sich aus der für die Veränderliche w umgeschriebenen Gleichung (30) und können für ein bestimmtes n und q ein für allemal berechnet werden. Wir setzen

$$\lambda_{1,p}(w) = \sum_{\nu=p}^q a_{\nu,p} w^{\nu} \quad \text{mit} \quad a_{p,p} = \frac{1}{p!} \quad (a_{\nu,p} = 0 \text{ für } \nu < p) \quad (33)$$

und fassen die $a_{\nu,p}$ in einer $(q+1)$ -reihigen quadratischen Matrix zusammen:

$$\mathfrak{A}_{n,q} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{10} & \frac{1}{1!} & 0 & \dots & 0 \\ a_{20} & a_{21} & \frac{1}{2!} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{q0} & a_{q1} & a_{q2} & \dots & \frac{1}{q!} \end{pmatrix}; \quad (34)$$

zum Beispiel erhalten wir für das Quadrat und $q = 2$:

$$\mathfrak{A}_{42} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{9}{2}(1-i) & 1 & 0 \\ -21i & \frac{9}{2}(1-i) & \frac{1}{2} \end{pmatrix}.$$

Ordnet man die zweite Summe in (5b*) um nach Potenzen von w bzw. $w^{(\mu)}$, so erhalten wir:

$$G(w) = \sum_{\mu=1}^n [M_1[w^{(\mu)}]]^{q+1} \sum_{v=0}^q d_{\mu,v} [w^{(\mu)}]^v + R_{\gamma+1}^*(w). \quad (5b**)$$

Die Koeffizienten $d_{\mu,v}$ ergeben sich aus:

$$\begin{pmatrix} d_{\mu 0} \\ d_{\mu 1} \\ \vdots \\ d_{\mu q} \end{pmatrix} = \mathfrak{A}_{n,q} \begin{pmatrix} \varepsilon_{\mu 0} & G_\mu^{(0)} \\ \varepsilon_{\mu 1} & G_\mu^{(1)} \\ \vdots & \vdots \\ \varepsilon_{\mu q} & G_\mu^{(q)} \end{pmatrix} \text{ mit } \varepsilon_{\mu,v} = e^{i \frac{2\pi}{n} v(\mu-1)} \text{ und } v = 0, 1, \dots, q. \quad (35)$$

Im Falle des Quadrats gehen die Faktoren $\varepsilon_{\mu,v}$ wieder in Potenzen von i über.

Die Dreiecksform der Matrix $\mathfrak{A}_{n,q}$ vereinfacht die Ausrechnung der Koeffizienten $d_{\mu,v}$ (bei Verwendung der Veränderlichen z statt w hätte man eine volle Matrix erhalten). Insbesondere gilt:

$$d_{\mu,0} = G_\mu^{(0)} = G_\mu.$$

Die Koeffizienten $d_{\mu,v}$ enthalten die Funktionswerte und die q ersten Ableitungen. Man hat also für eine zu interpolierende spezielle Funktion $G(w)$ an den Eckpunkten des Einheitspolygons diese Koeffizienten bereitzustellen.

Will man aus diesen Werten $d_{\mu,\nu}$ umgekehrt die Ableitungen $G_{\mu}^{(\rho)}$ selbst wieder errechnen, so kann dies infolge der Dreiecksgestalt der Matrix $\mathfrak{A}_{n,q}$ leicht durch Aufrechnung des gestaffelten Gleichungssystems erfolgen.

Liegt die spezielle Funktion $G(w)$ in dieser Weise vertafelt vor, so wird die Interpolation nach (5b**) durchgeführt. Dabei ist es zweckmäßig, die zweite Summe bei Nichtvorhandensein einer Potenztafel mit dem Horner-Schema auszuwerten. Als einziger Koeffizient müßte dann noch $[M_1(w)]^{q+1}$ im Einheitsquadrat vertafelt vorliegen (bei der früher gebrauchten Vorgehensweise mußten $q+1$ Koeffizienten vertafelt werden). Da aber dieser Koeffizient auch verhältnismäßig einfach berechnet werden kann, läßt sich diese Art der Interpolation mit erträglichem Aufwand auch für eine beliebige Argumentstelle durchführen.

Die Anwendung der Formel (5b**) erfordert, wenn die $d_{\mu,\nu}$, berechnet vorliegen und $[M_1(w)]^{q+1}$ vertafelt ist, genau so viele komplexe Multiplikationen wie (5a) bzw. (5b).

Eine Umschreibung der Formeln mit Mittelpunkt kann entsprechend vorgenommen werden.

LITERATURVERZEICHNIS

- [1] A. N. LOWAN und H. E. SALZER, *Coefficients for Interpolation within a Square Grid in the Complex Plane*, J. Math. Phys. 23, 156–166 (1944).
- [2] W. SCHNELL, *Interpolation im Komplexen*, Z. angew. Math. Mech. 30, 274 bis 275 (1950), und Diplomarbeit Technische Hochschule Darmstadt (1950).
- [3] CH. HERMITE, *Sur la formule d'interpolation de Lagrange*, J. reine angew. Math. 84, 64–69 (1878) = Œuvres, Bd. 3 (Gauthier-Villars, Paris 1912), S. 432–443.
- [4] A. WALTHER, H. UNGER und B. ASHENHURST, *Tables for Lagrange Hermite Complex Interpolation*, Harvard University Computation Laboratory (in Vorbereitung).
- [5] J. WELLSTEIN, *Vandermondesche Determinanten, Partialbruchzerlegungen, Interpolationsaufgabe von Lagrange-Hermite*, J. reine angew. Math. 166, 235–240 (1931).
- [6] N. E. NÖRLUND, *Vorlesungen über Differenzenrechnung* (Springer-Verlag, Berlin 1924).
- [7] G. KOWALEWSKI, *Interpolation und genäherte Quadratur* (B. G. Teubner, Leipzig 1932).
- [8] P. MONTEL, *Sur quelques propriétés des différences divisées*, J. Math. pures appl. [9] 16, 219–231 (1937).
- [9] A. N. LOWAN (Mathematical Tables Project, National Bureau of Standards), *Tables of Lagrangian Interpolation Coefficients* (New York, Columbia University Press), 1. Aufl. 1944, 2. Aufl. 1948.
- [10] A. SELOW, *Zur Berechnung und Anwendung der Zylinderfunktionen vom Index 1/3 mit komplexem Argument*, Dissertation Technische Hochschule Darmstadt (1949).

Summary

Increased attention has been recently devoted to the interpolation of an analytic function $g(z)$ in the complex plane. Only if the increment in a given

table is sufficiently small, $g(z)$ may be approximated by a quadratic or cubic polynomial. To get an interpolation polynomial $P(z)$ of higher degree we employ the extension of the Lagrange formula given by HERMITE, based on the values of the function and some of its derivatives:

$$P(z) = \sum_{\mu=1}^n \sum_{p=0}^q L_{\mu,p}(z) g^{(p)}(z_\mu).$$

Choosing the points z_μ in the corners of a regular polygon the relation for the coefficients $L_{\mu,p}$ may be found as follows:

$$L_{\mu,p}(z) = e^{i(2\pi/n)p(\mu-1)} L_{1,p}[z^{(\mu)}],$$

where

$$z^{(\mu)} = z e^{i(2\pi/n)(1-\mu)}.$$

It is convenient and advantageous to choose a square grid. Employing e.g. an interpolation polynomial of the degree 11, it is unnecessary to tabulate the twelve coefficients $L_{\mu,p}(z)$. You need only three, namely $L_{1,0}$, $L_{1,1}$ and $L_{1,2}$. The corner values together with the first and second derivatives of the special function $g(z)$ must be known. It should be noted, that in many cases the derivatives could be easily computed (i.e. by the differential equation).

(Eingegangen: 20. 4. 1951.)

Über die Instabilität von Methoden zur Integration gewöhnlicher Differentialgleichungen

Von HEINZ RUTISHAUSER, Zürich¹⁾

Trotz der bemerkenswerten Publikation von J. TODD²⁾, deren wesentliche Punkte weiter unten wiedergegeben werden, hat der Verfasser seither öfters Verfahren zur numerischen Integration von Differentialgleichungen beobachtet, die, obwohl zwar mit bestechend kleinem Abrechfehler behaftet, doch die große Gefahr der numerischen Instabilität in sich bergen. Ich will vorwegnehmen, daß bei den bewährten Verfahren von RUNGE-KUTTA und ADAMS (Extrapolationsverfahren), sofern sie richtig angewendet werden, diese Gefahr kaum besteht.

Es liegt in der Natur der Sache, daß man eine Differentialgleichung, wenn man sie numerisch lösen will, durch eine Differenzengleichung approximiert und diese löst. Um dabei die Schrittänge nicht allzu klein wählen zu müssen, bevorzugt man solche Differenzengleichungen, die die Differentialgleichung möglichst gut approximieren, dafür dann aber von höherer Ordnung sind als die ursprüngliche Differentialgleichung. Gerade darin liegt aber eine Gefahr, denn damit hat die Differenzengleichung auch eine größere Lösungsmannigfaltigkeit, und es ist gut möglich, daß die numerische Integration gerade eine der «eingeschleppten»

¹⁾ Institut für angewandte Mathematik der ETH.

²⁾ J. TODD, *Solution of Differential Equations by Recurrence Relations*, MTAC 4, 39–44 (1950).

LT2. APPLICATIONS OF THE LAPLACE TRANSFORM

Convolution. Consider a linear system in which the effect at the present time t of a stimulus $f(t_1)dt_1$ at the past time t_1 is proportional to the stimulus. On physical grounds we assume that the proportionality constant depends only on the elapsed time $t - t_1$ and hence is of the form $g(t - t_1)$. The effect at the present time t is therefore

$$f(t_1)g(t - t_1)dt_1.$$

Since the system is linear, the response to the history can be obtained by adding these effects and we are lead to the integral

$$(1) \quad w(t) = \int_0^t f(t_1)g(t - t_1)dt_1.$$

(The lower limit 0 means that the process is assumed to have started at time $t = 0$.)

The expression (1) is called the *convolution* of f and g . It gives the response at the present time t as a weighted superposition over the inputs at times $t_1 \leq t$. The weighting factor $g(t - t_1)$ characterizes the system and $f(t_1)$ characterizes the history of the input.

The function w in (1) is denoted by $f * g$.

The following theorem show that convolution in the t -domain corresponds to multiplication in the s -domain.

Theorem 1. If $f, g \in E$ then $f * g \in E$ and $\mathcal{L}[f * g] = \mathcal{L}[f]\mathcal{L}[g]$.

Proof. I leave the proof of $f * g \in E$ to the reader. By defining $f(t) = 0$ and $g(t) = 0$ for all negative t , we write

$$\begin{aligned} \mathcal{L}[f * g(t)] &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(t_1)g(t - t_1)dt_1 \right) e^{-st}dt \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(t - t_1)e^{-st}dt \right) f(t_1)dt_1. \end{aligned}$$

We set $t - t_1 = t_2$, $t = t_1 + t_2$, and the assertion follows. \square

The convolution plays a prominent role in the study of heat conduction, wave motion, plastic flow and creep, and in other branches of mathematical physics. It is also encountered in sociology, economics, ecology, and genetics in which an effect at a given time t_1 induces a delayed response at a later time t .

Example 2. Consider the problem

$$y'' + \omega^2 y = \omega^2 f(t), \quad y(0) = y'(0) = 0,$$

where ω is a constant and $f \in E$. Taking the Laplace transform yields that

$$\mathcal{L}y = \frac{\omega^2}{s^2 + \omega^2} \mathcal{L}[f],$$

or

$$\mathcal{L}[y] = \mathcal{L}[f]\mathcal{L}[g] = \mathcal{L}[f * g],$$

where $g(t) = \omega \sin \omega t$. By uniqueness, then $y = f * g$, or more explicitly

$$y(t) = \omega \int_0^t f(t_1) \sin \omega(t - t_1) dt_1.$$

The unit impulse function. We consider the response of a system to a narrow spike function (an impulse) that acts for a very short time but produces a large effect, e.g., a point charge or an electron point. To formulate the idea of an impulse, let a be a small positive constant and let $\delta_a(t)$ be the function defined as

$$(2) \quad \delta_a(t) = \begin{cases} \frac{1}{a} & \text{for } 0 \leq t < a, \\ 0 & \text{elsewhere.} \end{cases}$$

Note that the area under the curve is the unity while the function is nonzero only on a small interval. Its Laplace transform is

$$\mathcal{L}[\delta_a(t)] = \int_{0-}^a \frac{1}{a} e^{-st} dt = \frac{1 - e^{-sa}}{sa}.$$

It is a conceptual aid to introduce an expression $\delta(t)$ that describes the effect of $\delta_a(t)$ as $a \rightarrow 0$ and to say that $\mathcal{L}[\delta(t)] = 1$. The symbol $\delta(t)$ is called the *unit impulse function* or the Dirac delta distribution. is loosely thought of as a function on the real line which is zero everywhere except at the origin,

$$\delta(t) = 0 \quad \text{for } x \neq 0,$$

which is subject to

$$\int_{-\infty}^{\infty} \delta(t) dt = 1.$$

Let me say at the outset that $\delta(t)$ is *not* truly a function. Rigorous treatment of the Dirac delta requires the theory of distributions*.

Example 3 (The unit impulse response). We consider the following RLC- circuit problem with the rest initial conditions

$$(D^2 + 2D + 2)x = f(t), \quad x(0) = x'(0) = 0.$$

The *unit impulse response* is the solution $x(t)$ to a unit impulse function $\delta(t)$, that is, the solution to

$$(D^2 + 2D + 2)x = \delta(t), \quad x(0) = x'(0) = 0.$$

Taking the Laplace transform and using that $\mathcal{L}[\delta(t)] = 1$, we obtain in succession

$$\mathcal{L}[x] = \frac{1}{(s+1)^2 + 1}, \quad x(t) = u(t)e^{-t} \sin t.$$

It is immediate to see that the unit impulse input has a lasting effect.

We note that the solution x is continuous for all t and it satisfies the differential equation for $t \neq 0$. At $t = 0$, however, it satisfies neither the differential equation nor the initial conditions, and it is not even differentiable. Indeed, $x'(0-) = 0$ and $x'(0+) = 1$. The unit impulse thus produces a jump of magnitude 1 in $x'(t)$ at $t = 0$.

*A distribution is characterized, not by giving its value $\delta(t)$ at each t , but by giving its value $\delta\phi$ on a suitable class of functions ϕ . Distributions were introduced in mid 1930s by Sergei Sobolev, and independently, in late 1940s by Laurent Schwartz.

Test functions. The basis of the theory of $\delta(t)$ rests on the fact that $\delta(t)$ is not a function whose value is defined for each t , but is defined via the action of $\delta(t)$ on other functions $\phi(t)$, which is called *test functions*. In the general theory of distributions, test functions are assumed to have derivatives of all orders and have support on a finite interval. Here we require that $\phi(t)$ is continuously differentiable as many times as are needed.

The defining characteristic of the Dirac delta is

$$(3) \quad \int_{-\infty}^{\infty} \delta(t)\phi(t)dt = \phi(0).$$

If $\delta(t)$ were an ordinary function and if the integral in (3) were a ordinary Riemann integral, then a change of variable would give

$$\int_{-\infty}^{\infty} \delta(t-c)\phi(t)dt = \int_{-\infty}^{\infty} \delta(t)\phi(t+c)dt.$$

By definition in (3), the right side is $\phi(c)$. This is now taken as the definition of the left side, that is,

$$\int_{-\infty}^{\infty} \delta(t-c)\phi(t)dt = \phi(c).$$

Similarly, we defined the derivative of $\delta(t)$ by

$$\int_{-\infty}^{\infty} \delta'(t)\phi(t)dt = \int_{-\infty}^{\infty} \delta(t)\phi'(t)dt.$$

Approximate identity. The delta function can be viewed as the limit of a sequence of functions $\delta_a(t)$ in (2). The term *approximate identity* has a particular meaning in harmonic analysis, in relation to a limiting sequence to an identity element for the convolution operation. Another kind of nascent delta functions is the normal distribution

$$\delta_a(t) = \frac{1}{a\sqrt{\pi}} e^{-x^2/a^2}.$$

The transfer function. Let $p(D)$ be a linear differential operator with constant coefficient and with the characteristic polynomial $p(s)$. The DE

$$p(D)x = Ae^{st}$$

could be solved, in general, by trying the solution $x = Be^{st}$. Under the assumption that $p(s) \neq 0$, the result is $p(s)Be^{st} = Ae^{st}$, or $B = \frac{1}{p(s)}A$. The function $W(s) = 1/P(s)$ that transforms the input amplitude A into the output amplitude B is called the *transfer function*. The major property of a transfer function is given by

$$(\text{output}) = (\text{transfer function})(\text{input}).$$

Suppose now that we characterize the input f and output x by their Laplace transforms F and X . If $f \in E$ the rest solution of the equation $p(D)x = f$ satisfies

$$p(s)\mathcal{L}x = \mathcal{L}f \quad \text{or} \quad X = W(s)F,$$

where $W(s) = 1/p(s)$ is the transfer function as defined above. This allows an arbitrary input f . Note that the passage from F to X involves merely a multiplication, while the passage from f to x requires the solution of a differential equation. This is one of the advantages of working in the s -domain.

The problem of identification. Here we put in a known input, measure the output, and try to find the differential equation. This is a common view in bio-mathematics and in mathematical medicine. After measuring the response to a known regime of drug dosage, for example, one can try to discover the mechanism by which that response was produced.

It should be said at the outset that the most one can hope for is to find the coefficients in the operator $p(D)$, and if many choices of system parameters give the same $p(D)$ then the situation cannot be further disentangled by mere study of the input-output relation. As an illustration, the vibrations of a frictionless mass-spring system are characterized by two parameters m and k , the mass and the stiffness constant. However, even complete knowledge of the differential equation

$$mx'' + kx = kf(t)$$

yields only the ratio k/m , not k and m separately.

The differential operator $p(D)$ is determined by its characteristic polynomial $p(s)$ and successively $p(s)$ is determined by the transfer function $W(s) = 1/p(s)$. The input function $f(t) \in E$ is said to be nontrivial if its transform $F(s)$ is not identically zero. The output associated with any nontrivial input suffices for the complete determination of the transfer function $W(s)$. For example, the inputs $f_1(t) = u(t)$ and $f_2(t) = \delta(t)$ yields, respectively,

$$W(s) = sX_1(s) \quad \text{and} \quad W(s) = X_2(s).$$

The solutions corresponding to X_1 and X_2 are called, respectively, the unit step response and the unit impulse response. Note that the second equation states that *the transfer function is the unit impulse response*.

Lumped-parameter circuits. The voltage V across a capacitor is related to the charge Q by $V = Q/C$ where $1/C$ is the elastance. Since the current is $I = dQ/dt$ an integration gives

$$V = Q/C = S \int_0^t Idt + Q(0)/C.$$

The initial charge $Q(0)$ is typically 0 and this condition is assumed here.

If the capacitor is connected in series with a resistor of resistance R , an inductor of inductance L and a source of voltage $V(t)$ satisfy

$$L \frac{dI}{dt} + RI + S \int_0^t Idt = V(t).$$

The transform of this with $I(0) = 0$ is

$$(sL + R + s^{-1}S)\mathcal{L}[I] = \mathcal{L}[V].$$

The conditions $Q(0) = 0$ and $I(0) = 0$ are used here.

For a circuit with n independent loops a calculation of the same sort can be carried out for each loop. The result after transformation is a system of the form

$$\sum_{j=1}^n (sL_{ij} + R_{ij} + s^{-1}S_{ij})\mathcal{L}[I_j] = \mathcal{L}[V_i],$$

where I_j is the current and V_i is the impressed voltage associated with the j th loop. The resistances R_{ij} and elastances S_{ij} are 0 for $i \neq j$ but the inductances L_{ij} need not have this property because of the possibility of mutual inductance between branches.

In the matrix form, the above is written

$$V = ZI, \quad I = YV$$

where $Z = sL + R + s^{-1}S$ and $Y = (sL + R + s^{-1}S)^{-1}$. In the first case, I is considered to be input, V is output, and the transfer function Z is an impedance. In the second case V is considered to be input, I is output, and Y is an admittance.

Circuits that are characterized by numbers R_{ij}, L_{ij}, S_{ij} as above are said to be *lumped parameter*, to distinguish them from *distributed parameter* systems such as a waveguide or a submarine cable.

REFERENCES

- [1] Garret Birkhoff & Gian-Carlo Rota, *Ordinary Differential Equations*, 4th ed., New York: Wiley, 1978.
- [2] Ray Redheffer & Dan Port *Differential Equations, theory and applications*, Jones and Barlett publishers, Boston, 1991.

Problems.

1. If f is continuous at $t = c$, where $c > 0$, show that

$$\lim_{t \rightarrow c^-} u(t - c)f(t) = 0, \quad \lim_{t \rightarrow c^+} h(t - c)f(t) = f(c).$$

Hence $u(t - c)f(t)$ is continuous at c if and only if $f(c) = 0$. (Hint: Continuity at c means $\lim_{t \rightarrow c} f(t) = f(c)$.)

2. If $x = 0$ for $t < 0$ and $x'(t) = \delta(t)$ for $t > 0$ then the Laplace transform suggests that $s\mathcal{L}x = 1$. Assuming this, conclude that x agrees with the unit step function $u(t)$ except perhaps at $t = 0$. In that sense $\delta(t) = u'(t)$.

$$z = 1 + \frac{a_{2n-1}}{\bar{z}}, \quad \bar{z} = 1 + \frac{a_{2n}}{z}, \quad n \geq 1.$$

This gives

$$\begin{aligned} a_{2n-1} &= (z - 1)\bar{z}, \\ a_{2n} &= (\bar{z} - 1)z = \bar{a}_{2n-1}, \end{aligned}$$

and it is easily seen that all a_n lie on the boundary of the parabola. The theorem is now completely proved.

BIBLIOGRAPHY

1. J. F. Paydon, *Convergence regions and value regions for continued fractions*, this Bulletin, abstract 47-11-473.
2. W. T. Scott and H. S. Wall, *A convergence theorem for continued fractions*, Transactions of this Society, vol. 47 (1940), pp. 155-172.
3. ———, *Value regions for continued fractions*, this Bulletin, vol. 47 (1941), pp. 580-585.

THE RICE INSTITUTE

A TABLE OF COEFFICIENTS FOR NUMERICAL DIFFERENTIATION

ARNOLD N. LOWAN, HERBERT E. SALZER AND ABRAHAM HILLMAN¹

The following table lists the coefficients $A_{m,s}$ for $m = 1, 2, \dots, 20$ and $s = m, \dots, 20$ in Markoff's formula for the m th derivative in terms of advancing differences, namely

$$\omega^m f^{(m)}(x) = \sum_{s=m}^{n-1} (-1)^{m+s} A_{m,s} \Delta^s f(x) + (-1)^{m+n} \omega^n A_{m,n} f^{(n)}(\xi).$$

In this formula ω is the tabular interval and

$$A_{m,s} = (-1)^{m+s} m B_{s-m}^{(s)} / s(s-m)!$$

and $B_{s-m}^{(s)}$ is the $(s-m)$ th Bernoulli number of the s th order.

Presented to the Society, April 4, 1942 under the title *Coefficients of differences in the expansion of derivatives in terms of advancing differences*; received by the editors March 7, 1942.

¹ The results reported here were obtained in the course of the work done by the Mathematical Tables Project, Work Projects Administration, New York City.

If a function has been tabulated to sufficiently great accuracy and for some suitable interval of the argument along the real axis, the accompanying table may be used to generate the values of the derivatives which in turn may be employed to generate the values of the function in the complex plane within a region where the function is analytic.

The coefficients were computed from the recurrence formula

$$sA_{m,s} = (s - 1)A_{m,s-1} + mA_{m-1,s-1}$$

and checked by independent calculations using the identity

$$x(x + 1)(x + 2) \cdots (x + s - 1) \equiv s! \sum_{j=1}^s A_{j,s} x^j / j!$$

From the identity

$$(x + x^2/2 + x^3/3 + \cdots)^m \equiv A_{m,m} x^m + A_{m,m+1} x^{m+1} + \cdots$$

it was discovered that a prime p is not effectively present in the denominator of an $A_{m,s}$ for which $s < m + p - 1$. The cancellation of prime factors in accordance with this rule was a further check on the accuracy of the work.

The Markoff formula is used at the beginning and end of a table where advancing differences are the only types available. For a full discussion see L. M. Milne-Thomson, *The Calculus of Finite Differences*, chap. 7, pp. 157–159. According to Milne-Thomson the relative simplicity of the remainder term is another advantage over central difference formulae.

Comparison of the Markoff coefficients with central difference coefficients shows the latter to be much smaller and obviously more convenient for obtaining the derivatives of a polynomial sufficiently far away from the ends of a table. However for many important functions in applied mathematics such as Bessel, error, and gamma functions, use of the Markoff formula for a polynomial approximation of some fixed degree might yield a smaller total error due to the particular form of its remainder term.

The first few coefficients of the various formulae may be found in H. T. Davis, *Table of the Higher Mathematical Functions*, vol. 1, pp. 73–77; Whittaker and Robinson, *Calculus of Observations*, pp. 62–65, and in an article by W. S. Bickley *Numerical differentiation near the limits of a difference table*, Philosophical Magazine, (7), vol. 33 (1942), pp. 12–14. (This article lists coefficients of the first 12 derivatives up to those of the 12th difference.)

COEFFICIENTS $A_{m,s}$ IN MARKOFF'S EXPANSION

$m \backslash s$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$	$\frac{1}{10}$	$\frac{1}{11}$	$\frac{1}{12}$	$\frac{1}{13}$	$\frac{1}{14}$
2		1	1	$\frac{11}{12}$	$\frac{5}{6}$	$\frac{137}{180}$	7	363	761	7129	671	83711	6617	1145993
3			1	$\frac{3}{2}$	$\frac{7}{4}$	$\frac{15}{8}$	29	469	29531	1303	16103	190553	128977	9061
4				1	2	$\frac{17}{6}$	7	967	89	4523	7645	341747	412009	9301169
5					1	$\frac{5}{2}$	25	35	1069	285	31063	139381	1148963	355277
6						1	3	$\frac{23}{4}$	9	3013	781	242537	48035	1666393
7							1	$\frac{7}{2}$	$\frac{91}{12}$	$\frac{105}{8}$	4781	13321	314617	790153
8								1	4	$\frac{29}{3}$	$\frac{55}{3}$	10831	897	944311
9									1	$\frac{9}{2}$	12	$\frac{99}{4}$	$\frac{1747}{40}$	5551
10										1	5	$\frac{175}{12}$	$\frac{65}{2}$	491
11											1	$\frac{11}{2}$	$\frac{209}{12}$	1001
12												1	6	$\frac{41}{2}$
13													1	$\frac{13}{2}$
14														1

MATHEMATICAL TABLES PROJECT
70 Columbus Avenue

$$\omega^m D^m f(x) \sim \sum_{s=m}^{s=n} (-1)^{m+s} A_{m,s} \Delta^s f(x)$$

15	16	17	18	19	20	$\frac{s}{m}$
$\frac{1}{15}$	$\frac{1}{16}$	$\frac{1}{17}$	$\frac{1}{18}$	$\frac{1}{19}$	$\frac{1}{20}$	1
1171733	1195757	143327	42142223	751279	275295799	2
2702700	2882880	360360	110270160	2042040	775975200	
30946717	39646461	58433327	344499373	784809203	169704792667	3
17199000	22422400	33633600	201801600	467812800	102918816000	
406841	35118025721	4446371981	80847323107	2263547729	32262100943	
71280	6054048000	756756000	13621608000	378378000	5360355000	4
21939781	2065639	2195261857	371446039969	27566944753	31938836201	5
1496880	133056	134534400	21794572800	1556755200	1743565824	
22463	277382447	38101097	1356664151597	162356544377	694142313941	6
720	7983360	997920	32691859200	3632428800	14529715200	
899683	2271089	86853967	13195009	227663026369	2022480780283	
16200	34560	1140480	152064	2335132800	18681062400	7
35717	54576553	8424673	334947281	9764119	5013017410969	
432	518400	64800	2138400	52800	23351328000	8
515261	23915	76492463	21878439	4065163957	3975325483	9
5040	168	403200	89600	13305600	10644480	
2485	324509	59279	79243781	11795941	6063698587	10
24	2016	252	241920	26880	10644480	
30217	1199	494351	1513391	18843187	367394203	
360	8	2016	4032	34560	483840	11
105	26921	6341	5490071	976163	354467473	12
2	240	30	15120	1680	403200	
143	65	35269	46631	3965533	10596053	13
6		240	160	7560	12096	
7	329	238	136241	31521	6406481	
	12	3	720	80	8640	14

WORK PROJECTS ADMINISTRATION
New York City

COEFFICIENTS $A_{m,s}$ IN MARKOFF'S EXPANSION
 $\omega_m D^m f(x) \sim \sum_{s=m}^{s=n} (-1)^{m+s} A_{m,s} \Delta^s f(x)$

$m \backslash s$	15	16	17	18	19	20
15	1	$\frac{15}{2}$	$\frac{125}{4}$	$\frac{765}{8}$	$\frac{11519}{48}$	$\frac{50255}{96}$
16		1	8	$\frac{106}{3}$	114	$\frac{18017}{60}$
17			1	$\frac{17}{2}$	$\frac{119}{3}$	$\frac{1615}{12}$
18				1	9	$\frac{117}{4}$
19					1	$\frac{19}{2}$
20						1

NEW YORK CITY

2. ———, *On the solvability of non-linear functional equations*, Duke Math. J. (to appear).
3. ———, *Variational boundary value problems for quasi-linear elliptic equations of arbitrary order*, Proc. Nat. Acad. Sci. U.S.A. (to appear).
4. ———, *Variational boundary value problems for quasi-linear elliptic equations*. II, Proc. Nat. Acad. Sci. U.S.A. 50 (1963), 31–37.
5. ———, *Variational boundary value problems for quasilinear elliptic equations*. III, Proc. Nat. Acad. Sci. U.S.A. (to appear).
6. P. D. Lax and A. N. Milgram, *Parabolic equations*, Annals of Mathematics Study No. 33, Princeton, N. J., 1954, pp. 167–190.
7. G. J. Minty, *Monotone (non-linear) operators in Hilbert space*, Duke Math. J. 29 (1962), 341–346.
8. M. M. Vainberg and R. I. Kachurovski, *On the variational theory of non-linear operators and equations*, Dokl. Akad. Nauk 129 (1959), 1199–1202.
9. M. I. Visik, *Solution of a system of quasilinear equations having divergence form under periodic boundary conditions*, Dokl. Akad. Nauk 137 (1961), 502–505.
10. ———, *Boundary value problems for quasilinear strongly elliptic systems in divergence form*, Dokl. Akad. Nauk 138 (1961), 518–521.
11. ———, *Simultaneous quasilinear elliptic equations with lower order terms*, Dokl. Akad. Nauk 144 (1962), 13–16.
12. M. I. Visik and O. A. Ladyzenskaya, *Boundary problems for partial differential equations and some classes of operator equations*, Uspehi Mat. Nauk 11 (1956), 41–97.

INSTITUTE FOR ADVANCED STUDY AND
UNIVERSITY OF CHICAGO

ERRATUM, VOLUME 48

Arnold N. Lowan, Herbert E. Salzer and Abraham Hillman, *A table of coefficients for numerical differentiation*, pp. 920–924.

Page 924, $A_{m,s}$ for $m=18, s=20$, reads $\frac{117}{4}$, should read $\frac{177}{4}$.

Fréchet algebras and formal power series

by

GRAHAM R. ALLAN (Cambridge)

Abstract. The class of elements of *locally finite closed descent* in a commutative Fréchet algebra is introduced. Using this notion, those commutative Fréchet algebras in which the algebra $\mathbb{C}[[X]]$ may be embedded are completely characterized, and some applications to the theory of automatic continuity are given.

1. Introduction. We write \mathcal{F} for the algebra $\mathbb{C}[[X]]$ of all formal power series in a single variable X , with complex coefficients. (An elementary account of the algebraic theory of \mathcal{F} may be found in [7], Chapter 1, §1.) In 1972, the author gave in [1] necessary and sufficient conditions on a commutative Banach algebra A for \mathcal{F} to be embeddable in A (in a purely algebraic sense). This involved the introduction of a new notion, that of an element of *finite closed descent* in a Banach algebra (the property being used in [1], but not given a name until [2]).

We recall this notion: as in [1], Section 2, it is convenient to do this in a rather general context. A *topological algebra* will here be a non-zero complex algebra which is a Hausdorff topological vector space in which the ring multiplication is separately continuous. An *F-algebra* will be a complete metrizable topological algebra (in which case, the multiplication is necessarily jointly continuous [4]). Let A be a commutative topological algebra and let $x \in A$. Then x is said to have *finite closed descent* (FCD) if and only if, for some integer $m \geq 0$, Ax^{m+1} is dense in Ax^m . (We adopt the convention that, when $m = 0$, Ax^m means A , even when A has no identity element.) We also write $\delta(x)$ for the least integer m having this property, and may conventionally write $\delta(x) = \infty$ to indicate that an element does not have FCD.

1991 *Mathematics Subject Classification*: Primary 46J05; Secondary 46J45.

A brief version of this paper was given as a lecture at the conference on Topological Algebras, held at the Banach Center, 23–27 October 1995. The author wishes to thank the Mathematical Institute of the Polish Academy of Sciences for its support and hospitality during October 1995, when the work for this paper was carried out.

Notice that always, at one extreme, every invertible element x has $\delta(x) = 0$, while also the zero element has FCD, $\delta(0) = 1$. The most elementary properties of elements of FCD are conveniently recalled in a lemma.

LEMMA 1. *Let A be a commutative topological algebra and let $x, y \in A$. Then:*

- (i) if $\delta(x) = m < \infty$, then Ax^n is dense in Ax^m for all $n \geq m$;
- (ii) if x is nilpotent and if m is the least integer such that $x^m = 0$, then $\delta(x) = m$ provided A has a 1; if A does not have a 1, then $\delta(x) = m$ or $m - 1$ (and both cases occur);
- (iii) $\delta(xy) \leq \max(\delta(x), \delta(y))$;
- (iv) if A has a 1 and if the set of invertible elements is open (e.g. if A is a unital Banach algebra) then $\delta(x) = 0$ if and only if x is invertible;
- (v) if A does not have a 1 and if A_+ is the unitization of A then, for any $x \in A$, $\delta_{A_+}(x) = \delta_A(x)$ or $\delta_A(x) + 1$ (and both cases occur).

Proof. See [2], p. 462, remarks (1)–(5). (There the results were proved for a Banach algebra, but the weaker assumption that A is a topological algebra makes no essential difference.)

Next, for $x \in A$, we define the mapping $L_x : A \rightarrow A$ by $L_x(y) = xy$ ($y \in A$). We also write $I(x) = \bigcap_{n \geq 1} Ax^n$, so that $I(x)$ is an ideal of A (not in general closed); also, throughout the paper, q_x will denote the quotient homomorphism $q_x : A \rightarrow A/I(x)$.

Then we have the following elementary, but crucial, lemma.

LEMMA 2. *Let A be a commutative topological algebra and let $x \in A$ have FCD. Then L_x maps $I(x)$ bijectively onto itself.*

Proof. See [1], Lemma 1.

In the generality of Lemma 2, it may easily happen that $I(x) = 0$. The chief tool for obtaining non-trivial cases with $I(x) \neq 0$ (and for much else in this paper) is the Mittag-Leffler theorem on inverse limits. There is an interesting survey of the use of Mittag-Leffler methods in [10]. Because of its great importance for us, we shall recall the basic ideas.

Let $(X_n)_{n \geq 1}$ be a sequence of sets and, for each $n \geq 1$, let $d_n : X_{n+1} \rightarrow X_n$ be a mapping; we say that $(X_n; d_n)$ is an *inverse-limit* (or *projective-limit*) sequence. The inverse limit of the sequence, denoted by $\varprojlim(X_n; d_n)$, is the set of all elements $x = (x_n)_{n \geq 1}$ in the cartesian product $\prod_{n \geq 1} X_n$ such that $x_n = d_n(x_{n+1})$ (for all $n \geq 1$). We write $\pi_m : \prod_{n \geq 1} X_n \rightarrow X_m$ for the m th coordinate projection. If the X_n are Hausdorff topological spaces and the mappings d_n are continuous, then $\varprojlim(X_n; d_n)$ is a closed subset of the product space $\prod_{n \geq 1} X_n$; it is then given the subspace topology.

We have the following consequence of Lemma 2.

COROLLARY 1. *Under the conditions of Lemma 2, the following are equivalent for an element $a \in A$:*

- (i) $a \in I(x)$;
- (ii) there is a sequence $(a_n)_{n \geq 0}$ in A such that

$$a = a_0, \quad a_0 = a_1 x, \quad a_1 = a_2 x, \quad \dots, \quad a_n = a_{n+1} x, \quad \dots$$

Moreover, under the equivalent conditions (i) and (ii), the sequence (a_n) lies in $I(x)$ and is uniquely determined by a . The mapping $a \mapsto (a_n)_{n \geq 0}$ is an algebra-isomorphism between $I(x)$ and the inverse limit of the sequence

$$A \xleftarrow{L_x} A \xleftarrow{L_x} A \xleftarrow{L_x} \dots$$

Proof. If $a \in I(x)$ then, by Lemma 2, there is a unique sequence $(a_n)_{n \geq 0}$ in $I(x)$ such that $a_0 = a$ and $a_n = a_{n+1} x$ for all $n \geq 0$. But also, given any such sequence (a_n) in A , we have $a_n = a_{n+k} x^k$ for each $n \geq 0$ and $k \geq 1$, so that in fact $a_n \in I(x)$ for all n . The final statement of the corollary is then immediate from the definition of inverse limit.

We now recall the Mittag-Leffler theorem.

THEOREM 1. *Let $(X_n; d_n)_{n \geq 1}$ be an inverse-limit sequence in which each X_n is a complete metric space and each d_n is a continuous mapping with $d_n(X_{n+1})$ dense in X_n (for all $n \geq 1$). Then, for each m , $\pi_m(\varprojlim(X_n; d_n))$ is dense in X_m . In particular, $\varprojlim(X_n; d_n) \neq \emptyset$, provided that each $X_n \neq \emptyset$.*

Proof. See e.g. [6], Theorem 2.4, [10], Theorem 2.14.

COROLLARY 2. *Let A be an F -algebra and let $x \in A$ have FCD, say $\delta(x) = m$. Then $\overline{I(x)} = \overline{Ax^m}$. In particular, $I(x) = 0$ if and only if x is nilpotent.*

Proof. Just consider the inverse limit sequence

$$I_m \xleftarrow{L_x} I_m \xleftarrow{L_x} I_m \xleftarrow{L_x} \dots,$$

in which each space is $I_m = \overline{Ax^m}$ and each mapping is L_x (restricted to I_m). Then I_m is a closed subspace of the F -algebra A and so is a complete metric space. Each $L_x|_{I_m}$ is continuous and has dense range, since $\delta(x) = m$. The result is then immediate from Theorem 1. If $I(x) = 0$ then $I_m = 0$, so $x^{m+1} = 0$. Conversely, it is clear that $I(x) = 0$ if x is nilpotent.

A somewhat more elaborate version of the same idea gives:

LEMMA 3. *Let A be an F -algebra with 1 and let $x \in A$ have FCD, $\delta(x) = m$. Let $(a_n)_{n \geq 0}$ be a given sequence in A and define $\alpha_n = \sum_{k=0}^n a_k x^k$ ($n \geq 0$). Then there is a dense subset J_0 of Ax^{m+1} such that, for every $\beta \in J_0$, we have $\alpha_n - (\alpha_m + \beta) \in Ax^{n+1}$ for all $n \geq 0$.*

Proof. See [1], Lemma 2. A clearer explanation (of a slightly less general version) is given by Esterle in [10], Theorem 3.2. (In [1] we, rather perversely, deduced Corollary 2 from Lemma 3.)

PROPOSITION 1. *Let A be a commutative F -algebra with 1 and let $x \in A$ have FCD, $\delta(x) = m$. Then there is a unique unital homomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ such that $\Psi_x(X) = q_x(x)$. Moreover, Ψ_x is injective if and only if $x^m \notin Ax^{m+1}$.*

Proof. The existence of the unique homomorphism Ψ_x is [1], Lemma 3. If Ψ_x is not injective then, since $\{\mathcal{F}X^k : k \geq 0\}$ are precisely all the non-zero ideals of \mathcal{F} , it follows that $x^k \in I(x)$ for some $k \geq 0$. Then, by Lemma 2, $x^k = x^k j$, for a unique $j \in I(x)$. But also, for any other $i \in I(x)$, $i = x^k i'$ for some $i' \in I(x)$ and then $ij = (x^k i')j = x^k i' = i$. Since $I(x)$ is dense in Ax^m (by Corollary 2), $ax^m j = ax^m$ ($a \in A$) and so $Ax^m \subseteq I(x) \subseteq Ax^{m+1}$; in particular, $x^m \in Ax^{m+1}$.

Conversely, if $x^m \in Ax^{m+1}$ then, by an easy induction, $x^n \in Ax^n$ for all $n \geq m + 1$; hence $x^m \in I(x)$, so $\Psi_x(X^m) = 0$ and Ψ_x is not injective.

In the case when A is a Banach algebra, the converse to Proposition 1 also holds. Thus:

PROPOSITION 2. *If A is a commutative Banach algebra with 1, $x \in A$ and if there is a unital homomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ such that $\Psi_x(X) = q_x(x)$, then x has FCD.*

Proof. See [2], Theorem 2 (implication (b) \Rightarrow (a); the main ingredient in the proof was Theorem 1 of [1]).

In particular, if there is a homomorphism $\theta : \mathcal{F} \rightarrow A$ with $\theta(X) = x$, then, taking $\Psi_x = q_x \circ \theta$, we see that x must have FCD. In fact, the main result of [1] may be summarized as follows (where $\text{rad } A$ is the Jacobson radical of A):

THEOREM 2. *Let A be a commutative unital Banach algebra and let $x \in A$. Then the following are equivalent:*

- (i) *there is a unital homomorphism $\theta : \mathcal{F} \rightarrow A$ such that $\theta(X) = x$;*
- (ii) *$x \in \text{rad } A$ and x has FCD.*

Moreover, if the equivalent conditions (i) and (ii) hold, then θ is injective if and only if x is not nilpotent.

Proof. See [1], Theorems 1 and 2 (and [2], Theorem 2).

However, as we shall soon see, after recalling some definitions, if in Theorem 2 (or in Proposition 2) we require only that A be a commutative Fréchet algebra, then it is no longer necessary that x should have FCD. (See Example 1 below.) It turns out that in order to obtain sharp versions of these

theorems for Fréchet algebras, with necessary and sufficient conditions, we must consider elements that satisfy a weak form of the FCD condition, that we call “locally finite closed descent”. The definition will be given in the next section.

2. Fréchet algebras. A Fréchet algebra A is an F -algebra whose topology may be defined by a sequence $(p_n)_{n \geq 1}$ of submultiplicative seminorms. Without loss of generality, we may (and shall) take the sequence (p_n) to be increasing. Recall that a seminorm p on A is continuous if and only if for some $K > 0$ and integer $n \geq 1$, $p(x) \leq Kp_n(x)$ ($x \in A$). The basic theory of Fréchet algebras was introduced in [5] and [11]. The principal tool in the study of Fréchet algebras is a representation of A as an inverse limit of Banach algebras. We shall briefly describe this, in order to establish notation. (We are concerned here with commutative Fréchet algebras, though, for the most basic elements of the theory, the commutativity is not important.)

Thus, let A be a commutative Fréchet algebra, with its topology defined by the increasing sequence $(p_n)_{n \geq 1}$ of submultiplicative seminorms. For each n let $\pi_n : A \rightarrow A/\ker p_n$ be the quotient map; then $A/\ker p_n$ is naturally a normed algebra, normed by setting $\|\pi_n(x)\|_n = p_n(x)$ ($x \in A$). We let $(A_n; \|\cdot\|_n)$ be its completion, so that A_n is a commutative Banach algebra; henceforth we consider π_n as a mapping from A into A_n . (It is important to note that $\pi_n(A)$ is a dense subalgebra of A_n but that, in general, $\pi_n(A) \neq A_n$.) Since $p_n \leq p_{n+1}$, there is a, naturally induced, norm-decreasing homomorphism $d_n : A_{n+1} \rightarrow A_n$ such that $d_n \circ \pi_{n+1} = \pi_n$, for all n . Since $\text{im } d_n \supseteq \text{im } \pi_n$, it follows that $d_n(A_{n+1})$ is dense in A_n for each n . For an element $x \in A$, we shall usually write $x_n = \pi_n(x)$; it is then evident that, for each $x \in A$, the sequence $(x_n)_{n \geq 1}$ is an element of $\varprojlim(A_n; d_n)$.

The elementary, but fundamental, structure theorem for Fréchet algebras is:

THEOREM 3 (Arens–Michael isomorphism). *Let A be a (commutative) Fréchet algebra with a defining sequence of seminorms (p_n) . Then, with the above notation, the mapping $x \mapsto (x_n)_{n \geq 1}$ is a topological-algebra isomorphism of A with $\varprojlim(A_n; d_n)$.*

Proof. See [11], Theorem 5.1 (proved for more general locally multiplicatively convex algebras).

The main point of Theorem 3 should be emphasized: given elements $x_n \in A_n$ such that $x_n = d_n(x_{n+1})$ for all $n \geq 1$, there is a unique $x \in A$ such that $\pi_n(x) = x_n$ for all n . (It should be noted that what we write as A_n appears as \bar{A}_n in [11].) The inverse-limit representation of A given by Theorem 3 will be called an Arens–Michael representation of A .

EXAMPLES. 1. The algebra $\mathcal{F} = \mathbb{C}[[X]]$ has a natural Fréchet-algebra topology. For $f = \sum_{n \geq 0} \lambda_n X^n \in \mathcal{F}$ and each $m \geq 0$, define $p_m(f) = \sum_{k=0}^m |\lambda_k|$. It is readily checked that then (p_m) is an increasing sequence of submultiplicative seminorms on \mathcal{F} defining a Fréchet-algebra topology, say κ , on \mathcal{F} . It is called the topology of *coefficientwise convergence*. In [1], Lemma 2, Corollary 2, it was shown that κ is the unique F -algebra topology on \mathcal{F} . We may refer to κ as “the Fréchet topology of \mathcal{F} ”. We recall that the non-zero ideals of \mathcal{F} are just the principal ideals $\mathcal{F}X^k$ ($k \geq 0$); each of these is closed in \mathcal{F} , so that, in particular, the element X does not have FCD in $(\mathcal{F}; \kappa)$.

Thus, by considering the identity map $i : \mathcal{F} \rightarrow (\mathcal{F}; \kappa)$, we see that, for A to be a Fréchet algebra for which there exists a monomorphism $\theta : \mathcal{F} \rightarrow A$, it is not necessary that $x = \theta(X)$ should have FCD.

2. Let U be any open subset of \mathbb{C}^n and let $\mathcal{O}(U)$ be the algebra of all complex-valued holomorphic functions on U , with the usual topology of local uniform convergence. It is well known that $\mathcal{O}(U)$ is a Fréchet algebra in this topology. Write $U = \bigcup_{n \geq 1} K_n$, where each K_n is compact and $K_n \subseteq \text{int } K_{n+1}$, and set $p_n(f) = \sup\{|f(z)| : z \in K_n\}$. Then (p_n) is an increasing sequence of submultiplicative seminorms on $\mathcal{O}(U)$ that defines its topology.

3. We may, in the last example, take the algebra $C(U)$ of all continuous complex-valued functions on U , with the rest of the definition being analogous.

The discussion of these examples will be continued, and other examples introduced, after Proposition 3 below.

We now have the main new definition of the paper. Let A be a commutative Fréchet algebra and let $x \in A$. We say that x has *locally finite closed descent* (LFCD) if and only if, for each continuous submultiplicative seminorm p on A , x has FCD relative to the p -topology (i.e. there is some integer N , which may depend on the seminorm p , such that Ax^{N+1} is p -dense in Ax^N). Equivalently, if $A = \varprojlim(A_n; d_n)$ is an Arens–Michael representation of A as an inverse limit of Banach algebras, then $x \in A$ has LFCD if and only if, for each n , $x_n = \pi_n(x)$ has FCD in the Banach algebra A_n .

It is clear from Lemma 1(v) that, if A has no identity and $x \in A$, then x has LFCD in A if and only if it has LFCD relative to the unitization A_+ of A . Thus, for most purposes, we may, without loss of generality, assume that A is unital.

If we write $\delta_n(x) = \delta(x_n)$, then the fact that (p_n) is an increasing sequence implies that $\delta_n(x)$ is a non-decreasing sequence of positive integers. It is immediate that an element $x \in A$ has FCD if and only if it has LFCD and the sequence $\delta_n(x)$ is bounded (i.e. there is an integer N such that $A_n x_n^{N+1}$ is dense in $A_n x_n^N$ for all n).

In particular, if A is a Banach algebra, then $x \in A$ has LFCD if and only if it has FCD.

There is a special case of LFCD that we single out. An element x (of a commutative Fréchet algebra A) will be called *locally nilpotent* if and only if, for each continuous submultiplicative seminorm p on A , there is a positive integer N (depending on p) such that $p(x^N) = 0$. Again, it is clear that, if $A = \varprojlim(A_n; d_n)$ is an Arens–Michael representation of A , then x is locally nilpotent if and only if x_n is nilpotent for each n . Also, a locally nilpotent element is nilpotent if and only if there is some N such that $x_n^N = 0$ for all n . Again, if A is a Banach algebra, then nilpotence and local nilpotence are equivalent properties.

PROPOSITION 3. *Let x be a locally nilpotent element of a commutative Fréchet algebra A . Then $x \in \text{rad } A$.*

Proof. Let x be a locally nilpotent element of A . Then for each n , x_n is nilpotent, so $\text{Sp}_{A_n}(x_n) = \{0\}$. But $\text{Sp}_A(x) = \bigcup_{n \geq 1} \text{Sp}_{A_n}(x_n)$ (see [11], Theorem 5.3(a)), so $\text{Sp}_A(x) = \{0\}$ and $x \in \text{rad } A$.

EXAMPLES. 1. Let $\mathcal{F} = \mathbb{C}[[X]]$ with its Fréchet topology κ . Then X is locally nilpotent, for with $p_m(\sum \lambda_n X^n) = \sum_{n=0}^m |\lambda_n|$, it is clear that $p_m(X^n) = 0$ for all $n > m$. In particular, therefore, X has LFCD; but, as remarked in the earlier discussion of this example, it does not have FCD. Moreover, every element of \mathcal{F} has LFCD, since, for any $f \in \mathcal{F}$, either f is invertible (so $\delta(f) = 0$), or $f \in \mathcal{F}X$ and f is locally nilpotent.

2. Let U be a connected open subset of \mathbb{C}^n , $A = \mathcal{O}(U)$ in its standard Fréchet topology. Then we claim that A has no elements of LFCD, apart from the trivial cases of zero and the invertible elements, which always have FCD. (We remark that, if U were not connected, there would be other, more or less trivial, examples of elements of FCD. For example, we could take a function f that was identically zero on some components of U but nowhere zero on the remaining components.)

Suppose, then, that $f \in A$, $f \neq 0$ and f not invertible. Then there exists $a = (a_1, \dots, a_n) \in U$ such that $f(a) = 0$ but f is not identically zero on any neighbourhood of a . Let K be a compact polydisc centered at a , $K \subset U$, and let $p(f) = \sup_{z \in K} |f(z)|$; then p is a continuous submultiplicative norm on A . Let A_p be the completion of $(A; p)$; then A_p is the algebra of all those continuous functions on K that are holomorphic on $\text{int } K$. Elementary complex-variable theory (considering the Taylor series of f about a) shows that $\bigcap_{n \geq 1} A_p f^n = 0$. By Corollary 2 (since f is certainly not nilpotent), it follows that f does not have FCD in the Banach algebra A_p , and so f does not have LFCD in A .

3. Let U be an open subset of \mathbb{C}^n , $A = C(U)$ in its standard Fréchet topology. Then we claim that every element of A has FCD.

Indeed, for every compact $K \subset U$, $C(U)|K = C(K)$. The closed ideals of the Banach algebra $C(K)$ (in the uniform norm $\|\cdot\|_K$) are well known to correspond precisely to the closed subsets of K . In particular, for each $f \in A$ and each compact $K \subset U$, $f|K$ and $f^2|K$ generate the same closed ideal of $C(K)$. It follows that $\delta(f) \leq 1$.

4. Let $A = L_{loc}^1(\mathbb{R}^+)$ (see [8], §7), with convolution product and Fréchet topology defined by the seminorms (p_n) , where for each $n \geq 1$ and each $f \in A$, $p_n(f) = \int_0^n |f(t)| dt$. Then the corresponding quotient Banach algebras are $A_n \cong L^1[0, n] \cong V$ (the Volterra algebra $L^1[0, 1]$). Let $0 \neq f \in A$ and write $\alpha(f) = \inf \text{supp } f$, so that $\alpha(f)$ is the least $\alpha \geq 0$ such that $f(t) = 0$ (a.e.) on $[0, \alpha]$. The standard theory of V shows that Af is dense in A if and only if $\alpha(f) = 0$. If $\alpha(f) > 0$ then, since (again by a standard result) $\alpha(f^m) = m\alpha(f)$, it follows that, though f does not have FCD, f is locally nilpotent, and so has LFCD. Thus, every element of $L_{loc}^1(\mathbb{R}^+)$ has LFCD. (A convenient reference for the relevant properties of V is [8], Example 7.8, Theorem 7.9.)

5. Let $\mathbf{w} = (w^{(k)})_{k \geq 1}$ be an increasing sequence of radical algebra-weights on \mathbb{R}^+ . Thus each w_k is (for simplicity) a continuous function, $w_k : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \setminus \{0\}$ with $w_k(x+y) \leq w_k(x)w_k(y)$ ($x, y \in \mathbb{R}^+$) and such that $w_k(x)^{1/x} \rightarrow 0$ as $x \rightarrow \infty$. Then, with standard notation, $L^1(\mathbb{R}^+; w_k)$ is a radical Banach algebra (under convolution product). We define $R = L^1(\mathbb{R}^+; \mathbf{w}) = \bigcap_{k \geq 1} L^1(\mathbb{R}^+; w_k)$. We topologize R by the sequence of norms $(\|\cdot\|_k)$, where $\|f\|_k = \int_0^\infty |f(t)|w_k(t) dt$. Evidently, R is a radical Fréchet algebra, and $R_k \cong L^1(\mathbb{R}^+; w_k)$.

Each weight is sufficiently rapidly decreasing so that the function u , defined by $u(t) = 1$ ($t \geq 0$), is in each R_k , and it is well known that $R_k u$ is norm-dense in R_k . It follows that $\overline{R}u = R$, so that u has FCD in R . If $\alpha(f) > 0$ (with the notation of Example 4) then, for each k , $\bigcap_{n \geq 1} R_k f^n = 0$, so that f does not even have LFCD. For a general f with $\alpha(f) = 0$, we do not know whether Rf need be dense in R , because this problem is still open (for general radical weights) for the Banach algebras R_k . For the Fréchet-algebra case there seems to be the possibility that a given sequence of “bad” weights could be equivalent (in the sense of giving the same Fréchet topology) to a sequence of “good” weights. (For a discussion of the Banach algebras $L^1(\mathbb{R}^+; w)$ see e.g. [8], §7.)

6. Let $A = C^\infty(\mathbb{R}^+)$, with convolution product. We may define the standard Fréchet topology on A by the seminorms

$$p_k(f) = \sup_{0 \leq t \leq k} k \sum_{r=0}^k |f^{(r)}(t)|,$$

for all $f \in A$ and $k \geq 1$. (The factor k before the sum ensures that the (p_k)

form an increasing sequence of submultiplicative seminorms, relative to the convolution product.)

Again let $u(t) = 1$ ($t \geq 0$). Consideration of the order of vanishing of u^k (convolution power: $u^k(t) = t^{k-1}/(k-1)!$) at 0 shows that Au^{k+1} is not p_m -dense in Ru^k for any $m \geq k-1$; but it is p_{k-2} -dense. Thus u does not have FCD, but it does have LFCD (but is clearly not locally nilpotent).

Probably the definition of an element of LFCD looks, at first sight, highly artificial. We claim that, in fact, it is very natural, citing in evidence Theorems 4 and 7 below; also Theorem 6, which shows that, contrary to first appearances, the property of having LFCD is actually an algebraic property.

We conclude this section with some lemmas about seminorms on \mathcal{F} that will be needed later.

LEMMA 4. *Let p be a submultiplicative seminorm on \mathcal{F} . Then X has FCD relative to the p -topology.*

Proof. Either p is a norm, in which case the result follows from [1], Theorem 1.

Or p is a proper seminorm, i.e. $\ker p \neq 0$, so that $\ker p = \mathcal{F}X^m$ for some $m \geq 0$. In particular, $p(X^m) = 0$, so $\mathcal{F}X^{m+1}$ is p -dense in $\mathcal{F}X^m$.

LEMMA 5. *Let A be a commutative Fréchet algebra and let $x \in A$. Suppose that there exists a homomorphism $\Psi : \mathcal{F} \rightarrow A/I(x)$ such that $\Psi(X) = q_x(x)$. Then x has LFCD.*

Proof. Let $A = \varprojlim(A_n; d_n)$, in standard notation. Then $\pi_n : A \rightarrow A_n$ maps $I(x)$ into $I(x_n)$ and so induces a homomorphism $\tilde{\pi}_n : A/I(x) \rightarrow A_n/I(x_n)$. Then $\Psi_n = \tilde{\pi}_n \circ \Psi$ is a homomorphism from \mathcal{F} into $A_n/I(x_n)$ such that $\Psi_n(X) = q_n(x_n)$ (where $q_n : A_n \rightarrow A_n/I(x_n)$ is the quotient map). By Proposition 2, it follows that x_n has FCD in A_n . This holds for all n , so that x has LFCD in A .

COROLLARY 3. *If $\Theta : \mathcal{F} \rightarrow A$ is a homomorphism into the commutative Fréchet algebra A , then $\Theta(X)$ has LFCD in A .*

Proof. Let $x = \Theta(X)$, and note that $\Psi = q_x \circ \Theta$ is a homomorphism from \mathcal{F} to $A/I(x)$ such that $\Psi(X) = q_x(x)$, and apply the lemma.

Note that by a proper seminorm we mean a seminorm that is not a norm.

PROPOSITION 4. *Let $(q_n)_{n \geq 1}$ be an increasing sequence of proper submultiplicative seminorms on \mathcal{F} that separates the points of \mathcal{F} . Then the topology defined by the (q_n) is the Fréchet topology κ of \mathcal{F} .*

Proof. By assumption, for each n , $\ker q_n \neq 0$, so there is an integer $m(n) \geq 1$ such that $\ker q_n = \mathcal{F}X^{m(n)}$. Since (q_n) is increasing, the sequence

of integers $m(n)$ is also increasing; moreover, $m(n) \rightarrow \infty$ as $n \rightarrow \infty$, since the sequence (q_n) separates points.

Then, on the finite-dimensional subspace of polynomials with degree not exceeding $m(n) - 1$, the restriction of q_n is a norm that is equivalent, on this subspace, to the restriction of the standard seminorm $p_{m(n)-1}$ (see Example 1 following Theorem 3). But $\ker q_n = \ker p_{m(n)-1} = \mathcal{F}X^{m(n)}$, so that q_n and $p_{m(n)-1}$ are equivalent seminorms on \mathcal{F} . Hence the sequence of seminorms (q_n) defines the same topology on \mathcal{F} as the sequence $(p_{m(n)-1})$. But $m(n) \rightarrow \infty$, so this is just the unique Fréchet topology κ of \mathcal{F} .

3. Elements of locally finite closed descent. It will be shown that those properties of elements of FCD in a Banach algebra that were needed for the proofs of Theorem 2 and Proposition 1, may be extended to elements of LFCD in a commutative Fréchet algebra. First we must make some further study of elements of FCD in a Banach algebra.

Let A be a commutative Banach algebra and let $x \in A$ have FCD. We saw in Corollary 1 that $I(x) = \bigcap_{n \geq 1} Ax^n$ is naturally isomorphic to the inverse limit of the sequence

$$A \xleftarrow{L_x} A \xleftarrow{L_x} A \xleftarrow{L_x} \dots$$

Thus, although $I(x)$ is not, in general, closed in A , it carries a Fréchet topology, say τ_x , as an inverse limit of Banach spaces. Recalling (Lemma 2) that $L_x|I(x)$ maps $I(x)$ bijectively onto itself, and writing L_x^{-1} for the inverse bijection, we may describe τ_x explicitly by saying that, for a sequence (u_n) in $I(x)$ and element $u \in I(x)$, we have $u_n \rightarrow u$ in the topology τ_x if and only if $L_x^{-r}(u_n) \rightarrow L_x^{-r}(u)$ in norm, for every $r \geq 0$, as $n \rightarrow \infty$. In particular, the topology τ_x is stronger than the norm topology restricted to $I(x)$.

LEMMA 6. *Let $T : A \rightarrow B$ be a continuous homomorphism of commutative Banach algebras. Let $x \in A$ have FCD, and let $y = T(x)$. Then:*

- (i) *y has FCD in B and $T(I(x)) \subseteq I(y)$;*
- (ii) *$T|I(x) : I(x) \rightarrow I(y)$ is continuous for the Fréchet topologies τ_x, τ_y ;*
- (iii) *if $T(A)$ is norm-dense in B , then $T(I(x))$ is τ_y -dense in $I(y)$.*

Proof. (i) is trivial.

(ii) Let $r \geq 0$ and let $u \in I(x)$. Then $L_y^r(T(L_x^{-r}u)) = T(u)$, so that $T(L_x^{-r}u) = L_y^{-r}T(u)$, from which the continuity statement is clear. (We remark that this continuity may also be deduced from the closed graph theorem.)

(iii) Let $T(A)$ be norm-dense in B . We first show that $T(I(x))$ is norm-dense in $I(y)$. Suppose that $\delta(x) = m$; then, by Corollary 2, $I(x)$ is dense in Ax^m . Then, since $T(A)$ is dense in B , $T(I(x))$ is dense in By^m . But $T(I(x)) \subseteq I(y) \subseteq By^m$, so also $T(I(x))$ is norm-dense in $I(y)$.

But then, for any $v \in I(y)$, $r \geq 0$ and $\varepsilon > 0$, there is some $u \in I(x)$ such that $\|T(u) - L_y^{-r}(v)\| < \varepsilon$. Let $w = x^r u \in I(x)$; then $\|L_y^{-r}(T(w) - u)\| < \varepsilon$. This proves the density statement.

Now let A be a commutative Fréchet algebra, with an Arens–Michael representation $A = \varprojlim(A_n; d_n)$. Let $x \in A$ be an element of LFCD. Then, in the standard notation (see the beginning of §2), $x_n = \pi_n(x)$ has FCD in A_n for each n . Then, for each n , $d_n : A_{n+1} \rightarrow A_n$ has dense range and $d_n(x_{n+1}) = x_n$; by Lemma 6, $d_n(I(x_{n+1})) \subseteq I(x_n)$ and $d_n|I(x_{n+1}) : I(x_{n+1}) \rightarrow I(x_n)$ is continuous with dense range for the Fréchet topologies on these ideals. Moreover, d_n induces a homomorphism, say $\tilde{d}_n : A_{n+1}/I(x_{n+1}) \rightarrow A_n/I(x_n)$. We represent these mappings, together with the canonical inclusions $j_n : I(x_n) \rightarrow A_n$ and quotient maps $q_n : A_n \rightarrow A_n/I(x_n)$, in a commutative diagram; we write $\bar{d}_n = d_n|I(x_{n+1})$:

$$\begin{array}{ccccccc} 0 & & 0 & & 0 & & \dots \\ \downarrow & & \downarrow & & \downarrow & & \\ I(x_1) & \xleftarrow{\bar{d}_1} & I(x_2) & \xleftarrow{\bar{d}_2} & I(x_3) & \xleftarrow{\bar{d}_3} & \dots \\ \downarrow j_1 & & \downarrow j_2 & & \downarrow j_3 & & \\ A_1 & \xleftarrow{d_1} & A_2 & \xleftarrow{d_2} & A_3 & \xleftarrow{d_3} & \dots \\ \downarrow q_1 & & \downarrow q_2 & & \downarrow q_3 & & \\ A_1/I(x_1) & \xleftarrow{\tilde{d}_1} & A_2/I(x_2) & \xleftarrow{\tilde{d}_2} & A_3/I(x_3) & \xleftarrow{\tilde{d}_3} & \dots \\ \downarrow & & \downarrow & & \downarrow & & \\ 0 & & 0 & & 0 & & \end{array}$$

Each column is a short exact sequence of complex algebras. The middle row is an inverse-limit sequence giving an Arens–Michael representation of A . The maps in the top row are continuous, with dense range, for the respective Fréchet topologies, as explained above. The bottom row is just a sequence of complex algebras and homomorphisms, with no given topologies.

The following is the vital technical lemma concerning elements with LFCD.

LEMMA 7. *Let x be an element of LFCD in the commutative Fréchet algebra A . Then the Arens–Michael isomorphism $A \cong \varprojlim(A_n; d_n)$ induces isomorphisms:*

- (i) $I(x) \cong \varprojlim(I(x_n); \tilde{d}_n)$;
- (ii) $A/I(x) \cong \varprojlim(A_n/I(x_n); \tilde{d}_n)$.

Proof. (i) It is clear that the Arens–Michael isomorphism $A \cong \varprojlim(A_n; d_n)$, $u \mapsto (u_n)_{n \geq 1}$ (where $u_n = \pi_n(u)$), maps $I(x)$ injectively into $\varprojlim(I(x_n); \tilde{d}_n)$ (where $\tilde{d}_n = d_n|I(x_{n+1})$). The main point to be proved is that this mapping is *onto* the inverse limit.

Thus, let $u_n \in I(x_n)$ ($n \geq 1$), with $u_n = d_n(u_{n+1})$ for each n . Then the Arens–Michael isomorphism gives a unique $u \in A$ such that $\pi_n(u) = u_n$ ($n \geq 1$); we must prove that $u \in I(x)$.

By Lemma 2, for each $k \geq 1$, $n \geq 1$, there is a unique $v_{n,k} \in I(x_n)$ such that $x_n^k v_{n,k} = u_n$. But $u_n = d_n(u_{n+1}) = d_n(x_{n+1}^k v_{n+1,k}) = x_n^k d_n(v_{n+1,k})$. Now $d_n(v_{n+1,k}) \in d_n(I(x_{n+1})) \subseteq I(x_n)$; by the uniqueness, $d_n(v_{n+1,k}) = v_{n,k}$ ($n \geq 1$). Hence, by the Arens–Michael isomorphism, there is a unique element $v_k \in A$ with $\pi_n(v_k) = v_{n,k}$ ($n \geq 1$). But then, for all n , $\pi_n(x^k v_k) = x_n^k v_{n,k} = u_n$, so that $x^k v_k = u$. Thus $u \in Ax^k$ for all $k \geq 1$, i.e. $u \in I(x)$, as was to be proved.

(ii) As in (i), there is a homomorphism, say $T : A \rightarrow \varprojlim(A_n/I(x_n); \tilde{d}_n)$, namely $T(u) = (q_n(u_n))_{n \geq 1}$ (where q_n is the quotient map $A_n \rightarrow A_n/I(x_n)$). Then $u \in \ker T$ if and only if $u_n \in I(x_n)$ for every n , i.e. if and only if $u \in I(x)$, by part (i). We thus have a naturally induced injective homomorphism, say $\tilde{T} : A/I(x) \rightarrow \varprojlim(A_n/I(x_n); \tilde{d}_n)$; again, the problem is to show that \tilde{T} maps onto this inverse limit.

Thus let, say, $\xi_n \in A_n/I(x_n)$ with $\tilde{d}_n(\xi_{n+1}) = \xi_n$ for all $n \geq 1$. For each n take $a_n \in A_n$ such that $q_n(a_n) = \xi_n$. Then, for each n , $q_n d_n(a_{n+1}) = \tilde{d}_n q_{n+1}(a_{n+1}) = \tilde{d}_n(\xi_{n+1}) = \xi_n = q_n(a_n)$, so that $a_n - d_n(a_{n+1}) \in I(x_n)$ for each n . We now seek to modify the sequence a_n to a sequence b_n such that $b_n - a_n \in I(x_n)$ and $d_n(b_{n+1}) = b_n$ for each n . Thus we want to find, say $z_n \in I(x_n)$ ($n \geq 1$), such that $a_n + z_n = d_n(a_{n+1} + z_{n+1})$. Therefore, for each n we define the mapping $f_n : I(x_{n+1}) \rightarrow I(x_n)$ by setting

$$f_n(z) = d_n(z) + d_n(a_{n+1}) - a_n \quad (n \geq 1).$$

Note that, since $d_n(I(x_{n+1})) \subseteq I(x_n)$ and $a_n - d_n(a_{n+1}) \in I(x_n)$ for each n , the mapping f_n does map $I(x_{n+1}) \rightarrow I(x_n)$. Also, by Lemma 6, f_n is continuous with dense range for the Fréchet topologies on $I(x_{n+1})$ and $I(x_n)$. Hence, by the Mittag-Leffler theorem (Theorem 1), there is a sequence $(z_n)_{n \geq 1}$ such that $z_n \in I(x_n)$ and $a_n + z_n = d_n(a_{n+1} + z_{n+1})$ for all $n \geq 1$. We then set $b_n = a_n + z_n$; so, for each n , $q_n(b_n) = q_n(a_n) = \xi_n$ and also $(b_n) \in \varprojlim(A_n; d_n)$. There is thus a unique $b \in A$ such that $\pi_n(b) = b_n$ for all n , and thus $T(b) = (\xi_n)_{n \geq 1}$. This completes the proof.

COROLLARY 4. Let A be a commutative Fréchet algebra, and let $x \in A$ have LFCD. Then:

- (i) L_x maps $I(x)$ bijectively onto itself;
- (ii) $I(x) = 0$ if and only if x is locally nilpotent.

Proof. (i) Let $u \in I(x)$ and suppose that $L_x(u) = 0$. Then, with our standard notation, $L_{x_n}(u_n) = 0$ (for all n). By Lemma 2, $u_n = 0$ (for all n) and so $u = 0$. Thus $L_x|I(x)$ is injective.

Now let $v \in I(x)$; so $v_n \in I(x_n)$ for all n and, again by Lemma 2, for each n , there is a unique element $u_n \in I(x_n)$ such that $L_{x_n}(u_n) = v_n$. From the uniqueness statement, it follows that $u_n = d_n(u_{n+1})$ ($n \geq 1$). Hence, by Lemma 7(i), there is a unique $u \in I(x)$ such that $\pi_n(u) = u_n$ for each n . It is immediate that $L_x(u) = v$, which completes the proof of (i).

(ii) If x is locally nilpotent, then each x_n is nilpotent, so $I(x_n) = 0$ for all n , and thus $I(x) = 0$, by Lemma 7(i).

Conversely, let $I(x) = 0$. By the Mittag-Leffler theorem, applied to $I(x) = \varprojlim(I(x_n); \tilde{d}_n)$, we deduce that $\pi_n(I(x))$ is dense in $I(x_n)$ for the Fréchet topology, for each n . So, for each n , $I(x_n) = 0$, and thus x_n is nilpotent, by Corollary 2. Hence x is locally nilpotent.

THEOREM 4. Let A be a commutative, unital Fréchet algebra and let $x \in A$. The following are equivalent:

- (i) there is a unital homomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ such that $\Psi_x(X) = q_x(x)$;
- (ii) x has LFCD.

Moreover, in case (i) and (ii) hold, the homomorphism Ψ_x is uniquely determined; it is injective if and only if, for all m , $x^m \notin Ax^{m+1}$.

Proof. (i) \Rightarrow (ii). This is Lemma 5.

(ii) \Rightarrow (i). Let $x \in A$ have LFCD. Then, for each n , x_n has FCD. By Proposition 1, there is a unique unital homomorphism $\Psi_n : \mathcal{F} \rightarrow A_n/I(x_n)$ such that $\Psi_n(X) = q_n(x_n)$. The uniqueness property then implies that, with the notation of Lemma 7, $\Psi_n = \tilde{d}_n \circ \Psi_{n+1}$ ($n \geq 1$). There is then a unique homomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ such that $\tilde{T} \circ \Psi_x(f) = (\Psi_n(f))_{n \geq 1}$ ($f \in \mathcal{F}$), where $\tilde{T} : A/I(x) \rightarrow \varprojlim(A_n/I(x_n); \tilde{d}_n)$ is the isomorphism given by Lemma 7(ii). In particular, $\tilde{T}\Psi_x(X) = (\Psi_n(X))_{n \geq 1} = (q_n(x_n))_{n \geq 1} = \tilde{T}(q_x(x))$, i.e. $\Psi_x(X) = q_x(x)$.

Now suppose that Ψ_x is not injective. Then $\ker \Psi_x$ is a non-zero ideal of \mathcal{F} , so, for some $m \geq 1$, $q_x(x^m) = \Psi_x(X^m) = 0$. It follows that $x^m \in I(x) \subseteq Ax^{m+1}$, i.e. $x^m \in Ax^{m+1}$ for some $m \geq 1$.

Conversely, suppose that $x^m \in Ax^{m+1}$ for some $m \geq 1$. Then $Ax^m = Ax^{m+1}$, so $Ax^m = Ax^n$ for all $n \geq m$ and $I(x) = Ax^m$. But then $X^m \in \ker \Psi_x$, so Ψ_x is not injective.

COROLLARY 5. If $x \in \text{rad } A$ and x has LFCD, then Ψ_x is injective if and only if x is not nilpotent.

Proof. Clearly, if x is nilpotent then Ψ_x is not injective. Conversely, if Ψ_x is not injective then, by Theorem 4, $x^m \in Ax^{m+1}$ for some $m \geq 1$. Thus $x^m = ax^{m+1}$ for some $a \in A$, i.e. $x^m(1 - ax) = 0$, so $x^m = 0$, since $1 - ax$ is invertible in A , because $x \in \text{rad } A$.

THEOREM 5. Let x be a locally nilpotent, non-nilpotent element of a commutative unital Fréchet algebra A . Then there is a unique homomorphism $\Theta_x : \mathcal{F} \rightarrow A$ such that $\Theta_x(X) = x$. Moreover, Θ_x is injective, $\text{im } \Theta_x$ is a closed subalgebra of A and $\Theta_x : \mathcal{F} \rightarrow \text{im } \Theta_x$ is an isomorphism of Fréchet algebras.

Proof. Since x is locally nilpotent, $I(x) = 0$ (by Corollary 4(ii)), so that the mapping $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ of Theorem 4 becomes a (unique) homomorphism $\Theta_x : \mathcal{F} \rightarrow A$ such that $\Theta_x(X) = x$. Since x is non-nilpotent, Θ_x is injective.

Let $A_0 = \text{im } \Theta_x$; if (p_n) is an increasing sequence of seminorms defining the topology of A , let $q_n = p_n|_{A_0}$ ($n \geq 1$). If, for some n , q_n were a norm, then the locally nilpotent element x would be actually nilpotent—which is not allowed. Thus, each q_n is a proper seminorm on A_0 and so, by Proposition 4, the seminorms $(q_n \circ \Theta_x)$ on \mathcal{F} define the standard Fréchet topology κ of \mathcal{F} . The result follows.

As a corollary, we have the following curious characterization of \mathcal{F} as a Fréchet algebra.

COROLLARY 6. Let A be a unital Fréchet algebra. Then A is isomorphic to \mathcal{F} if and only if it is generated, as a Fréchet algebra, by some element that is locally nilpotent but not nilpotent.

Proof. If the Fréchet algebra A is generated by the locally nilpotent, non-nilpotent element x , then A is commutative and the homomorphism Θ_x of Theorem 5 has $\text{im } \Theta_x = A$.

Theorem 4 has a consequence for the theory of automatic continuity. (The result is an extension of Theorem 1 of [2].) It has the surprising consequence that, for an element x of a commutative Fréchet algebra to have LFCD is, in fact, an *algebraic* property. This last remark does, of course, follow at once from Theorem 4, but in fact we have the following:

THEOREM 6. Let A and B be commutative Fréchet algebras and let $T : A \rightarrow B$ be a homomorphism, not necessarily continuous. Let $x \in A$ have LFCD. Then $T(x)$ has LFCD.

Proof. Without loss of generality, we may assume that A and B are unital.

Let $y = T(x)$. Then $T(I(x)) \subseteq I(y)$, so there is a homomorphism $\tilde{T} : A/I(x) \rightarrow B/I(y)$ such that $\tilde{T}q_x = q_yT$ and, in particular, $\tilde{T}(q_x(x)) = q_y(y)$.

If x has LFCD then, by Theorem 4, there is a unital homomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ such that $\Psi_x(X) = q_x(x)$. But then $\tilde{T}\Psi_x : \mathcal{F} \rightarrow B/I(y)$ maps X to $q_y(y)$ so, by the reverse implication of Theorem 4, y has LFCD in B .

Remark. We note in particular that if x is locally nilpotent in A , then $T(x)$ has LFCD in B . If T were continuous then $T(x)$ would, of course, also be locally nilpotent, but this stronger consequence does not generally hold when T is merely a homomorphism. This follows from Theorem 2 of [1], since there exists, for example, an injective homomorphism, say θ , from $\mathcal{F} \rightarrow V_+$ (the unitization of the Volterra algebra). Then $\theta(X)$ is not nilpotent, hence, since V_+ is a Banach algebra, it is not locally nilpotent. But X is a locally nilpotent element of \mathcal{F} .

The following proposition may have some interest in relation to the still unsolved “Michael problem”, which is to determine whether every character on a (commutative) Fréchet algebra need be continuous.

PROPOSITION 5. Let x be an element of LFCD in a commutative Fréchet algebra A . Then $q_x(x) \in \text{rad } A/I(x)$. In particular, if φ is a character on A , continuous or not, such that $\ker \varphi \supseteq I(x)$, then $\varphi(x) = 0$.

Proof. By Theorem 4, there is a homomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ such that $\Psi_x(X) = q_x(x)$. Since $X \in \text{rad } \mathcal{F}$, it follows that $q_x(x) \in \text{rad } A/I(x)$.

If φ is a character on A such that $\ker \varphi \supseteq I(x)$, then there is a unique homomorphism $\varphi_0 : A/I(x) \rightarrow \mathbb{C}$ such that $\varphi = \varphi_0 q_x$. Since $q_x(x) \in \text{rad } A/I(x)$, it follows that $\varphi(x) = \varphi_0 q_x(x) = 0$.

4. Embedding \mathcal{F} in Fréchet algebras. We now turn to the problem of characterizing those commutative Fréchet algebras in which \mathcal{F} may be embedded. This is to generalize Theorems 1 and 2 of [1]. In the earlier paper, it was not initially clear that \mathcal{F} could be embedded in *any* Banach algebra, whereas even \mathcal{F} itself is already a Fréchet algebra. Nevertheless, the solution to the problem of describing *all* those commutative Fréchet algebras in which \mathcal{F} may be embedded does include the earlier result as a special case.

THEOREM 7. Let A be a commutative Fréchet algebra and let $x \in A$. The following are equivalent:

- (i) there is some unital, injective homomorphism $\Theta_x : \mathcal{F} \rightarrow A$ such that $\Theta_x(X) = x$;
- (ii) $x \in \text{rad } A$, x has LFCD, but x is not nilpotent.

Moreover, in case the equivalent conditions (i) and (ii) hold, then Θ_x is unique if and only if x is locally nilpotent (i.e. if and only if $I(x) = 0$). In general, for any $f \in \mathcal{F}$ that is transcendental over $\mathbb{C}[X]$, we may define Θ_x so that $\Theta_x(f)$ is any chosen element of the coset $q_x^{-1}(\Psi_x(f))$ of $I(x)$.

(Here, $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ is the unique homomorphism such that $\Psi_x(X) = q_x(x)$, given by Theorem 4; by Corollary 5, Ψ_x is injective.)

P r o o f. From Theorem 4, we already know that (i) implies (ii) (the deduction that $x \in \text{rad } A$ being immediate). Moreover, towards proving that (ii) implies (i), Corollary 5 shows that there is a unique injective homomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ such that $\Psi_x(X) = q_x(x)$. It follows that, if the homomorphism Θ_x of the present theorem exists, then it must satisfy $q_x\Theta_x = \Psi_x$, i.e. it must be a lift of Ψ_x .

Also, the case when x is locally nilpotent (and, by Proposition 3, this is precisely the case $I(x) = 0$) is covered by Theorem 5, including the statement that uniqueness of Θ_x follows from the local nilpotence of x .

It thus remains to prove that, when x has LFCD but is not locally nilpotent, then the monomorphism $\Psi_x : \mathcal{F} \rightarrow A/I(x)$ may be lifted to a homomorphism $\Theta_x : \mathcal{F} \rightarrow A$ such that $\Theta_x(X) = x$. (There are also the final remarks on the extent of non-uniqueness to be proved.) We thus have to extend to Fréchet algebras the result of [1], Theorem 2, but now with the assumption that x has LFCD, rather than FCD. Fortunately, we are able to make use of some of the lemmas from the earlier paper.

We start, necessarily, with the homomorphism, say θ_0 , defined on $\mathbb{C}[X] \subset \mathcal{F}$ such that $\theta_0(X) = x$; clearly $q_x\theta_0 = \Psi_x|\mathbb{C}[X]$.

Now suppose, more generally, that we have a unital homomorphism $\theta_0 : \mathcal{F}_0 \rightarrow A$, defined on some subalgebra \mathcal{F}_0 of \mathcal{F} , $\mathcal{F}_0 \supseteq \mathbb{C}[X]$, such that both $\theta_0(X) = x$ and $q_x\theta_0 = \Psi_x|\mathcal{F}_0$. Notice that, since Ψ_x is injective (see above), θ_0 must also be injective. (This point is important since, although the ideal structure of \mathcal{F} is very simple, that of \mathcal{F}_0 may be much richer.) The idea now is to show that, if $\mathcal{F}_0 \neq \mathcal{F}$, then θ_0 has a proper extension, say $\theta_1 : \mathcal{F}_1 \rightarrow A$, where the subalgebra $\mathcal{F}_1 \supsetneq \mathcal{F}_0$ and $q_x\theta_1 = \Psi_x|\mathcal{F}_1$. A standard application of Zorn's lemma will then complete the proof.

Thus, given $\theta_0 : \mathcal{F}_0 \rightarrow A$, as in the last paragraph, with $\mathcal{F}_0 \neq \mathcal{F}$, let $f \in \mathcal{F} \setminus \mathcal{F}_0$.

Case 1: f is transcendental over \mathcal{F}_0 , i.e. if, for any polynomial $P(Y) \in \mathcal{F}_0[Y]$, with coefficients from \mathcal{F}_0 , we define $P(f)$ in the natural way, then the mapping $P(Y) \mapsto P(f)$ is an algebra-isomorphism from $\mathcal{F}_0[Y]$ onto $\text{alg}(\mathcal{F}_0, f)$, the subalgebra of \mathcal{F} generated by \mathcal{F}_0 and f . In this case we may choose any $a \in A$ such that $q_x(a) = \Psi_x(f)$, and it is then elementary that defining $\theta_1(P(f)) = P(a)$ (for all $P(Y) \in \mathcal{F}_0[Y]$) gives an extension $\theta_1 : \text{alg}(\mathcal{F}_0, f) \rightarrow A$ of θ_0 that satisfies $\theta_1(f) = a$ and $q_x\theta_1 = \Psi_x|\text{alg}(\mathcal{F}_0, f)$.

Case 2: f is algebraic over \mathcal{F}_0 , i.e. for some integer $N \geq 1$ and elements g_0, \dots, g_N of \mathcal{F}_0 , with $g_N \neq 0$, we have $P(f) = 0$, where $P(Y) = g_0 + g_1Y + \dots + g_NY^N \in \mathcal{F}_0[Y]$. We now make use of Lemmas 4 and 5 of [1]. For each $n \geq 1$ we have the homomorphism, say, $\theta_n = \pi_n\theta_0 : \mathcal{F}_0 \rightarrow A_n$ such that $\theta_n(X) = \pi_n(x) = x_n$ and $q_n\theta_n = \Psi_n|\mathcal{F}_0$ (where $q_n : A_n \rightarrow A_n/I(x_n)$ is a quotient mapping, and $\Psi_n : \mathcal{F} \rightarrow A_n/I(x_n)$ is the unique

unital homomorphism such that $\Psi_n(X) = q_n(x_n)$). Then, by Lemmas 4 and 5 of [1], for each n , θ_n has a unique extension, say, $\theta'_n : \text{alg}(\mathcal{F}_0, f) \rightarrow A_n$ such that $q_n\theta'_n = \Psi_n|\text{alg}(\mathcal{F}_0, f)$. The uniqueness statement then means that, for each $n \geq 1$, we have $\theta'_n = d_n\theta'_{n+1}$. Hence, using the Arens–Michael isomorphism, there is a unique homomorphism $\theta : \text{alg}(\mathcal{F}_0, f) \rightarrow A$ such that $q_x\theta = \Psi_x|\text{alg}(\mathcal{F}_0, f)$, given by $\theta(g) = (\theta'_n(g))_{n \geq 1} \in \varprojlim(A_n; d_n) \cong A$. This proves that an algebraic extension step may be carried out uniquely.

A standard application of Zorn's lemma completes the construction.

The remarks on the extent of non-uniqueness are also clear. For, if $f \in \mathcal{F}$ is transcendental over the polynomial algebra $\mathbb{C}[X]$, we may make the extension of θ_0 from $\mathbb{C}[X]$ to $\text{alg}(\mathbb{C}[X], f)$ the first step of the construction. This step then comes under Case 1 above, so that we may choose $\Theta_x(f)$ to be any element $a \in A$ such that $q_x(a) = \Psi_x(f)$. (Since $\ker q_x = I(x) \neq 0$, the choice is not unique.) Theorem 7 is proved.

There is a further consequence in automatic continuity, which extends the theorem of [3].

THEOREM 8. *Let A, B be commutative unital Fréchet algebras. Suppose that:*

- (i) *A has a point derivation of infinite order, $(d_n)_{n \geq 0}$ (at some continuous character d_0), with $d_1 \neq 0$;*
- (ii) *$\text{rad } B$ contains an element with LFCD that is not locally nilpotent.*

Then there is a discontinuous homomorphism from A to B .

P r o o f. Define $\theta_0 : A \rightarrow \mathcal{F}$ by $\theta_0(x) = \sum_{n \geq 0} d_n(x)X^n$ ($x \in A$). Then θ_0 is a homomorphism.

Since $d_1(1) = 0$ and $d_1 \neq 0$, there is some $a \in A$ with $d_0(a) = 0$, $d_1(a) \neq 0$. Then $\theta_0(a)$ is a formal power series of order 1, so there is a unique automorphism α of \mathcal{F} with $X = (\alpha\theta_0)(a)$. Let $\theta = \alpha\theta_0$; then $\theta : A \rightarrow \mathcal{F}$ is a homomorphism with $\theta(a) = X$.

For any polynomial p , clearly $\theta(p(a)) = p(X)$. But also $\theta(e^a) = e^X$ namely, if $c_n(X) = \sum_{k=0}^n X^k/k!$ ($n = 0, 1, \dots$), then, for each n , $e^a - c_n(a) \in Aa^{n+1}$, so that $\theta(e^a) - c_n(X) \in \mathcal{F}X^{n+1}$, and $\theta(e^a) - e^X \in \bigcap_{n \geq 1} \mathcal{F}X^n = 0$, i.e. $\theta(e^a) = e^X$.

As an element of \mathcal{F} , the series e^X is transcendental over $\mathbb{C}[X]$ (exercise). So if $b \in \text{rad } B$, where b is not locally nilpotent but has LFCD, then, by Theorem 7, there is a homomorphism $\Psi : \mathcal{F} \rightarrow B$ such that $\Psi(X) = b$ but $\Psi(e^X) \neq e^b$. Then the homomorphism $T = \Psi \circ \theta : A \rightarrow B$ has $T(a) = b$ but $T(e^a) = \Psi(e^X) \neq e^b$; in particular, T is discontinuous.

EXAMPLE. Let $A = \mathcal{O}(\mathbb{C})$, the algebra of entire functions in one variable (which even has an infinite-order point derivation at every character), and let B be the unitization of any of the algebras in Examples 4, 5, 6 following

Proposition 3. Then Theorem 8 gives a discontinuous homomorphism from A to B .

References

- [1] G. R. Allan, *Embedding the algebra of formal power series in a Banach algebra*, Proc. London Math. Soc. (3) 25 (1972), 329–340.
- [2] —, *Elements of finite closed descent in a Banach algebra*, J. London Math. Soc. (2) 7 (1973), 462–466.
- [3] —, *A remark in automatic continuity theory*, Bull. London Math. Soc. 12 (1980), 452–454.
- [4] R. F. Arens, *Linear topological division algebras*, Bull. Amer. Math. Soc. 53 (1947), 632–630.
- [5] —, *A generalization of normed rings*, Pacific J. Math. 2 (1952), 455–471.
- [6] —, *Dense inverse limit rings*, Michigan Math. J. 5 (1958), 169–182.
- [7] H. Cartan, *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes*, Hermann, Paris, 1963.
- [8] H. G. Dales, *Automatic continuity: a survey*, Bull. London Math. Soc. 10 (1978), 129–183.
- [9] J. Esterle, *Elements for a classification of commutative radical Banach algebras*, in: Radical Banach Algebras and Automatic Continuity, J. Bachar et al. (eds.), Lecture Notes in Math. 975, Springer, 1983, 4–65.
- [10] —, *Mittag-Leffler methods in the theory of Banach algebras and a new approach to Michael's problem*, in: Contemp. Math. 32, Amer. Math. Soc., 1984, 107–129.
- [11] E. A. Michael, *Locally multiplicatively-convex topological algebras*, Mem. Amer. Math. Soc. 11 (1953; third printing 1971).

Department of Pure Mathematics and Mathematical Statistics
 University of Cambridge
 16 Mill Lane
 Cambridge CB2 1SB
 United Kingdom
 E-mail: g.r.allan@pmms.cam.ac.uk

Received January 29, 1996
 Revised version March 6, 1996

(3601)

Multiplicative functionals and entire functions

by

KRZYSZTOF JAROSZ (Edwardsville, Ill., and Warszawa)

Abstract. Let \mathcal{A} be a complex Banach algebra with a unit e , let T , φ be continuous functionals, where T is linear, and let F be a nonlinear entire function. If $T \circ F = F \circ \varphi$ and $T(e) = 1$ then T is multiplicative.

1. Introduction. If T is a multiplicative functional on a complex Banach algebra \mathcal{A} with a unit e then $T(e) = 1$, and for any invertible element x of \mathcal{A} we have $T(x) \neq 0$. A. M. Gleason [5] and, independently, J. P. Kahane & W. Żelazko [7] proved that the above property characterizes multiplicative functionals. In fact, they proved even a stronger result:

THEOREM 1. *If T is a continuous linear functional on a complex unital Banach algebra \mathcal{A} such that $T(e) = 1$ and $T(\exp x) \neq 0$ for $x \in \mathcal{A}$, then T is multiplicative.*

The above statement can be rephrased in the following equivalent way.

THEOREM 2. *If T is a continuous linear functional on a complex unital Banach algebra \mathcal{A} with $T(e) = 1$, and there is a complex valued function φ on \mathcal{A} such that*

$$(1) \quad T(\exp x) = \exp(\varphi(x)) \quad \text{for } x \in \mathcal{A},$$

then T is multiplicative.

R. Arens [1] asked if the exponential function in (1) can be replaced by any other entire function F , that is, whether

$$(2) \quad T \circ F = F \circ \varphi$$

1991 *Mathematics Subject Classification*: Primary 46J05; Secondary 46H05, 46H30, 46J15, 46J20.

Research was supported in part by a grant from the International Research & Exchanges Board, with funds provided by the National Endowment for the Humanities and the U.S. State Department.

Manley-Rowe

Some General Properties of Nonlinear Elements— Part I. General Energy Relations*

J. M. MANLEY†, SENIOR MEMBER, IRE, AND H. E. ROWE‡, ASSOCIATE, IRE

Summary—Two independent equations relating the average powers at the different frequencies in nonlinear inductors and capacitors are derived. These equations are a consequence only of the assumption of a single-valued characteristic for the nonlinear element. They are independent of the particular shape of this characteristic, of the power levels at the various frequencies, and of the external circuit in which the nonlinear reactor is connected.

These general energy relations give information regarding the gain and stability of nonlinear reactor modulators and demodulators, and consequently of magnetic and dielectric amplifiers, without requiring detailed information about these devices. The utility of these equations is illustrated by discussing the gain and stability of the simplest types of nonlinear reactor modulators and demodulators and harmonic generators.

This analysis for nonlinear inductors and capacitors is extended to include the effects of hysteresis in the nonlinear characteristic in the special case where the operating hysteresis loop is no more than double-valued.

A similar analysis applied to nonlinear resistors yields two equations relating the reactive powers at the different frequencies, rather than the real powers as above. The interpretation of reactive power under these conditions is not clear.

INTRODUCTION

ANY AMPLIFIER can be considered to be a modulator in that the input wave causes variations in the energy flowing from the amplifier's energy source. The modulator must deliver more power to an output load than the input wave delivers to the modulator if it is to provide useful gain.

If the energy source of a modulator is direct, for example a battery, the output wave is, ideally, a replica of the input wave. Examples are: the direct electron current in the plate-cathode path of a vacuum tube is varied by the field caused by applying a wave to the grid; the direct battery current in a carbon microphone is varied by mechanical motion of the diaphragm; the direct collector-emitter current in a transistor is varied by the base-emitter current, which is proportional to an applied wave. Depending on the nature of the modulator it may exhibit a gain or a loss.

On the other hand, if the energy source does not have zero frequency, but is an alternator, referred to as the local oscillator, the output is not a replica of the input, but occurs at a different frequency than the input, and consequently the device is called a modulator, demodulator, or a mixer. This shift in frequency occurs because new frequencies are generated when waves of two dif-

ferent frequencies, both different from zero, are applied to a nonlinear element. These frequencies are harmonics and sum and difference combinations of the applied frequencies. Again, some of these devices exhibit a modulating gain and some a loss, depending on their nature. Examples are: magnetic amplifiers,¹ dielectric amplifiers,² and copper oxide modulators.

Although modulators with ac energy sources are quite similar to the more conventional type of amplifier, whose energy is supplied by a dc source, their special properties are less widely known. The main purpose of Part I of this paper is to derive some general power relations which govern nonlinear reactor modulators; *i.e.*, modulators whose nonlinear element is a nonlinear inductor or capacitor. These relations consist of two independent equations relating the powers at the different frequencies in such an element; their only restriction is that the nonlinear characteristic, voltage-charge or flux-current for the nonlinear capacitor or inductor respectively, be single-valued. They do not depend on the detailed shape of this characteristic or on the external circuit in which the nonlinear element is connected. They are independent of the levels at the various frequencies; they are thus not confined to the small signal case, in which all signal levels are much lower than the local oscillator level, but hold true in general.

These results are useful in considering the gain and the stability of nonlinear reactor modulators and demodulators,³ and hence of magnetic and dielectric amplifiers, which may be regarded as carrier systems using a nonlinear reactor modulator. The question of stability is similar to that which arises with conventional feedback amplifiers; *i.e.*, modulators with a dc carrier, which may oscillate under certain operating conditions. With nonlinear reactor modulators, instability corresponds to the production of additional frequencies, which may or may not be the normal signal frequencies, with only the local oscillator frequency applied. In general these new frequencies are incommensurable with the local oscillator frequency and may be greater or less than this frequency; but in special cases

* J. M. Manley, "Some general properties of magnetic amplifiers," PROC. IRE, vol. 39, pp. 242-251; March, 1951.

† W. P. Mason and R. F. Wick, "Ferroelectrics and the dielectric amplifier," PROC. IRE, vol. 42, pp. 1606-1620; November, 1954.

‡ In the remainder of this paper the term *modulator* will be used to denote a device in which the output signal frequency is higher than the input signal frequency, while the term *demodulator* will denote a device in which the output signal frequency is lower than the input signal frequency, when specific devices are under consideration. The term *modulator* will also be used, as above, to denote the general class of devices in which new frequencies are produced.

* Original manuscript received by the IRE, January 31, 1956. A shortened version of this paper was presented at a symposium on nonlinear circuit analysis, sponsored by the Polytechnic Inst. of Brooklyn, N. Y., on April 25, 26, and 27, 1956.

† Bell Telephone Labs., Inc., Murray Hill, N. J.

‡ Bell Telephone Labs., Inc., Holmdel, N. J.

these new frequencies may be rationally related to the local oscillator frequency, yielding frequencies that are subharmonics or multiples of subharmonics.

The utility of these results is illustrated by applying them to the simplest nonlinear reactor modulators and demodulators, in which only one of the principal sidebands of the applied signal frequency about the local oscillator frequency is allowed to carry a significant amount of power. Under these conditions the powers at the various frequencies will be directly proportional to the frequencies. There will be two different situations to consider, depending on the relative positions of the local oscillator and the signal frequencies. In the stable case the maximum available gain is equal to the ratio of the output frequency to input frequency; consequently, modulators have gain while demodulators have loss. In the potentially unstable case the gain is unlimited, any gain from 0 to ∞ being attainable for either the modulator or the demodulator; however, the gain as a modulator exceeds the gain as a demodulator, as in the stable case.

The remainder of Part I discusses certain related results. The analysis is extended to include the effects of hysteresis in the characteristic of nonlinear reactors under somewhat more restricted conditions, which insure that the operating hysteresis loop is no more than double-valued. Finally, a similar analysis applied to nonlinear resistors yields analogous relations among the reactive powers at the various frequencies. The interpretation of reactive power in this situation is not clear, and these results seem to be less important than those of the nonlinear reactor analysis.

Part II of this paper (to be submitted later) entitled "Small-Signal Theory," is concerned with the application of the well-known small signal analysis⁴ to the determination of further properties of these devices.

HISTORICAL NOTE

To the best knowledge of the writers, the first calculations of the reactions of currents of different frequencies on one another in a reactance modulator was made by R. V. L. Hartley of the Bell Telephone Labs. in 1916. He assumed a simple cubic characteristic for a biased iron-cored inductance modulator, and assumed that the external circuit permitted currents at only four frequencies, the two applied frequencies and their sum and difference frequencies, to flow through the nonlinear inductance. There are two important results of his analysis:

1) The high and low frequency sources applied to the modulator supply power unequally, the ratio of these two powers being greater than the ratio of their corresponding frequencies. Thus, if one source has a much higher frequency than the other, it will supply most of

⁴ W. R. Bennett, "A general review of linear varying parameter and non-linear circuit analysis," *PROC. IRE*, vol. 38, pp. 259-263; March, 1950.

the power to the modulator, the low-frequency source supplying very little power.

2) The flow of power out of the modulator at the sum frequency absorbs power from both generators; *i.e.*, introduces a positive resistance into both source circuits. However, the flow of power at the difference frequency introduces a positive resistance into the high frequency source circuit, a negative resistance into the low frequency source circuit.

The first result shows that a modulating gain is possible with this type of modulator when the local oscillator frequency is appreciably higher than the signal frequency. The second result shows that under certain conditions instability or oscillation is possible, with power emerging from all of the signal terminals with only the local oscillator applied to the modulator. These output frequencies will, in general, be incommensurable with the applied carrier frequency, but in special cases may be rationally related, being subharmonics or multiples of subharmonics. This suggests that by operating close to instability, very high gains may be obtained as either a modulator or as a demodulator.

The first application of these results was to the coil modulators being studied then for possible use in carrier telephone systems. The experiments bore out the analysis in that good modulating gains could be obtained and that instabilities resulted when attempts were made to operate a demodulator with gain. They were also used to explain the generation of subharmonics and other nonharmonic frequencies,^{5,6} and the occurrence of gain and the source of occasional instability in magnetic amplifiers.¹ Later, Hartley analyzed a particular form of capacitance modulator and obtained similar results,⁷ which were verified experimentally.⁸ He suggested that similar processes could occur within atoms and thereby explain the Raman effect.^{9,6}

These results were derived by J. M. Manley under much more general conditions.¹⁰ His analysis assumes only that the characteristic of the nonlinear reactor is single-valued; *i.e.*, that there be no hysteresis; otherwise the shape of the nonlinear characteristic is arbitrary. The derivation includes an arbitrary number of sidebands and is valid for arbitrary power levels at the various frequencies.

This analysis is the principal topic of Part I of this paper. It is extended to include the effects of hysteresis in the nonlinear characteristic for the special case in which the operating hysteresis loop is no more than double-valued.

⁵ J. M. Manley and E. Peterson, "Negative resistance effects in saturable reactor circuits," *Trans. AIEE*, vol. 65, pp. 870-881; 1946.

⁶ E. Peterson, "Atomic Physics and Circuit Theory," *Bell Labs. Record*, vol. 7, p. 231; February, 1929.

⁷ R. V. L. Hartley, "Oscillations in systems with non-linear reactance," *B.S.T.J.*, vol. 15, pp. 424-440; July, 1936.

⁸ L. W. Hussey and L. R. Wrathall, "Oscillations in an electro-mechanical system," *B.S.T.J.*, vol. 15, pp. 441-445; July, 1936.

⁹ R. V. L. Hartley, "A wave mechanism of quantum phenomena," *Phys. Rev.*, vol. 33, p. 289; February, 1929.

¹⁰ Unpublished memoranda; 1933-1935.

GENERAL ENERGY RELATIONS IN NONLINEAR REACTORS

In this section we derive J. M. Manley's power relations for hysteresisless nonlinear capacitors and inductors,¹⁰ which relate the powers at the different frequencies in such an element. The analysis will be carried out for the nonlinear capacitor only, since the corresponding analysis for the nonlinear inductor is almost identical and yields identical final results.

The characteristic of the nonlinear capacitor is given by specifying the voltage as some arbitrary function of the charge:

$$v = f(q), \quad (1)$$

where q is the charge on the nonlinear capacitor and v is the voltage across it. In particular, for a linear capacitor (1) becomes

$$v = \frac{1}{C} \cdot q$$

where C is the capacitance. In the present section hysteresis is excluded and so $f(q)$ is a single-valued function, but otherwise its shape is arbitrary; there are no further restrictions on the analysis. In a later section hysteresis is considered in the special case for which $f(q)$ is at most a double-valued function. Note that the characteristic of the nonlinear capacitor could be given equally well by specifying the charge as some arbitrary but single-valued function of the voltage; the subsequent analysis would then be the dual of that presented below. In the corresponding nonlinear inductor analysis the flux and current must be similarly related by an arbitrary single-valued function.

In general, all of the frequencies $f_{mn} = mf_1 + nf_0$ will be present, where m and n take on all integral values, positive, negative, and zero. The two fundamental frequencies f_1 and f_0 are assumed to be incommensurable and positive. In later applications of these results f_1 and f_0 will be taken as the two applied frequencies, and the other frequencies as the resulting sidebands. However, this identification is not essential to the analysis, and others may be convenient.

For conceptual purposes, the nonlinear capacitor may be thought of as connected in the circuit of Fig. 1, in which each frequency flows in a separate external circuit. Each of these circuits contains in series a voltage generator of the appropriate frequency, a load impedance, and an ideal filter which presents a short circuit to the desired frequency and an open circuit to all other frequencies. The voltages of all generators in this equivalent circuit except those corresponding to the sources driving the nonlinear capacitor are, of course, set equal to zero. Thus, in the case discussed above where only two generators at the frequencies f_1 and f_0 are connected to the nonlinear capacitor, only these two branches in the equivalent circuit will contain voltage generators, all other branches containing passive load

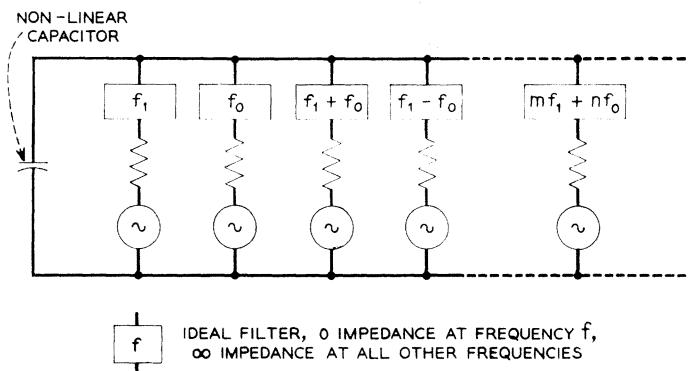


Fig. 1—Illustrative circuit with a nonlinear capacitor.

impedances. This circuit is only for purposes of illustration, and does not enter into the analysis in any way.

The charge q flowing into the nonlinear capacitor may be written as a double Fourier series,

$$q = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} Q_{m,n} e^{i(mx+ny)} \quad (2)$$

$$\begin{aligned} x &= \omega_1 t & \omega_1 &= 2\pi f_1 \\ y &= \omega_0 t & \omega_0 &= 2\pi f_0. \end{aligned} \quad (3)$$

Since q is real,

$$Q_{m,n} = Q_{-m,-n}^* \quad Q_{-m,-n} = Q_{m,n}^{*11} \quad (4)$$

The variables x and y are initially considered as independent, taking on any values in the $x-y$ -plane. They are subsequently replaced by the substitution of (3), so that only values along a straight line in the $x-y$ plane appear in the final results. This method was first used by W. R. Bennett in calculating modulation products.¹²

Taking the total derivative of (2) with respect to time, we obtain the current i flowing into the nonlinear capacitor.

$$i = \frac{dq}{dt} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} I_{m,n} e^{i(mx+ny)} \quad (5)$$

$$I_{m,n} = j(m\omega_1 + n\omega_0)Q_{m,n} \quad (6)$$

$$I_{m,n} = I_{-m,-n}^* \quad I_{-m,-n} = I_{m,n}^*. \quad (7)$$

Next, the voltage v is assumed to be a single-valued, periodic function of x and y , so that it may be represented by a double Fourier series similar to (2). In the present case where the characteristic $f(q)$ is single-valued, v will satisfy these conditions, in view of (2) for q . In the analysis of hysteresis in a later section it will be shown that under the restrictions which insure that $f(q)$ is at most double-valued, v will again satisfy these conditions. Thus,

$$v = f(q) = f[q(x, y)] = F(x, y) \quad (8)$$

where in both cases $F(x, y)$ is single-valued and periodic in x and y . Consequently v may be expanded in a double Fourier series.

¹¹ The * indicates the complex conjugate.

¹² W. R. Bennett, "New results in the calculation of modulation products," *B.S.T.J.*, vol. 12, pp. 228-243; April, 1933.

$$v = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} V_{m,n} e^{j(mx+ny)} \quad (9)$$

$$V_{m,n} = V_{-m,-n}^* \quad V_{-m,-n} = V_{m,n}^*. \quad (10)$$

The coefficients are given by

$$V_{m,n} = \frac{1}{4\pi^2} \int_0^{2\pi} dy \int_0^{2\pi} dx \cdot F(x, y) e^{-j(mx+ny)}. \quad (11)$$

Let us multiply both sides of (11) by $jmQ_{m,n}^*$ and sum from $-\infty$ to $+\infty$ over m and n . Then, interchanging the order of integration and summation on the right-hand side, (11) becomes

$$\begin{aligned} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jmQ_{m,n}^* V_{m,n} &= \frac{1}{4\pi^2} \int_0^{2\pi} dy \int_0^{2\pi} dx F(x, y) \\ &\quad \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jmQ_{m,n}^* e^{-j(mx+ny)}. \end{aligned} \quad (12)$$

The double summation inside the integral may be easily identified. Taking (2) for q and differentiating with respect to x ,

$$\frac{\partial q}{\partial x} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jmQ_{m,n} e^{j(mx+ny)}. \quad (13)$$

Using (4), this may be written as

$$\frac{\partial q}{\partial x} = - \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} jmQ_{m,n}^* e^{-j(mx+ny)}, \quad (14)$$

which is just the negative of the double summation inside the integral of (12). Further, solving (6) for $Q_{m,n}$, taking the complex conjugate, and substituting into the left-hand side of (12), this latter equation may be written as

$$\begin{aligned} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mV_{m,n} I_{m,n}^*}{mf_1 + nf_0} \\ = \frac{1}{2\pi} \int_0^{2\pi} dy \int_0^{2\pi} dx \cdot \frac{\partial q}{\partial x} \cdot F(x, y). \end{aligned} \quad (15)$$

Since $(\partial q/\partial x)dx$ is just dq with y held constant, the variable in the second integral in (15) may be changed from x to q . Referring to (8), (15) becomes

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mV_{m,n} I_{m,n}^*}{mf_1 + nf_0} = \frac{1}{2\pi} \int_0^{2\pi} dy \int_{q(0,y)}^{q(2\pi,y)} f(q) dq. \quad (16)$$

The limits on the second integral indicate that the variation of q is determined by allowing x to vary from 0 to 2π , holding y constant.

A similar analysis may be performed in which the roles of x and y are interchanged. Both sides of (9) are multiplied by $jnQ_{m,n}^*$ and summed from $-\infty$ to $+\infty$ over m and n . Interchanging the order of summation and integration on the right-hand side, the resulting double summation is identified as $-(\partial q/\partial y)$; the remainder of the analysis is identical to that given above. Thus, the analog of (16) becomes

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{nV_{m,n} I_{m,n}^*}{mf_1 + nf_0} = \frac{1}{2\pi} \int_0^{2\pi} dx \int_{q(x,0)}^{q(x,2\pi)} f(q) dq, \quad (17)$$

where the limits on the second integral indicate that the variation of q is determined by allowing y to vary from 0 to 2π with x constant.

While all of the above equations have included both positive and negative frequencies, there can be no physical distinction between a positive and a negative frequency of the same magnitude. Therefore, we wish to combine appropriate pairs of terms on the left-hand sides of (16) and (17), relating the quantities $V_{m,n} I_{m,n}^*$ to the average powers associated with the various frequencies.

Defining

$$S_{m,n} = W_{m,n} + jX_{m,n} = 2V_{m,n} I_{m,n}^* \quad (18)$$

$$S_{m,n} = S_{-m,-n}^* \quad S_{-m,-n} = S_{m,n}^* \quad (19)$$

we have

$$W_{m,n} = V_{m,n} I_{m,n}^* + V_{m,n}^* I_{m,n} = W_{-m,-n} \quad (20)$$

$$X_{m,n} = j(V_{m,n}^* I_{m,n} - V_{m,n} I_{m,n}^*) = -X_{-m,-n}. \quad (21)$$

A particular frequency includes both the positive and the negative components $m'f_1 + n'f_0$ and $-(m'f_1 + n'f_0)$, where m' and n' have been chosen so that $m'f_1 + n'f_0 > 0$ (the alternate choice would have been $-m'$, $-n'$). Then $S_{m',n'}$, $W_{m',n'}$, and $X_{m',n'}$ are respectively the vector, real, and reactive powers flowing into the nonlinear element at this frequency. While this choice of indexes is necessary in order for $S_{m,n}$ and $X_{m,n}$ to agree with the usual definitions of vector and reactive power, in view of (20), $W_{m,n}$ will be the real or average power flowing into the nonlinear element at this frequency for either choice of indexes.

Inspection of (16) and (17) shows that only the real powers $W_{m,n}$ will appear in the nonlinear reactor case, and consequently only (20) of the present discussion will be used in this section; the reactive powers will appear in the corresponding analysis of the nonlinear resistor. Eq. (20) for the real power may be easily derived by taking the time average of the product of (5) and (9).

Thus, combining pairs of terms corresponding to the positive and negative components of each frequency, taking care to include each term only once, and making use of (20), (16), and (17) become:

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mW_{m,n}}{mf_1 + nf_0} = \frac{1}{2\pi} \int_0^{2\pi} dy \int_{q(0,y)}^{q(2\pi,y)} f(q) dq \quad (22)$$

$$\sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{nW_{m,n}}{mf_1 + nf_0} = \frac{1}{2\pi} \int_0^{2\pi} dx \int_{q(x,0)}^{q(x,2\pi)} f(q) dq, \quad (23)$$

where $W_{m,n}$ is the average power flowing into the nonlinear capacitor at the frequencies $\pm |mf_1 + nf_0|$. Note that the summations extend over different ranges in these two equations. To obtain terms of a similar form on the left-hand sides of both equations, they may be rewritten so that positive frequencies always appear in the denominator and the corresponding values of m and

n appear as subscripts on W , making use of (20). For example, the term corresponding to the frequency $3f_1 - 2f_0$ (assuming $3f_1 > 2f_0$) in (22) is

$$\frac{3W_{+3,-2}}{3f_1 - 2f_0}.$$

The term corresponding to this frequency in (23) is written

$$\frac{2W_{-3,+2}}{-3f_1 + 2f_0} = -\frac{2W_{+3,-2}}{3f_1 - 2f_0}.$$

We next make use of the requirement that there be no hysteresis in the nonlinear characteristic, noting that since this assumption has not yet been introduced, (22) and (23) are valid for a more general type of characteristic. These equations provide the starting point for the treatment of hysteresis in a later section.

The limits on the second integrals in (22) and (23) indicate that the variation of q is determined by allowing x and y respectively to vary from 0 to 2π , holding the other variable constant. Since q is periodic in x and y and consequently returns to its initial value at the end of these integrations, and since $f(q)$ is single-valued, these two integrals on q will be identically zero for all values of y and x respectively. Consequently the right-hand sides of these equations are equal to zero if the nonlinear characteristic is single-valued, and they become

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mW_{m,n}}{mf_1 + nf_0} = 0. \quad (24)$$

$$\sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{nW_{m,n}}{mf_1 + nf_0} = 0. \quad (25)$$

Eqs. (24) and (25) are the final results that have been sought. They provide two independent relations among the powers flowing into the nonlinear capacitor at the various frequencies; they apply equally well to the case of a nonlinear inductor, the analysis being almost identical to that given above. These results are a consequence only of the single-valuedness of the characteristic of the nonlinear capacitor or inductor. They are quite remarkable in that they are independent of the shape of this characteristic and of the power levels. We note that since they hold true in general, they must apply to a linear capacitor, in which no modulation effects take place. In this case $W_{m,n} = 0$ for all m and n , and (24) and (25) remain satisfied.

Since, in the absence of hysteresis no power may be dissipated in the nonlinear reactor, the sum of all the powers flowing into it at the different frequencies must be zero. This result is readily obtained from (24) and (25) by multiplying (24) by f_1 , (25) by f_0 , and adding the two together.

Certain general properties of nonlinear reactor modulators and demodulators may be deduced from these relations. We assume that two generators are connected

to the device at frequencies f_1 and f_0 . Since power can enter the nonlinear element from only the two generators, all of the $W_{m,n}$ must be negative or zero except $W_{1,0}$ and $W_{0,1}$. Since (24) contains $W_{1,0}$ but not $W_{0,1}$ and (25) contains $W_{0,1}$ but not $W_{1,0}$, we may solve these two equations for the respective generator powers in terms of the sideband powers. Then, as Manley and Peterson^{1,5} have pointed out, these equations show that the flow of a sum frequency (m and n both positive) yields a positive contribution to the power entering the nonlinear element from both generators, while the flow of a difference frequency (m and n of opposite sign) yields a positive contribution to one generator power and a negative contribution to the other. Stated differently, the flow of a sum frequency introduces positive resistances in both source circuits, while flow of a difference frequency introduces a positive resistance in one source circuit and a negative resistance in the other. Of course the values of the various powers in these equations can be determined only by a detailed circuit analysis of the nonlinear element and the external circuit.

While at least one of the generator powers $W_{1,0}$ and $W_{0,1}$ must be positive, (24) and (25) show that it is possible for one of them to be negative. There are consequently two possible situations: 1) $W_{1,0}$ and $W_{0,1}$ both positive. Power flows into the nonlinear element from both generators, and the device is said to be stable. 2) Either $W_{1,0}$ or $W_{0,1}$ negative. Power flows into the nonlinear element from one generator, but is returned to the other generator by the nonlinear element. The device is said to be potentially unstable.

The reason for these definitions is that if the external positive resistance in the generator circuit to which the nonlinear element presents a negative resistance is reduced until the net resistance of this circuit is zero, the generator can be removed and current of approximately the same frequency will continue to flow in this circuit, generating sidebands with the remaining generator, as before. The exact frequency of operation is determined by the tuning of the external circuits. Thus, new frequencies in general incommensurable with a single applied frequency can be generated by a nonlinear capacitor or inductor under these conditions, several different types of oscillation being possible.⁵

In considering the behavior of the device as a modulator or as a demodulator, one generator represents the local oscillator and the other the applied signal. One or more of the resulting sidebands may be taken as the useful signal output. The signal input and output terminals may thus be regarded as externally available, while the local oscillator is regarded as an internal part of the modulator. Instability can arise as a result of negative resistance introduced in either the local oscillator circuit or in the applied signal circuit.

Consider first instability in the local oscillator circuit. Under normal conditions the applied signal level is so far below the local oscillator level that even though it introduces a negative resistance in the local oscillator

circuit, the losses in this circuit are sufficient to prevent instability. However, under these conditions if the input signal level is raised until it becomes comparable with the local oscillator level, the total resistance in the local oscillator circuit can become negative, with the resulting instability and the production of new frequencies unrelated to the signal and local oscillator frequencies.

Our primary interest is in the case where the signal levels are small compared to the local oscillator level. Consequently instability in the local oscillator circuit is ignored, and the device is regarded as potentially unstable only if a negative resistance appears in the applied signal circuit, so that it is possible to have power emerging at all of the signal frequencies with only the local oscillator applied and passive impedances connected to the signal terminals; otherwise the device is considered to be stable. Regarding the local oscillator circuit as an internal part of the modulator, this definition of stability is in accord with the usual definition adopted for transducers. However, the possibility of instability caused by negative resistance introduced in the local oscillator circuit must be kept in mind for extreme operating conditions.

In order to simplify the analysis of modulation problems it is customary to assume that currents (or voltages) at only the most important frequencies are allowed to flow through (or exist across) the nonlinear element, all the other frequencies being suppressed by ideal filters. In the present analysis of nonlinear reactors it may be assumed that all frequencies except the ones of interest are terminated in an open circuit, short circuit, or pure reactance so that $W_{m,n} = 0$ for all but the wanted frequencies. While this is a useful procedure, in practice it must be kept in mind that although the powers of all of the unwanted frequencies may be quite small, they are not likely to be strictly zero; and some of these unwanted components can cause regeneration or even oscillation under unfavorable conditions, greatly modifying the characteristics of the device. This method is adopted in the next section, where the simplest nonlinear reactor devices are discussed.

NONLINEAR REACTOR MODULATORS, DEMODULATORS, AND HARMONIC GENERATORS^{1,10}

The general results of the previous section are best illustrated by some simple examples. Consider nonlinear reactor modulators and demodulators in which only one of the principal sidebands of the signal about the carrier is allowed to carry a significant amount of power; all other sidebands are assumed to be reactively terminated. Fig. 2 illustrates the frequencies involved in each of the two types of devices considered. Recalling that in the previous section f_1 and f_0 were taken as the two applied frequencies, the local oscillator frequency is designated as f_l . We have avoided using the symbol f_0 in order to eliminate confusion, because we wish to allow any of the signal frequencies to be the signal input in order to discuss the different modulators and demodu-

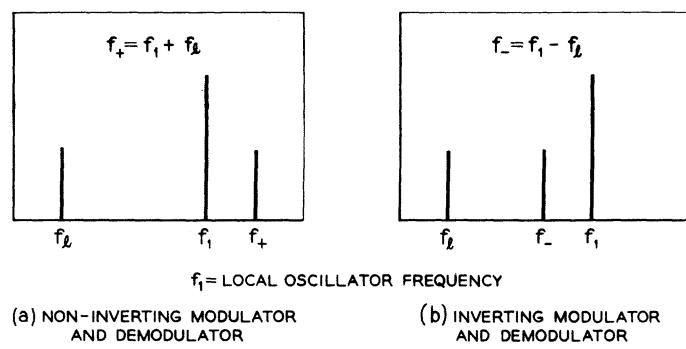


Fig. 2—Signal and local oscillator frequencies for modulators and demodulators.

lators. Consequently f_l indicates the lowest signal frequency, small compared to the local oscillator frequency, while f_+ and f_- are equal to $f_l \pm f_1$ respectively. For the modulators the signal input is at f_l , the signal output at either f_+ or f_- ; for the demodulators the signal input is at either f_+ or f_- , the signal output at f_l .

The two cases illustrated in Fig. 2 are distinguished by whether the local oscillator frequency lies between the two signal frequencies or whether it is greater than either of the signal frequencies. The choice of names *noninverting* and *inverting* is based on what the device does to the signal spectrum. This is best understood by considering not a single signal frequency as shown but a narrow band of signal frequencies. Then it is easily seen that the modulator and demodulator illustrated in Fig. 2(a) do not invert the signal spectrum, while those of Fig. 2(b) do invert the signal spectrum.

Taking the noninverting modulator and demodulator, whose signal and local oscillator frequencies are shown in Fig. 2(a), and assuming that the powers at all but the three frequencies shown are equal to zero, (24) and (25) become

$$\frac{W_1}{f_1} + \frac{W_+}{f_+} = 0, \quad (26)$$

$$\frac{W_l}{f_l} + \frac{W_+}{f_+} = 0, \quad (27)$$

where W_l , W_+ , and W_1 represent the powers flowing into the nonlinear reactor at the frequencies f_l , f_+ , and f_1 respectively. For the modulator W_l and W_1 are positive, representing power flowing into the nonlinear reactor, while W_+ is negative, representing the useful power output of the device. For the demodulator the situation is reversed, W_+ being positive and W_l and W_1 negative.

The power gain G_p is easily found from (27) for the noninverting modulator and demodulator. This gain is defined as the ratio of the power delivered to a specified load impedance to the power absorbed by the input of a transducer.¹³ The two power gains are

$$G_{p+} = \frac{f_+}{f_l}, \text{ noninverting modulator power gain} \quad (28)$$

¹³ "American standard definitions of electrical terms," ASA C42; August, 1953.

$$G_{p+1} = \frac{f_l}{f_+}, \text{ noninverting demodulator power gain} \quad (29)$$

where the subscripts show the direction of travel of the signal, the first subscript indicating the input terminals, the second, the output terminals. The power gains are independent of the load impedances, which is not true for all transducers.

The power gain is seen to be simply the ratio of the output frequency to the input frequency; the modulator and demodulator gains are reciprocal. Thus a noninverting modulator operating between widely separated frequencies has a substantial power gain; a small amount of input power at the frequency f_l controls a large amount of local oscillator power at the frequency f_1 , the sum of these powers appearing as useful output at the frequency f_+ . In contrast a noninverting demodulator operating between widely separated frequencies has a corresponding loss; most of the input power at the frequency f_+ is wasted in the local oscillator circuit, very little of it emerging as useful output at the frequency f_l . These results hold true for any single-valued nonlinear characteristic and for any external circuit conditions, assuming that all unwanted frequencies are terminated in lossless circuits.

Since these power gains are positive the device must be stable, according to the discussion of the preceding section; that is, power cannot be delivered to purely passive terminations at both of the signal terminals (f_l and f_+) simultaneously with only the local oscillator applied to the nonlinear reactor. Consequently, the power gain of the noninverting modulator and demodulator is equal to the maximum available gain,¹³ which is equal to the transducer gain,¹³ defined as the ratio of the power delivered to the load impedance to the available power of the source, for a conjugate match at both source and load terminals. We note that the instability in the local oscillator circuit discussed in the preceding section can occur only in the demodulator, and then only if the input signal power at the frequency f_+ is not small compared to the local oscillator power.

Next, consider the inverting modulator and demodulator, whose signal and local oscillator frequencies are shown in Fig. 2(b). Assuming as before that the powers at all unwanted frequencies are equal to zero, (24) and (25) become

$$\frac{W_1}{f_1} + \frac{W_-}{f_-} = 0 \quad (30)$$

$$\frac{W_l}{f_l} - \frac{W_-}{f_-} = 0. \quad (31)$$

These differ from (26) and (27) for the noninverting device in the negative sign in the second equation. From (31) the power gains are

$$G_{p-1} = -\frac{f_-}{f_l}, \text{ inverting modulator power gain.} \quad (32)$$

$$G_{p-1} = -\frac{f_l}{f_-}, \text{ inverting demodulator power gain.} \quad (33)$$

The power gains are again reciprocal, but in this case are the negative of the ratio of the output frequency to input frequency, in contrast with (28) and (29) for the noninverting case. For both the inverting modulator and demodulator W_1 is positive, W_l and W_- negative; the input power from the local oscillator at the frequency f_l flows out of the nonlinear reactor at the two signal frequencies, most of it appearing at the higher signal frequency f_- , only a small amount appearing at the lower signal frequency f_l . Thus the inverting device is potentially unstable and under certain conditions will oscillate, power emerging from both of the signal terminals into passive terminations with only the local oscillator applied.

Consequently, the transducer gain for either the inverting modulator or demodulator can take on any value between zero and infinity, depending on the external impedances at the signal terminals. The resistive component of input impedance at either signal frequency is negative, and so neither the source nor the load can be matched. Since, for the inverting device, most of the input power from the local oscillator appears at the high signal frequency, we would expect that for fixed terminal impedances the modulator gain will greatly exceed the demodulator gain, as with the noninverting device, although in contrast to this latter case these two gains can both be made arbitrarily large by operating close to the point of instability. Finally, we note that instability in the local oscillator circuit cannot arise in this case.

As a final example, consider the application of (24) and (25) to a nonlinear reactor harmonic generator. Assuming the power input $W_{1,0}$ at the fundamental frequency f_1 and the power output $-W_{m,0}$ at the harmonic frequencies mf_1 , setting all other powers equal to zero,

$$W_{1,0} = \sum_{m=2}^{\infty} -W_{m,0}. \quad (34)$$

The total harmonic power output is equal to the fundamental power input, in agreement with the fact that the nonlinear reactor is assumed lossless. If the load impedances at the unwanted harmonic frequencies can be made zero, infinite, or purely reactive, the power gain to the desired harmonic can approach 1.

HYSTeresis

In this section we extend the analysis of nonlinear reactors given in the third section to include the effects of hysteresis in the special case for which there are no departures from the principal single-frequency hysteresis loop established by the local oscillator alone. Thus, the path traversed in the $q-v$ plane is assumed to be, at most, double-valued.

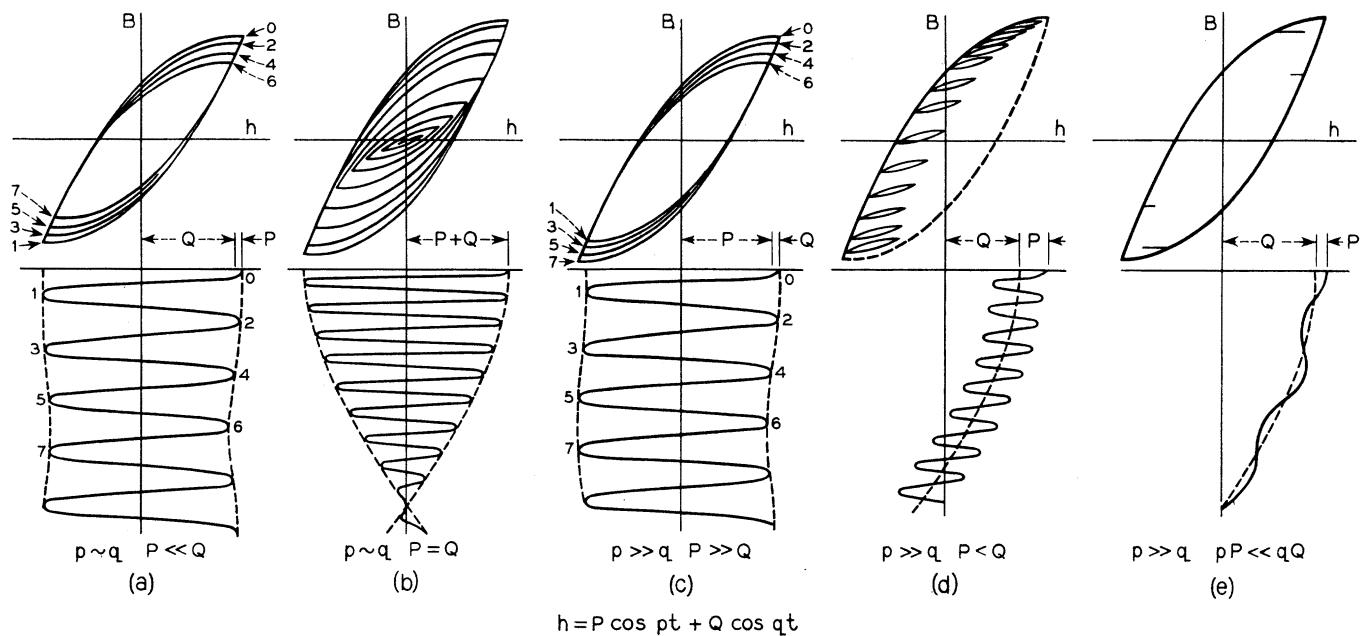


Fig. 3—Types of hysteresis loops characteristic of a two-frequency magnetizing force (taken from R. M. Kalb and W. R. Bennett,¹⁴ Fig. 1).

R. M. Kalb and W. R. Bennett¹⁴ have considered the problem of a nonlinear inductor with hysteresis when currents at only two frequencies flow in the element and when the operation is far from saturation. They give a detailed discussion of the various ways in which complex hysteresis loops can be formed and the corresponding effects on the operation of the device.

The present analysis differs from Kalb and Bennett's in three principal respects: 1) More than two driving frequencies are involved. 2) The level of the local oscillator at the frequency f_1 is assumed to be large enough to drive the nonlinear element well into saturation, so that the $q-v$ characteristic will be single-valued over appreciable regions near its ends. The signal levels at the other frequencies are assumed to be small compared to the local oscillator level. 3) It is necessary to restrict this analysis by ruling out complex hysteresis loops and requiring the $q-v$ characteristic to be at most double-valued in order to obtain any simple general results. The restrictions imposed in 2) and 3) above simplify the analysis in two ways: The voltage v , while a double-valued function of the charge q , is a single-valued function of x and y in (8), so that (22) and (23) remain valid in this case. The integrals on the right-hand sides of (22) and (23) may be simply evaluated in terms of the area of the hysteresis loop, as shown below.

The possible types of departure from the principal single-frequency hysteresis loop for the case of two applied frequencies and operation far from saturation are illustrated in Fig. 3, taken from a previous reference.¹⁴ These show the general types of complex hysteresis

loops which must be considered in the present analysis, although since many driving frequencies may be present, the corresponding complex hysteresis loops may be combinations of the types shown in Fig. 3.

Taking account of the fact that the local oscillator level far exceeds the other signal levels, there are two principal types of departure from the main hysteresis loop that must be considered. The first type, illustrated in Fig. 3(a) and (c), corresponds to a slow modulation of the single-frequency hysteresis loop. There are no reversals between successive maxima of the hysteresis loop, but the magnitudes of these maxima vary in a periodic manner. Consequently the successive branches of the complex hysteresis loop are all quite similar but are shifted a small amount from each other on each successive cycle of the local oscillator. This can occur either because of small high frequency components whose frequencies are near the local oscillator frequency f_1 , which modulate the amplitude of the local oscillator at a slow rate, or because of small low frequency components, which shift the average value of the local oscillator waveform at a slow rate. Since in the present analysis the local oscillator is assumed to drive the nonlinear device well into saturation, and since the other signal levels are assumed to be small compared to the local oscillator, this type of departure from the main loop cannot occur and is eliminated from further consideration.

The second type of departure from the principal loop is illustrated in Fig. 3(d) and (e). Here there are small reversals between the principal maxima of the complex hysteresis loop, which result in the formation of many minor loops inside the principal loop; the principal maxima all have approximately the same magnitude. This phenomenon is caused by the presence of small

¹⁴ R. M. Kalb and W. R. Bennett, "Ferromagnetic distortion of a two-frequency wave," *B.S.T.J.*, vol. 14, pp. 322-359; April, 1935.

components of much higher frequency than the local oscillator. This type of departure from the main loop must also be ruled out if the conditions of the present analysis are to be satisfied; this can be done in two ways:

1) The device may be assumed to have the property that small reversals in the direction of travel as the principal hysteresis loop is traversed do not cause departures from this principal loop, in the form of minor loops.

2) If the assumption, 1) above, is not tenable, then further restrictions must be placed on the signal levels so that such reversals cannot occur in the open portion of the principal hysteresis loop; *i.e.*, the part of the loop that is double-valued. This sort of restriction is illustrated in Fig. 3(e) for the case of only one high frequency component and the nature of the necessary restriction on levels is given for this case. The restrictions in the general case are similar, merely requiring that the sum of the maximum slopes of all of the high frequency components must be less than the slope of the local oscillator driving function at the ends of the open portion of the main loop. One special case of interest in which this restriction must be satisfied is that discussed in previous sections in which currents at only the local oscillator frequency f_1 and the three signal frequencies f_0 and $f_1 \pm f_0$ are allowed to flow in the nonlinear element (or the dual case where voltages at only these four frequencies are allowed to exist across it). Then none of the high frequency components exist, and no minor loops can be formed.

Subject to the above conditions, the $q-v$ characteristic may be treated as, at most, a double-valued function. While the signal levels must be fairly small compared to the local oscillator to satisfy these conditions, they need not be vanishingly small, as in the small signal analysis of Part II. Consequently this might be called a quasi-small-signal analysis.

Returning to the analysis of the third section, in (1) we now regard v as a double-valued function of q . In the subsequent analysis the lower branch of this function must be chosen when the charge at the local oscillator frequency f_1 is increasing, the upper branch when it is decreasing. For example, if in (2) $Q_{1,0} = Q_{-1,0}$, so that both are real and the local oscillator charge at frequency f_1 is a cosine function, then for $0 < x < \pi$ the upper branch of the $q-v$ characteristic must be chosen, while for $\pi < x < 2\pi$ the lower branch is selected. In (8) while v is a double-valued function of q , it is a single-valued function of x and y , as it must be if v is to be expanded into a double Fourier series. Thus, hysteresis does not alter the analysis leading up to (22) and (23).

The integrals on the right-hand sides of (22) and (23) were found to be identically equal to zero in the third section for the case of no hysteresis, $v=f(q)$ single-valued. The introduction of hysteresis, with $v=f(q)$ double-valued, requires reconsideration of these integrals.

Consider first the integral of (22). With $y=\text{constant}$, q will travel entirely around the hysteresis loop as x goes

from 0 to 2π . Thus the integral on q yields a constant, independent of y , equal to the area of the hysteresis loop, which is called h . The integral on y cancels the factor $1/2\pi$, and (22) becomes

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mW_{m,n}}{mf_1 + nf_0} = h \quad (35)$$

where

$$h = \oint f(q) dq, \text{ area of hysteresis loop} \quad (36)$$

h is equal to the energy dissipated in the nonlinear reactor in one transit around the hysteresis loop, which occurs in a time equal to one period of the local oscillator frequency f_1 . Therefore the average power H dissipated in the nonlinear reactor is

$$H = f_1 \cdot h \quad (37)$$

and so (35) may be written

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mW_{m,n}}{mf_1 + nf_0} = \frac{H}{f_1} \quad (38)$$

where H is the average power dissipated in hysteresis.

Next, consider the integral of (23). With x constant, q will travel back and forth along the same branch of the hysteresis loop, returning to its initial value without enclosing any area, as y goes from 0 to 2π . Thus, this integral on q is identically equal to zero for all x , and the right-hand side of (23) remains equal to zero, as in the case of no hysteresis, yielding

$$\sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{nW_{m,n}}{mf_1 + nf_0} = 0. \quad (39)$$

Eqs. (38) and (39) give the general power relations in nonlinear reactors with hysteresis, subject to the restrictions discussed above, which insure that the hysteresis loop is no more than double-valued. We must remember that f_1 has been taken as the local oscillator frequency. Examining (38), we see that the hysteresis term may be transferred to the left-hand side and combined with the local oscillator term, $W_{1,0}/f_1$, to give $(W_{1,0} - H)/f_1$. Referring to (24) and (25), we see that the only difference from the results for the hysteresisless case is that the power lost in hysteresis is subtracted from the input power from the local oscillator. Thus, the power lost in hysteresis may be considered to come only from the local oscillator circuit, subject to the restriction that the hysteresis loop is no more than double-valued.

Consequently the power gains of the nonlinear reactor modulators and demodulators discussed above will remain the same while the power gain of the harmonic generator will be reduced in the presence of hysteresis.

The presence of complex hysteresis loops would complicate these calculations considerably, as illustrated by Kalb and Bennett's analysis.¹⁴ Since no general results could be obtained, this case is not considered.

NONLINEAR RESISTOR

A general analysis for the nonlinear resistor, similar to that given above for nonlinear reactors, may be carried out. However, the results of the nonlinear resistor analysis are relations among the reactive powers at the various frequencies, rather than among the real powers, as in the case of nonlinear reactors; further, the frequencies do not enter these equations.

The nonlinear resistor characteristic may be given by specifying the voltage v as some arbitrary single-valued function of the current i .

$$v = f(i). \quad (40)$$

As before, the current is written as a double Fourier series.

$$i = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} I_{m,n} e^{i(mx+ny)} \quad (41)$$

$$x = \omega_1 t \quad \omega_1 = 2\pi f_1$$

$$y = \omega_0 t \quad \omega_0 = 2\pi f_0 \quad (42)$$

$$I_{m,n} = I_{-m,-n}^* \quad I_{-m,-n} = I_{m,n}^* \quad (43)$$

Since the characteristic is single-valued, v is also periodic in x and y and may be written

$$v = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} V_{m,n} e^{i(mx+ny)} \quad (44)$$

$$V_{m,n} = V_{-m,-n}^* \quad V_{-m,-n} = V_{m,n}^*. \quad (45)$$

Proceeding as in the nonlinear reactor case, and making use of (18)–(21), the results corresponding to (24) and (25) are

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} mX_{m,n} = 0 \quad (46)$$

$$\sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} nX_{m,n} = 0 \quad (47)$$

with $X_{m,n}$ given by (21).

As with the nonlinear capacitor results, the summations in (45) and (46) extend over different ranges. Consequently, it is necessary to rewrite these results to obtain similar terms in the two equations, making use of (21) and making the choice of subscripts m,n that insures that $mf_1 + nf_0 > 0$, so that the resulting X 's will represent the reactive powers flowing into the nonlinear resistor at the various frequencies.

These relations among the reactive powers in a nonlinear resistor do not appear to be particularly useful. The interpretation of reactive power in this situation is not clear. Note that reactive power in a nonlinear resistor is not conserved, as was real power in the nonlinear capacitor.

DISCUSSION

The general energy relations derived above provide two independent equations relating the powers at the different frequencies in nonlinear capacitors and inductors. These results are independent of the detailed shape of the nonlinear characteristic, of the levels at the various frequencies, and of the external circuits. They show which sidebands in a nonlinear reactor tend to make the device stable and which tend to make it unstable, and in the latter case in which generator circuit the negative resistance will appear. They thus provide useful information about the gain and stability of nonlinear reactor modulators, including such devices as magnetic and dielectric amplifiers, and indicate the conditions under which the production of subharmonic and other frequencies incommensurable with the applied frequency is possible in a nonlinear reactor. These equations would be useful in checking the results of detailed analysis of special nonlinear circuits.

The discussion of the two simplest types of nonlinear reactor modulators and demodulators, illustrated in Fig. 2, shows further properties of these devices. The non-inverting device is stable and gives the best performance when matched; it has a high modulator gain and demodulator loss. The inverting device is potentially unstable, and so can not be matched; while the modulator gain again greatly exceeds the demodulator gain, either can become as large as desired by operating close enough to the point of instability. This suggests that although a reasonable modulator gain may be obtained with the inverting device, gain as a demodulator can be obtained only by operating so close to instability that the resulting bandwidth will be narrow and the sensitivity¹⁵ of the gain to changes in the terminating impedances will be high.

In the small signal analysis of Part II (to be submitted later) the bandwidth, sensitivity, and the terminal impedances of these devices are determined for the special case in which the signal levels are much smaller than the local oscillator level.

¹⁵ H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co., Inc., New York, N.Y.; 1945.



Some General Properties of Nonlinear Elements.

II. Small Signal Theory*

H. E. ROWE†, ASSOCIATE MEMBER, IRE

Summary—The simplest types of nonlinear capacitor modulators, demodulators, and negative conductance amplifiers, in which components at only two signal frequencies are present, are studied by means of the well-known small-signal analysis. The results of this analysis of course agree with the general energy relations of Part I,¹ but in addition give the gain, bandwidth, terminal admittances, and sensitivity (to changes in the terminal admittances or in the local oscillator drive) of these devices, and show the way in which these quantities depend on the amount of nonlinearity. In general, the bandwidth of all of these devices approaches zero as the nonlinearity approaches zero.

Three cases are considered; f_1 is the local oscillator frequency.

1) Noninverting modulator and demodulator—signal frequencies f_1 and $f_+ = f_1 + f_1$. This device is stable and yields maximum gain with matched source and load. Under matched conditions the gain is equal to the ratio of output to input frequency; for widely separated signal frequencies the modulator has substantial gain, the demodulator an equal loss. Only a relatively small amount of nonlinearity is required to attain a bandwidth equal to the low-signal frequency. Since source and load are matched the sensitivity is zero.

2) Inverting modulator and demodulator—signal frequencies f_1 and $f_- = f_1 - f_1$. This device is potentially unstable; its input conductance is negative and match is impossible. The modulator gain again exceeds the demodulator gain, the ratio of the two being the same as in the noninverting case; but now both gains may be made as large as desired at the expense of narrow bandwidth and high sensitivity. For gains equal to the ratio of output to input frequency (as in the matched noninverting device) the bandwidth is slightly smaller than in the noninverting case; again a relatively small amount of nonlinearity yields a bandwidth equal to the low-signal frequency. For a larger amount of nonlinearity, substantially greater gains can be obtained with moderate bandwidth and sensitivity. However, a demodulator with high gain must have a very narrow bandwidth and a high sensitivity.

3) Inverting negative conductance amplifier. The negative input conductance of the inverting device discussed above may be used to provide amplification at a single frequency. Substantial gain at moderate bandwidth and sensitivity may be obtained at either signal frequency; as with the inverting modulator and demodulator, high gain implies narrow bandwidth and high sensitivity. A large nonlinearity is desirable to achieve the maximum bandwidth.

Finally, the general energy relations of Part I for nonlinear reactors are shown to apply to lossless linear variable reactors for arbitrary signal levels if the powers at the different frequencies in these equations include mechanical as well as electrical power.

INTRODUCTION

In Part I¹ two independent equations relating the powers at the different frequencies in nonlinear capacitors and inductors were derived. These general energy relations are independent of the nonlinear characteristic, the external circuit, and the signal levels at

* Original manuscript received by the IRE, December 20, 1957; revised manuscript received February 6, 1958.

† Bell Telephone Labs., Inc., Holmdel, N. J.

¹ J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—Part I. General energy relations," PROC. IRE, vol. 44, pp. 904-913; July, 1956.

the various frequencies. These relations have been derived in a simpler way for quantum mechanical systems such as Masers.² Alternate classical derivations for nonlinear capacitors and inductors have been given.^{3,4}

Although these results give useful information about the general behavior of various nonlinear reactor devices, they are, of course, not a substitute for a detailed analysis of a particular device. While in general a complete analysis will be difficult to carry out, one important case in which it may be accomplished readily is the well-known small-signal case,⁵⁻⁹ in which the levels of the signal components applied to the nonlinear element are much smaller than the level of the local oscillator.

The present paper uses the standard small-signal analysis to discuss the behavior of the simplest nonlinear reactor devices, used to illustrate the general energy relations in Part I, in which only one of the principal sidebands of the applied signal about the local oscillator frequency is allowed to carry a significant amount of power. The results in the small-signal case of course agree with the general energy relations, but in addition yield expressions for the gain, bandwidth, impedances, and sensitivity (to changes in the terminal admittances or in the local oscillator drive) for the various modulators, demodulators, and negative conductance amplifiers. These results may be of interest in connection with the various recent microwave devices of this type using *p-n* junctions and ferrites.¹⁰⁻¹² As in Part I, the analysis will be carried out for the nonlinear capacitor only, since the corresponding analysis for the nonlinear inductor is very similar.

² M. T. Weiss, "Quantum derivation of energy relations analogous to those for nonlinear reactances," PROC. IRE, vol. 45, pp. 1012-1013; July, 1957.

³ B. Salzberg, "Masers and reactance amplifiers—basic power relations," PROC. IRE, vol. 45, pp. 1544-1545; November, 1957.

⁴ C. H. Page, "Frequency conversion with nonlinear reactance," *J. Res. Natl. Bur. Standards*, vol. 58, pp. 227-236; May, 1957.

⁵ L. C. Peterson and F. B. Llewellyn, "The performance and measurement of mixers in terms of linear-network theory," PROC. IRE, vol. 33, pp. 458-476; July, 1945.

⁶ A. van der Ziel, "On the mixing properties of nonlinear condensers," *J. Appl. Phys.*, vol. 19, pp. 999-1006; November, 1948.

⁷ H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, N. Y.; 1948.

⁸ C. F. Edwards, "Frequency conversion by means of a nonlinear admittance," *Bell Sys. Tech. J.*, vol. 35, pp. 1403-1416; November, 1956.

⁹ S. Duinker, "General properties of frequency-converting networks," Dissertation, Technical University of Delft, Netherlands; June, 1957.

¹⁰ A. Uhlig, "Two-terminal *p-n* junction devices for frequency conversion and computation," PROC. IRE, vol. 44, pp. 1183-1191; September, 1956.

¹¹ H. Suhl, "Proposal for a ferromagnetic amplifier in the microwave range," *Phys. Rev.*, vol. 106, pp. 384-385; April 15, 1957.

¹² M. T. Weiss, "A solid-state microwave amplifier and oscillator using ferrites," *Phys. Rev.*, vol. 107, p. 317; July 1, 1957.

For convenience the frequencies involved in the two types of devices to be considered are shown in Fig. 1 (taken from Fig. 2, Part I), together with the corresponding general energy relations and modulator and demodulator power gains, given in Part I. The notation and terminology will remain the same. f_1 is the local oscillator frequency, f_i the lowest signal frequency (normally small compared to the local oscillator frequency for modulators and demodulators), $f_{\pm} = f_1 \pm f_i$. For modulators the signal input is at f_i , the signal output at either f_+ or f_- ; for demodulators the signal input is at either f_+ or f_- , the signal output at f_i . The W 's represent power flowing into the nonlinear element at the corresponding frequencies indicated by the subscripts. G_p represents power gain, defined as the ratio of power delivered to a specified load impedance to the power absorbed by the input of a transducer,¹³ the subscripts indicate the direction of travel of the signal. The names inverting and noninverting were chosen because these devices respectively do and do not invert the signal spectrum when used as modulators or demodulators. A detailed discussion of the operation of these devices is given in Part I.

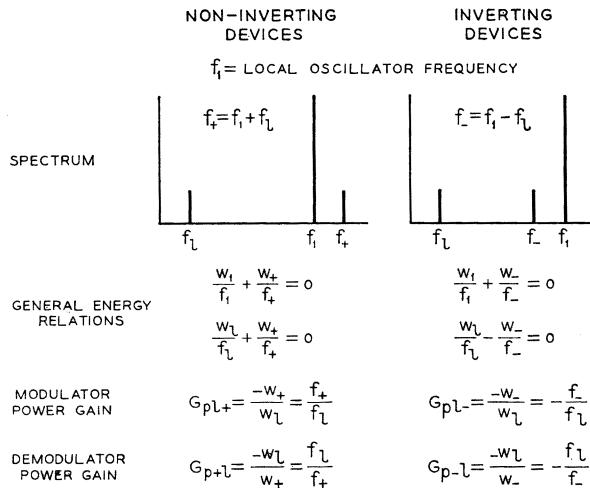


Fig. 1—Signal and local oscillator frequencies, general energy relations, and power gains for nonlinear reactor devices.

Finally, the general energy relations of Part I for nonlinear reactors are shown to hold true for any lossless linear variable reactor if, assuming the reactor to be varied mechanically, both electrical and mechanical power at every frequency are included in the analysis.

SMALL-SIGNAL ANALYSIS⁵⁻⁹

In this section the well-known small-signal analysis is summarized briefly for a nonlinear capacitor. The charge q will be some function of the voltage v ,

$$q = f(v). \quad (1)$$

Let q_1 and v_1 be the carrier components of charge and

¹³ ASA C42, "American Standard Definitions of Electrical Terms," August, 1953.

voltage, consisting of fundamental components and harmonics of the local oscillator frequency f_1 . Then in the absence of signal

$$q_1 = f(v_1). \quad (2)$$

Let the signal components of charge and voltage be given by δq and δv , small compared to the carrier components q_1 and v_1 . Then,

$$\delta q = f'(v_1) \cdot \delta v \quad f'(v) = \frac{df(v)}{dv}. \quad (3)$$

Now $f'(v_1)$, which may be thought of as an equivalent linear time-varying capacitance, will be periodic, of fundamental frequency f_1 . Therefore, it may be written as a Fourier series,

$$f'(v_1) = \sum_{n=-\infty}^{\infty} C_n e^{inx} \quad x = 2\pi f_1 t \quad (4)$$

$$C_n = \frac{1}{2\pi} \int_0^{2\pi} f'(v_1) e^{-inx} dx \\ C_n = C_{-n}^*. \quad (5)$$

Thus, as far as the signal components δq and δv are concerned, the nonlinear capacitor may be replaced by an equivalent linear variable capacitor whose variation with time is determined by the nonlinear characteristic and by the local oscillator waveform. The origin of time has been arbitrary up to now; we can always choose it in such a way that

$$C_1 = C_{-1} = \text{positive real}, \quad (6)$$

and in the following we will assume that this has been done.

For our special case we now assume that the nonlinear capacitor is terminated in an ideal filter such that signal voltages at only the three frequencies f_i , $f_+ = f_1 + f_i$, and $f_- = f_1 - f_i$ are allowed to exist across the nonlinear capacitor. Then,

$$\delta v = V_+^* e^{-j(x+y)} + V_-^* e^{-j(x-y)} + V_i^* e^{-iy} \\ + V_i e^{iy} + V_- e^{j(x-y)} + V_+ e^{j(x+y)} \quad (7)$$

$$\delta q = Q_+^* e^{-j(x+y)} + Q_-^* e^{-j(x-y)} + Q_i^* e^{-iy} \\ + Q_i e^{iy} + Q_- e^{j(x-y)} + Q_+ e^{j(x+y)} + \dots \quad (8)$$

$$x = 2\pi f_1 t \quad y = 2\pi f_1 t \quad (9)$$

As indicated, δq will contain additional components at all of the frequencies $m f_1 \pm f_i$, $m = 2, 3, \dots$. However, since the ideal filter suppresses the voltages at these frequencies, no power can flow except at the principal signal frequencies f_i , $f_1 \pm f_i$, and so these other components play no part in the present analysis. Substituting (4), (7), and (8) into (3), we find that the final results for the components of interest can be written in the usual matrix form

¹⁴ The * indicates the complex conjugate.

$$\begin{bmatrix} Q_-^* \\ Q_I \\ Q_+ \end{bmatrix} = \begin{bmatrix} C_0 & C_{-1} & C_{-2} \\ C_1 & C_0 & C_{-1} \\ C_2 & C_1 & C_0 \end{bmatrix} \begin{bmatrix} V_-^* \\ V_I \\ V_+ \end{bmatrix}. \quad (10)$$

Since $I = j2\pi f \cdot Q$ and $I^* = -j2\pi f \cdot Q^*$, we have from (10) the well-known relation

$$\begin{bmatrix} I_-^* \\ I_I \\ I_+ \end{bmatrix} = \begin{bmatrix} -j2\pi f_- C_0 & -j2\pi f_- C_{-1} & -j2\pi f_- C_{-2} \\ j2\pi f_I C_1 & j2\pi f_I C_0 & j2\pi f_I C_{-1} \\ j2\pi f_+ C_2 & j2\pi f_+ C_1 & j2\pi f_+ C_0 \end{bmatrix} \begin{bmatrix} V_-^* \\ V_I \\ V_+ \end{bmatrix}. \quad (11)$$

Eq. (11) describes the small-signal behavior of the nonlinear capacitor when all three signal frequencies are present. For the noninverting devices of Fig. 1, we assume the voltage at the frequency $f_- = f_1 - f_I$ is suppressed by the ideal filter. Setting $V_-^* = 0$ in (11), and noting (6),

Noninverting case

$$\begin{bmatrix} I_I \\ I_+ \end{bmatrix} = \begin{bmatrix} j2\pi f_I C_0 & j2\pi f_I C_1 \\ j2\pi f_+ C_1 & j2\pi f_+ C_0 \end{bmatrix} \begin{bmatrix} V_I \\ V_+ \end{bmatrix}. \quad (12)$$

In the alternate case the voltage at the frequency $f_+ = f_1 + f_I$ is suppressed, and setting $V_+ = 0$ in (11),

Inverting case

$$\begin{bmatrix} I_I \\ I_-^* \end{bmatrix} = \begin{bmatrix} j2\pi f_I C_0 & j2\pi f_I C_1 \\ -j2\pi f_- C_1 & -j2\pi f_- C_0 \end{bmatrix} \begin{bmatrix} V_I \\ V_-^* \end{bmatrix}. \quad (13)$$

Since C_0 and C_1 are real, the admittance matrices of (12) and (13) are pure imaginary.

GAIN AND TERMINAL ADMITTANCES FOR A FOUR-POLE WITH A PURE IMAGINARY ADMITTANCE MATRIX

Consider a linear four-pole characterized by a pure imaginary admittance matrix, such as those of (12) and (13).

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} jB_{11} & jB_{12} \\ jB_{21} & jB_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \quad (14)$$

The nonlinear capacitor may be treated as such a four-pole; the two signal frequencies are assumed to be brought out at two different ports by the ideal filter associated with the nonlinear element. The powers flowing into the network at the two ports are

$$W_1 = \frac{1}{2} \operatorname{Re} V_1 I_1^*, \quad W_2 = \frac{1}{2} \operatorname{Re} V_2 I_2^*. \quad (15)$$

Then from (14) we have

$$\begin{aligned} \frac{W_1}{B_{12}} + \frac{W_2}{B_{21}} &= \frac{1}{2} \operatorname{Re} \left[\frac{V_1^*}{B_{12}} \frac{V_2^*}{B_{21}} \right] \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \\ &= \frac{1}{2} \operatorname{Re} \left[\frac{V_1^*}{B_{12}} \frac{V_2^*}{B_{21}} \right] \begin{bmatrix} jB_{11} & jB_{12} \\ jB_{21} & jB_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0. \end{aligned}$$

Therefore,

$$\frac{W_1}{B_{12}} + \frac{W_2}{B_{21}} = 0$$

$$G_{p12} = \frac{-W_2}{W_1} = \frac{B_{21}}{B_{12}} \quad G_{p21} = \frac{-W_1}{W_2} = \frac{B_{12}}{B_{21}}. \quad (16)$$

From (12) and (13), therefore, we have for the nonlinear capacitor:

Noninverting case

Inverting case

Energy relation

$$\frac{W_I}{f_I} + \frac{W_+}{f_+} = 0 \quad \frac{W_I}{f_I} - \frac{W_-}{f_-} = 0$$

Modulator power gain

$$G_{pl+} = \frac{-W_+}{W_I} = \frac{f_+}{f_I} \quad G_{pl-} = \frac{-W_-}{W_I} = -\frac{f_-}{f_I}$$

Demodulator power gain

$$G_{p+l} = \frac{-W_I}{W_+} = \frac{f_I}{f_+} \quad G_{p-l} = \frac{-W_I}{W_-} = -\frac{f_I}{f_-}. \quad (17)$$

Thus, the small-signal results agree with the general energy relations of Part I, shown in Fig. 1.

Next consider the four-pole terminated in admittances Y_1 and Y_2 and driven by current sources, as shown in Fig. 2. g_1 and g_2 represent the internal generator conductance and the load conductance, and B_1 and B_2 represent the terminal susceptances added for matching purposes. The input admittances are given by

$$\begin{aligned} Y_{1 \text{ in}} &= jB_{11} + \frac{B_{12}B_{21}}{g_2 + j(B_2 + B_{22})} \\ Y_{2 \text{ in}} &= jB_{22} + \frac{B_{12}B_{21}}{g_1 + j(B_1 + B_{11})}. \end{aligned} \quad (18)$$

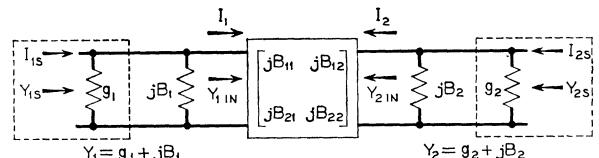


Fig. 2—Linear four-pole driven by current sources.

The admittances seen by the current sources are

$$\begin{aligned} Y_{1s} &= Y_1 + Y_{1 \text{ in}} \\ &= g_1 + j(B_1 + B_{11}) + \frac{B_{12}B_{21}}{g_2 + j(B_2 + B_{22})} \\ Y_{2s} &= Y_2 + Y_{2 \text{ in}} \\ &= g_2 + j(B_2 + B_{22}) + \frac{B_{12}B_{21}}{g_1 + j(B_1 + B_{11})}. \end{aligned} \quad (19)$$

Next we determine the transducer gains G_t ,¹³ defined as the ratio of the power delivered to the load to the available power of the source, for various types of devices. We consider two cases separately.

Case 1

$$\underline{B_{12}B_{21} > 0.}$$

These conditions correspond to the noninverting nonlinear capacitor devices (12). From (16),

$$G_p > 0. \quad (20)$$

For all terminal conditions (18) show that

$$\operatorname{Re} Y_{1\text{ in}} > 0 \quad \operatorname{Re} Y_{2\text{ in}} > 0, \quad (21)$$

so that these devices are always stable.

We next determine the transducer gains for the signal input and output at opposite ports (corresponding to the transducer gains for the noninverting modulator and demodulator). From (16) the power delivered to the load equals the power gain G_p times the power entering the input port of the four-pole. The latter is found in the usual way in terms of the admittances (18) and (19). Thus,

$$G_{t12} = G_{p12} \frac{4 \operatorname{Re} \frac{Y_{1\text{ in}}}{g_1}}{\left| \frac{Y_{1s}}{g_1} \right|^2},$$

$$G_{t21} = G_{p21} \frac{4 \operatorname{Re} \frac{Y_{2\text{ in}}}{g_2}}{\left| \frac{Y_{2s}}{g_2} \right|^2}, \quad (22)$$

where the corresponding power gains G_p are given in (16), or for the nonlinear capacitor in (17). From (20) and (21),

$$G_t > 0. \quad (23)$$

The maximum value of G_t , defined as the maximum available gain, is attained for conjugate match at both ports; the terminal admittances must then satisfy the relation

$$[g_1 + j(B_1 + B_{11})] \cdot [g_2 - j(B_2 + B_{22})] = B_{12}B_{21}$$

or

$$\angle [g_1 + j(B_1 + B_{11})] = \angle [g_2 + j(B_2 + B_{22})]$$

$$|g_1 + j(B_1 + B_{11})| \cdot |g_2 + j(B_2 + B_{22})| = B_{12}B_{21}. \quad (24)$$

In this case (22) becomes $G_t = G_p$, and thus the maximum available gain is equal to the power gain.

Eq. (24) indicates that a wide class of possible terminal admittances will yield a conjugate match at input and output ports. Indeed, for any output admittance whatever, a corresponding input admittance can be found that will match both input and output, and vice versa. The reason for this is easily seen. A four-pole of the special type given in (14) (all terms pure imaginary) must satisfy the general energy relation of (16) or (17).

Since the power gains are independent of the terminal admittances, the transducer gain is maximized and made equal to the power gain by adjusting the generator (load) admittance for match, for any load (generator) admittance. The choice from the infinitely many possible pairs of terminal admittances must be made on grounds other than gain, such as maximizing the bandwidth of the device.

Case 2

$$\underline{B_{12}B_{21} < 0.}$$

These conditions correspond to the inverting nonlinear capacitor devices (13). From (16),

$$G_p < 0. \quad (25)$$

Eq. (18) shows that for all terminal conditions

$$\operatorname{Re} Y_{1\text{ in}} < 0 \quad \operatorname{Re} Y_{2\text{ in}} < 0, \quad (26)$$

so these devices are potentially unstable. Under certain terminal conditions they will oscillate, so that a small-signal analysis is invalid. If the device is stable, the transducer gains G_{t12} and G_{t21} (corresponding to the inverting modulator and demodulator transducer gains) are again given by (22).

Since from (25) and (26) both G_p and $\operatorname{Re} Y_{\text{in}}$ are negative, (22) shows that

$$G_t > 0 \quad (27)$$

as before. In this case, since the input conductance will always be negative (26), match is never possible, and the transducer gain can be made as large as desired by operating close enough to instability, making the factor $|Y_s/g|^2$ in the denominator of (22) small enough.

The operation discussed so far has assumed a generator connected to one port of the four-pole and a load connected to the other. In the nonlinear capacitor devices the separate ports, of course, correspond to the different signal frequencies; the gains correspond to operation as a modulator or demodulator. However, in the present case, ($B_{12}B_{21} < 0$, corresponding to the inverting case), another type of operation is of interest in which gain at a single port (or signal frequency) is obtained by utilizing the negative input conductance of the device.

Assume that a circulator¹⁶ is connected between the generator and the four-pole (or nonlinear capacitor) as shown in Fig. 3, so that the incident and the reflected waves may be separated easily. W_i represents the power in the incident wave, equal to the available power of the source; W_r represents the power in the reflected wave, transmitted to the load by the circulator. W_1 represents the power entering port 1 and W_2 the power entering port 2, as before (both will be negative in the present

¹⁶ A. G. Fox, S. E. Miller, and M. T. Weiss, "Behavior and applications of ferrites in the microwave region," *Bell Sys. Tech. J.*, vol. 34, pp. 5-103; January, 1955.

case). The negative conductance transducer gain at port 1, considering W_r as the useful output, is given by

$$G_{t11} = \frac{W_r}{W_i}.$$

Since the input conductance is negative, the reflection coefficient is greater than unity, and power gain is obtained.

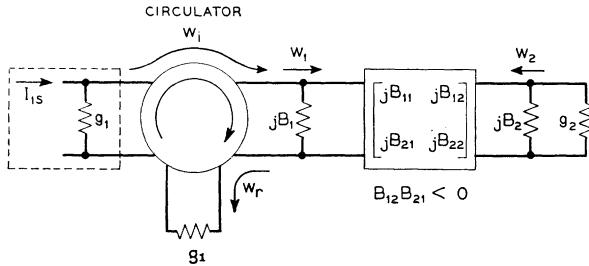


Fig. 3—Linear four-pole operated as a negative conductance amplifier.

From previous discussion the power and transducer gains considering $-W_2$ as the useful output are defined by

$$G_{p12} = \frac{-W_2}{W_1} \quad G_{t12} = \frac{-W_2}{W_i},$$

given in (16) and (22). Conservation of power requires that

$$W_i = W_r + W_1.$$

Then,

$$G_{t11} = 1 - \frac{W_1}{W_i} = 1 - \frac{G_{t12}}{G_{p12}} \quad (28)$$

with a similar expression for G_{t22} , the corresponding negative conductance gain at port 2. From (22) we finally obtain

$$G_{t11} = 1 - \frac{4 \operatorname{Re} \frac{Y_{1\text{in}}}{g_1}}{\left| \frac{Y_{1s}}{g_1} \right|^2} \quad (29)$$

$$G_{t22} = 1 - \frac{4 \operatorname{Re} \frac{Y_{2\text{in}}}{g_2}}{\left| \frac{Y_{2s}}{g_2} \right|^2}.$$

Since $\operatorname{Re} Y_{\text{in}} < 0$ (26), (29) shows that

$$G_t > 1. \quad (30)$$

By making $|Y_s/g|^2$, in the denominator of the second term, sufficiently small the gain can be made as large as desired.

Eq. (29) applies equally well to the first case discussed above ($B_{12}B_{21} > 0$, corresponding to the noninverting case). Here, since $\operatorname{Re} Y_{\text{in}} > 0$, (29) states that $0 < G_t < 1$. Of course we do not expect to obtain gain in this way from a device with a positive input conductance.

NONLINEAR CAPACITOR MODULATORS AND DEMODULATORS

Noninverting Case

Consider a nonlinear capacitor whose terminal admittances consist of a parallel combination of conductance and inductive susceptance, as shown in Fig. 4. Here the two signal frequencies f_l and f_+ are assumed to be brought out at separate terminals by the ideal filter associated with the nonlinear element. The g 's represent the internal generator conductance and the load conductance.

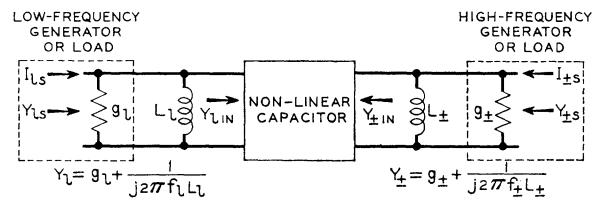


Fig. 4—Nonlinear capacitor with parallel resonant terminal networks.

We now select the terminal admittances to yield a symmetric single-peaked response curve of maximum gain and bandwidth. The terminal admittances must be matched at midband (24); to obtain a symmetric characteristic they are made parallel resonant at midband with their associated self-susceptances of the nonlinear capacitor matrix. Thus,

$$\angle [g_1 + j(B_1 + B_{11})] = \angle [g_2 + j(B_2 + B_{22})] = 0$$

$$B_1 = -B_{11} \quad B_2 = -B_{22}$$

$$g_1 g_2 = B_{12} B_{21}. \quad (31)$$

Denoting the midband frequencies by the subscript m ,

$$f_{l_m} = \text{midband frequency},$$

$$f_l = f_{l_m} + \delta f,$$

$$\delta f = 0 \text{ at midband.} \quad (32)$$

The terminal inductances L_l and L_+ and the terminal conductances g_l and g_+ in Fig. 4 are given by

$$L_l = \frac{1}{4\pi^2 f_{lm}^2 C_0} \quad L_+ = \frac{1}{4\pi^2 f_{+m}^2 C_0};$$

$$g_l g_+ = 4\pi^2 C_1^2 f_{lm} f_{+m}. \quad (33)$$

We may then write $Y_{l_m} = Y_l + Y_{l\text{in}}$, the admittances seen by the current sources in Fig. 4, in the following convenient form.

$$\begin{aligned} \frac{Y_{ls}}{g_l} &= 1 + jQ_l F_l + \frac{\left(1 + \frac{\delta f}{f_{lm}}\right) \left(1 + \frac{\delta f}{f_{+m}}\right)}{1 + jQ_l F_l} \\ \frac{Y_{+s}}{g_+} &= 1 + jQ_+ F_+ + \frac{\left(1 + \frac{\delta f}{f_{lm}}\right) \left(1 + \frac{\delta f}{f_{+m}}\right)}{1 + jQ_+ F_+}, \end{aligned} \quad (34)$$

where

$$\begin{aligned} Q_l &= \frac{2\pi f_{lm} C_0}{g_l} \quad Q_+ = \frac{2\pi f_{+m} C_0}{g_+} \\ F_l &= \frac{f_l}{f_{lm}} - \frac{f_{lm}}{f_l} = \left(1 + \frac{\delta f}{f_{lm}}\right) - \frac{1}{\left(1 + \frac{\delta f}{f_{lm}}\right)} \\ F_+ &= \frac{f_+}{f_{+m}} - \frac{f_{+m}}{f_+} = \left(1 + \frac{\delta f}{f_{+m}}\right) - \frac{1}{\left(1 + \frac{\delta f}{f_{+m}}\right)}. \end{aligned} \quad (35)$$

From (33),

$$Q_l Q_+ = \left(\frac{C_0}{C_1}\right)^2. \quad (36)$$

Maximum bandwidth is obtained for equal load conductances.

$$g_l = g_+ = g = 2\pi C_1 \sqrt{f_{lm} f_{+m}}. \quad (37)$$

By choosing other matched terminal admittances of the form shown in Fig. 4, in accordance with (24), symmetric double-peaked response curves and a wide variety of asymmetric response curves may be obtained. However, the present choice permits the simplest discussion.

In further discussion only the narrow-band case will be considered. Here,

$$\begin{aligned} \left|\frac{\delta f}{f_{lm}}\right| &\ll 1 \\ \left|\frac{\delta f}{f_{+m}}\right| &\ll 1. \end{aligned} \quad (38)$$

Thus we may set the numerator of the second term on the right-hand side of the two equations in (34) equal to 1. The input admittance of the nonlinear capacitor now may be represented by the equivalent circuit shown in Fig. 5, consisting of a parallel resonant circuit and a series resonant circuit in parallel. For the special terminal admittances of (33) both circuits are resonant at midband and the two resistors are equal; the restriction on the terminal conductances in (33) requires that the product of the separate bandwidths of the parallel and

series resonant circuits of the equivalent circuit of Fig. 5 must be a constant. For equal terminal conductances, (37), these two bandwidths become equal. This condition can be seen to maximize the bandwidth over which the input power to the nonlinear capacitor remains constant, and hence to result in the greatest operating bandwidth for the nonlinear capacitor used as a modulator or as a demodulator.

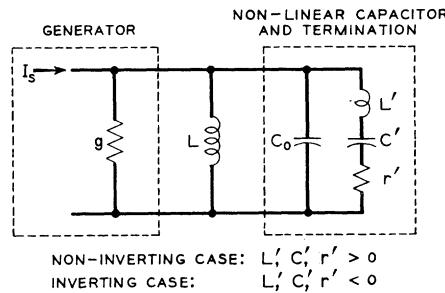


Fig. 5—Equivalent circuit for input admittance of nonlinear capacitor.

Making the terminal conductances equal, (37), and using the narrow-band approximation (38), (34)–(36) become

$$\begin{aligned} \frac{Y_{ls}}{g} &= \frac{Y_{+s}}{g} = (1 + jx) + \frac{1}{(1 + jx)} \\ x &= \frac{4\pi C_0}{g} \delta f = 2 \frac{C_0}{C_1} \frac{1}{\sqrt{f_{lm} f_{+m}}} \delta f. \end{aligned} \quad (39)$$

The modulator and demodulator transducer gains G_t are now determined in the narrow-band case from (22) and (39),

$$G_t = G_p \frac{\frac{1}{1 + jx}}{\left|1 + jx + \frac{1}{1 + jx}\right|^2} = G_p \frac{1}{1 + \left(\frac{x}{\sqrt{2}}\right)^4} \quad (40)$$

with corresponding subscripts on G_t and G_p . The power gains G_p are given by (17) and (32) as

$$\begin{aligned} G_{pl+} &= \frac{f_+}{f_l} = \frac{f_{+m}}{f_{lm}} \frac{1 + \frac{\delta f}{f_{+m}}}{1 + \frac{\delta f}{f_{lm}}} \\ G_{p+l} &= \frac{f_l}{f_+} = \frac{f_{lm}}{f_{+m}} \frac{1 + \frac{\delta f}{f_{lm}}}{1 + \frac{\delta f}{f_{+m}}}. \end{aligned} \quad (41)$$

In the narrow-band approximation we neglect the frequency variation of G_p due to the second factor of (41), and substituting into (40) finally obtain for G_t

¹⁸ The Q 's of (34)–(36) and (51)–(53) should not be confused with the Q 's of (8) and (10).

$$\begin{aligned} G_{tl+} &= \frac{f_{+m}}{f_{lm}} \cdot F(x) & G_{tl+} &= \frac{f_{lm}}{f_{+m}} \cdot F(x) \\ F(x) &= \frac{1}{1 + \left(\frac{x}{\sqrt{2}}\right)^4} & x &= 2 \frac{C_0}{C_1} \frac{\delta f}{\sqrt{f_{lm}f_{+m}}} \end{aligned} \quad (42)$$

for the modulator and demodulator gains, respectively. At midband (42) becomes

$$\left. \begin{aligned} G_{tl+} &= G_{pl+} = \frac{f_{+m}}{f_{lm}} \\ G_{tl+} &= G_{p+l} = \frac{f_{lm}}{f_{+m}} \end{aligned} \right\} \text{for } x = 0. \quad (43)$$

The 3-db bandwidth for both modulator and demodulator now is determined by setting $G_t = \frac{1}{2}G_p$.

$$\begin{aligned} F(x) &= \frac{1}{2} \\ x &= \pm \sqrt{2}. \end{aligned} \quad (44)$$

Then, from (42),

$$\delta f = \pm \frac{1}{2} \frac{C_1}{C_0} \sqrt{2f_{lm}f_{+m}} \quad (45)$$

so that the 3-db bandwidth B for both the noninverting modulator and demodulator becomes

$$\begin{aligned} B &= \frac{C_1}{C_0} \sqrt{2f_{lm}f_{+m}} \\ \frac{B}{f_{lm}} &= \frac{C_1}{C_0} \sqrt{2 \frac{f_{+m}}{f_{lm}}}. \end{aligned} \quad (46)$$

The present narrow-band discussion depends on having a small fractional bandwidth at the low-signal frequency, as stated in (38). Therefore, from (46) we must have

$$\frac{C_1}{C_0} \ll \sqrt{\frac{1}{2} \frac{f_{lm}}{f_{+m}}}. \quad (47)$$

The ratio C_1/C_0 may be taken as a figure of merit for the nonlinear element. If the element becomes linear this quantity approaches 0; the maximum value it can attain for a nonlinear element whose $q-v$ characteristic must always have a positive slope is 1, although this maximum value can be attained only for a highly idealized nonlinear capacitor.¹⁷ Eq. (46) shows that even for a very modest amount of nonlinearity the fractional bandwidth at the low-signal frequency will be large for the usual case where the high-signal frequency is much greater than the low-signal frequency, $(f_+/f_l) \gg 1$. In fact the narrow-band analysis will be invalid for many cases of practical interest, although it gives a general indication of the response. Of course we may always return to

(34) and perform an exact calculation for the cases of interest, but this will not be done here.

Finally, we consider the sensitivity of the device at midband to small changes in either of the terminal conductances or in the nonlinear capacitor parameter C_1 . The sensitivity S is defined as the fractional change in G_t divided by the corresponding fractional change in the parameter causing the variation in gain.¹⁸ Thus, zero sensitivity is desirable, and a high sensitivity undesirable in a device. In the present case the device is matched at midband, and so

$$S = \frac{dG_t/G_t}{dg/g} = \frac{dG_t/G_t}{dC_1/C_1} = 0. \quad (48)$$

Let us consider briefly one numerical example. Let

$$\begin{aligned} f_{lm} &= 100 \text{ mc} & f_{+m} &= 20,000 \text{ mc} \\ f_1 &= 19,900 \text{ mc, local oscillator frequency} \end{aligned}$$

Then the modulator gain at mid-band is

$$G_{tl+} = 200 \quad 10 \log G_{tl+} = +23 \text{ db.}$$

The demodulator gain is

$$G_{tl+} = \frac{1}{200} \quad 10 \log G_{tl+} = -23 \text{ db.}$$

The 3-db bandwidth is given by

$$B = 2000 \frac{C_1}{C_0} \text{ mc} \quad \left(\frac{B}{100} \right)_{\text{mc}} = 20 \frac{C_1}{C_0}.$$

For a value of $C_1/C_0 = 1/20$, corresponding to a very small nonlinearity, the fractional bandwidth at the low-signal frequency, given by the narrow-band approximation, equals unity. Thus, for moderate nonlinearity we expect wide bandwidths from these devices.

Inverting Case

We now repeat the above analysis for the inverting case. Again consider the nonlinear capacitor terminated at the signal frequencies f_l and f_- as shown in Fig. 4. The terminal admittances are now chosen to yield a symmetric single-peaked response curve; they are again made parallel resonant at midband with their associated self-susceptances of the nonlinear capacitor matrix, as in (31). Now, however, the input conductance of the nonlinear capacitor is negative and match is impossible, and so the terminal conductances are no longer restricted by a condition such as the last equation in (31). Denoting the midband frequencies by f_{lm} and f_{-m} as before,

$$\begin{aligned} f_l &= f_{lm} + \delta f \\ f_- &= f_{-m} - \delta f \\ \delta f &= 0 \text{ at midband.} \end{aligned} \quad (49)$$

¹⁷ Torrey and Whitmer, *op. cit.*, pp. 410-411. The proof here is given for a nonlinear conductance, but applies equally well to the present case of a nonlinear capacitance.

¹⁸ H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co., Inc., New York, N. Y.; 1945.

Then the terminal inductances L_i and L_- in Fig. 4 are given by

$$L_i = \frac{1}{4\pi^2 f_{lm}^2 C_0} \quad L_- = \frac{1}{4\pi^2 f_{-m}^2 C_0}. \quad (50)$$

The admittances seen by the current sources in Fig. 4, $\underline{Y}_{ls} = Y_i + Y_{in}$, may then be written

$$\begin{aligned} \frac{Y_{ls}}{g_i} &= 1 + jQ_i F_i - \frac{\alpha \left(1 + \frac{\delta f}{f_{lm}}\right) \left(1 - \frac{\delta f}{f_{-m}}\right)}{1 - jQ_i F_-} \\ \frac{Y_{-s}}{g_-} &= 1 + jQ_i F_- - \frac{\alpha \left(1 + \frac{\delta f}{f_{lm}}\right) \left(1 - \frac{\delta f}{f_{-m}}\right)}{1 - jQ_i F_i}, \end{aligned} \quad (51)$$

where

$$\begin{aligned} Q_i &= \frac{2\pi f_{lm} C_0}{g_i} \quad Q_- = \frac{2\pi f_{-m} C_0}{g_-} \\ F_i &= \frac{f_i}{f_{lm}} - \frac{f_{lm}}{f_i} = \left(1 + \frac{\delta f}{f_{lm}}\right) - \frac{1}{\left(1 + \frac{\delta f}{f_{lm}}\right)} \\ F_- &= \frac{f_-}{f_{-m}} - \frac{f_{-m}}{f_-} = \left(1 - \frac{\delta f}{f_{-m}}\right) - \frac{1}{\left(1 - \frac{\delta f}{f_{-m}}\right)} \quad (52) \\ \alpha &= \frac{4\pi^2 f_{lm} f_{-m} C_1^2}{g_i g_-} = Q_i Q_- \left(\frac{C_1}{C_0}\right)^2. \quad (53) \end{aligned}$$

As discussed below, the parameter α determines the midband gain. For a fixed value of α , maximum bandwidth is again obtained for equal load conductances, as in the noninverting case.

$$g_i = g_- = g = 2\pi C_1 \sqrt{\frac{f_{lm} f_{-m}}{\alpha}}. \quad (54)$$

Considering again the narrow-band case,

$$\begin{aligned} \left| \frac{\delta f}{f_{lm}} \right| &\ll 1 \\ \left| \frac{\delta f}{f_{-m}} \right| &\ll 1, \end{aligned} \quad (55)$$

we set the numerator of the second term on the right-hand side of (51) equal to 1. Thus, the input admittance of the nonlinear capacitor is again represented by the equivalent circuit of Fig. 5, except that all three elements of the series resonant circuit are negative. For the terminal admittances of (50) both circuits are resonant at midband. In this case, the parameter α determines whether this network is stable or unstable,

$$\begin{aligned} 0 < \alpha < 1 &\text{ stable} \\ \alpha > 1 &\text{ unstable}, \end{aligned} \quad (56)$$

as shown by a Nyquist diagram. While this method of studying the stability is not strictly valid, it is assumed to give a reliable indication of stability.

For a fixed value of α the product of the separate bandwidths of the parallel and series circuits of Fig. 5 must be a constant. For equal terminal conductances (54) these two bandwidths become equal. This condition can be seen to maximize the bandwidth over which the power returned by the nonlinear capacitor to the source remains constant, and hence to yield the greatest operating bandwidth for the inverting modulator or demodulator.

For equal terminal conductances (54), in the narrow-band case (55) the admittances (51)–(53) become

$$\begin{aligned} \frac{Y_{ls}}{g} &= \frac{Y_{-s}}{g} = (1 + jx) - \frac{\alpha}{(1 + jx)} \\ x &= \frac{4\pi C_0}{g} \delta f = 2 \frac{C_0}{C_1} \sqrt{\frac{\alpha}{f_{lm} f_{-m}}} \delta f. \end{aligned} \quad (57)$$

Under stable operating conditions, $0 < \alpha < 1$, the modulator and demodulator transducer gains are given by (22) and (57);

$$G_t = G_p \frac{\frac{4 \operatorname{Re} \frac{-\alpha}{1+jx}}{\left|1+jx-\frac{\alpha}{1+jx}\right|^2}}{1+x^2} = -G_p \frac{\frac{4\alpha}{(1-\alpha)^2}}{1+\frac{(3+\alpha^2)}{(1-\alpha)^2}x^2 + \frac{(3+2\alpha)}{(1-\alpha)^2}x^4 + \frac{1}{(1-\alpha)^2}x^6}. \quad (58)$$

Making the usual narrow-band approximations for G_p (17), and noting that G_p is negative, we obtain for the modulator and demodulator gains

$$\begin{aligned} G_{tl-} &= \frac{f_{-m}}{f_{lm}} \frac{4\alpha}{(1-\alpha)^2} G(x, \alpha), \\ G_{t-l} &= \frac{f_{lm}}{f_{-m}} \frac{4\alpha}{(1-\alpha)^2} G(x, \alpha) \\ G(x, \alpha) &= \frac{1+x^2}{1+\frac{(3+\alpha^2)}{(1-\alpha)^2}x^2 + \frac{(3+2\alpha)}{(1-\alpha)^2}x^4 + \frac{1}{(1-\alpha)^2}x^6}, \\ x &= 2 \frac{C_0}{C_1} \sqrt{\frac{\alpha}{f_{lm} f_{-m}}} \delta f. \end{aligned} \quad (59)$$

The midband gains are no longer equal to the ratio of output-to-input frequency, but become

$$\left. \begin{aligned} G_{tl-} &= \frac{f_{-m}}{f_{lm}} \cdot \frac{4\alpha}{(1-\alpha)^2} \\ G_{t-l} &= \frac{f_{lm}}{f_{-m}} \cdot \frac{4\alpha}{(1-\alpha)^2} \end{aligned} \right\} \text{for } x = 0. \quad (60)$$

The factor containing α varies from 0 to ∞ as α varies from 0 to 1, so that the midband transducer gain may be as large as we please. The 3-db bandwidth in both cases is again found by solving

$$G(x, \alpha) = \frac{1}{2} \quad (61)$$

for x and substituting into (59) to determine the bandwidth as a function of α . This cannot be done in closed form because of the complexity of $G(x, \alpha)$, so we will study two special cases to illustrate the general behavior.

First, consider the behavior close to instability, where α is almost equal to 1. Then, the solution to (61) is found to be

$$x = \pm \frac{1}{2}(1 - \alpha). \quad (62)$$

From (59),

$$\delta f = \pm \frac{1}{4} \frac{C_1}{C_0} \sqrt{f_{lm} f_{-m}} (1 - \alpha), \quad (63)$$

so the 3-db bandwidth for both the inverting modulator and demodulator becomes

$$\left. \begin{aligned} B &= \frac{1}{2} \frac{C_1}{C_0} \sqrt{f_{lm} f_{-m}} (1 - \alpha) \\ \frac{B}{f_{lm}} &= \frac{1}{2} \frac{C_1}{C_0} \sqrt{\frac{f_{-m}}{f_{lm}}} (1 - \alpha) \end{aligned} \right\} \text{for } (1 - \alpha) \ll 1. \quad (64)$$

This is similar to (46) for the noninverting case, except for the factor $(1 - \alpha)$. From (64) and (60) we see that as $\alpha \rightarrow 1$ and instability is approached, the gain increases to infinity and the bandwidth decreases to zero. We have

$$\left. \begin{aligned} \sqrt{G_{tl-}} \cdot B &= \frac{C_1}{C_0} f_{-m} \\ \sqrt{G_{t-l}} \cdot B &= \frac{C_1}{C_0} f_{lm} \end{aligned} \right\} \text{for } (1 - \alpha) \ll 1, \quad (65)$$

showing the way in which bandwidth may be exchanged for gain.

Second, consider the behavior when the midband gains are equal to the ratio of output to input frequency, as in the noninverting case. From (60),

$$\frac{4\alpha}{(1 - \alpha)^2} = 1 \quad \alpha = 0.172. \quad (66)$$

The solution to (61) is then,

$$x = \pm 0.513, \quad (67)$$

so that from (59),

$$\delta f = \pm \frac{1.24}{2} \frac{C_1}{C_0} \sqrt{f_{lm} f_{-m}} \quad (68)$$

and the 3-db bandwidth for both modulator and demodulator becomes

$$\left. \begin{aligned} B &= 1.24 \frac{C_1}{C_0} \sqrt{f_{lm} f_{-m}} \\ \frac{B}{f_{lm}} &= 1.24 \frac{C_1}{C_0} \sqrt{\frac{f_{-m}}{f_{lm}}} \end{aligned} \right\} \text{for } \alpha = 0.172. \quad (69)$$

Comparison with (46) shows, for equal midband gains and signal frequencies, the inverting device will have slightly less bandwidth than the noninverting device.

Finally, the sensitivities at midband to small changes in either of the terminal conductances or in the parameter C_1 are given by

$$S_g = \frac{dG_t/G_t}{dg/g} = -\frac{1+\alpha}{1-\alpha}, \quad S_{C_1} = \frac{dG_t/G_t}{dC_1/C_1} = -2S_g. \quad (70)$$

As $\alpha \rightarrow 1$ and instability is approached the sensitivity $\rightarrow \infty$.

Let us take as an example the same signal frequencies used in the previous section to illustrate the noninverting modulator and demodulator, and consider the behavior of the inverting modulator and demodulator for several values of α .

α	$f_{lm} = 100 \text{ mc}$ $f_{-m} = 20,000 \text{ mc}$ $f_1 = 20,100 \text{ mc}$, local oscillator frequency				Sensitivity (terminal conductance) S_g
	Modulator Gain $10 \log G_{tl-}$	Demodulator Gain $10 \log G_{t-l}$	Bandwidth B		
0.172	+23 db	-23 db	1752 C_1/C_0 mc	-1.414	
0.9	+48.5 db	+2.5 db	70.7 C_1/C_0 mc	-19	
0.99	+69 db	+23 db	7.07 C_1/C_0 mc	-199	

In the first case, $\alpha = 0.172$, the modulator and demodulator gains are simply the ratio of output to input frequency, as in the noninverting case. The bandwidth of the inverting modulator and demodulator is slightly less than that of the corresponding noninverting device. By operating close enough to instability the midband gain can be made as large as desired. However, as $\alpha \rightarrow 1$ the bandwidth will become very small and the sensitivity very large. Thus, for $\alpha = 0.99$ an infinitesimal decrease of x per cent in either source or load conductance will cause an increase in gain of $199x$ per cent; a 1 per cent decrease in either terminal conductance or a $\frac{1}{2}$ per cent increase in C_1 (caused perhaps by a slight variation in the local oscillator drive applied to the nonlinear element) will cause the device to become unstable.

The possibility of exchanging bandwidth for gain in the inverting case may permit us to use to better advantage a nonlinear element with a large nonlinearity as an inverting modulator or demodulator than as a noninverting device; in the above example, if C_1/C_0 is close to 1, substantially greater gains at moderate bandwidths can be obtained if a moderate sensitivity is not objectionable. However, we see that a demodulator with high gain will have a very narrow band and a high sensitivity, and such a device would not appear to be too practical.

NONLINEAR CAPACITOR NEGATIVE CONDUCTANCE AMPLIFIERS

We now consider the use of an inverting nonlinear capacitor device as a single-frequency amplifier, by utilizing its negative input conductance. A circulator is assumed connected at the input port corresponding to the frequency to be amplified, as shown in Fig. 3, so that the incident and the (amplified) reflected waves may be separated. The discussion in the last section regarding the choice of terminal admittances to yield a symmetric single-peaked over-all response curve of maximum gain and bandwidth applies equally well to the present case. Thus (49)–(57) remain valid for the negative conductance amplifier; and from (29) and (57), or from (28) and (59) we have

$$G_{ttt} = G_{t--} = 1 + \frac{4\alpha}{(1-\alpha)^2} G(x, \alpha) \quad (71)$$

with $G(x, \alpha)$ as given in (59). The response is the same at both the low and the high signal frequency. At midband the gains become

$$G_{ttt} = G_{t--} = 1 + \frac{4\alpha}{(1-\alpha)^2} = \left(\frac{1+\alpha}{1-\alpha}\right)^2 \text{ for } x=0. \quad (72)$$

This type of operation will be of interest only when α is close to 1 so that substantial gains can be obtained; under these conditions the term unity in (71) and (72) may be neglected compared to the second term, so that the 3-db bandwidths are given approximately by

$$G(x, \alpha) = \frac{1}{2}, \quad (73)$$

which is the same as (61) for the modulator-demodulator bandwidths. Therefore, the bandwidth as a negative conductance amplifier is the same as (64),

$$B = \frac{1}{2} \frac{C_1}{C_0} \sqrt{f_{lm} f_{-m}} (1-\alpha) \text{ for } (1-\alpha) \ll 1; \quad (74)$$

and the gain-bandwidth product is

$$\sqrt{G_t} \cdot B = \frac{C_1}{C_0} \sqrt{f_{lm} f_{-m}} \text{ for } (1-\alpha) \ll 1. \quad (75)$$

The midband sensitivities to small variations in either terminal conductance or in C_1 are

$$S_g = \frac{dG_t/G_t}{dg/g} = - \frac{4\alpha}{1+\alpha} \cdot \frac{1}{1-\alpha} \rightarrow - \frac{2}{1-\alpha} \text{ for } (1-\alpha) \ll 1,$$

$$S_{C_1} = -2S_g, \quad (76)$$

the same as (70) for α close to 1.

In considering a numerical example we choose signal frequencies different than those used in discussing modulators and demodulators, since there is no reason now for one of the signal components to be a low (IF) frequency.

$f_{lm} = 10,000 \text{ mc}$ $f_{-m} = 20,000 \text{ mc}$
 $f_1 = 30,000 \text{ mc}$, local oscillator frequency

α	Negative Conductance Gain $10 \log G_{ttt} = 10 \log G_{t--}$	Bandwidth B	Sensitivity (terminal conductance) S_g
0.9	+25.6 db	707 C_1/C_0 mc	-18.95
0.99	+46.0 db	70.7 C_1/C_0 mc	-199.0

Once again as $\alpha \rightarrow 1$ the gain and sensitivity approach infinity, the bandwidth approaches zero. If the nonlinear element has a reasonably large nonlinearity, moderate gains and bandwidths at either signal frequency can be obtained, with of course a moderate sensitivity to variations in the terminal conductances or in the parameter C_1 .

GENERAL ENERGY RELATIONS FOR LINEAR VARIABLE REACTORS

Since the general energy relations of Part I for nonlinear elements are valid for arbitrary levels at the various frequencies present, they of course must hold when the signal levels are small, as shown in the special case considered here. This has been shown in general for the small-signal case, including all of the frequencies mf_1 , $mf_1 \pm f_0$, by using the general matrix equations of the equivalent linear variable reactor.⁹

We may show, in a simple way, that these general energy relations apply to any lossless linear variable reactor varied periodically in time, for arbitrary signal levels. Consider linear inductors and capacitors that are varied mechanically and have no mechanical or electrical losses. Then (24) and (25) of Part I¹ will remain valid if W_{mn} is now taken to be the sum of both the electrical and mechanical power supplied to the variable reactor at the frequency $|mf_1 + nf_0|$. For convenience let the fundamental frequency at which the reactance is varied be f_1 , and the fundamental frequency of the applied electrical signal f_0 . Then for $n \neq 0$, W_{mn} will contain only electrical energy; mechanical energy can occur only for $n=0$.

First we note that such a variable reactor may be replaced by an equivalent nonlinear reactor, plus an associated local oscillator, which for small signals is identical as far as the signal-frequency terminals are concerned. The nonlinear characteristic and the local oscillator driving waveform, consisting of the frequencies mf_1 , must be chosen so that the equivalent time-varying reactance is identical with that of the original variable reactor, and so that the signal components are small compared to the local oscillator component. There are, of course, a great many ways of doing this, and consequently, there is not necessarily any correspondence whatever between the mechanical powers in the variable reactor and the local oscillator powers in the equivalent nonlinear reactor at the frequencies mf_1 ; further, the limit on the signal levels in order that they will be small compared to the local oscillator level will

obviously depend on the choice of equivalent nonlinear element and local oscillator waveform.

The equivalent nonlinear reactor is, of course, governed by the general energy relations (24) and (25) of Part I. However, (25) contains only signal-frequency terms, since the $n=0$ terms $W_{m,0}$ are absent. Since the two devices are equivalent as far as the signal frequencies are concerned, (25) must apply also to the linear variable reactor for small signals. But since (25) is linear in the $W_{m,n}$ and since the variable reactor is linear in the electrical signals, (25) is valid for arbitrary signal levels. The first energy relation (24) may be obtained now from (25) and the conservation of electrical and mechanical energy.

Thus, the general energy relations of Part I for nonlinear reactors apply to linear variable reactors as well if the $W_{m,n}$ are suitably interpreted to include mechanical as well as electrical power.¹⁹

DISCUSSION

The derivation of the general energy relations for nonlinear capacitors and inductors in Part I depended only on having a single-valued nonlinear characteristic; consequently, these relations tell us nothing about how the amount of nonlinearity affects the operation of the various devices. The present small-signal analysis shows, for two types of nonlinear reactor devices, the limitations imposed by the shape of the nonlinear characteristic. While the energy relations and the gains are independent of this characteristic, the bandwidths are limited and depend on the amount of nonlinearity; as the nonlinearity goes to zero so does the bandwidth. In contrast, for a nonlinear resistor the gain depends on the amount of nonlinearity and must be less than 1, while the bandwidth is limited only by the external circuits.

In the usual case of widely separated signal frequencies the noninverting modulator will have a high gain and the demodulator a high loss. Only a relatively small amount of nonlinearity will be required to attain a bandwidth equal to the low-signal frequency. Since this is a stable device and yields maximum gain for matched terminal admittances, the gain cannot be increased at the expense of bandwidth, as is possible with the inverting device.

The inverting device is potentially unstable and has a negative input conductance; consequently, it can never be matched. Its gain as a modulator or demodulator may be varied from 0 to ∞ by changing its terminal

admittances; the ratio of modulator to demodulator gain remains the same as for the noninverting device. The price of high gain is narrow bandwidth and high sensitivity to changes in the terminal admittances or to changes in C_1 (due to variations in the local oscillator drive). If the modulator and demodulator gains are made the same as in the corresponding noninverting case, so that the modulator gain and demodulator loss are equal, the bandwidth will be almost as large as in the corresponding noninverting modulator and demodulator, so that again only a relatively small amount of nonlinearity is required to yield a bandwidth equal to the low-signal frequency. If a larger nonlinearity is available, then substantially greater gains can be obtained at moderately wide bandwidths and moderate sensitivities. If, however, the gains are increased to the point where the demodulator has a high gain, the bandwidth will be severely limited and the sensitivity very large.

As a negative conductance amplifier, a nonlinear capacitor in the inverting case can provide amplification at a single frequency with moderate gain, bandwidth, and sensitivity; the larger the nonlinearity the larger the bandwidth. Here again gain and bandwidth may be exchanged within wide limits, but very high gains are accompanied by narrow bandwidth and high sensitivity.

Various extensions in the present analysis may be made. First, terminal admittances other than the simple ones studied here may be of interest; increased bandwidth can be obtained if a double-peaked characteristic is acceptable. If several of these devices are cascaded the usual stagger tuning techniques may be of use. Next, the present narrow-band analysis may be extended to the wide-band case in a straightforward manner. Finally, the case where both sidebands $f_1 \pm f_i$ are present may be treated in a similar manner,²⁰ since the terminations at all three signal frequencies now must be specified, the analysis contains several additional parameters.²¹

ACKNOWLEDGMENT

The author wishes to thank J. M. Manley and M. T. Weiss for many helpful discussions and for their detailed suggestions during the preparation of this paper.

²⁰ M. E. Hines, forthcoming paper.

²¹ A related paper by S. Bloom and K. K. N. Chang, "Theory of parametric amplification using nonlinear reactances," *RCA Rev.*, vol. 18, pp. 578-593; December, 1957, came to the author's attention after the present paper had been submitted. It considers the gain and bandwidth of an inverting negative conductance amplifier whose nonlinear element has a quadratic characteristic under different terminal conditions.

¹⁹ A. E. Siegman, private communication.



Power-Frequency Relations in Nonlinear Two-Poles*

JACEK KUDREWICZ†

Summary—In the first part of the paper, some properties of the coefficients of the Fourier-series function $f[v(t)]$ are presented, where $f(v)$ is an arbitrary continuous function and $v(t)$ is an almost-periodic function.

In the second part of the paper, the power-frequency relations for nonlinear two-poles are obtained on the basis of these properties. These power-equilibrium equations are the generalizations of Groszkowski's and Manley-Rowe's equations. The presented, generalized form of these equations comprise as well the elements having loop-like characteristics, nonlinear inductance, capacitance with losses, and others.

THE EQUATIONS describing the relations between the imaginary powers in a resistance nonlinear two-pole were given by J. Groszkowski whose principle of equilibrium of imaginary harmonic powers has since found wide application in the theory of frequency generation and stabilization.¹ Many years later J. M. Manley and E. H. Rowe proposed a similar principle for the effective powers in some special types of lossless nonlinear two-poles.² Manley's and Rowe's equations have been applied in modulation theory and in the theory of magnetic and capacitance amplifiers, and were generalized to cases where the two-pole is excited with signals of several incommensurable frequencies.^{3,4}

Analogically, Groszkowski's equations have been generalized to the case where the signals may be presented by Fourier's integral.⁵

On considering the ideas that produced the power-equilibrium equations, it might be presumed that they resulted from certain general properties of the Fourier coefficients for almost-periodical functions and that, consequently, the very essence of these equations rests on formal mathematical relations which were given a highly elegant and useful physical interpretation. The present investigation is an attempt to bring the Groszkowski and Manley-Rowe principles under a common point of view, and to generalize them into more complex two-poles.

* Received August 13, 1963.

† Institute of Fundamental Technical Problems, Polish Academy of Sciences, Warsaw, Poland.

¹ J. Groszkowski, "Wytwarzanie drgań elektrycznych," ("Generation of Electric Oscillations"), Państwowe Wydawnictwa Techniczne, Warsaw, Poland; 1959. (In Polish.)

² J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—Part I: general relations," Proc. IRE, vol. 44, pp. 904-913; July, 1956.

³ C. Yen, "Generalized energy relations of nonlinear reactive elements," Proc. IRE (Correspondence), vol. 48, p. 253; February, 1960.

⁴ P. A. Clavier, "The Manley-Rowe relations," Proc. IRE (Correspondence), vol. 46, p. 1781; October, 1959.

⁵ J. Groszkowski, "On the extension of the principle of imaginary power balance of harmonics to circuits with continuous spectra," Bull. Acad. Polon. Sci., Classe IV, vol. 2, pp. 131-135, No. 3; 1954.

In communication practice we may find two-poles in the form of pure inductance, capacitance, or nonlinear resistance, as well as such elements, which simultaneously posses the resistive, inductive and capacitive character, but are not simple combinations of uniform two-poles. For example, a two-pole, for which the relationship of current-to-voltage may be written by

$$u(t) = f \left[ri(t) + l \frac{di(t)}{dt} \right],$$

where f is a continuous nonlinear function, while r and l are numerical coefficients, is similar in form to all elements having loop-like characteristics. The generalized form of power-equilibrium equations presented below also comprise the two-poles of that type.

In order to distinguish the properties of Fourier series from the properties of two-poles, this work has been divided into two parts. The first deals with mathematical problems, the second with physical arrangements. The basic idea of the presented work has already been published.⁶

COEFFICIENTS OF THE FOURIER SERIES OF ALMOST-PERIODIC FUNCTIONS

1) An arbitrary almost-periodic function according to H. Bohr⁷ can be represented in the form of a uniformly-convergent trigonometric series

$$v(t) = \sum_{\nu} v_{\nu} e^{i\omega_{\nu} t} \quad (1)$$

where $\omega_1, \omega_2, \omega_3, \dots$ is a sequence of real numbers. The coefficients v_{ν} can be calculated from the formula

$$v_{\nu} = (v, e^{i\omega_{\nu} t}) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T v(\tau) e^{-i\omega_{\nu} \tau} d\tau, \quad (2)$$

the parentheses $(,)$ denoting the scalar product.

In special cases, (2) can be represented in finite form. Thus, e.g., if $v(t)$ is a periodic function of period $T = 2\pi/\omega_0$ we have $\omega_{\nu} = n\omega_0$ and $v_{\nu} = v_n$ in (1), and (2) becomes

$$v_n = \frac{1}{T} \int_0^T v(\tau) e^{-in\omega_0 \tau} d\tau = \frac{1}{2\pi} \int_0^{2\pi} v(\xi) e^{-in\xi} d\xi \quad (3)$$

with

$$v(\xi) = v(\xi).$$

⁶ J. Kudrewicz, "On energy relations in non-linear two poles," Bull. Acad. Polon. Sci., Classe IV, vol. 10, pp. 603-611, No. 10; 1962.

⁷ H. Bohr, "Almost Periodic Functions," Chelsea Publishing Co., New York, N. Y.; 1951.

We now proceed to derive a formula for the Fourier coefficients for the function

$$v(t) = \sum_{n,m} v_{nm} e^{i(n\omega_1+m\omega_2)t}, \quad (4)$$

which is a special case of the one given by (1). We introduce the notation $\omega_1 t = x$, $\omega_2 t = y$, so that

$$v(t) = v(x, y) = \sum_{n,m} v_{nm} e^{i(nx+my)} = \sum_n \left(\sum_m v_{nm} e^{imy} \right) e^{inx},$$

consequently, $v(x, y)$ is of the form of a Fourier series of two variables. Treating it first as series-vs- e^{inx} we may determine the coefficients of its expansion according to

$$\sum_m v_{nm} e^{imy} = \frac{1}{2\pi} \int_0^{2\pi} v(\xi, y) e^{-in\xi} d\xi.$$

Then

$$v_{nm} = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{1}{2\pi} \int_0^{2\pi} v(\xi, \eta) e^{-in\xi} d\xi \right) e^{-im\eta} d\eta.$$

Finally, we get

$$v_{nm} = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} v(\xi, \eta) e^{-i(n\xi+m\eta)} d\xi d\eta. \quad (5)$$

If ω_1/ω_2 is an irrational number there is but one function $v(x, y)$ corresponding exactly to each of the functions of the form given by (4). Eq. (5) is readily generalized to the case of $\omega_r = n_1\omega_1 + n_2\omega_2 + \dots + n_k\omega_k$. The respective coefficients of (1) are now given by k -fold integrals. Hence, if

$$v(t) = v(x_1, x_2, \dots, x_k) = \sum v_{n_1 \dots n_k} e^{i(n_1x_1 + \dots + n_kx_k)} \quad (6)$$

where

$$x_i = \omega_i t,$$

we have

$$v_{n_1 \dots n_k} \simeq \frac{1}{(2\pi)^k} \int_0^{2\pi} \dots \int_0^{2\pi} v(\xi_1, \xi_2, \dots, \xi_k) \cdot e^{-i(n_1\xi_1 + \dots + n_k\xi_k)} d\xi_1 \dots d\xi_k. \quad (7)$$

As in the previous case there is a one-to-one correspondence between $v(t)$ and $v(x_1 \dots x_k)$ if, and only if, all numbers $\omega_1 \dots \omega_k$ are incommensurable.

2) Let f/v be an arbitrary continuous function of one variable defined for all v . On replacing v by an arbitrary almost-periodic function $v(t)$ we obtain the almost-periodic function

$$f[v(t)] = \sum_\mu f_\mu e^{i\omega_\mu t}, \quad (8)$$

where

$$f_\mu = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f[v(\tau)] e^{-i\omega_\mu \tau} d\tau. \quad (9)$$

In some particular cases (9) can be reduced to finite form in the same way as (2) was reduced to the form (3), (5) or (7).

Let us first consider the simplest case where $v(t) = e^{i\omega_0 t} + e^{-i\omega_0 t}$. Here $f[v(t)]$ is a periodic function of fundamental pulsation ω_0 ; therefore, it may be written in the form

$$f[v(t)] = \sum_n f_n e^{in\omega_0 t}. \quad (10)$$

We obtain the same form of $f[v(t)]$ if $v(t)$ is an arbitrary periodic function of fundamental pulsation ω_0 . In the space of all almost-periodic functions, the subspace of periodic functions is invariant with respect to the operation f . Consequently, if $v(t)$ is a periodic function of fundamental pulsation ω_0 , the coefficients f_n of the series of (10) can be calculated from the formula

$$f_n = \frac{1}{2\pi} \int_0^{2\pi} f[v_*(\xi)] e^{-in\xi} d\xi. \quad (11)$$

Let us now consider the more highly involved case where $v(t)$ is of the form of (4) and ω_1/ω_2 is an irrational number. Here, too, the subspace of functions defined by the vectors $e^{i(n\omega_1+m\omega_2)t}$ is invariant with respect to the operation f and $f[v(t)]$ is of the form

$$f[v(t)] = \sum_{n,m} f_{nm} e^{i(n\omega_1+m\omega_2)t}, \quad (12)$$

the coefficients f_{nm} being given by the formula

$$f_{nm} = \frac{1}{2\pi^2} \int_0^{2\pi} \int_0^{2\pi} f[v(\xi, \eta)] e^{-i(n\xi+m\eta)} d\xi d\eta. \quad (13)$$

More generally, the function $v(t)$ can be of the form (6) with the $\omega_1, \omega_2, \dots, \omega_k$ incommensurable. In this case, the Fourier series of $f[v(t)]$ will contain components with the pulsations $n_1\omega_1 + \dots + n_k\omega_k$ for $n_1, \dots, n_k = 0, \pm 1, \dots$ and the formula for the coefficients $f_{n_1 \dots n_k}$ will be derived from (7) in the same way as (13) was derived from (5).

3) We now proceed to the case where both $v(t)$ and $f(v)$ are real functions. In the Fourier expansions of $v(t)$ and $f[v(t)]$, the coefficients corresponding to mutually inverse pulsations will be conjugate; thus, $v_\nu = \bar{v}_{-\nu}$, $\bar{f}_\nu = f_{-\nu}$, with $\omega_\nu = -\omega_{-\nu}$.

For simplification, let us assume $v(t)$ to possess a continuous derivative

$$v'(t) = \sum_\nu j\omega_\nu v_\nu e^{i\omega_\nu t}$$

and the series on the right-hand side of the equation to be uniformly convergent.

Let $v(t)$ be a periodic function of fundamental pulsation ω_0 .

$$v(t) = v_*(\omega_0 t) = \sum_n v_n e^{in\omega_0 t}. \quad (14)$$

Since the subspace of periodic functions is invariant with respect to the operator f , the function $f[v(t)]$ will have the form of (10), and the coefficients f_n will be given by (11). The sum of the series

$$\sum_n n f_n \bar{v}_n = \frac{1}{2\pi} \int_0^{2\pi} f[v_*(\xi)] \sum_n n \bar{v}_n e^{-jn\xi} d\xi$$

may be calculated in the following manner. As

$$\sum (-jn) \bar{v}_n e^{-jn\xi} = \bar{v}_*'(\xi) = v_*'(\xi)$$

we have ultimately

$$\begin{aligned} \sum_n n f_n \bar{v}_n &= \frac{j}{2\pi} \int_0^{2\pi} f[v_*(\xi)] v_*'(\xi) d\xi \\ &= \frac{j}{2\pi} (F[v_*(2\pi)] - F[v_*(0)]), \end{aligned}$$

where F denotes an indefinite integral of f . Since 2π is the period of function V_* the right-hand side of the preceding equation vanishes, and

$$\sum_n n f_n \bar{v}_n = 0. \quad (15)$$

Since $f_{-n} \bar{v}_{-n} = \bar{f}_n v_n$, (15) may be rewritten into the form

$$\sum_{n=0}^{\infty} n \operatorname{Im}(f_n \bar{v}_n) = 0.$$

Let us now consider the case of $v(t)$ given by (4), ω_1/ω_2 being an irrational number. $f[v(t)]$ now takes the form of (12) and the coefficients f_{nm} are given by (13). The sum of the series

$$\sum_{n,m} n f_{nm} \bar{v}_{nm}$$

may be calculated by inverting the order of summation and integration in (13), which gives us

$$\begin{aligned} \sum_{n,m} n f_{nm} \bar{v}_{nm} &= \frac{j}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} f[v(\xi, \eta)] \sum_{n,m} (-jn) \bar{v}_{nm} e^{-j(n\xi+m\eta)} d\xi d\eta \\ &= \frac{j}{4\pi^2} \int_0^{2\pi} d\eta \int_0^{2\pi} \frac{dF(v)}{dv} \frac{\partial v(\xi, \eta)}{\partial \xi} d\xi \\ &= \frac{j}{4\pi^2} \int_0^{2\pi} (F[v(2\pi, \eta)] - F[v(0, \eta)]) d\eta. \end{aligned} \quad (16)$$

The last integrand vanishes because the function (ξ, η) has a period 2π with respect to ξ . Hence, ultimately

$$\sum_{n,m} n f_{nm} \bar{v}_{nm} = 0; \quad (17)$$

and there exists an equality

$$\sum_{n,m} m f_{nm} \bar{v}_{nm} = 0, \quad (18)$$

since it may similarly be shown that

$$\sum_{n,m} m f_{nm} \bar{v}_{nm} = \frac{j}{4\pi^2} \int_0^{2\pi} (F[v(\xi, 2\pi)] - F[v(\xi, 0)]) d\xi.$$

Using the symmetry of coefficients of f_{nm} and v_{nm} , (17) and (18) may be written in the form of

$$\sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} n \operatorname{Im}(f_{nm} \bar{v}_{nm}) = \sum_{n=-\infty}^{\infty} \sum_{m=1}^{\infty} m \operatorname{Im}(f_{nm} \bar{v}_{nm}) = 0.$$

Eqs. (17) and (18) are now readily generalized to the case of $v(t)$ presenting the form (6) and of incommensurable $\omega_1, \omega_2, \dots, \omega_k$.

Making use of the formula for the coefficients f_{n_1, n_2, \dots, n_k} and carrying out transformations analogous to those of (16) we have

$$\sum_{n_1 \dots n_k} n_l f_{n_1 \dots n_k} \bar{v}_{n_1 \dots n_k} = 0 \quad l = 1, 2 \dots k. \quad (19)$$

Power equilibrium equations for nonlinear two-poles

1) We are interested in steady-state effects, assuming the voltage and current in the two-pole to be almost-periodic functions

$$u(t) = \sum_{\nu} u_{\nu} e^{j\omega_{\nu} t} \quad (20a)$$

$$i(t) = \sum_{\nu} i_{\nu} e^{j\omega_{\nu} t}. \quad (20b)$$

(We assume $u(t)$ and $i(t)$ to be real so that $u_{-\nu} = \bar{u}_{\nu}$ and $i_{-\nu} = \bar{i}_{\nu}$, for $-\omega_{-\nu} = \omega_{\nu}$.)

Consider an arbitrary nonlinear two-pole with admittance operator Y , i.e. assume as given a law ascribing the current

$$i(t) = Y u(t)$$

to each value of the voltage $u(t)$. We define the power developed in the network Y as the scalar product of the voltage $u(t)$ and current $i(t)$:

$$p = (i, u) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T i(t) \bar{u}(t) dt. \quad (21)$$

Since both $i(t)$ and $u(t)$ are real-value functions, p is real and represents the effective power developed within the two-pole. By the Fourier expansion (20a), (20b) of $u(t)$ and $i(t)$ and the properties of the scalar product, we have

$$p = (i, u) = \sum_{\nu} i_{\nu} \bar{u}_{\nu} = \sum_{\nu} p_{\nu}, \quad (22)$$

with p_{ν} denoting by definition the product $i_{\nu} \bar{u}_{\nu}$.

Henceforth we shall refer to $p_{\nu} = i_{\nu} \bar{u}_{\nu}$ as the complex power component corresponding to a signal of pulsation ω_{ν} . Obviously $p_{\nu} = \bar{p}_{-\nu}$ whence

$$p = p_0 + 2 \sum_{\nu > 0} \operatorname{Re} p_{\nu}.$$

The last equation expresses the well-known law for summing up the active power values of signals at various frequencies. If, in a particular case, the signals are periodic, then p will be the active power value averaged for the whole time period. (It is also worth stressing that (28) holds if, for arbitrary different indices ν and μ , we have $\omega_\nu \neq \omega_\mu$.)

2) Let us now define the class of two-poles for which the power-equilibrium equations will be derived. Before all we shall assume that this class will comprise the simplest nonlinear two-poles, such as nonlinear resistance, inductance, as well as capacitance. The admittance operators of such two-poles are defined, respectively, by the following:

$$i(t) = f[u(t)], \quad (23a)$$

$$i(t) = f\left[\int_{-\infty}^t u(\tau)d\tau\right], \quad (23b)$$

$$i(t) = \frac{d}{dt}f[u(t)] \quad (23c)$$

where $f(v)$ is a real, nonlinear, continuous, or differentiable function for all real arguments v .

The linear two-poles formed of RLC elements which do not vary in time will be ranked among two-poles of the class considered. As we know, the admittance operator of such a two-pole has the form of

$$Y_u(t) = \sum_v Y(j\omega_v)u_v e^{j\omega_v t} \quad (24)$$

where $Y(j\omega)$ is the admittance of a two-pole, as occurs in usual symbolic calculus.

In the case of two-poles with concentrated RLC parameters, $Y(j\omega)$ is a rational function of the variable $j\omega$ having real coefficients.

Let $[A(j\omega)]$ denote any operator of the form

$$[A(j\omega)]x(t) = \sum_v A(j\omega_v)x_v e^{j\omega_v t}$$

with

$$x(t) = \sum_v x_v e^{j\omega_v t} \quad (25)$$

defined on the subset of almost-periodic functions for which the operation yields an almost-periodic function. Accordingly, we shall denote the admittance operator of (24) by $[Y(j\omega)]$. It is readily shown that

$$\begin{aligned} [A(j\omega)]x(t) &= \int_0^\infty k(\tau)x(t - \tau)d\tau \\ &= \int_{-\infty}^t k(t - \tau)x(\tau)d\tau \end{aligned} \quad (26)$$

where the function $k(\tau)$ is the inverse Fourier transform of $A(j\omega)$. If in the classic sense $k(\tau)$ is nonexistent, it should be interpreted as having the meaning of a distribution.

The following equalities may serve as particular cases of (26).

$$[j\omega]x(t) = \frac{dx(t)}{dt},$$

$$\left[\frac{1}{j\omega}\right]x(t) = \int_{-\infty}^t x(\tau)d\tau.$$

To the considered class of two-poles we finally may add all two-poles for which the relation of current-to-voltage is defined by

$$i(t) = Yu(t) = [A(j\omega)]f[B(j\omega)]u(t). \quad (27)$$

It may be easily checked that the considered class also comprises such two-poles as "nonlinear capacitance with losses" for which it may be written

$$i(t) = C \frac{d}{dt}f(u) + Gf(u) = [j\omega C + G]fu(t). \quad (28)$$

Then, "nonlinear inductance with losses in series" or "shunted losses," i.e. either

$$i(t) = f\left[\frac{1}{l} \int_{-\infty}^t e^{-r(t-\tau)/l}u(\tau)d\tau\right] = f\left[\frac{1}{j\omega l + r}\right]u(t), \quad (29)$$

or

$$i(t) = f\left[\frac{1}{l} \int_{-\infty}^t u(\tau)d\tau + gu(t)\right] = f\left[\frac{1}{j\omega l} + g\right]u(t), \quad (30)$$

as well as several more complex cases.

We now proceed to derive the energy relations for our network and make the following substitution:

$$v(t) = [B(j\omega)]u(t) = \sum_v v_v e^{j\omega_v t},$$

where

$$v_v = B(j\omega_v)u_v.$$

Then,

$$i(t) = \sum_v A(j\omega_v)f_v e^{j\omega_v t},$$

with

$$f_v = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f[v(\tau)]e^{-j\omega_v \tau} d\tau.$$

Hence,

$$p_v = A(j\omega_v)f_v \bar{u}_v = \frac{A(j\omega_v)}{\bar{B}(j\omega_v)} f_v \bar{u}_v$$

and finally

$$f_v \bar{u}_v = \frac{\bar{B}(j\omega_v)}{A(j\omega_v)} p_v. \quad (31)$$

The calculations carried out as above will be the correct ones if $u(t)$, $v(t)$, $f[v(t)]$ and $i(t)$ are almost-periodic

functions. Now it is not a difficult task to formulate, with adequate correctness, the requirements for considerations carried out as above. For example, it is quite sufficient to assume the functions $A(j\omega)$ and $B(j\omega)$ as continuous and bounded for all real values of ω , $f(v)$ as a continuous function, and $u(t)$ as any almost-periodic function. These assumptions may be considerably weakened, assuming the functions $A(j\omega)$ and $B(j\omega)$ are bounded to only such pulsations ω , which occur in the signal spectrum of $u(t)$ and $f[v(t)]$. For the sake of problem simplicity we shall not talk over various variants of these assumptions.

3) Let us first assume our network to operate in a circuit admitting signals of pulsations $n\omega_0$ ($\omega_0 \neq 0$, $n = 0, \pm 1, \dots$) only. Eq. (31) now becomes

$$f_n \tilde{v}_n = \frac{\bar{B}(jn\omega_0)}{A(jn\omega_0)} p_n \quad n = 0, \pm 1, \pm 2, \dots \quad (32)$$

where the relationship between f_n and $v_*(x) = \sum v_n e^{inx}$ is given by (11). Accordingly, by (15)

$$\sum_n n \frac{\bar{B}(jn\omega_0)}{A(jn\omega_0)} p_n = 0 \quad (33)$$

which is the general form of the relationship between the complex powers in a nonlinear two-terminal network.

In a particular case of nonlinear resistance we shall have $A(j\omega) = B(j\omega) = 1$, and the power-equilibrium equation will take the form of

$$\sum_n n p_n = 2j \sum_{n=1}^{\infty} n \operatorname{Im}(i_n \tilde{u}_n) = 0. \quad (34)$$

This equation was first formulated by J. Groszkowski in 1932. It is in wide use while defining the relationship between the frequency of free oscillations and harmonic contents in signals produced by self-excited generators.

Numerous examples of the application of that principle are found in Groszkowski.¹ In practice we find the elements which behave nearly like nonlinear resistances, yet showing some abilities to store energy, e.g., that of the electric field or of magnetic field as an example may serve the two-poles for which the relationship between current and voltage may be defined by either

$$i(t) = Yu(t) = f\left(u(t) + \epsilon \frac{du}{dt}\right) \quad (35)$$

or

$$i(t) = Yu(t) = f\left(u(t) + \epsilon \int_{-\infty}^t u(\tau) d\tau\right). \quad (36)$$

The characteristic of such networks is most commonly represented in the form of a graph of the parametric equations $u(t) = \sin \omega_0 t$, $i(t) = Yu(t)$. The curve is usually a closed loop (generally depending on the parameter ω_0). This method of presenting an element is not an ac-

curate one because of hysteresis loop-form which essentially depends upon the type of signal applied. Hence there is a tendency to find full equations describing a given two-pole.

It is easy to verify that, for example, in the case of given two-poles with admittance operators defined by (35) and (36), the power-equilibrium equations will take the forms

$$\sum_{n=1}^{\infty} n \operatorname{Im}(i_n \tilde{u}_n) = \epsilon \omega_0 \sum_{n=1}^{\infty} n^2 \operatorname{Re}(i_n \tilde{u}_n), \quad (37)$$

$$\sum_{n=1}^{\infty} n \operatorname{Im}(i_n \tilde{u}_n) = -\frac{\epsilon}{\omega_0} \sum_{n=1}^{\infty} \operatorname{Re}(i_n \tilde{u}_n). \quad (38)$$

Groszkowski has proved that the power-equilibrium equations for two-poles having loop characteristics will be of the form

$$\sum_{n=1}^{\infty} n \operatorname{Im}(i_n \tilde{u}_n) = p, \quad (39)$$

where p is the power which is equal to the area enclosed by the hysteresis loop drawn on a u , i coordinate system. Eqs. (37), (38) show the way in which the hysteresis field for various two-poles depends upon active powers of harmonics.

4) Let us now consider the more involved case where signals of pulsations $n_1\omega_1 + n_2\omega_2$ (ω_1/ω_2 being an irrational number) can appear at the terminals. Eq. (31) becomes

$$f_{nm} \tilde{v}_{nm} = \frac{\bar{B}(jn\omega_1 + jm\omega_2)}{A(jn\omega_1 + jm\omega_2)} p_{nm} \quad (40)$$

and the coefficients f_{nm} are given by (13). Thus (17) and (18) lead to

$$\sum_n n \frac{\bar{B}(jn\omega_1 + jm\omega_2)}{A(jn\omega_1 + jm\omega_2)} p_{nm} = 0 \quad (41a)$$

$$\sum_m m \frac{\bar{B}(jn\omega_1 + jm\omega_2)}{A(jn\omega_1 + jm\omega_2)} p_{nm} = 0. \quad (41b)$$

This is the generalized form of power-equilibrium equations for nonlinear two-poles excited with signals having two basic frequencies, ω_1 and ω_2 . In a particular case of nonlinear capacitance and nonlinear inductance, we may have, respectively,

$$i(t) = \frac{d}{dt} f[u(t)] \quad (42a)$$

and

$$i(t) = f\left[\int_{-\infty}^t u(\tau) d\tau\right] \quad (42b)$$

and hence $\bar{B}(j\omega)/A(j\omega) = \pm 1/j\omega$.

On taking into account that $p_{nm} = \bar{p}_{-n,-m}$ we have, by (41a), (41b)

$$\sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{n \operatorname{Re} p_{nm}}{n\omega_1 + m\omega_2} = 0, \quad (43a)$$

$$\sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{m \operatorname{Re} p_{nm}}{n\omega_1 + m\omega_2} = 0 \quad (43b)$$

being the Manley-Rowe equations² of distribution of effective powers in a nonlinear reactance two-pole [as given by (42a), (42b)].

Numerous examples of application of these equations for solving various problems of the modulation theory and of magnetic and capacitance amplifiers were given by Manley and others.

We still wish to draw attention to the generalization of the Manley-Rowe equations for the case of two-poles given by equations of the type

$$i(t) = Gf(u) + C \frac{d}{dt} j(u). \quad (44)$$

This is the form of equation common in "nonlinear" condensers presenting losses. Eq. (38) is a special case of (28) for $A(j\omega) = G + j\omega C$ and $B(j\omega) = 1$; consequently, by (35), the power-equilibrium equations now become

$$\sum_{n,m} \frac{n p_{nm}}{G + j(n\omega_1 + m\omega_2)C} = 0 \quad (45a)$$

$$\sum_{n,m} \frac{m p_{nm}}{G + j(n\omega_1 + m\omega_2)C} = 0 \quad (45b)$$

which represent the Manley-Rowe equations as generalized to two-poles with losses.²

If for all natural numbers n and m there holds an inequality

$$\delta_{nm} = \frac{G}{(n\omega_1 + m\omega_2)C} \ll 1,$$

then (45a), (45b) may be simplified to

$$\begin{aligned} & \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{n}{n\omega_1 + m\omega_2} (\operatorname{Re} p_{nm} - \delta_{nm} \operatorname{Im} p_{nm}) \\ & \approx \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{m}{n\omega_1 + m\omega_2} (\operatorname{Re} p_{nm} - \delta_{nm} \operatorname{Im} p_{nm}) \approx 0. \end{aligned}$$

The power-equilibrium equations are in a high degree generalized; they do not depend upon the character of the nonlinear function $f(u)$, nor do they give much information about two-pole properties. They only represent the power distribution of signals of various frequencies without defining the magnitudes of these powers.

5) To conclude, let us consider the case when signals with pulsations $n_1\omega_1 + n_2\omega_2 + \dots + n_k\omega_k$ (where $\omega_1,$

$\omega_2, \dots, \omega_k$ are incommensurable) appear at the terminals. Eq. (31) now becomes

$$f_{n_1 \dots n_k} \bar{p}_{n_1 \dots n_k} = \frac{\bar{B}(jn_1\omega_1 + \dots + jn_k\omega_k)}{A(jn_1\omega_1 + \dots + jn_k\omega_k)} p_{n_1 \dots n_k} \quad (46)$$

and the respective relationships between the powers are obtained by substituting the foregoing into (19).

$$\sum_{n_1 \dots n_k} n_l \frac{\bar{B}(jn_1\omega_1 + \dots + jn_k\omega_k)}{A(jn_1\omega_1 + \dots + jn_k\omega_k)} p_{n_1 \dots n_k} = 0, \quad l = 1, 2, \dots, k. \quad (47)$$

6) So far we have considered the two-poles with admittance operator, of the form shown in (28). If the function $f(v)$ is a monotonic one, then there does exist the inverse function f^{-1} and the two-pole has an impedance operator defined by

$$u(t) = Zi(t) = [B^{-1}(j\omega)]f^{-1}[A^{-1}(j\omega)]i(t).$$

Nevertheless, it may happen that the function f is not a monotonic one and there is no impedance operator for such a two-pole. Similarly, there exist two-poles having an impedance operator of form

$$Zi(t) = [B^{-1}(j\omega)]\phi[A^{-1}(j\omega)]i(t), \quad (48)$$

but no admittance operator. Because of this ϕ is not a monotonic function, and therefore has no function inversely related to it. It is not difficult to verify that the above derived power-equilibrium equations are correct for two-poles presented by (48) also.

7) In conclusion, it must be remembered that the generalized equations for two-poles with admittance operator, comprising a few nonlinear functions f_a, f_b, \dots , encounter some difficulties. The relations obtained do not relate to two-poles for which, e.g.,

$$Y = [A(j\omega)]f_a + [B(j\omega)]f_b \text{ or } Y = f_a[A(j\omega)]f_b \text{ and so on.}$$

Nevertheless, with some assumptions, the use of power-equilibrium equations may be widened considerably. To discuss these cases we shall introduce a concept of energetically-equivalent two-poles.

Definition: Two-poles having admittance operators Y_1 and Y_2 we shall denote as energetically equivalent if, in every sub-space of almost-periodic functions, their power-equilibrium equations remain identical. From the generalized form of power-equilibrium equations, cf. (47), comes the following conclusion: if

$$Y_1 = [A_1(j\omega)]f_1[B_1(j\omega)] \quad (49a)$$

and

$$Y_2 = [A_2(j\omega)]f_2[B_2(j\omega)] \quad (49b)$$

then the two-poles Y_1 and Y_2 are energetically equivalent if and only if

$$\frac{\bar{B}_1(j\omega)}{A_1(j\omega)} = k \frac{\bar{B}_2(j\omega)}{A_2(j\omega)} \quad (50)$$

where $k \neq 0$ and is any settled number. The fact concerning energetical equivalence does not depend at all upon the function form of f_1, f_2 . Nonlinear capacitance and nonlinear inductance, both having the admittance operator defined by (42a), (42b), may serve as an example of energetically-equivalent two-poles.

Theorem:

Two energetically equivalent two-poles in series and in shunted connection is a two-pole energetically equivalent to their components.

Proof:

The following proof is for the case of a shunted connection; the proof for the connection in series is analogous.

1) Write down the power equilibrium equations for the two two-poles Y_1 and Y_2 . To shorten the entry designate the determinants collection n_1, n_2, \dots, n_k as v and $\omega_v = n_1\omega_1 + \dots + n_k\omega_k$ for Y_1 :

$$\sum_v n_l \frac{\bar{B}_1(j\omega_v)}{A_1(j\omega_v)} i_{1v} \bar{u}_{1v} = 0 \quad l = 1, 2, \dots, k \quad (51a)$$

and for Y_2 :

$$\sum_v n_l \frac{\bar{B}_2(j\omega_v)}{A_2(j\omega_v)} i_{2v} \bar{u}_{2v} = 0 \quad l = 1, 2, \dots, k. \quad (51b)$$

In shunted connection $u_{1v} = u_{2v}$.

2) Multiply both sides of (51b) by k .

3) Add them to (51a) using the equality (50). The following equation is obtained:

$$\sum_v n_l \frac{\bar{B}_1(j\omega_v)}{A_1(j\omega_v)} (i_{1v} + i_{2v}) \bar{u}_{1v} = 0. \quad (52)$$

4) The expression for the resulting two-pole $(i_{1v} + i_{2v}) \bar{u}_{1v} = p_v$ is a component of complex power corresponding to frequency ω_v . From this form of equation comes the resultant two-pole energetically-equivalent to the two-poles Y_1 and Y_2 .

CONCLUSION

The Manley and Rowe equations in their classic formulation are related to any two-pole, which results when nonlinear inductance and capacitance are in series and in shunted connection.

GLOSSARY OF SYMBOLS

$v(t)$ = almost-periodic function.

v_v (and $v_n, v_{nm}, v_{n_1 \dots n_k}$) = Fourier coefficients of $v(t)$.

ω_v = pulsation.

$f(v)$ = continuous function.

f_v (and $f_n, f_{nm}, f_{n_1 \dots n_k}$) = Fourier coefficients of $f[v(t)]$.

$F(v)$ = infinite integral of $f(v)$.

$u(t)$ = voltage.

$i(t)$ = current.

u_v (and $u_n, u_{nm}, u_{n_1 \dots n_k}$) = Fourier coefficients of $u(t)$.

i_v (and $i_n, i_{nm}, i_{n_1 \dots n_k}$) = Fourier coefficients of $i(t)$.

Y = admittance operator.

p_v (and $p_n, p_{nm}, p_{n_1 \dots n_k}$) = complex power corresponding to the signal of pulsation ω_v (and $n\omega_0, n\omega_1 + m\omega_2, n_1\omega_1 + \dots + n_k\omega_k$).

p = power developed in the two-pole.

$Y(j\omega)$ = admittance.

$[A(j\omega)], [B(j\omega)]$ = symbols of the operators.

$A(j\omega), B(j\omega)$ = rational functions of the variable $j\omega$.

$A^{-1}(j\omega) = 1/A(j\omega), B^{-1}(j\omega) = 1/B(j\omega)$.

$k(\tau)$ = inverse Fourier transform of $A(j\omega)$.

C, G, r, l, g, ϵ , = coefficients.

f_a, f_b, ϕ, f_1, f_2 , = nonlinear functions.

f^{-1} = inverse function.

$j = \sqrt{-1}$.

References

- 1 SPANG, H. A., and SCHULTHEISS, P. M.: 'Reduction of quantizing noise by use of feedback', *IRE Trans.*, 1962, CS-10, p. 373
- 2 DAS, J., and CHATTERJEE, P. K.: 'Simple delay-line p.c.m. coder'. See p. 21
- 3 FARUQUI, M. N.: 'A slope-quantized binary pulse code modulation'. To be published in *Electronic Engng.*

PROOF OF THE MANLEY-ROWE RELATIONS FROM QUANTUM CONSIDERATIONS

We feel it important to point out that the communication by J. Brown having the above title¹ does not constitute a proof. This applies also to the modified proof proposed by Brown^{2,3} in response to Valko and Anderson, who have likewise questioned the validity of the original derivation.

For the moment, let us disregard the fact that the simultaneous measurement of frequency (of the quanta) and of time (of quantum arrival) is not possible with arbitrary accuracy, but is subject to an uncertainty relation, since the two are conjugate quantities. Then from a knowledge of one particular quantum flow rate A_{mn} (referring to the notation used by Brown in Reference 3), denoting the flow of quanta of frequency $mf_1 + nf_2$, one could find a time T , such that $A_{mn}T = A'_{mn}$ is an integer. There is no *a priori* reason to assume, however, that in the same period T the number of quanta of another combination frequency $m'f_1 + n'f_2$ would also be integral. Thus, unless very special conditions prevail, $A'_{m'n'}$ will generally be nonintegral for most index pairs $m'n'$ and finite times T . This invalidates Brown's original proof. Furthermore, we doubt that a limiting process can be devised whereby it can be shown that all $A'_{m'n'}$ are integral for infinite T , even though they are nonintegral for any finite T . Thus the burden of the modified proof still rests with Brown.

In addition, we have to consider the fundamental uncertainties inherent in simultaneous frequency and time determinations. In his response to Valko,² Brown suggests quite correctly that there are philosophic difficulties in defining exactly the integral number of quanta emitted at a given frequency, since the time necessary to define that frequency (and to distinguish it from arbitrarily closely lying high-order combination frequencies) is infinite. On the other hand, the counting of quanta requires a localisation in time of events and, consequently, bandwidth. These facts make questionable any physical proof which rests on the premise that the number of quanta may be determined for frequencies of zero linewidth. We thus feel that Brown's defence to Valko is an indictment of his own approach.

Another insight into the nature of the Manley-Rowe relations and into the failure of Brown's proof may be gained if we consider the coefficients A_{mn} (or $A_{mn}T$) no longer as numbers which may be rational or irrational, but rather as functions of frequency.

One might argue within the context of Brown's proof that it is not even necessary for A_{mn} to be integers. Instead, a sufficient assumption for this type of proof would be the statement that A_{mn} are independent of frequencies f_1 and f_2 , because then one could differentiate his eqn. 2 with respect to f_1 and f_2 , a procedure permitted, since f_1 and f_2 are independent variables, and obtain directly the Manley-Rowe relations (eqns. 4 of Reference 1).

Here again, however, we do not find any basis for such an *a priori* assumption. Quite to the contrary, much experimental evidence indicates that the rates of frequency-conversion effects in nonlinear-reactive systems where the Manley-Rowe relations apply (e.g. parametric amplifiers, Raman lasers, optical mixing and many others) do indeed depend on frequency. Thus, in general, the coefficients A_{mn} are functions of the frequencies f_1 and f_2 .*

To what extent are the A_{mn} arbitrary functions of both frequencies? One obvious restriction is energy conservation (eqn. 2 of Reference 1), which we rewrite in the form

$$A_{10}f_{10} + A_{01}f_{01} + \sum_{j=1}^{\infty} A_{m_j n_j} (m_j f_{10} + n_j f_{01}) = 0 \quad (1)$$

where m_j, n_j are pairs of integers, excluding 0, 1 and 1, 0.

A second restriction results from, or is equivalent to, the Manley-Rowe relations. It is only one additional restriction, because the second Manley-Rowe relation derives from energy conservation and the first Manley-Rowe relation. In our notation, the Manley-Rowe relations are

$$A_{10} + \sum_{j=1}^{\infty} m_j A_{m_j n_j} = 0 \quad \dots \dots \dots \dots \dots \dots \quad (2)$$

$$A_{01} + \sum_{j=1}^{\infty} n_j A_{m_j n_j} = 0 \quad \dots \dots \dots \dots \dots \dots \quad (3)$$

We now may use energy conservation and the Manley-Rowe relations to derive the most general frequency dependence of the A_{mn} . This, of course, is a reversed argument. It is no longer attempted to prove the Manley-Rowe relations. Rather, it is shown what necessary requirements the coefficients A_{mn} have to satisfy, such that the Manley-Rowe relations be true. We derive these requirements by differentiating eqn. 1 with respect to f_{10} and f_{01} and observing conditions 2 and 3. This leads to

$$\frac{\partial A_{10}}{\partial f_{10}} f_{10} + \frac{\partial A_{01}}{\partial f_{10}} f_{01} + \sum_{j=1}^{\infty} \frac{\partial A_{m_j n_j}}{\partial f_{10}} (m_j f_{10} + n_j f_{01}) = 0 \quad \dots \dots \dots \dots \quad (4)$$

$$\frac{\partial A_{10}}{\partial f_{01}} f_{10} + \frac{\partial A_{01}}{\partial f_{01}} f_{01} + \sum_{j=1}^{\infty} \frac{\partial A_{m_j n_j}}{\partial f_{01}} (m_j f_{10} + n_j f_{01}) = 0 \quad \dots \dots \dots \dots \quad (5)$$

One verifies that the general solution of eqns. 4 and 5, compatible with eqn. 1, can be given in the form

$$A_{m_j n_j} = g_j(f_{10}, f_{01}) \quad \dots \dots \dots \dots \dots \dots \quad (6a)$$

where the g_j are arbitrary functions of their two independent variables, and

$$A_{10} = - \sum_{j=1}^{\infty} m_j g_j(f_{10}, f_{01}) \quad \dots \dots \dots \dots \dots \dots \quad (6b)$$

$$A_{01} = - \sum_{j=1}^{\infty} n_j g_j(f_{10}, f_{01}) \quad \dots \dots \dots \dots \dots \dots \quad (6c)$$

Of the various possibilities to represent the solutions of eqns. 6, the one given has the advantage of showing directly that energy conservation and the Manley-Rowe relations are satisfied individually for each index j , i.e. for nonvanishing power flow at only three frequencies, characterised, respectively, by the indices 0, 1; 1, 0; and m_j, n_j . This in turn shows that the general Manley-Rowe relations involving an infinite summation can be obtained by summing up the Manley-Rowe relations valid for such frequency triplets.

It may be added that the high degree of arbitrariness expressed in eqns. 6 cannot be reduced in general, but is eliminated only if particular interactions are considered.

In conclusion, we have shown that Brown's approach does not offer a valid proof of the Manley-Rowe relations. For a success of this approach, one would have to show through other considerations that the coefficients A_{mn} can be written in the form of eqns. 6 or its equivalents. This, however, is tantamount to proving the Manley-Rowe relations.

What valid proofs are there, then? It is well to appreciate that one can, at least in principle, construct ideal transducers which change one form of energy into another, while conserving power and frequency. This is significant, since in their usual classical formulation the Manley-Rowe relations are power/frequency relations. Thus, once the uniqueness of power/frequency relations has been established, it suffices to derive the Manley-Rowe relations by considering only one particular type of interaction.

Examples for such proofs are Manley and Rowe's original derivation, in which they treat nonlinear reactances within

* Manley and Rowe⁵ have sketched a proof based on classical network theory, in which they construct an especially chosen ensemble of networks, such that the quantities A_{mn} (in the present notation) are frequency-independent.

the framework of network theory,^{4,5} a treatment in a similar spirit by Salzberg,⁶ a consideration of momentum and energy conservation in the case of waves reflected from moving obstacles, due to Pierce,⁷ the treatment of electromagnetic-wave propagation in nonlinear dielectrics by Haus,⁸ and similarly by Bloembergen and coworkers,⁹ and considerations of torque equality in the shafts of an ideal mechanical differential by Penfield¹⁰ and Seidel;¹¹ many other prototype interactions could be considered for derivation of the Manley-Rowe relations—for example, the conventional induction motor^{10,11} having three distinguished rotational speeds of stator field, rotor motion and slippage.

Within the framework of quantum theory, there is the treatment due to Weiss,¹² in which he showed the equivalence of the Manley-Rowe relations with transition processes, conserving the individual level populations within a three (or more)-level system; Bennett¹³ gave another network theoretical proof, discussed the consequences of embedding the nonlinear elements within a linear network, and, in the realm of quantum theory, showed that the Manley-Rowe relations are equivalent to two physical principles, that of energy conservation and that of detailed balance; finally, the Manley-Rowe relations are, in fact, equivalent to the accepted rules of second quantisation.¹¹

E. O. SCHULZ-DUBOIS

H. SEIDEL

*Bell Telephone Laboratories Inc.
Murray Hill, NJ, USA*

17th December 1965

References

- 1 BROWN, J.: 'Proof of the Manley-Rowe relations from quantum considerations', *Electronics Letters*, 1965, 1, p. 23
- 2 VALKO, I. P., and BROWN, J.: 'Proof of the Manley-Rowe relations from quantum considerations', *ibid.*, 1965, 1, p. 129
- 3 ANDERSON, B. D., and BROWN, J.: 'Proof of the Manley-Rowe relations from quantum considerations', *ibid.*, 1965, 1, p. 199
- 4 MANLEY, J. M., and ROWE, H. E.: 'Some general properties of nonlinear elements', *Proc. Inst. Radio Engrs.*, 1965, 44, p. 904
- 5 MANLEY, J. M., and ROWE, H. E.: 'General energy relations in nonlinear reactances', *ibid.*, 1959, 47, p. 2115
- 6 SALZBERG, B.: 'Masers and reactance amplifiers—basic power relations', *ibid.*, 1957, 45, p. 1544
- 7 PIERCE, J. R.: 'Use of the principle of conservation of energy and momentum in connection with the operation of wave-type parametric amplifiers', *J. Appl. Phys.*, 1959, 30, p. 1341
- 8 HAUS, H. A.: 'Power-flow relations in lossless nonlinear media', *IRE Trans.*, 1959, MTT-6, p. 317
- 9 ARMSTRONG, J. A., BLOEMBERGEN, N., DUCUING, T., and PERSHAN, P. S.: 'Interactions between light waves in a nonlinear dielectric', *Phys. Rev.*, 1962, 127, p. 1918
- 10 PENFIELD, P.: 'Frequency power formulas', (MIT Press and Wiley, New York and London, 1960)
- 11 SEIDEL, H.: To be published
- 12 WEISS, M. T.: 'Quantum derivations of energy relations analogous to those for nonlinear reactances', *Proc. Inst. Radio Engrs.*, 1957, 45, p. 1012
- 13 BENNETT, W. R.: 'Amplification in nonlinear reactive networks' in 'Proceedings of the symposium on active networks and feedback systems' (Polytechnic Press, New York, 1960)

OBSERVATION OF MICROWAVE BEATS IN A PARALLEL-PLATE TRANSMISSION-LINE PHOTOMIXER

A parallel-plate transmission-line photomixer has been used to produce microwave beats at frequencies up to 30–40 Gc/s. These were derived by photoelectric mixing of the optical modes of a ruby laser.

A distributed emission photomixer, which consists of a section of parallel-plate transmission line, one plate of which forms the photocathode, has been used to produce microwave beats at frequencies from about 10 Gc/s up to 30–40 Gc/s. A photograph of one of the photomixers used is shown in Fig. 1.

The photocathode is illuminated along its length with coherent laser light at two closely spaced optical frequencies. Hence the photocurrent flow between the two plates of the transmission line contains a component at the difference frequency ω whose peak emission density can be written as

I . Its phase distribution along the photocathode depends on the angles of incidence of the laser beams.¹ If the phase constant of the current emission β_c is made equal to that of the circuit wave β , the circuit wave grows linearly with distance L along the transmission line. The average circuit-wave power is given by

$$P = \frac{1}{2} I^2 |\Phi(\omega\tau)|^2 Z_0 L^2 \frac{\sin^2(\beta - \beta_c) \frac{L}{2}}{\left\{ (\beta - \beta_c) \frac{L}{2} \right\}^2} \quad . . . \quad (1)$$

Z_0 is the characteristic impedance of the line; $\Phi(\omega\tau)$ is a transit-line reduction factor for the time of flight of electrons τ from cathode to anode. This reduction is rather large for the highest frequencies produced in the present experiments:

$$|\Phi(\omega\tau)|^2 \approx 0.1$$

The photomixers used incorporated S-1 or S-20 photocathodes. Under the conditions of the experiment in which the photocathode was illuminated close to grazing incidence

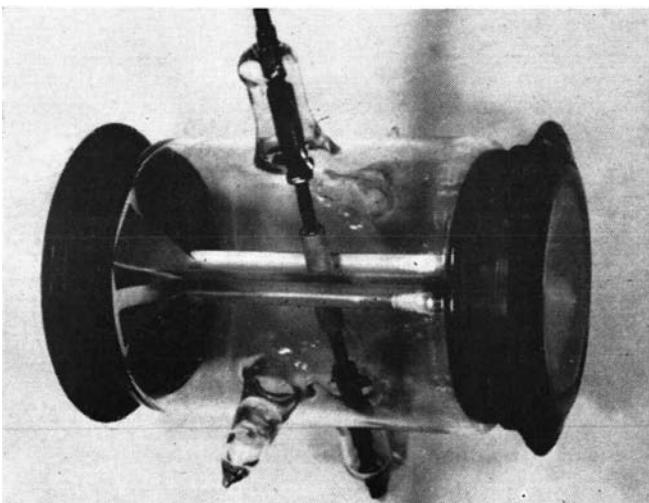


Fig. 1 Distributed emission photomixer with S-1 photocathode

by the laser beams, both types of cathode gave currents of the same magnitude. The quantum efficiency η at normal incidence is much greater for the S-20 photocathode ($\eta(S-20) = 0.04$; $\eta(S-1) = 0.003$), but the emission from the S-20 photocathode diminishes rapidly as grazing incidence is approached, whereas the emission from the S-1 photocathode does not vary greatly with angle owing to the roughness of its surface. The transmission lines are 8 cm long, 8 mm wide, with a separation of 3 mm. Each end of the line is opened out into a horn which feeds the microwave radiation through a glass end window. The ruby lasers used in the experiments consisted of 2 and 3 in rods with dielectric coating on one end, the other mirror being separate from the rod. Pump strengths of up to 3000 J were used. The laser beam is incident onto the photocathode at angles up to about 3° from grazing.

Microwave beats have been observed in frequency bands around 10, 20 and 30 Gc/s. The frequency bandwidth of the detected beats is determined by the frequency response of the detector crystals. In all cases direct detection has been used. A lower limit in frequency is set by the waveguides of the detection systems used in the experiments (6.6 Gc/s, 14 Gc/s, 21.2 Gc/s). No exact measurements of the frequency of the microwave beats have yet been made.

A typical observation of beats is shown in Fig. 2 at frequencies in the 30–40 Gc/s region (Qband). This shows about $250\ \mu s$ of a ruby-laser pulse. The upper trace shows the spikes of d.c. emission in the photomixer, and the lower shows the Qband beats. The largest of the current-emission spikes is about 200 mA. A voltage of 3 kV is applied across the tube. The largest of the Qband beats corresponds to a received power of about $20\ \mu W$. However, the output coupling from the tube is very poor, and a power substantially greater than $20\ \mu W$ may be actually generated along the transmission line.

Proof: By using the following Lemmas and by induction.

Lemma 1 (from the Law of Combinations¹):

$$\begin{aligned} SPC(m, p) A_m \\ = SPC(m - 1, p) A_{m-1} \\ + A_m \cdot SPC(m - 1, p - 1) A_{m-1} \end{aligned}$$

Lemma 2 (from the Laws of Absorption^{2, 3}):

$$\begin{aligned} SPC(m, 1) A_m \cdot SPC(m - 1, 1) A_{m-1} \\ = SPC(m - 1, 1) A_{m-1} \end{aligned}$$

$$\begin{aligned} \text{e.g. } SPC(3, 1) A_3 \cdot SPC(2, 1) A_2 \\ = (a_1 + a_2 + a_3) \cdot (a_1 + a_2) \\ = (a_1 + a_2) \\ = SPC(2, 1) A_2 \end{aligned}$$

Lemma 3 (from the Absorption and Idempotent Laws):

$$\begin{aligned} SPC(m, p) A_m \cdot SPC(m, p - 1) A_m \\ = SPC(m, p) A_m \end{aligned}$$

$$\begin{aligned} \text{e.g. } SPC(3, 2) A_3 \cdot SPC(3, 1) A_3 \\ = (a_1 \cdot a_2 + a_2 \cdot a_3 + a_3 \cdot a_1) \\ \quad \cdot (a_1 + a_2 + a_3) \\ = (a_1 \cdot a_2 + a_2 \cdot a_3 + a_3 \cdot a_1) \\ = SPC(3, 2) A_3 \end{aligned}$$

The proof of the duality theorem by induction consists of two steps:

Step 1: The theorem is true for $p = m$ from the definition:

$$\overline{SPC(m, m) A_m} = SPC(m, 1) A_m$$

Step 2: We must show that Step 1 implies the truth of the theorem for $p = m - 1$ and hence for all p such that $m \geq p \geq 1$. By Lemma 1,

$$\begin{aligned} SPC(m, m - 1) A_m \\ = SPC(m - 1, m - 1) A_{m-1} \\ + a_m \cdot SPC(m - 1, m - 2) A_{m-1} \end{aligned}$$

Thus,

$$\begin{aligned} \overline{SPC(m, m - 1) A_m} \\ = \overline{SPC(m - 1, m - 1) A_{m-1}} \cdot a_m \\ + \overline{SPC(m - 1, m - 1) A_{m-1}} \\ \cdot \overline{SPC(m - 1, m - 2) A_{m-1}} \\ = a_m \cdot SPC(m - 1, 1) A_{m-1} \\ + SPC(m - 1, 1) A_{m-1} \\ \cdot \overline{SPC(m - 1, m - 2) A_{m-1}} \end{aligned}$$

Similarly,

$$\begin{aligned} \overline{SPC(m - 1, m - 2) A_{m-1}} \\ = a_{m-1} \cdot SPC(m - 2, 1) A_{m-2} \\ + SPC(m - 2, 1) A_{m-2} \\ \cdot \overline{SPC(m - 2, m - 3) A_{m-2}} \\ \dots \dots \dots \end{aligned}$$

$$\begin{aligned} \overline{SPC(3, 2) A_3} \\ = a_3 \cdot SPC(2, 1) A_2 \\ + SPC(2, 1) A_2 \cdot \overline{SPC(2, 1) A_2} \\ \overline{SPC(2, 1) A_2} = SPC(2, 2) A_2 \end{aligned}$$

Assembling and using Lemmas 2 and 3 we obtain

$$\begin{aligned} \overline{SPC(m, m - 1) A_m} \\ = a_m \cdot SPC(m - 1, 1) A_{m-1} \\ + a_{m-1} \cdot SPC(m - 2, 1) A_{m-2} \\ + \dots \dots \dots \\ + a_3 \cdot SPC(2, 1) A_2 \\ + SPC(2, 2) A_2 \\ = SPC(m, 2) \end{aligned}$$

and hence the implication is proved.

Thus the theorem is valid for every positive integral value of p such that $m \geq p \geq 1$, and

$$\overline{SPC(m, p) A_m} = SPC(m, m - 1 + p) A_m$$

Discussion

The above theorem could be restated as

$$\overline{SPC(m, p) A_m} = \overline{SPC(m, q) A_m}$$

where $p + q = m + 1$.

Case 1: m even: then $p + q$ is odd and hence p cannot be equal to q .

Case 2: m odd: then $p + q$ is even and we could find $p = q$ such that

$$\overline{SPC(m, p) A_m} = SPC(m, p) A_m$$

which are self duals. Thus, self duality is a property of sets containing odd number of elements $2n + 1$, say.

It is interesting to note that the functions $SPC(3, 2) A_3$, $SPC(5, 3) A_5$, \dots , $SPC(2n + 1, n + 1) A_{2n+1}$ are self duals; these are easily recognised as majority logic functions.

Further, conventional OR and AND logic units are the functions $SPC(2, 1) A_2$ and $SPC(2, 2) A_2$, respectively, which are duals. This suggests that we could call $SPC(m, p) A_m$ and $SPC(m, m - p + 1) A_m$ pseudo-AND and pseudo-OR logic units, corresponding to the cases $2p > m + 1$ and $2p < m + 1$, respectively. When $2p = m + 1$, these units are majority logic units, which are self duals.

It is seen from Lemma 1 that

$$\begin{aligned} SPC(3, 2) A_3 \\ = SPC(2, 2) A_2 + SPC(2, 1) A_2 \cdot a_3 \end{aligned}$$

Hence, AND and OR logic functions could be realised using the majority-logic-function setting a_3 as 0 or 1, respectively, which is a known result.

Incidentally, it is worth observing that the function $SPC(m, p) A_m$ defines a threshold logic with threshold p , and hence its dual has a threshold $(m - p + 1)$.

These dual-logic units play a role in the design of a two-layer error-correcting replicative-logic system. These details are to be presented elsewhere.

E. V. KRISHNAMURTHY 26th February 1965

E. BODFISH

Institute for Computer Research

University of Chicago, Chicago

Ill., USA

References

1 RIORDAN, J.: 'An introduction to combinatorial analysis' (Wiley, 1958)

2 CULBERTSON, J. T.: 'Mathematics and logic for digital devices' (Von Nostrand, 1958)

3 BARTREE, T. C., LEBOW, I. L., and REED, I. S.: 'Theory and design of digital machines' (McGraw-Hill, 1962)

Discussions of the behaviour of parametric amplifiers and related devices are greatly simplified by the Manley-Rowe relations,¹ which provide information on power transfers between the different frequencies arising in the circuit. Weiss has shown that the Manley-Rowe relations can be quickly derived in two special cases by a quantum argument.² The present note provides a general proof, which can be readily extended to more complicated situations than those discussed by Manley and Rowe. The only quantum result used in the proof is that the magnitude of an energy quantum at frequency f is hf , where h is Planck's constant.

Suppose a nonlinear reactance is coupled to energy sources or sinks at a set of frequencies $mf_1 + nf_2$, where m, n are restricted to integral values. It is implicit in this assumption that the ratio f_2/f_1 is irrational, for otherwise it would be possible to express all the frequencies as harmonics of one fundamental frequency. Suppose further that the net power input to the reactance is P_{mn} at frequency $mf_1 + nf_2$; P_{mn} is positive if a power source is operating at this frequency and negative if a power sink is connected in the circuit. The power input P_{mn} can be regarded as provided by A_{mn} quanta per second, each quantum carrying energy $h(mf_1 + nf_2)$. Hence

$$P_{mn} = hA_{mn}(mf_1 + nf_2) \dots \dots \dots (1)$$

Each of the numbers A_{mn} must necessarily be integral.

Since the reactance does not absorb power,

$$\sum_{m, n} P_{mn} = \sum_{m, n} hA_{mn}(mf_1 + nf_2) = 0 \dots \dots \dots (2)$$

where the summations are taken over all pairs of values m, n for which there are sources or sinks. We obtain from eqn. 2 the relation:

$$\sum_{m, n} mA_{mn} + \frac{f_2}{f_1} \sum_{m, n} nA_{mn} = 0 \dots \dots \dots (3)$$

Since the ratio f_2/f_1 is irrational and m, n, A_{mn} are all integers, this equation can only be satisfied if both sums vanish; i.e. if

$$\sum_{m, n} mA_{mn} = \sum_{m, n} nA_{mn} = 0 \dots \dots \dots (4)$$

Substitution for A_{mn} in terms of P_{mn} from eqn. 1 gives

$$\sum_{m, n} \frac{mP_{mn}}{mf_1 + nf_2} = \sum_{m, n} \frac{nP_{mn}}{mf_1 + nf_2} \dots \dots \dots (5)$$

i.e. the Manley-Rowe relations in their usual form.

Extension of the Manley-Rowe relations

This derivation can readily be extended to cover any combinations of frequencies which are not harmonically related. Suppose the sources or sinks are operative at frequencies

$$\sum_{m, n} m_i f_i$$

where all the ratios f_i/f_j are irrational. Retracing the steps of the above argument as far as eqn. 3, we find, with an

PROOF OF THE MANLEY-ROWE RELATIONS FROM QUANTUM CONSIDERATIONS

A general proof using a quantum argument is presented of the Manley-Rowe relations, which provide information on power transfers between different frequencies arising in the circuits of parametric amplifiers and related devices.

obvious extension in notation,

$$\begin{aligned} & \sum_{m_1, m_2, \dots, m_N} m_1 A_{m_1} m_2 \dots m_N \\ & + \frac{f_2}{f_1} \sum_{m_1, m_2, \dots, m_N} m_2 A_{m_1} m_2 \dots m_N + \dots \\ & + \frac{f_N}{f_1} \sum_{m_1, m_2, \dots, m_N} m_N A_{m_1} m_2 \dots m_N \\ & = 0 \dots (6) \end{aligned}$$

Since all ratios f_i/f_j are irrational, each of the sums must vanish to satisfy the equation, and the following set of conditions is obtained:

$$\begin{aligned} & \sum_{m_1, m_2, \dots, m_N} \frac{m_i P_{m_1} m_2 \dots m_N}{m_1 f_1 + m_2 f_2 + \dots + m_N f_N} \\ & = 0; i = 1, 2, \dots, N \dots (7) \end{aligned}$$

I am indebted to an IEE referee for drawing my attention to Reference 2.

J. BROWN 25th February 1965
Indian Institute of Technology
New Delhi-16, India

References

- 1 MANLEY, J. M., and ROWE, H. E.: 'Some general properties of non-linear elements', *Proc. Inst. Radio Engrs.*, 1956, 44, p. 904
- 2 WEISS, M. T.: 'Quantum derivation of energy relations analogous to those for non-linear reactances', *ibid.*, 1957, 45, p. 1012

FIELD EMISSION AT 9000 Gc/s IN A SUPERCONDUCTING CAVITY

The nonlinear current/voltage characteristic of the field-emission diode can be used for frequency multiplication of microwaves.^{1, 2, 3} Field-emission cathodes generally have the shape of very sharp needles. The electrons are emitted from the tip, which has a radius of curvature of the order of 10^{-5} cm. The accelerating field at the surface of the cathode is of the order of 10^7 V/cm. It is concentrated round about the emitting tip; the extent is several tip radii. An application up to very high frequencies should therefore be

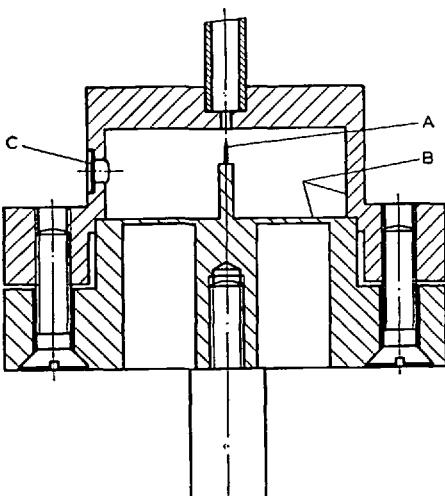


Fig. 1 Re-entrant cavity with a field-emission cathode

possible. Unfortunately, the field-emission cathode has properties which have so far made it impracticable for applications. These properties are the high beam resistance, the instability and short life at poor vacuum conditions and the wide angle of the beam.

In microwave devices the field-emission beam must interact with the electromagnetic field. To solve the matching problem, it was proposed to make the beam resistance low by working at high pulsed-power levels. It was also proposed to use multiple cathodes.¹ We went the other way, making the interaction resistance high, using superconducting cavities. By this means we could obtain a good match of a single cathode beam to the field with a continuous-wave power of less than 1 W. By use of superconductors it is possible to make microwave cavities with Q -factors several orders of magnitude higher than those of normal conducting cavities.^{4, 5} In our experiments we used lead as the superconductor. As the field-emission cathode A (Fig. 1) a sharply etched tungsten wire was used, mounted on the central post of a re-entrant cavity. A lead layer was deposited on the inner surface B of the cavity; it was finally vacuum-tight-soldered and evacuated. Power was coupled through a mica window C. During the measurement, the cavity was immersed in liquid helium. The resonance frequency and coupling could be varied without taking the cavity out of the liquid-helium bath.

Field emission in the cavity was detected by the dependence of Q on the input power P . It is not possible to determine $Q(P)$ of the cavity loaded by field emission directly by measuring the 3db bandwidth or the time constant of the resonator, because the definitions are valid only for a linear network. Instead, we determined $Q(P)$ from the intrinsic $Q(0)$, measured at very low power levels below the threshold for field emission, and from the change of the coupling coefficient by loading the cavity. This method is possible, because the coupling coefficient is proportional to $Q(P)$. From the results, we could calculate the effective beam resistance and the effective beam current.

The current/voltage relation of a field-emission diode can be described in the simplified form:

$$I = C \exp(-\sigma/V)$$

where the constants C and σ depend on the diode geometry and the cathode work function. Under the given circumstances, the following approximation² is valid for the modulated electron beam

$$I(t) = I_m \exp[-\frac{1}{2}\alpha(\omega t)^2]$$

where I_m is the peak current, ω is the modulation frequency, $\alpha = 2\sigma/V_m$ and V_m is the voltage amplitude. From the effective beam current, which is the r.m.s. value, we can calculate the peak current by the relation

$$I_m^2 = \bar{I}^2 / (2\pi\alpha)$$

Plotting $\log I_m$ versus $1/V_m$, we get a linear dependence, which is typical for a field-emission discharge. This relationship could be confirmed by our measurements, as shown in Fig. 2. With a maximum input power of 440 mW, we determined a $Q(440)$ of 68000, whereas the intrinsic $Q(0)$ was 190000; this means that 64% of the input power was coupled into the field-emission beam.

Furthermore, we observed a quite remarkable property of the field-emission device. Even at large current densities of the emission and very poor vacuum (10^{-2} mm Hg), neither instabilities nor vacuum arc occurred. The field emission remained constant up to pressures at which gas discharge ignited. The reason for this is easy to realise. In the alternating field, it is not the ions but the polarised particles which are moving towards the cathode. This implies that no ionic bond is produced at the cathode surface. The cathode has not been flashed in vacuum at all and is contaminated with several layers of oxygen. Helium, the only existing gas at 4.2°K , will therefore certainly not be adsorbed at the cathode. The cathode

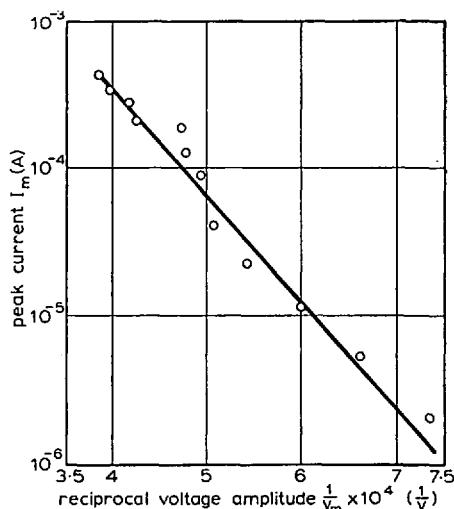


Fig. 2 Field-emission plot experimentally confirmed by measurement of $Q(P)$

cannot be destroyed by vacuum arc because it emits with constant power.

H. ZIMMER 3rd March 1965
Philips Zentrallaboratorium GMBH
Hamburg, Germany

References

- 1 CHARBONNIER, F. M., BARBOUR, J. P., GARRETT, L. F., and DYKE, W. P.: 'Basic and applied studies of field emission at microwave frequencies', *Proc. Inst. Elect. Electronics Engrs.*, 1963, 51, p. 991
- 2 FONTANA, J. R., and SHAW, H. J.: 'Microwave devices with field emission cathodes', *Trans. Amer. Inst. Elect. Engrs.*, 1962, 81, Pt. I, p. 43
- 3 ZIMMER, H.: 'Ein harmonischer Generator zur Erzeugung von Frequenzen über 300 GHz', Philips, Unsere Forschung in Deutschland, 1964, p. 136
- 4 MAXWELL, E.: 'Superconducting resonant cavities', *Progr. Cryogenics*, 1964, 4, p. 123
- 5 ZIMMER, H.: 'Ein supraleitender Resonator hoher Gute bei 9375 MHz', Philips, Unsere Forschung in Deutschland, 1964, p. 134

When do the Manley–Rowe relations really hold?

B. D. O. Anderson, B.Sc., B.E.

Synopsis

Difficulties arising because of a standard assumption in the use of the Manley–Rowe relations, that generator frequencies are incommensurable, are discussed from a physical and mathematical viewpoint, and relations are developed which are valid for any generator frequencies, by making realistic assumptions concerning the generators.

The use of these relations is shown to be equivalent to the use of the standard relations when the assumption is dropped that filters used in the circuitry have infinitely narrow bandwidths.

1 Introduction

The Manley–Rowe relations,¹ now with us for about ten years, have been tested in their application by parametric amplifiers, converters, and the like. Since the original derivation, other derivations^{2,3} have appeared, and, to the best of the author's knowledge, all these derivations require certain frequencies to be incommensurable; i.e. they cannot be expressed as the ratio of two integers.

At first glance, such a requirement seems, *per se*, not entirely in accord with physical reality. The question whether incommensurability is a valid assumption was raised by the author in Reference 4; in this paper I suggest a different requirement that has a more physical motivation, and rederive the relations to incorporate the modification.

While signals at incommensurate frequencies are inevitably (phase) incoherent, this may or may not be the case with waveforms at commensurate frequencies. We propose here that the incommensurability requirement of Manley and Rowe can more accurately be thought of as a requirement that all generators in the circuit should be incoherent, without any restriction on the frequencies of the generators.

To be sure, this implies, in the case of harmonically related generators, that their instantaneous frequencies should be 'almost always' incommensurable (as the irrational numbers are infinitely denser than the rational numbers). But it seems we cannot conclude immediately that the Manley–Rowe relations are valid, since they were derived in terms of steady-state frequencies.

Accordingly, the approach will be to argue on physical grounds that the incommensurability condition should be replaced by an assumption of the incoherency of the generators (Section 2); in Section 3 a modified version of the original proof of the relations is presented, and in Section 4 the practical application of the proof is discussed, resolving points that seem to have been inadequately covered in the literature.

2 Physical considerations

This Section will primarily be concerned with what might be termed the 'causative' elements in a circuit to which it is desired to apply the Manley–Rowe relations, the causative elements being the sources or generators connected to the network.

Sinusoidal generators are common enough devices in practical and theoretical engineering work; commonly their physical behaviour may be modelled by assuming they can be represented by an impedance in series with a voltage source $e = E_0 \sin \omega t$. For most applications this model permits a mathematical analysis, the results of which correspond to physical reality; i.e. the model is perfectly satisfactory. But it is a model, and, in a given situation, the model

may therefore not be perfectly satisfactory, because it is not an accurate enough description of the physical object. For example, it may be important to take into consideration the fact that we cannot in practice produce a discrete line spectrum from a generator; the generator must have been constructed and then switched on at some finite time, and accordingly its frequency spectrum will be nondiscrete. As a second example, and one that will concern us more here, we note that it is impossible to construct a generator of completely stable (i.e. constant) frequency. To say, as we shall, that a generator has frequency ω is really to say that its instantaneous output frequency is approximately ω , and averages out at ω . We could then very reasonably choose to model the generator voltage by $e = \text{Re}(E_0 e^{j\phi} e^{j\omega t})$, where ϕ is a phase angle that varies slowly in time, with a mean zero.

If we consider several generators operating simultaneously, at frequencies $\omega_1, \omega_2, \dots, \omega_p$, where there are no restrictions on the incommensurability or otherwise of ω_i , and if all the generators are generating incoherently, i.e. there is no electrical connection between them that serves as a synchronising device, it is clear that we could model the generators using voltages $e_i = \text{Re}(E_{0i} e^{j\phi_i} e^{j\omega_i t})$, where the ϕ_i could be assumed to vary independently. Thus, in addition to $\bar{\phi}_i = 0$ for each i , we have, for example, $\bar{\phi}_i \bar{\phi}_j = 0$ for each i, j with $i \neq j$, where the averaging is assumed to take place over a sufficiently long time. It is, moreover, reasonable to assume that the random processes described by ϕ_i are ergodic; i.e. ensemble and time averages are the same.

Of course, if there is synchronisation between the generators, this model is invalid. For example, if two generators could be described by $e_1 = \text{Re}(E_1 e^{j\phi_1} e^{j\omega_1 t})$ and $e_2 = \text{Re}(E_2 e^{j2\phi_1} e^{j2\omega_1 t})$, so that $\phi_2 = 2\phi_1$, the expected value of $\phi_1 \phi_2$ is evidently nonzero.

Taking another step toward a consideration of the basic problem, let us now consider a circuit containing a number of sources, sinks and a nonlinear reactive element, without loss of generality assumed capacitive. We shall assume that the sources are a set of incoherent generators of frequencies $\omega_i (i = 1, 2, \dots, p)$. The voltages and currents present in the circuit, and the charge on the capacitor, can be regarded not merely as the superposition of various harmonics, but as the superposition of various harmonics which are somehow dependent on ϕ_i .

If merely one generator is present, we can suppose that the current i through the reactive element is given by

$$i = \sum_{m_1=-\infty}^{+\infty} I_{m_1} e^{jm_1 \phi} e^{jm_1 \omega_1 t} \quad (1)$$

Here m_1 runs through all positive and negative integers, and I_{m_1} is a complex number giving the magnitude of the m_1 th harmonic when the dependence on the phase ϕ , which is a random variable, is separated out. It is not hard to see that the dependence on ϕ is precisely that given, by performing a simple Fourier expansion and recalling that ϕ is assumed to be slowly varying.

Paper 4976 E, first received 20th September and in revised form 20th December 1965
Mr. Anderson is with Stanford Electronics Laboratories, Stanford, Calif., USA

If two generators of the same frequency are present (where by same frequency is really meant the same nominal frequency, but independent, slowly time-varying phases) it is intuitively clear that, for example, the fundamental component of i will certainly contain two components, one depending on ϕ_1 and one depending on ϕ_2 . Harmonics may similarly be broken down into subcomponents, and we may, in general, write

$$i = \sum_{m_1=-\infty}^{+\infty} \sum_{m_2=-\infty}^{+\infty} I_{m_1 m_2} e^{j(m_1 \phi_1 + m_2 \phi_2)} e^{j(m_1 + m_2) \omega_1 t} \quad (2)$$

The coefficients $I_{m_1 m_2}$ in eqn. 2 can be calculated in terms of i in a manner stated explicitly in the next Section. A similar expansion applies for v , the voltage across the nonlinear reactance.

Consider now power flow at, for example, the second-harmonic frequency. This power flow can be computed as the sum of a number of component power flows. For example, the component of i associated with $e^{+2j\phi_1}$, namely $I_{20} e^{2j\phi_1}$, and the component of v associated with $e^{+2j\phi_1}$, namely $V_{20} e^{2j\phi_1}$, will combine together to give a contribution $\text{Re}(I_{20}^* V_{20})$; the component of i associated with $e^{j(\phi_1 + \phi_2)}$, namely $I_{11} e^{j(\phi_1 + \phi_2)}$, and the similar component of v , namely $V_{11} e^{j(\phi_1 + \phi_2)}$, give a contribution $\text{Re}(I_{11}^* V_{11})$, and so on. There will, however, be no power flow associated with 'mixed' components of v and i . Consider, for example, the $e^{2j\phi_1}$ component of v and the $e^{j(\phi_1 + \phi_2)}$ component of i ; the power associated with these will be $\text{Re}(I_{11}^* V_{20} e^{j(\phi_1 - \phi_2)})$ averaged over the ensemble of ϕ_1 and ϕ_2 . Since ϕ_1 and ϕ_2 are independent, however, $\exp j(\phi_1 - \phi_2) = \exp j\phi_1 \exp (-j\phi_2) = 0$.

More generally, the power flow at the r th harmonic frequency may be considered to be the sum of a number of different power flows, one associated with each distinct pair of integers m_1, m_2 , so that $m_1 + m_2 = r$.

At this stage let us observe that, in the classic Manley-Rowe derivation, for two generators, the generator frequencies ω_1 and ω_2 are assumed incommensurable, in order that the various harmonics $m_1 \omega_1 + m_2 \omega_2$ will be distinct (and thus the power flows associated with each will be distinct) for distinct integer pairs m_1, m_2 . We have shown that, for identical ω_1 and ω_2 , and it follows for general ω_1 and ω_2 as long as we maintain our incoherence assumption, associated with a distinct integer pair m_1, m_2 there is defined a power flow at frequency $m_1 \omega_1 + m_2 \omega_2$, irrespective of whether other power flows are possible at this frequency. The requirement of a unique power flow associated with a pair m_1, m_2 is automatically fulfilled in the incommensurable case (as only one pair of m_1, m_2 can define any one frequency). Thus the incoherency assumption is principally for dealing with the general case, where generator frequencies are not necessarily incommensurable.

Accordingly the preceding analysis suggests the following. The incommensurability of generator frequencies in the Manley-Rowe relation is more meaningfully described by saying that the generators are incoherent, i.e. their phases are independent random variables. In the next Section we give a derivation, essentially a modification of Manley and Rowe's proof, of the relations under the incoherence, rather than the incommensurability, assumption.

3 Derivation of the relations

Consider a circuit containing a nonlinear capacitance, together with time-invariant passive elements and incoherent generators at frequencies $\omega_1, \omega_2, \dots, \omega_p$. Suppose the phases of the generators are $\phi_1, \phi_2, \dots, \phi_p$. These generators will establish various values of network variables throughout the circuit at frequencies $\sum m_i \omega_i$ (m_i integers). In particular, we

may write for the instantaneous charge q on the nonlinear capacitor

$$q = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \dots \sum_{m_p=-\infty}^{\infty} Q_{m_1 m_2 \dots m_p} e^{j(m_1 \phi_1 + m_2 \phi_2 + \dots + m_p \phi_p)} e^{j(m_1 + m_2 + \dots + m_p) \omega_1 t} \quad (3)$$

or, in briefer notation,

$$q = \sum_{\{m_i\}} Q_{\{m_i\}} e^{j \sum m_i \phi_i} e^{j \sum m_i \omega_i t} \quad (4)$$

where $Q_{\{m_i\}}$ is determined from q through

$$Q_{\{m_i\}} = \frac{1}{(2\pi)^p} \underbrace{\int_0^{2\pi} \int_0^{2\pi} \dots \int_0^{2\pi}}_p \frac{qe^{-j \sum m_i \phi_i} e^{-j \sum m_i \omega_i t}}{d(\omega_1 t) d(\omega_2 t) \dots d(\omega_p t)} \quad (5)$$

In order to avoid difficulties associated with time-averaging, we assume an ensemble average over ϕ_i here, as in the discussion of power flows in Section 2.

If, for example, $p = 2$ and $\omega_1 = \omega_2$, then Q_{02}, Q_{11} and Q_{20} define three of infinitely many distinct components of q , all associated with the single frequency $2\omega_1$.

Note that, since q is real, we require

$$Q_{\{m_i\}} = Q_{\{-m_i\}}^* \quad (6)$$

We may perform similar expansions on i and v , the current through and the voltage across the capacitance, and obtain:

$$i = \sum_{\{m_i\}} I_{\{m_i\}} e^{j \sum m_i \phi_i} e^{j \sum m_i \omega_i t} \quad (7)$$

where, since $i = dq/dt$, using the slowness of the ϕ_i variation,

$$I_{\{m_i\}} = j Q_{\{m_i\}} \sum_i m_i \omega_i \quad (8a)$$

Also

$$V_{\{m_i\}} = \frac{1}{(2\pi)^p} \underbrace{\int_0^{2\pi} \int_0^{2\pi} \dots \int_0^{2\pi}}_p \frac{ve^{-j \sum m_i \phi_i} e^{-j \sum m_i \omega_i t}}{d(\omega_1 t) d(\omega_2 t) \dots d(\omega_p t)} \quad (8b)$$

Now we may follow Manley and Rowe and multiply both sides of eqn. 8b by $j m_1 Q_{\{m_i\}}^*$. Thus, assuming interchangeability of summation with integration and the averaging operation,

$$\sum_{\{m_i\}} j m_1 Q_{\{m_i\}}^* V_{\{m_i\}} = \frac{1}{(2\pi)^p} \underbrace{\int_0^{2\pi} \int_0^{2\pi} \dots \int_0^{2\pi}}_p \frac{v \sum_{\{m_i\}} j m_1 Q_{\{m_i\}}^* e^{-j \sum m_i \phi_i} e^{-j \sum m_i \omega_i t}}{d(\omega_1 t) d(\omega_2 t) \dots d(\omega_p t)} \quad (9)$$

Using eqn. 3, the reality of q , and eqn. 6, the integrand on the right-hand side of eqn. 9 simplifies so that, with interchange of integration and averaging,

$$\sum_{\{m_i\}} j m_1 Q_{\{m_i\}}^* V_{\{m_i\}} = - \frac{1}{(2\pi)^p} \underbrace{\int_0^{2\pi} \int_0^{2\pi} \dots \int_0^{2\pi}}_p v \frac{dq}{d(\omega_1 t)} d(\omega_1 t) d(\omega_2 t) \dots d(\omega_p t) \quad (10)$$

The derivation is now identical to that of Reference 1. The integral on the right-hand side is readily found to be zero, while the use of eqn. 8a and some easy manipulations on the left-hand side yield

$$\begin{aligned} \sum_{\{m_i\}} j m_1 Q_{\{m_i\}}^* V_{\{m_i\}} &= - \sum_{\{m_i\}} \frac{m_1 V_{\{m_i\}} I_{\{m_i\}}^*}{\sum_i m_i \omega_i} \\ &= - \sum_{\{m_i\}} \frac{m_1 P_{\{m_i\}}}{\sum_i m_i \omega_i} \end{aligned}$$

where $P_{\{m_i\}}$ is the power associated with the components of

v and i that are multiplied by $e^{+j\sum m_i \phi_i}$ in the Fourier decompositions of these quantities.

Thus from eqn. 10 we obtain

$$\sum_{\{m_i\}} \frac{m_i P_{\{m_i\}}}{\sum_i m_i \omega_i} = 0 \quad \quad (11)$$

More generally, it is clear that

$$\sum_{\{m_i\}} \frac{m_k P_{\{m_i\}}}{\sum_i m_i \omega_i} = 0 \quad (k = 1, 2, \dots, p) \quad \quad (12)$$

These are the Manley-Rowe relations as originally derived, provided $P_{\{m_i\}}$ is suitably interpreted in the case where some of the ω_i are not incommensurable.

4 Practical considerations

Consider a circuit where the generator frequencies are nominally not incommensurable. For example, we may have a generator at frequency ω_1 , and one at frequency $\omega_2 = 2\omega_1$, and we may be interested in the power in a sink at frequency $\omega_3 = 3\omega_1$. The relations in eqn. 12 apply, if we recognise that the power at frequency ω_3 is made up of an infinite set of subcomponents, e.g. P_{30}, P_{11}, P_{5-1} etc. In general, we will be unable to find all these components individually, or their sum, by using eqn. 12. How then can eqn. 12 be used in a practical situation?

The answer lies in the fact that even with incommensurable frequencies ω_1, ω_2 a difficulty analogous to that described above can arise, and the elimination of the difficulty, though not always possible, can be carried out in roughly corresponding fashions both when ω_1 and ω_2 are incommensurable and when they are not. Accordingly we shall consider the difficulty that arises in the incommensurable case. The extension to more than two generators is simple, and thus for convenience we shall restrict discussion to the two-generator case.

It is customary, in the usual Manley-Rowe derivations, to predicate the use of ideal filters—ideal in the sense that one and only one frequency is passed. The incoherence of the generated voltages will imply a finite spectral width for each source, and this in turn will require the use of nonideal filters. Moreover, in practice, any filter used will have a finite bandwidth, under any sensible definition of bandwidth, and accordingly we are led to drop the ideal filter assumption. But now, even if ω_1 and ω_2 are incommensurable, any filter that is designed to pass, for example, $\omega_1 + \omega_2$ must also pass components at frequency $m_1\omega_1 + m_2\omega_2$, where m_1 and m_2 are such that $m_1\omega_1 + m_2\omega_2$ and $\omega_1 + \omega_2$ are approximately equal.

Practically, these other components will act to all intents and purposes like noise perturbing the filter output signal at frequency $\omega_1 + \omega_2$. Practically also, the narrower the filter passband, the higher will be the values of m_1 and m_2 which enable $m_1\omega_1 + m_2\omega_2$ to fall within the passband of the filter, and thus also the smaller the magnitude of the perturbing signal. Thus, when we write the Manley-Rowe relations applying, for example, to an upconvertor, and assume that power flow occurs at only three discrete frequencies, we are inherently neglecting power flows at frequencies close to the three considered, on the grounds that the magnitudes of these power flows are negligible.

Of course, this may not be the case in a given situation, but in theory we can certainly analyse a given situation and estimate whether this is the case. Note, however, that the exact nature of the reactance nonlinearity will now be important in such an estimation.

Having dealt with the case of incommensurable ω_1 and ω_2 , it is now relatively easy to see what happens when ω_1 and ω_2 are commensurable. Under these conditions there will be infinitely many components of power at frequency $\omega_1 + \omega_2$, with smaller magnitudes the larger m_1 and m_2 are, where $m_1\omega_1 + m_2\omega_2 = \omega_1 + \omega_2$. Now suppose ω_2 is changed, by

PROC. IEE, Vol. 113, No. 4, APRIL 1966

an extremely small amount, to a frequency ω'_2 incommensurable with ω_1 . Then, with the same m_1, m_2 pairs as above, $m_1\omega_1 + m_2\omega'_2$ will approximately equal $\omega_1 + \omega_2$, differing more from $\omega_1 + \omega_2$ for the larger values of m_1 and m_2 . In any case, a filter whose passband is centred at $\omega_1 + \omega_2$ will pass as many significant components in the incommensurable case as there are significant components in the commensurable case if ω'_2 is chosen sufficiently close to ω_2 .

Our conclusion is therefore that the problem associated with the commensurable frequency case, namely that we cannot determine all components of power at a given frequency, is paralleled by a problem in the incommensurable frequency case, namely that we cannot determine all components of power in a given frequency band.

In both cases we are forced to make an engineering approximation in order to perform calculations, namely that the magnitudes of the embarrassing components are negligible. The validity of this assumption will in part depend on the nature of the reactance nonlinearity and, of course, on the actual frequencies involved.

For example, if we had a signal at frequency ω_1 , a pump at frequency $\omega_2 = 9/8\omega_1$, and desired to examine the output at frequency $\omega_3 = 1/8\omega_1$, in terms of the preceding notation, we see that output power is made up of components including, for example, P_{-11}, P_{8-7} etc., with all possible components save P_{-11} corresponding to large m_1 and/or m_2 . Accordingly we might assume that the only significant component was P_{-11} , and then apply eqn. 12. If ω_2 were not $9/8\omega_1$, but a frequency extremely close to $9/8\omega_1$, so that ω_1 and ω_2 were incommensurable, then we should have to make precisely the same assumption in this case, if the output filter bandwidth were such that it passed, for example, P_{8-7} .

5 Conclusions

To summarise, we observe that the incommensurability condition of the Manley-Rowe relations is more accurately described as a condition on the sources that they be 'incoherent' generators. The use of solely incommensurable frequencies in earlier derivations meant that this was automatically the case; by explicitly postulating the requirement, commensurable frequencies can be considered too. With this assumption the derivation of Manley and Rowe can be modified to yield the result (eqn. 12). The difficulties associated with the existence of an infinite number of power components at certain frequencies, when the exciting frequencies are not incommensurable, highlight an approximation normally made in the application of the relations, namely that certain frequency components are negligible in magnitude. Moreover, this approximation has been masked by the assumption of filters with infinitely narrow passbands.

The validity of the approximation will, in part, depend on the nature of the nonlinearity of the reactive element; qualitatively the approximation amounts to disregarding perturbations of important frequency components by 'noise-like' signals.

6 Acknowledgment

This work was carried out under Joint Services Contract Nonr-225(83). The author would like to acknowledge the financial assistance of the Services Canteens Trust Fund, an agency of the Australian Government, and the United States Educational Foundation in Australia, for a Fulbright travelling grant.

7 References

- 1 MANLEY, J. M., and ROWE, H. E.: 'Some general properties of nonlinear elements', *Proc. Inst. Radio Engrs.*, 1956, 44, p. 904
- 2 RICHARDS, P. I.: 'Manual of mathematical physics' (Pergamon, 1959)
- 3 BROWN, J.: 'Proof of the Manley-Rowe relations from quantum considerations', *Electronics Letters*, 1965, 1, p. 23
- 4 ANDERSON, B. D.: 'Proof of the Manley-Rowe relations from quantum considerations', *ibid.* 1965, 1, p. 199

Conservation relations and the generalized Manley–Rowe relation for the plasma–maser in magnetized plasma

Bipuljyoti Saikia

Centre of Plasma Physics, Saptaswahid Path, Dispur, Guwahati 781 006, India

M. Nambu

Department of Applied Physics, Faculty of Engineering, Kyushu University, Ropponmatsu, Fukuoka 810, Japan

S. Bujarbarua

Centre of Plasma Physics, Saptaswahid Path, Dispur, Guwahati 781 006, India

(Received 23 June 1994; accepted 19 December 1994)

The plasma–maser interaction among Langmuir waves, Bernstein waves, and the electron distribution function on an equal footing is studied. The total momentum and energy conservation relations between electron kinetic energy and wave energy is satisfied for magnetized plasma. The Manley–Rowe relation in its original form is violated because the resonant electrons carry nonlinear momentum and energy. The plasma–maser process satisfies the generalized Manley–Rowe relation which includes the effective action from electrons in addition to the original one from waves. © 1995 American Institute of Physics.

I. INTRODUCTION

The plasma–maser effect is a new nonlinear process in plasma turbulence, first described some time ago.¹ The process has potential importance to interpret numerous radiation phenomena in space and laboratory plasmas.² The main idea of the plasma–maser is to use nonlinear dissipative forces³ for the wave amplification. A slightly different version of the plasma–maser is pointed out in Ref. 4.

The plasma–maser effect occurs when nonresonant as well as resonant plasma oscillations are present. The resonant waves are those for which the Čerenkov resonance condition ($\omega - \mathbf{k} \cdot \mathbf{v} = 0$) is satisfied, while the nonresonant waves are those for which both the Čerenkov and the scattering resonance conditions are not satisfied ($\Omega - \mathbf{K} \cdot \mathbf{v} \neq 0$, $\Omega - \omega - (\mathbf{K} - \mathbf{k}) \cdot \mathbf{v} \neq 0$). Here, ω and Ω are the frequencies of the resonant and nonresonant waves, respectively, and \mathbf{k} and \mathbf{K} are the corresponding wave numbers. Thus, in the plasma–maser process, the three components, i.e., the resonant,⁵ nonresonant⁶ waves and the electrons^{7,8} are strongly coupled to each other. Despite intensive studies on the maser instability,^{1,9} all of the previous works except Ref. 10 are not self-consistent in the sense that at least one component is assumed to be time independent.

The plasma–maser process works at least for two cases. The first one is the open system¹¹ with particles/energy supplied from outside the system. The second case is the magnetized plasmas with a symmetry breaking factor, i.e., finite Larmor radius effect.¹² The polarization term is very sensitive to the symmetry properties of the problem; its vanishing contribution in the case of an unmagnetized plasma is connected with causality as well as homogeneity and isotropy.¹³

Here, we consider the nonlinear interaction among the resonant modes (Langmuir waves), the nonresonant modes (Bernstein waves), and the electron distribution function on an equal footing. This paper shows that the energy and mo-

mentum conservation relations among the three components are exactly satisfied for plasma–maser effect in magnetized plasmas. On the other hand, the Manley–Rowe relation¹⁴ in its original form does not hold. However, the generalized Manley–Rowe relations which include the effective action due to resonant electrons are satisfied. In contrast to the unmagnetized case,⁹ the polarization term is the main nonlinear contribution for magnetized plasma considered here.

The organization of the paper is as follows. In Sec. II, the basic physical quantities of the problem are summarized. The momentum and energy conservation relations for the resonant, nonresonant modes, and the electrons are studied in Sec. III. The violation of the Manley–Rowe relation in its original form is discussed and the generalized Manley–Rowe relation is proposed in Sec. IV. Finally, discussions and conclusions of the present investigation are given in Sec. V.

II. PLASMA–MASER INTERACTION

We consider a homogeneous magnetized plasma in the presence of an enhanced Langmuir wave turbulence driven by a weak electron beam drifting with velocity v_0 through a background plasma of Maxwellian electrons and ions with the following distribution functions:

$$\begin{aligned} f_{0e}(\mathbf{v}) &= (1 - \delta) \left(\frac{m}{2\pi T_b} \right)^{3/2} e^{-mv_\perp^2/2T_b} e^{-mv_\parallel^2/2T_b} \\ &\quad + \delta \left(\frac{m}{2\pi T} \right)^{3/2} e^{-mv_\perp^2/2T} e^{-m(v_\parallel - v_0)^2/2T}, \\ f_{0i}(\mathbf{v}) &= \left(\frac{M}{2\pi T_i} \right)^{3/2} e^{-Mv^2/2T_i}, \end{aligned} \quad (1)$$

where $f_{0e}(\mathbf{v})$ is the electron distribution function, T_b and $T(>T_b)$ are temperatures for the background and beam plasmas, and $\delta=n_b/n_0 \ll 1$ is the ratio of the beam electron density to the background plasma density. Here, $v_0 \gg (2T/m)^{1/2}$ is assumed; f_{0i} is the ion distribution function.

According to previous results,¹⁵ the nonlinear dielectric function of the Bernstein waves (nonresonant waves) in the presence of the Langmuir wave turbulence (resonant waves) is

$$\epsilon_h^n(\mathbf{K}, \Omega) = \epsilon_{dh}(\mathbf{K}, \Omega) + \epsilon_{ph}(\mathbf{K}, \Omega), \quad (2)$$

where $\epsilon_{dh}(\mathbf{K}, \Omega)$ is the direct coupling term,

$$\begin{aligned} \epsilon_{dh}(\mathbf{K}, \Omega) = & \left(\frac{\omega_{pe}}{K_\perp} \right)^2 \left(\frac{e}{m} \right)^2 \sum_{n,a,s} \sum_{\mathbf{k}} |E_l(\mathbf{k}, \omega)|^2 \\ & \times \int \frac{J_a^2(x)}{\Omega - a\Omega_e} \frac{\partial}{\partial v_\parallel} \frac{J_s^2(x)}{\Omega - \omega + k_\parallel v_\parallel - s\Omega_e} \\ & \times \left(\frac{\partial}{\partial v_\parallel} \frac{J_n^2(x)}{\Omega - n\Omega_e} \frac{n\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} \right. \\ & \left. + \frac{s\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} \frac{1}{-\omega + k_\parallel v_\parallel + i0} \frac{\partial}{\partial v_\parallel} \right) f_{0e} d\mathbf{v}, \end{aligned} \quad (3)$$

where $\epsilon_{ph}(\mathbf{K}, \Omega)$ is the polarization coupling term,

$$\begin{aligned} \epsilon_{ph}(\mathbf{K}, \Omega) = & \left(\frac{\omega_{pe}}{K_\perp} \right)^2 \left(\frac{e}{m} \right)^2 \sum_{\mathbf{k}} |E_l(\mathbf{k}, \omega)|^2 \\ & \times \frac{\omega_{pe}^2}{\epsilon_{0h}(\mathbf{K}-\mathbf{k}) |\mathbf{K}-\mathbf{k}|^2} [(A+B) \times (C+D)] \end{aligned} \quad (4)$$

with

$$\begin{aligned} A = & \sum_{a,s} \int \frac{J_a^2(x)}{\Omega - a\Omega_e} \frac{\partial}{\partial v_\parallel} \frac{J_s^2(x)}{\Omega - \omega + k_\parallel v_\parallel - s\Omega_e} \\ & \times \left(\frac{s\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} - k_\parallel \frac{\partial}{\partial v_\parallel} \right) f_{0e} d\mathbf{v}, \\ B = & \sum_a \int \frac{J_a^2(x)}{\Omega - a\Omega_e} \left(\frac{a\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} - k_\parallel \frac{\partial}{\partial v_\parallel} \right) \\ & \times \frac{1}{\omega - k_\parallel v_\parallel + i0} \frac{\partial}{\partial v_\parallel} f_{0e} d\mathbf{v}, \end{aligned} \quad (5)$$

$$\begin{aligned} C = & \sum_{n,s} \int \frac{J_s^2(x)}{\Omega - \omega + k_\parallel v_\parallel - s\Omega_e} \frac{\partial}{\partial v_\parallel} \\ & \times \frac{J_n^2(x)}{\Omega - n\Omega_e} \frac{n\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} f_{0e} d\mathbf{v}, \end{aligned}$$

$$\begin{aligned} D = & \sum_s \int \frac{J_s^2(x)}{\Omega - \omega + k_\parallel v_\parallel - s\Omega_e} \frac{s\Omega_e}{v_\perp} \\ & \times \frac{\partial}{\partial v_\perp} \frac{1}{-\omega + k_\parallel v_\parallel + i0} \frac{\partial}{\partial v_\parallel} f_{0e} d\mathbf{v}, \end{aligned}$$

and

$$\begin{aligned} \epsilon_{0h}(\mathbf{K}-\mathbf{k}) = & 1 + \frac{4\pi e^2}{m|\mathbf{K}-\mathbf{k}|^2} \sum_s \int \frac{J_s^2(x)}{\Omega - \omega + k_\parallel v_\parallel - s\Omega_e} \\ & \times \left(\frac{s\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} - k_\parallel \frac{\partial}{\partial v_\parallel} \right) f_{0e} d\mathbf{v}, \end{aligned} \quad (6)$$

where $\Omega, \mathbf{K}(=K_\perp, 0, 0)$ and $\omega, \mathbf{k}(=0, 0, k_\parallel)$ are the frequencies and wave numbers of the Bernstein and Langmuir waves, respectively, and $E_l(\mathbf{k}, \omega)$ is the electric field of the Langmuir waves. $J_a(x)$ is the Bessel function, where $x = K_\perp v_\perp / \Omega_e$. Other symbols involved in the above expressions have their usual meaning.

The growth rate of the Bernstein wave due to the plasma-maser is,

$$\gamma_h(\mathbf{K}, \Omega) = \frac{\text{Im } \epsilon_h^n(\mathbf{K}, \Omega) + \frac{1}{2} [\partial^2 \epsilon_{0h}(\mathbf{K}, \Omega, t) / \partial \Omega \partial t]}{\partial \epsilon_{0h}(\mathbf{K}, \Omega, t) / \partial \Omega}, \quad (7)$$

where Im shows the imaginary part and $\epsilon_{0h}(\mathbf{K}, \Omega, t)$ is the linear dielectric function of the Bernstein wave given by,

$$\begin{aligned} \epsilon_{0h}(\mathbf{K}, \Omega) = & 1 + \left(\frac{\omega_{pe}}{K_\perp} \right)^2 \\ & \times \sum_n \int \frac{J_n^2(x)}{\Omega - n\Omega_e} \frac{n\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} f_{0e} d\mathbf{v}. \end{aligned} \quad (8)$$

The second term in the numerator of Eq. (7) shows the reverse absorption effect of the plasma-maser, which originates from the slow time change of the medium due to quasilinear process between the resonant mode and the electrons.

Equation (2) shows that the nonlinear dielectric function of the nonresonant waves in a magnetized plasma consists of two terms: the direct and the polarization mode coupling terms. It is well known that for nonresonant electrostatic waves, the polarization mode coupling term vanishes in an unmagnetized plasma.^{12,16} But, for a magnetized plasma considered in this paper, the polarization term gives the most dominant contribution to the nonlinear dielectric function. From Eqs. (3), and (8), it can be shown that,¹⁵

$$\text{Im } \epsilon_{dh}(\mathbf{K}, \Omega) = 0 \quad \text{and} \quad \frac{1}{2} \frac{\partial^2 \epsilon_{0h}(\mathbf{K}, \Omega)}{\partial \Omega \partial t} = 0. \quad (9)$$

Accordingly, the polarization term [$\epsilon_{ph}(\mathbf{K}, \Omega)$] gives the main contribution to the growth rate [Eq. (7)] of the Bernstein waves. This growth rate is given by

$$\frac{\gamma_h(\mathbf{K}, \Omega)}{\Omega} = \sum_a \sum_{\mathbf{k}} \frac{W_l(\mathbf{k}, \omega)}{NT} \frac{\omega_{pe}^6 2a^2 \sqrt{\pi} \delta}{\Omega^4 \beta_a \Omega_e^2} \frac{bk_{||}}{\epsilon_{0h}(\mathbf{K}-\mathbf{k}) |\mathbf{K}-\mathbf{k}|^2 K_\perp^2 \rho_e^2 a_e^2 |k_{||}|} \frac{(k_{||} v_0 - \omega)}{k_{||} v_e} I_a(K_\perp^2 \rho_e^2) e^{-K_\perp^2 \rho_e^2} e^{-m(\omega/k_{||} - v_0)^2/2T} \\ \times \left[1 + \frac{2\Omega^2(\Omega^2 + a^2 \Omega_e^2)}{(\Omega^2 - a^2 \Omega_e^2)^2} \frac{c}{b} + \frac{\Omega^2 I_a(K_\perp^2 a_e^2) e^{-K_\perp^2 a_e^2}}{(\Omega^2 - a^2 \Omega_e^2) b} \left(1 - \frac{4\Omega^2}{\Omega^2 - a^2 \Omega_e^2} \right) \right], \quad (10)$$

with

$$W_l(\mathbf{k}, \omega) = \frac{|E_l(\mathbf{k}, \omega)|^2}{8\pi} \omega \frac{\partial}{\partial \omega} \epsilon_{0l}(\mathbf{k}, \omega) = \frac{|E_l(\mathbf{k}, \omega)|^2}{4\pi}, \\ \beta_a = \frac{2\omega_{pe}^2 I_a(K_\perp^2 a_e^2) e^{-K_\perp^2 a_e^2}}{K_\perp^2 a_e^2 \Omega_e^2}, \\ a_e = \left(\frac{T_b}{m \Omega_e^2} \right)^{1/2}, \quad \rho_e = \left(\frac{T}{m \Omega_e^2} \right)^{1/2}, \quad (11) \\ b = \int_0^\infty J_0^2(x) J_n^2(x) f_{0e}(v_\perp) 2\pi v_\perp dv_\perp, \\ c = \int_0^\infty J_s^2(x) J_n^2(x) f_{0e}(v_\perp) 2\pi v_\perp dv_\perp.$$

The rate of change of energy of the Bernstein waves may be defined as,

$$\frac{\partial}{\partial t} \sum_{\mathbf{k}} W_h = 2 \sum_{\mathbf{k}} \gamma_h(\mathbf{K}, \Omega) W_h. \quad (12)$$

Next, we consider the rate of decrease of the Langmuir wave energy density corresponding to the nonlinear emission of Bernstein mode radiation. The nonlinear dielectric function of the Langmuir wave in the presence of the Bernstein mode turbulence is calculated using standard procedure as,

$$\epsilon_l^n(\mathbf{k}, \omega) = \epsilon_{dl}(\mathbf{k}, \omega) + \epsilon_{pl}(\mathbf{k}, \omega), \quad (13)$$

where $\epsilon_{dl}(\mathbf{k}, \omega)$ is the direct nonlinear coupling term,

$$\epsilon_{dl}(\mathbf{k}, \omega) = -\frac{\omega_{pe}^2}{k_{||}} \left(\frac{e}{m} \right)^2 \sum_{\mathbf{K}} |E_h(\mathbf{K})|^2 \sum_{p, q, g, h} \int \left(\frac{1}{\omega - k_{||} v_{||} + \Omega_e + i0} \frac{e^{i\phi}}{2} + \frac{1}{\omega - k_{||} v_{||} - \Omega_e + i0} \frac{e^{-i\phi}}{2} \right) \\ \times \frac{\partial}{\partial v_\perp} \frac{J_p(x) J_q(x) e^{i(p-q)\phi}}{\omega - \Omega - k_{||} v_{||} + q\Omega_e} \left(\frac{q\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} \frac{1}{\omega - k_{||} v_{||} + i0} \frac{\partial}{\partial v_{||}} + \frac{\partial}{\partial v_{||}} \frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{h\Omega_e - \Omega} \frac{h\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} \right) f_{0e} d\mathbf{v} \quad (14)$$

and $\epsilon_{pl}(\mathbf{k}, \omega)$ is the polarization mode coupling term,

$$\epsilon_{pl}(\mathbf{k}, \omega) = \frac{\omega_{pe}^2}{k_{||}} \left(\frac{e}{m} \right)^2 \sum_{\mathbf{K}} |E_h(\mathbf{K})|^2 \sum_{p, q, g, h} \int \left\{ \frac{k_{||}}{\omega - k_{||} v_{||} + i0} \frac{\partial}{\partial v_{||}} \frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{\Omega - h\Omega_e} \frac{h\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} \right. \\ \left. - \left(\frac{1}{\omega - k_{||} v_{||} + \Omega_e + i0} \frac{e^{i\phi}}{2} + \frac{1}{\omega - k_{||} v_{||} - \Omega_e + i0} \frac{e^{-i\phi}}{2} \right) \frac{\partial}{\partial v_\perp} \left[\frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{\Omega - h\Omega_e} \frac{h\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} - \frac{J_p J_q e^{i(p-q)\phi}}{\omega - \Omega - k_{||} v_{||} + q\Omega_e} \right. \right. \\ \left. \times \left(k_{||} \frac{\partial}{\partial v_{||}} - \frac{q\Omega_e}{v_\perp} \frac{\partial}{\partial v_\perp} \right) \right] \left. f_{0e} d\mathbf{v} \right\} \frac{\omega_{pe}^2}{\epsilon_{0l}(\mathbf{k}-\mathbf{K}) |\mathbf{k}-\mathbf{K}|^2} \int \frac{J_p(x) J_q(x) e^{i(p-q)\phi}}{\omega - \Omega - k_{||} v_{||} + g\Omega_e} \left(\frac{q\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} \frac{1}{\omega - k_{||} v_{||} + i0} \frac{\partial}{\partial v_{||}} \right. \\ \left. + \frac{\partial}{\partial v_{||}} \frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{h\Omega_e - \Omega} \frac{h\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} \right) f_{0e} d\mathbf{v}, \quad (15)$$

where $E_h(\mathbf{K}, \Omega)$ is the electric field of the Bernstein wave, and ϕ is the azimuthal angle. After performing the ϕ integration, we obtain,

$$\epsilon_{dl}(\mathbf{k}, \omega) = 0, \quad (16)$$

$$\epsilon_{pl}(\mathbf{k}, \omega) = \frac{\omega_{pe}^2}{k_{||}} \left(\frac{e}{m} \right)^2 \sum_{\mathbf{K}} |E_h(\mathbf{K}, \Omega)|^2 \sum_{p, g} \int \frac{k_{||}}{\omega - k_{||} v_{||} + i0} \frac{\partial}{\partial v_{||}} \frac{J_g^2(x)}{\Omega - g\Omega_e} \frac{g\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} f_{0e} v_\perp dv_\perp dv_{||} \frac{4\pi^2 \omega_{pe}^2}{\epsilon_{0l}(\mathbf{k}-\mathbf{K}) |\mathbf{k}-\mathbf{K}|^2} \\ \times \int \frac{J_p^2(x)}{\omega - \Omega - k_{||} v_{||} + p\Omega_e} \left(\frac{p\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} \frac{1}{\omega - k_{||} v_{||} + i0} \frac{\partial}{\partial v_{||}} + \frac{\partial}{\partial v_{||}} \frac{J_g^2(x)}{g\Omega_e - \Omega} \frac{g\Omega_e}{K_\perp v_\perp} \frac{\partial}{\partial v_\perp} \right) f_{0e} v_\perp dv_\perp dv_{||}. \quad (17)$$

Thus, it is seen that the evolution of the resonant mode waves is also determined by the polarization mode coupling term. The growth (damping) rate of the resonant Langmuir wave due to the plasma-maser interaction is,

$$\gamma_l(\mathbf{k}, \omega) = -\frac{\text{Im } \epsilon_l^h(\mathbf{k}, \omega)}{\partial \epsilon_{0l}(\mathbf{k}, \omega) / \partial \omega}, \quad (18)$$

where $\epsilon_{0l}(\mathbf{k}, \omega)$ is the linear dielectric function of the Langmuir wave given by

$$\epsilon_{0l}(\mathbf{k}, \omega) = 1 + \frac{\omega_{pe}^2}{k_{\parallel}} \int \frac{1}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} d\mathbf{v}. \quad (19)$$

From Eqs. (17)–(19), we obtain the damping rate of the Langmuir wave as

$$\begin{aligned} \frac{\gamma_l(\mathbf{k}, \omega)}{\omega} = & -\sum_{\mathbf{K}, g} \frac{W_h(\mathbf{K}, \Omega)}{NT} \frac{\omega_{pe}^6 2g^2 \sqrt{\pi} \delta}{\Omega^4 \beta_g \Omega_e^2} \frac{bk_{\parallel}}{\epsilon_{0l}(\mathbf{k}-\mathbf{K}) |\mathbf{k}-\mathbf{K}|^2 K_{\perp}^2 \rho_e^2 a_e^2 |k_{\parallel}|} \frac{(k_{\parallel}v_0 - \omega)}{k_{\parallel}v_e} I_g(K_{\perp}^2 \rho_e^2) e^{-K_{\perp}^2 \rho_e^2} e^{-m(\omega/k_{\parallel} - v_0)^2/2T} \\ & \times \left[\frac{2\Omega^2 I_g(K_{\perp}^2 a_e^2) e^{-K_{\perp}^2 a_e^2}}{\omega^2 b} + 1 + \frac{2\Omega^2 (\Omega^2 + g^2 \Omega_e^2)}{(\Omega^2 - g^2 \Omega_e^2)^2} \frac{c}{b} + \frac{\Omega^2 I_g(K_{\perp}^2 a_e^2) e^{-K_{\perp}^2 a_e^2}}{(\Omega^2 - g^2 \Omega_e^2) b} \left(1 - \frac{4\Omega^2}{\Omega^2 - g^2 \Omega_e^2} \right) \right] \end{aligned} \quad (20)$$

with

$$W_h(\mathbf{K}, \Omega) = \frac{|E_h(\mathbf{K}, \Omega)|^2}{8\pi} \Omega \frac{\partial}{\partial \Omega} \epsilon_{0h}(\mathbf{K}, \Omega) = \sum_g \frac{|E_h(\mathbf{K}, \Omega)|^2}{8\pi} \frac{2g^2 \Omega_e^2 \Omega^2 \beta_g}{(\Omega^2 - g^2 \Omega_e^2)^2}. \quad (21)$$

The rate of change of energy of the Langmuir mode waves may be defined as

$$\frac{\partial}{\partial t} \sum_{\mathbf{k}} W_l = 2 \sum_{\mathbf{k}} \gamma_l(\mathbf{k}, \omega) W_l. \quad (22)$$

Next, we determine the nonlinear effect of the resonant (Langmuir) and nonresonant (Bernstein) waves simultaneously on the evolution of the electron distribution function (the inverse plasma-maser effect^{7,8}). This can be written as

$$\frac{\partial f_{0e}}{\partial t} = \left(\frac{\partial f_{0e}}{\partial t} \right)_I + \left(\frac{\partial f_{0e}}{\partial t} \right)_{II}, \quad (23)$$

where

$$\left(\frac{\partial f_{0e}}{\partial t} \right)_I = \frac{|e|}{m} \sum_{\mathbf{K}} \left\langle \mathbf{E}_h(-\mathbf{K}, -\Omega) \cdot \frac{\partial}{\partial \mathbf{v}} \Delta f_h(\mathbf{K}, \Omega) \right\rangle \quad (24)$$

and

$$\left(\frac{\partial f_{0e}}{\partial t} \right)_{II} = \frac{|e|}{m} \sum_{\mathbf{k}} \left\langle \mathbf{E}_l(-\mathbf{k}, -\omega) \cdot \frac{\partial}{\partial \mathbf{v}} \Delta f_l(\mathbf{k}, \omega) \right\rangle. \quad (25)$$

In Eqs. (24) and (25) $\langle \dots \rangle$ shows the ensemble average over the low-frequency resonant and the high-frequency nonresonant waves, respectively. Thus the time derivative of f_{0e} is composed of two parts. The first one comes from the high-frequency nonlinear response [$\Delta f_h(\mathbf{K}, \Omega)$], and the second one originates from the low-frequency nonlinear response [$\Delta f_l(\mathbf{k}, \omega)$] of the electrons. From Eq. (24), we obtain

$$\begin{aligned} \left(\frac{\partial f_{0e}}{\partial t} \right)_I = & -i \left(\frac{e}{m} \right)^4 \sum_{\mathbf{K}, \mathbf{k}} |E_h(\mathbf{K}, \Omega)|^2 |E_l(\mathbf{k}, \omega)|^2 \frac{1}{K_{\perp}} \frac{\partial}{\partial v_x} \frac{J_a(x) J_b(x) e^{i(a-b)\phi}}{\Omega - b\Omega_e} \left[\frac{\partial}{\partial v_{\parallel}} \frac{J_s(x) J_t(x) e^{i(s-t)\phi}}{\Omega - \omega + k_{\parallel}v_{\parallel} - t\Omega_e} \right. \\ & \times \left(\frac{\partial}{\partial v_{\parallel}} \frac{J_n(x) J_l(x) e^{i(n-l)\phi}}{\Omega - l\Omega_e} \frac{l\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} + \frac{t\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - \frac{1}{-\omega + k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \right) \left. f_{0e} \right] \\ & + i \left(\frac{e}{m} \right)^4 \sum_{\mathbf{K}, \mathbf{k}} |E_h(\mathbf{K}, \Omega)|^2 |E_l(\mathbf{k}, \omega)|^2 \frac{1}{K_{\perp}} \frac{\partial}{\partial v_x} \frac{J_a(x) J_b(x) e^{i(a-b)\phi}}{\Omega - b\Omega_e} \left[\frac{\partial}{\partial v_{\parallel}} \frac{J_s(x) J_t(x) e^{i(s-t)\phi}}{\Omega - \omega + k_{\parallel}v_{\parallel} - t\Omega_e} \left(\frac{t\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - k_{\parallel} \frac{\partial}{\partial v_{\parallel}} \right) \right. \\ & + \left(\frac{b\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - k_{\parallel} \frac{\partial}{\partial v_{\parallel}} \right) \frac{1}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \left. f_{0e} \right] \frac{\omega_{pe}^2}{\epsilon_{0h}(\mathbf{K}-\mathbf{k}) |\mathbf{K}-\mathbf{k}|^2} \int \frac{J_s(x) J_t(x) e^{i(s-t)\phi}}{\Omega - \omega + k_{\parallel}v_{\parallel} - t\Omega_e} \\ & \times \left(\frac{\partial}{\partial v_{\parallel}} \frac{J_n(x) J_l(x) e^{i(n-l)\phi}}{\Omega - l\Omega_e} \frac{l\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} + \frac{t\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - \frac{1}{-\omega + k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \right) f_{0e} d\mathbf{v}. \end{aligned} \quad (26)$$

Similarly, from Eq. (25), we obtain

$$\begin{aligned}
\left(\frac{\partial f_{0e}}{\partial t} \right)_{\text{II}} = & -i \left(\frac{e}{m} \right)^4 \sum_{\mathbf{K}, \mathbf{k}} |E_l(\mathbf{k}, \omega)|^2 |E_h(\mathbf{K}, \Omega)|^2 \frac{\partial}{\partial v_{\parallel}} \left(\frac{1}{\omega - k_{\parallel}v_{\parallel} + \Omega_e + i0} \frac{e^{i\phi}}{2} + \frac{1}{\omega - k_{\parallel}v_{\parallel} - \Omega_e + i0} \frac{e^{-i\phi}}{2} \right) \\
& \times \frac{\partial}{\partial v_{\perp}} \frac{J_p(x) J_q(x) e^{i(p-q)\phi}}{\omega - \Omega - k_{\parallel}v_{\parallel} + q\Omega_e} \left[\frac{q\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} \frac{1}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} + \frac{\partial}{\partial v_{\parallel}} \frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{h\Omega_e - \Omega} \frac{h\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right] f_{0e} \\
& + i \left(\frac{e}{m} \right)^4 \sum_{\mathbf{K}, \mathbf{k}} |E_l(\mathbf{k}, \omega)|^2 |E_h(\mathbf{K}, \Omega)|^2 \frac{\partial}{\partial v_{\parallel}} \left\{ \frac{k_{\parallel}}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{\Omega - h\Omega_e} \frac{h\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right. \\
& \left. - \left(\frac{1}{\omega - k_{\parallel}v_{\parallel} + \Omega_e + i0} \frac{e^{i\phi}}{2} + \frac{1}{\omega - k_{\parallel}v_{\parallel} - \Omega_e + i0} \frac{e^{-i\phi}}{2} \right) \frac{\partial}{\partial v_{\perp}} \left[\frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{\Omega - h\Omega_e} \frac{h\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right. \right. \\
& \left. \left. - \frac{J_p(x) J_q(x) e^{i(p-q)\phi}}{\omega - \Omega - k_{\parallel}v_{\parallel} + q\Omega_e} \left(k_{\parallel} \frac{\partial}{\partial v_{\parallel}} - \frac{q\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right) \right] \right\} f_{0e} \frac{\omega_{pe}^2}{\epsilon_{0l}(\mathbf{k} - \mathbf{K}) |\mathbf{k} - \mathbf{K}|^2} \int \frac{J_p(x) J_q(x) e^{i(p-q)\phi}}{\omega - \Omega - k_{\parallel}v_{\parallel} + q\Omega_e} \\
& \times \left[\frac{q\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} \frac{1}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} + \frac{\partial}{\partial v_{\parallel}} \frac{J_g(x) J_h(x) e^{i(g-h)\phi}}{h\Omega_e - \Omega} \frac{h\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right] f_{0e} d\mathbf{v}. \tag{27}
\end{aligned}$$

III. ENERGY-MOMENTUM CONSERVATION RELATIONS

We study conservation relations for the nonlinear interaction between resonant mode (Langmuir waves) and electrons. First, we consider the total momentum density conservation relation. The rate of change of electron momentum density may be expressed as

$$\frac{d}{dt} \mathbf{P}_e = \int m(v_x \hat{x} + v_y \hat{y} + v_{\parallel} \hat{z}) \left(\frac{\partial f_{0e}}{\partial t} \right) d\mathbf{v} = \hat{x} m \int v_x \left(\frac{\partial f_{0e}}{\partial t} \right)_I d\mathbf{v} + \hat{z} m \int v_{\parallel} \left(\frac{\partial f_{0e}}{\partial t} \right)_{\text{II}} d\mathbf{v} = \hat{x} \frac{d}{dt} P_{eh} + \hat{z} \frac{d}{dt} P_{el}, \tag{28}$$

where P_{eh} and P_{el} represent the electron momentum due to the high- and the low-frequency nonlinear motion of the electrons, respectively. Inserting Eq. (26) into the first part on the right-hand side of Eq. (28) and performing partial integration, we obtain

$$\begin{aligned}
\frac{d}{dt} P_{eh} = & m \int v_x \left(\frac{\partial f_{0e}}{\partial t} \right)_I d\mathbf{v} \\
= & \frac{e^4}{m^3} \sum_{\mathbf{K}, \mathbf{k}} |E_h(\mathbf{K}, \Omega)|^2 |E_l(\mathbf{k}, \omega)|^2 \frac{1}{K_{\perp}} \text{Im} \left\{ \int \frac{J_a^2(x)}{\Omega - a\Omega_e} \left[\frac{\partial}{\partial v_{\parallel}} \frac{J_s^2(x)}{\Omega - \omega + k_{\parallel}v_{\parallel} - s\Omega_e} \left(\frac{\partial}{\partial v_{\parallel}} \frac{J_n^2(x)}{\Omega - n\Omega_e} \frac{n\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right. \right. \right. \\
& \left. \left. \left. + \frac{s\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - \frac{1}{-\omega + k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \right) f_{0e} \right] d\mathbf{v} - \int \frac{J_a^2(x)}{\Omega - a\Omega_e} \left[\frac{\partial}{\partial v_{\parallel}} \frac{J_s^2(x)}{\Omega - \omega + k_{\parallel}v_{\parallel} - s\Omega_e} \left(\frac{s\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - k_{\parallel} \frac{\partial}{\partial v_{\parallel}} \right) \right. \\
& \left. \left. + \left(\frac{a\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - k_{\parallel} \frac{\partial}{\partial v_{\parallel}} \right) \frac{1}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \right] f_{0e} d\mathbf{v} \frac{\omega_{pe}^2}{\epsilon_{0h}(\mathbf{k} - \mathbf{K}) |\mathbf{k} - \mathbf{K}|^2} \int \frac{J_s^2(x)}{\Omega - \omega + k_{\parallel}v_{\parallel} - s\Omega_e} \right. \\
& \left. \times \left(\frac{\partial}{\partial v_{\parallel}} \frac{J_n^2(x)}{\Omega - n\Omega_e} \frac{n\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} + \frac{s\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} \frac{1}{-\omega + k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \right) f_{0e} d\mathbf{v} \right\} = \sum_{\mathbf{K}} K_{\perp} \text{Im} \epsilon_h^n(\mathbf{K}, \Omega) \frac{|E_h(\mathbf{K}, \Omega)|^2}{4\pi}. \tag{29}
\end{aligned}$$

In a similar manner, inserting Eq. (27) into the second part on the right-hand side of Eq. (28) and performing partial integration (and the integration over ϕ), we obtain

$$\begin{aligned}
\frac{d}{dt} P_{el} = & m \int v_{\parallel} \left(\frac{\partial f_{0e}}{\partial t} \right)_{\text{II}} d\mathbf{v} \\
= & -i \frac{e^4}{m^3} \sum_{\mathbf{K}, \mathbf{k}} |E_l(\mathbf{k}, \omega)|^2 |E_h(\mathbf{K}, \Omega)|^2 \frac{4\pi^2 \omega_{pe}^2}{\epsilon_{0l}(\mathbf{k} - \mathbf{K}) |\mathbf{k} - \mathbf{K}|^2} \text{Im} \left[\int \frac{k_{\parallel}}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \frac{J_g^2(x)}{\Omega - g\Omega_e} \frac{g\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} f_{0e} v_{\perp} dv_{\perp} dv_{\parallel} \right. \\
& \times \int \frac{J_p^2(x)}{\omega - \Omega - k_{\parallel}v_{\parallel} + p\Omega_e} \left(\frac{p\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} \frac{1}{\omega - k_{\parallel}v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} + \frac{\partial}{\partial v_{\parallel}} \frac{J_g^2(x)}{g\Omega_e - \Omega} \frac{g\Omega_e}{K_{\perp}v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right) f_{0e} v_{\perp} dv_{\perp} dv_{\parallel} \left. \right] \\
= & \sum_{\mathbf{k}} k_{\parallel} \text{Im} \epsilon_l^n(\mathbf{k}, \omega) \frac{|E_l(\mathbf{k}, \omega)|^2}{4\pi}. \tag{30}
\end{aligned}$$

The momentum density of the Bernstein wave is defined by,

$$\mathbf{P}_h = \sum_{\mathbf{K}} \frac{\mathbf{K}}{\Omega} W_h.$$

From which we obtain,

$$\begin{aligned} \frac{d}{dt} P_h &= -2 \sum_{\mathbf{K}} \frac{K_{\perp}}{\Omega} \gamma_h(\mathbf{K}, \Omega) W_h \\ &= -2 \sum_{\mathbf{K}} \frac{K_{\perp}}{\Omega} \frac{\text{Im } \epsilon_h^n(\mathbf{K}, \Omega)}{\partial \epsilon_{0h} / \partial \Omega} \frac{|E_h(\mathbf{K}, \Omega)|^2}{8\pi} \Omega \frac{\partial \epsilon_{0h}}{\partial \Omega} = - \sum_{\mathbf{K}} K_{\perp} \text{Im } \epsilon_h^n(\mathbf{K}, \Omega) \frac{|E_h(\mathbf{K}, \Omega)|^2}{4\pi} = - \frac{d}{dt} P_{eh}. \end{aligned} \quad (31)$$

The momentum density of Langmuir turbulence is defined by

$$\mathbf{P}_l = \sum_{\mathbf{k}} \frac{\mathbf{k}}{\omega} W_l,$$

from which we obtain

$$\frac{d}{dt} P_l = \sum_{\mathbf{k}} \frac{k_{\parallel}}{\omega} \frac{\partial}{\partial t} W_l = -2 \sum_{\mathbf{k}} \frac{k_{\parallel}}{\omega} \frac{\text{Im } \epsilon_l^n(\mathbf{k}, \omega)}{\partial \epsilon_{0l} / \partial \omega} W_l = - \sum_{\mathbf{k}} k_{\parallel} \text{Im } \epsilon_l^n(\mathbf{k}, \omega) W_l = - \frac{d}{dt} P_{el}. \quad (32)$$

Equations (28), (31), and (32) clearly show that the total momentum density is conserved for the plasma-maser interaction in a magnetized plasma:

$$\frac{d}{dt} (\mathbf{P}_e + \mathbf{P}_h + \mathbf{P}_l) = 0. \quad (33)$$

Next, we consider the total energy conservation relation. The rate of change of total electron kinetic energy density may be expressed as

$$\frac{d}{dt} T_e = \int \frac{m}{2} (v_x^2 + v_y^2 + v_{\parallel}^2) \left(\frac{\partial f_{0e}}{\partial t} \right) d\mathbf{v} = \frac{m}{2} \int v_x^2 \left(\frac{\partial f_{0e}}{\partial t} \right)_I d\mathbf{v} + \frac{m}{2} \int v_{\parallel}^2 \left(\frac{\partial f_{0e}}{\partial t} \right)_{II} d\mathbf{v} = \frac{d}{dt} T_{eh} + \frac{d}{dt} T_{el}, \quad (34)$$

where T_{eh} and T_{el} represent the electron kinetic energy due to the high- and low-frequency nonlinear motion of the electrons, respectively. Inserting Eq. (26) into the first part on the right-hand side of Eq. (34) and performing partial integration, we obtain

$$\begin{aligned} \frac{d}{dt} T_{eh} &= \frac{m}{2} \int v_x^2 \left(\frac{\partial f_{0e}}{\partial t} \right)_I d\mathbf{v} \\ &= \frac{e^4}{m^3} \sum_{\mathbf{K}, \mathbf{k}} |E_h(\mathbf{K}, \Omega)|^2 |E_l(\mathbf{k}, \omega)|^2 \frac{1}{K_{\perp}} \text{Im} \left\{ \int v_{\perp} \left(\frac{e^{i\phi} + e^{-i\phi}}{2} \right) \frac{J_a(x) J_b(x) e^{i(a-b)\phi}}{\Omega - b\Omega_e} \left[\frac{\partial}{\partial v_{\parallel}} \frac{J_s(x) J_t(x) e^{i(s-t)\phi}}{\Omega - \omega + k_{\parallel} v_{\parallel} - t\Omega_e} \right. \right. \\ &\quad \times \left(\frac{\partial}{\partial v_{\parallel}} \frac{J_n(x) J_l(x) e^{i(n-l)\phi}}{\Omega - l\Omega_e} \right) \frac{l\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} + \frac{t\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} \frac{1}{-\omega + k_{\parallel} v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \left. \right] f_{0e} d\mathbf{v} \\ &\quad - \int v_{\perp} \left(\frac{e^{i\phi} + e^{-i\phi}}{2} \right) \frac{J_a(x) J_b(x) e^{i(a-b)\phi}}{\Omega - b\Omega_e} \left[\frac{\partial}{\partial v_{\parallel}} \frac{J_s(x) J_t(x) e^{i(s-t)\phi}}{\Omega - \omega + k_{\parallel} v_{\parallel} - t\Omega_e} \left(\frac{t\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} - k_{\parallel} \frac{\partial}{\partial v_{\parallel}} \right) + \left(\frac{b\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\parallel}} \right. \right. \\ &\quad \left. \left. - k_{\parallel} \frac{\partial}{\partial v_{\parallel}} \right) \frac{1}{\omega - k_{\parallel} v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \right] f_{0e} d\mathbf{v} \frac{\omega_{pe}^2}{\epsilon_{0h}(\mathbf{K} - \mathbf{k}) |\mathbf{K} - \mathbf{k}|^2} \int \frac{J_s(x) J_t(x) e^{i(s-t)\phi}}{\Omega - \omega + k_{\parallel} v_{\parallel} - t\Omega_e} \left(\frac{\partial}{\partial v_{\parallel}} \frac{J_n(x) J_l(x) e^{i(n-l)\phi}}{\Omega - l\Omega_e} \right. \\ &\quad \times \left. \left. \frac{l\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} + \frac{t\Omega_e}{v_{\perp}} \frac{\partial}{\partial v_{\perp}} \frac{1}{-\omega + k_{\parallel} v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \right) f_{0e} d\mathbf{v} \right\}. \end{aligned} \quad (35)$$

Using the identity,

$$\int v_{\perp} \left(\frac{e^{i\phi} + e^{-i\phi}}{2} \right) \frac{J_a J_b e^{i(a-b)\phi}}{\Omega - b\Omega_e} [\dots] d\mathbf{v} = \int v_{\perp} \left(\frac{J_{a+1} + J_{a-1}}{2} \right) \frac{J_b e^{i(a-b)\phi}}{\Omega - b\Omega_e} [\dots] d\mathbf{v} = \frac{a\Omega_e}{K_{\perp}} \int \frac{J_a J_b e^{i(a-b)\phi}}{\Omega - b\Omega_e} [\dots] d\mathbf{v},$$

we obtain from Eq. (35)

$$\frac{d}{dt} T_{eh} = \sum_{\mathbf{K}} a\Omega_e \operatorname{Im} \epsilon_h^n(\mathbf{K}, \Omega) \frac{|E_h(\mathbf{K}, \Omega)|^2}{4\pi}. \quad (36)$$

Inserting Eq. (27) into the second part on the right-hand side of Eq. (34) and performing partial integration (and the integration over ϕ), we obtain

$$\begin{aligned} \frac{d}{dt} T_{el} &= \frac{m}{2} \int v_{\parallel}^2 \left(\frac{\partial f_{0e}}{\partial t} \right)_{\text{II}} d\mathbf{v} = -i \frac{e^4}{m^3} \sum_{\mathbf{K}, \mathbf{k}} |E_l(\mathbf{k}, \omega)|^2 |E_h(\mathbf{K}, \Omega)|^2 \frac{4\pi^2 \omega_{pe}^2}{\epsilon_{0l}(\mathbf{k} - \mathbf{K}) |\mathbf{k} - \mathbf{K}|^2} \\ &\quad \times \operatorname{Im} \left[\int \frac{k_{\parallel} v_{\parallel}}{\omega - k_{\parallel} v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} \frac{J_g^2(x)}{\Omega - h\Omega_e} \frac{h\Omega_e}{K_{\perp} v_{\perp}} \frac{\partial}{\partial v_{\perp}} f_{0e} v_{\perp} dv_{\perp} dv_{\parallel} \int \frac{J_p^2(x)}{\omega - \Omega - k_{\parallel} v_{\parallel} + p\Omega_e} \right. \\ &\quad \times \left. \left(\frac{p\Omega_e}{K_{\perp} v_{\perp}} \frac{\partial}{\partial v_{\perp}} \frac{1}{\omega - k_{\parallel} v_{\parallel} + i0} \frac{\partial}{\partial v_{\parallel}} + \frac{\partial}{\partial v_{\parallel}} \frac{J_g^2(x)}{g\Omega_e - \Omega} \frac{g\Omega_e}{K_{\perp} v_{\perp}} \frac{\partial}{\partial v_{\perp}} \right) f_{0e} v_{\perp} dv_{\perp} dv_{\parallel} \right] \end{aligned} \quad (37)$$

which on further integration gives,

$$\frac{d}{dt} T_{el} = \sum_{\mathbf{k}} \omega \operatorname{Im} \epsilon_l^n(\mathbf{k}, \omega) W_l. \quad (38)$$

From Eq. (12), the rate of change of Bernstein wave energy for $\Omega \sim a\Omega_e$ is given by

$$\begin{aligned} \frac{\partial}{\partial t} \sum_{\mathbf{k}} W_h &= 2 \gamma_h(\mathbf{K}, \Omega) W_h \\ &= -2 \sum_{\mathbf{K}} \frac{\operatorname{Im} \epsilon_h^n(\mathbf{K}, \Omega)}{\partial \epsilon_{0h} / \partial \Omega} \frac{|E_h(\mathbf{K}, \Omega)|^2}{8\pi} a\Omega_e \frac{\partial \epsilon_{0h}}{\partial \Omega} \\ &= -\sum_{\mathbf{K}} a\Omega_e \operatorname{Im} \epsilon_h^n(\mathbf{K}, \Omega) \frac{|E_h(\mathbf{K}, \Omega)|^2}{4\pi} \\ &= -\frac{d}{dt} T_{eh}. \end{aligned} \quad (39)$$

From Eq. (22), the rate change of Langmuir wave energy is,

$$\begin{aligned} \frac{\partial}{\partial t} \sum_{\mathbf{k}} W_l &= 2 \sum_{\mathbf{k}} \gamma_l(\mathbf{k}, \omega) W_l \\ &= -2 \sum_{\mathbf{k}} \frac{\operatorname{Im} \epsilon_l^n(\mathbf{k}, \omega)}{\partial \epsilon_{0l} / \partial \omega} W_l \\ &= -\sum_{\mathbf{k}} \omega \operatorname{Im} \epsilon_l^n(\mathbf{k}, \Omega) W_l = -\frac{d}{dt} T_{el}. \end{aligned} \quad (40)$$

Equations (34), (39), and (40) show that the total energy density is conserved in the plasma-maser process in a magnetized plasma:

$$\frac{d}{dt} (T_e + W_h + W_l) = 0. \quad (41)$$

From the above observations we conclude that the energy-momentum conservation relations are exactly satisfied for the plasma-maser interaction in a magnetized plasma.

IV. VIOLATION OF THE MANLEY-ROWE RELATION FOR PLASMA WAVES AND THE GENERALIZED MANLEY-ROWE RELATION

In this section we consider the Manley-Rowe relation (conservation of the total plasmon number density) for the plasma-maser instability in a magnetized plasma. This relation holds for the standard mode-mode coupling processes in weak turbulence theory, viz., the three-wave resonance and the nonlinear Landau resonance. However, as is demonstrated below, this relation does not hold for the plasma-maser interaction. The plasmon numbers for the Langmuir and Bernstein mode waves are defined as

$$N_l(\mathbf{k}, \omega) = \frac{W_l(\mathbf{k}, \omega)}{\omega}, \quad (42)$$

$$N_h(\mathbf{K}, \Omega) = \frac{W_h(\mathbf{K}, \Omega)}{\Omega}. \quad (43)$$

From Eq. (20) we obtain the rate of change of the plasmon number for the Langmuir wave as

$$\begin{aligned}
\frac{\partial}{\partial t} \sum_{\mathbf{k}} N_l(\mathbf{k}, \omega) &= \frac{2}{\omega} \sum_{\mathbf{k}} \gamma_l(\mathbf{k}, \omega) W_l(\mathbf{k}, \omega) \\
&= - \sum_g \sum_{\mathbf{k}} W_l(\mathbf{k}, \omega) \sum_{\mathbf{K}} \frac{W_h(\mathbf{K}, \Omega)}{NT} \frac{4\omega_{pe}^6 g^2 \sqrt{\pi} \delta}{\Omega^4 \beta_g \Omega_e^2} \frac{bk_{\parallel}}{\epsilon_{0l}(\mathbf{k}-\mathbf{K}) |\mathbf{k}-\mathbf{K}|^2 K_{\perp}^2 \rho_e^2 a_e^2 |k_{\parallel}|} \frac{(k_{\parallel} v_0 - \omega)}{k_{\parallel} v_e} I_g(K_{\perp}^2 \rho_e^2) \\
&\quad \times e^{-K_{\perp}^2 \rho_e^2} e^{-m(\omega/k_{\parallel} - v_0)^2/2T} \left[\frac{2\Omega^2 I_g(K_{\perp}^2 a_e^2) e^{-K_{\perp}^2 a_e^2}}{\omega^2 b} + 1 + \frac{2\Omega^2 (\Omega^2 + g^2 \Omega_e^2)}{(\Omega^2 - g^2 \Omega_e^2)^2} \frac{c}{b} + \frac{\Omega^2 I_g(K_{\perp}^2 a_e^2) e^{-K_{\perp}^2 a_e^2}}{(\Omega^2 - g^2 \Omega_e^2) b} \right. \\
&\quad \left. \times \left(1 - \frac{4\Omega^2}{\Omega^2 - g^2 \Omega_e^2} \right) \right]. \tag{44}
\end{aligned}$$

From Eq. (10) we obtain the rate of change of plasmon number for the Bernstein wave as,

$$\begin{aligned}
\frac{\partial}{\partial t} \sum_{\mathbf{K}} N_h(\mathbf{K}, \Omega) &= \frac{2}{\Omega} \sum_{\mathbf{K}} \gamma_h(\mathbf{K}, \Omega) W_h(\mathbf{K}, \Omega) \\
&= \sum_a \sum_{\mathbf{K}} W_h(\mathbf{K}, \Omega) \sum_{\mathbf{k}} \frac{W_l(\mathbf{k}, \omega)}{NT} \frac{4\omega_{pe}^6 a^2 \sqrt{\pi} \delta}{\Omega^4 \beta_a \Omega_e^2} \frac{bk_{\parallel}}{\epsilon_{0h}(\mathbf{K}-\mathbf{k}) |\mathbf{K}-\mathbf{k}|^2 K_{\perp}^2 \rho_e^2 a_e^2 |k_{\parallel}|} \frac{(k_{\parallel} v_0 - \omega)}{k_{\parallel} v_0} \\
&\quad \times I_a(K_{\perp}^2 \rho_e^2) e^{-K_{\perp}^2 \rho_e^2} e^{-m(\omega/k_{\parallel} - v_0)^2/2T} \left[1 + \frac{2\Omega^2 (\Omega^2 + a^2 \Omega_e^2)}{(\Omega^2 - a^2 \Omega_e^2)^2} \frac{c}{b} + \frac{\Omega^2 I_a(K_{\perp}^2 a_e^2) e^{-K_{\perp}^2 a_e^2}}{(\Omega^2 - a^2 \Omega_e^2) b} \right. \\
&\quad \left. \times \left(1 - \frac{4\Omega^2}{\Omega^2 - a^2 \Omega_e^2} \right) \right]. \tag{45}
\end{aligned}$$

Thus, it is clear from Eqs. (44) and (45) that the Manley–Rowe relations are violated for the plasma waves in the plasma–maser interaction in a magnetized plasma, because

$$\left| \frac{\partial}{\partial t} \sum_{\mathbf{k}} N_l(\mathbf{k}, \omega) \right| > \left| \frac{\partial}{\partial t} \sum_{\mathbf{K}} N_h(\mathbf{K}, \Omega) \right|. \tag{46}$$

The Manley–Rowe relation in its original form [Eqs. (42) and (43)] is defined only for wave action. It is shown that the total wave actions are not conserved [Eq. (46)]. Thus, the Manley–Rowe relations in the original form are violated. However, if we define the effective change of action due to electron kinetic energy as

$$\frac{\partial}{\partial t} N_e = \frac{\partial}{\partial t} \sum_{\mathbf{k}} \frac{T_{el}}{\omega} + \frac{\partial}{\partial t} \sum_{\mathbf{K}} \frac{T_{eh}}{\Omega}, \tag{47}$$

we then find

$$\frac{\partial}{\partial t} \sum_{\mathbf{k}} N_l + \frac{\partial}{\partial t} \sum_{\mathbf{K}} N_h + \frac{\partial}{\partial t} N_e = 0. \tag{48}$$

The above equation is clearly satisfied from Eqs. (36) and (38). Accordingly, we may say that Eq. (48) is the generalized Manley–Rowe relation for the plasma–maser effect. The introduction of the last term on the left-hand side of Eq. (48) is clear because the electrons carry energy and momentum nonlinearly in the plasma–maser process. It should be

mentioned here that the violation of the Manley–Rowe relation in its original form is already pointed out for an open system,¹⁷ a dissipative system,¹⁸ and a nonstationary system.¹⁹ The violation of the Manley–Rowe relation in the plasma–maser interaction is expected because the process involves direct acceleration of the electrons. Recently, Johnston *et al.*²⁰ have shown that total action for waves is not conserved if the electrons are directly accelerated by waves. This result is markedly different from the standard nonlinear Landau resonance^{21,22} where the Manley–Rowe relation is satisfied. Actually, it is well known that the Manley–Rowe relation is valid only in the absence of energy dissipation process,^{18,23} so that one cannot expect that this remains true in the case of plasma–maser interaction which always coexists with a dissipation process (Landau or cyclotron resonance).

V. DISCUSSION AND CONCLUSION

We have shown that the total momentum and energy conservation relations among the resonant modes, nonresonant modes, and the electrons are exactly satisfied for the plasma–maser process in a magnetized plasma [Eqs. (33) and (41)]. However, the Manley–Rowe relation in its original form is violated because the electrons carry nonlinear momentum and energy [Eq. (46)]. Instead of the original Manley–Rowe relation, the plasma–maser satisfies the generalized Manley–Rowe relation for magnetized plasma, de-

veloped in Sec. IV, which includes the effective action due to electrons in addition to the wave actions [Eq. (48)].

As has been mentioned earlier, the plasma-maser process is effective at least for two systems: (1) an open system for which a free energy source from outside the system is available,¹¹ and (2) magnetized plasmas with a symmetry breaking factor.^{12,13} Here, we study the conservation relations and the Manley-Rowe relation for the plasma-maser process in a magnetized plasma to clarify the basic physical principles of the process. The unique feature of the process, that it does not require population inversion for the energy up-conversion from the low-frequency mode to the high-frequency mode, is justified by the violation of the Manley-Rowe relation for plasma waves.

The anomalous high-frequency radiation in the presence of enhanced low-frequency fluctuations is often reported in magnetized plasmas. The extraordinary mode radiation from the electrostatic lower-hybrid turbulences is observed in space²⁴ and by computer simulation.²⁵ According to them, the electrons accelerated along the magnetic field by the lower-hybrid turbulences radiate the high-frequency electromagnetic waves. The plasma-maser effect predicts most of the characteristics of the observations. It should be mentioned here that similar high-frequency radiation is also observed in laboratory experiments.²⁶

According to the recent experiments in the ionosphere,²⁷ stimulated electromagnetic emission, such as up-converted broadband continuum ordinary mode radiation, cannot be explained in terms of the existing mode coupling processes. The plasma-maser process, on the other hand, predicts the energy up-conversion with broad continuum spectrum. Thus, the ordinary mode radiation with broad spectrum may be generated by the plasma-maser effect. Furthermore, the concurrent excitation of Langmuir waves in turbulent heating experiments, in which ion sound wave activity is much enhanced, is a typical phenomenon of the laboratory plasma.²⁸ The new mode coupling process may play an important role in the interpretation of numerous anomalous radiation phenomena in magnetized plasmas.

ACKNOWLEDGMENT

The work of one of the authors (BS) is supported by the Council of Scientific and Industrial Research (CSIR), India.

- ¹M. Nambu, *Laser Part. Beams* **1**, 427 (1983).
- ²K. Ronmark and L. Stenflo, *Planet. Space Sci.* **24**, 904 (1976).
- ³K. K. Sarma, S. N. Sarma, M. Nambu, and T. Hada, *Phys. Rev. A* **43**, 5555 (1991).
- ⁴Yu. M. Aliev, O. M. Tolokachev, and L. Stenflo, *Phys. Rev. A* **38**, 4899 (1987).
- ⁵M. Nambu, *J. Phys. Soc. Jpn.* **54**, 2361 (1985).
- ⁶S. V. Vladimirov, *Phys. Rev. E* **49**, R997 (1994).
- ⁷V. S. Krivitskii and S. V. Vladimirov, *J. Plasma Phys.* **46**, 209 (1991).
- ⁸V. S. Krivitskii and S. V. Vladimirov, *J. Plasma Phys.* **46**, 219 (1991).
- ⁹V. S. Krivitskii, V. N. Tsytovich, and S. V. Vladimirov, *Phys. Rep.* **218**, 141 (1992).
- ¹⁰M. Nambu and T. Hada, *Phys. Fluids B* **5**, 742 (1993).
- ¹¹S. B. Isakov, V. S. Krivitskii, and V. N. Tsytovich, *Zh. Eksp. Teor. Fiz.* **90**, 933 (1986) [*Sov. Phys. JETP* **63**, 545 (1986)].
- ¹²M. Nambu, S. Bujarbarua, and S. N. Sarma, *Phys. Rev. A* **35**, 798 (1987).
- ¹³M. Nambu, S. V. Vladimirov, and H. Schamel, *Phys. Lett. A* **178**, 400 (1993).
- ¹⁴J. M. Manley and H. E. Rowe, *Proc. IRE* **44**, 904 (1956).
- ¹⁵M. Nambu, S. N. Sarma, and K. K. Sarma, *Phys. Rev. A* **45**, 7456 (1992).
- ¹⁶B. J. Saikia, B. K. Saikia, S. Bujarbarua, and M. Nambu, *Pramana-J. Phys.* **42**, 435 (1994).
- ¹⁷L. Krilin, *J. Plasma Phys.* **12**, 365 (1974).
- ¹⁸K. Papadopoulos, K. Ko, and V. Tripathi, *Phys. Rev. Lett.* **51**, 463 (1983).
- ¹⁹T. Tanaka, *J. Phys. Soc. Jpn.* **62**, 2351 (1993).
- ²⁰T. W. Johnston, P. Bertrand, A. Ghizzo, M. Shoucri, E. Fijalkow, and M. R. Feix, *Phys. Fluids B* **4**, 2523 (1992).
- ²¹R. C. Davidson, *Methods in Nonlinear Plasma Theory* (Academic, New York, 1972).
- ²²A. A. Galeev and R. Z. Sagdeev, *Basic Plasma Physics I*, edited by A. A. Galeev and R. N. Sudan (North-Holland, Amsterdam, 1983), p. 677.
- ²³G. Gryning and P. R. Berman, *Phys. Rev. A* **43**, 3994 (1991).
- ²⁴M. Malingre, R. Pottelette, N. Dubouloz, P. A. Lindqvist, G. Holmgren, and B. Aparicio, *Geophys. Res. Lett.* **19**, 1339 (1992).
- ²⁵K. G. McClements, R. Bingham, J. J. Su, J. M. Dawson, and D. S. Spicer, *Astrophys. J.* **409**, 465 (1993).
- ²⁶T. Mieno, M. Oertl, R. Hatakeyama, and N. Sato, *Phys. Lett. A* **184**, 445 (1994).
- ²⁷B. Thide, in *Proceeding of International Summer School on Space Plasma Physics* (Radioophysical Research Institute, Nizhniy Novgorod, Russia, 1993), p. 35.
- ²⁸H. de Kluiver, N. F. Perepelkin, and A. Hirose, *Phys. Rep.* **199**, 281 (1991).

Parametric-Basics

Low-Noise Amplifiers—Then and Now

James J. Whelehan, *Life Fellow, IEEE*

Invited Paper

Abstract—The evolution in the performance of low-noise amplifiers (LNAs) has been dynamic over the past years. From the early LNAs that were complex, large, and heavy, to the present day InP high electron-mobility transistors that have virtually transformed the industry by their performance and extension into frequency bands that were not even considered in the past. This paper will hopefully summarize the transformation that has occurred in the LNA field, viewing where they were in the past, and where they are now.

Index Terms—Cryogenics, FETs, masers, parametric amplifiers, transistors.

I. INTRODUCTION

OVER THE past number of years, the advances made in low-noise technology have been dynamic. From the early 1980s, the technological innovations achieved in this expanding field not only resulted in improved devices, subsystems, and systems, but also enabled new markets to be developed such as the cellular market, Direct TV, satellite communications, Internet communications, etc. We are now accustomed to a brand new vocabulary that has been ingrained in us. The vocabulary of the 1980s for low-noise front-ends and, in particular, low-noise amplifiers (LNAs), was parametric amplifiers (paramps), field-effect transistors (FETs), microwave amplification by stimulated emission of radiation (MASERS), cryogenically cooled amplifiers, thermoelectrically cooled (TE) amplifiers, bipolar transistors, etc. Fig. 1 shows the state-of-the-art noise performance of these LNAs [1]. While the low-noise performance of the LNAs were extremely good then, the design engineer still had to make some very complex system trades. Many LNAs were large, heavy, and consumed a lot of power. Satellite ground terminals were one application where low-noise performance, light weight, low power, and high reliability were simultaneously required, but not always achieved. In many cases, the LNAs were designed for lower noise figures than required since the amplifiers and their ancillary support equipment could not be located in very close proximity to the antenna. Relatively long interconnecting lines that could have substantial loss were used between them. Consequently, higher cost LNAs with noise performance lower than was really required were necessary to meet the system requirements. With the development of three-terminal devices that has continually lowered the noise threshold coupled with a significant reduction in size, weight, and power, system

performance and capability has been vastly improved. We see LNAs now being used in systems well beyond the traditional boundaries that were considered their role in the past. LNA technology has expanded into the millimeter-wave frequency band and is now approaching the submillimeter-wave portion of the frequency spectrum. Expanded bandwidths across the entire frequency spectrum have been obtained and are now commonly included in our designs. We now have a new vocabulary that includes as a minimum high electron-mobility transistors (HEMTs), pseudomorphic high electron-mobility transistors (p-HEMTs), indium phosphide high electron-mobility transistors (InP HEMTs), microwave/millimeter-wave integrated circuits (MMICs), high-temperature superconductivity (HTSC), and expanding. We anticipate newer and more advanced devices to be developed to meet the needs of tomorrow's systems. We have already witnessed the explosive growth in wireless and satellite communication systems.

But first, let us consider where we started. In the 1980s, LNAs were developed for the microwave and low millimeter-wave region. Mixers were generally the front-end component of choice in the millimeter- and submillimeter-wave region. We now have practical low-noise millimeter-wave satellite systems operating around 30 GHz (TDRS H, I, J) and above, collision-avoidance systems used on buses and cars that operate around 77 GHz [2]. LNAs for remotely sensing the Earth's environment from a satellite platform have been developed at 200 GHz. Significant progress to improve low-noise performance, enhance system capability while simultaneously making the systems smaller, lighter weight, lower power, broader bandwidth, and increasing the frequency of operation has been made. Multichip module (MCM) millimeter-wave modules (Fig. 2) have been developed for space applications using MMIC amplifiers. With these new technologies, phased arrays with their unique capability are becoming more practical and affordable. HTSC is an area that has tremendous promise if higher temperature material could be developed beyond the present 100-K threshold. Low noise has been and still is a multidimensional growth area. We will explore a few of the paths taken by LNAs from the 1980s to today. We will consider some of the new developments that have been made, and ponder where we are heading and how these low-noise developments will impact future applications and the new system architectures.

II. MASERS

The technological advances, particularly in LNAs, that we had seen from the end of World War II were dynamic and

Manuscript received November 15, 2001.

The author is with JJW Consulting Inc., Hatfield, PA 19440 USA.

Publisher Item Identifier S 0018-9480(02)01971-3.

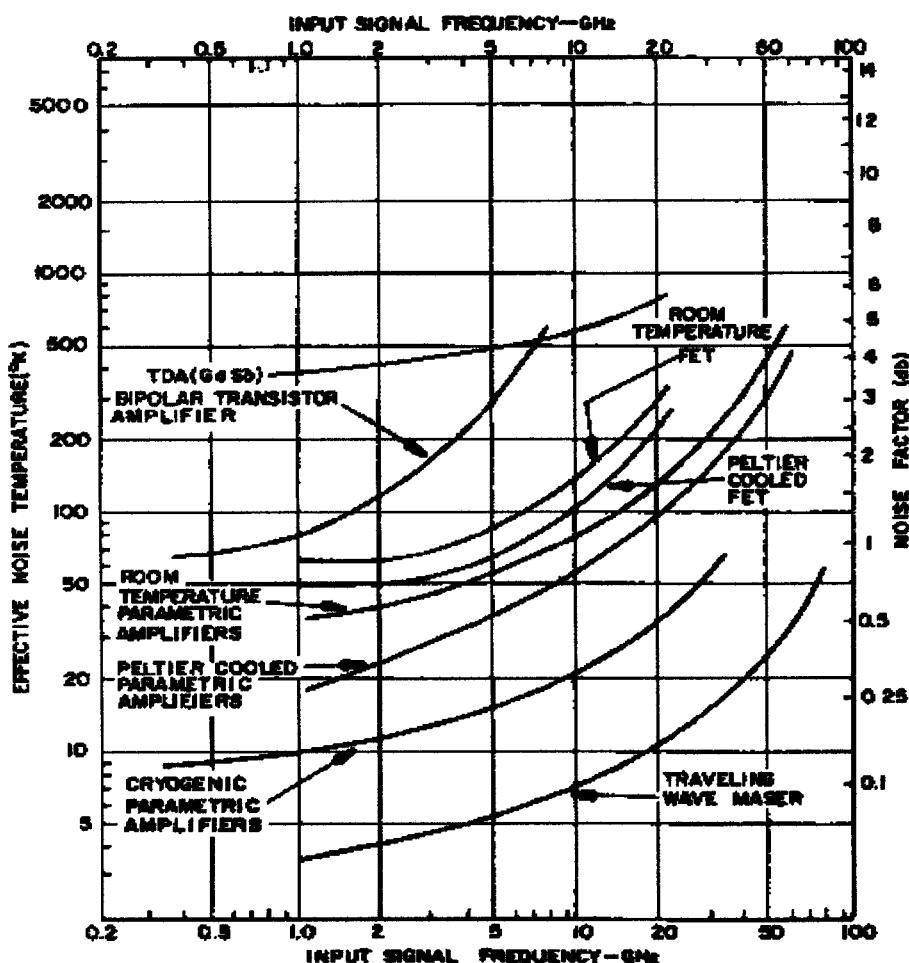


Fig. 1. State-of-the-art performance of LNAs.

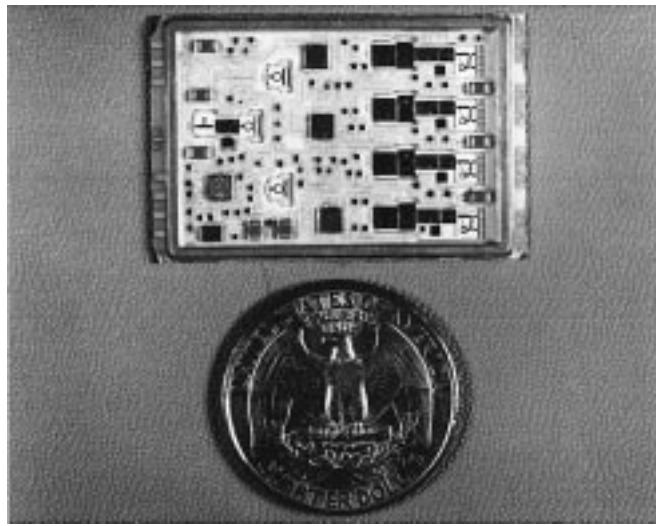


Fig. 2. Millimeter-wave MCM module (courtesy of the EDO Corporation).

enabled us to design systems that were only speculative in the past. In the early 1980s, ultra-low-noise performance generally required amplifiers to be cryogenically cooled to liquid nitrogen temperatures (77 K) and even lower temperatures (4 K). Cryogenically cooled amplifiers were extensively used in satellite communication ground terminals due to the limited transmitter

power available in the spacecraft, and by the radio astronomers who required highly sensitive receivers to explore the wonders of the universe. They were also developed for the Jet Propulsion Laboratory (JPL) deep space network (DSN) to maintain the long communication links with the spacecraft that were being launched to explore the solar system. Most of the LNAs had relatively narrow bandwidths (about 5%–15%). However, when the ultimate in low-noise performance was required, the only LNA that could satisfy these demands was the MASER, a narrow-band LNA (~5%–10%). However, this LNA had to be cryogenically cooled to liquid-helium temperatures (\approx 4 K). Both traveling-wave masers (TWM) and cavity masers were originally developed. However, the instantaneous bandwidth of the TWM was broader and became the MASER of choice in most cases. Some of the early MASERS were physically immersed in a bath of liquid helium, but the operational duty cycle was very limited. With the availability of closed-cycle refrigerators, the MASER became more acceptable. A MASER is a very complex amplifier that is large, heavy, and consumed lots of power (kilowatts). Most of this power was used for the cryogenic refrigerator that cooled the maser structure to 4.5 K. The TWM MASER amplifies microwave signals that are propagating along the length of a tuned ruby crystal (other crystals were also used) in a slow-wave structure. The ruby crystal is cooled to a bath temperature of about 4.5 K. The

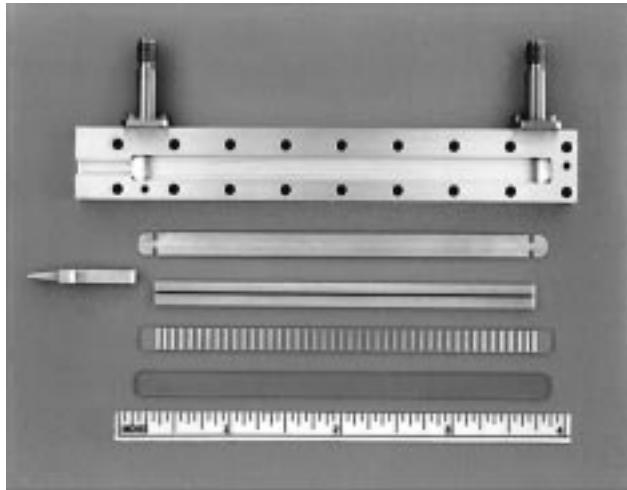


Fig. 3. *X*-band TWM assembly (courtesy of the Jet Propulsion Laboratory, California Institute of Technology).

physical details of an *X*-band TWM structure are shown in Fig. 3 [3]. The noise temperature (T_e) of the MASER structure is directly proportional to its operating temperature, as shown by the following equation [4]:

$$T_e = T_B/I$$

where

- T_e noise temperature of the TWM;
- T_B physical temperature of the ruby crystal;
- I inversion ratio-constant determined by the material.

The noise temperature of a cavity maser and the TWM MASER are essentially the same for a high gain case when a relatively low-loss slow-wave structure is used. An *X*-band MASER structure that operates at a bath temperature of 4.5 K, with an inversion ratio of approximately 2.8, has a theoretical noise temperature of 1.6 K. The noise temperature of the LNA measured at the room-temperature flange is approximately 3.5 K. As the bath temperature of the MASER structure is raised or lowered, the noise temperature is raised or lowered proportionally. Since the MASER structure is mounted on a cold station at 4.5 K, low-loss input and output lines coupled with high thermal isolation were required to achieve the low-noise temperature at the room-temperature input flange. Many innovative low-noise techniques were used over the years to minimize the input line loss, such as gold flashed stainless-steel lines, gapped waveguide structures in the cryogenic environment, antenna horns attached to the intermediate cold station (77 K), etc. In addition, ancillary lines that included the waveguide for the pump source, bias lines for the magnet, etc. were also required. The entire structure was enclosed in a vacuum chamber to minimize the heat load. A photograph of an *X*-band MASER [5] without the compressor and ancillary equipment box is shown in Fig. 4 and demonstrates the complexity of this LNA. An *S*-band MASER operating at 2.2 GHz had a noise temperature of approximately 2 K at its input terminal, a gain of 30 dB, with a bandwidth of 20 MHz. An *X*-band MASER had a noise temperature of 3.5 K at the feedhorn aperture, a gain of 40 dB, and a bandwidth of 100 MHz. As the technology has

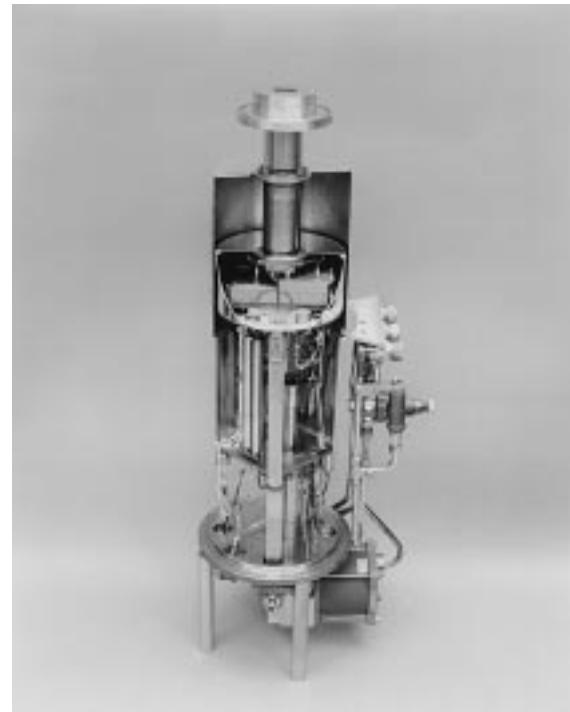


Fig. 4. *X*-band TWM internal cryogenic structure (courtesy of the Jet Propulsion Laboratory, California Institute of Technology).

progressed, *Ka*-band cavity masers are becoming the preferred design approach because of lower pump power requirements and smaller volume when compared to the TWM. For these units, noise temperatures of 5 K with a gain of 35 dB and a bandwidth of 85 MHz have been achieved in an open-cycle liquid-helium system. For future systems, it appears that cavity masers could be the preferred choice over TWMs for future DSN applications [5].

The active element for the maser was a ruby crystal, although other crystals were also used, that was placed in a very high magnetic field, normally provided by a superconducting magnet that was also operated at 4.5 K. The superconducting magnet was used to minimize the thermal load on the refrigerator. It normally was an air-core solenoid made of niobium-zirconium wire. Some of the magnets were as large as 7.5 in in diameter, 5 in in length, and weighed about 3 lb. The magnet generated a magnetic field of about 4500 G with a current of 6.5 A.

Improvements in MASER technology that improved RF performance and its size has been achieved over the years. However, new technologies have started to compete successfully with it. Some of the MASER amplifiers are being supplemented by cryogenically cooled HEMTs that operate at 12 K and lower. While still having the complexity of cryogenic cooling, these refrigerators are less expensive, much smaller, and consume a lot less power. Cryogenically cooled HEMTs have been developed that are now being used in the JPL's DSN. For example, an *X*-band InP HEMT operating in a 6-K refrigerator achieved a noise temperature of 4.5 K. The DSN is now using both MASERS and cryogenically cooled HEMTs in their ground terminals. While the physical structure of the HEMT cryogenically cooled amplifier is still complex, broader instantaneous bandwidths and almost comparable noise

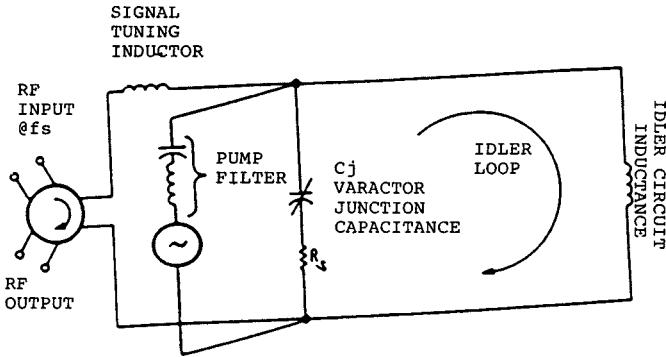


Fig. 5. Paramp circuit diagram.

performance can be obtained. At *Ka*-band, a noise temperature of 16 K has been achieved with an InP HEMT cooled to 6 K. However, further improvements for the HEMTs are anticipated, and, in the future, cryogenically cooled HEMTs may come very close to the noise performance that is now obtained only with the MASER.

III. PARAMPS

In addition to the MASER, the paramp became the amplifier of choice for many low-noise applications. The paramp not only provided extremely good low-noise performance at room temperature, but gave excellent performance when cooled to cryogenic temperatures. Two basic types of paramps were used, i.e., the nondegenerate and degenerate amplifier. The nondegenerate amplifier, a single-sideband amplifier, was commonly used for communication and radar applications. The degenerate amplifier was used in double-sideband applications where the signal was broad-band noise used in radio astronomy applications. Paramps maintained their LNA lead position until the FETs, HEMTs, etc., excelled in their low-noise performance and other attributes. However, the paramp did serve a very unique low-noise market that could not have been served by any other amplifier at the time.

Paramps use the nonlinear capacitance variation of a reverse-biased varactor diode to provide the basic amplification mechanism. As shown in Fig. 5, the varactor is “pumped” at a frequency f_p , which is usually an order of magnitude higher than the signal frequency f_s (for a nondegenerate amplifier). The paramp requires an “idler” circuit that allows current to flow at the difference frequency that is generated by mixing the pump and the signal frequency in the nonlinear capacitance. Through this mechanism, energy from the pump circuit is transferred to the signal frequency, thereby providing gain. Actually, the pumped varactor appears as a negative resistance at the signal frequency. A nonreciprocal three-port ferrite isolator with good isolation shown in Fig. 5 is used to separate the input signal from the output signal.

The noise temperature T_e of the nondegenerate paramp (less the circulator loss) is shown as follows: [6]:

$$T_e = \left[\frac{\frac{f_s}{f_i} + \frac{f_s f_i}{M^2}}{1 - \frac{f_s f_i}{M^2}} \right] T_D$$

where

- f_s signal frequency;
- f_i idler frequency = $f_p - f_s$;
- M varactor figure-of-merit = a/f_c ;
- a varactor nonlinearity ratio;
- f_c varactor cutoff frequency;
- T_D varactor junction temperature.

The varactor cutoff frequency is defined by the following expression:

$$f_c = 1/2\pi R_s C_o$$

where

- R_s varactor equivalent series resistance
- C_o effective operating junction capacitance.

GaAs Schottky varactors with high self-resonant frequencies (>30 GHz) had operating cutoff frequencies in excess of 500 GHz. Varactors generally used GaAs as the intrinsic semiconductor material. The varactors had excellent performance at both room and cryogenic temperatures and a variety of paramps were available for the system design engineer. These LNAs were designed to operate at 20 K, 77 K, thermoelectrically cooled to -40 °C, as well as room temperature. However, similar to the MASER, paramps were relatively complex designs. The paramp generally required a very high-frequency pump source that operated at a millimeter-wave frequency. The pump source had to be temperature stabilized since changes of 0.1 dB in pump power resulted in 1.0-dB changes in gain. In addition, the circulator and paramp mount were generally integrated to control the impedance match between them to achieve unconditional stability. The circulator and paramp also had to be temperature stabilized. In addition, depending on the type of paramp used, they could be relatively large, heavy, and consumed a lot of power. A photograph of a two-stage cryogenically cooled paramp (without the compressor) that operated at 20 K is shown in Fig. 6. Over the operating frequency range of 3.7–4.2 GHz that was used for a satellite communication ground terminal, this LNA had a gain of 30 dB and noise temperature around 15 K. As varactor cutoff frequencies and circulator insertion loss improved, noise temperatures that could be obtained at room temperature were lowered. At the same time, transmitter power in the satellites was increasing. With the exception of the radio astronomers, the trend was to develop smaller, lighter weight, more economical paramps that could still meet the G/T_{op} system requirements. Thermoelectric coolers were also becoming more efficient and reliable. As a result, many new paramp designs were being developed using thermoelectric coolers where the paramp mount and circulator were cooled to about -40 °C. A thermoelectrically cooled paramp (Fig. 7) that was cooled to -38 °C had a noise temperature of 30 K over the 3.7–4.2-GHz frequency band. Comparing these two LNAs (Fig. 6 versus Fig. 7), it is quite obvious that the size and weight for the TE cooled paramp was much less than for the cryogenically cooled paramp. Since the TE cooled paramp met the ground-station requirements, they were used in the new systems and were replacements for the cryogenically cooled units. Paramps operated successfully to about 20 GHz and a little beyond. Paramps were developed that

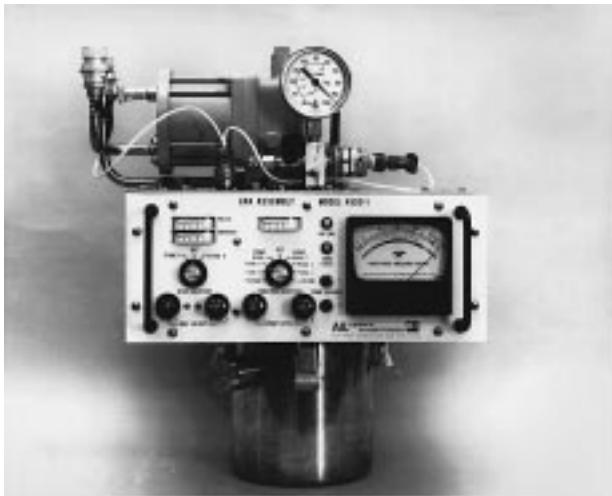


Fig. 6. Cryogenically cooled paramp (courtesy of the EDO Corporation).

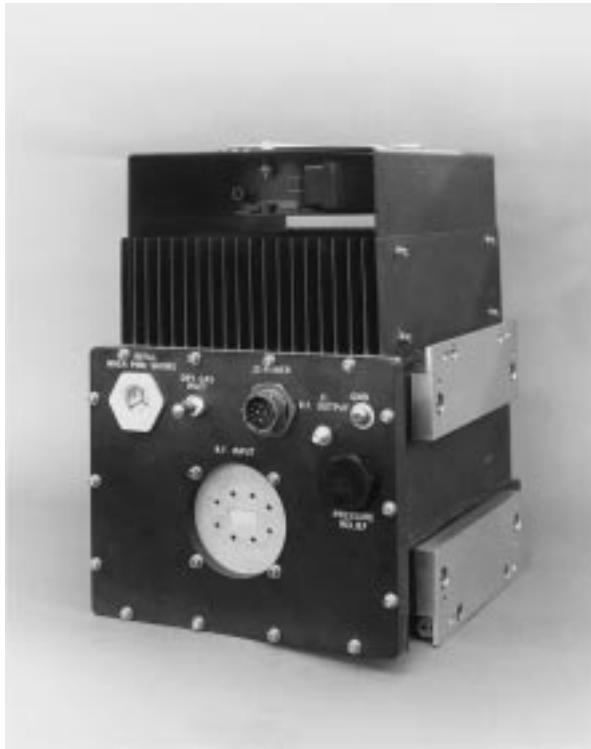


Fig. 7. Thermoelectrically cooled paramp (courtesy of the EDO Corporation).

operated at 60 GHz [7] and as high as 94 GHz [8], but these were the exception, and not the rule. However, the coming of age of the FETs, HEMTs, etc., was to change all of this. It was not long before the room-temperature and TE-cooled HEMT became the workhorse for many old applications, and enabled the LNAs of choice for new ones, thereby replacing the paramp.

Since a paramp requires a nonlinear capacitance for its mode of operation, it is basically a relatively narrow-band device. However, the paramp was capable of providing much greater bandwidths than the MASER. To achieve the broadest possible instantaneous bandwidth, the balanced amplifier was developed that used the self-resonant frequency of the varactor for the idler circuit. This paramp had a bandwidth of approximately 10%–15%.

One application that tested the performance and reliability of the paramp was the *S*-band communication link for the space shuttle. The space shuttle was required to communicate with the tracking ground terminals during liftoff and in orbit either directly or through the TDRSS satellite. The *S*-band communication antennas are located at the front-end of the shuttle. The *S*-band preamplifier assembly (LRU) is located on a shelf away from the antennas and, therefore, has a relatively long interconnecting line between them. The LRU had a transmit/receive diplexer, a redundant paramp followed by a bipolar amplifier. Size prevented the paramp from being integrated more closely with the antenna, and a lower noise LNA was required to meet the performance objectives. The paramp was the only LNA at the time that could meet these noise-figure requirements. The paramp was pumped at 50 GHz with a thermally stabilized Gunn oscillator. The Gunn oscillator was isolated from the paramp mount by a small section of gold flashed plastic waveguide. The four-port circulator used fiberglass ground planes that were gold plated. This innovative design served the dual purpose of both reducing the power consumption to achieve fast warmup, as well as meeting the low insertion loss to meet the noise-figure requirements. The redundant paramp performed extremely well over these many years. However, due to obsolescence of many of the component parts required for the paramp, a replacement for the paramp was required. After over 20 years of successful operation, the redundant paramp is finally starting to be replaced by a high reliability redundant HEMT amplifier that meets the 1.3-dB noise figure at the highest operating temperature. However, until the retrofit is complete, the majority of shuttles today still operate with a paramp in the *S*-band communication link. A photograph of a single paramp and its HEMT replacement that will be eventually installed in the space shuttle LRU is shown in Fig. 8. The noise figure for the HEMT amplifier is about the same as the paramp, but the size, weight, and power have been significantly reduced. This figure clearly demonstrates that not only has the RF performance been at least equaled, but that the LNA size and weight has been greatly reduced. Power dissipation has also been significantly reduced. The reliability of the system is also increased due to the significant reduction in the number of components and junction operating temperature. With further improvements in MMIC technology to improve overall RF performance, a further reduction in size, and providing more functions on a single chip, the LNA could eventually be integrated with the antenna to achieve optimum performance. The new technology introduced by the FETs, HEMTs, and MMIC components were rapidly changing LNA performance. These new technologies were to open the way for many new applications that were not even considered or possible during the brief age of paramps.

IV. FETs AND HEMTs

The new technological developments in LNAs that started before the 1980s has rapidly progressed to the present day. This continuous improvement in three-terminal devices (i.e., FETs, HEMTs, etc.) has led to better performance in noise temperature, gain, bandwidth, frequency of operation, power

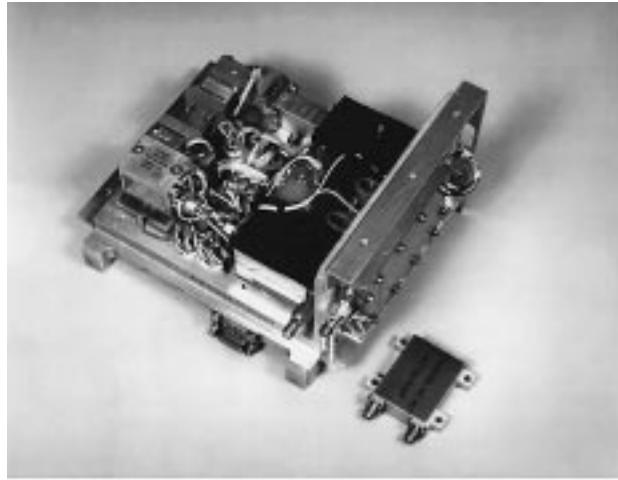


Fig. 8. Space shuttle paramp and its HEMT replacement (courtesy of the EDO Corporation).

dissipation, size, and reliability, as a minimum. With these developments, we have greatly expanded the capabilities of both our present and new systems. The performance of FETs in the early 1980s has been shown in Fig. 1. Prior to that time, GaAs FETs were developed for low-noise performance, clearly having better performance than the bipolar transistor. MESFETs and HEMTs were also under development. For example, in 1972 [9], it was reported that FETs had achieved a F_{max} in excess of 30 GHz. A typical device at 4 GHz with a G_{max} of 11 dB had a noise figure of 3 dB. At 8 GHz, a transistor achieved a G_{max} of 7 dB with a noise figure of 4 dB. From these early developments, rapid improvements in semiconductor materials and processes led to greatly improved performance. If we consider the equation for minimum noise figure for FETs, we can readily determine where improvements in device parameters are required.

For FETs, the minimum noise figure is described by [10]

$$F_{\text{min}} = 1 + k_1 f C_{gs} [(R_g + R_s)/gm]^{1/2}$$

where

- F_{min} minimum noise figure;
- k_1 fitting factor;
- f frequency of operation;
- C_{gs} gate/source capacitance;
- gm transconductance;
- R_g gate resistance;
- R_s source resistance.

As we can see, a reduction in C_{gs} as well as increasing the transconductance is beneficial to reducing the minimum noise figure. In fact, transconductance has been dramatically increased over the years. This increase has been achieved as we went from FETs, to HEMTs, to p-HEMTs, and then to InP HEMTs. By 1974, MESFETs (GaAs FETs with Schottky-barrier gates) were reported that combined low noise with high gain and high dynamic range [11]. A broad-band amplifier with a gain of 20 dB over the entire 8.0–12.0-GHz frequency band with a noise figure of 5.5 dB was developed. We were then introduced to the HEMTs. HEMTs have higher cutoff frequencies than FETs due to their higher electron mobility and,

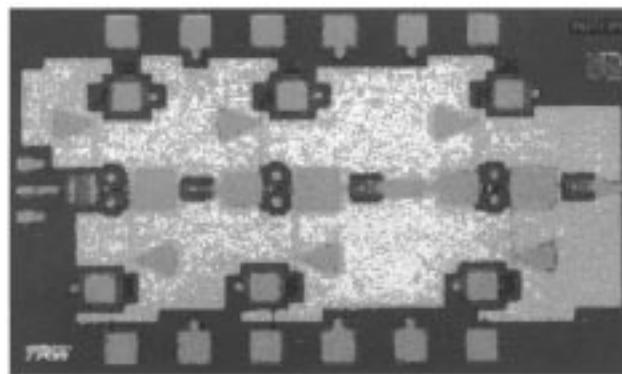


Fig. 9. Three-stage MMIC LNA at 183 GHz [12].

therefore, has a lower noise figure. We were then acquainted with the p-HEMTs, and now the InP HEMTs. As the devices improved, the achievable noise figure in the microwave region was continuously lowered, while, at the same time, the frequency of operation was moving higher due to the increase in transconductance and the reduction in gate length. In addition, the devices could be cryogenically cooled and, therefore, extremely low-noise temperatures were achieved. The high transconductance (≈ 1700 mS/mm) for the InP HEMTs has increased the operating frequency range, gain bandwidth, and provided the low-noise performance for LNAs. InP HEMTs have a gm of about 1000 mS/mm, which is almost twice the value of p-HEMTs. Recently, the performance of a MMIC LNA that operates at 183 GHz with a gain of 20 dB over a 30-GHz bandwidth has been reported [12]. The noise figure of the LNA was less than 5.5 dB. A photograph of the three-stage MMIC LNA is shown in Fig. 9. The amplifier used $0.08\text{-}\mu\text{m}$ gate InP MMIC technology. The device has a transconductance that is greater than 1000 mS/mm, a cutoff frequency above 200 GHz, and an oscillation frequency above 400 GHz. Similar performance is also being reported by other researchers [13]. Remarkable progress has been made in LNAs over these years and we expect to have even more breakthroughs in the near future.

It is virtually impossible to enumerate all the advances that have been made in this paper. For this reason, we will highlight some of the achievements that have been reported. In Table I, [13], [14], we have tabulated the performance of both discrete/MMIC LNAs that have been reported. This is not a complete list, but does show what has been accomplished in a relatively short time.

The HEMTs, p-HEMTs, and InP HEMTs devices have been continuously improved to significantly lower the noise performance of LNAs. By cooling these LNAs in TE coolers or cryogenic systems, low-noise performance has been dramatically improved. As stated previously, cryogenically cooled HEMTs are now being used in the JPL DSN and are coming close to competing with the MASER. At X-band, an InP HEMT cooled to an operating temperature of 6 K had a noise temperature of 4.5 K at the aperture of the feedhorn [5]. At 32 GHz, a cryogenically cooled HEMT had a noise temperature of 16 K when cooled to the same 6 K. A MMIC LNA at 90 GHz had a noise temperature of 62 K when cooled to 27 K [27]. A broad-band

TABLE I

SUMMARY OF DISCRETE/MMIC LNAs					
FREQ GHZ	NOISE dB	GAIN- dB	YEAR	CRYO	REF.
2.3-2.5	0.4	35	1993	No	15
7.0-11.0	1	21	1993	No	16
19.0-22.0	1.1	38	1995	No	17
20-25	0.4	35	2000	YES	18
43.0-46.0	1.9	22	1995	No	19
50	2.8	9	1994	No	20
60	2.2	22.8	2000	No	21
56.0-60.0	3.2	15	1992	No	22
56.6-64.0	2.7	25	1993	No	23
60-80	2.3	25	2001	No	24
73-77	0.4	25	2001	YES	24
75.0- 110.0	6	23	1993	No	25
92.0-96.0	3.3	20	1995	No	26
82-106	1.6	35	2000	YES	18
183	5.5	24	2001	No	12

InP MMIC LNA that covered the 65–110 GHz had a noise temperature of 45 K across the operating bandwidth. This amplifier had a noise temperature of 30 K at a frequency of 102 GHz [28]. We can only speculate where further improvements in transistor technology will eventually lead us. However, if we review what has been recently accomplished, we can only look to the future with a great deal of anticipation.

V. CONCLUSION

We have seen dramatic improvements over these few years, not only in lowering noise-figure performance, but also in extending the frequency to which LNAs now operate, the broad bandwidths, as well as reductions in size, weight, and power. We can only speculate on what the future will bring. Further reductions in noise figure coupled with lower power dissipation is one of the thrusts today. However, advances in other disciplines will also drive the performance of the low-noise LNA if the technological developments come to pass. For example, if HTSC could eventually operate at room temperature, not only would we see dramatic improvements in LNAs, but also in the system architectures. However, this is but one of many future development to which we can look forward.

ACKNOWLEDGMENT

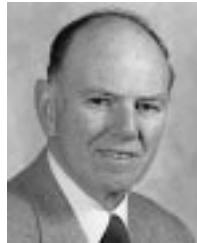
The author acknowledges the helpful support of S. Petty, Jet Propulsion Laboratory (JPL), Pasadena, CA, and H. Paczkowski, EDO Corporation, Deer Park, NY, in supplying the photographs for this paper. The author would also like to

acknowledge the help provided by P. M. Smith, BAE Systems, Nashua, NH.

REFERENCES

- [1] S. Okwit, "An historical view of the evolution of low-noise concepts and techniques," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, pp. 1068–1082, Sept. 1984.
- [2] W. Menzel *et al.*, "A 77 GHz FM/CW radar frontend with a low-profile, low-loss printed antenna," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1999, pp. 1484–1488.
- [3] J. S. Shell *et al.*, "Ruby masers for maximum G/T_{op} ," *Proc. IEEE*, vol. 82, pp. 796–810, May 1994.
- [4] J. A. De Gruyl *et al.*, "Noise considerations in broad-band traveling-wave masers," *IEEE Trans. Microwave Theory Tech.*, vol. 16, pp. 586–595, Sept. 1968.
- [5] J. Bautista *et al.*, "DSN low noise amplifiers in the new millennium," *TMOD Technol. Sci. Program News*, vol. 82, no. 5, Jan. 2001.
- [6] P. P. Penfield and R. P. Rafuse, *Varactor Applications*. Cambridge, MA: MIT Press, 1962.
- [7] J. Whelehan *et al.*, "A nondegenerate millimeter wave parametric amplifier with a solid state pump source," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1973, pp. 75–77.
- [8] H. C. Okean *et al.*, "Low noise, 94 GHz parametric amplifier development," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1973, pp. 72–79.
- [9] L. Besser, "Design considerations of a 3.1 to 3.5 GHz GaAs FET feedback amplifier," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1972, pp. 230–232.
- [10] H. Fukui, "Design of microwave GaAs MESFET's for broad-band low-noise amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-27, pp. 643–650, July 1979.
- [11] A. Liechti *et al.*, "Design and performance of microwave amplifiers with GaAs Schottky-gate field-effect transistors," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-22, pp. 510–517, May 1974.
- [12] R. Rajir *et al.*, "183 GHz low noise amplifier module with 5.5 dB noise figure for the Conical Scanning Microwave Imager Sounder (CMIS) program," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 2001.
- [13] P. M. Smith *et al.*, "Advances in InP HEMT technology for high frequency applications," presented at the IEEE GaAs IC Symp. Plenary Session, Oct. 2001.
- [14] P. M. Smith, "Status of InP HEMT technology for microwave receiver applications," *IEEE Trans. Microwave Theory Tech.*, vol. 44, pp. 2328–2333, Dec. 1996.
- [15] S. E. Rosenbaum *et al.*, "A 2 GHz three stage AlInAs-InP HEMT MMIC low noise amplifier," *IEEE Microwave Guided Lett.*, vol. 3, pp. 265–267, Aug. 1993.
- [16] S. E. Rosenbaum *et al.*, "A 7 to 11 GHz AlInAs/GaInAs/InP MMIC low noise amplifier," in *IEEE MTT-S Microwave Symp. Dig.*, vol. 2, 1993, pp. 1103–1104.
- [17] K. H. G. Duh *et al.*, "Application of GaAs and InP-based HEMT technology to MILSATCOM systems," in *IEEE MILCOM Conf.*, vol. 2, San Diego, CA, 1995, pp. 731–736.
- [18] M. W. Pospieszalski *et al.*, "Design and performance of wideband, low-noise, millimeter-wave amplifiers for microwave anisotropy probe radiometers," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, 2000, pp. 25–28.
- [19] R. Isobe *et al.*, "Q and V-band MMIC chip set using 0.1 μm millimeter wave low noise InP HEMTs," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 3, 1995, pp. 1133–1136.
- [20] Y. Umeda *et al.*, "Sensitivity analysis of 50 GHz MMIC LNA on gate recess depth with InAlAs/InGaAs/InP HEMTs," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, 1994, pp. 123–126.
- [21] A. Fujihara *et al.*, "High performance 60-GHz coplanar MMIC LNA using InP heterojunction FET's with AlAs/InAs super lattice layer," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, 2000, pp. 21–24.
- [22] R. T. Webster *et al.*, "Monolithic InP HEMT V-band low noise amplifier," *IEEE Microwave Guided Wave Lett.*, vol. 2, pp. 236–238, June 1992.
- [23] R. Lai *et al.*, "A high performance and low dc power V-band MMIC LNA using 0.1 μm InGaAs/InAlAs/InP HEMT technology," *IEEE Microwave Guided Wave Lett.*, vol. 3, pp. 447–449, Dec. 1993.

- [24] P. Kangaslahti *et al.*, "Low noise amplifiers in InP technology for pseudo correlating millimeter wave radiometer," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 3, 2001, pp. 1959–1962.
- [25] H. Wang *et al.*, "A monolithic 75–110 GHz balanced InP based HEMT amplifier," *IEEE Microwave Guided Lett.*, vol. 3, pp. 381–383, Oct. 1993.
- [26] G. I. Ng *et al.*, "A fully passivated ultra low noise W-band monolithic InGaAs/InAlAs/InP HEMT amplifier," in *IEEE Microwave Millimeter Wave Monolithic Circuits Symp. Dig.*, 1995, pp. 63–66.
- [27] M. Pospieszalski, "Millimeter-wave waveguide bandwidth cryogenically-coolable InP HEMT amplifiers," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 3, 1997, pp. 1285–1288.
- [28] S. Weinreb *et al.*, "W-band InP wideband MMIC LNA with 30 K noise temperature," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, 1999, pp. 101–104.



James J. Whelehan (S'59–M'60–SM'84–F'88–LF'96) received the B.E.E. degree and the M.S. degree in electrical engineering from the Polytechnic Institute of Brooklyn (PIB), Brooklyn, NY, in 1959 and 1965, respectively. He also took additional courses at PIB and CW Post, Brookville, NY, working toward the Ph.D. degree.

In 1959, he joined the staff of the Airborne Instruments Laboratory (AIL), where he was involved in the design and development of LNAs. While with AIL, he was responsible for the design and develop-

ment of microwave and millimeter-wave receivers, radiometers and specialized components for electronic warfare (EW), ground, and space applications. He was also responsible for developing many of these products for military and commercial applications. When he left AIL in 1995 as their Director of Technology, he was responsible for the design and development of receivers and transmitters from the microwave through the infrared/electrooptic (IR/EO) frequency spectrum. He then formed his own consulting company, JJW Consulting Inc., Hatfield, PA, where he specializes in the development of microwave and millimeter-wave receivers. He has authored or co-authored over 30 technical papers.

Mr. Whelehan headed the Technical Committee on low noise for the IEEE Microwave Theory and Techniques Society (IEEE MTT-S) for over 20 years. He was also the technical co-chairman of the 1988 IEEE MTT-S International Microwave Symposium.

Chapter 12

Parametric Amplifiers and Oscillator

A device exhibiting a negative conductance, such as a tunnel diode, can be utilized to construct an amplifier and oscillator. A laser is also categorized as a negative conductance oscillator as we have seen in the previous chapter. There is another class of amplifier and oscillator, which is based on non-linear susceptances and known as a parametric amplifier/oscillator.

For instance, a reverse-biased pn junction has the non-linear charge-voltage characteristic due to the voltage-dependent capacitance. The mixing occurs between the three frequency components of signal, idler and pump waves in such a nonlinear element and the energy flows from a strong pump wave to weak signal and idler waves. This flow of the power from the pump to the signal introduces the negative conductance into the signal circuit. In optical spectral domain, the atomic dipole moment, driven by an intense pump laser, features a similar non-linearity and is capable of amplifying weak signal and idler waves.

12.1 Non-Degenerate Parametric Amplifier

12.1.1 Principle of Operation

An equivalent circuit for a non-degenerate parametric amplifier is shown in Fig. 12.1. A nonlinear capacitor is surrounded by three parallel LCR circuits, which represent the signal, idler and pump circuits, respectively.

The charge q on the nonlinear capacitance is a function of the voltage across its terminals. Using the Taylor series expansion, the charge may be expressed in the form:

$$q(t) = a_1 v(t) + a_2 v^2(t) + a_3 v^3(t) + \dots . \quad (12.1)$$

When all the coefficients except the first and second terms are zero, the charge varies quadratically with the voltage,

$$q(t) = C v(t) + a_2 v^2(t) , \quad (12.2)$$

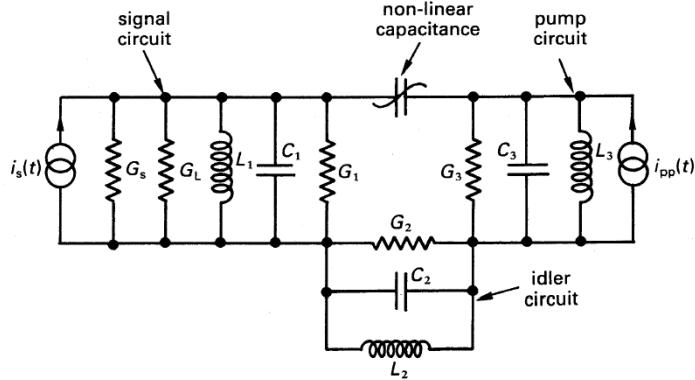


Figure 12.1: An equivalent circuit of a non-degenerate parametric amplifier.

where a_1 is just replaced by the linear capacitance C . The current flowing in the nonlinear capacitance is

$$i(t) = \frac{dq(t)}{dt} = C \frac{dv(t)}{dt} + 2a_2 v(t) \frac{dv(t)}{dt} , \quad (12.3)$$

where the voltage across the nonlinear capacitance consists of the signal, idler and pump waves at angular frequencies ω_1 , ω_2 , and ω_3 , respectively,

$$\begin{aligned} v(t) &= v_1(t) + v_2(t) + v_3(t) \\ &= V_1 \cos(\omega_1 t + \phi_1) + V_2 \cos(\omega_2 t + \phi_2) + V_3 \cos(\omega_3 t + \phi_3) . \end{aligned} \quad (12.4)$$

The angular frequencies in Eq. (12.4) satisfy

$$\omega_3 = \omega_1 + \omega_2 , \quad (12.5)$$

$$\omega_i = 1/\sqrt{L_i(C_i + C)} . \quad (12.6)$$

Using Eq. (12.4) in Eq. (12.3), we obtain the expression for the current,

$$i(t) = i_1(t) + i_2(t) + i_3(t) , \quad (12.7)$$

where

$$i_1(t) = -\omega_1 C V_1 \sin(\omega_1 t + \phi_1) - \omega_1 a_2 V_2 V_3 \sin(\omega_1 t + \phi_3 - \phi_2) , \quad (12.8)$$

$$i_2(t) = -\omega_2 C V_2 \sin(\omega_2 t + \phi_2) - \omega_2 a_2 V_1 V_3 \sin(\omega_2 t + \phi_3 - \phi_1) , \quad (12.9)$$

$$i_3(t) = -\omega_3 C V_3 \sin(\omega_3 t + \phi_3) - \omega_3 a_2 V_1 V_2 \sin(\omega_3 t + \phi_1 + \phi_2) . \quad (12.10)$$

Equations (12.8)-(12.10) can be rewritten as

$$i_1(t) = C \frac{dv_1(t)}{dt} + \frac{a_2 V_2 V_3}{V_1} \left\{ \cos(\phi_3 - \phi_2 - \phi_1) \frac{dv_1(t)}{dt} - \omega_1 v_1(t) \sin(\phi_3 - \phi_2 - \phi_1) \right\} , \quad (12.11)$$

$$i_2(t) = C \frac{dv_2(t)}{dt} + \frac{a_2 V_1 V_3}{V_2} \left\{ \cos(\phi_3 - \phi_2 - \phi_1) \frac{dv_2(t)}{dt} - \omega_2 v_2(t) \sin(\phi_3 - \phi_2 - \phi_1) \right\} , \quad (12.12)$$

$$i_3(t) = C \frac{dv_3(t)}{dt} + \frac{a_2 V_1 V_2}{V_3} \left\{ \cos(\phi_3 - \phi_2 - \phi t_1) \frac{dv_3(t)}{dt} + \omega_3 v_3(t) \sin(\phi_3 - \phi_2 - \phi_1) \right\} . \quad (12.13)$$

Taking the Fourier transform of Eqs. (12.11)-(12.13), we obtain the admittances Y_i ($i = 1, 2, 3$) seen by the signal, idler and pump circuits:

$$Y_1 = \frac{I_1(j\omega)}{V_1(j\omega)} = j\omega_1 C + j\omega_1 a_2 \frac{V_2 V_3}{V_1} \exp[j(\phi_3 - \phi_2 - \phi_1)] , \quad (12.14)$$

$$Y_2 = \frac{I_2(j\omega)}{V_2(j\omega)} = j\omega_2 C + j\omega_2 a_2 \frac{V_1 V_3}{V_2} \exp[j(\phi_3 - \phi_2 - \phi_1)] , \quad (12.15)$$

$$Y_3 = \frac{I_3(j\omega)}{V_3(j\omega)} = j\omega_3 C + j\omega_3 a_2 \frac{V_1 V_2}{V_3} \exp[-j(\phi_3 - \phi_2 - \phi_1)] . \quad (12.16)$$

The current-voltage relations for the three circuits are given by

$$I_s(j\omega) = \left\{ G_T + j\omega_1 a_2 \frac{V_2 V_3}{V_1} \exp[j(\phi_3 - \phi_2 - \phi_1)] \right\} V_1(j\omega) , \quad (12.17)$$

$$O = \left\{ G_2 + j\omega_2 a_2 \frac{V_1 V_3}{V_2} \exp[j(\phi_3 - \phi_2 - \phi_1)] \right\} V_2(j\omega) , \quad (12.18)$$

$$I_P(j\omega) = \left\{ G_3 + j\omega_3 a_2 \frac{V_1 V_2}{V_3} \exp[-j(\phi_3 - \phi_2 - \phi_1)] \right\} V_3(j\omega) . \quad (12.19)$$

Here $I_s(j\omega)$ and $I_P(j\omega)$ are the Fourier transforms of the input signal and pump currents, respectively, and $G_T = G_s + G_L + G_1$. The LC circuit resonant condition Eq. (12.6) is used.

By eliminating V_2 and V_3 from Eq. (12.17) using Eqs. (12.18) and (12.19), we obtain the admittance of the signal circuit,

$$Y_s = G_T - G = G_T - \frac{\omega_1 \omega_2 a_2^2}{G_2 G_3^2} \frac{|I_P(j\omega)|^2}{\left[1 + \frac{\omega_2 \omega_3}{G_2 G_3} a_2^2 V_1^2 \right]^2} . \quad (12.20)$$

There emerges a negative conductance due to the nonlinear capacitance driven by the pump wave at ω_3 . If V_1 satisfies the condition,

$$\frac{\omega_2 \omega_3}{G_2 G_3} a_2^2 V_1^2 \ll 1 , \quad (12.21)$$

the negative conductance is independent of the signal input and the linear parametric amplification is realized.

12.1.2 Power Gain

The power gain G of the non-degenerate parametric amplifier is given by the ratio of the power delivered to the load G_L to the input power to the source G_s :

$$\begin{aligned} G &= \frac{G_L V_1^2}{(|I_s|^2 / 4G_s)} \\ &= \frac{4G_s G_L}{|Y_s|^2} . \end{aligned} \quad (12.22)$$

When there is no pump ($|I_P(j\omega)| = 0$), the amplifier has no gain ($Y_s = G_T$). When the pump current reaches the threshold:

$$|I_P(j\omega)|^2 = \frac{G_T G_2 G_3^2}{\omega_1 \omega_2 a_2^2} , \quad (12.23)$$

the system becomes unstable ($|Y_s| \rightarrow 0$) and the amplifier starts to oscillate. Between these two extreme conditions, linear amplification of the input signal is provided as far as the signal is not too strong, i.e. Eq. (12.21) is satisfied.

12.1.3 Noise Figure

The noise in a parametric amplifier is generated by the circuit conductance G_s , G_1 and G_2 . The noise generated by the pump circuit conductance G_3 can be normally neglected because the pump current $i_P(t)$ is usually very large and well approximated as a noise-free sinusoidal wave. The noise from the load conductance G_L is ignored, because it is usually taken into account in the following state.

Equation (12.17) suggests that a voltage fluctuation ΔV_2 across the idler circuit at frequency ω_2 results in a current fluctuation ΔI_s in the signal circuit at frequency ω_1 . The spectral density of the voltage fluctuation ΔV_2 is given by

$$\overline{S_{V_2}(\omega)} = \begin{cases} 4k_B T / G_2 & (\text{thermal limit}) \\ 2\hbar\omega_2 / G_2 & (\text{quantum limit}) \end{cases} . \quad (12.24)$$

The spectral density of the induced current fluctuation ΔI_s is given by

$$\overline{S_{I_{s2}}(\omega)} = \omega_1^2 |C'|^2 \overline{S_{V_2}(\omega)} , \quad (12.25)$$

where

$$|C'| = a_2 V_3 \simeq a_2 |I_P(j\omega)| / G_3 . \quad (12.26)$$

The second equality is obtained by neglecting the gain saturation effect Eq. (12.21).

The spectral densities of the current generators associated with G_s and G_1 are

$$\overline{S_{I_{ss}}(\omega)} = \begin{cases} 4k_B T G_s & \\ 2\hbar\omega_1 G_s & \end{cases} , \quad (12.27)$$

$$\overline{S_{I_{s1}}(\omega)} = \begin{cases} 4k_B T G_1 & \\ 2\hbar\omega_1 G_1 & \end{cases} . \quad (12.28)$$

Since there is no correlation between these three noise sources, the noise figure of the amplifier in the thermal limit can be written as

$$\begin{aligned} F &= \frac{\overline{S_{I_{ss}}(\omega)} + \overline{S_{I_{s1}}(\omega)} + \overline{S_{I_{s2}}(\omega)}}{\overline{S_{I_{ss}}(\omega)}} \\ &= 1 + \frac{G_1}{G_s} + \frac{\omega_1^2 |C'|^2}{G_2 G_s} . \end{aligned} \quad (12.29)$$

From Eqs. (12.20) and (12.26), we can express $|C'|^2$ in terms of the negative conductance G ,

$$|C'|^2 = \frac{G_2 G}{\omega_1 \omega_2} . \quad (12.30)$$

Using Eq. (12.30) in Eq. (12.29), the noise figure is expressed as

$$F = 1 + \frac{G_1}{G_s} + \frac{\omega_1}{\omega_2} \frac{G}{G_s} . \quad (12.31)$$

The noise figure can be reduced to one (ideal amplification) by achieving the negligible internal loss in the signal circuit ($G_1 \ll G_s$) and the large ratio of $\omega_1/\omega_2 \ll 1$.

The noise figure of the amplifier in the quantum limit is, on the other hand, given by

$$F = 1 + \frac{G_1}{G_s} + \frac{G}{G_s} . \quad (12.32)$$

In a high gain amplifier $G \simeq G_s$ ($G_1, G_L \ll G_s$) at the quantum limit, the minimum noise figure is $F_{min} = 2(3$ dB) instead of $F_{min} = 1(0$ dB) at the thermal limit.

12.2 Degenerate Parametric Amplifier

12.2.1 Principle of Operation

When the signal and idler waves have identical frequencies, such a parametric amplifier is called a degenerate parametric amplifier and has a unique characteristic. Consider a swing driven by a person (Fig. 12.2(a)). During one-half cycle (left to right) of the swing, the person makes a full one cycle (up-down-up). The frequency of the driving person (pump) and that of the driven swing (signal) satisfy $\omega_P = 2\omega_s$. Figure 12.2(b) is an equivalent LCR circuit of the swing, in which the driving action of the person is represented by the nonlinear capacitor.

The circuit equations are given by

$$\begin{aligned} -I &= \frac{d}{dt} Q = \frac{d}{dt} C V \\ V &= R I + L \frac{d}{dt} I \end{aligned} \quad (12.33)$$

Eliminating the current I from Eq. (12.33), we obtain

$$\left[\frac{d^2}{dt^2} + \frac{R}{L} \frac{d}{dt} + \frac{1}{LC} \right] V = 0 \quad (12.34)$$

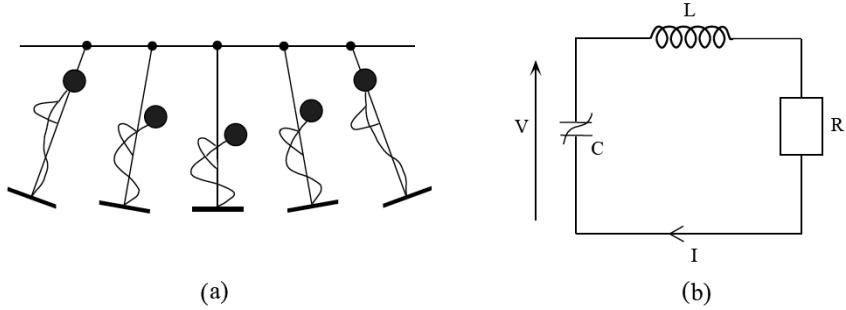


Figure 12.2: A swing driven by a person and an equivalent circuit.

The solution of a damped harmonic oscillator expressed by Eq. (12.34) is

$$V = V_0 \exp\left(-\frac{Rt}{2L}\right) \exp\left(\pm j\sqrt{\omega_0^2 - \frac{R^2}{4L^2}}t\right) . \quad (12.35)$$

If the capacitance is modulated at the pump frequency ω_P as

$$C = C_0 [1 - \Delta C \sin(\omega_P t + \phi)] , \quad (12.36)$$

Equation (12.34) is modified to

$$\left\{ \frac{d}{dt^2} + \frac{R}{L} \frac{d}{dt} + \omega_0^2 \left[1 + \frac{\Delta C}{C_0} \sin(\omega_P t + \phi) \right] \right\} V \simeq 0 , \quad (12.37)$$

where $\omega_0 = 1/\sqrt{LC_0}$ and it is assumed $\Delta C \ll C_0$. If we assume the solution of Eq. (12.37) has the form,

$$V = Re \{ 2V_0 \exp(\alpha t) \exp(j\omega t) \} , \quad (12.38)$$

we obtain

$$\begin{aligned} 2V_0 \exp(\alpha t) & \left[\left(\alpha^2 - \omega^2 + \frac{R}{L} \alpha + \omega_0^2 \right) \cos(\omega t) - \left(2\omega\alpha + \frac{R}{L}\omega \right) \sin(\omega t) \right] \\ & = -2V_0 \exp(\alpha t) \frac{\omega_0^2 \Delta C}{2C_0} [\sin \omega t \cos \phi + \cos \omega t \sin \phi] . \end{aligned} \quad (12.39)$$

By comparing the $\cos \omega t$ and $\sin \omega t$ terms in both sides of Eq. (12.39), we have the equations which determine the new oscillation frequency ω and amplification/attenuation coefficient α :

$$\omega^2 = \omega_0^2 + \alpha^2 + \frac{R}{L}\alpha + \frac{\omega_0^2 \Delta C}{2C_0} \sin \phi , \quad (12.40)$$

$$2\alpha = \frac{\omega_0 \Delta C}{2C_0} \cos \phi - \frac{R}{L} . \quad (12.41)$$

If $\phi = 0$ and $\Delta C > \frac{2RC_0}{\omega_0 L}$, we have a growing solution ($\alpha > 0$). The energy is provided to the signal from the pump. If $\phi = \frac{\pi}{2}$ or $\frac{3\pi}{2}$, there is no energy exchange between the pump and signal waves. If $\phi = \pi$, we have an attenuating solution ($\alpha < -\frac{R}{L}$). The energy is extracted from the signal and transferred to the pump.

12.2.2 Phase Sensitive Amplifier

Changing the pump phase from $\phi = 0$ to $\phi = \pi$ in Eq. (12.36) corresponds to shifting the capacitance modulation by half a pump period, which is equivalent to one-quarter signal period. That is, one quadrature amplitude of the signal wave corresponding to the $\phi = 0$ solution is amplified by a gain coefficient $2\alpha = \frac{\omega_0 \Delta C}{2C_0} - \frac{R}{L}$ but the other quadrature amplitude corresponding to the $\phi = \pi$ solution is deamplified by an attenuation coefficient $2\alpha' = -\frac{\omega_0 \Delta C}{2C_0} - \frac{R}{L}$. This type of operation is called a phase sensitive amplifier. If the signal wave is expressed by the two quadrature amplitudes a_1 and a_2 as

$$E_s = a_1 \cos \omega_s t + a_2 \sin \omega_s t , \quad (12.42)$$

and the pump phase is set to amplify the $\cos \omega_s t$ component and deamplify the $\sin \omega_s t$ component, the two-kinds of input signals with isotropic (phase insensitive) noise are transformed to the squeezed state as shown in Fig. 12.3. When the input noise is dominated by thermal noise, the process is called thermal noise squeezing or classical squeezing. When the input noise is dominated by quantum mechanical zero-point noise, the process is called quantum noise squeezing.

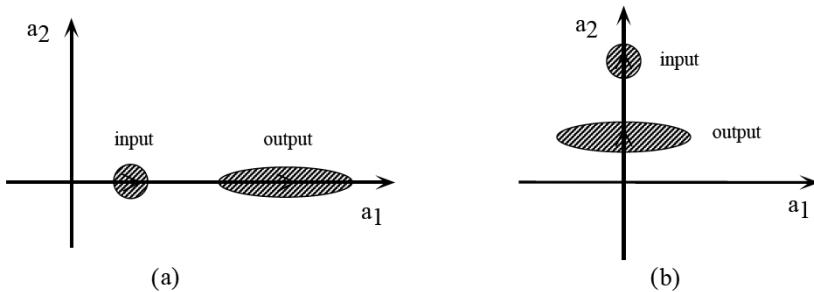


Figure 12.3: The input and output signals of a degenerate parametric amplifier.

12.3 Quantum Limit of a Linear Amplifier

The simplified input-output relations for a degenerate parametric amplifier are given by

$$b_{s1} = \sqrt{G} a_{s1} , \quad (12.43)$$

$$b_{s2} = \frac{1}{\sqrt{G}} a_{s2} , \quad (12.44)$$

where $a_{s1}(b_{s1})$ and $a_{s2}(b_{s2})$ are the $\cos \omega_s t$ and $\sin \omega_s t$ components of the input (output) signal waves. Equation (12.43) indicates that the amplification of one quadrature component does not introduce any additional noise. The noise figure of this amplifier is

$$F = \frac{\langle \Delta b_{s1}^2 \rangle}{G \langle \Delta a_{s1}^2 \rangle} = 1 \quad (0 \text{dB}) . \quad (12.45)$$

The sacrifice of the noise-free amplification is the loss of the signal information stored in the other quadrature a_{s2} since that quadrature component is deamplified.

The input-output relations for a nondegenerate parametric amplifier are given by

$$b_{s1} = \sqrt{G}a_{s1} + \sqrt{G-1}a_{i1} , \quad (12.46)$$

$$b_{s2} = \sqrt{G}a_{s2} - \sqrt{G-1}a_{i2} , \quad (12.47)$$

where a_{i1} and a_{i2} are the $\cos\omega_i t$ and $\sin\omega_i t$ components of the input idler wave. The nondegenerate parametric amplifier allows the extraction of the two quadrature information simultaneously, but the amplifier introduces the additional noise. The minimum noise figure in this case is

$$F = \frac{\langle \Delta b_{s1}^2 \rangle}{G \langle \Delta a_{s1}^2 \rangle} = 2 \quad (3 \text{ dB}) , \quad (12.48)$$

where it is assumed that the signal and idler carries the identical noise, i.e. $\langle \Delta a_{s1}^2 \rangle = \langle \Delta a_{i1}^2 \rangle$. This typical situation corresponds to the case that the input signal wave is in a coherent state and the input idler wave is in a vacuum state (no input), where $\langle \Delta a_{s1}^2 \rangle = \langle \Delta a_{i1}^2 \rangle = \frac{1}{4}$.

A microwave nondegenerate parametric amplifier is dominated by thermal noise rather than quantum noise. In such a case, cooling the idler input port to below the noise equivalent temperature of the signal channel is effective to reduce the noise figure. Indeed, the noise figure of close to 0 dB is achieved in a microwave nondegenerate parametric amplifier by this technique.

Bibliography

- [1] A. von der Ziel, J. Appl. Phys. **19**, 999 (1948).
- [2] J. M. Manley and H. E. Rowe, Proc. IRE **47**, 2115 (1959).
- [3] M. Uenohara, Low Noise Amplification, Handbuch der Physik, vol. 23 (Springer, Berlin, 1962).
- [4] J. A. Giordmaine and R. C. Miller, Phys. Rev. Lett. **14**, 973 (1965).
- [5] S. E. Harris, M. K. Oshman and R. L. Byer, Phys. Rev. Lett. **18**, 732 (1967).
- [6] H. A. Haus and J. A. Mullen, Phys. Rev. **128**, 2407 (1962).
- [7] C. M. Caves, Phys. Rev. **D23**, 1693 (1981).

Microwave Parametric Frequency Dividers With Conversion Gain

Zainabolhoda Heshmati, *Student Member, IEEE*, Ian C. Hunter, *Fellow, IEEE*, and Roger D. Pollard, *Fellow, IEEE*

Abstract—A novel active parametric frequency divider configuration using coupled microstrip transmission lines and two balanced pseudomorphic HEMTs (pHEMTs) is presented. The analysis of the divide-by-2 circuit presented applies the principles of subharmonic generation using a nonlinear reactance to an active semiconductor device such as a pHEMT. A 2–1-GHz active analog frequency divider is designed and fabricated, with measurements showing a 20% bandwidth, 13.5-dB conversion gain, and harmonic rejection levels of more than 22 dBc. A maximum conversion gain of 18 dB is also achieved. These higher conversion efficiencies and the ability to cascade dividers allow for higher order division ratios to be achieved with the same topology.

Index Terms—Analog frequency dividers, coupled transmission lines, nonlinear circuits, parametric subharmonic generation.

I. INTRODUCTION

MODERN communication systems require sophisticated frequency generation circuits in which frequency synthesizers play an important role. Frequency division is a key function of such circuits.

Frequency dividers have applications in many systems, from communications to military applications. They are also essential to a variety of microwave systems. Classic examples include applications involving straight frequency counting, phase-locked loops (PLLs), and phase noise reduction [1]. Frequency dividers also have bandwidth compression capabilities, which opens up the possibility of transferring wide microwave bands, to the region where the power of digital logic can be used for processing [2].

There have been a number of different microwave frequency divider concepts described in the literature. Each of these individual concepts can be separated into two basic categories: digital and analog. Digital dividers are capable of broadband performances up into the microwave frequency range (up to 40 GHz), but their power consumption increases with frequency (several watts dc power at millimeter wavelengths). Analog dividers feature lower power consumption, simpler circuit design, and higher operating frequencies, which makes them attractive for communications purposes.

Manuscript received April 6, 2007; revised June 30, 2007. This work was supported by the British Ministry of Defence, Electro-Magnetic Remote Sensing Defence Technology Centre.

The authors are with the Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering, The University of Leeds, Leeds LS2 9JT, U.K. (e-mail: eenzh@leeds.ac.uk; i.c.hunter@leeds.ac.uk; r.d.pollard@leeds.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMTT.2007.906490

Among the various analog solutions [1], parametric frequency dividers [3], [4] offer simpler circuit configuration and broader bandwidth in comparison with other alternatives such as the regenerative [5]–[7] or injection-locked dividers [8]–[10]. Parametric division is a process in which a subharmonic oscillation is generated from the excitation of a nonlinear reactive element. The most common nonlinear reactive element used in parametric dividers is the varactor diode, which exploits the asymmetrical voltage-dependent depletion layer capacitance of an abrupt junction diode [11]. The basic theory of device operation is presented in [12] and [13]. Some divider designs based on empirical techniques are described in [4], [14], and [15].

Due to losses within the varactor diodes, amplifiers are generally required to recover the input signal level. Although original divider proposals had conceived the idea of a general division by N , subsequent divider designs have been mainly divide-by-2 configurations. To obtain greater division ratios, frequency dividers have been cascaded. However, this increases hardware requirements and results in the need for more amplification. The design presented here employs pseudomorphic high electron-mobility transistor (pHEMT) devices to perform parametric frequency division and amplification simultaneously at microwave frequencies.

A novel design has been developed and tested for an active parametric frequency divide-by-2 circuit with a large division bandwidth and high conversion gain. The presented circuit design consists of a balanced topology using microstrip coupled transmission lines, which is more amenable to integration and can be implemented in contemporary silicon technology.

II. PARAMETRIC FREQUENCY DIVISION

The term *parametric* has become associated with a class of amplifying and frequency-converting devices, which utilize the properties of nonlinear or time-varying reactances. These reactances channel energy from an ac source to a useful load and are capable of power conversion from one frequency to another.

Solid-state frequency dividers can be realized both in passive and active configurations, either employing passive devices with reactive nonlinearities (e.g., diodes under reversed bias) or using active devices (e.g., MESFET, HEMT, HBT, etc.) biased in a strongly nonlinear operating region. Unlike passive systems, active dividers exhibit conversion gain and, therefore, do not require amplifiers to recover the output signal. The availability of active devices exhibiting conversion gain at frequencies extending well into the millimeter-wave region with considerable bandwidth is pushing towards the active solution. Moreover, this

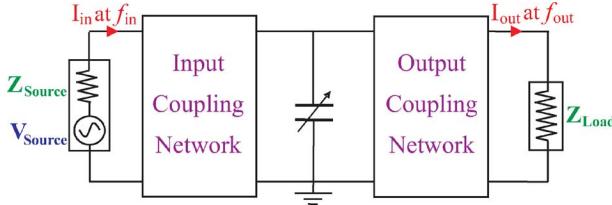


Fig. 1. Block diagram of a parametric frequency divider circuit.

choice has the obvious advantage of allowing functional integration in a single technology if monolithic implementation is attempted.

Existing documentation describe two parametric divider topologies, both implemented using varactors: filter-based dividers [15] and balanced dividers [16]. However, the two topologies share the same basic principles. The primary performance discriminator between the two types of dividers is bandwidth related.

The filter-based divider is limited in bandwidth because of the high- Q selective filters, but offers the best possibility of producing a minimum threshold design due to its narrow bandwidth and single varactor loss. On the other hand, the balanced design is more amenable to wideband performance and has better transient response. The balanced divider, which requires dual varactors, has been analyzed by Harrison [16]. He shows that the even/odd modality of the device lends itself to closed-form solutions, and has documented, in conjunction with Kalivas [4], [17], several practical dividers, but all are passive designs, which require high input powers and show large conversion loss.

The analysis presented here is in some ways similar to that of Harrison [4], but replaces the nonlinear varactor with a pHEMT. The input nonlinear junction capacitance of the pHEMT operates as a varactor divider producing subharmonics. The device transconductance g_m simultaneously provides amplification.

Higher order 2^n division can also be obtained by cascading several dividers. The associated gain of the active device compensates for the conversion loss and enables frequency dividers to be cascaded without much concern about signal degradation and power loss.

III. BASIC PARAMETRIC DIVIDER THEORY

The basic principles of operation of parametric frequency dividers are based on the excitation of a nonlinear reactive element (in the simple case: a varactor). Fig. 1 shows a block diagram of a simple parametric frequency divider using a varactor (non-linear voltage-dependent capacitor). Like other nonlinear reactances, varactors can generate power not only at harmonics, but at subharmonics of the pumping frequency. The circuit consists of two input and output coupling networks, which, for frequency division, must be properly adjusted. The coupling networks act like filters, which, in their simplest form, consist of resonant LC networks tuned at the required input and output frequency, as shown in Fig. 2. Theoretically, if the varactor is ideal and the filters are extremely selective (i.e., high Q), then it is possible to achieve very high efficiencies with this circuit.

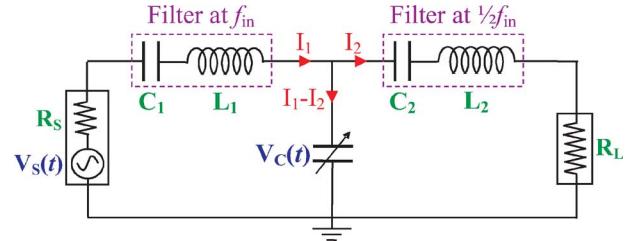


Fig. 2. Simple lumped element varactor frequency divide-by-2 circuit.

The divide-by-2 circuit consists of a shunt varactor nested between two resonant LC network filters. The input filter is tuned to the input pump frequency f_{in} , and allows the pump energy to reach the varactor. The filter also isolates the source from the subharmonic oscillations occurring at the diode. The output LC filter is tuned to the subharmonic frequency $(1/2)f_{in}$ and prevents the pump energy from reaching the load.

For the divider circuit in Fig. 2, we assume that some subharmonic current is already present in the noise for it to grow in magnitude, rather like an oscillator. This is the essential difference between a frequency multiplier and divider. In a frequency multiplier, there will also always be some harmonics present, regardless of the magnitude of the input, but frequency dividers have a power threshold level below which subharmonics cannot exist. This threshold level is the power required to overcome the losses in the output circuit for subharmonic oscillations to occur. This is the fundamental difference between a frequency divider and subharmonic oscillator.

The analysis of this circuit will provide an insight into the general operation of parametric frequency dividers. The circuit is analyzed in terms of the input and output loops as

$$\begin{aligned} V_s(t) &= V_1(t) + V_c(t) \\ &= i_1 R_S + L_1 \frac{di_1}{dt} + \frac{1}{C_1} \int i_1 dt + V_c(t) \end{aligned} \quad (1)$$

and

$$\begin{aligned} V_c(t) &= V_2(t) \\ &= i_2 R_L + L_2 \frac{di_2}{dt} + \frac{1}{C_2} \int i_2 dt \end{aligned} \quad (2)$$

where $V_1(t)$ and $V_2(t)$ are the voltages across the input and output RLC networks, respectively. The voltage across the varactor is

$$V_c(t) = \Phi - m (q(t) - q_0)^2 \quad (3)$$

where

$$m = \frac{1}{4C_{j0}^2\Phi}$$

since

$$C = \frac{dq}{dv} = \frac{C_{j0}}{(1 - v/\Phi)^{\frac{1}{2}}}. \quad (4)$$

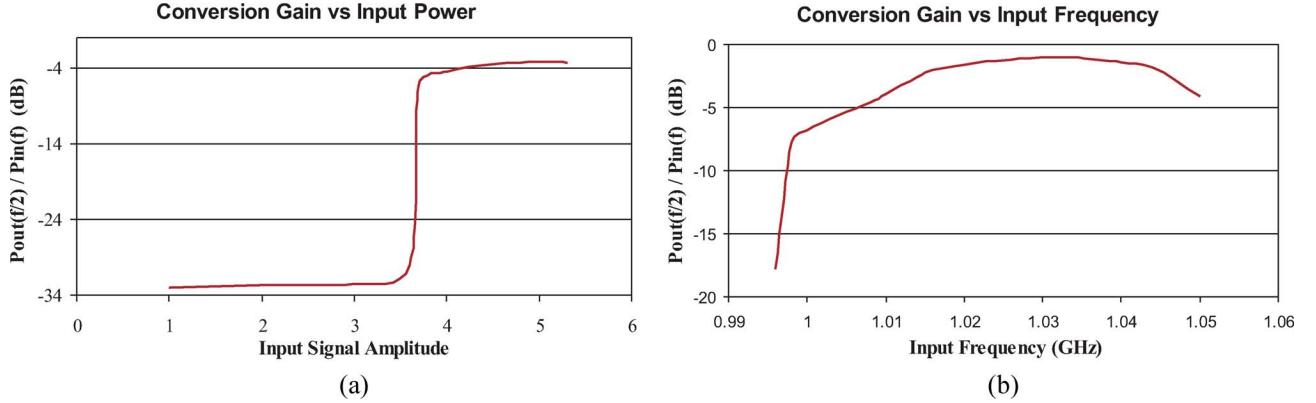


Fig. 3. Simulink simulation results showing: (a) input signal threshold level and (b) working frequency bandwidth of the divider circuit.

Also,

- C_{j0} zero-voltage junction capacitance;
- Φ junction in-built potential;
- q_Φ charge at $v = \Phi$;
- $q(t)$ instantaneous charge on the varactor;
- $v(t)$ instantaneous voltage across varactor.

The circuit equations can be represented in MATLAB Simulink using a mathematical block diagram. This is then simulated with different input parameters for a divide-by-2 circuit. These simulations provide the numerical solution to the nonlinear circuit equations of the basic parametric frequency divider and are used to qualitatively explain the behavior of the divider circuit.

The results clearly show frequency division at particular power levels for a 1–0.5-GHz divider circuit. Fig. 3(a) shows the existence of an input signal threshold level in the divider operation, and the operating frequency bandwidth is shown in Fig. 3(b).

These principles of a nonlinear reactance frequency divider are now applied to a parametric pHEMT frequency divider with a balanced topology. We can write a simplified model for the active device (in this case: pHEMT), which contains this type of reactive nonlinearity on its input, as well as the output current source that provides the gain. The active divider circuit can then be designed based on the above varactor parametric design principles, and the output current source will provide the simultaneous gain. This analogy has been adapted throughout the paper.

IV. BALANCED CIRCUIT ANALYSIS

The balanced circuit, as with other parametric divider topologies, requires two loops in the circuit, which are resonant at the fundamental input frequency f_{in} and the subharmonic output frequency $(1/2)f_{in}$. The circuit is constructed in such a way as to minimize coupling of the input frequency to the output port. These requirements can be satisfied by a microstrip resonant structure consisting of a pair of symmetrical coupled transmission lines in combination with two nonlinear reactive elements

Conversion Gain vs Input Frequency

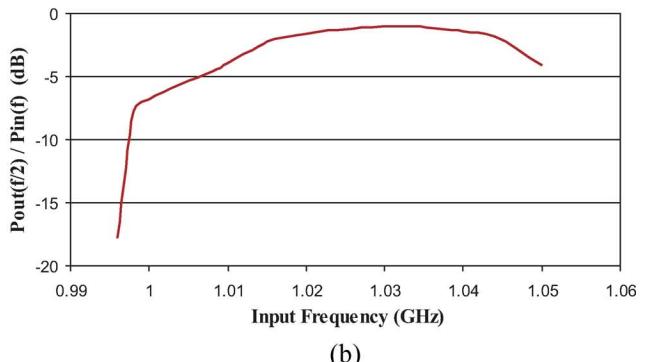


Fig. 4. Basic balanced coupled line frequency divider circuit.

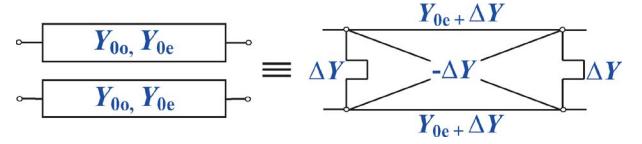


Fig. 5. Simplified equivalent circuit of a pair of coupled lines.

[i.e., the nonlinear input capacitance of the active device, represented in Fig. 4 as $C(v)$].

An input signal at f_{in} entering the microstrip lines is divided between the two lines so they have potentials equal in magnitude and sign. This excites the nonlinear capacitances in phase and the propagation is determined by the even-mode admittance Y_{0e} .

The odd-mode resonance of the coupled lines supports oscillations at $(1/2)f_{in}$. This resonance behavior is determined by the odd-mode admittance, Y_{0o} of the coupled lines. Due to the nonlinear coupling mechanism between the even and odd mode, energy is transferred from f_{in} to $(1/2)f_{in}$ via the nonlinear capacitances.

A simplified analysis of coupled transmission lines is given by Sato and Cristal [18] leading to the equivalent circuit shown in Fig. 5 for the coupled lines, where $\Delta Y = (Y_{0o} - Y_{0e})/2$.

This method applied to the divider circuit of Fig. 4, and leads to the equivalent circuit of Fig. 6. Each of the coupled lines are reactively loaded with the pHEMT with a capacitance of C_0 at the input bias voltage V_{gs} .

The following analysis is based on two separate modes of excitation for the circuit of Fig. 6, i.e., even and odd mode, for which the line of symmetry is shown. Node A is the common input point to the coupled lines. In the even mode, a finite voltage at frequency f_{in} is presented at A and an open circuit exists across the line of symmetry. Looking in from the

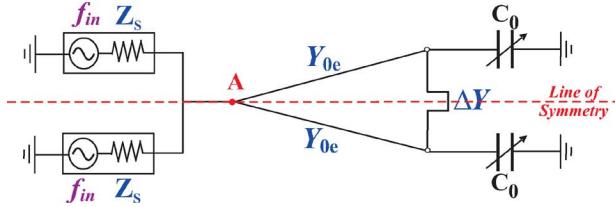


Fig. 6. Equivalent circuit of the balanced parametric divider.

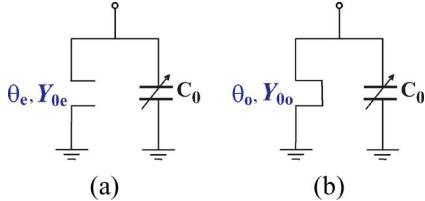


Fig. 7. (a) Even- and (b) odd-mode equivalent resonance circuits.

output, the equivalent circuit is the parallel combination of the nonlinear capacitance and the open-circuit stub of admittance Y_{0e} and electrical length θ_e at f_{in} , as shown in Fig. 7(a). Resonance condition, therefore, at the fundamental frequency $f_{\text{in}} = \omega_{\text{in}}/2\pi$, is

$$Z_{0e} \cot \theta_e \omega_{\text{in}} C_0 = -1 \quad (5)$$

and θ_e will be the even-mode electrical length of the coupled lines.

Similarly in the odd-mode case, zero voltage exists at node A with a short circuit across the line of symmetry. This reduces the equivalent circuit in the odd mode to that shown in Fig. 7(b). The condition for resonance at the subharmonic frequency $(1/2)f_{\text{in}}$ is

$$j \frac{\omega_{\text{in}}}{2} C_0 = \frac{j Y_{0o}}{\tan \theta_o}.$$

This gives

$$Z_{0o} \tan \theta_o \frac{\omega_{\text{in}}}{2} C_0 = 1 \quad (6)$$

as the odd-mode resonance condition with θ_o as the odd-mode electrical length of the coupled lines.

Equations (5) and (6) form the two resonance conditions and together will force only two frequency components of current to flow in the nonlinearities. From these two equations, we derive the ratio of the even- and odd-mode impedances as

$$\frac{Z_{0e}}{Z_{0o}} = \frac{-\tan \theta_e \tan \theta_o}{2}.$$

In order for the lines to be physically realizable [19],

$$\frac{Z_{0e}}{Z_{0o}} > 1. \quad (7)$$

Since the value of C_0 can be extracted from measurements of the active devices, suitable values of θ_e and θ_o are chosen to satisfy conditions (5)–(7). Given the frequency of operation and substrate parameters, the physical dimensions and spacing of the coupled lines can be derived from Z_{0e} and Z_{0o} . For maximum bandwidth, the resonator size should be minimized.

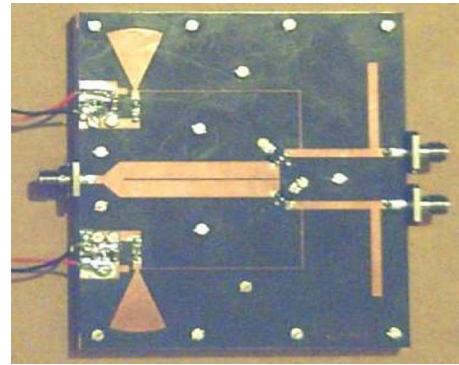


Fig. 8. Balanced active divide-by-2 circuit.

V. PRACTICAL EXAMPLE

A 2–1-GHz frequency divider is designed in microstrip, on a Rogers RT Duroid 5880. The active device used in this study is the Filtronic FPD1500SOT89 pHEMT, which has a gate length of $0.25 \mu\text{m}$. The small-signal input capacitance of the pHEMT, measured versus the input bias voltage V_{gs} , showed a capacitance value C_0 , of approximately 3 pF at the required bias voltage.

With substrate parameters of $\epsilon_r = 2.2$ and thickness of 0.787 mm, θ_e and θ_o are chosen to be 130° and 65° , respectively. This gives $Z_{0e} = 31.6 \Omega$ and $Z_{0o} = 24.7 \Omega$. The coupled line dimensions are then calculated to have widths of 5.1 mm, and lengths of 38.8 mm with a spacing of 0.38 mm between the two lines.

Fig. 8 shows the fabricated pHEMT frequency divider. The input gate bias of the pHEMTs (V_{gs}) is applied via a bias-tee to the input of the circuit. The drain bias (V_{ds}) is applied on the output of the pHEMTs via two high-impedance lines, each $\lambda/4$ at the output frequency of 1 GHz. The radial stubs produce broadband shorts at the junction of each high impedance line.

The nonlinear model TOM3 [20] is used to simulate the biased active devices in Agilent's ADS software to analyze the output impedances. The pHEMT is biased at a gate voltage of $V_{\text{gs}} = -0.6 \text{ V}$ and a drain bias of $V_{\text{ds}} = +3 \text{ V}$, and showed an output impedance of $17.613 + j8.44 \Omega$, which is conjugately matched to the 50Ω output to maximize the power delivered to the load at the output frequency of 1 GHz. The matching network consists of a length of transmission line and an open-circuit stub, as shown in Fig. 8. Two capacitors are also used as dc blocks on the output.

The input port can also be matched to 50Ω in order to reduce any reflection from the circuit. This will also reduce the threshold power, as more power will be delivered to the circuit with optimum match.

VI. MEASUREMENTS

A minimum level of input power (P_{in}) is required in order for frequency division to occur. For the 2–1-GHz active parametric divider, the threshold input level is -14 dBm , at which power frequency division commences abruptly. This is clearly shown in Fig. 9, where the output power is plotted against the input power P_{in} . As P_{in} increases beyond this level, the bandwidth increases.

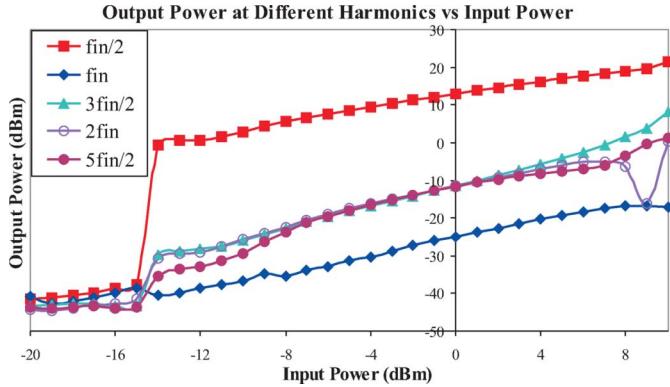


Fig. 9. Results for 2–1-GHz frequency divider showing the output power versus the input power.

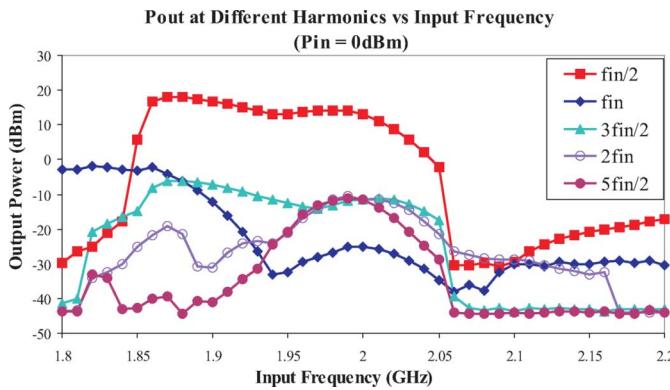


Fig. 10. Results of the divider showing the output power versus the input frequency.

The output power is optimized by bias adjustments to $V_{gs} = -1$ V and $V_{ds} = +3.15$ V. The output frequency response and working bandwidth of the divider, tested at $P_{in} = 0$ dBm, are shown in Fig. 10. A wide bandwidth of approximately 20% is achieved with harmonic rejection levels of more than 22 dBc. The average gain produced by this active divider is 13.5 dB at $f_{in} = 2$ GHz, with a maximum gain of 18 dB occurring at $f_{in} = 1.88$ GHz for $P_{in} = 0$ dBm.

To the best of the authors' knowledge, the results show a higher conversion gain compared to other types of analog frequency dividers. The reduced bandwidth is a tradeoff for conversion gain. Also, since we have considered the resonance condition of the fundamental frequency in the design of the coupled lines structure, the bandwidth is reduced compared to the varactor frequency divider of Harrison [4]. This has the added benefit of improved spurious performance, as well as reduced signal-to-noise ratio performance.

VII. CONCLUSIONS

A practical and efficient approach for the design of an active microwave frequency divider has been demonstrated. The parametric frequency division method has been applied to a balanced

divider configuration using nonlinear semiconductor devices in order to achieve a parametric frequency divider at microwave frequencies, which exhibits conversion gain. A 2–1-GHz frequency divider was designed and fabricated with demonstrated working bandwidth of 200 MHz and a maximum conversion gain of 18 dB. Harmonic rejection levels of more than 22 dBc were achieved.

ACKNOWLEDGMENT

The authors thank the Electro-Magnetic Remote Sensing (EMRS) Defence Technology Centre (DTC), for supporting this study.

REFERENCES

- [1] I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*, 2nd ed. Hoboken, NJ: Wiley, 2003.
- [2] W. D. Cornish, "Microwave frequency dividers, devices and applications," in *Proc. Military Microw. Conf., Microw. Exhibitions, Publishers Ltd.*, Oct. 1980, pp. 13–18.
- [3] J. W. McConnel, "A parametric frequency divider using hyperabrupt junction diode capacitance," *IEEE J. Solid-State Circuits*, vol. SSC-3, no. 3, pp. 311–312, Sep. 1968.
- [4] R. G. Harrison, "A broadband frequency divider using varactors," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-25, no. 12, pp. 1055–1059, Dec. 1977.
- [5] R. L. Miller, "Fractional-frequency generators utilizing regenerative modulation," *Proc. IRE*, vol. 37, no. 7, pp. 446–457, Jul. 1939.
- [6] C. Rauscher, "Regenerative frequency division with a GaAs FET," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-32, no. 11, pp. 1461–1468, Nov. 1984.
- [7] E. S. Ferre-Pikal and F. L. Walls, "Microwave regenerative frequency dividers with low phase noise," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 46, no. 1, pp. 216–219, Jan. 1999.
- [8] H. R. Rategh and T. H. Lee, "Superharmonic injection-locked frequency dividers," *IEEE J. Solid-State Circuits*, vol. 34, no. 6, pp. 813–821, Jun. 1999.
- [9] X. Zhang, X. Zhou, B. Aliener, and A. S. Daryoush, "A study of subharmonic injection locking for local oscillators," *IEEE Microw. Guided Wave Lett.*, vol. 2, no. 3, pp. 97–99, Mar. 1992.
- [10] R. J. Betancourt-Zamora, S. Verma, and T. H. Lee, "1-GHz and 2.8-GHz CMOS injection-locked ring oscillator prescalers," in *VLSI Circuits Tech. Symp. Dig.*, Jun. 2001, pp. 47–50.
- [11] H. A. Watson, *Microwave Semiconductor Devices and Their Circuit Applications*. New York: McGraw-Hill, 1969.
- [12] J. Hilibrand and R. Beam, "Semiconductor diodes in parametric subharmonic oscillators," *RCA Rev.*, pp. 229–253, Jun. 1959.
- [13] P. Penfield and R. P. Rafuse, *Varactor Applications*. Cambridge, MA: MIT Press, 1962.
- [14] Z. Nativ, "The application of a frequency multiplier design method to the design of microwave parametric dividers," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-35, no. 2, pp. 189–194, Feb. 1987.
- [15] G. R. Sloan, "The modeling, analysis, and design of filter-based parametric frequency dividers," *IEEE Trans. Microw. Theory Tech.*, vol. 41, no. 2, pp. 224–228, Feb. 1993.
- [16] R. G. Harrison, "Theory of the varactor frequency divider," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Boston, MA, Jun. 1983, pp. 203–205.
- [17] G. A. Kalivas and R. G. Harrison, "A new slotline-microstrip frequency halver," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 1985, pp. 683–686.
- [18] R. Sato and E. G. Cristal, "Simplified analysis of coupled transmission-line networks," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-18, no. 3, pp. 122–131, Mar. 1970.
- [19] D. M. Pozar, *Microwave Engineering*, 2nd ed. New York: Wiley, 1998.
- [20] R. B. Hallgren and P. H. Litzenberg, "TOM3 capacitance model: Linking large- and small-signal MESFET models in SPICE," *IEEE Trans. Microw. Theory Tech.*, vol. 47, no. 5, p. 556, May 1999.



Zainabolhoda Heshmati (S'07) received the B.Eng. Hons. degree (first class) in electronics and telecommunications engineering from the University of Bradford, Bradford, U.K., in 2003, and is currently working toward the Ph.D. degree in electronic engineering at The University of Leeds, Leeds, U.K.

Her research interests include nonlinear microwave circuits, subharmonic generation, microwave frequency dividers, and parametric circuits.

Ms. Heshmati was the recipient of the Institution of Electrical Engineers (IEE), U.K., Best Student Prize (2003). She was also the recipient of Electro-Magnetic Remote Sensing (EMRS) Defense Technology Centre (DTC) sponsorship for postgraduate research.



Ian C. Hunter (M'92–SM'94–F'07) received the B.Sc. (first-class honors) and Ph.D. degrees from The University of Leeds, Leeds, U.K., in 1978 and 1981, both in electrical engineering.

He has worked in industry for Aercom Industries Inc., Sunnyvale, CA, and KW Engineering, Oakland, CA, and for Filtronic Components Ltd., Leeds, U.K., during which time he developed broadband microwave filters for electronic warfare (EW) applications. From 1995 to 2001, he was with Filtronic Comtek, where he was involved with advanced filters for cellular radio. He is currently a Professor with the School of Electronic and Electrical Engineering, The University of Leeds, where he teaches circuit

theory, electromagnetism, and microwave engineering. He authored *Theory and Design of Microwave Filters* (IEE Press, 2001). His research includes linear and nonlinear microwave filters. He also investigates applications of microwaves in biology.



Roger D. Pollard (M'77–SM'91–F'97) holds the Agilent Technologies Chair in High Frequency Measurements, is currently Dean of the Faculty of Engineering, and was Head of the School of Electronic and Electrical Engineering with The University of Leeds, Leeds, U.K., where he has been a faculty member since 1974. Since 1981, he has been a consultant to Agilent Technologies (previously the Hewlett-Packard Company), Santa Rosa, CA. He has authored or coauthored over 100 technical papers. He holds three patents.

Prof. Pollard is a Chartered Engineer and Fellow of the Institution of Electrical Engineers (IEE) (U.K.). He was elected to the Royal Academy of Engineering in 2005. He is an active IEEE volunteer, serving a term as president of the IEEE Microwave Theory and Techniques Society (IEEE MTT-S) in 1998. He has chaired numerous IEEE committees, including Strategic Planning for the Publications, Services and Products Board, Technical Activities Board Products Committee, and the Electronic Products Committee, responsible for the development and introduction of IEEE-Xplore(r). He was the Technical Program Committee (TPC) co-chair for the 2006 IEEE MTT-S International Microwave Symposium (IMS).

RF CMOS Parametric Downconverters

Sebastian Magierowski, *Member, IEEE*, Jean-Francois Bousquet, *Student Member, IEEE*,
Zhixing Zhao, *Associate Member, IEEE*, and Takis Zourntos, *Senior Member, IEEE*

Abstract—Parametric amplifiers are absent today from the majority of electronics applications. This is especially the case for parametric downconverters (PDCs). Coupled with the increasing emphasis on millimeter-wave applications and the cost of transistor scaling, the time may be right to reconsider these circuits. By employing coupled-mode theory, we arrive at a general description of PDCs. Consequently, a simple mixer is proposed that achieves gain at reduced pumping frequencies without resorting to sub-harmonics. The implications of this design for quadrature receiver systems are shown. Fundamental gain and noise limits are derived indicating the ability to operate at sub-5-dB noise figures (NFs) with very low-power requirements. Measurements on an accumulation-mode varactor in 130-nm CMOS technology indicate the necessary pumping and biasing regimes needed to approach these limits. Finally, a compact 30-GHz PDC design with 2-dB NF is discussed.

Index Terms—MOS varactor, parametric amplifier, parametric mixer, RF integrated circuit (RFIC).

I. INTRODUCTION

DUCTANCE-BASED amplifiers and frequency converters—parametric circuits—generally outperform their technological contemporaries. Historically, they have been known to operate at higher frequencies, introduce less noise, and consume less power than concurrent conductance-based devices operating at room-temperature and realized in either the vacuum or the solid-state. Despite these advantages, parametric circuits are scarcely present in today’s design landscape, an absence that is near total in the silicon integrated circuit (IC) sphere.

Yet, modern CMOS IC technology provides an excellent platform for the application of parametric circuits to emerging millimeter and sub-millimeter wave systems. In fact, a symbiotic relationship can easily be envisioned whereby parametric approaches profit from the simplified pumping, filtering, and digital calibration opportunities afforded by CMOS technology, while the latter benefits from a circuit technique capable of extending its operational range beyond the capabilities of its transistor devices.

Manuscript received September 10, 2009; revised November 13, 2009. First published February 17, 2010; current version published March 12, 2010. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

S. Magierowski, J.-F. Bousquet, and Z. Zhao are with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB, Canada T2N 1N4 (e-mail: smagiero@ucalgary.ca; jfbousqu@ucalgary.ca; zhao@ucalgary.ca).

T. Zourntos is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128 USA (e-mail: takis@neo.tamu.edu).

Digital Object Identifier 10.1109/TMTT.2010.2040332

This paper attempts to strengthen this case by considering the performance of parametric downconverters (PDCs) in a 130-nm CMOS technology for heterodyne receiver applications. First, a technique is proposed that allows the PDC to maintain gain at reduced pumping frequencies without resorting to sub-harmonic methods. This structure is discussed in [1], although the present work employs coupled-mode theory to cast it in more general terms and extends it to a dual-pumped structure. This phase of the work is a contribution to the parametric circuit literature in general where reports on downconversion mixers are relatively rare. However, this development is also important when considering the PDC for CMOS implementation as it simplifies the design of the MOS-based local oscillator (LO) needed to drive the mixer. Second, this work considers the application of modern technology to PDCs in the form of accumulation-mode MOS varactors (AMOSVs). Since interest in parametric techniques withered long before the AMOSVs emergence into the CMOS mainstream, PDCs remain to fully exploit its merits. Third, a specific PDC topology is discussed as an illustration of the manner in which parametric circuits can be utilized in a CMOS technology.

This paper is organized as follows. Section II briefly discusses early developments of PDCs and outlines recent advances in parametric circuits. A discussion of PDC operational principles follows in Section III with a focus on the simplest topologies, the lower sideband downconverter (LSBDC) and the upper sideband downconverter (USBDC). Herein, a simple architecture, the double-sideband downconverter (DSBDC), is introduced as a useful means of lowering the LO frequency without sacrificing gain. Section IV details the gain and noise operation of the LSBDC and DSBDC, highlighting their performance potential. Special attention is paid to the DSBDC, which exhibits interesting behavior when considered for use in a quadrature receiver. The ability of AMOSVs in a 130-nm CMOS technology to satisfy the gain and noise needs of PDCs is the topic of Section V, which considers optimal bias and pump conditions. Finally, a basic means of implementing a PDC in an IC setting is considered in Section VI.

II. PARAMETRIC DOWNCONVERSION: BACKGROUND

Parametric circuits can address a great number of fundamental RF circuit operations including amplification, mixing, (sub) harmonic generation, and oscillation [2]. Several reasons prompt us to consider only the downconversion function in this paper. For one, the ongoing reliance on heterodyne receivers for supra-10-GHz wireless maintains the popularity of this architecture. Further, a sufficiently sensitive mixer can significantly ease the performance requirements of the low-noise amplifier that typically precedes it. This is typically out of the question for conductance-based mixers, but the low-noise properties of

reactance-based downconverters allow the system designer to consider this option.

The observation of reactance-based downconversion was first publicized by North in the mid-1940s [3] for a 10-GHz–30-MHz mixer composed of welded-contact germanium crystals. The operation of these devices was explained in terms of a variable capacitance by Torrey in [4]. Despite being the first diode-based mixers to operate with a conversion gain, these early results exhibited poor noise performance, especially when operated under gain conditions, and seem to have maligned research into the PDC ever since [5], [6].

The success of North's diode as a rectifier may be partly to blame for its failure as a parametric mixer. The innovative manufacturing technique applied to the welded-contact crystal resulted in a structure that tracked the diode equation over six decades of current, three orders of magnitude greater than other devices at the time [7]. Thus, under the influence of the LO, the variation of the diffusion capacitance, ideally an exponential function of applied voltage, was substantial. By choosing to bias his device in the forward region of operation, North modulated this capacitance and experienced parametric action. Unfortunately, this choice of bias also increases the shot noise inside the diode. It is likely that the parametric action converted these fluctuations to the IF and was an important culprit behind the poor noise behavior recorded by the early researchers.

A decade after Torrey's discussion on parametric conversion, high-quality diffused silicon junctions had emerged [6], the theory of reactive amplifiers had expanded [8], and engineers began to emphasize the use of reverse-biased diodes in parametric circuits [9]. As a result, parametric upconverters and the reflection amplifiers began achieving record-breaking noise performance, spurring significant research into these circuits. However, work on PDCs seems not to have reignited. With the emergence of the GaAs MESFET [10], interest in electronic parametric design in general began to wane.

Today, researchers have considered the advantages of parametric action for submillimeter-wave upconversion [11] and frequency division/multiplication [12]. The parametric amplifier has also been reconsidered in an IC CMOS context [13]. This work exemplifies a number of advantages that can be accrued by combining the parametric approach with IC technology. It exploits a three-terminal inversion-mode varactor to realize a discrete-time amplifier, which frees the circuit from the burden of filtering. Compared to completely transistor-based approaches, this technique achieves very low-noise and high-linearity amplification for minuscule power requirements.

III. PARAMETRIC DOWNCONVERSION: OPERATION

Any nonlinear resistive or reactive circuit element can act as a mixer. When downconverting an input at frequency ω_{RF} to a lower frequency ω_{IF} , mixers can employ either an upper sideband (USB) approach in which the LO frequency $\omega_{LO} = \omega_{RF} - \omega_{IF}$, or a lower sideband (LSB) approach, where $\omega_{LO} = \omega_{RF} + \omega_{IF}$. Resistive mixers are relatively insensitive to this choice, but PDCs are profoundly affected by it.

Sections III-A–D examine different PDC arrangements that differ only by the choice of ω_{LO} relative to ω_{RF} . Fig. 1 illus-

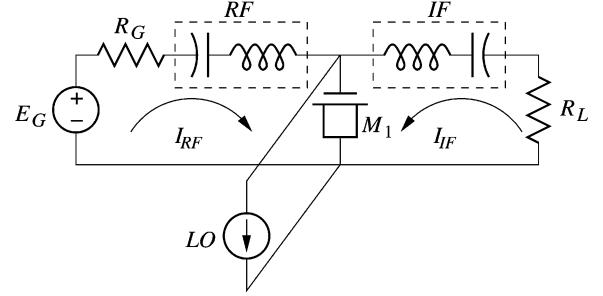


Fig. 1. Basic PDC structure, a varactor pumped by a large signal LO coupling two filtered loads.

trates a canonical PDC setup applicable to all three topologies where an LO current-source supplies power to and modulates the varactor M_1 . The bandpass filtered RF and IF loops to either side of M_1 limit the signal frequencies to which the varactor is exposed. Although all mixers embedded in radio systems require filtering components, for PDCs this goes beyond signal management and bears heavily on the operation of the circuit itself.

A. USBDC

The parametric USBDC follows a standard mixing approach by employing an LO with $\omega_{LO} < \omega_{RF}$. The immediate drawback to this configuration is the apparent absence of conversion gain. Specifically, according to the Manley–Rowe relations [8], we can summarize the power flow in this circuit with

$$\frac{P_{RF}}{\omega_{RF}} + \frac{P_{IF}}{\omega_{IF}} = 0 \quad (1)$$

$$\frac{P_{LO}}{\omega_{LO}} - \frac{P_{IF}}{\omega_{IF}} = 0 \quad (2)$$

where P_{RF} is the power delivered to M_1 by the RF signal, $-P_{IF}$ is the power flowing out of M_1 to the IF port, and P_{LO} is the power delivered to M_1 by the LO. Rewriting (1), we see that

$$-P_{IF} = \frac{\omega_{IF}}{\omega_{RF}} P_{RF}. \quad (3)$$

Apparently, for this arrangement, the lower the IF signal frequency, the greater the signal attenuation. The remainder of the RF signal power is delivered to the LO.

In concert with the Manley–Rowe analysis, we can also examine parametric circuits from the perspective of coupled-mode analysis [14]–[16], an approach that better details the energy exchange between resonators in time-varying circuits and provides more direct insight on the conditions needed to achieve gain in PDCs.

Even a somewhat artificial coupled-mode formulation in the context of the circuit in Fig. 2 is useful in establishing insight on PDC performance. In this circuit example, the varactor is shown coupling two resonators, one with a resonant frequency of ω_{RF} , the other with a resonant frequency of ω_{IF} .

Before accounting for the time-varying varactor capacitance, the coupled-mode analysis seeks to obtain a description of the energy dynamics of isolated resonators. This is obtained by first

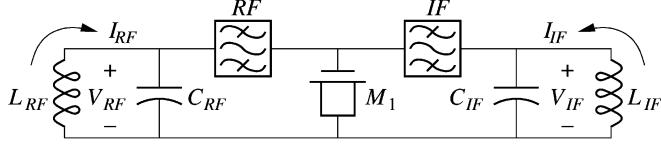


Fig. 2. Ideal scenario outlining the means of mode coupling as mitigated by a MOS varactor.

noting the relation between current and voltage variables of an isolated resonator (the RF tank is used as an example)

$$\frac{dI_{RF}}{dt} = -\frac{V_{RF}}{L_{RF}} \quad (4)$$

$$\frac{dV_{RF}}{dt} = \frac{I_{RF}}{C_{RF} + C_0} \quad (5)$$

where C_0 is the average capacitance of the varactor M_1 . These relations can be transformed into a pair of decoupled expressions

$$\frac{da_{RF}}{dt} = j\omega_{RF}a_{RF} \quad (6)$$

$$\frac{da_{RF}^*}{dt} = j\omega_{RF}a_{RF}^* \quad (7)$$

obtained by multiplying (5) by $\pm j\omega_{RF}(C_{RF} + C_0)$, adding it to (4) and multiplying the sum by the normalization constant $\sqrt{L_{RF}/2}$.

The new variables a_{RF} and its complex-conjugate a_{RF}^* are referred to as the *normal modes* of the RF resonator. These variables are proportional to the square root of the resonator energy and can be viewed as a pair of independent counter-rotating vectors expressed as

$$a_{RF}(t) = \frac{\sqrt{L_{RF}}}{2} [I_{RF}(t) + j\omega_{RF}(C_{RF} + C_0)V_{RF}(t)]. \quad (8)$$

The normal modes a_{IF} and a_{IF}^* can be derived for the IF tank in an identical manner. In terms of the normal modes, the total energy of the RF tank is

$$E_{RF} = |a_{RF}(t)|^2 + |a_{RF}^*(t)|^2.$$

Most importantly, the normal modes allow us to describe the energy dynamics of an isolated resonator with a pair of decoupled equations. As we will see shortly, this transformation makes it much easier to gauge the effect of a time-varying coupling between resonators.

Having established relations for the dynamics of energy in isolated resonators, we can now consider the effect of coupling between them. The normal modes of the USBDC and the manner in which they are coupled are shown in Fig. 3(a). As indicated, in the USBDC, ω_{LO} is such that a coupling between co-rotating (i.e., a_{RF} to a_{IF} and a_{RF}^* to a_{IF}^*) modes only is established. In short, we can rewrite (6) as

$$\frac{da_{RF}}{dt} = j\omega_{RF}a_{RF} + c_{uRF}a_{IF}e^{j\omega_{LO}t} \quad (9)$$

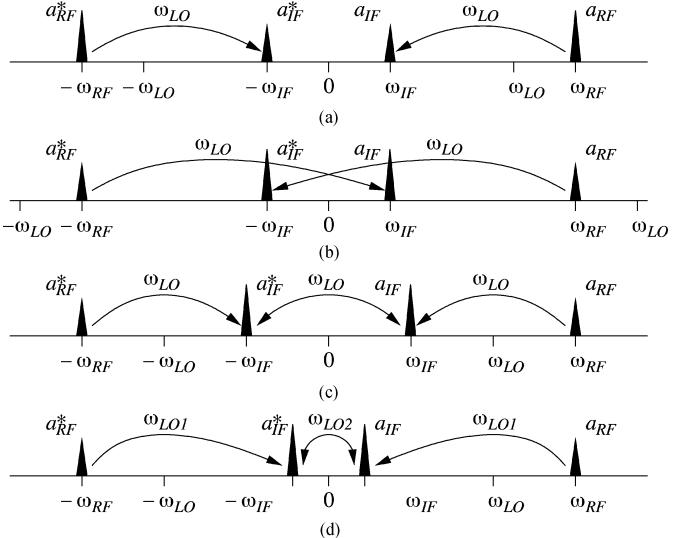


Fig. 3. Coupling between modes of the: (a) USB PDC, (b) LSB PDC, (c) double-sideband PDC, and (d) dual-pump PDC.

where c_{uRF} represents the coupling established between RF and IF modes. Equation (9) can be rigorously derived from circuit analysis, but is also recognizable as a general description of coupling between wave phenomena. Similarly, the expression for the IF normal mode is

$$\frac{da_{IF}}{dt} = j\omega_{IF}a_{IF} + c_{uIF}a_{RF}e^{-j\omega_{LO}t}. \quad (10)$$

The coupling coefficients c_{uRF} and c_{uIF} are defined by the details of the system; however, as (16) will show, an important relationship is established between these variables depending on the normal modes that are being coupled. Continuing, we can transform (9) and (10) into [14]

$$\frac{dA_{RF}}{dt} = c_{uRF}A_{IF} \quad (11)$$

$$\frac{dA_{IF}}{dt} = c_{uIF}A_{RF} \quad (12)$$

where A_{RF} and A_{IF} are slowly varying components of the respective modes. Specifically,

$$a_{RF}(t) = A_{RF}(t)e^{j\omega_{RF}t} \quad (13)$$

$$a_{IF}(t) = A_{IF}(t)e^{j\omega_{IF}t}. \quad (14)$$

Since the normal modes are energy variables they can be used to restate (3) of the Manley–Rowe relations as

$$-\frac{d}{dt} \frac{|a_{IF}|^2}{\omega_{IF}} = \frac{d}{dt} \frac{|a_{RF}|^2}{\omega_{RF}}. \quad (15)$$

Applying this to (9) and (10), we can show that

$$\frac{c_{uRF}}{\omega_{RF}} = -\frac{c_{uIF}^*}{\omega_{IF}} \quad (16)$$

which implies that the USBDC's normal modes are without exponentially increasing solutions. We can see this directly by sub-

stituting (16) into (11) and combining the result with (12) to produce

$$\frac{d^2 A_{IF}}{dt^2} = \lambda A_{IF} \quad (17)$$

where

$$\lambda = -\frac{\omega_{RF}}{\omega_{IF}} |c_{uIF}|^2. \quad (18)$$

This shows that the method of coupling between RF and IF in the USBDC is not capable of increasing the energy in the IF resonator, and hence, is referred to as a passive coupling scheme [14]. In other words, by modulating the varactor capacitance, the pump facilitates frequency conversion, but it does so in a manner that prevents the transfer of LO power to the IF normal modes.

B. LSBDC

Increasing the LO frequency such that $\omega_{RF} < \omega_{LO}$ results in the possibility of achieving a net gain between the RF and IF ports. Specifically, the Manley–Rowe relations state that the amount of IF power flowing into the converter is proportional to the RF input

$$P_{IF} = \frac{\omega_{IF}}{\omega_{RF}} P_{RF} \quad (19)$$

while the amount of IF power flowing out is

$$-P_{IF} = \frac{\omega_{IF}}{\omega_{LO}} P_{LO}. \quad (20)$$

This result shows that, unlike the USBDC, the LSBDC allows the LO to transfer a portion of its power P_{LO} to the IF port. Since the LO power is much greater than the input power, the opportunity for gain is present in this circuit. However, the relation between RF input and IF output power is not explicit in these results, an indication of the potential instability of this configuration [8]. Further, the gain that can be achieved with the LSBDC comes at the cost of a higher pump frequency, a problem for cheaper CMOS oscillators.

As implied by (20), the LSBDC is an example of active coupling [14], where the pump not only facilitates frequency conversion, but, in the process, also transfers some of its power to the signal. Referring this time to Fig. 3(b), we see that the increased ω_{LO} setting of the LSBDC establishes a link between counter-rotating vectors a_{RF} and a_{IF}^* such that, in analogy to (9) and (10),

$$\frac{da_{RF}}{dt} = j\omega_{RF}a_{RF} + c_{lRF}a_{IF}^*e^{j\omega_{LO}t} \quad (21)$$

$$\frac{da_{IF}^*}{dt} = -j\omega_{IF}a_{IF}^* + c_{lIF}a_{RF}e^{-j\omega_{LO}t}. \quad (22)$$

In this case, we can show this time with the aid of (19) that, instead of (16), the LSBDC's coupling coefficients relate as

$$\frac{c_{lRF}}{\omega_{RF}} = \frac{c_{lIF}^*}{\omega_{IF}}. \quad (23)$$

This relationship introduces exponentially increasing terms into the solutions for the modes of the LSBDC. This result reinforces the relationship stated in (20), but it also distills a general design

insight: *the coupling of counter-rotating modes leads to conversion gain*. We now exploit this property to suggest a simple PDC, which achieves gain at a reduced LO frequency.

C. DSBDC

An examination of Fig. 3(a) and (b) indicates that a direct link between counter-rotating modes (i.e., between a_{RF} and a_{IF}^*) is out of the question when $\omega_{LO} < \omega_{RF}$, but, as illustrated in Fig. 3(c), these diagrams suggest that a link consisting of a two-step conversion, but only one pump frequency, is possible if we choose $\omega_{LO} = 2\omega_{RF}/3$. This coupling can be formally stated as

$$\frac{da_{RF}}{dt} = j\omega_{RF}a_{RF} + c_{dRF}a_{IF}e^{j\omega_{LO}t} \quad (24)$$

$$\frac{da_{IF}}{t} = j\omega_{IF}a_{IF} + c_{dIF}a_{RF}e^{-j\omega_{LO}t} + k_{dIF}a_{IF}^*e^{j\omega_{LO}t} \quad (25)$$

$$\frac{da_{IF}^*}{dt} = -j\omega_{IF}a_{IF}^* + c_{dIF}a_{RF}^*e^{j\omega_{LO}t} + k_{dIF}^*a_{IF}e^{-j\omega_{LO}t} \quad (26)$$

$$\frac{da_{RF}^*}{dt} = -j\omega_{RF}a_{RF}^* + c_{dRF}a_{IF}^*e^{-j\omega_{LO}t} \quad (27)$$

where k_{dIF} is the coupling component between the IF counter-rotating modes.

In this system, ω_{LO} initiates a downconverting link from a_{RF} to a_{IF} , which forms a passive coupling between co-rotating modes characteristic of the USBDC. However, as indicated in Fig. 3(c), our choice of ω_{LO} simultaneously establishes a link between counter-rotating modes a_{IF} and a_{IF}^* , thus, according to the insights developed earlier, an active coupling is present in our PDC that makes conversion gain possible. The Manley–Rowe relations cannot be used to comment directly on this architecture because they hold only for incommensurate frequency combinations.

As a shorthand in alignment with the previous two acronyms, we refer to this unit as the DSBDC for the fact that the signal operated on by the LO appears both below and above ω_{LO} . The apparent benefits of this approach include the ability to achieve gain for lower ω_{LO} without the need for any structural changes to the basic PDC topologies already discussed. In this sense, the DSBDC configuration represents a compact solution. Further, the DSBDC achieves its operation without resorting to subharmonic excitation, which is normally associated with a sacrifice in noise performance. The ramifications of this choice of pumping, in addition to the standard LSBDC approach, are explored in more detail from a circuit theory perspective in Section IV.

D. DPDC

The DSBDC is not the only means by which ω_{LO} may be lowered and gain maintained without the need to resort to subharmonic pumping or additional filters (i.e., idlers). Indeed, the two-tiered downconversion and gain procedure of the DSBDC can be generalized to arbitrary ω_{IF} if two LO sources at ω_{LO1} and ω_{LO2} are employed as illustrated in Fig. 3(d). For such a dual-pump downconverter (DPDC), we assume that $\omega_{LO2} < \omega_{LO1} < \omega_{RF}$. In order to couple counter-rotating modes a_{IF}

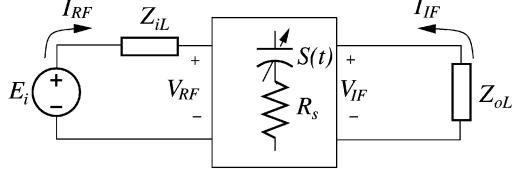


Fig. 4. LSBDC with associated port variables.

and a_{IF}^* , we note from Fig. 3(d) that $\omega_{LO2}/2 + \omega_{LO1}$ must equal to ω_{RF} . Therefore, we have the general relationship

$$\omega_{LO2} = 2(\omega_{RF} - \omega_{LO1}) \quad (28)$$

between the DPDC's pump frequencies. Compared to the DSBDC, the addition of a second pump complicates the down-converter design, although, from the heterodyne radio systems perspective, the availability of a second LO is necessary anyway to complete the conversion to baseband.

IV. GAIN AND NOISE BEHAVIOR

To estimate the gain and noise behavior of the circuits discussed above, we turn to a conversion matrix analysis [17] of the LSBDC and the DSBDC. The results presented here will allow AMOSV measurements discussed in Section V to indicate PDC performance.

The two-port system under consideration is shown in Fig. 4, where E_i is the RF driving signal source. A varactor is always the centerpiece of solid-state parametric circuits and an equivalent series-RC circuit is used to model it. In this approach, the port parameters, V_{RF} , V_{IF} , I_{RF} , and I_{IF} are treated as small signals with only the pumping LO allowed to control the variable capacitance $C(t)$. In this analysis, we characterize the varactor's reactive component in terms of elastance $S(t) = 1/C(t)$ such that we have the Fourier series expression

$$S(t) = \sum_{n=-\infty}^{\infty} S_n e^{jn\omega_{LO}t}. \quad (29)$$

Filters are assumed to have established the requisite ω_{RF} and ω_{IF} frequencies at the input and output ports, respectively, while remaining circuit losses and reactive contributors can be embedded in the input and output impedances $Z_{iL} = R_{iL} + jX_{iL}$ and $Z_{oL} = R_{oL} + jX_{oL}$.

A. LSBDC

The conversion matrix relation of the LSBDC is

$$\begin{bmatrix} V_{RF} \\ V_{IF}^* \end{bmatrix} = \begin{bmatrix} R_s + S_0/j\omega_{RF} & -S_1/j\omega_{IF} \\ S_1^*/j\omega_{RF} & R_s - S_0/j\omega_{IF} \end{bmatrix} \begin{bmatrix} I_{RF} \\ I_{IF}^* \end{bmatrix} \quad (30)$$

where R_s is the series equivalent resistance of the pumped varactor, while S_0 and S_1 are the varactor's average and first harmonic elastance, respectively. Using (30) in combination with the loop expressions (see Fig. 4)

$$V_{RF} = E_{RF} - Z_{iL}I_{RF} \quad (31)$$

$$V_{IF}^* = -Z_{oL}^*I_{IF}^* \quad (32)$$

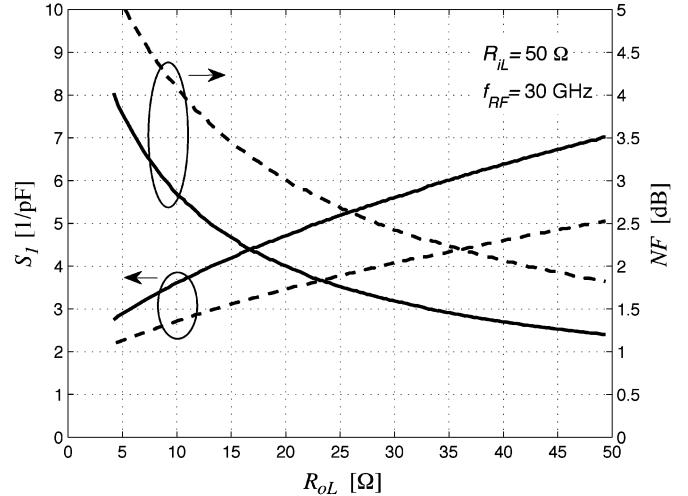


Fig. 5. Plot of S_1 necessary for an LSBDC to achieve 10-dB transducer gain as a function of output resistance R_{oL} for a 30-GHz input and 10-GHz output. Also shown is the NF associated with the S_1 settings. The dashed NF curve pertains to the dashed S_1 setting, while the solid NF curve corresponds to the S_1 values denoted by the solid line.

we can derive the conversion gain of the mixer. In this case, we consider the transducer gain assuming reactive components have been resonated out at the generator and load

$$G_{LSBDC} = \frac{4R_{oL}R_{iL}|S_1|^2}{(\omega_{RF}R'_oR'_i - |S_1|^2/\omega_{IF})^2} \quad (33)$$

where $R'_o = R_{oL} + R_s$ and $R'_i = R_{iL} + R_s$. The central role of the pumped elastance harmonic S_1 is apparent here. The dependence of this variable on pumping and bias for a $0.13-\mu\text{m}$ CMOS AMOSV will be discussed in Section IV-B. The elastance harmonic also bears significant influence on the mixer's noise factor F , which, as calculated from the conversion matrix description, is

$$F_{LSBDC} = 1 + \frac{\overline{e_{nIF}^2} + \overline{e_{nRF}^2} (|S_1|/\omega_{RF}R'_i)^2}{\overline{e_{n,i}^2} (|S_1|/\omega_{RF}R'_i)^2} \quad (34)$$

where $\overline{e_{nIF}^2} = \overline{e_{nRF}^2} = 4kT R_s \Delta f$ are the noise sources at IF and RF associated with the varactor loss and where $\overline{e_{n,i}^2} = 4kT R_{iL} \Delta f$ is the RF input noise source. Since PDCs only operate properly with their filtering already in place, the SSB and DSB noise factor distinction does not apply [18].

The gain and noise behaviors of this circuit using measured elastance ranges indicative of IC AMOSVs (discussed in Section V) are shown in Fig. 5. These results have been corroborated by harmonic balance simulations using Agilent's Advanced Design System (ADS) simulator. All the curves pertain to an LSBDC pumped at 40 GHz while converting a 30-GHz input to a 10-GHz IF with 10-dB transducer gain and 50-Ω generator resistance. A pessimistic R_s of 5-Ω is assumed, which corresponds to a device Q of roughly 20. Remaining circuit losses can be assumed to be lumped into R_{oL} and R_{iL} .

The curves described by the left-ordinate pertain to the pumped S_1 values needed to sustain 10-dB gain. We see that there are two possibilities (solid and dashed curves in Fig. 5) for each R_{oL} , a consequence of the tradeoffs between the LSBDC's

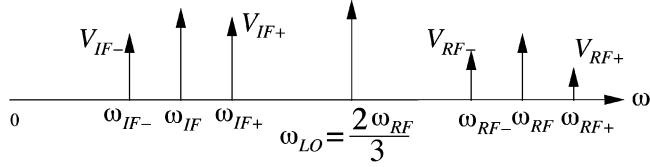


Fig. 6. Frequency components of the DSBDC's port voltages.

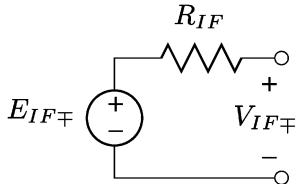


Fig. 7. Thevenin equivalent of the DSBDC IF output port.

open circuit voltage gain and negative output resistance. The advantage of running with a higher S_1 for a given gain is improved noise performance (at the cost of increased power consumption, which we discuss in Section V). As expected from a reactive converter, the noise figure (NF) associated with the necessary S_1 settings is excellent.

B. DSBDC

Fig. 3(c) indicates how the DSBDC couples RF, IF, and their conjugate modes. As done above for the LSBDC, we can represent this in terms of the port parameters in Fig. 4, although now we must expand the number of signals that are coupled as illustrated in Fig. 3(c). In particular, the conversion matrix expression of the DSBDC is

$$\begin{bmatrix} V_{RF-} \\ V_{IF-} \\ V_{IF+}^* \\ V_{RF+}^* \end{bmatrix} = \begin{bmatrix} R_s + \frac{S_0}{j\omega_{RF}} & \frac{S_1}{j\omega_{IF}} & \frac{-S_2}{j\omega_{IF}} & \frac{-S_3}{j\omega_{RF}} \\ \frac{S_1^*}{j\omega_{RF}} & R_s + \frac{S_0}{j\omega_{IF}} & \frac{-S_1}{j\omega_{IF}} & \frac{-S_2}{j\omega_{RF}} \\ \frac{S_2^*}{j\omega_{RF}} & \frac{S_1^*}{j\omega_{IF}} & R_s - \frac{S_0}{j\omega_{IF}} & \frac{-S_1}{j\omega_{RF}} \\ \frac{S_3^*}{j\omega_{RF}} & \frac{S_2^*}{j\omega_{IF}} & \frac{-S_1}{j\omega_{IF}} & R_s - \frac{S_0}{j\omega_{RF}} \end{bmatrix} \times \begin{bmatrix} I_{RF-} \\ I_{IF-} \\ I_{IF+}^* \\ I_{RF+}^* \end{bmatrix}.$$

The subscripts denote the frequency associated with each signal and the sideband it occupies relative to a carrier frequency. This nomenclature is elaborated in Fig. 6; for example, signals just below the IF carrier are denoted with $_{IF-}$, while signals just above the IF carrier are denoted with $_{IF+}$. In the expression above, we also employ a narrowband approximation $\omega_{RF} \approx \omega_{RF-} \approx \omega_{RF+}$ and $\omega_{IF} \approx \omega_{IF-} \approx \omega_{IF+}$.

With the average reactive circuit contributions resonated out (assuming appropriate IF load terminations) and retaining only the elastance fundamental S_1 , we can derive an approximate Thevenin equivalent for the DSBDC's IF port, as illustrated in Fig. 7. As with the LSBDC, the values of this circuit are derived

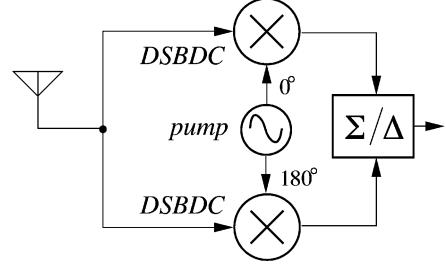


Fig. 8. DSBDC-based noncoherent PAM receiver.

using the conversion matrix in conjunction with (31) and (32). First, the equivalent IF resistance is

$$R_{IF} \approx R_s + \frac{3|S_1|^2}{\omega_{RF}^2 R'_i} - \frac{9|S_1|^2 R'_i}{\omega_{RF}^2 R'_o R'_i + 3|S_1|^2} \quad (35)$$

where R'_i and R'_o are as defined earlier. The equivalent IF voltage depends on the sideband under consideration, as well as the strength of the RF driving signal, specifically

$$E_{IF\mp} \approx \frac{S_1^* E_{i\mp}}{j\omega_{RF} R'_i} + \frac{3S_1^2 E_{i\pm}^*}{\omega_{RF}^2 R'_o R'_i + 3|S_1|^2} \quad (36)$$

where $E_{IF\mp}$ refers to a double-sided output at $\omega_{IF\mp}$, while $E_{i\mp} = E_{RF}$ refers to a double-sided RF input at $\omega_{RF\mp}$. Both sidebands of the RF input signal (E_{i-} and E_{i+}) contribute to each downconverted sideband (E_{IF-} and E_{IF+}) and the contribution of each input sideband is weighted by S_1^* and S_1^2 , respectively.

To aid with an upcoming analysis of the DSBDC's role in radio systems, we retain only the phase sensitive elements and distill (36) to the relation

$$E_{IF\mp} \propto |S_1| e^{j(\pi/2 + \theta_{LO})} E_{i\mp} + k|S_1|^2 e^{-j2\theta_{LO}} E_{i\pm}^* \quad (37)$$

where θ_{LO} is the phase of the LO/pump and k is a proportionality constant standing for the relative weighting between the summands of (36). Equation (37) highlights the dependence of the DSBDC's gain on the phase of the pump, which influences the use of this PDC in receivers. For instance, an M -PAM receiver processes a double-sided RF input of the form

$$E_{i-} = E_{i+} = |E_i| e^{j\theta_i} \quad (38)$$

where θ_i is the phase of the carrier. In this case, (37) becomes

$$\frac{E_{IF\mp}}{|E_i|} \propto e^{j(\pi + \theta_{LO} + \theta_i)} (|S_1| + k|S_1|^2 e^{-j(\pi + 3\theta_{LO} + 2\theta_i)}). \quad (39)$$

We see that the magnitude of the downconversion gain indicated by (39) depends on θ_{LO} and θ_i . In a digital receiver, we can minimize this problem by controlling the LO with the phase correction typically extracted in the baseband of a PAM receiver. In fact, this is a standard approach in modern coherent digital receivers [19]. Alternatively we can eliminate this issue by employing a differential DSBDC arrangement, as shown in Fig. 8, whose operation is independent of the LO phase.

Insight on this arrangement can be obtained from an inspection of (39), which shows that summing or subtracting the outputs of two DSBDCs, one pumped with θ_{LO} and the other with

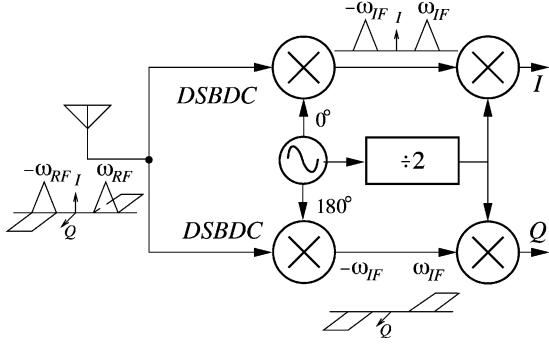


Fig. 9. DSBDC-based heterodyne QPSK receiver alleviating the need for quadrature pumping.

$\theta_{\text{LO}} + \pi$, eliminates the dependence of the gain magnitude on the signal and LO phase. Specifically, adding two versions of (39) that are 180° out of phase results in

$$\frac{E_{\text{IF}\mp}}{|E_i|} \Big|_{\Sigma} \propto 2k|S_1|^2 e^{-j(2\theta_{\text{LO}}+\theta_i)} \quad (40)$$

while the difference produces

$$\frac{E_{\text{IF}\mp}}{|E_i|} \Big|_{\Delta} \propto -2|S_1| e^{j(\theta_{\text{LO}}+\theta_i)} \quad (41)$$

and, as expected, the magnitude of the received signal becomes independent of phase.

The cost of this structure, increased component count and pump complexity, is relatively small in the context of an IC technology where component matching is good and differential oscillations are readily implemented with standard cross-coupled topologies.

The DSBDC also affords a means of processing quadrature phase-shift keying (QPSK) signals without requiring a quadrature LO. From (39), note that the gain of a DSBDC fluctuates between maximum and minimum values as θ_i varies by $\pi/2$ and θ_{LO} varies by $\pi/3$. Thus, if a DSBDC processing a signal with some phase $\theta_{i,I}$ attains maximum gain by realizing some net phase ϕ

$$3\theta_{\text{LO}} + 2\theta_{i,I} = \phi \bmod 2\pi \quad (42)$$

then, for a second DSBDC pumped 180° out of phase with the first and processing a signal with some phase $\theta_{i,Q}$, to achieve maximum gain, we need the equivalence

$$3(\theta_{\text{LO}} \pm n\pi) + 2\theta_{i,Q} = \phi \bmod 2\pi \quad (43)$$

$$3\theta_{\text{LO}} + 2\theta_{i,Q} = (\phi \mp \pi) \bmod 2\pi \quad (44)$$

where n is an odd integer. The second DSBDC can achieve this net phase if the signal that it processes has the phase characteristics $\theta_{i,Q} = \theta_{i,I} \mp \pi/2$. Hence, quadrature input signals may be individually processed by two DSBDCs driven by a differential LO, as shown in Fig. 9, for a heterodyne downconversion scheme. Appearing further down the receive chain, the two IF mixers following the DSBDC pair can be realized using more common transistor-based topologies.

The double-sided transducer gain expression for the DSBDC under optimal pumping conditions can be written in terms of (35) and (36)

$$\begin{aligned} G_{\text{DSBDC}} &= \frac{2R_o L R_{iL}}{(R_o L + R_{iL})^2} \frac{|E_{\text{IF}-}|^2 + |E_{\text{IF}+}|^2}{|E_i|^2} \\ &= R_o L R_{iL} \\ &\times \left\{ \frac{3\omega_{\text{RF}}|S_1|^3 + \omega_{\text{RF}}|S_1|R'_i(3|S_1| + \omega_{\text{RF}}R'_o)}{\omega_{\text{RF}}^4(R'_o R'_i)^2 + 9|S_1|^2(|S_1|^2 - [R'_i \omega_{\text{RF}}]^2)} \right\}^2 \end{aligned} \quad (45)$$

where a double-sided RF input E_i is assumed as per (38).

The NF of the DSBDC can be expressed as

$$F_{\text{DSBDC}} = 1 + \frac{G_{\text{IF}} \overline{e_{n\text{IF}}^2} + G_{\text{RF}} \overline{e_{n\text{RF}}^2}}{G_{\text{RF}} \overline{e_{n,i}^2}} \quad (46)$$

where

$$G_{\text{IF}} = 1 + \frac{3|S_1^*|^2}{(\omega_{\text{RF}}R'_o + 3|S_1|^2/\omega_{\text{RF}}R'_i)^2} \quad (47)$$

$$G_{\text{RF}} = \frac{|S_1|^2}{\omega_{\text{RF}}^2 R'_i^2} + \frac{3|(S_1^*)^2|^2}{(\omega_{\text{RF}}^2 R'_i R'_o + 3|S_1|^2)^2}. \quad (48)$$

The above expressions include the conversion of noise from sidebands $\omega_{\text{IF}-}$, $\omega_{\text{IF}+}$, $\omega_{\text{RF}-}$, and $\omega_{\text{RF}+}$ to the output.

Similar to Fig. 5, a plot of the requisite S_1 and ensuing NF for a 30–10-GHz 10-dB gain DSBDC pumped at only 20 GHz is shown in Fig. 10. The gain is that of a summing DSBDC, as shown in Fig. 8, and hence, independent of pump phase. For the settings examined in this paper, the DSBDC can achieve superior noise and gain at lower S_1 settings than the LSBDC. However, these advantages do not occur simultaneously. That is, very low NF's can be realized, but at high S_1 , settings or low values of S_1 can achieve the requisite gain, but at more significant noise sacrifice than for the LSBDC. The choice of operating regime will depend on radio system requirements.

V. VARACTOR DEVICES

The expressions presented in Section IV underscore the need to maximize S_1 as a general means of improving the gain and noise performance of PDCs. In this section, we examine the extent to which modern silicon IC technology stands to satisfy this need. Specifically the AMOSV, which has become commonplace in CMOS technology, is considered. Compared to pn-junction varactors, the AMOSV typically has a higher Q , is easier to bias, and can sustain sub-harmonic mixing much more effectively [20].

The representative elastance obtained from S -parameter measurements on a variety of 0.13- μm CMOS AMOSV devices is shown in Fig. 11. Results for two different channel dopings are presented as a function of the gate-source voltage drop across the varactor. Although the doped device in this case is not helpful (it pertains to a low-leakage transistor doping with increased channel concentration), modern technologies

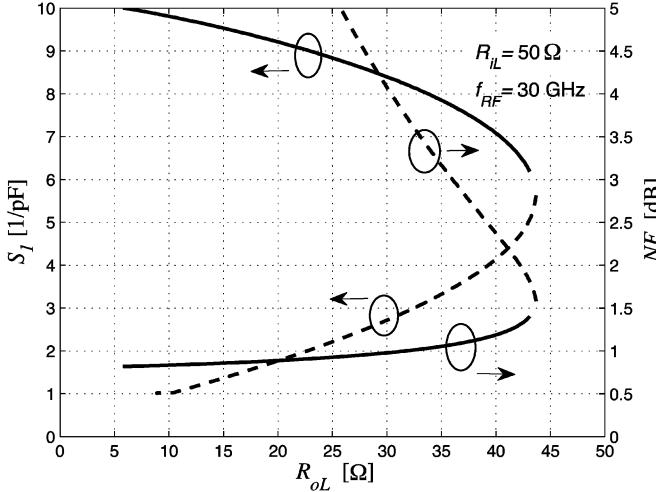


Fig. 10. Plot of the S_1 necessary for a DSBDC to achieve 10-dB transducer gain as a function of output resistance R_{oL} for a 30-GHz input and a 10-GHz output. As in Fig. 5, the dashed NF curve pertains to the dashed S_1 setting with a similar relationship for the solid curves.

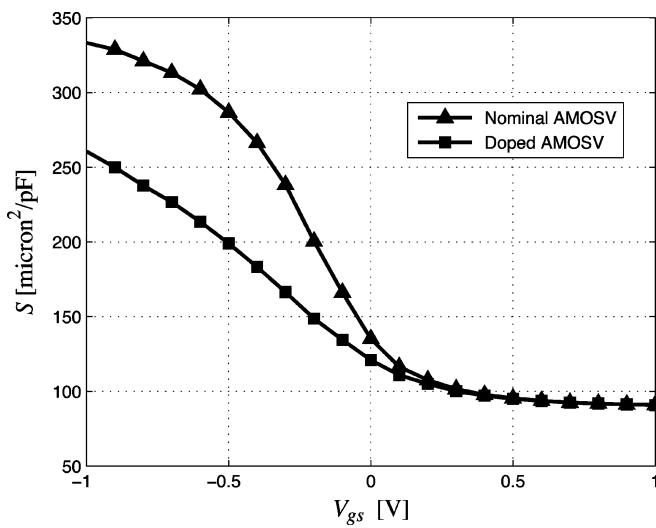


Fig. 11. Measured AMOSV elastance as a function of gate-source voltage. Results for two different channel dopings are presented.

offer native and low-voltage options that could be of significant benefit to AMOSVs for parametric circuits.

The fundamental elastance S_1 is extracted from these measurements for the nominal AMOSV, and is shown in the Fig. 12 contour plot. In essence, this is a sine-input describing function of the AMOSV. S_1 is normalized to device area and measured in units of $\mu\text{m}^2/\text{pF}$ with the abscissa denoting device biasing V_{gs} and the ordinate denoting the pump voltage amplitude V_{LO} . The solid line in Fig. 12 indicates the locus of bias conditions that maximize the S_1 -to- V_{LO} ratio. This optimum region lies roughly between the maximum and minimum elastance regions and is weakly dependent on the applied LO voltage amplitude.

As seen in Fig. 13, the optimum bias line corresponds to an average unit capacitance C'_0 of roughly $4.5 \text{ fF}/\mu\text{m}^2$ requiring device areas of at least $10 \mu\text{m}^2$ to help establish reasonable match conditions for the RF mixer. This issue actually eases as

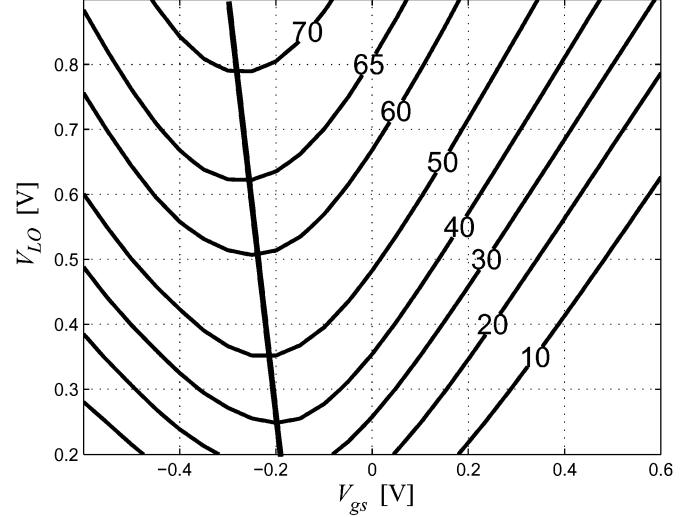


Fig. 12. Contour plot of S_1 in units of $\mu\text{m}^2/\text{pF}$ for an AMOSV in a $0.13\text{-}\mu\text{m}$ CMOS technology as a function of LO voltage pumping amplitude and gate-source bias.

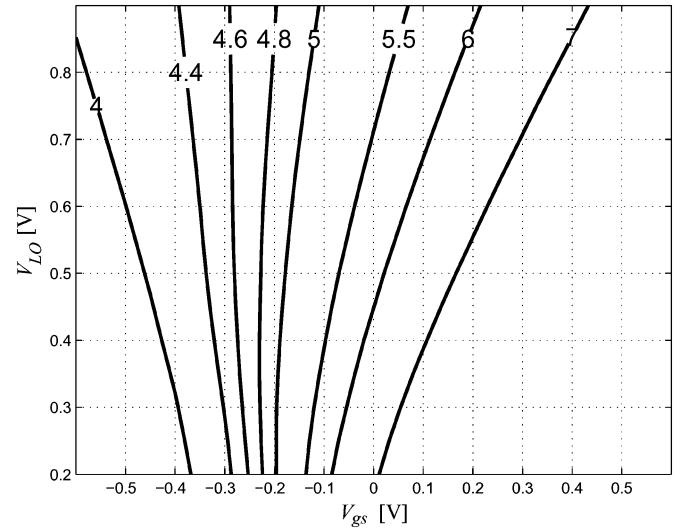


Fig. 13. Contour plot of the average capacitance per unit area $C'_0[\text{fF}/\mu\text{m}^2]$ of the pumped varactor as a function of LO amplitude and bias conditions for an AMOSV in a $0.13\text{-}\mu\text{m}$ CMOS technology.

operational frequencies increase and lower average capacitance values are employed.

Thus, it seems that S_1 's up to 7 pF^{-1} can be achieved in a CMOS technology under reasonable pumping conditions, putting the results presented in Section IV in a broader context. This consideration eases as operational frequencies increase since lower average capacitances are required of the pumped varactor.

Finally, we comment on the power consumption of PDCs. Unlike transistor-based designs, PDCs make more efficient use of the LO by using it as both a power source and a frequency reference. Further, no dc power needs to be expanded to bias the varactor mixing element. The average power required to pump the varactor is

$$P_{LO} = \frac{|V_{LO}|^2 R_s \omega_{LO}^2}{2 R_s^2 \omega_{LO}^2 + 2 S_0^2} \quad (49)$$

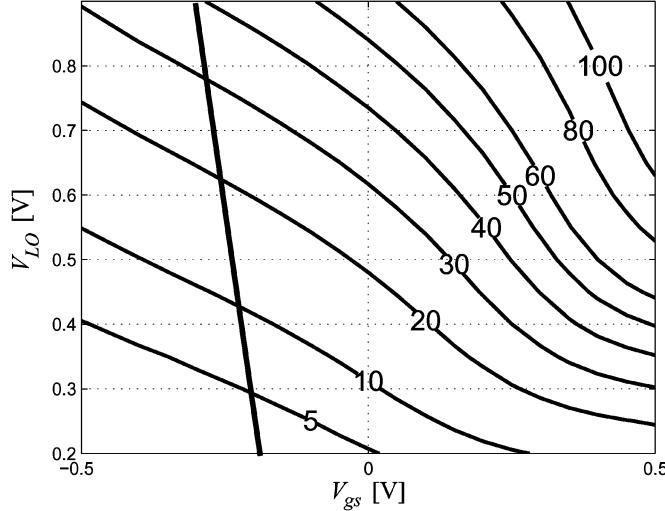


Fig. 14. Contour plot of the LO power in μW needed to pump a $10-\mu\text{m}^2$ AMOSV at 20 GHz.

where S_0 is the average varactor elastance and V_{LO} is the voltage amplitude of the sinusoidal pump signal. The LO power in μW needed to pump a $10-\mu\text{m}^2$ device at 20 GHz under various bias and pumping voltages is shown in Fig. 14. This case envisions a DSBDC pumped at 20 GHz while converting a 30-GHz input to a 10-GHz IF. The straight line corresponds to the one present in Fig. 12. Although somewhat idealized, compared to transistor-based mixers, this power consumption is minuscule, varying from about 5 to 40 μW (four times less than LSBDCs), and provides a large margin for implementation specifics.

VI. BASIC PDC CIRCUIT CONFIGURATION

Thus far we have discussed the underlying operation of PDCs, their performance potential, their influence on radio systems, and AMOSV properties available to realize them. We now turn to the description of a possible PDC implementation in silicon IC technology. To the authors' knowledge, an IC continuous-time RF PDC has not yet been reported in modern CMOS technology. Although the parametric paradigm is popular in optical devices and has met some success in electromechanical systems, the only prominent fully electronic realization is a $0.25-\mu\text{m}$ CMOS low-frequency discrete-time parametric amplifier first presented in [21].

Part of the difficulty with integrating parametric circuits is the need to realize the associated filters. In the past, designers met this challenge with waveguide and transmission line components that precluded realization in IC form even for microstip-based designs [22]. However, recent advances in CMOS towards high-quality lumped and distributed passives are changing this, and allow for the realization of improved filtering components directly on-chip [23].

This development is augmented by the presence of transistors in IC technology, which, as controlled-current sources, allow signals to be coupled into and out of parametric amplifiers relatively efficiently. Further, IC technology allows the pump to be directly interfaced to the varactor, thus foregoing the need

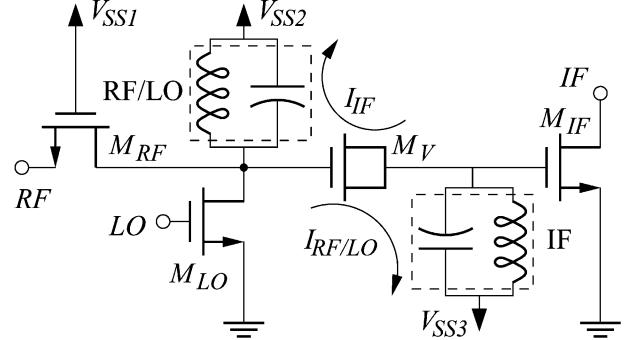


Fig. 15. Simple MOS-based receiver employing a PDC.

for intricate matching circuitry. These advantages were absent when work on parametric circuits was last vigorous.

As an example of how PDCs can be realized in an IC technology, we consider the receiver in Fig. 15. The PDC in this case consists of the AMOSV M_V along with associated filters that can be realized as LC resonators. The RF input and LO pump are both coupled to the PDC via transistors M_{RF} and M_{LO} . The RF/LO filter ensures that both signals drive M_V and simultaneously establishes a return path for the IF current, as indicated in the figure. The RF/LO filter also allows the bias for M_{RF} , M_{LO} , and M_V to be fed into the circuit. Similarly, the IF filter ensures that a return path is present for the RF/LO current while coupling in the bias for M_V and M_{IF} . The transistor M_{IF} extracts the downconverted signal with minimal influence on the PDC's output port. Although the load at the drain of M_{IF} is not specified, a typical scenario may employ this transistor as the transconductance element in a single-balanced mixer, thus making low-impedance field-effect transistor (FET) source terminals its likely termination.

By isolating the dc signals in the RF and IF stages of this receiver, the series-coupling PDC (i.e., M_V forms a series connection between RF and IF ports) simplifies the biasing in this circuit. In this configuration, the PDC is also less sensitive to shunt parasitics, which can be absorbed into the surrounding filters.

In designing the circuit, the RF/LO filter is selected to peak between the RF and LO input frequencies, while the IF filter is set to resonate near ω_{IF} . Once the RF, LO, and IF transistors are attached, the quality factors of these filters are largely set by the losses associated with the constituent passives, as well as those present at the gate and drains of the accompanying transistors. M_V is then inserted and a conversion-matrix analysis, as discussed above, is carried out. The dimensions of M_V are iterated until a desired gain and pumping current are achieved. Note that, in this case, the loads R_{iL} and R_{oL} are the series equivalent losses of the RF/LO and IF filters, respectively, and no effort has been taken to establish a reactive match.

Example analysis and simulation results appear in Fig. 16 for the voltage gain G_v and NF associated with the LSBDC-based receiver shown in Fig. 15. Specifically, G_v refers to the gain from the source of M_{RF} (assuming a 10-mS transconductance) to the gate of M_{IF} , while the NF is calculated from the drain of M_{RF} to the gate of M_{IF} . Analysis results (solid lines) based on the procedures discussed in Section IV compare favorably

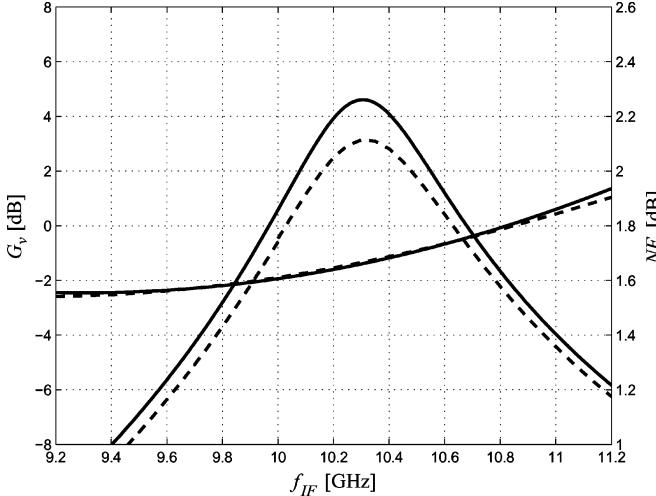


Fig. 16. Voltage gain G_v and NF of a 130-nm LSBDC converting a 30-GHz input to a 10-GHz output. Dashed lines indicate ADS simulation results.

with harmonic balance ADS simulations (dashed lines) run on a physical charge-based analog hardware description language (AHDL) model of the AMOSV available from the foundry (the transistors are idealized to current sources and their parasitics are included in the filter structures to simplify analysis). The converter produces a 10-GHz IF for a 30-GHz input and the net quality factors of the RF/LO and IF filters are 6 and 10, respectively. M_V has dimensions of $L = 0.36 \mu\text{m}$ and $W = 1 \mu\text{m}$ and is arranged in eight stripes with each stripe consisting of seven gate fingers. This PDC's simulated input-referred 1-dB compression point (excluding the RF transistor) for a 30-GHz input is -16 dBm with a third-order intermodulation intercept point (IIP3) of -3.2 dBm .

The LO power needed to drive the entire PDC (including filters) is 3.1 mW . This compares quite favorably to 130-nm transistor-based mixers, which require both LO and dc power to operate [24]. Further, rearranging this circuit into a DSBDC configuration and thus cutting ω_{LO} in half promises to reduce the present consumption by a factor of 4. Still, referring to Fig. 14, it should be noted that the drawn pump power is much higher than that needed by the varactor. The difference is substantially the result of losses in the filters and can be reduced by improving filter quality. For instance, doubling the IF filter quality factor to 20 approximately quarters the pump power necessary to maintain the present levels of gain.

VII. CONCLUSION

PDCs are a rarity among RF communications blocks. This was the case even at the height of research into parametric circuits. However, given the increasing expectation of meeting millimeter-wave applications with silicon technology, plus the growing difficulty of scaling CMOS transistors, parametric circuits now have the potential to be an excellent complement to mainstream RF design techniques. As discussed in this paper, PDCs can operate at low noise and at low power, and under the right conditions, can facilitate signal gain as well.

This paper has discussed PDCs from the coupled-mode perspective and used this approach to develop insight on the

characteristics of PDCs with gain. With the coupled-mode description, we can quickly identify parametric circuits capable of gain by noting whether their time-varying components couple counter-rotating modes. This allowed a simple pictorial analysis of PDCs, which led to the proposal of a mixer that achieves gain at a lower pumping frequency: the DSBDC.

The fundamental gain and noise limitations of this mixer along with the more traditional LSBDC were explored in the context of the converter elastance. Further, the pumping and bias conditions needed to maximize varactor elastance in a 130-nm CMOS technology, while minimizing LO requirements were discussed. The gain peculiarities of the DSBDC were explored as well and architectures proposed, which simplify the LO requirements in quadrature communicators. Finally, a means of embedding a PDC in an IC as a series-coupled structure was discussed.

ACKNOWLEDGMENT

The authors wish to acknowledge CMC Microsystems, Kingston, ON, Canada, for IC fabrication.

REFERENCES

- [1] S. Magierowski, T. Zourntos, J.-F. Bousquet, and Z. Zhao, "Compact parametric downconversion using MOS varactors," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2009, pp. 1377–1380.
- [2] M. E. Hines, "The virtues of nonlinearity—Detection, frequency conversion, parametric amplification and harmonic generation," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-32, no. 1, pp. 1097–1104, Jan. 1984.
- [3] H. Q. North, "Properties of welded contact germanium rectifiers," *J. Appl. Phys.*, vol. 17, pp. 912–923, Nov. 1946.
- [4] H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers*, ser. MIT Radiat. Lab. New York: McGraw-Hill, 1948, vol. 15.
- [5] G. F. Herrmann, M. Ueno-hara, and A. Uhliir, Jr., "Noise figure measurements on two types of variable reactance amplifiers using semiconductor diodes," *Proc. IRE*, vol. 46, no. 6, pp. 1301–1303, Jun. 1958.
- [6] A. E. Bakanowski, N. G. Cranna, and A. Uhliir, Jr., "Diffused silicon nonlinear capacitors," *IEEE Trans. Electron Devices*, vol. ED-6, no. 10, pp. 384–390, Oct. 1959.
- [7] R. V. Pound, *Microwave Mixers*, ser. MIT Radiat. Lab. New York: McGraw-Hill, 1948, vol. 16.
- [8] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—Part I. General energy relations," *Proc. IRE*, vol. 44, no. 7, pp. 904–913, Jul. 1956.
- [9] A. Uhliir, Jr., "The potential of semiconductor diodes in high-frequency communications," *Proc. IRE*, vol. 46, no. 6, pp. 1099–1115, Jun. 1958.
- [10] C. A. Mead, "Schottky barrier gate field effect transistor," *Proc. IEEE*, vol. 54, no. 2, pp. 307–308, Feb. 1966.
- [11] D. S. Kurtz, J. L. Hesler, T. W. Crowe, and R. M. Weikle, II, "Submillimeter-wave sideband generation using varactor schottky diodes," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 1, pp. 2610–2617, Jan. 2002.
- [12] A. Suarez and R. Melville, "Simulation-assisted design and analysis of varactor-based frequency multipliers and dividers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 3, pp. 1166–1179, Mar. 2006.
- [13] S. Ranganathan and Y. Tsividis, "Discrete-time parametric amplification based on a three-terminal MOS varactor," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2087–2093, Dec. 2003.
- [14] W. H. Louisell, *Coupled Mode and Parametric Electronics*. New York: Wiley, 1960.
- [15] H. Haus and W. Huang, "Coupled-mode theory," *Proc. IEEE*, vol. 79, no. 10, pp. 1505–1518, Oct. 1991.
- [16] C. S. Teoh and L. E. Davis, "A coupled pendula system as an analogy to coupled transmission lines," *IEEE Trans. Educ.*, vol. 39, no. 4, pp. 548–557, Nov. 1996.
- [17] S. A. Maas, *Nonlinear Microwave Circuits*. New York: IEEE Press, 1996.
- [18] S. A. Maas, *Microwave Mixers*, 2nd ed. Boston, MA: Artech House, 1993.

- [19] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York: Wiley, 1997.
- [20] S. Magierowski, H. Chan, and T. Zourntos, "Subharmonically pumped RF CMOS paramp," *IEEE Trans. Electron Devices*, vol. 55, no. 2, pp. 601–608, Feb. 2008.
- [21] S. Ranganathan and Y. Tsividis, "A MOS capacitor-based discrete-time parametric amplifier with 1.2 V output swing and 3 μ W power dissipation," in *Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2003, pp. 406–407.
- [22] P. Bura, R. Camisa, W. Pan, S. Yuan, and A. Block, "Design considerations for an integrated 1.8-GHz parametric amplifier," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-16, no. 7, pp. 424–428, Jul. 1968.
- [23] C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Brodersen, "Millimeter-wave CMOS design," *IEEE J. Solid-State Circuits*, vol. 40, pp. 144–155, Jan. 2005.
- [24] L. Zhang and Z. Yu, "Transformer-based low voltage design of 17 GHz mixers with on-chip balun using 0.13- μ m standard CMOS process," in *Proc. IEEE Conf. Electron Devices and Solid-State Circuits*, Nov. 2007, pp. 457–460.



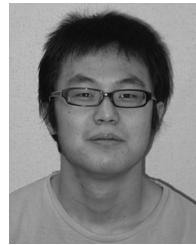
Sebastian Magierowski (M'04) received the B.A.Sc. degree in electrical engineering from the University of Windsor, Windsor, ON, Canada, in 1994, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1997, and 2004, respectively. While working toward the Ph.D. degree, he held internships with Nortel Technology and PMC-Sierra Inc. related to integrated device modeling and mixed-signal design. Since 2004, he has been a faculty member with the Department of Electrical and Computer Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada. His research interests include RFIC design, nonlinear circuits, low-power analog integrated computation, and multiuser radio architectures.



low-power electronics.

Jean-Francois Bousquet (S'07) was born in Montreal, QC, Canada, in 1977. He received the B.Sc. degree in electrical engineering from the Ecole Polytechnique de Montreal, Montreal, QC, Canada, in 2000, the M.Sc. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 2006, and is currently working toward the Ph.D. degree in electrical and computer engineering at the University of Calgary.

His research interests are currently focused on the development of efficient radio architectures using



Zhixing Zhao (S'09–A'09) was born in Dali, Yunnan, China, in 1984. He received the B.Sc. degree in electronic science and engineering from Jilin University, Changchun, China, in 2007, and is currently working toward the M.Sc. degree in electrical and computer engineering at the University of Calgary, Calgary, AB, Canada.

He is currently with the Fully Integrated Systems Hardware Laboratory (FISH Laboratory), University of Calgary, where he is engaged in the development of integrated parametric circuits at high frequency.

His current research interest concerns fully-integrated low-noise parametric circuits millimeter-wave applications.



Takis Zourntos (S'91–M'00–SM'08) received the B.A.Sc. degree in engineering science and M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 1993, 1996 and 2003, respectively.

From 1998 to 2001, he was a co-founder and System Designer with the Protolinx Corporation, Toronto, ON, Canada. Since 2003, he has been a faculty member with the Department of Electrical and Computer Engineering, Texas A&M University, College Station. His research interests include robotics, mixed-signal ICs, and the application of nonlinear control and dynamical systems theory to a wide range of problems.

Dr. Zourntos is a Registered Professional Engineer in the Province of Ontario. He was the recipient of the 2003 IEEE Students' Choice Award for Best Electrical Engineering Professor at Texas A&M University.

Sensitivity Enhancement of Remotely Coupled NMR Detectors Using Wirelessly Powered Parametric Amplification

Chunqi Qian,* Joseph Murphy-Boesch, Stephen Dodd, and Alan Koretsky

A completely wireless detection coil with an integrated parametric amplifier has been constructed to provide local amplification and transmission of MR signals. The sample coil is one element of a parametric amplifier using a zero-bias diode that mixes the weak MR signal with a strong pump signal that is obtained from an inductively coupled external loop. The NMR sample coil develops current gain via reduction in the effective coil resistance. Higher gain can be obtained by adjusting the level of the pumping power closer to the oscillation threshold, but the gain is ultimately constrained by the bandwidth requirement of MRI experiments. A feasibility study here shows that on a NaCl/D₂O phantom, ²³Na signals with 20 dB of gain can be readily obtained with a concomitant bandwidth of 144 kHz. This gain is high enough that the integrated coil with parametric amplifier, which is coupled inductively to external loops, can provide sensitivity approaching that of direct wire connection. *Magn Reson Med* 68:989–996, 2012. © 2012 Wiley Periodicals, Inc.

Key words: parametric amplifier; varactor; wireless

It has long been a practice in the magnetic resonance community to use small coils for localized spectroscopy and imaging (1,2), for example catheter coils for interventional MRI (3). In the case of implanted or catheter coils, it has been common to run a hard-wired connection from the coil to the preamplifier, sometimes over long distances, leading to sensitivity losses and RF heating along the cables. Mutual inductive coupling between an internal coil and an external pick-up loop can be used to eliminate wires; however, the separation between the external loop and the receive coil must be close if sensitivity is to be maintained (2,4). It would be advantageous to amplify the MRI signal first before it is transmitted wirelessly to an external coil. Most transistor-based low-noise amplifiers require a DC power source that is difficult to provide wirelessly. Parametric amplifiers (5,6), on the other hand, can obtain power from an externally applied RF magnetic field. The pumping signal can be used either to amplify the weak NMR signal at the same frequency or to mix and amplify the signal to a higher frequency. Parametric amplifiers were widely used for low-noise amplification prior to the

advent of high-frequency transistors. There has been renewed interest in using parametric amplifiers as MR detectors (7–10) and traveling wave amplifiers in the context of metamaterials (11–13).

Common implementations of a parametric amplifier use a reverse-biased varactor and this generally requires a wire connection to bias the diode. In this work, a parametric amplifier with a zero-biased varactor is demonstrated without bias connections, and amplification can be integrated within a compact, wireless sample circuit. Although the varactor is inaccessible for tuning, the parametric amplifier can provide sufficient gain even though the detection circuit is detuned from the Larmor frequency, as long as the frequency and power of the pumping signal are adjusted to compensate for this detuning. A recently reported wireless design used the upper side-band of a parametric amplifier for gain in phased-array detectors (8). However, our design utilizes the lower mixing sideband that can provide higher gains of up to 27 dB, and pumping power consumption is reduced to only 0 dBm using the zero-biased diode. The integrated detector and amplifier can have greatly enhanced sensitivity when it is remotely coupled to an external loop.

MATERIALS AND METHODS

Figure 1 illustrates passive and active modes for detecting with inductively coupled resonators. In Fig. 1a, the NMR signal is detected with a passive resonator that is coupled inductively to a larger, external loop. The signal-to-noise (power) ratio detected with the external loop is

$$\left(\frac{S}{N}\right)_0 = \frac{S_0 E_1}{N_1 E_1 + N_{Lp1}}. \quad [1]$$

Here, S_0 is the total signal power generated by the nuclear spins, N_1 is the noise power originating from the local resonator, N_{Lp1} is the noise power originating from the external loop, and E_1 is the transmission efficiency, i.e., the power delivered to the pick-up loop per unit power originating from the local resonator. If the local resonator has a power gain of G_1 , the signal-to-noise ratio becomes

$$\left(\frac{S}{N}\right)_1 = \frac{S_0 E_1 G_1}{N_1 F_1 E_1 G_1 + N_{Lp1}}. \quad [2]$$

where F_1 is the noise factor of the amplification process. If the gain is sufficient such that $E_1 G_1 \gg 1$, the noise of

Laboratory of Functional and Molecular Imaging, National Institute of Neurological Disorder and Stroke, National Institutes of Health, Bethesda, Maryland, USA.

Grant sponsors: NIH, NINDS (Intramural Research Program)

*Correspondence to: Chunqi Qian, 10 Center Drive Room 3D17, National Institutes of Health, Bethesda, MD 20892. E-mail: qianc2@ninds.nih.gov

Received 3 May 2011; revised 2 October 2011; accepted 5 October 2011.

DOI 10.1002/mrm.23274

Published online 13 January 2012 in Wiley Online Library (wileyonlinelibrary.com).

© 2012 Wiley Periodicals, Inc.

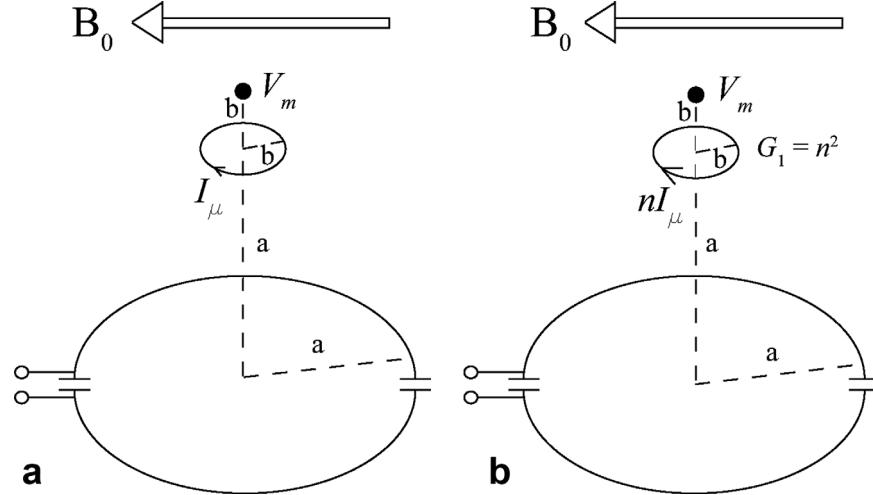


FIG. 1. Functional description of the isolated parametric amplifier. **a:** A small volume (V_m) of sample is directly detected by an embedded local resonator of radius b at a distance separation of b . The nuclei spins induce a current flow I_μ in the local resonator, which is then inductively coupled to an external pick-up loop at a distance separation of a ($b \ll a$). **b:** The current flow is increased by a factor of n due to parametric frequency mixing, and the embedded local resonator has a power gain G_1 of n^2 .

the local resonator can be amplified to a level much larger than that of the external loop, and the signal-to-noise ratio is approximately S_0/N_1F_1 , which is comparable to that of a local resonator with a direct wire connection.

A local resonator with gain was implemented as a triple frequency parametric resonator shown in Fig. 2a. To promote mixing, strong currents flow through the varactor C_2 at each of its resonance frequencies, ω_{10} , ω_{20} , ω_{30} . The circuit is constructed such that $\omega_{10} + \omega_{20} \approx \omega_{30}$ with ω_{10} close to the Larmor frequency ω_1 . The weak NMR signal at ω_1 is mixed with a strong pumping input at ω_3 to produce an amplified output at $\omega_2 = \omega_3 - \omega_1$. The output signal at ω_2 is mixed again with the pumping signal at ω_3 to provide a secondary output with gain at ω_1 . This secondary signal manifests itself as an increased current flow at ω_1 . Previous analysis of parametric amplifiers (5) addressed the case where the circuit is tuned precisely to its driving frequency, i.e., $\omega_{10} = \omega_1$, $\omega_{20} = \omega_2$, and $\omega_{30} = \omega_3$. The analysis can be extended to include the off-resonance behavior (see Appendix), which allows for the situation when the resonator is isolated and cannot be retuned. To optimize the gain, two conditions should be satisfied: first, the frequency matching condition

$$\frac{2Q_1(\omega_1 - \omega_{10})}{\omega_{10}} \approx \frac{2Q_2(\omega_2 - \omega_{20})}{\omega_{20}} \equiv \kappa \quad [3a]$$

where Q_1 and Q_2 are the quality factors at ω_{10} and ω_{20} , respectively; second, the resistance matching condition

$$R_N \equiv \frac{M^2}{C_2^2(1 - M^2)^2 \omega_1 \omega_2 R_2(1 + \kappa^2)} \approx R_1 \quad [3b]$$

where C_2 is the varactor capacitance at zero bias, M is the modulation index defined by $C_2(t) = C_2(1 + 2M \cos \omega_3 t)$, R_1 and R_2 are the effective resistance of the circuit loop containing the varactor at ω_{10} and ω_{20} ,

respectively, and κ is the ratio defined in Eq. 3a. Equation 3b defines the “negative resistance” R_N that can cancel the circuit resistance R_1 , and this is done by adjusting the modulation index M . When $R_1 = R_N$, $M = M_0$ and the parametric resonator begins to oscillate. To avoid oscillation, $R_1 > R_N$ is required, and the power gain can be estimated from the extent to which the resonator deviates from the oscillation condition

$$G_1 = \left(\frac{R_1}{R_1 - R_N} \right)^2 \approx \frac{1}{4} \left(\frac{M_0}{M_0 - M} \right)^2. \quad [4]$$

Figure 2b shows the prototype resonator that was constructed. This configuration was chosen to ensure that significant current passed through the diode at each frequency. It was constructed on a planar Teflon substrate of 1.6-mm thickness. A single, large, rectangular inductor $24 \times 8 \text{ mm}^2$ in size was cut from single-clad Cu PCB, with a single break at the center of one long conductor for the varactor diode. Two parallel circuits were added at each end of the rectangle for triple resonance. The circuit resonance ω_{10} was chosen to be close to the ^{23}Na frequency and ω_{20} was chosen close to the ^1H frequency. The X-channel of the spectrometer was used to observe ^{23}Na signals, and the proton-channel was used to observe the lower side-band of the up-converted frequency. The sample used for all experiments was a 3-mm tube filled with 1 M NaCl/D₂O solution. The resonance frequencies of the resonator when loaded with the sample were, $\omega_{10} = 131.7$ MHz, $\omega_{20} = 498.7$ MHz and $\omega_{30} = 628.4$ MHz, and the quality factors were $Q_1 = 93$, $Q_2 = 48$ and $Q_3 = 60$. There were two separate single-tuned loops on the periphery that were weakly coupled to this resonator. Loop 1 coupled with L_2 to excite the sodium spins at ω_1 , and loop 3 coupled with L_2 for the pumping signal at ω_3 .

For NMR experiments, loop 3 was placed on the same plane as the resonator, and loop 1 was placed directly above the rectangular inductor such that it could be used for both excitation and detection at ω_1 (Fig. 2c). ω_1 was

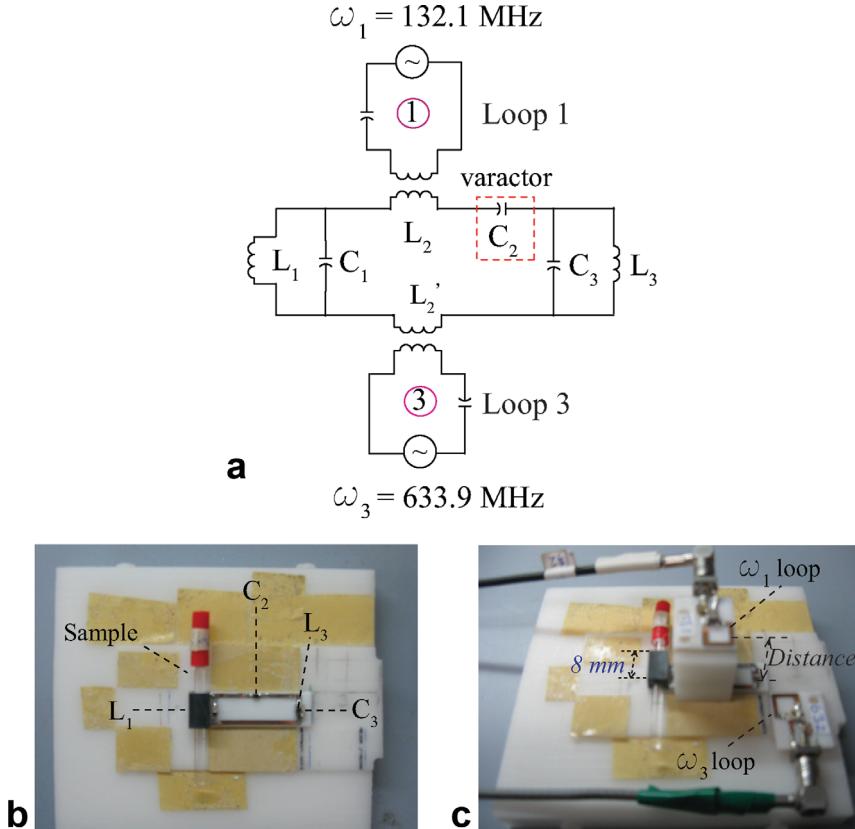


FIG. 2. **a:** The circuit diagram of the integrated coil with parametric amplifier and two peripheral loops labeled by 1, 3. **b:** A snapshot of the resonator showing the sample placement. The actual components used for the triple frequency resonator were: $L_1 = 90$ nH (Coilcraft 132-09SM), $C_1 = 3.6$ pF (ATC-100A), L_2 is a rectangular copper loop with a dimension of 24×8 mm 2 , C_2 is a varactor (Skyworks Inc., SMV-1139) with a zero-bias capacitance of 8 pF, $L_3 = 1.65$ nH (Coilcraft 0906-2KL), $C_3 = 33$ pF (ATC-100B). The volume dimensions of the entire resonator were $30 \times 8 \times 6$ mm 3 . L_1 contains a glass cylindrical sample tube with the red cover. **c:** For the actual NMR experiments, the triple frequency resonator couples with loop 1 and loop 3 through the inductor L_2 . Loop 3 is placed in the same plane as the resonator. Loop 1 is placed directly above the resonator at a certain distance separation, and it is used for both excitation and detection at ω_1 . NMR and MRI experiments were performed using an 11.7 T horizontal magnet (Magnex Inc., Oxford, UK) equipped with an Avance 3 spectrometer console (Bruker Inc., Billerica, MA). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

132.1 MHz which was the ^{23}Na Larmor frequency at 11.7 T, and the correct operating condition was found by empirically adjusting the pumping signal to a proper frequency and power level so that the sharp oscillation peak overlapped the anticipated spectrum peak. The optimal pumping frequency was observed to be $\omega_3 = 633.9$ MHz. According to $\omega_2 = \omega_3 - \omega_1$, ω_2 was 501.8 MHz, which was in good agreement with the frequency predicted by Eq. 3a. The pumping power level was then reduced and spectra and images were acquired.

RESULTS

The parametric amplifier is predicted to have a bandwidth that is inversely proportional to the square root of the power gain $\sqrt{G_1}$ (5). To test this relation using a zero-biased diode, bench measurements were performed on the parametric resonator with Loop 3 driving the pump signal, and a double pick-up loop placed directly above the resonator was connected to a network analyzer to measure gain and bandwidth. A reference curve was first measured using no pumping power, and the remain-

ing curves were obtained by gradually increasing the pumping power until it reached $\sim 95\%$ of the oscillation value. As is shown in Fig. 3a, when the gain increases, the bandwidth decreases. The relative height of each curve reflects the power gain G_1 of the circuit at ω_1 , and this is plotted in Fig. 3b to show the reciprocal relationship between the bandwidth and $\sqrt{G_1}$. When the resonator reaches a gain of 20 dB ($1/\sqrt{G_1} = 0.1$), the bandwidth at -3 dB is 144 kHz, which is sufficient for most MRI experiments.

Gain measurements were also obtained from the parametric circuit using ^{23}Na NMR. The modulation index M_0 was determined from the point of oscillation, and the ratio $M_0/(M_0 - M)$ was monitored as the gain of the NMR signal was being measured (Note: M is not the magnetization of the nuclear spins). In Fig. 3c the magnitude of the signal gain, $\sqrt{G_1} - 1$, was plotted against $M_0/(M_0 - M)$ to show the linear relationship predicted by Eq. 4. $\sqrt{G_1}$ was obtained by taking the ratio of spectra intensities at ω_1 with and without parametric amplification. In the plot, we use $\sqrt{G_1} - 1$ instead of $\sqrt{G_1}$ to remove the source contribution to the output signal. This is an

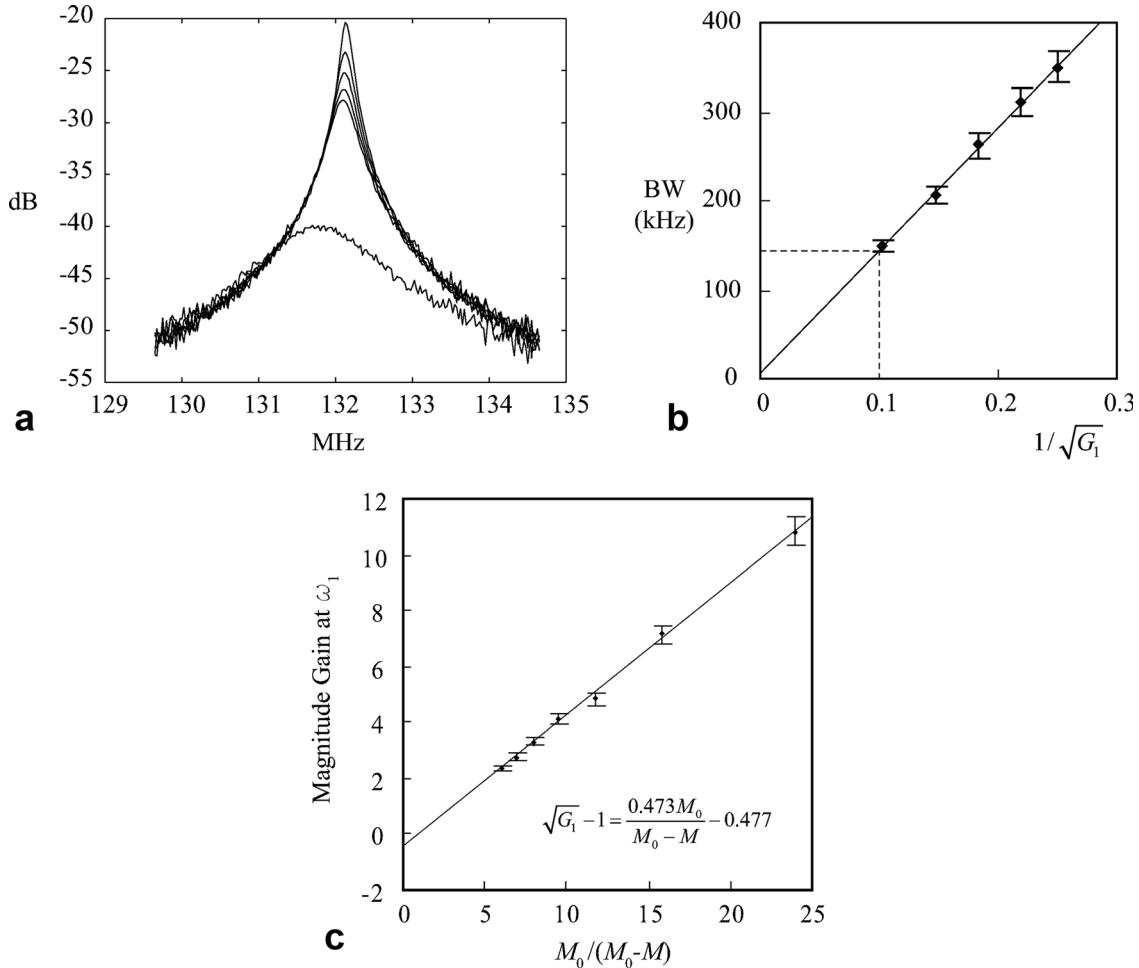


FIG. 3. **a:** The bench measurement of the S_{12} transmission curve with a double pick-up loop connected to an HP-8715 network analyzer. The bottom-most curve is measured without pumping power. As the pumping power is increased, the peak position rises and the bandwidth decreases. The five curves correspond to the pumping power settings that are 1.16, 1.00, 0.84, 0.66, and 0.46 dB below the oscillation condition. **b:** The linear dependence of bandwidth on the reciprocal of the magnitude gain. For each curve shown in Fig. 3a, the bandwidth is measured for the -3 dB points with respect to the peak value, and the power gain (in dB) is obtained by subtracting the peak value with the corresponding S_{12} value of the bottom-most curve at that particular frequency. **c:** The magnitude gain as measured by NMR experiments using the setup detailed in Fig. 2c. M_0 is the modulation index required for oscillation, M is the modulation index used in NMR experiments, and the vertical axis represents the magnitude gain at the output frequency ω_1 excluding the contribution from the unamplified signal.

important modification to Eq. 4, especially when the gain is low. The line in the figure has a slope of 0.473, which is in good agreement with the theoretical value 0.5 in Eq. 4, and the intercept at the vertical axis is close to the slope, which reconfirms $\sqrt{G_1} = 1$ when $M = 0$.

In Fig. 4a, the sensitivity of the parametric resonator is compared with that of a single tuned coil matched to $50\ \Omega$ and connected by cable to the spectrometer preamp; both used the same sample and identical sample coils. The parametric resonator was positioned a short distance (1.1 cm) away from the external pickup loop, and its gain was set to 20 dB. The spectrum and images obtained by the parametric resonator are displayed in the left column of Fig. 4a, and those obtained by the direct connected coil are displayed in the right column. The comparison of spectral intensity in the first row demonstrates that at short distance separation, the parametric resonator is only 10% less sensitive than a coil with a direct connection. The comparison of MRI images

in the bottom rows demonstrates that the parametric resonator can be used to obtain high quality images. Figure 4b demonstrates the sensitivity advantage of parametric amplification under weak coupling conditions where the distance separation between the external pickup loop and the parametric resonator was 2.8 cm. Images collected in the absence of pumping power are shown in the left column, and images acquired in the presence of pumping power are shown in the right column. The pumping power was ~ 0.4 dB below oscillation. The images in the bottom row are 16 dB more sensitive than the images in the top row. This greatly enhanced sensitivity demonstrates the advantage of a parametric resonator over a passive resonator when signals are remotely detected.

Figure 5 shows the dependence of sensitivity on the separation between the resonator and external pick-up loop. For each distance, the pumping power was adjusted to three levels: 1.70, 0.46, and 0.20 dB below

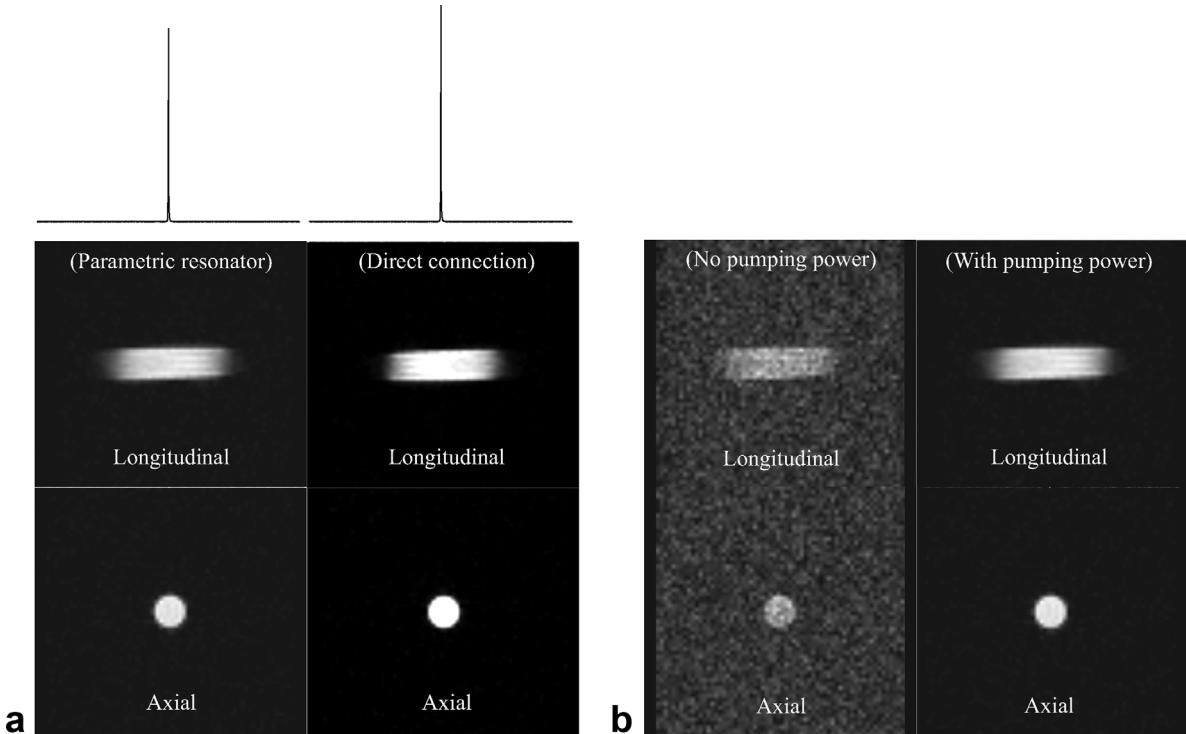


FIG. 4. **a:** A performance comparison between the parametric resonator and a local resonator with direct wire connection. The images in the left column were acquired by the parametric resonator according as the experimental setup detailed in Fig. 2c, with the distance separation between the resonator and the external pick-up loop to be 1.1 cm. The images in the right column were acquired by the same inductor coil matched to $50\ \Omega$ and directly connected to the spectrometer preamp. The first row shows the intensity comparison for single pulse experiments when the peak height is normalized to the same noise level. The second row and third row show the comparison of longitudinal and axial slices of 3D FLASH experiments. The sample used in all experiments is 1 M NaCl/D₂O solution contained in 3-mm tube, and the parameters for FLASH images are: TR = 100 ms, TE = 3.1 ms, NEX = 1, FOV = $1.9 \times 1.9 \times 1.9\text{ cm}^3$, matrix size $64 \times 64 \times 64$. **b:** A 3D FLASH images acquired by the weakly coupled parametric resonator without amplification (left column) and with 20 dB gain (right column). The distance separation between the parametric resonator and the external pick-up loop is 2.8 cm. Other experimental parameters are the same as those in Fig. 4a.

oscillation, which corresponded to gain levels of 10, 20, and 27 dB, respectively. The sensitivity measured under each condition was then normalized against a reference spectrum obtained with the single-tuned coil having a direct cable connection. The normalized sensitivities of Na spectra acquired without parametric amplification are plotted as the blue curve at the bottom, which is described by Eq. 1 and falls rapidly at the rate of -8.4 dB/cm . As pumping power is applied, the gain overcomes the attenuation at short distances, and the sensitivity curves flatten. Most notable is the sensitivity profile at 27 dB gain (red curve), which remains above -3.4 dB for distance separations up to 3.7 cm. This distance separation is much larger than the dimension of the resonator and the pick-up loop. No experiments were attempted beyond 27 dB gain, because the -3 dB bandwidth falls to 67 kHz according to Fig. 3b, which is a lower limit for many MRI experiments. For larger separations, the attenuation eventually dominates, and the sensitivity curves fall off at the same rate as the passive circuit.

DISCUSSION

A detector with an integrated parametric amplifier has been constructed to demonstrate enhanced detection sen-

sitivity of remotely coupled detectors. The detection coil had a larger dimension than the pick-up loop, but according to the principle of reciprocity, this is equivalent to the case when the parametric resonator has a small dimension, as would be used for implantation, and the external loop could have a much larger dimension. The radius of the external pick-up loop approximates the required penetration depth, while the dimensions of the local resonator should be large enough to have sufficient field-of-view and small enough to fit into the internal space within the tissue. A parametric resonator on the mm scale should be straightforward to make.

The pumping signal used to power the parametric circuit creates an additional heat source besides the NMR excitation pulse (14). For our prototype circuit, the output power level from the pumping loop was 13 dBm, which yielded an effective power received by the resonator of 0 dBm after attenuation was taken into account. A zero-biased varactor requires less pumping power owing to its greater nonlinearity, and this power is far less than that required by a low-noise transistor amplifier and too small to induce local heating. If the resonator is implanted such that a passively coupled circuit experiences a 30 dB reduction in signal-to-noise ratio, the pumping power required to recover this sensitivity can

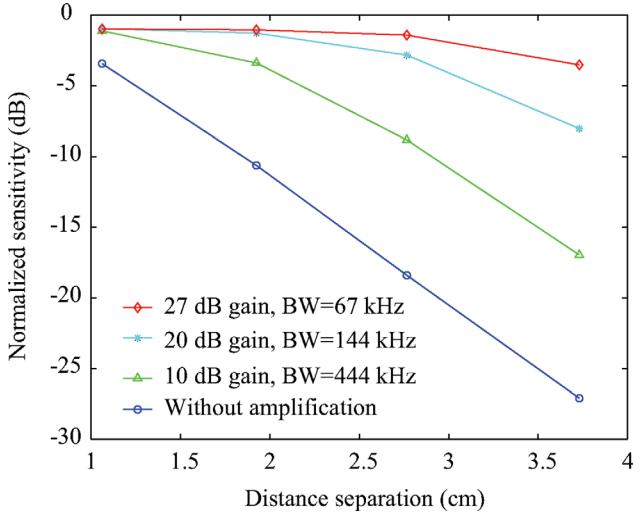


FIG. 5. Dependence of the normalized detection sensitivity on the separation between the parametric resonator and the output loop. The resonator has overall dimensions of $3 \times 0.8 \times 0.6 \text{ cm}^3$, while the rectangular output receiving loop has a dimension of $6.5 \times 6.5 \text{ mm}^2$. The blue curve represents the normalized sensitivity when signal is detected without parametric amplification, the green curve represents the normalized sensitivity when pumping power is 1.70 dB below oscillation and achieves a gain of 10 dB, the cyan curve represents the normalized sensitivity when the pumping power is 0.46 dB below oscillation and achieves a gain of 20 dB, and the red curve represents the normalized sensitivity when the pumping power is 0.20 dB below oscillation and achieves a gain of 27 dB. The bandwidth at each gain level is estimated from the linear relation obtained in Fig. 3b.

also be estimated as 30 dBm, which is a relatively low power. However, further studies will be needed to evaluate the RF safety of the wirelessly powered parametric amplifier for in vivo applications (15). In this context, according to Eq. 3b, the modulation index M_0 for oscillation is proportional to $1/\sqrt{Q_1 Q_2}$, thus the required pumping power is proportional to $1/(Q_1 Q_2)$. Therefore, when a parametric resonator with smaller size is constructed, it is important to design for good quality factors to limit the required power.

In summary, a parametric resonator with a zero-biased varactor circuit operating efficiently in a detuned condition has been demonstrated. A simple triple frequency resonator can wirelessly harvest external power and amplify signals for detection via inductive coupling. Excellent gain with acceptable noise figure was achieved. The parametric resonator should find applications in catheter and implanted coils where a wire connection is unfavorable or impossible. It can serve as a detector to couple to traveling wave MRI devices (16,17) when the imaging object is remotely excited by an antenna. It may also increase the sensitivity of specialized probes where the detection coil needs to reorient (18) or rotate (19) with the sample, making wire connections difficult to implement.

ACKNOWLEDGMENT

The authors thank Afonso Silva for carefully reading the article.

APPENDIX

The derivation below is an extension of theories outlined by Collin (5), and it includes the case where the circuit is tuned off-resonance. The triple frequency resonator (Fig. 2a) is a Foster network equivalent of the three-branch network described by Collin.

To understand how the NMR signal at frequency ω_1 is amplified and converted to a different frequency ω_2 , the current in the varactor is expanded as

$$i \approx \frac{dq(V_3)}{dt} + \frac{d}{dt} \left(\left(\frac{\partial q}{\partial V} \right)_{V=0} V \right) = \frac{dq(V_3)}{dt} + \frac{d}{dt} (C_2(t)V), \quad [\text{A1}]$$

where q is the charge accumulated in the varactor by a certain voltage, V_3 is the pumping voltage across the varactor at frequency ω_3 , and V is the sum of the much smaller signal voltages at the two lower resonance frequencies. If the pumping voltage is small compared with the threshold voltage of the varactor diode ($\sim 0.7 \text{ V}$), the time dependence of the capacitance can be expressed as $C_2(t) = C_2(1 + 2M \cos \omega_3 t)$, where M is the modulation index and ω_3 is the pumping frequency. The real signal voltages across the diode can be expressed in the complex forms

$$\begin{aligned} V_1 &= (V_s e^{j\omega_1 t} + V_s^* e^{-j\omega_1 t})/2 \\ V_2 &= (V_i e^{j\omega_2 t} + V_i^* e^{-j\omega_2 t})/2, \end{aligned}$$

where V_1 is the time-dependent voltage at the signal frequency ω_1 with a magnitude of V_s , and V_2 is the time-dependent voltage at the idler frequency ω_2 with a magnitude of V_i . By substituting V with $V_1 + V_2$ in Eq. A1, the current becomes

$$\begin{aligned} i &= \frac{dq(V_3)}{dt} + \frac{1}{2} \frac{d}{dt} C_2(1 + 2M \cos \omega_3 t) \\ &\quad \times (V_s e^{j\omega_1 t} + V_s^* e^{-j\omega_1 t} + V_i e^{j\omega_2 t} + V_i^* e^{-j\omega_2 t}). \quad [\text{A2}] \end{aligned}$$

The first term in the above equation is the current at the pumping frequency with no relation to the signal current, and it will be neglected in the following analysis. The current terms can similarly be expressed in the complex forms

$$\begin{aligned} I_1 &= (I_s e^{j\omega_1 t} + I_s^* e^{-j\omega_1 t})/2 \\ I_2 &= (I_i e^{j\omega_2 t} + I_i^* e^{-j\omega_2 t})/2, \end{aligned}$$

where I_1 is the time-dependent current at the signal frequency ω_1 with a magnitude of I_s , and I_2 is the time-dependent current at the idler frequency ω_2 with a magnitude of I_i . If the circuit is constructed to have a bandwidth narrow enough to reject all high-order frequency components, Eq. [A2] can be expanded as

$$\begin{aligned} I_s e^{j\omega_1 t} + I_s^* e^{-j\omega_1 t} + I_i e^{j\omega_2 t} + I_i^* e^{-j\omega_2 t} \\ = -jC_2 M V_i \omega_1 e^{-j\omega_1 t} + jC_2 M V_i^* \omega_1 e^{j\omega_1 t} + jC_2 V_s \omega_1 e^{j\omega_1 t} \\ - jC_2 V_s^* \omega_1 e^{-j\omega_1 t} - jC_2 M V_s \omega_2 e^{-j\omega_2 t} + jC_2 M V_s^* \omega_2 e^{j\omega_2 t} \\ + jC_2 V_i \omega_2 e^{j\omega_2 t} - jC_2 V_i^* \omega_2 e^{-j\omega_2 t}. \end{aligned}$$

These terms can be collected to form the admittance matrix

$$\begin{pmatrix} I_i^* \\ I_s \end{pmatrix} = \begin{pmatrix} -jC_2\omega_2 & -jC_2\omega_2 M \\ jC_2\omega_1 M & jC_2\omega_1 \end{pmatrix} \begin{pmatrix} V_i^* \\ V_s \end{pmatrix}. \quad [\text{A3}]$$

By inverting Eq. A3, we obtain the impedance matrix

$$\begin{pmatrix} V_i^* \\ V_s \end{pmatrix} = \frac{1}{1-M^2} \begin{pmatrix} -\frac{1}{j\omega_2 C_2} & -\frac{M}{j\omega_1 C_2} \\ \frac{M}{j\omega_2 C_2} & \frac{1}{j\omega_1 C_2} \end{pmatrix} \begin{pmatrix} I_i^* \\ I_s \end{pmatrix}. \quad [\text{A4}]$$

Applying Kirchhoff's Voltage Law to the complete circuit, the equivalent voltage source in series with the varactor at frequency ω_1 can be expressed as

$$\begin{aligned} \xi_1 &= V_s + I_s \left(\frac{j\omega_1 L_1}{1 - C_1 \omega_1^2 L_1} + \frac{j\omega_1 L_3}{1 - C_3 \omega_1^2 L_3} + R_1 + j\omega_1 L_2 \right) \\ &= I_s \left(\frac{j\omega_1 L_1}{1 - C_1 \omega_1^2 L_1} + \frac{j\omega_1 L_3}{1 - C_3 \omega_1^2 L_3} + R_1 + j\omega_1 L_2 \right. \\ &\quad \left. - \frac{j}{C_2(1-M^2)\omega_1} \right) - \frac{jMI_i^*}{C_2(1-M^2)\omega_2} \\ &\equiv I_s(R_1 + jX_1) - \frac{jMI_i^*}{C_2(1-M^2)\omega_2}. \end{aligned} \quad [\text{A5}]$$

$$I_s = -\frac{jC_2^2(1-M^2)^2\xi_1(X_2\omega_1+jR_2\omega_1)\omega_2}{-M^2+jC_2^2(1-M^2)^2(X_1R_2-X_2R_1)\omega_1\omega_2+C_2^2(1-M^2)^2(X_1X_2+R_1R_2)\omega_1\omega_2}.$$

The power gain at frequency ω_1 can be evaluated as

$$\begin{aligned} G_1 &= \frac{|I_s|^2 R_1}{\xi_1^2 R_1 / (R_1^2 + X_1^2)} \\ &= \frac{C_2^4(1-M^2)^4(X_1^2 + R_1^2)(X_2^2 + R_2^2)\omega_1^2\omega_2^2}{(-M^2 + C_2^2(1-M^2)^2(X_1X_2 + R_1R_2)\omega_1\omega_2)^2 + C_2^4(1-M^2)^4(X_1R_2 - X_2R_1)^2\omega_1^2\omega_2^2}. \end{aligned} \quad [\text{A7}]$$

For the general case where the reactance X_1 and X_2 are not equal to zero, both terms in the denominator of Eq. A7 need to be small to obtain high gain. As a result, the parametric amplifier should satisfy the frequency matching condition

$$\frac{X_1}{R_1} = \frac{2Q_1(\omega_1 - \omega_{10})}{\omega_{10}} \approx \frac{X_2}{R_2} = \frac{2Q_2(\omega_2 - \omega_{20})}{\omega_{20}} \equiv \kappa \quad [\text{A8}]$$

where Q_1 and Q_2 are quality factors of the resonator at frequencies ω_{10} and ω_{20} , respectively, and the resistance matching condition

$$R_N \equiv \frac{M^2}{C_2^2(1-M^2)^2\omega_1\omega_2R_2(1+\kappa^2)} \approx R_1 \quad [\text{A9}]$$

where most of the circuit resistance R_1 is cancelled by the negative resistance R_N . Equations A8 and A9 are summarized as Eq. 3 in the main text. The resonance frequencies, ω_{10} and ω_{20} , are fixed values determined by the resonator's tuning in the absence of the pumping signal. ω_1 is defined by the Larmor frequency of the

This voltage source is the signal induced by nuclei spin precession. $Z_1 = R_1 + jX_1$ is the effective impedance of the circuit loop that contains the varactor at frequency ω_1 . Similarly, the equivalent voltage source at frequency ω_2 can be written

$$\begin{aligned} \xi_2^* &= V_i^* + I_i^* \left(-\frac{j\omega_2 L_1}{1 - C_1 \omega_2^2 L_1} - \frac{j\omega_2 L_3}{1 - C_3 \omega_2^2 L_3} + R_2 - j\omega_2 L_2 \right) \\ &= I_i^* \left(-\frac{j\omega_2 L_1}{1 - C_1 \omega_2^2 L_1} - \frac{j\omega_2 L_3}{1 - C_3 \omega_2^2 L_3} + R_2 - j\omega_2 L_2 \right. \\ &\quad \left. + \frac{j}{C_2(1-M^2)\omega_2} \right) + \frac{jMI_s}{C_2(1-M^2)\omega_1} \\ &\equiv I_i^*(R_2 - jX_2) + \frac{jMI_s}{C_2(1-M^2)\omega_1}, \end{aligned} \quad [\text{A6}]$$

where $Z_2 = R_2 + jX_2$ is the impedance of the circuit loop that contains the varactor at frequency ω_2 . Equations A5 and A6 are similar to Eq. 11.34 of Collin's description (5). If the only source of voltage in the circuit is the signal from nuclear spins, then $\xi_2 = 0$ and I_s can be solved from Eq. A5 and Eq. A6 to be

nuclei of interest. Because $\omega_2 = \omega_3 - \omega_1$, the idler frequency ω_2 is a variable that can be adjusted to satisfy Eq. A8. In the presence of the pumping signal, ω_2 can be considered as the frequency at which the effective reactance of the circuit is zero. When the frequency matching condition of Eq. A8 is precisely met, the power gain G_1 in Eq. [A7] simplifies to

$$G_1 = \frac{R_1^2}{(R_1 - R_N)^2}. \quad [\text{A10}]$$

This expression can also be understood in terms of Joule's rule: given an input e.m.f., the power gain is obtained by squaring the ratio of currents that pass through the circuit resistance R_1 without and with pumping.

The "negative resistance" R_N defined in Eq. A9 is experimentally controlled by the modulation index M , and it can be estimated from the extent to which the resonator deviates from the oscillation condition. If we define $f_1(M) \equiv R_1 - R_N$, where R_N is a function of M , and if oscillation occurs when $f_1(M_0) = 0$, $f_1(M)$ at the

reduced pumping power level can be calculated from the Taylor expansion around the oscillation point

$$f_1(M) \approx f_1(M_0) + \frac{\partial f(M_0)}{\partial M}(M - M_0) \\ \approx \frac{-2M_0(M - M_0)}{C_2^2 R_2 \omega_1 \omega_2 (1 + A^2/R^2)} = \frac{2R_1(M_0 - M)}{M_0}. \quad [A11]$$

By solving for R_N based on Eq. A11, Eq. A10 can be expressed as

$$G_1 \approx \frac{1}{4} \left(\frac{M_0}{M_0 - M} \right)^2. \quad [A12]$$

Eq. A12 indicates that the power gain will become high as M approaches the oscillation value M_0 . This equation is summarized as Eq. 4 of the main text.

REFERENCES

1. Murphy-Boesch J, Koretsky AP. An in vivo NMR probe circuit for improved sensitivity. *J Magn Reson* 1983;54:526–532.
2. Schnall MD, Barlow C, Subramanian VH, Leigh JS. Wireless implanted magnetic-resonance probes for in vivo NMR. *J Magn Reson* 1986;68:161–167.
3. Atalar E, Bottomley PA, Ocali O, Correia LCL, Kelemen MD, Lima JAC, Zerhouni EA. High resolution intravascular MRI and MRS by using a catheter receiver coil. *Magn Reson Med* 1996;36:596–605.
4. Quick HH, Kuehl H, Kaiser G, Bosk S, Debatin JF, Ladd ME. Inductively coupled stent antennas in MRI. *Magn Reson Med* 2002;48: 781–790.
5. Collin RE. Foundations for microwave engineering. New York: McGraw-Hill; 1966. pp 541–570.
6. Uenohara M. Noise consideration of the variable capacitance parametric amplifier. *Proc Inst Radio Eng* 1960;48:169–179.
7. Berke HR. Solid State NMR Probe. U.S. Pat. 4,672,972 (Issued June 16, 1987).
8. Martius S, Heir O, Vester M, Biber S, Nistler J. Wireless Local Coil Signal Transmission Using a Parametric Upconverter. In: Proceedings of the 17th Annual Meeting of ISMRM, Honolulu, Hawaii, USA, 2009. p. 2934.
9. Cork P, Hulbert AP, Hunt J. Parametric Amplifier Device. U.S. Patent Application No. 12/753132, Publication No. 20100253349 (published Oct. 7, 2010).
10. Qian C, Murphy-Boesch J, Dodd S, Koretsky A. Integrated Detection, Amplification and Wireless transmission of MRI Signals Using a Parametric Amplifier. In: Proceedings of the 19th Annual Meeting of ISMRM, Montreal, Quebec, Canada, 2011. p. 626.
11. Sydoruk O, Shamonina E, Solymar L. Parametric amplification in coupled magnetoinductive waveguides. *J Phys D-Appl Phys* 2007;40: 6879–6887.
12. Syms RRA, Solymar L, Young IR. Three-frequency parametric amplification in magneto-inductive ring resonators. *Metamaterials* 2008;2: 122–134.
13. Wiltshire MCK, Pendry JB, Young IR, Larkman DJ, Gilderdale DJ, Hajnal JV. Microstructured magnetic materials for RF flux guides in magnetic resonance imaging. *Science* 2001;291:849–851.
14. Busch M, Vollmann W, Bertsch T, Wetzler R, Bornstedt A, Schnackenburg B, Schnorr J, Kivelitz D, Taupitz M, Gronemeyer D. On the heating of inductively coupled resonators (stents) during MRI examinations. *Magn Reson Med* 2005;54:775–782.
15. U.S. Department of Health and Human Services FDA, Center for Devices and Radiological Health. Guidance for the submission of premarket notifications for magnetic resonance diagnostic devices. Rockville, MD: US DHHS FDA; 1998.
16. Brunner DO, De Zanche N, Frohlich J, Paska J, Pruessmann KP. Travelling-wave nuclear magnetic resonance. *Nature* 2009;457: 994–992.
17. Webb AG, Collins CM, Versluis MJ, Kan HE, Smith NB. MRI and localized proton spectroscopy in human leg muscle at 7 Tesla using longitudinal traveling waves. *Magn Reson Med* 2010;63:297–302.
18. Qian CQ, Pines A, Martin RW. Design and construction of a contactless mobile RF coil for double resonance variable angle spinning NMR. *J Magn Reson* 2007;188:183–189.
19. Sakellaris D, Le Goff G, Jacquinot JF. High-resolution, high-sensitivity NMR of nanolitre anisotropic samples by coil spinning. *Nature* 2007;447:694–697.

QIC 890/891, Module 4: Microwave Parametric Amplification in Superconducting Qubit Readout experiments



**Instructor: Daryoush Shiri,
Postdoctoral fellow
IQC, University of Waterloo
IQC, June 2015
Session1, session 2**

Agenda

- ▶ Up converter amplifiers and Manley-Row relations
- ▶ Negative resistance amplifiers
- ▶ Travelling Wave (distributed) Parametric Amplifiers (TWPA)
- ▶ Parametric Phase Locked Oscillator (PPLO) or Parametron
- ▶ Quantum circuit theory of JPA:
- ▶ Circuit models of Josephson junction and SQUIDS
- ▶ Quantum analysis of parametric amplification
- ▶ Bandwidth enhancing schemes (Travelling wave JPA, Pumpistor)
- ▶ Application examples (two out of three examples will be covered):
 - ▶ Vacuum squeezing
 - ▶ Quantum feedback
 - ▶ PPLO for phase qubit readout

Electrical Systems ... History

3

- 1st proposed by Lord Rayleigh (*Phil. Mag. S. 5, 1883*):
 - Energy extracted from a source which suitably drives an energy storage element (K: stiffness of a spring) → Amplification of oscillation amplitude
- Aldert van der Ziel (*Journal of Applied Physics, vol. 19, 999, 1948*)
 - 1st Analysis of a nonlinear capacitance
- 1st Realizations by Weiss, Suhl, Hines, Adler (During 1957)
 - Using nonlinear effects in Ferrites
- Gain, Bandwidth (BW) and Noise of Parametric Amplifiers (PA)
 - H. Heffner, G. Wade, Stanford University, 1958
- **MAIN ADVANTAGE:**
 - **LOW NOISE** (No active or resistive element)

Historical Excursus

222

IRE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES

March

Parametric Devices and Masers: An Annotated Bibliography*

1960

E. MOUNT† AND B. BEGG‡

PREFACE

THIS bibliography is restricted to books and periodical articles published prior to October, 1959.

No attempt has been made to include the voluminous material to be found in technical reports, patents or similar sources.

Although the greatest portion of the bibliography

how it may be used in two solid-state amplifiers (Bloembergen's 3 energy-level type and Suhl's ferromagnetic type).

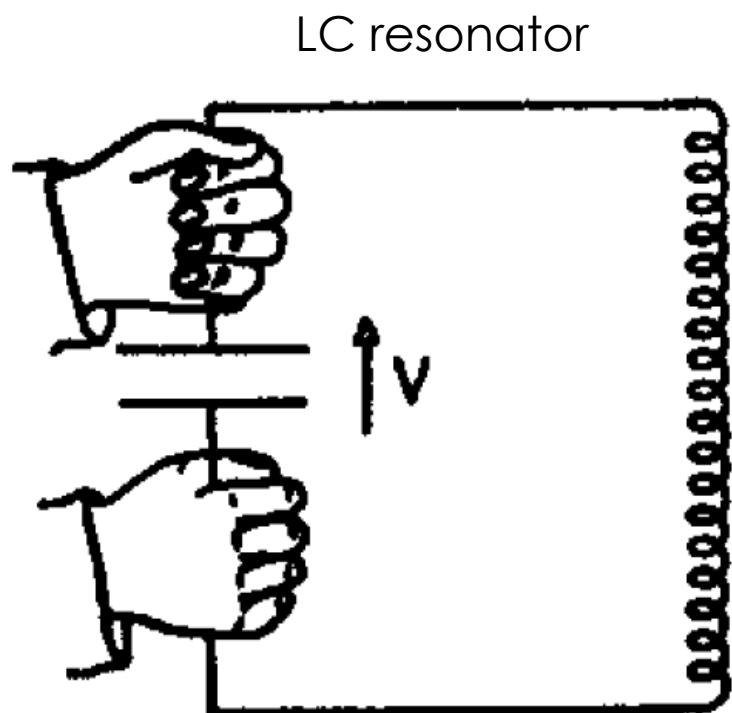
[2] Beam, R. E. and Brodwin, M. E.
Report of advances in microwave theory and techniques in U.S.A.—1958.

IRE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, MTT-7(3): 308-327; July, 1959.

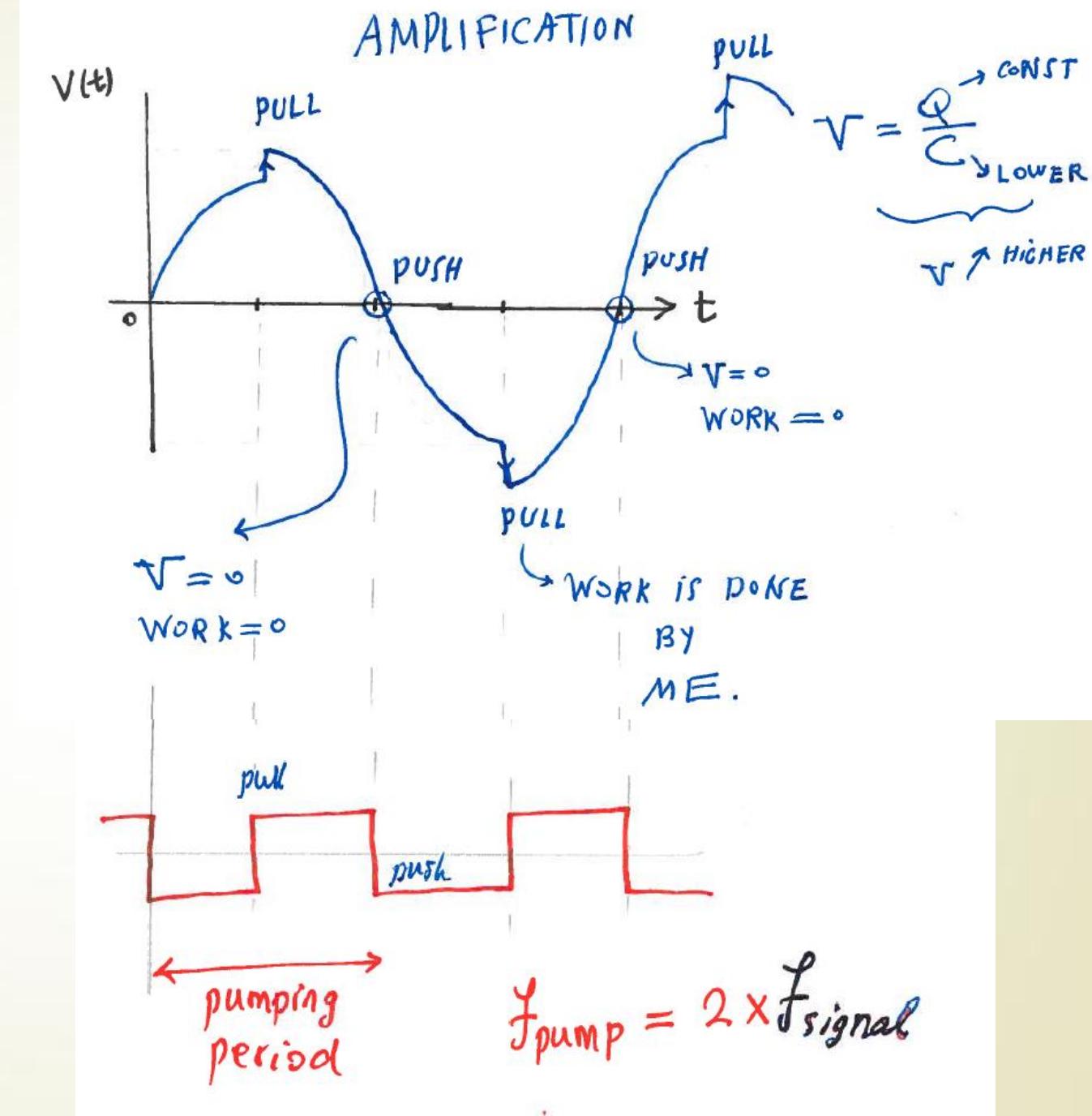
Includes a review of developments in masers and parametric devices, and cites 55 references of the U. S. publications in 1959.

5

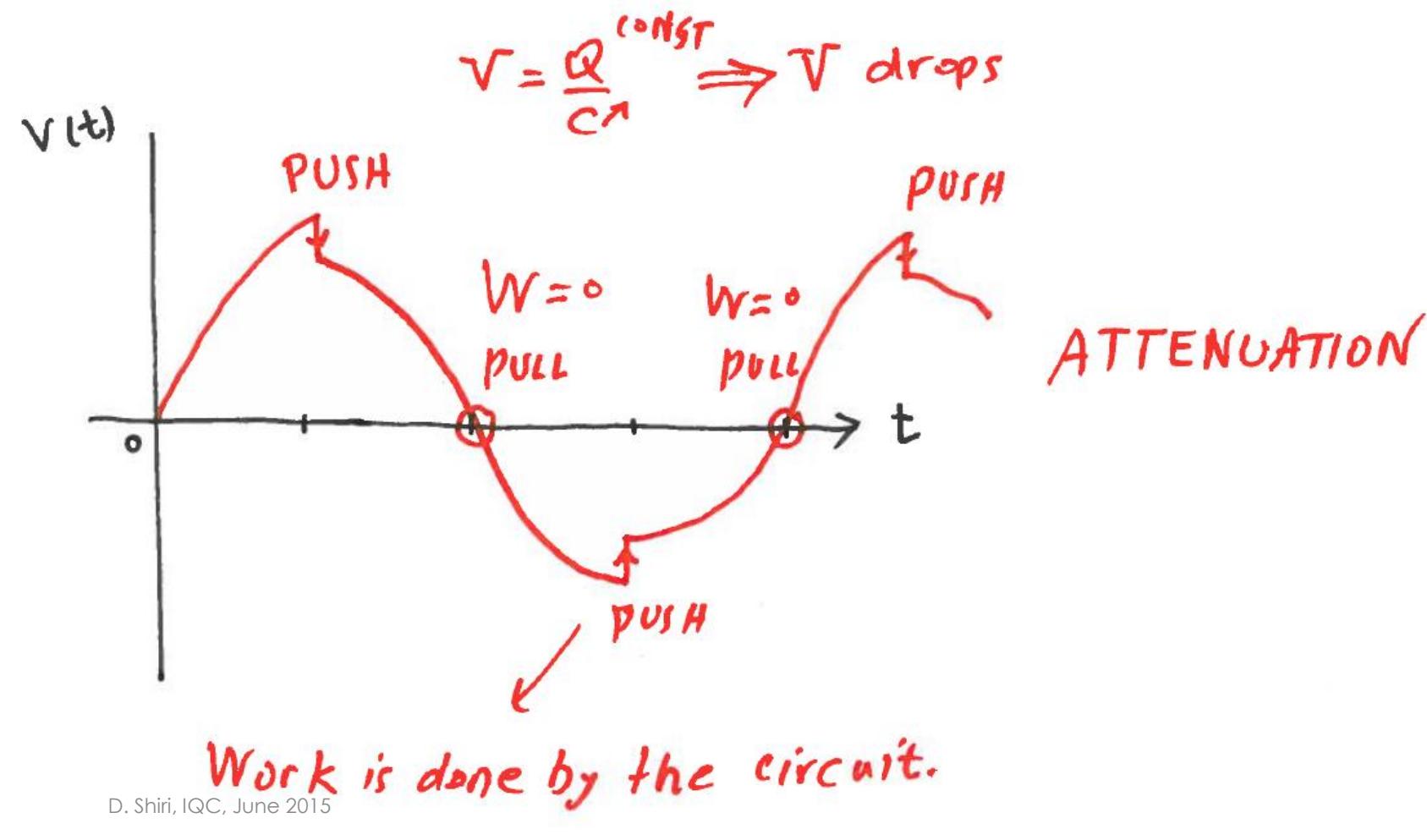
Amplification



D. Shiri, IQC, June 2015



Attenuation (De-amplification)

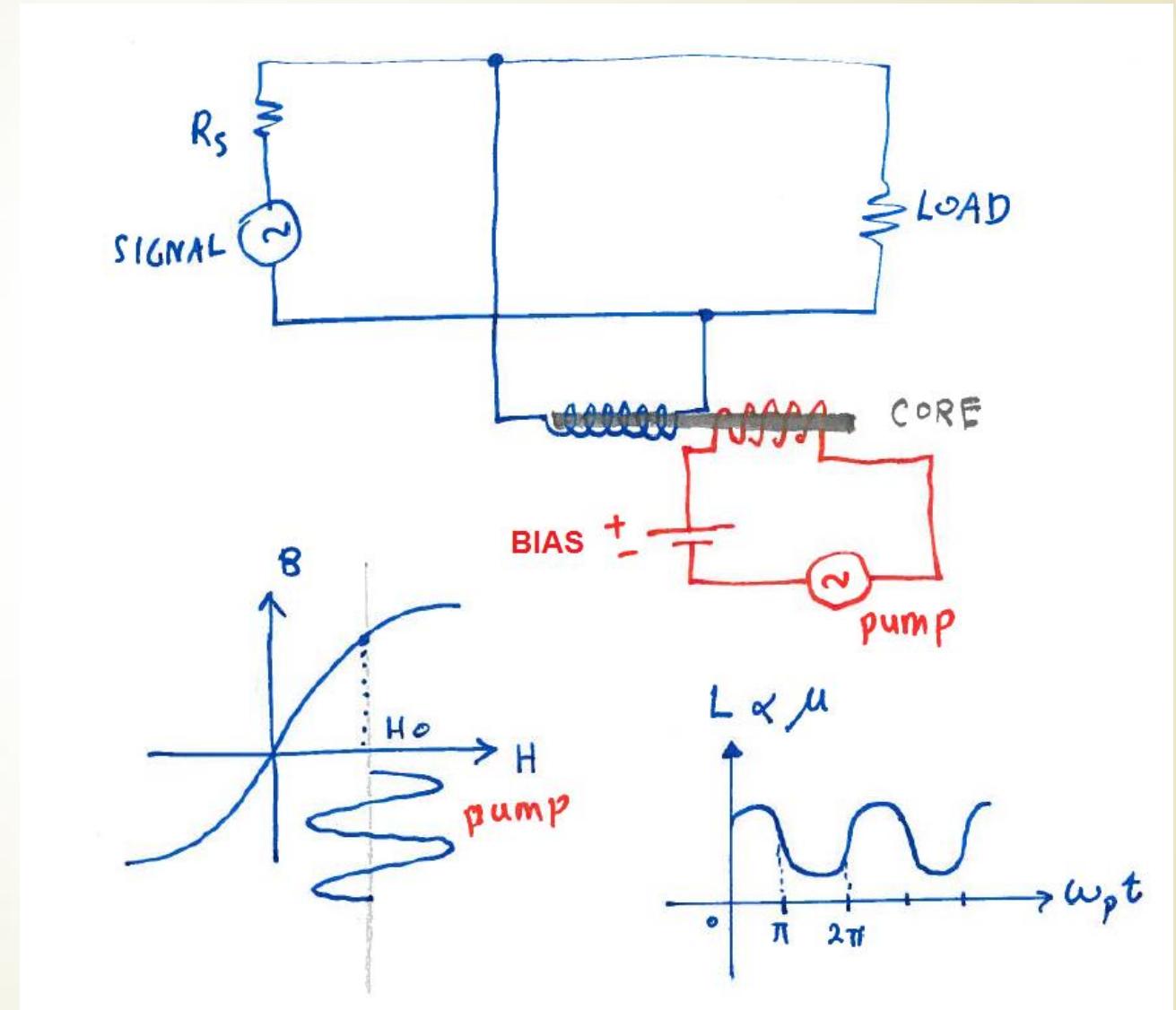
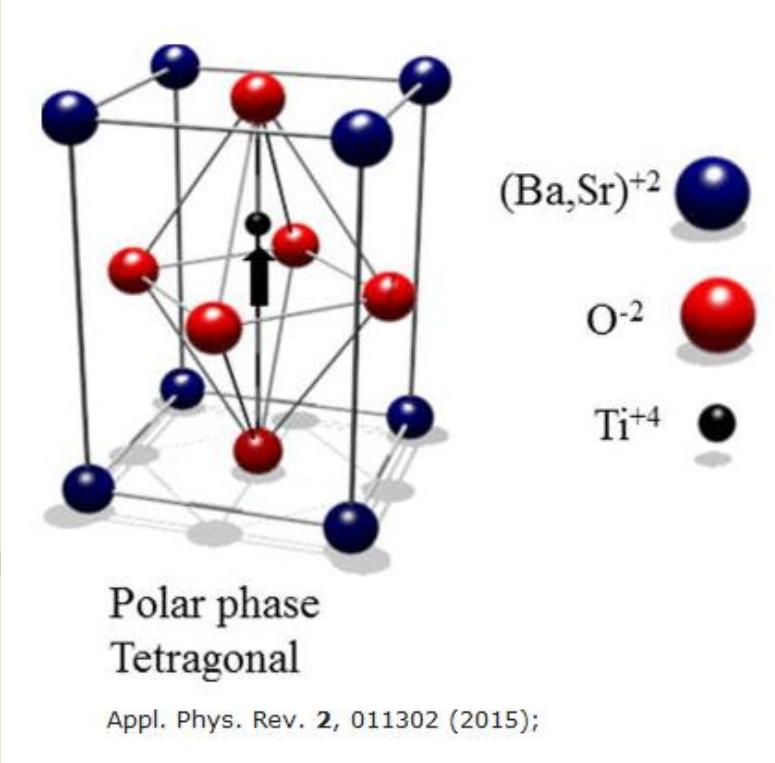


How to change the parameter?

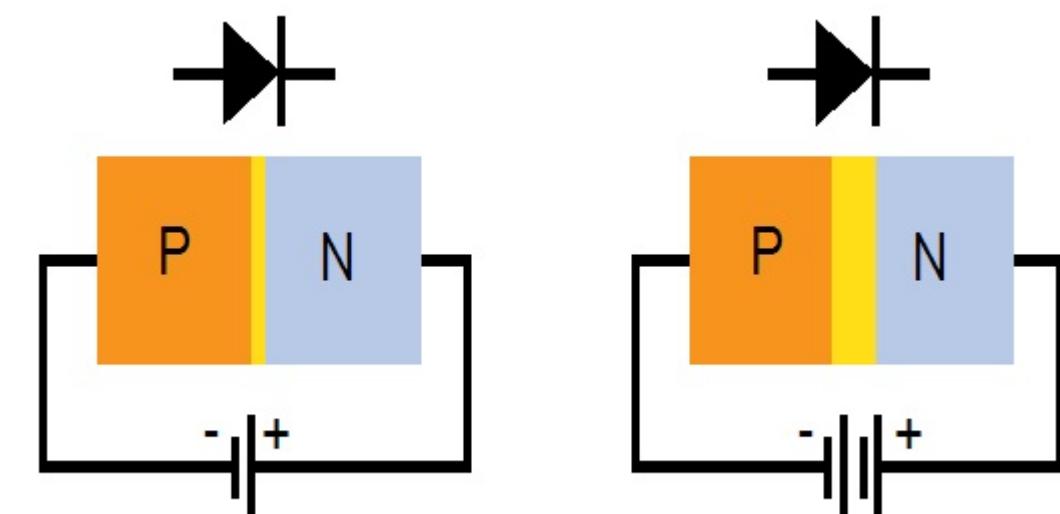
7

- ▶ Use a device to Push & Pull capacitor plates at right time
- ▶ $C(\text{capacitance}) = \epsilon \frac{A}{d}$
 - ▶ Ferroelectric: Change dielectric constant (ϵ) with applied voltage (BaTiO_3 crystal)
 - ▶ Change (d) using electric field e.g. VARACTOR
 - ▶ A reverse biased semiconductor pn junction diode
- ▶ $L(\text{inductance}) L = \mu . \text{geometry()}$
 - ▶ Ferrimagnetic e.g. BaFeO compounds as a core of a coil (change of μ permeability with current)
 - ▶ Use nonlinear inductance inherent in S-I-S junctions (Josephson junctions)

Material parameters



Varactor



$$C_J = \frac{C_{J0}}{\left(1 - \frac{V_A}{V_{bi}}\right)^{\frac{1}{m+2}}}$$

$m = 0$ abrupt junction

$m = 1$ graded junction

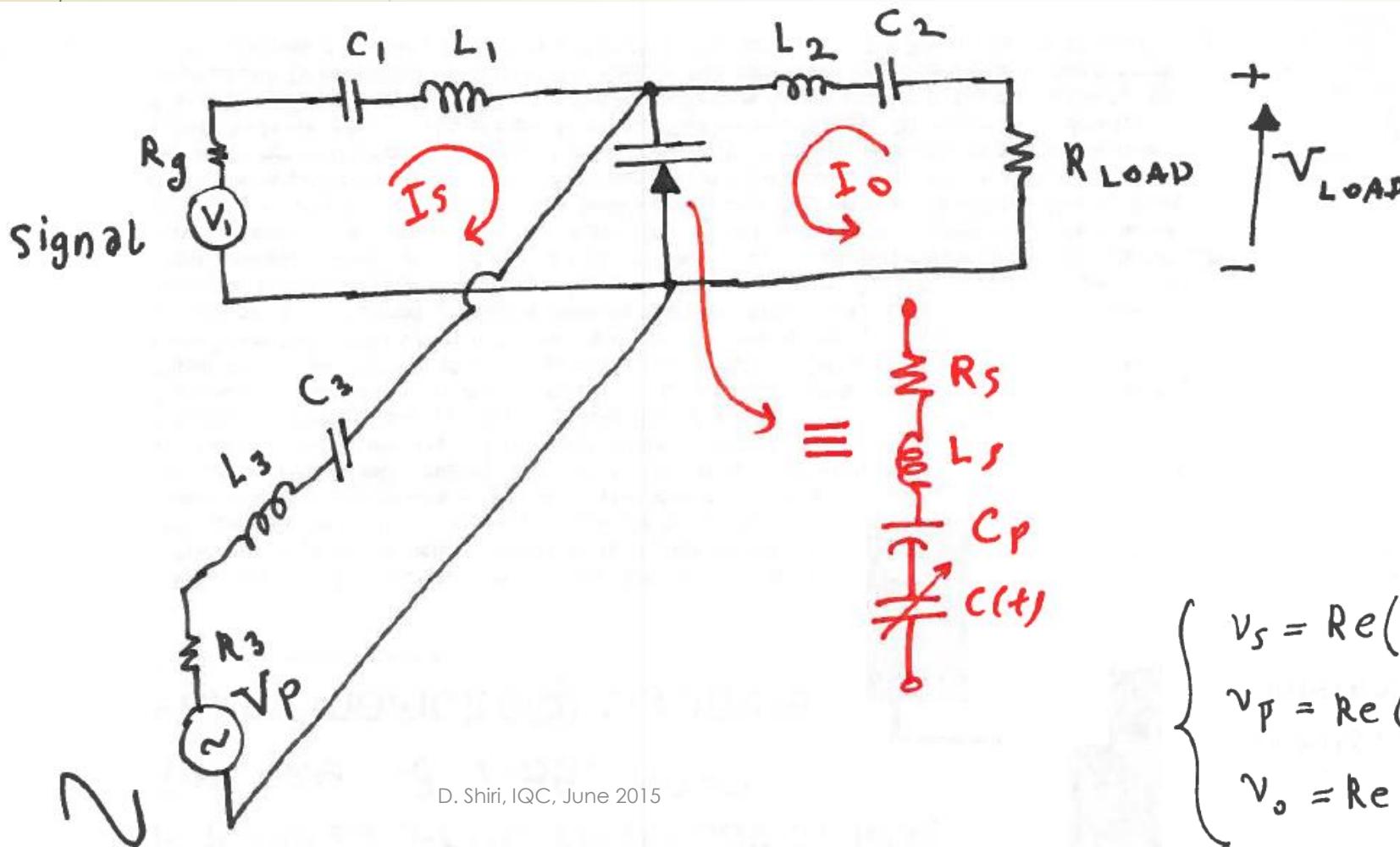
$$V(t) = V_0 \cos(\omega_{pump} t)$$

$$C(t) = \sum_n C_n \cos(n\omega_{pump} t)$$

(1) Parametric Up Converter

10

- ▶ Inject Signal at f_{signal} , Pump at f_{pump}
- ▶ Collect output voltage at $f_{\text{pump}} + f_{\text{signal}}$



D. Shiri, IQC, June 2015

$$\left\{ \begin{array}{l} v_s = \operatorname{Re}(V_s e^{j\omega_s t}) = \frac{1}{2} (V_s e^{j\omega_s t} + V_s^* e^{-j\omega_s t}) \\ v_p = \operatorname{Re}(V_p e^{j\omega_p t}) \\ v_o = \operatorname{Re}(V_o e^{j\omega_o t}) \end{array} \right.$$

(1) Capacitance Model



Modelling Capacitance Variation

$$\left\{ \begin{array}{l} C(t) = C_0 (1 + 2\gamma_1 \cos \omega_3 t) = C_0 + \gamma_1 C_0 [e^{j\omega_3 t} + e^{-j\omega_3 t}] \\ \textcircled{1} \quad \omega_3 = \omega_{\text{pump}} \\ \text{in some textbooks } \gamma_1 = M \end{array} \right.$$

C_0, γ_1 real quantities

$$i = \frac{dQ}{dt} = \frac{d}{dt} [C(t)V(t)]$$

D. Shirm, IQC, June 2015

$$(V_s e^{j\omega_3 t} + V_s^* e^{-j\omega_3 t} + V_o e^{j\omega_3 t} + V_o^* e^{-j\omega_3 t})$$

(1) Parametric Up Converter

12

$$i = \frac{dQ}{dt} = \frac{d}{dt} [C(t)V(t)]$$



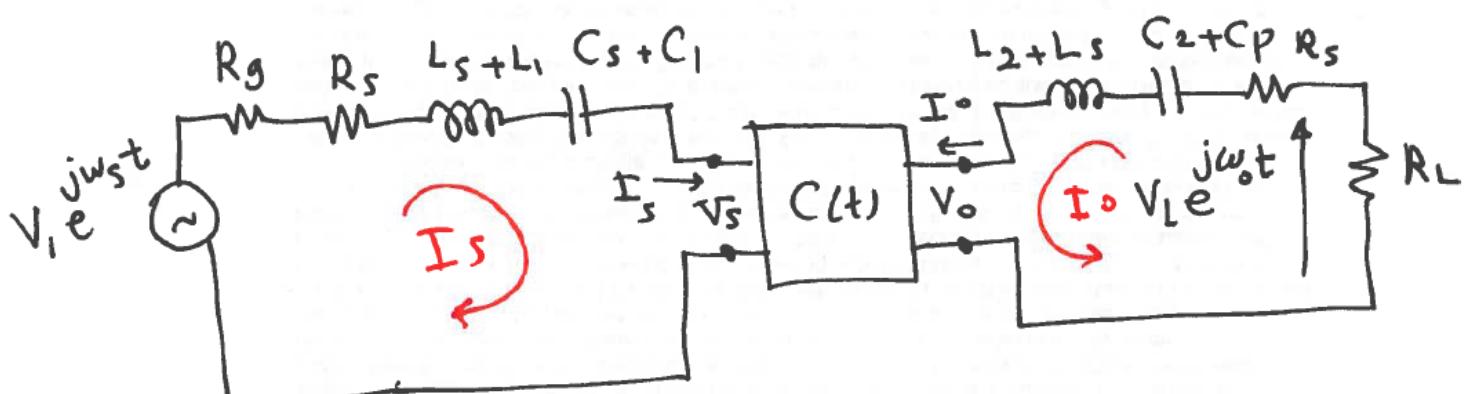
$$\begin{bmatrix} I_s \\ I_o \end{bmatrix} = \begin{pmatrix} j\omega_s C_0 & j\omega_s C_0 M \\ j\omega_o C_0 M & j\omega_o C_0 \end{pmatrix} \begin{pmatrix} V_s \\ V_o \end{pmatrix}$$

$\underbrace{\qquad\qquad\qquad}_{U}$

Current @ $\omega_s \rightarrow i_s = \frac{1}{2}(I_s e^{j\omega_s t} + I_s^* e^{-j\omega_s t})$

current @ $\omega_o \rightarrow i_o = \frac{1}{2}(I_o e^{j\omega_o t} + I_o^* e^{-j\omega_o t})$

$$\begin{bmatrix} V_s \\ V_o \end{bmatrix} = U^{-1} \begin{bmatrix} I_s \\ I_o \end{bmatrix}$$

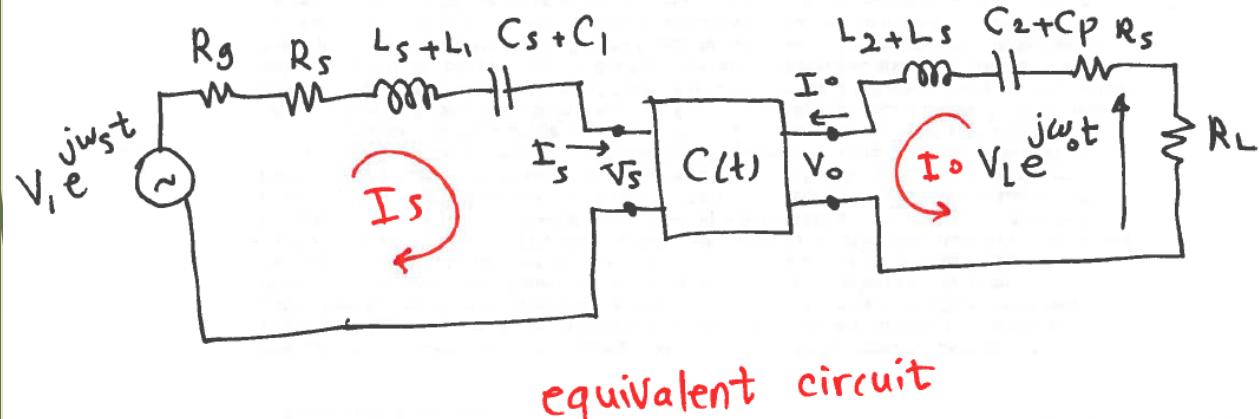


D. Shiri, IQC, June 2015

equivalent circuit

(1) Parametric Up Converter

13



$$\text{power gain} = \frac{|I_o|^2 R_L / 2}{V_i^2 / 2 R_g} = \frac{4 |I_o|^2 R_L R_g}{V_i^2}$$

$$G_{MAX} = \frac{\omega_o}{\omega_s} \frac{\delta}{(1 + \sqrt{1 + \delta})^2}$$

$$\delta = \left(\frac{\omega_s}{\omega_o} \right) (M Q)^2 \quad \text{gain-degradation factor}$$

effective quality factor of VARACTOR

$$R_s \rightarrow 0 \quad Q \rightarrow \infty \Rightarrow \delta \rightarrow \infty$$

$$G_{MAX} \rightarrow \frac{\omega_o}{\omega_s} = \frac{\omega_{pump} + \omega_{sig}}{\omega_{sig}}$$

Maximum power gain

$$= 1 + \frac{\omega_{pump}}{\omega_{sig}}$$

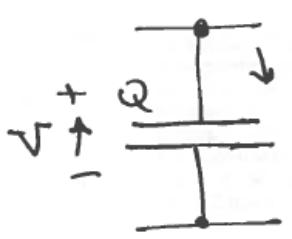
$$Q = \frac{1}{R_s \omega_s (1 - M^2) C_o}$$

It is difficult to increase this ☹

(1) MANLEY-ROW relations

14

- Conservation of energy (net average power going IN and OUT is ZERO)



$$I = \sum_{n,m=-\infty}^{+\infty} I_{n,m} e^{j(n\omega_1 + m\omega_2)t}$$

$$V = \sum_{n,m=-\infty}^{+\infty} V_{nm} e^{j(n\omega_1 + m\omega_2)t}$$

$$P_{nm} = V_{nm} I_{nm}^* + V_{nm}^* I_{nm}$$

$$\sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} P_{nm} = 0$$

$$\sum_{n,m=-\infty}^{+\infty} \frac{n\omega_1 + m\omega_2}{n\omega_1 + m\omega_2} P_{nm} = 0$$

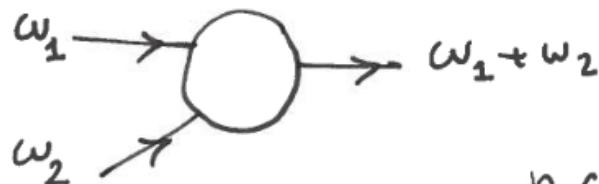
$$\underbrace{\omega_1 \sum_{nm} \frac{n P_{nm}}{n\omega_1 + m\omega_2}} + \underbrace{\omega_2 \sum_{nm} \frac{m P_{nm}}{n\omega_1 + m\omega_2}} = 0$$

$$ab + a^\dagger b^\dagger + ab^\dagger + ba^\dagger$$

Reminds us of Jaynes-Cummings Hamiltonian and Rotating-Wave Approximation (RWA)

(1) MANLEY-ROW ... parametric up converter

15



$$n \in \{0, 1\}$$
$$m \in \{0, 1\}$$

$$\left\{ \begin{array}{l} \frac{P_{10}}{\omega_1} + \frac{P_{11}}{\omega_1 + \omega_2} = 0 \\ \frac{P_{01}}{\omega_2} + \frac{P_{11}}{\omega_1 + \omega_2} = 0 \end{array} \right.$$

$$\text{power gain} = -\frac{P_{11}}{P_{10}} = \frac{\omega_1 + \omega_2}{\omega_1} = 1 + \frac{\omega_2}{\omega_1}$$

↑ pump
↑ signal

We obtained this before

D. Shiri, IQC, June 2015

Recall that we picked up sum of input frequencies at parametric up converter

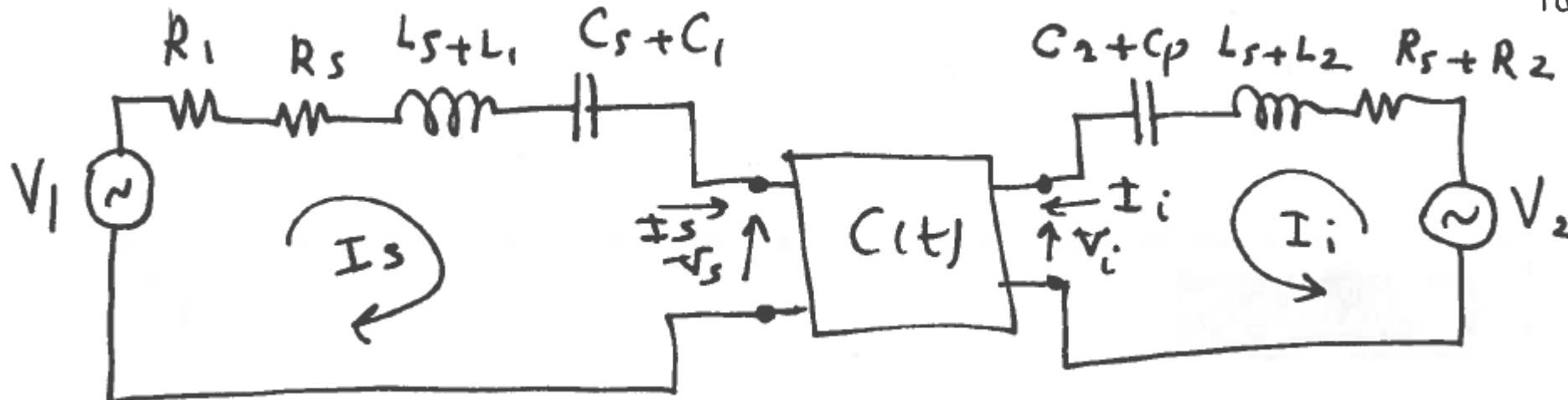
(2) Negative resistance PA

16

- ▶ Inject Signal at f_{signal} , Pump at f_{pump}
- ▶ Collect output voltage at $f_{\text{pump}} - f_{\text{signal}} = f_{\text{idler}}$ or f_{signal}

$$\omega_i = \omega_{\text{pump}} - \omega_{\text{signal}}$$

idler



$$\begin{pmatrix} I_i \\ I_s^* \end{pmatrix} = \underbrace{\begin{pmatrix} j\omega_i C_0 & j\omega_i C_0 M \\ -j\omega_s C_0 M & -j\omega_s C_0 \end{pmatrix}}_{\Omega} \begin{pmatrix} V_i \\ V_s^* \end{pmatrix}$$

Assume $V_2 = 0 \rightarrow \text{find } \underline{I_i} \rightarrow$

$$G = \frac{|I_i|^2 R_2 / 2}{|V_1|^2 / 2 R_1} = \frac{4 R_1 R_2 |\underline{I_i}|^2}{|V_1|^2}$$

@ $\omega_{\text{idler}} = \omega_p - \omega_s$

(2) Negative resistance PA, ... GAIN

17

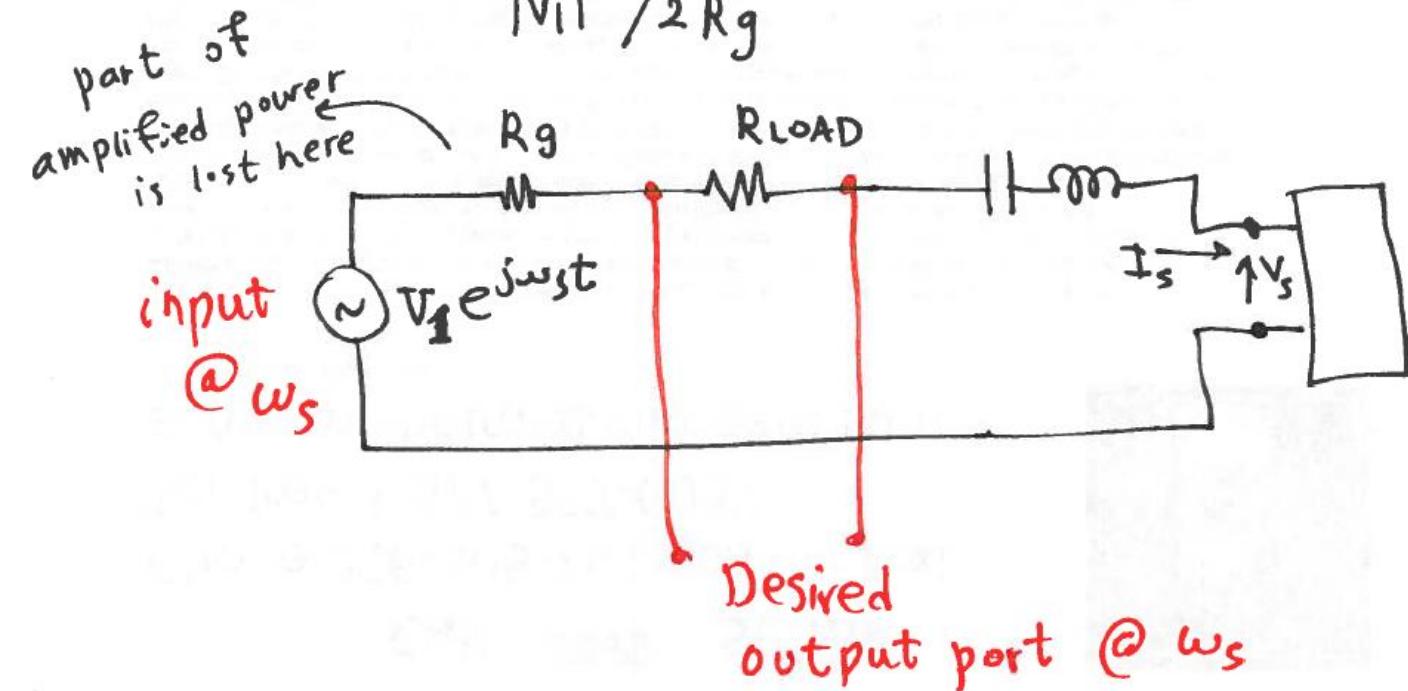
$$G_o = \frac{4R_1 R_2 [w_i (1 - M^2) C_0 \tilde{R}]^2}{M^2 (R_1 + R_s - \tilde{R})}$$

$$\tilde{R} = \frac{M^2}{(R_2 + R_s) w_i w_s (1 - M^2)^2 C_0^2}$$

$$G_o \rightarrow \infty \text{ if } \tilde{R} \rightarrow R_1 + R_s$$

if you like to amplify w_s

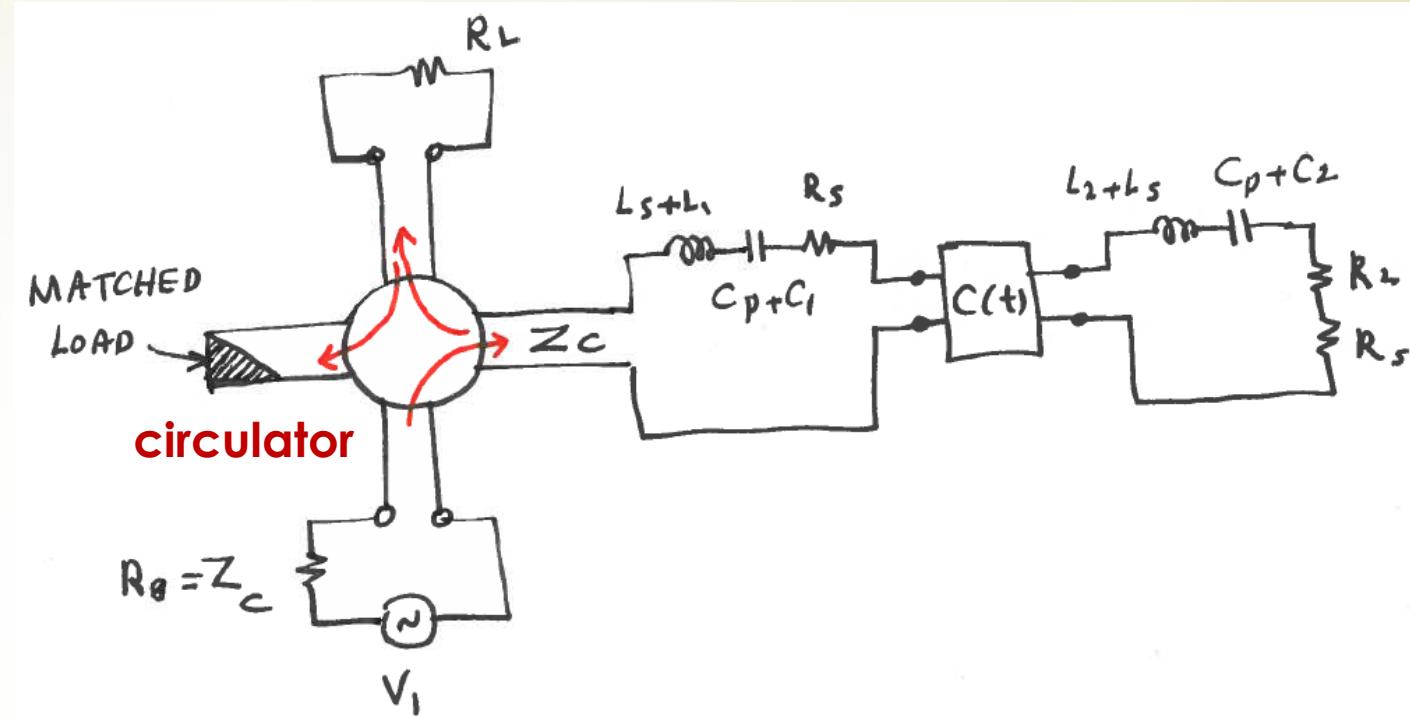
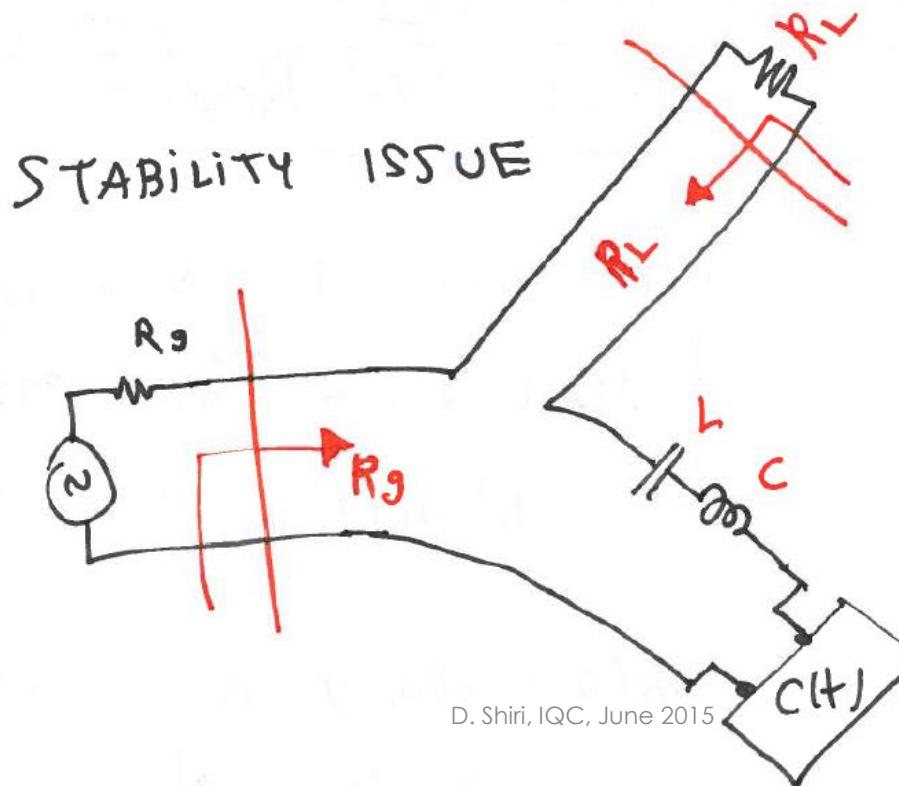
$$G = \frac{|I_s|^2 R_L / 2}{|V_i|^2 / 2 R_g}$$



(2) Negative resistance PA, ... Stability

18

$$G = \frac{4 R_g R_L}{(R_g + R_L + R_s - \tilde{R})}$$



- ☺ Power gain is **6 dB** higher since:
- ☺ Amplified power is completely diverted to R_L

(2) Negative resistance PA, ... Stability

19

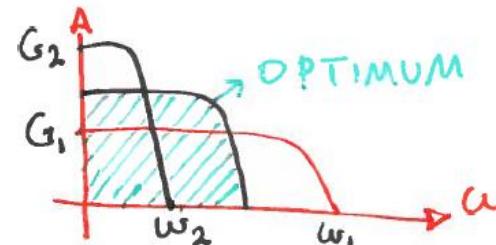
$$G_o = |\Gamma|^2 = \left| \frac{Z_{in} - Z_c}{Z_{in} + Z_c} \right|^2 = \frac{(R_s - R - R_L)^2}{(R_s + R_L - R)^2}$$

reflection coeff

$$\boxed{Z_c = R_L}$$
$$Z_{in} = R_s - \tilde{R}$$

@ resonance

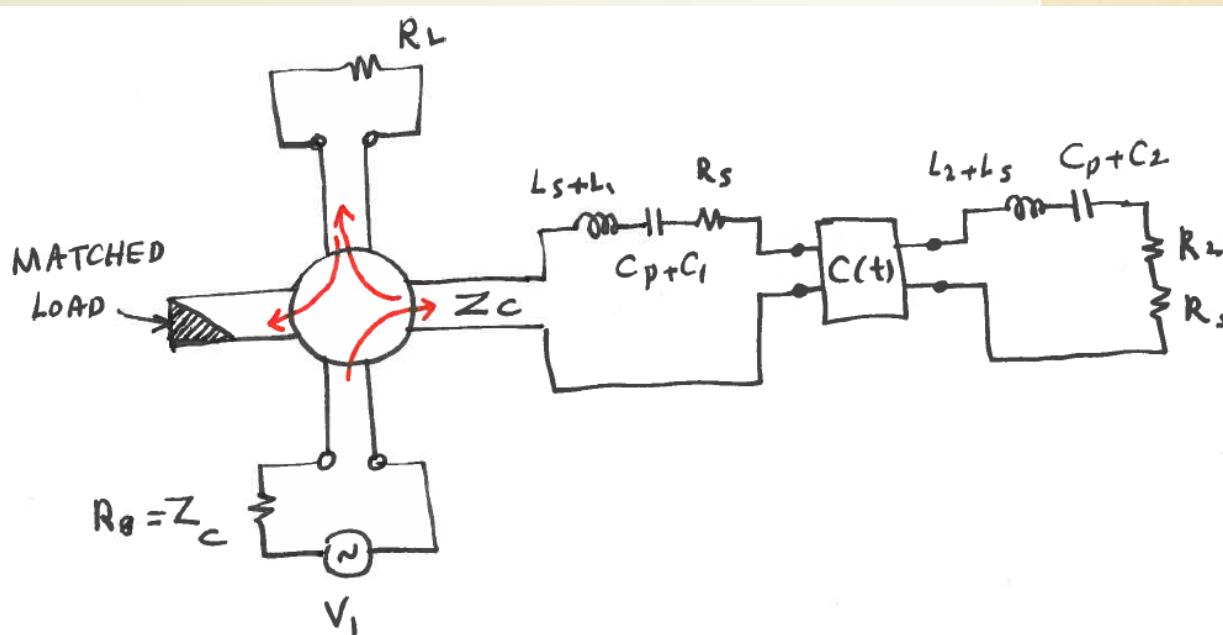
$$\underset{\text{MAX}}{(gain \cdot \text{BandWidth})} = \frac{M}{2} \sqrt{\frac{\omega_i}{\omega_s}}$$



THIS IS NOT SO GOOD.

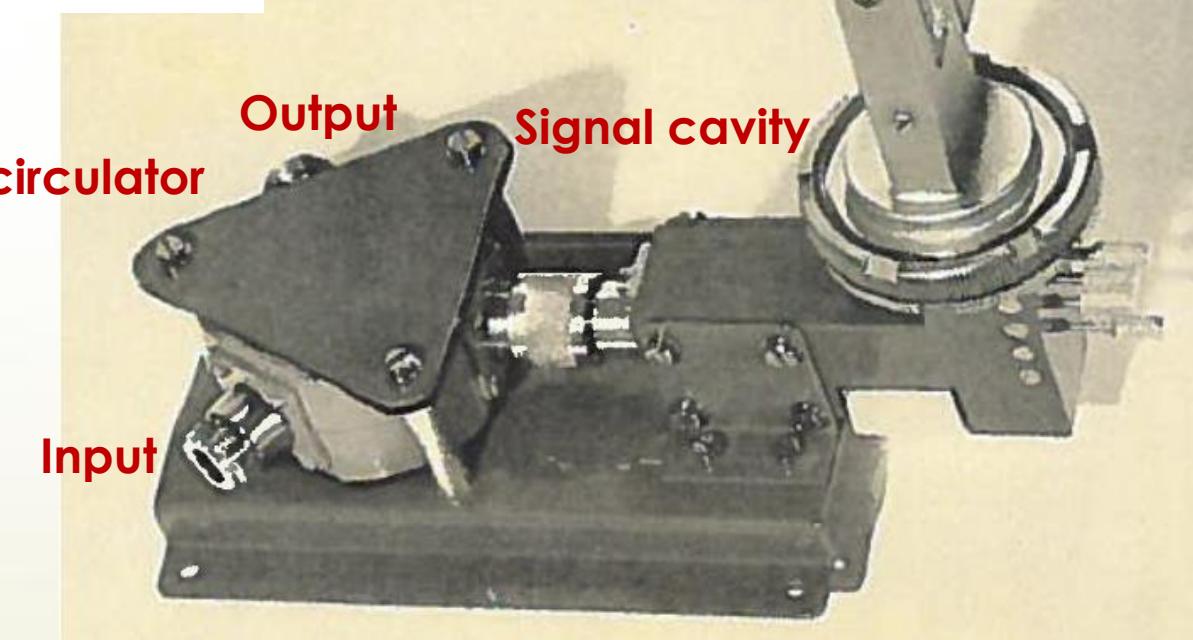
Implementation Example

20



$f_{\text{signal}} = 3 \text{ GHz}$, $f_{\text{pump}} = 10 \text{ GHz}$

Pump generator:
Klystron

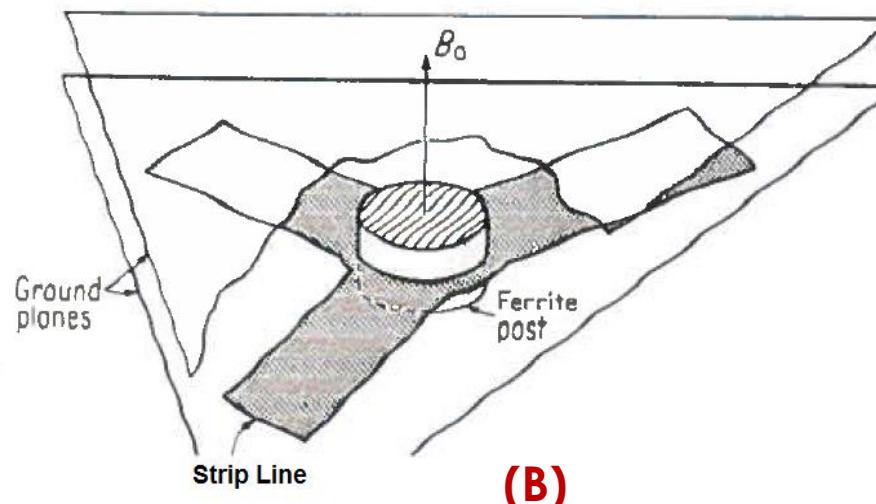
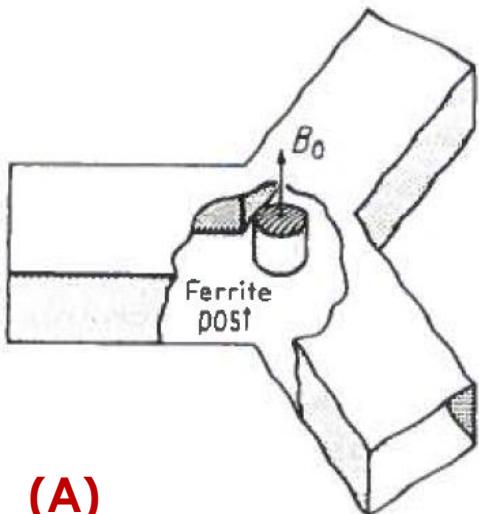


REF: D. G. Tucker, Circuits with periodically-varying parameters, D. Van Nostrand Inc. 1964.
D. Shiri, IQC, June 2015

A few words about Circulators

21

- They work based on **Faraday Rotation**
- Lack of **Reciprocity** (due to magnetic field, B)
- Breaking of **time reversal symmetry** with B
- An analogy with **Hall Effect**



(A)

(B)

D. Shiri, IQC, June 2015

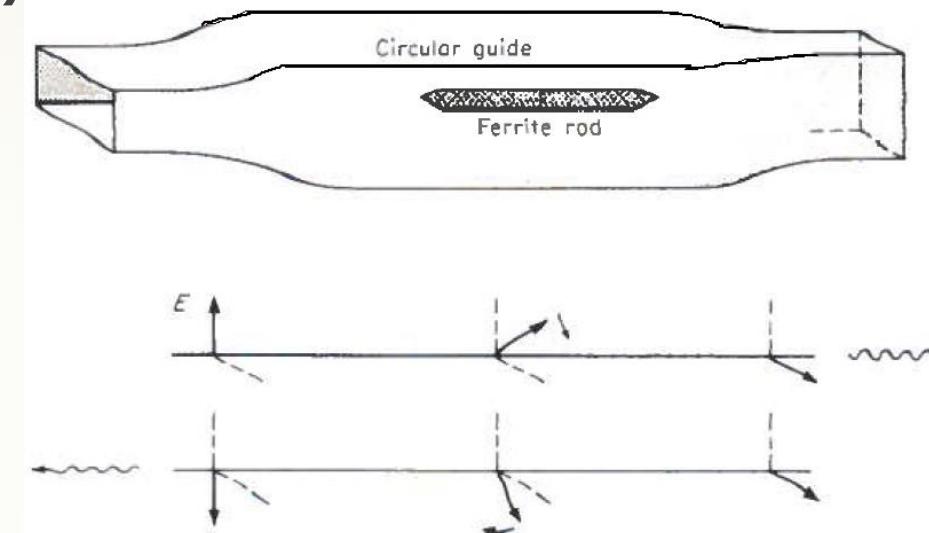
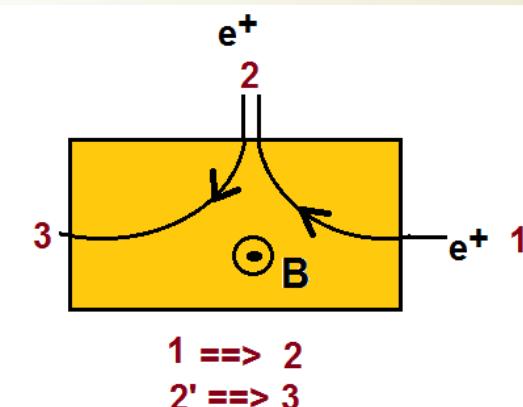


FIGURE 6.46
A gyrator without a twist section.



(2) Negative resistance PA, a sample

22

► $f_{\text{pump}} = 9.2 \text{ GHz}$, $f_{\text{idler}} = 7.9 \text{ GHz}$ (Ref: Robert E. Collin's book, 2nd Edition)

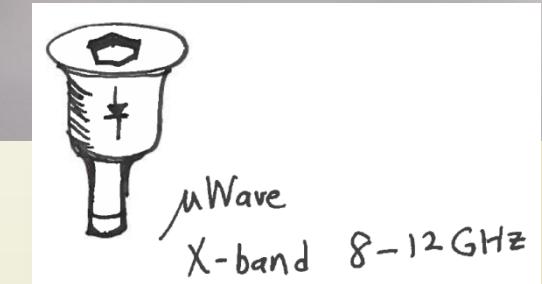
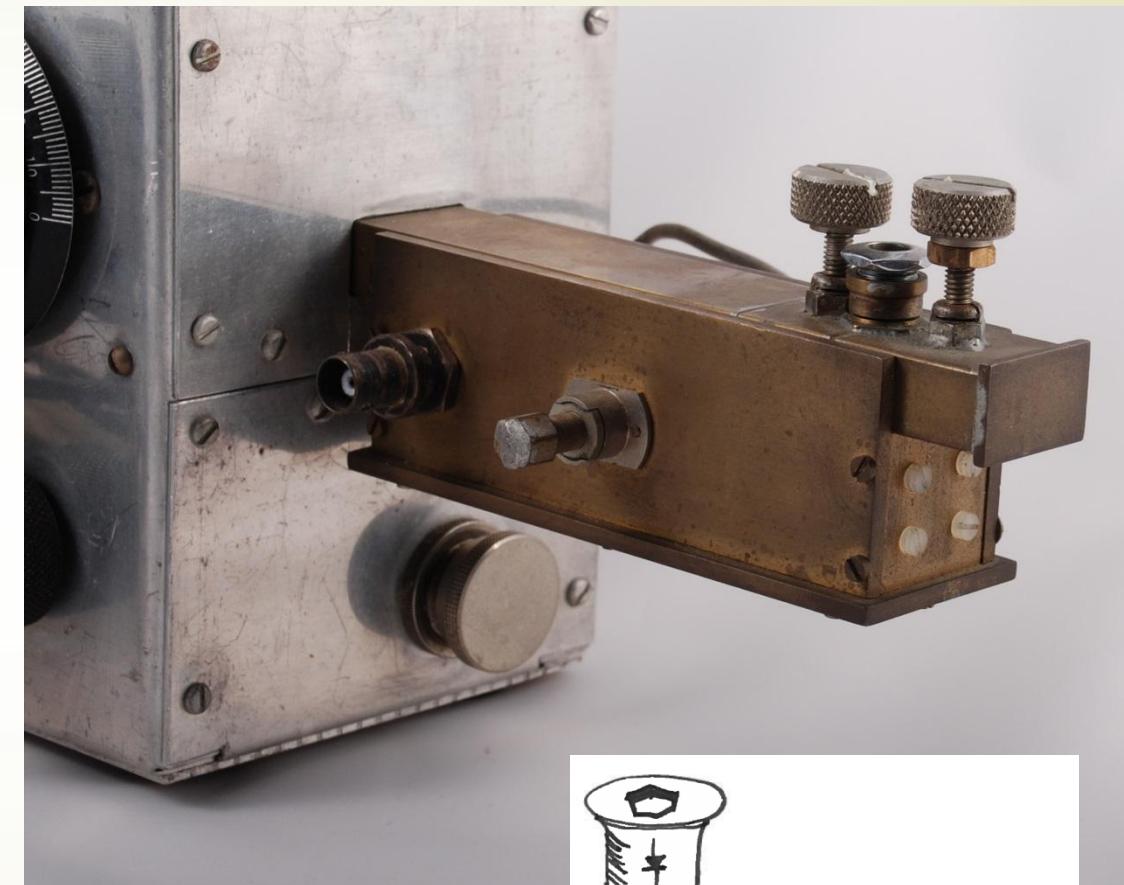
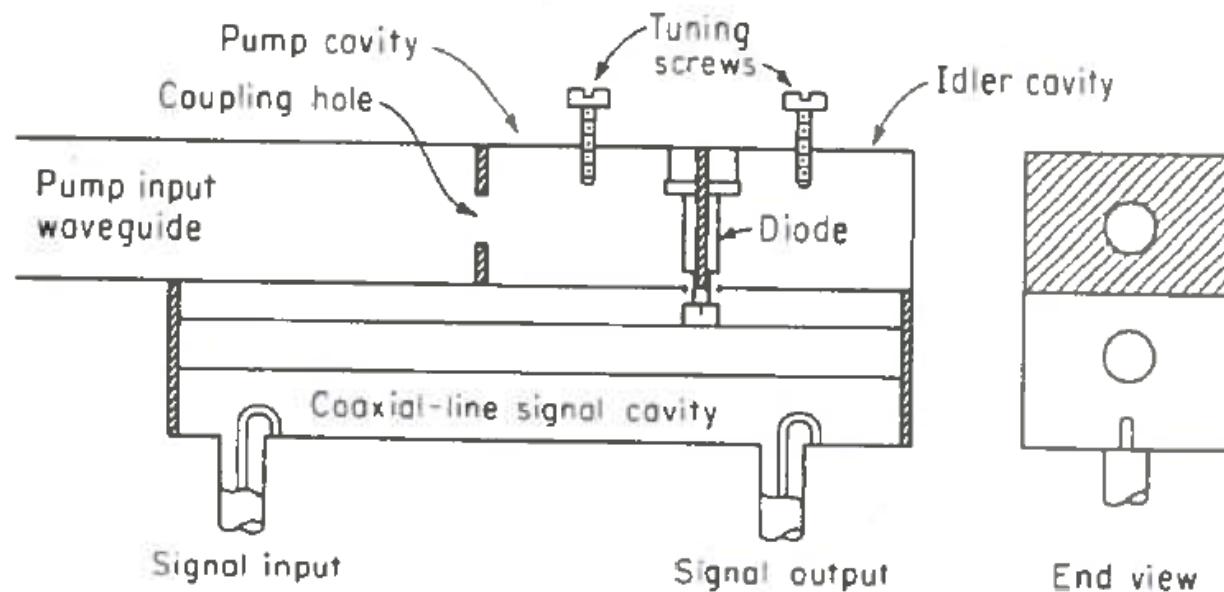


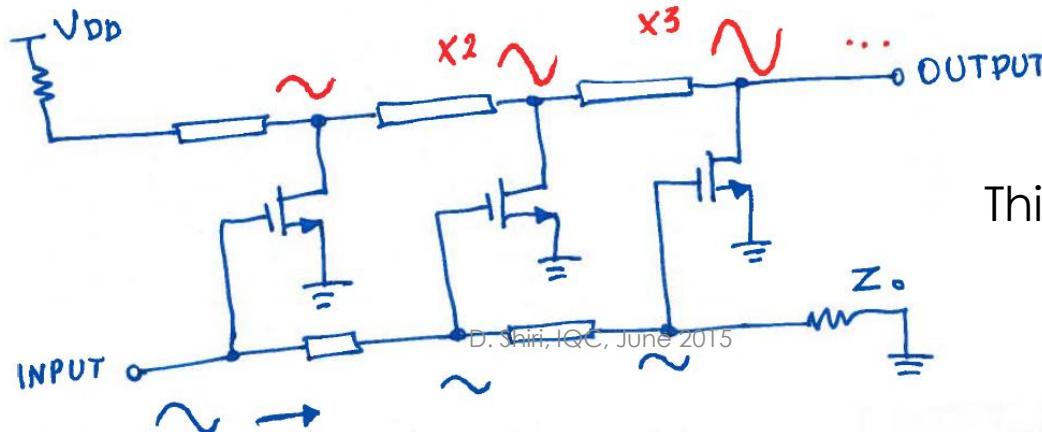
FIGURE 11.9
A microwave negative-resistance parametric amplifier.

Travelling Wave Amplifiers (TWA)

23

- Invented during 1939-1948 by:
 - H. A. Wheeler, Ginzton, Pierce, **Hewlett**, ...
 - H. W. Bode, Kompfner, et al
- **Disadvantages** of PA:
 - Limited Gain.BW product
 - Negative Resistance Amplifier
 - (stability issues of being one-port)
 - Limited $\omega_{\text{pump}}/\omega_{\text{signal}}$

AMPLIFICATION WHILE TRAVELLING



This has active device → NOISY

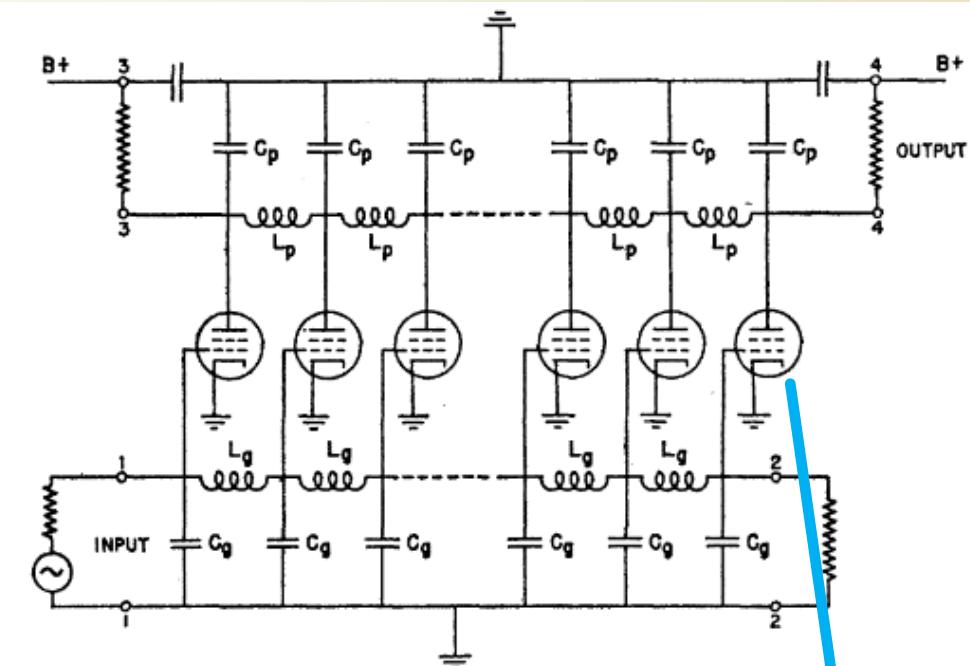


Fig. 1—Basic distributed amplifier.

Ref: Ginzton et al, Proceedings of the IRE, v.36, 8, 956 (1948).

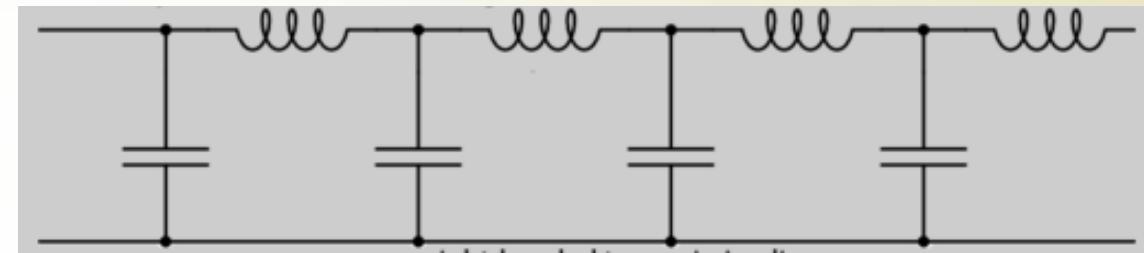


Travelling Wave parametric Amplifiers (TWA)

... Analysis

- ▶ Use parametric varying (L or C of a transmission line)
- ▶ Proposed by A. L. Cullen, **Half a page paper** in Nature, vol.181, 332 (1958).

$$L = L_0(1 + \gamma \sin(2\omega t - 2\beta x))$$



$$i = i_0 \exp(\alpha x) \sin(\omega t - \beta x + \varphi)$$

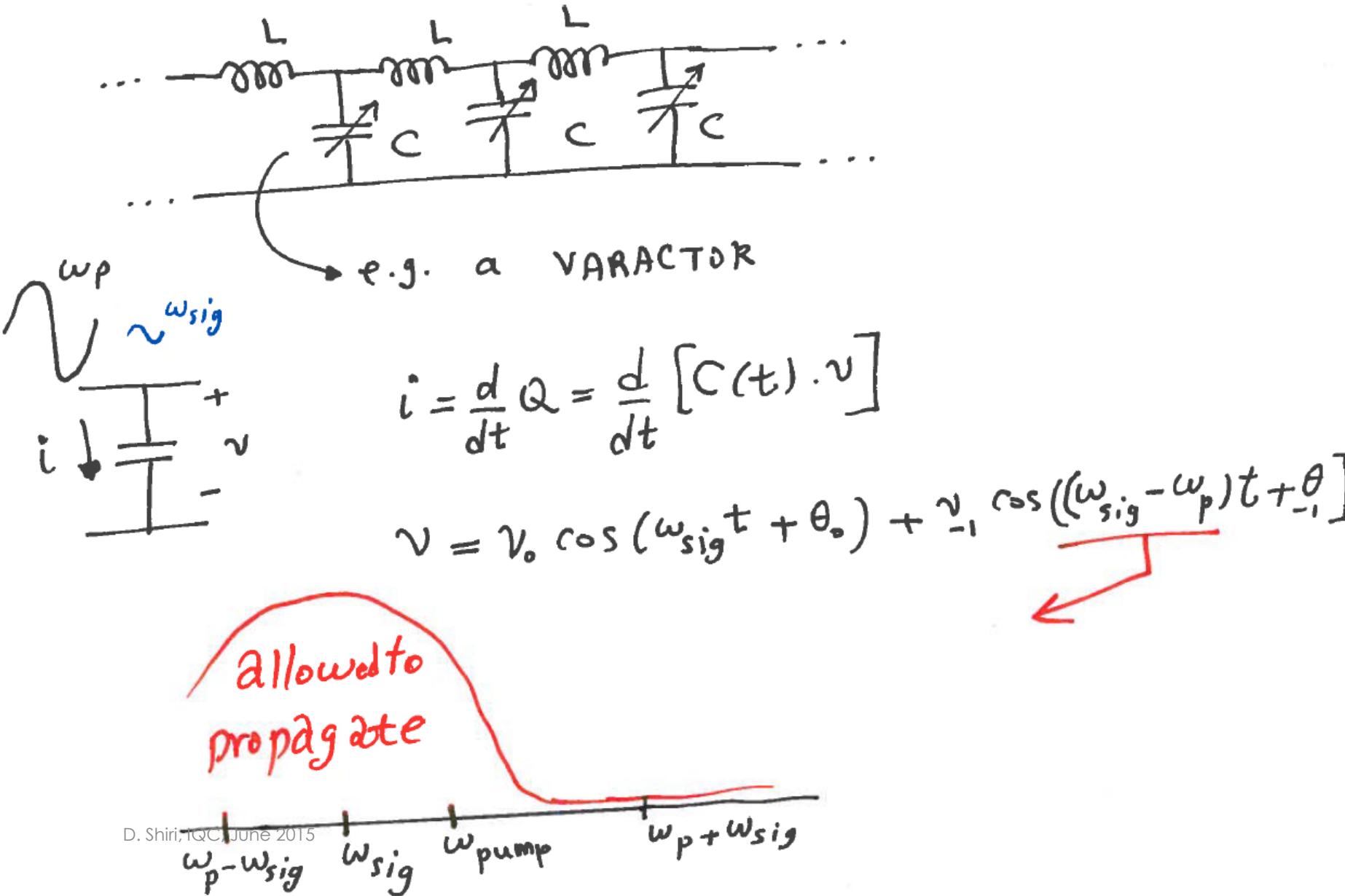
$$\alpha = \omega \frac{\gamma}{4} \cos(2\varphi) \sqrt{L_0 C_0}$$

$$\text{if } \varphi = \frac{\pi}{2} \implies \alpha = \omega \frac{\gamma}{4} \sqrt{L_0 C_0}$$

Let's learn the analysis done
by D. G. Tucker,
University of Birmingham, 1964

TWPA ... Analysis

25



TWPA ... Analysis

26

Since pump is a LARGE signal

$$C(t) = C_0 + C_1 \cos \omega_p t$$

Using phasor notation

$$@ \omega_{sig} \quad i_o = j\omega_{sig} C_0 v_o + \frac{1}{2} j\omega_{sig} C_1 v_{-1}$$

$$@ \omega_{sig} - \omega_{pump} \quad i_{-1} = j(\omega_{sig} - \omega_p) C_0 v_{-1} + \frac{1}{2} j(\omega_{sig} - \omega_p) C_1 v_o$$

TWPA ... Analysis

27

Admittance of $C(t)$ @ ω_{sig}

$$Y_0 = \frac{i_o}{v_o} = j\omega_{\text{sig}}C_0 + \frac{1}{2} j\omega_{\text{sig}}C_1 \left(\frac{v_{-1}}{v_o} \right)$$

Admittance of $C(t)$ @ $\omega_{\text{sig}} - \omega_{\text{pump}}$

$$Y_{-1} = \frac{i_{-1}}{v_{-1}} = j(\omega_s - \omega_p)C_0 + \frac{1}{2} j(\omega_s - \omega_p)C_1 \left(\frac{v_o}{v_{-1}} \right)$$

We tune the phase so that

v_{-1} is $\frac{\pi}{2}$ ahead of v_o

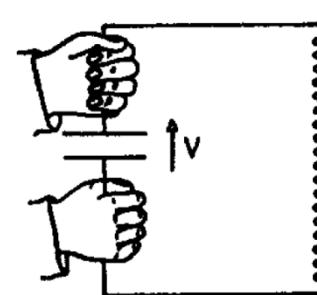


i.e. $\frac{v_{-1}}{v_o} = j \left| \frac{v_{-1}}{v_o} \right|$

$$\frac{v_o}{v_{-1}} = -j \left| \frac{v_o}{v_{-1}} \right|$$

D. Shiri, IQC, June 2015

Recall This experiment



TWPA ... Analysis

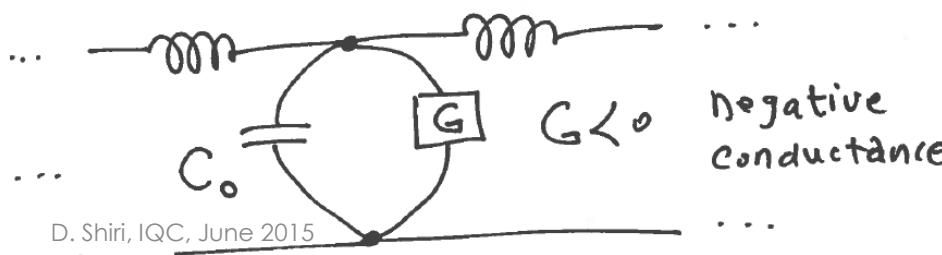
28

$$\left\{ \begin{array}{l} Y_0 = -\frac{1}{2} \omega_{sig} C_1 \left| \frac{V_1}{V_0} \right\rangle + j \omega_{sig} C_0 \\ Y_{-1} = \frac{1}{2} (\underbrace{\omega_{sig} - \omega_p}_{\text{Note that } \omega_{sig} < \omega_{pump}}) C_1 \left| \frac{V_0}{V_{-1}} \right\rangle + j (\underbrace{\omega_{sig} - \omega_p}_{\text{Note that } \omega_{sig} < \omega_{pump}}) C_0 \end{array} \right.$$

Note that $\omega_{sig} < \omega_{pump}$

$$\boxed{\omega_{sig} - \omega_p < 0}$$

Equivalent circuit of a section



D. Shiri, IQC, June 2015

TWPA ... Analysis (GAIN)

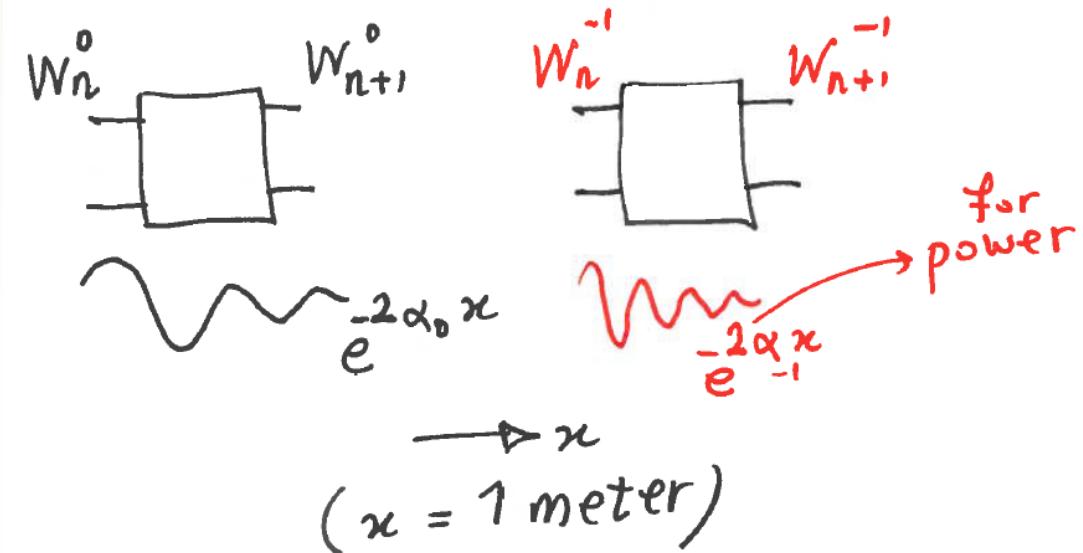
29

Manley-Row relations:

$$n\omega_{\text{sig}} + m\omega_{\text{pump}} \quad \frac{W_0}{\omega_{\text{sig}}} + \frac{W_{-1}}{\omega_{\text{sig}} - \omega_p} = 0$$

$\rightarrow n = 1$

$$\boxed{\frac{W_0}{\omega_{\text{sig}}} = \frac{W_{-1}}{\omega_{\text{pump}} - \omega_{\text{sig}}}}$$



$$\frac{W_n^0}{W_n^{-1}} = \frac{W_{n+1}^0}{W_{n+1}^{-1}} = \frac{\omega_{\text{sig}}}{\omega_{\text{pump}} - \omega_{\text{sig}}}$$

\Rightarrow

$$\alpha_0 = \alpha_{-1} = \alpha$$

TWPA ... Analysis (GAIN)

30

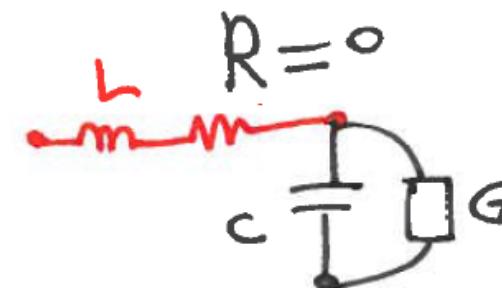
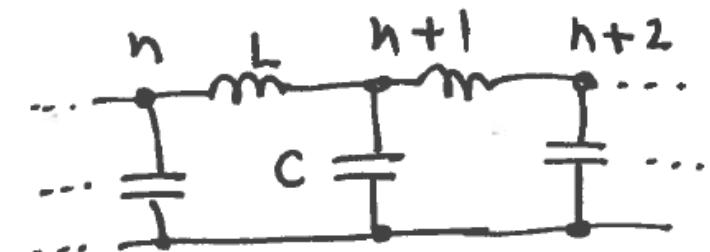
FIND α :

Recall from Pozar's book:

$$\text{Voltage} \rightarrow \frac{V_{n+1}}{V_n} = e^P = e^{\alpha + j\beta}$$

{ α : attenuation
 β : propagation

$$P = \sqrt{(R+j\omega L)(G+j\omega C)}$$



TWPA ... Analysis (GAIN)

31

$$@ \omega_{sig} \rightarrow P_o = j\omega_{sig} \sqrt{L C_0} \sqrt{1 + j \frac{1}{2} \frac{C_1}{C_0} \left| \frac{v_o}{v_{-1}} \right|}$$

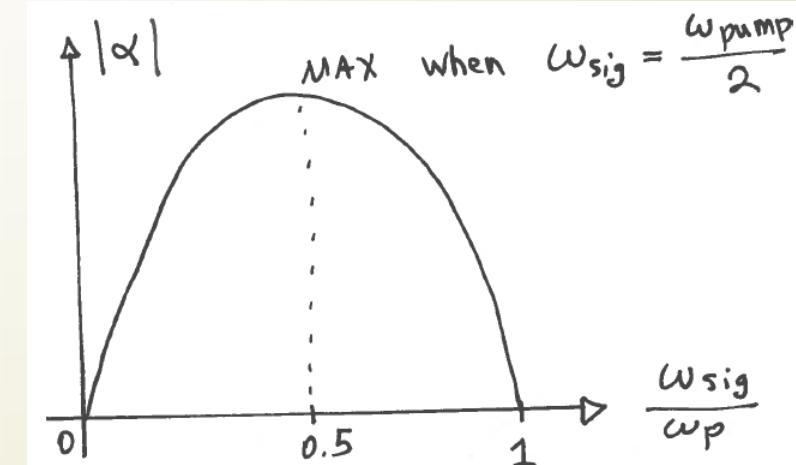
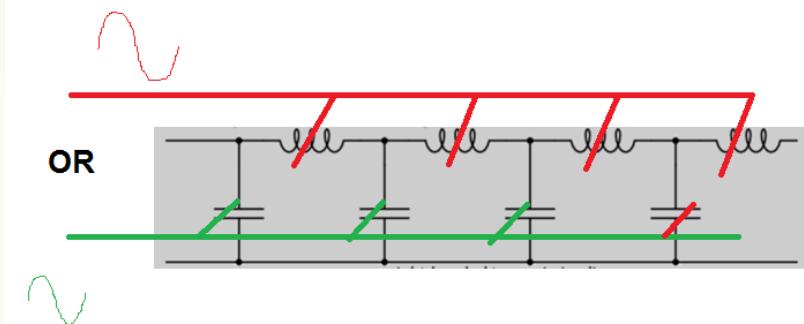
$$@ \omega_p - \omega_{sig} \rightarrow P_{-1} = j(\omega_p - \omega_{sig}) \sqrt{L C_0} \sqrt{1 + j \frac{1}{2} \frac{C_1}{C_0} \left| \frac{v_o}{v_{-1}} \right|}$$

$$\text{real}(P_o) = \text{real}(P_{-1}) = \alpha$$

$$\alpha = -\frac{1}{4} \sqrt{\omega_{sig} (\omega_p - \omega_{sig})} \cdot \sqrt{\frac{L}{C_0} \cdot C_1}$$

GAIN Z_0

D. Shiri, IQC, June 2015



Amplifiers-Basics



Modern Amplifier Terms Defined

1 dB compression point defines the output level at which the amplifier's gain is 1 dB less than the small signal gain, or is compressed by 1 dB.

Conditionally stable amplifier refers to an amplifier which will oscillate under particular load or source impedance (VSWR) conditions, an undesirable situation.

Cp (Process capability) Process capability broadly defined as the specification width (S) divided by the Process width (P) and is an indication of the spread of the process. Specification width "S" is the difference between the upper specification limit (USL) and the lower specification limit (LSL). Assuming the process to be Gaussian, its standard deviation can be denoted by sigma(σ).

Process width is defined as 6 times sigma (3 sigma on each side of mean). For example, if the USL and LSL of noise figure of an amplifier are 6.9 and 6.0 dB, then S is 0.9 dB. If the standard deviation (σ) is 0.1 dB, then P is 0.6 dB. Cp, the process capability, is 0.9/0.6 and is 1.5.

When Cp is 1, then 99.73% of the units pass specs and the process produces 0.27%. rejects. When the value of Cp increases, the number of rejects reduce dramatically. Percent defects are no longer used at higher values of Cp, instead parts per million (ppm) is used to describe the number of rejects. For example, when Cp is 1.5, rejects are 5 ppm and when Cp is 2 the rejects are 0.002 ppm (In this case process width is $\pm 6\sigma$ and the process is called a 6σ process). All the above numbers are based on the assumption that the center of the spec limits and the center of the process are the same. When this not true, Cp does not provide complete information.

Cpk (Process capability of a non-centered process) Cp does not take into account non centering of the process and hence is of minimal value in practice. In this case another term, called Cpk, is used and takes into account any non-centering of the process and is defined as

$$Cpk = Cp(1 - K), \text{ where:}$$

$$K = \left| \frac{D - \bar{X}}{S/2} \right| \text{ and } D = \text{design center}$$

$$\text{Cpk is also defined as : } Cpk = \left| \frac{NSL - \bar{X}}{3\sigma} \right|$$

where NSL is the nearest spec limit, \bar{X} is the mean of the process, and the vertical lines indicate that Cpk is always a positive number. Cp and Cpk are equal for a centered process. Cpk is also useful for defining processes with single-sided specs. For example, noise figure of an amplifier has only an upper spec limit and active directivity a lower spec limit. In deriving Cpk, one should

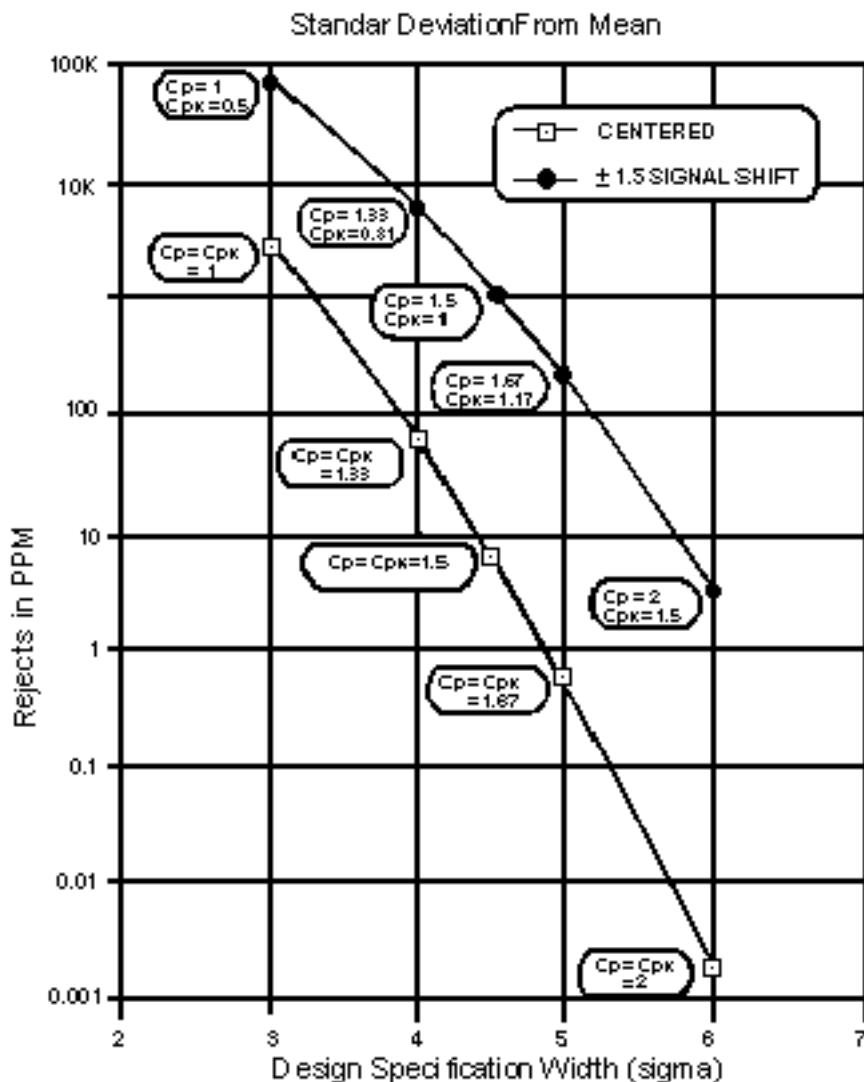
make sure that \overline{X} has a meaningful value, such as between spec limits when both spec limits are present. For single-sided specs, \overline{X} should be below the upper spec limit or above the lower spec limit. The graph below shows the number of rejects for various values of Cpk.

Directivity (active) is defined as the difference between isolation and forward gain in dB. It is an indication of the isolation of the source from the load, or how much the load impedance affects the input impedance and the source impedance affects the output impedance. The higher the active directivity (in dB), the better the isolation.

Dynamic range is the power range over which an amplifier provides useful linear operation, with the lower limit dependent on the noise figure and the upper level a function of the 1 dB compression point.

Gain flatness indicates the variation of an amplifier's gain characteristic over the full frequency response range at a given temperature expressed in dB.

Gain (forward gain, G) for RF amplifiers is the ratio of output power to input power, specified in the small signal, linear gain region, with a signal applied at the input. Gain in dB is defined as, $G_{dB} = 10 \log_{10} G$.



Harmonic distortion is produced by nonlinear amplifier operation and appears in the form of output signal frequencies at integral multiples of the input signal frequency. Since harmonic distortion is influenced by input power level it is generally specified in terms of the relative level for the harmonics to the fundamental signal power.

Isolation is the ratio of the power applied to the output of the amplifier to that measured at the input of the amplifier.

Linearity of an amplifier signifies its output power is a linear function of the input power. A linear amplifier produces at its output an amplified replica of the input signal with negligible or no harmonic generation.

Maximum signal level refers to the largest CW or pulse RF signal that can be safely applied to an amplifier's input. Exceeding the specified limit can result in noise figure degradation, increased distortion, gain reduction, and/or amplifier burnout.

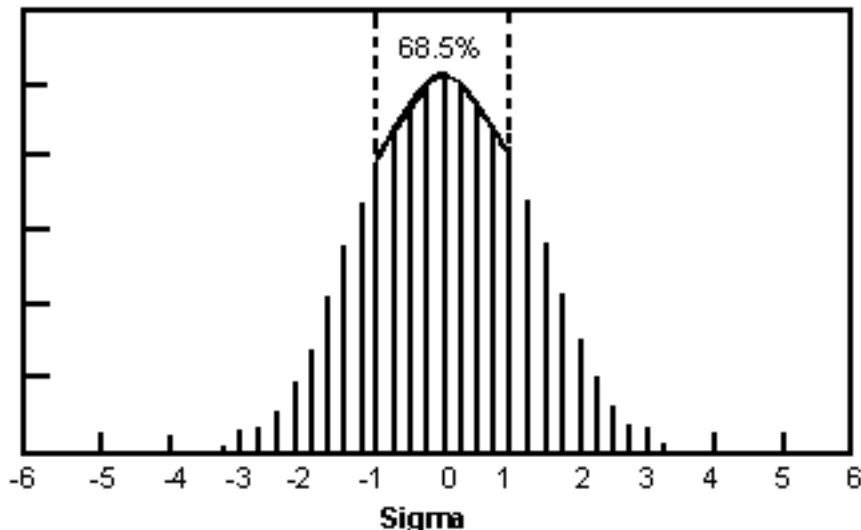
Noise factor is the ratio of signal-to-noise power ratio at an amplifier's input to the signal-to-noise power ratio at the output. Noise figure NF in dB is related to noise factor F by

$$NF = 10 \log_{10} F \text{ in dB}$$

Return loss(RL) is the ratio of reflected power to incident power at the RF port of an amplifier, expressed in dB by return loss ($RL = -20 \log |\rho|$, where ρ = reflection coefficient).

Reverse gain is the ratio of power measured at the input of an amplifier to the applied power at the output of an amplifier, also known as isolation.

Sigma is the statistical term for the Standard Deviation of a distribution. All things differ one from another to a greater or lesser degree. The measure of how much a distribution varies around the average or middle value \bar{x} is a measure of its standard deviation. Most distributions take a bell-shaped curve appearance, called a normal distribution. The distance from the center of a normal distribution curve to where the curve changes direction, or its point of inflection, is called the standard deviation or the Greek letter sigma; one sigma denotes one standard deviation.



Sigma performance, SP, tells how far the nearest spec limit is from the average, compared with the sigma value. For example, if the spec limit is 6 sigma away, the process has $SP = 6$. Thus, $5P$ equals 3 times the CpK value.

"Skinny" sigma refers to a small deviation. The relationship between the process standard deviation and the width of the specification reflects the sigma. For example, in a well-controlled process the nature of deviation from one unit to the next is small and most units thus fall well within the spec limits; the narrow variation indicate a "skinny" sigma. In a process that is not tightly controlled, units will vary from one spec extreme to the other or even exceed the spec limits; here the standard deviation and sigma are wide. What causes some confusion is the fact that 6 sigma is skinny while 2 sigma is wide; 6 sigma means the spec limits are much further away from the distribution.

Stability of an amplifier is an indication of its tendency to oscillate or generate a signal at its output without an applied input.

Two-Tone Third-order intercept point Two-tone third-order intercept point is a measure of third-order products generated by a second signal arriving at the input of a device such as an amplifier. If F_1 and F_2 are the frequencies of the two signals arriving at the input, the amplifier generates intermod products at its output due to inherent nonlinearity, in the form $\pm m*F_1 \pm n*F_2$ where m and n are positive integers which can assume any value from 1 to infinity. The order of the intermod is defined as $m + n$. Hence $2*F_1 - F_2$, $2*F_2 - F_1$, $3*F_1$ and $3*F_2$ are third-order products by definition. The first two products are called two-tone third-order products as they are generated when two tones are applied simultaneously at the input and the latter two are called single-tone third-order products.

For example, if 100 and 101 MHz are the frequencies of two applied signals, then 99 and 102 MHz are the two-tone third-order products and 300 and 303 MHz are single-tone third-order products. Two-tone third-order products are very close to the desired signals and are very difficult to filter out. Hence they are of great importance in system design.

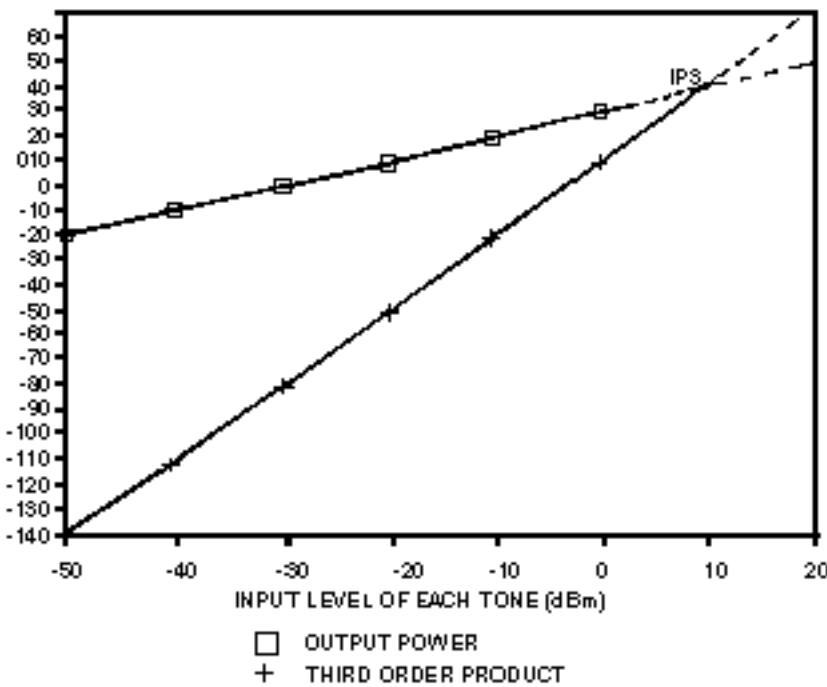
In the linear region, third-order products decrease/increase by 3 dB for every 1 dB decrease/increase of input power, and output signal power decreases/increases by a dB for every dB of input power. When drawn on a X-Y graph, with input power on X-axis and output power on the Y-axis, third-order products fall on a straight line with a slope of 3 and signal power on a straight line with a slope 1 as shown below. By extending the linear portions the two lines, they intercept at a point. The X co-ordinate and the Y co-ordinate of this point are called the input and output intercept point, and the two differ by an amount equal to the small-signal gain of the amplifier. Output intercept point, $IP3(dBm)$ can also be calculated using a simple formula.

$$IP3(dBm)_{out} = P_{out}(dBm) + A/2$$

where P_{out} (dBm) is the output power of each tone in dBm and "A" is the difference of output power and intermod level in dB. Input intercept point is obtained by substituting $P_{in}(dBm)$ for $P_{out}(dBm)$ in the above formula. Single-tone and two-tone third-order intercept points differ by a

fixed amount but have the same slope.

Amplifier Third-Order Intercept Point



Unconditionally stable refers to an amplifier that will not oscillate regardless of load or source impedance.

VSWR (voltage standing wave ratio) is related to return loss(RL) by the following:

$$\text{VSWR} = \frac{1 + 10^{-\text{RL}/20}}{1 - 10^{-\text{RL}/20}}$$

14 Often-Asked Questions About Amplifiers

Q. What is the effect of using a 50-ohm amplifier in a 75-ohm system?

A. When a 75-ohm load is seen from an ideal 50-ohm amplifier or vice-versa, a 1.5:1 VSWR results which alters gain, output return loss, and gain flatness in real-life amplifiers. If active directivity (defined as isolation minus gain) is low, a change in load impedance will result in a change of input impedance and a change in source impedance will result in a change in output impedance. Hence, a 75-ohm load on an amplifier with low directivity will affect the input

impedance of an amplifier. Maximum transfer of power may not occur.

However, in many applications, the mismatch may not be objectionable. For specific performance details, the 50-ohm amplifier should be tested under 75-ohm conditions. Contact the factory or sales rep. for computer automated performance data (CAPD) on the amplifier you are considering.

Q. What is output VSWR and what is its significance?

A. Output VSWR is a measure of how much power is reflected back from the amplifier's output port when an external signal is applied to that port. VSWR varies from a theoretical value of 1:1 for a perfect match to greater than 20:1 for total mismatch. Since loads in practical applications vary with frequency, maximum power and gain flatness also will deviate from what is specified. If the amplifier is connected to its load by a cable and all three have different impedance, then multiple reflections between the amplifier and its load can occur resulting in greater variation in frequency response. In general, the output impedance (characterized by output VSWR) is the source impedance of the following device.

Q. How is output VSWR measured?

A. A simple setup using a directional coupler is shown below.

First establish the 0 dB reference as follows. Apply the input signal to the directional coupler output port as shown. Apply a short circuit to the coupler's input port and measure the power at the coupled port. Then replace the short with an open circuit and note the reading at the coupled port. The average of the two readings is the 0 dB reference. Next substitute the open circuit with a 50 ohm load. Note the reading; this will give you the measurement range of the setup. Remove the 50 ohm load and replace it with the DUT. Measure how far the reflected signal is from the 0 dB reference; this is the output return loss (R.L.). To convert output return loss to VSWR, use the formula:

$$\text{VSWR} = \frac{1 + 10^{(\text{-R.L./20})}}{1 - 10^{(\text{-R.L./20})}}$$

```
graph LR; SG[SIGNAL GENERATOR] --> COUPLER[COUPLER]; COUPLER -- OUT --> AMP[AMP (DUT)]; COUPLER -- CPL IN --> SA[SA]; AMP -- IN --> PS[DC POWER SUPPLY]; AMP -- OUT --> LO[50 OHM]; SA --> LO
```

For more accurate measurements, use a scalar or vector network analyzer.

Q. What is the relationship between reflection coefficient, VSWR, and output return loss?

A. The voltage reflection coefficient (ρ) is the ratio of the reflected to incident voltage in an amplifier or device. The theoretical reflection coefficient varies from zero for a perfect match to one for a total mismatch. Magnitude of reflection coefficient and VSWR are related by

$$|\rho| = \frac{VSWR - 1}{VSWR + 1}$$

Return loss is related to the magnitude of reflection coefficient by

$$R.L. = -20 \log_{10} |\rho| \quad \text{or}$$

$$R.L. = -20 \log_{10} \frac{VSWR - 1}{VSWR + 1}$$

Q. To improve matching, can I use a resistive pad between amplifier stages?

A. Of course. But at the expense of overall gain, noise figure, and/or output power. The higher the gain of the first stage and the lower the value of the attenuator, the less the degradation of noise figure. Overall noise figure is calculated as follows:

$$F = F_1 + \left(\frac{L-1}{G_1} \right) + \frac{(F_2 - 1)L}{G_1} = F_1 + \frac{F_2 L - 1}{G_1}$$

Where F_1, F_2 are noise factors of first, and second amplifiers, and L is the loss of the pad.

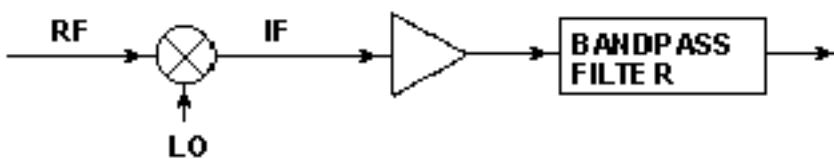
Noise figure in dB = $10 \log_{10} F$, where F is noise factor

Loss in dB = $10 \log_{10} L$, where L is loss factor

Gain in dB = $10 \log_{10} G$, where G is gain factor

Q. What is the significance of an amplifier's directivity characteristic in a system design?

A. Directivity is the difference between isolation and gain. Directivity is an indication of how the impedance mismatch at the amplifier's output affects the input.



In receiver applications, a filter following a wideband amplifier reflects undesired signals back to the amplifier and then to the mixer stage. If the amplifier provides high directivity, the reflected signals reaching the mixer will be much lower in magnitude and thus have little inter-action at the mixer stage.

Another common application is two-tone, third-order IM testing, where the two-tone signals must be well isolated; amplifiers with high directivity are used between source and combiner.

For relatively high RF frequencies, isolators can be used but they are expensive; for frequencies below 1 GHz, they are difficult to find.

High-directivity amplifiers, such as Mini-Circuits' MAN-AD series, are recommended for such

applications.

Q. Can I obtain higher power output by paralleling amplifiers?

A. Yes, but it's not as simple as merely connecting the two inputs and the two outputs in parallel. It involves judicious use of power hybrids with proper amplitude and phase balance and power levels, as well as amplifiers well matched for gain and phase characteristics. Quite candidly, it's generally advisable to specify an amplifier that meets the high-power requirements.

Q. I want to vary an amplifier's gain. Can I adjust the amplifier's supply voltage to achieve an AGC effect?

A. It's not recommended. An amplifier is designed to operate at a given supply voltage and its performance specifications (gain, power output, saturation, frequency response, etc.) are based on the stated supply voltage. Boost the supply voltage too much and gamble on the amplifier burning up; reduce the supply voltage too low and expect the performance specs to deviate considerably.

Variable-gain amplifiers are available, such as Mini-Circuits' ZFL-1000G and ZFL-1000GH, to accomplish your objective.

Q. I'm working with a 12-volt system and am considering your +30 dbm ZHL-series amplifiers. The specs indicate a +15 volt supply is required. How will performance be affected with 12-rather than 15-volt DC input?

A. Mini-Circuits' RF Signal-Processing Handbook, Vol. 3 on Amplifiers provides performance figures for three DC power input levels: +15, +16, and +12 volts. Need a copy? Ask your local Mini-Circuits' sales representative or contact the factory for a free copy.

Q. Is there a simple way to estimate second- and third-order intercept point for an amplifier?

A. As a general rule, the second-order point is 18 to 20 dB above the 1 dB compression point while the third-order intercept point is 10 dB above the 1 dB compression point.

Q. My application involved injecting sharp RF pulses to an amplifier. How can I tell whether the amplifier's peak power limits will be exceeded?

A. Here's a conservative rule-of-thumb estimate for a 50-ohm amplifier. (1) Take the amplifier's maximum power spec. (2) Convert from dBm to W. (3) Multiply by 100. (4) Take the square root, and use this figure as the maximum peak voltage that can be applied.

Q. When an amplifier is used in a test setup, is there such a thing as a safe sequence to connect the amplifier's input, output, and supply voltage to avoid damage?

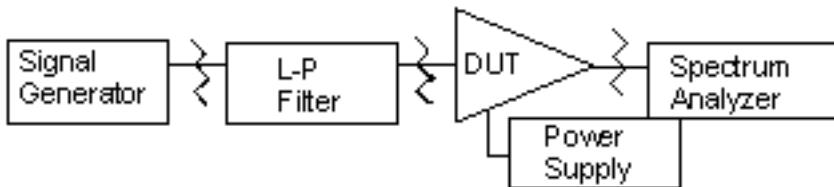
A. Yes, there is a recommended procedure. Begin by connecting the load, then the DC supply, and finally the RF input signal. When finished, first disconnect the RF input, then the DC power, and finally the load.

Q. Please sketch the test setup and describe the test procedure used to measure an amplifier's second-order intercept point.

A. Block diagram of a set up for measuring amplifier harmonics is shown below. Second harmonic of an amplifier is quantified and specified as second order intercept point (IP2). In the linear region of the amplifier, if second harmonic is A2 dB below fundamental, this IP2 is given by

$$IP2(\text{dBm}) = P_{\text{out}}(\text{dBm}) + A_2$$

In the block diagram, a low-pass filter is provided to attenuate second harmonics of the generator 10 to 20 dB below that generated by the amplifier. Sufficient attenuation should be provided at the amplifier output to prevent spectrum analyzer from generating harmonics.



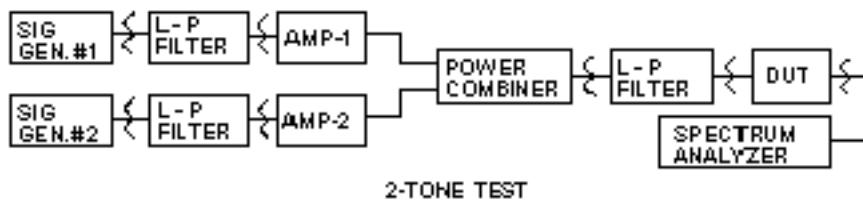
SECOND-ORDER INTERCEPT TEST

Q. *Describe the procedure for measuring 2-tone 3-order intercept point of an amplifier.*

A. Block diagram of a setup for measuring two-tone third-order intercept point is shown below.

If F1 & F2 are the frequencies of the two tones, then 2F1 – F2 and 2F2 – F1 are the third-order products. The set up should ensure that second harmonic of F1 & F2 are at least 10 to 20 dB below the third-order products to be measured. Care also should be taken to prevent F1 & F2 interaction and generation of third-order products. Amplifiers 1 and 2 are selected such that they have high directivity. This provides the desired isolation of the generators. If A3 is the level of the third-order product below the desired signal, then the output third-order intercept point is given as

$$IP3(\text{dBm}) = P_{\text{out}}(\text{dBm}) + A_3/2$$



EECS 142



Integrated Circuits for Communication

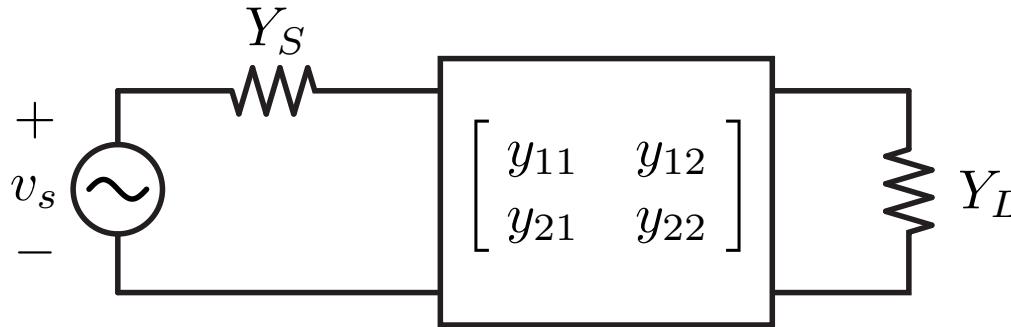
Lecture 4: Two-Port Circuits and Power Gain

Prof. Ali M. Niknejad

University of California, Berkeley

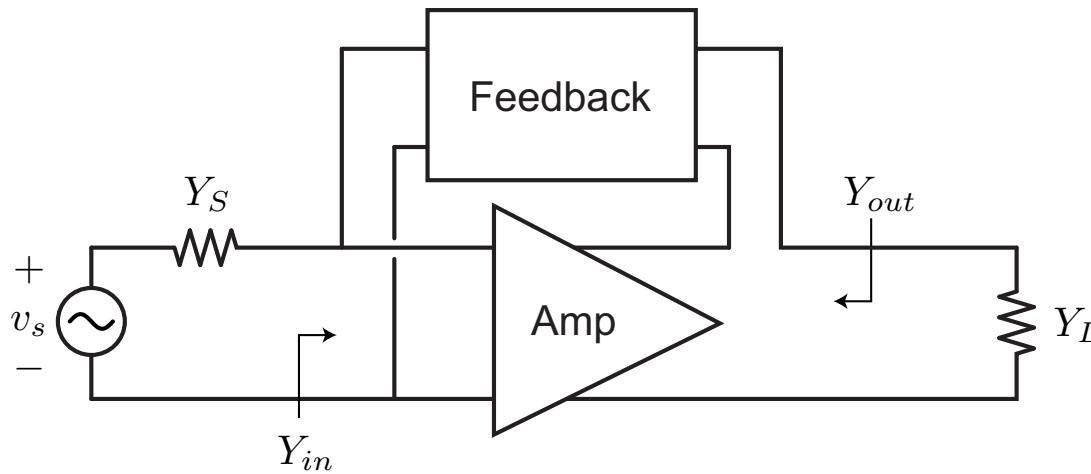
Copyright © 2005 by Ali M. Niknejad

A Generic Amplifier



- Consider the generic two-port amplifier shown above.
 - Note that any two-port linear and time-invariant circuit can be described in this way.
- We can use any two-port parameter set, including admittance parameters Y , impedance parameters Z , or hybrid or inverse-hybrid parameters H or G .

Choosing Two-Port Parameters



- The choice of parameter set is usually determined by convenience. For instance, if shunt feedback is applied, Y parameters are most convenient, whereas series feedback favors Z parameters. Other combinations of shunt/series can be easily described by H or G .
- $ABCD$ parameters are useful for cascading two-ports.

Y Parameters

- We'll primarily use the Y parameters

$$\begin{pmatrix} i_1 \\ i_2 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

- But in fact the choice depends largely on convenience. Often the form of feedback determines the best choice.
- All 2-port parameters are equivalent. Many of the results that we derive carry in terms of Y-parameters can be applied to other two-port parameters (input impedance, output impedance, gain, etc).

Admittance Parameters

- Notice that y_{11} is the short circuit input admittance

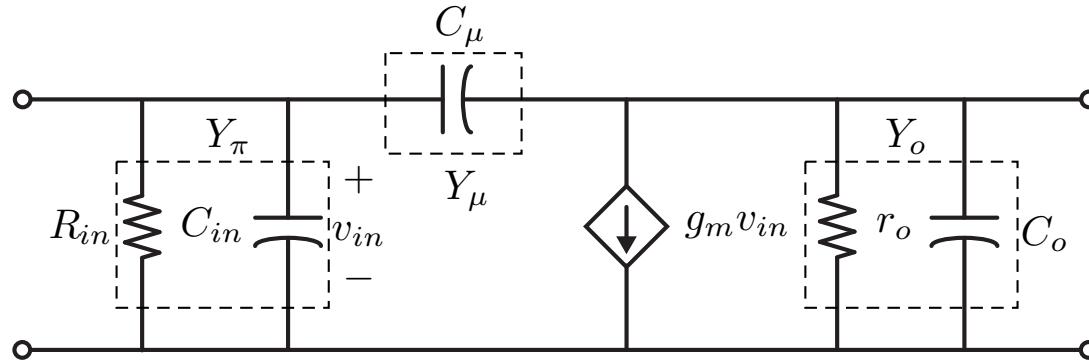
$$y_{11} = \left. \frac{i_1}{v_1} \right|_{v_2=0}$$

- The same can be said of y_{22} . The forward transconductance is described by y_{21}

$$y_{21} = \left. \frac{i_2}{v_1} \right|_{v_2=0}$$

- whereas the reverse transconductance is described by y_{12} .
- If a two-port amplifier is unilateral, then $y_{12} = 0$

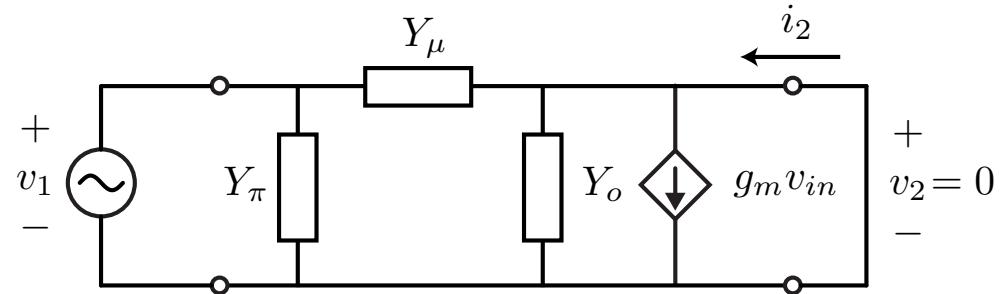
Hybrid- Π Admittance Parameters



- Let's compute the Y parameters for the common hybrid-II model

$$y_{11} = y_\pi + y_\mu$$

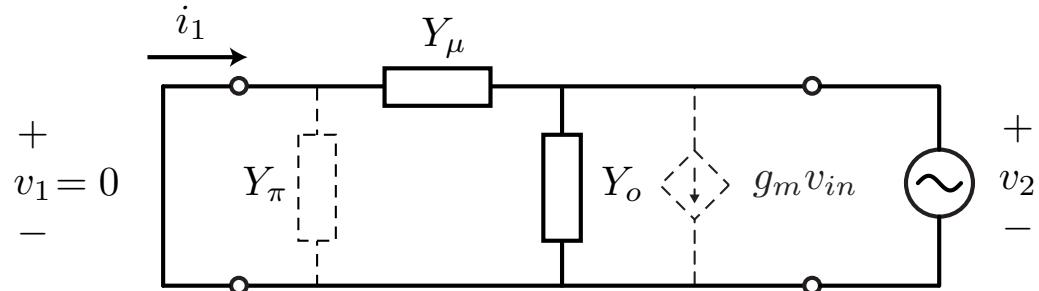
$$y_{21} = g_m - y_\mu$$



Admittance Parameters (cont)

$$y_{22} = y_o + y_\mu$$

$$y_{12} = -y_\mu$$



- Note that the hybrid- π model is unilateral if $y_\mu = sC_\mu = 0$. Therefore it's unilateral at DC.
- A good amplifier has a high ratio $\frac{y_{21}}{y_{12}}$ because we expect the forward transconductance to dominate the behavior

Why Use Two-Port Parameters?

- The parameters are generic and independent of the details of the amplifier → can be a single transistor or a multi-stage amplifier
- High frequency transistors are more easily described by two-port parameters (due to distributed input gate resistance and induced channel resistance)
- Feedback amplifiers can often be decomposed into an equivalent two-port unilateral amplifier and a two-port feedback section
- We can make some very general conclusions about the “optimal” power gain of a two-port, allowing us to define some useful metrics

Voltage Gain and Input Admittance

- Since $i_2 = -v_2 Y_L$, we can write

$$(y_{22} + Y_L)v_2 = -y_{21}v_1$$

- Which leads to the “internal” two-port gain

$$A_v = \frac{v_2}{v_1} = \frac{-y_{21}}{y_{22} + Y_L}$$

- Check low freq limit for a hybrid-II: $A_v = -g_m Z_o || Z_L$ ✓
- The input admittance is easily calculated from the voltage gain

$$Y_{in} = \frac{i_1}{v_1} = y_{11} + y_{12} \frac{v_2}{v_1}$$

$$Y_{in} = y_{11} - \frac{y_{12}y_{21}}{y_{22} + Y_L}$$

Output Admittance

- By symmetry we can write down the output admittance by inspection

$$Y_{out} = y_{22} - \frac{y_{12}y_{21}}{y_{11} + Y_S}$$

- Note that for a unilateral amplifier $y_{12} = 0$ implies that

$$Y_{in} = y_{11}$$

$$Y_{out} = y_{22}$$

- The input and output impedance are de-coupled!

External Voltage Gain

- The gain from the voltage source to the output can be derived by a simple voltage divider equation

$$A'_v = \frac{v_2}{v_s} = \frac{v_2}{v_1} \frac{v_1}{v_s} = A_v \frac{Y_S}{Y_{in} + Y_S} = \frac{-Y_S y_{21}}{(y_{22} + Y_L)(Y_S + Y_{in})}$$

- If we substitute and simplify the above equation we have

$$A'_v = \frac{-Y_S y_{21}}{(Y_S + y_{11})(Y_L + y_{22}) - y_{12}y_{21}}$$

- Verify that this makes sense at low frequency for hybrid-II:

$$A'_v(DC) = \frac{-Y_S y_{21}}{(Y_S + y_{11})(Y_L + y_{22})} = \frac{Z_{in}}{Z_{in} + Z_S} \times -g_m R_L || r_o$$

Feedback Amplifiers and Y -Params

- Note that in an ideal feedback system, the amplifier is unilateral and the closed loop gain is given by $\frac{y}{x} = \frac{A}{1+Af}$
- We found last lecture that the voltage gain of a general two-port driven with source admittance Y_S is given by

$$A'_v = \frac{-Y_S y_{21}}{(Y_S + y_{11})(Y_L + y_{22}) - y_{12}y_{21}}$$

- If we unilaterize the two-port by arbitrarily setting $y_{12} = 0$, we have an “open” loop forward gain of

$$A_{vu} = A'_v|_{y_{12}=0} = \frac{-Y_S y_{21}}{(Y_S + y_{11})(Y_L + y_{22})}$$

Identification of Loop Gain

- Re-writing the gain A'_v by dividing numerator and denominator by the factor $(Y_S + y_{11})(Y_L + y_{22})$ we have

$$A'_v = \frac{\frac{-Y_S y_{21}}{(Y_S + y_{11})(Y_L + y_{22})}}{1 - \frac{y_{12} y_{21}}{(Y_S + y_{11})(Y_L + y_{22})}}$$

- We can now see that the “closed” loop gain with $y_{12} \neq 0$ is given by

$$A'_v = \frac{A_{vu}}{1 + T}$$

- where T is identified as the loop gain

$$T = A_{vu}f = \frac{-y_{12}y_{21}}{(Y_S + y_{11})(Y_L + y_{22})}$$

The Feedback Factor and Loop Gain

- Using the last equation also allows us to identify the feedback factor

$$f = \frac{Y_{12}}{Y_S}$$

- If we include the loading by the source Y_S , the input admittance of the amplifier is given by

$$Y_{in} = Y_S + y_{11} - \frac{y_{12}y_{21}}{Y_L + y_{22}}$$

- Note that this can be re-written as

$$Y_{in} = (Y_S + y_{11}) \left(1 - \frac{y_{12}y_{21}}{(Y_S + y_{11})(Y_L + y_{22})} \right)$$

Feedback and Input/Output Admittance

- The last equation can be re-written as

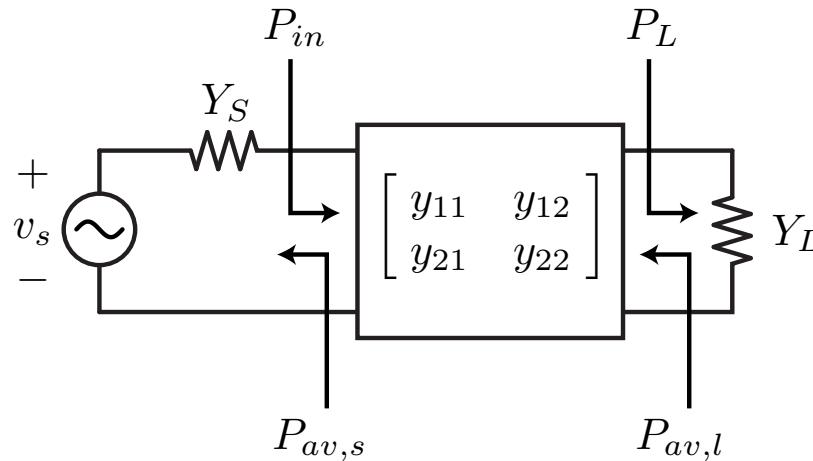
$$Y_{in} = (Y_S + y_{11})(1 + T)$$

- Since $Y_S + y_{11}$ is the input admittance of a unilateral amplifier, we can interpret the action of the feedback as raising the input admittance by a factor of $1 + T$.
- Likewise, the same analysis yields

$$Y_{out} = (Y_L + y_{22})(1 + T)$$

- It's interesting to note that the same equations are valid for series feedback using Z parameters, in which case the action of the feedback is to boost the input and output impedance.

Power Gain



- We can define power gain in many different ways. The *power gain* G_p is defined as follows

$$G_p = \frac{P_L}{P_{in}} = f(Y_L, Y_{ij}) \neq f(Y_S)$$

- We note that this power gain is a function of the load admittance Y_L and the two-port parameters Y_{ij} .

Power Gain (cont)

- The *available power gain* is defined as follows

$$G_a = \frac{P_{av,L}}{P_{av,S}} = f(Y_S, Y_{ij}) \neq f(Y_L)$$

- The available power from the two-port is denoted $P_{av,L}$ whereas the power available from the source is $P_{av,S}$.
- Finally, the *transducer gain* is defined by

$$G_T = \frac{P_L}{P_{av,S}} = f(Y_L, Y_S, Y_{ij})$$

- This is a measure of the efficacy of the two-port as it compares the power at the load to a simple conjugate match.

Derivation of Power Gain

- The power gain is readily calculated from the input admittance and voltage gain

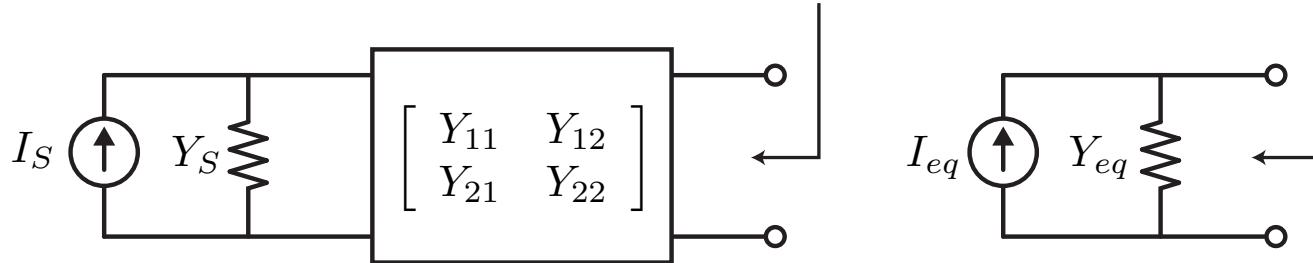
$$P_{in} = \frac{|V_1|^2}{2} \Re(Y_{in})$$

$$P_L = \frac{|V_2|^2}{2} \Re(Y_L)$$

$$G_p = \left| \frac{V_2}{V_1} \right|^2 \frac{\Re(Y_L)}{\Re(Y_{in})}$$

$$G_p = \frac{|Y_{21}|^2}{|Y_L + Y_{22}|^2} \frac{\Re(Y_L)}{\Re(Y_{in})}$$

Derivation of Available Gain



- To derive the available power gain, consider a Norton equivalent for the two-port where (short port 2)

$$I_{eq} = I_2 = Y_{21}V_1 = \frac{Y_{21}}{Y_{11} + Y_S} I_S$$

- The Norton equivalent admittance is simply the output admittance of the two-port

$$Y_{eq} = Y_{22} - \frac{Y_{21}Y_{12}}{Y_{11} + Y_S}$$

Available Gain (cont)

- The available power at the source and load are given by

$$P_{av,S} = \frac{|I_S|^2}{8\Re(Y_S)} \quad P_{av,L} = \frac{|I_{eq}|^2}{8\Re(Y_{eq})}$$

$$G_a = \left| \frac{I_{eq}}{I_S} \right|^2 \frac{\Re(Y_S)}{\Re(Y_{eq})}$$

$$G_a = \left| \frac{Y_{21}}{Y_{11} + Y_S} \right|^2 \frac{\Re(Y_S)}{\Re(Y_{eq})}$$

Transducer Gain Derivation

- The transducer gain is given by

$$G_T = \frac{P_L}{P_{av,S}} = \frac{\frac{1}{2}\Re(Y_L)|V_2|^2}{\frac{|I_S|^2}{8\Re(Y_S)}} = 4\Re(Y_L)\Re(Y_S) \left| \frac{V_2}{I_S} \right|^2$$

- We need to find the output voltage in terms of the source current. Using the voltage gain we have and input admittance we have

$$\left| \frac{V_2}{V_1} \right| = \left| \frac{Y_{21}}{Y_L + Y_{22}} \right|$$

$$I_S = V_1(Y_S + Y_{in})$$

$$\left| \frac{V_2}{I_S} \right| = \left| \frac{Y_{21}}{Y_L + Y_{22}} \right| \frac{1}{|Y_S + Y_{in}|}$$

Transducer Gain (cont)

$$|Y_S + Y_{in}| = \left| Y_S + Y_{11} - \frac{Y_{12}Y_{21}}{Y_L + Y_{22}} \right|$$

- We can now express the output voltage as a function of source current as

$$\left| \frac{V_2}{I_S} \right|^2 = \frac{|Y_{21}|^2}{|(Y_S + Y_{11})(Y_L + Y_{22}) - Y_{12}Y_{21}|^2}$$

- And thus the transducer gain

$$G_T = \frac{4\Re(Y_L)\Re(Y_S)|Y_{21}|^2}{|(Y_S + Y_{11})(Y_L + Y_{22}) - Y_{12}Y_{21}|^2}$$

Comparison of Power Gains

- It's interesting to note that *all* of the gain expression we have derived are in the exact same form for the impedance, hybrid, and inverse hybrid matrices.
- In general, $P_L \leq P_{av,L}$, with equality for a matched load. Thus we can say that

$$G_T \leq G_a$$

- The maximum transducer gain as a function of the load impedance thus occurs when the load is conjugately matched to the two-port output impedance

$$G_{T,max,L} = \frac{P_L(Y_L = Y_{out}^*)}{P_{av,S}} = G_a$$

Comparison of Power Gains (cont)

- Likewise, since $P_{in} \leq P_{av,S}$, again with equality when the two-port is conjugately matched to the source, we have

$$G_T \leq G_p$$

- The transducer gain is maximized with respect to the source when

$$G_{T,max,S} = G_T(Y_{in} = Y_S^*) = G_p$$

3

MICROWAVE TRANSISTOR AMPLIFIER DESIGN

3.1 INTRODUCTION

This chapter develops some basic principles used in the analysis and design of microwave transistor amplifiers. Based on the S parameters of the transistor and certain performance requirements, a systematic procedure is developed for the design of a microwave transistor amplifier.

The most important design considerations in a microwave transistor amplifier are stability, power gain, bandwidth, noise, and dc requirements. This chapter deals mainly with the problems of stability and power gain in narrow-band amplifiers. VSWR considerations are also discussed. Low-noise amplifiers, broadband amplifiers, and power amplifiers are discussed in Chapter 4.

A design usually starts with a set of specifications and the selection of the proper transistor. Then a systematic mathematical solution, aided by graphical methods, is developed to determine the transistor loading (i.e., the source and load reflection coefficients) for a particular stability and gain criteria. An unconditionally stable transistor will not oscillate with any passive termination. On the other hand, a design using a potentially unstable transistor requires some analysis and careful considerations so that the passive terminations produce a stable amplifier.

Design procedures for both unilateral and bilateral transistors, based on stability and gain requirements, are described. Both passive and active dc bias networks for BJTs and GaAs FETs are analyzed. It is important to select the correct dc operating point and the proper dc network topology in order to obtain the desired ac performance.

3.2 POWER GAIN EQUATIONS

Several power gain equations appear in the literature and are used in the design of microwave amplifiers. Figure 3.2.1 illustrates a microwave amplifier signal flow graph and the different powers used in gain equations. The transducer power gain G_T , the power gain G_p (also called the *operating power gain*), and the available power gain G_A are defined as follows:

$$G_T = \frac{P_L}{P_{AVS}} = \frac{\text{power delivered to the load}}{\text{power available from the source}}$$

$$G_p = \frac{P_L}{P_{IN}} = \frac{\text{power delivered to the load}}{\text{power input to the network}}$$

and

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{\text{power available from the network}}{\text{power available from the source}}$$

The expressions for G_T , G_p , and G_A were already derived in (2.6.14), (2.6.15), (2.6.18), and (2.6.22)—namely,

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_{IN}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.2.1)$$

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (3.2.2)$$

$$G_p = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.2.3)$$

$$G_A = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{OUT}|^2} \quad (3.2.4)$$

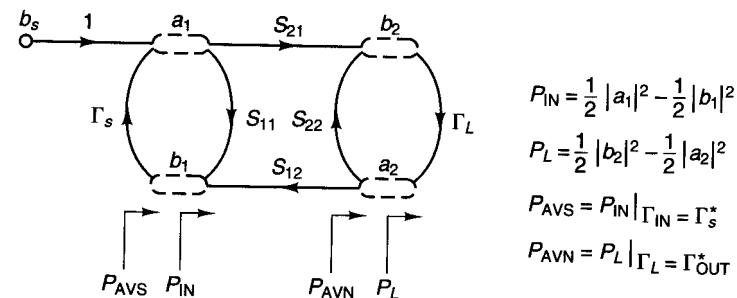


Figure 3.2.1 Different power definitions.

$$\Gamma_{IN} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \quad (3.2.5)$$

$$\Gamma_{OUT} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \quad (3.2.6)$$

G_T is a function of Γ_s , Γ_L , and the S parameters of the transistor (i.e., $G_T = f(\Gamma_s, \Gamma_L, [S])$), $G_p = f(\Gamma_L, [S])$, and $G_A = f(\Gamma_s, [S])$.

In terms of the amplifier shown in Fig. 3.2.2, the input matching network transforms the generator impedance Z_1 (usually 50Ω) to the impedance Z_s , or in other words to the source reflection coefficient Γ_s . The output matching network transforms the impedance Z_2 (usually 50Ω) to the load impedance Z_L or to the load reflection coefficient Γ_L . Observe that we use the nomenclature Z_s and Γ_s for the source impedance and source reflection coefficient at the input port of the transistor, and Z_L and Γ_L for the load impedance and load reflection coefficient at the output port of the transistor.

The values of Γ_s , Γ_L , and the S parameters of the transistor determine the gains of the amplifier according to (3.2.1) through (3.2.4).

The passive matching networks produce values of Γ_s and Γ_L such that $|\Gamma_s| < 1$ and $|\Gamma_L| < 1$. In other words, the resistive part associated with Z_s and Z_L is positive. However, from (3.2.5) and (3.2.6) it is possible that for certain values of the S parameters (where $|\Gamma_s| < 1$ and $|\Gamma_L| < 1$) that $|\Gamma_{IN}| > 1$ and/or $|\Gamma_{OUT}| > 1$. When $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$, the input or output ports of the transistor present a negative resistance and oscillations can occur. Obviously, this is a situation that we must avoid in amplifier design.

Example 3.2.1

(a) The input and output matching networks in Fig. 3.2.2 are designed to produce $\Gamma_s = 0.5|120^\circ$ and $\Gamma_L = 0.4|90^\circ$. Determine G_T , G_A , and G_p if the S parameters of the transistor are

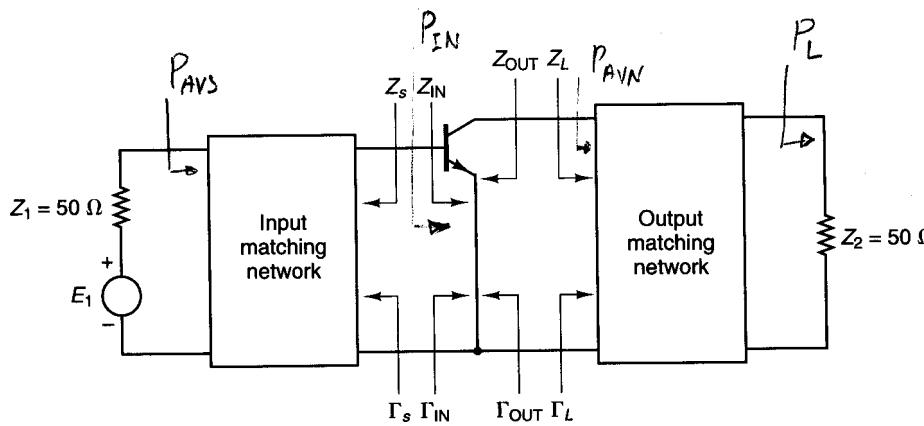


Figure 3.2.2 A microwave amplifier diagram.

$$S_{11} = 0.6| -160^\circ$$

$$S_{12} = 0.045| 16^\circ$$

$$S_{21} = 2.5| 30^\circ$$

$$S_{22} = 0.5| -90^\circ$$

(b) Calculate P_{AVS} , P_{IN} , P_{AVN} , and P_L in Fig. 3.2.2 if $E_1 = 10|0^\circ$, $Z_1 = 50 \Omega$, and $\Gamma_2 = 50 \Omega$.

(c) Calculate $(VSWR)_{in}$ and $(VSWR)_{out}$.

Solution. (a) From (3.2.5) and (3.2.6) we obtain

$$\Gamma_{IN} = 0.6| -160^\circ + \frac{0.045| 16^\circ(2.5| 30^\circ)0.4| 90^\circ}{1 - 0.5| -90^\circ(0.4| 90^\circ)} = 0.627| -164.6^\circ$$

and

$$\Gamma_{OUT} = 0.5| -90^\circ + \frac{0.045| 16^\circ(2.5| 30^\circ)0.5| 120^\circ}{1 - 0.6| -160^\circ(0.5| 120^\circ)} = 0.471| -97.63^\circ$$

Then, from (3.2.1), (3.2.3), and (3.2.4) it readily follows that

$$G_T = \frac{1 - (0.5)^2}{|1 - 0.627| -164.6^\circ(0.5| 120^\circ)|^2} \frac{(2.5)^2}{|1 - 0.5| -90^\circ(0.4| 90^\circ)|^2} = 9.43 \quad (\text{or } 9.75 \text{ dB})$$

$$G_p = \frac{1}{1 - (0.627)^2} \frac{(2.5)^2}{|1 - 0.5| -90^\circ(0.4| 90^\circ)|^2} = 13.51 \quad (\text{or } 11.31 \text{ dB})$$

and

$$G_A = \frac{1 - (0.5)^2}{|1 - 0.6| -160^\circ(0.5| 120^\circ)|^2} \frac{(2.5)^2}{|1 - (0.471)|^2} = 9.55 \quad (\text{or } 9.8 \text{ dB})$$

Since the values of $G_p = P_L/P_{IN}$ and $G_T = P_L/P_{AVS}$ differ by 1.56 dB (i.e., $11.31 \text{ dB} - 9.75 \text{ dB} = 1.56 \text{ dB}$), it follows that the input power is less than the power available from the source. In fact, recall (2.7.16) and (2.7.29)—namely,

$$P_{IN} = P_{AVS} M_s \quad (3.2.7)$$

and

$$G_T = G_p M_s \quad (3.2.8)$$

where from (2.7.17) the source mismatch factor is

$$M_s = \frac{[1 - (0.5)^2][1 - (0.627)^2]}{|1 - 0.5| -160^\circ(0.627| -164.6^\circ)|^2} = 0.6983 \quad (\text{or } -1.56 \text{ dB})$$

Observe how G_T and G_p are related by (3.2.8). That is, $9.43 = 13.51(0.6983)$ or $9.75 \text{ dB} = 11.31 \text{ dB} - 1.56 \text{ dB}$.

Also, since $G_A = P_{AVN}/P_{AVS}$ and $G_T = P_L/P_{AVS}$ are fairly close (i.e., 9.8 dB and 9.75 dB, respectively), the power delivered to the load is close to the power available from the network. In fact,

$$P_L = P_{AVN}M_L \quad (3.2.9)$$

and

$$G_T = G_A M_L \quad (3.2.10)$$

where from (2.7.25) the load mismatch factor is

$$M_L = \frac{[1 - (0.4)^2][1 - (0.471)^2]}{|1 - 0.471| - 97.63^\circ(0.4|90^\circ|)} = 0.9874 \quad (\text{or } -0.055 \text{ dB})$$

Observe how G_T and G_A are related by (3.2.10). That is, $9.43 = 9.55(0.9874)$ or 9.75 dB = 9.8 dB - 0.055 dB.

(b) The power available from the source is

$$P_{AVS} = \frac{E_1^2}{8 \operatorname{Re}[Z_1]} = \frac{10^2}{8(50)} = 0.25 \text{ W}$$

Then, using (3.2.7) the input power is

$$P_{IN} = P_{AVS}M_s = 0.25(0.6983) = 0.1745 \text{ W}$$

The power delivered to the load can be calculated using the definition of G_T . That is,

$$P_L = G_T P_{AVS} = 9.43(0.25) = 2.358 \text{ W}$$

It can also be calculated using the definition of G_P . That is,

$$P_L = G_P P_{IN} = 13.51(0.1745) = 2.358 \text{ W}$$

The power available from the network can be calculated using (3.2.9). That is,

$$P_{AVN} = \frac{P_L}{M_L} = \frac{2.358}{0.9874} = 2.39 \text{ W}$$

It can also be calculated using the definition of G_A . That is,

$$P_{AVN} = G_A P_{AVS} = 9.55(0.25) = 2.39 \text{ W}$$

(c) Since the mismatch factor M_s has been calculated in part (a), it is simple to use (2.8.1) and (2.8.2) to calculate $(VSWR)_{in}$. That is,

$$(VSWR)_{in} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} = \frac{1 + \sqrt{1 - M_s}}{1 - \sqrt{1 - M_s}} = \frac{1 + \sqrt{1 - 0.6983}}{1 - \sqrt{1 - 0.6983}} = 3.44$$

Similarly, using (2.8.4) and (2.8.5) we obtain

$$(VSWR)_{out} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} = \frac{1 + \sqrt{1 - M_L}}{1 - \sqrt{1 - M_L}} = \frac{1 + \sqrt{1 - 0.9874}}{1 - \sqrt{1 - 0.9874}} = 1.25$$

3.3 STABILITY CONSIDERATIONS

The stability of an amplifier, or its resistance to oscillate, is a very important consideration in a design and can be determined from the S parameters, the matching networks, and the terminations. In a two-port network, oscillations are possible when either the input or output port presents a negative resistance. This occurs when $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$, which for a unilateral device occurs when $|S_{11}| > 1$ or $|S_{22}| > 1$. For example, a unilateral transistor is a transistor where $S_{12} = 0$ (or its effect so small that it can be set equal to zero). If $S_{12} = 0$, it follows from (3.2.5) and (3.2.6) that $|\Gamma_{IN}| = |S_{11}|$ and $|\Gamma_{OUT}| = |S_{22}|$. Hence, if $|S_{11}| > 1$ the transistor presents a negative resistance at the input, and if $|S_{22}| > 1$ the transistor presents a negative resistance at the output.

The two-port network shown in Fig. 3.3.1 is said to be unconditionally stable at a given frequency if the real parts of Z_{IN} and Z_{OUT} are greater than zero for all passive load and source impedances. If the two-port is not unconditionally stable, it is potentially unstable. That is, some passive load and source terminations can produce input and output impedances having a negative real part.

In terms of reflection coefficients, the conditions for unconditional stability at a given frequency are

$$|\Gamma_s| < 1 \quad (3.3.1)$$

$$|\Gamma_L| < 1 \quad (3.3.2)$$

$$|\Gamma_{IN}| = \left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| < 1 \quad (3.3.3)$$

and

$$|\Gamma_{OUT}| = \left| S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \right| < 1 \quad (3.3.4)$$

where, of course, all coefficients are normalized to the same characteristic impedance Z_o .

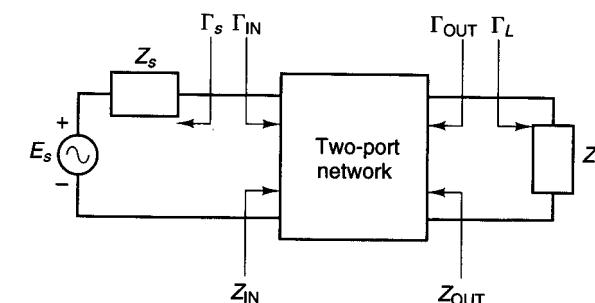


Figure 3.3.1 Stability of two-port networks.

Equations (3.3.1) and (3.3.2) state that the source and load are passive, while (3.3.3) and (3.3.4) state that the input and output impedances must also be passive (i.e., no negative resistance is associated with their real parts).

The solutions of (3.3.1) to (3.3.4) give the required conditions for the two-port network to be unconditionally stable. However, before we discuss the intricacies of the necessary and sufficient conditions for unconditional stability, a graphical analysis of (3.3.1) to (3.3.4) is presented. The graphical analysis is especially useful in the analysis of potentially unstable transistors.

When the two-port in Fig. 3.3.1 is potentially unstable, there may be values of Γ_s and Γ_L (i.e., source and load impedances) for which the real parts of Z_{IN} and Z_{OUT} are positive. These values of Γ_s and Γ_L (i.e., regions in the Smith chart) can be determined using the following graphical procedure.

First, the regions where values of Γ_L and Γ_s produce $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$ are determined, respectively. Setting the magnitude of (3.3.3) and (3.3.4) equal to 1 and solving for the values of Γ_L and Γ_s shows that the solutions for Γ_L and Γ_s lie on circles (called *stability circles*) whose equations are given by

$$\left| \Gamma_L - \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right| \quad (3.3.5)$$

and

$$\left| \Gamma_s - \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12}S_{21}}{|S_{11}|^2 - |\Delta|^2} \right| \quad (3.3.6)$$

where

$$\Delta = S_{11}S_{22} - S_{12}S_{21}$$

The derivations of (3.3.5) and (3.3.6) are given in Appendix A, Section A.2. For completeness, the reader can find in Appendix A, Section A.1, a review of circle equations and bilinear transformations.

The radii and centers of the circles where $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$ in the Γ_L plane and Γ_s plane, respectively, are obtained from (3.3.5) and (3.3.6), namely

Γ_L values for $|\Gamma_{IN}| = 1$ (Output Stability Circle):

$$r_L = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right| \quad (\text{radius}) \quad (3.3.7)$$

$$C_L = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \quad (\text{center}) \quad (3.3.8)$$

Γ_s values for $|\Gamma_{OUT}| = 1$ (Input Stability Circle):

$$r_s = \left| \frac{S_{12}S_{21}}{|S_{11}|^2 - |\Delta|^2} \right| \quad (\text{radius}) \quad (3.3.9)$$

$$C_s = \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \quad (\text{center}) \quad (3.3.10)$$

With the S parameters of a two-port device at one frequency, the expressions (3.3.7) to (3.3.10) can be calculated and plotted on a Smith chart, and the set of values of Γ_L and Γ_s that produce $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$ can be easily observed. Figure 3.3.2 illustrates the graphical construction of the stability circles where $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$. On one side of the stability circle boundary, in the Γ_L plane, we will have $|\Gamma_{IN}| < 1$ and on the other side $|\Gamma_{IN}| > 1$. Similarly, in the Γ_s plane on one side of the stability circle boundary, we will have $|\Gamma_{OUT}| < 1$ and on the other side $|\Gamma_{OUT}| > 1$.

Next we need to determine which area in the Smith chart represents the stable region—in other words, the region where values of Γ_L (where $|\Gamma_L| < 1$) produce $|\Gamma_{IN}| < 1$ and where values of Γ_s (where $|\Gamma_s| < 1$) produce $|\Gamma_{OUT}| < 1$. To this end, we observe that if $Z_L = Z_o$, then $\Gamma_L = 0$ and from (3.2.5) $|\Gamma_{IN}| = |S_{11}|$. If the magnitude of S_{11} is less than 1, then $|\Gamma_{IN}| < 1$ when $\Gamma_L = 0$. That is, the center of the Smith chart in Fig. 3.3.2a represents a stable operating point, because for $\Gamma_L = 0$ it follows that $|\Gamma_{IN}| < 1$. On the other hand, if $|S_{11}| > 1$ when $Z_L = Z_o$, then $|\Gamma_{IN}| > 1$ when $\Gamma_L = 0$ and the center of the Smith chart represents an unstable operating point. Figure 3.3.3 illustrates the two cases discussed. The shaded area represents the values of Γ_L that produce a stable operation. Similarly, Fig. 3.3.4 illustrates stable and unstable regions for Γ_s .

For unconditional stability any passive load or source in the network must produce a stable condition. From a graphical point of view, for $|S_{11}| < 1$ and $|S_{22}| < 1$, we want the stability circles shown in Figs. 3.3.3a and 3.3.4a to fall completely outside (or to completely enclose) the Smith chart. The case in which the stability circles fall completely outside the Smith chart is illustrated in Fig. 3.3.5. Therefore, the conditions for unconditional stability for all passive sources and loads can be expressed in the form

$$|C_L| - r_L > 1 \quad \text{for } |S_{11}| < 1 \quad (3.3.11)$$

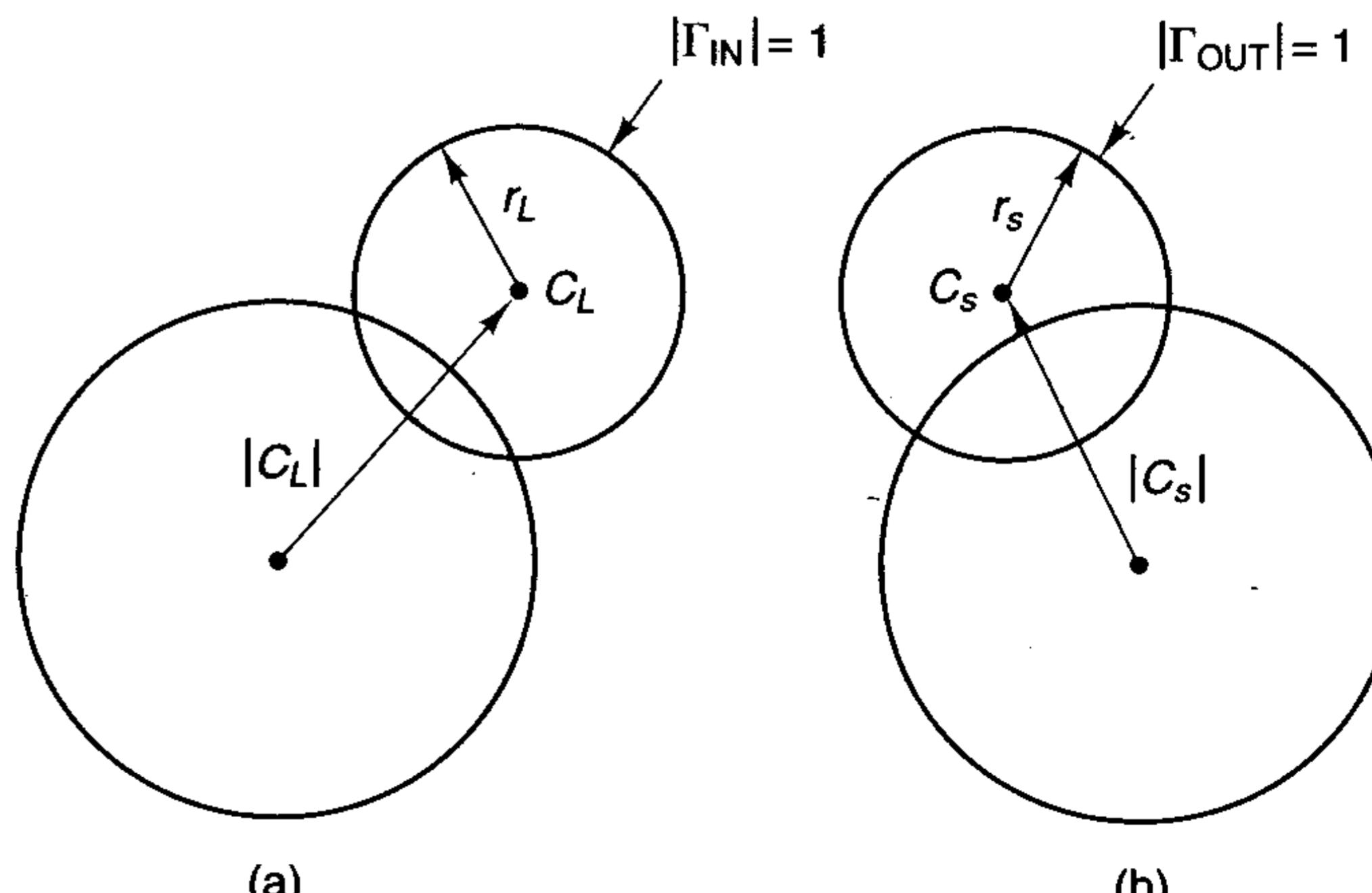


Figure 3.3.2 Stability circle construction in the Smith chart: (a) Γ_L plane; (b) Γ_s plane.

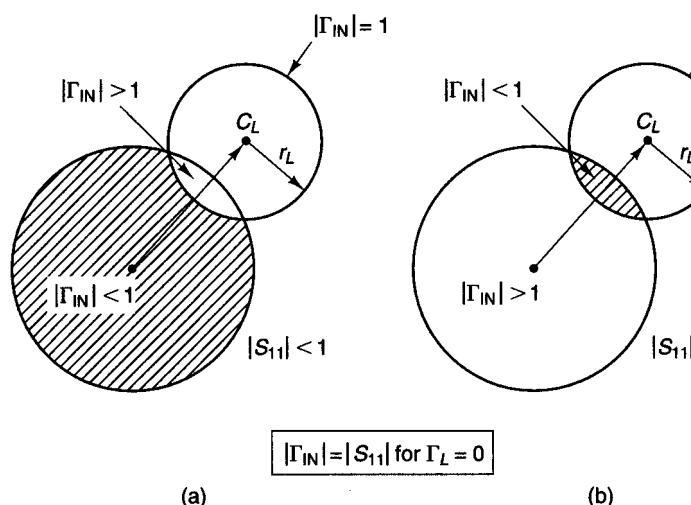


Figure 3.3.3 Smith chart illustrating stable and unstable regions in the Γ_L plane.

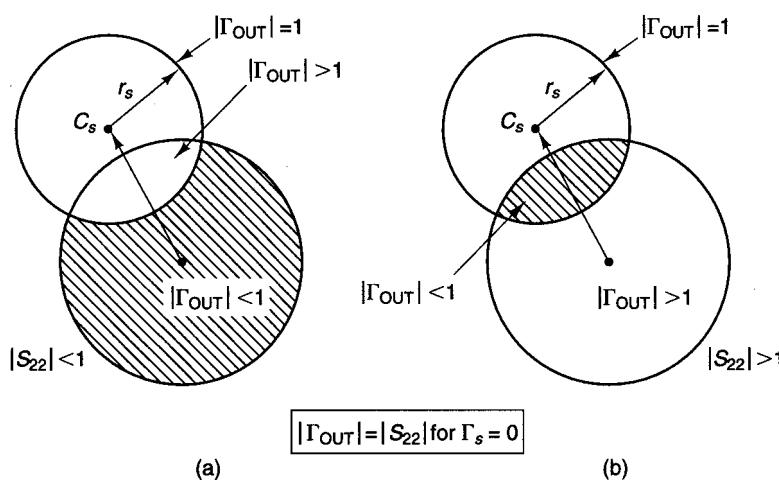


Figure 3.3.4 Smith chart illustrating stable and unstable regions in the Γ_s plane.

and

$$|C_s| - r_s > 1 \quad \text{for } |S_{22}| < 1 \quad (3.3.12)$$

If either $|S_{11}| > 1$ or $|S_{22}| > 1$, the network cannot be unconditionally stable because the termination $\Gamma_L = 0$ or $\Gamma_s = 0$ [see (3.3.3) and (3.3.4)] will produce $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$.

We now return to the necessary and sufficient conditions for a two-port to be unconditionally stable. A straightforward but somewhat lengthy manipu-

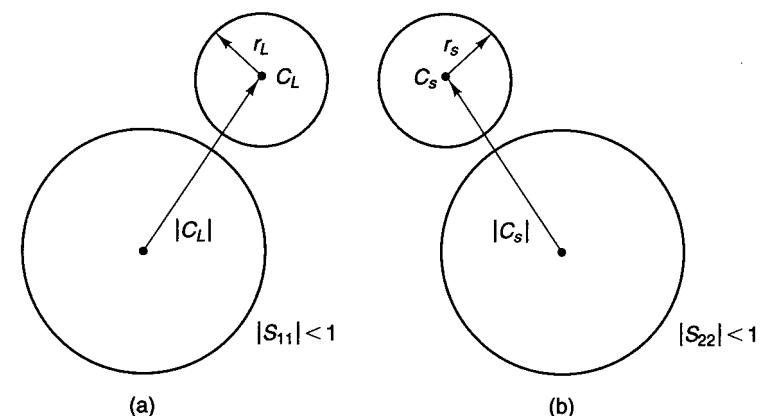


Figure 3.3.5 Conditions for unconditional stability: (a) Γ_L plane; (b) Γ_s plane.

lation of (3.3.1) to (3.3.4) results in the following necessary and sufficient conditions for unconditional stability (see Appendix B):

$$K > 1 \quad (3.3.13)$$

and

$$1 - |S_{11}|^2 > |S_{12}S_{21}| \quad (3.3.14)$$

$$1 - |S_{22}|^2 > |S_{12}S_{21}| \quad (3.3.15)$$

where

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} \quad (3.3.16)$$

and

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \quad (3.3.17)$$

There are other ways of expressing the necessary and sufficient conditions for unconditional stability [3.1]. Adding (3.3.14) and (3.3.15) gives

$$2 - |S_{11}|^2 - |S_{22}|^2 > 2|S_{12}S_{21}| \quad (3.3.18)$$

Since

$$|\Delta| = |S_{11}S_{22} - S_{12}S_{21}| \leq |S_{11}S_{22}| + |S_{12}S_{21}|$$

we use (3.3.18) to obtain

$$|\Delta| < |S_{11}S_{22}| + 1 - \frac{1}{2}|S_{11}|^2 - \frac{1}{2}|S_{22}|^2$$

$$|\Delta| < 1 - \frac{1}{2}(|S_{11}| - |S_{22}|)^2$$

or simply

$$|\Delta| < 1$$

Hence, a convenient way of expressing the necessary and sufficient conditions for unconditional stability is

$$K > 1 \quad (3.3.19)$$

and

$$|\Delta| < 1 \quad (3.3.20)$$

In this textbook we use (3.3.19) and (3.3.20) to test for unconditional stability.

Another way that the necessary and sufficient conditions for unconditional stability are found in the literature is (see Appendix C)

$$K > 1$$

and

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 > 0 \quad (3.3.21)$$

From a theoretical point of view, a two-port network can have any value of K and $|\Delta|$. From a practical point of view, most microwave transistors produced by manufacturers are either unconditionally stable or potentially unstable with $K < 1$ and $|\Delta| < 1$. In fact, in potentially unstable transistors most practical values of K are such that $0 < K < 1$. These potentially unstable transistors have source and load stability circles that intersect the boundary of the Smith chart (e.g., see Figs. 3.3.3a and 3.3.4a).

Negative values of K in the range $-1 < K < 0$ result in most of the Smith chart being unstable. Some transistor configurations (e.g., some CB configurations) used in oscillator designs are potentially unstable with negative values of K .

Example 3.3.1

The S parameters of a BJT at $V_{CE} = 15$ V and $I_C = 15$ mA at $f = 500$ MHz, 1 GHz, 2 GHz, and 4 GHz are as follows:

f (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
0.5	$0.761[-151^\circ]$	$0.025[31^\circ]$	$11.84[102^\circ]$	$0.429[-35^\circ]$
1	$0.770[-166^\circ]$	$0.029[35^\circ]$	$6.11[89^\circ]$	$0.365[-34^\circ]$
2	$0.760[-174^\circ]$	$0.040[44^\circ]$	$3.06[74^\circ]$	$0.364[-43^\circ]$
4	$0.756[-179^\circ]$	$0.064[48^\circ]$	$1.53[53^\circ]$	$0.423[-66^\circ]$

Determine the stability. If the transistor is potentially unstable at a given frequency, draw the input and output stability circles.

Solution. At $f = 500$ MHz it follows from (3.3.16) and (3.3.17) that $K = 0.482$ and $|\Delta| = 0.221[-123^\circ]$. Therefore, for this transistor at 500 MHz we have $K < 1$ and $|\Delta| < 1$. Since $K < 1$, the transistor is potentially unstable. Using (3.3.10), the center of the input stability circle is

$$C_s = \frac{(0.761[-151^\circ] - 0.221[-123^\circ](0.429[35^\circ]))^*}{(0.761)^2 - (0.221)^2} = 1.36[157.6^\circ]$$

and from (3.3.9) the radius is

$$r_s = \frac{|0.025[31^\circ](11.84[102^\circ])|}{(0.761)^2 - (0.221)^2} = 0.558$$

Similarly, from (3.3.8) and (3.3.7) the center and radius of the output stability circles are $C_L = 2.8[57.86^\circ]$ and $r_L = 2.18$.

At $f = 1$ GHz, we find that $K = 0.857$ and $|\Delta| = 0.173[-162.9^\circ]$. Therefore, since $K < 1$, the transistor is potentially unstable at 1 GHz. The center and radius of the input stability circle at 1 GHz are $C_s = 1.28[169^\circ]$ and $r_s = 0.315$; and for the output stability circle $C_L = 2.62[51.3^\circ]$ and $r_L = 1.71$.

At $f = 2$ GHz, we find that $K = 1.31$ and $|\Delta| = 0.174[160^\circ]$. Since $K > 1$ and $|\Delta| < 1$, it follows from (3.3.19) and (3.3.20) that the transistor is unconditionally stable at 2 GHz.

At $f = 4$ GHz, we find that $K = 1.535$ and $|\Delta| = 0.226[121^\circ]$. Therefore, since $K > 1$ and $|\Delta| < 1$, the transistor is unconditionally stable at 4 GHz.

The stability circles are plotted in Fig. 3.3.6 at $f = 500$ MHz and $f = 1$ GHz.

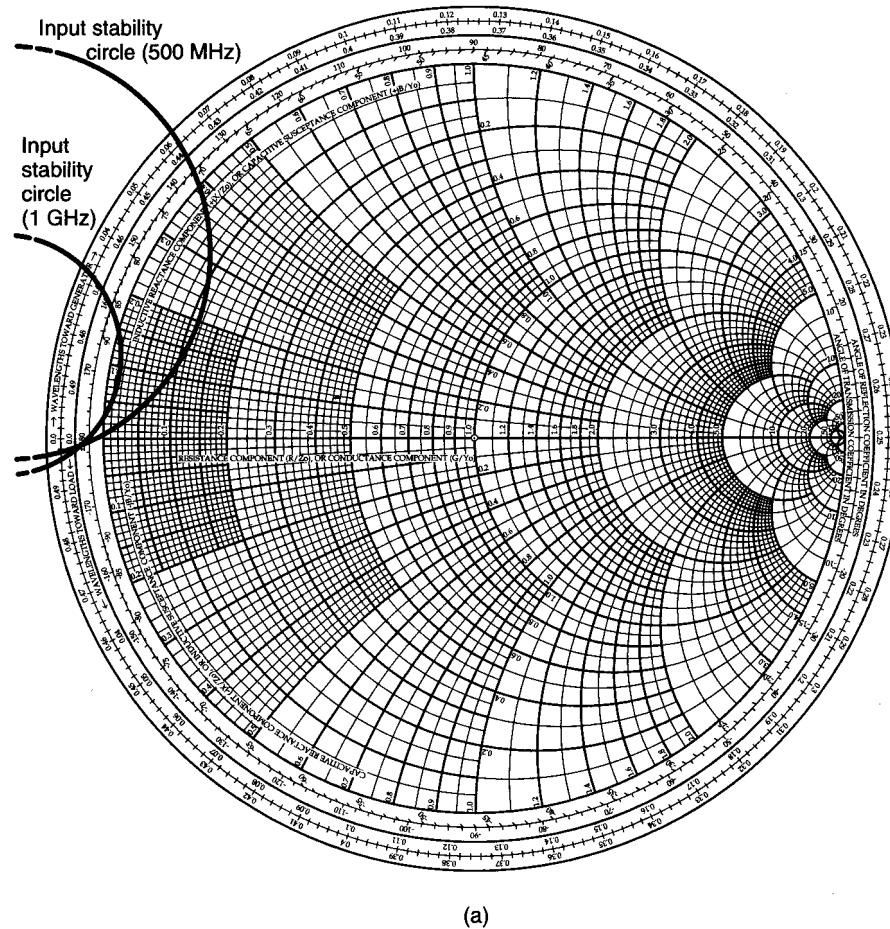
Manufacturers do not fabricate transistors with $K > 1$ and $|\Delta| > 1$. However, the addition of certain feedback networks or certain terminations to a transistor can create a potentially unstable two-port network with $K > 1$ and $|\Delta| > 1$. For example, a two-port network whose S parameters are $S_{11} = 0.75[-60^\circ]$, $S_{21} = 6[90^\circ]$, $S_{22} = 0.5[60^\circ]$, and $S_{12} = 0.3[70^\circ]$ has $K = 1.344$ and $|\Delta| = 2.156$. This two-port network has $K > 1$; however, it is potentially unstable because $|\Delta| > 1$. In fact, the input and output stability circles, from (3.3.7) to (3.3.10), are located at $C_s = 0.1[107.4^\circ]$, $r_s = 0.44$, $C_L = 0.26[-36.3^\circ]$, and $r_L = 0.41$, which can be drawn inside the Smith chart to show the unstable regions. The evaluation of B_1 in (3.3.21) gives $B_1 = -3.34$. Hence, when $|\Delta| > 1$ it follows that $B_1 < 0$. This shows that when the condition (3.3.20) for stability is not satisfied, neither is the condition (3.3.21).

A two-port network is said to be unilateral when $S_{12} = 0$. In a unilateral two-port network, $\Gamma_{IN} = S_{11}$ and $\Gamma_{OUT} = S_{22}$. Hence, we have unconditional stability if $|S_{11}| < 1$ and $|S_{22}| < 1$ for all passive source and load terminations. In fact, from (3.3.16) and (3.3.17), with $S_{12} = 0$ we have $K = \infty$ and $|\Delta| = S_{11}S_{22}$, and it follows from (3.3.19) that

$$1 - |S_{11}|^2 - |S_{22}|^2 + |S_{11}S_{22}|^2 > 0$$

or

$$(1 - |S_{11}|^2)(1 - |S_{22}|^2) > 0$$



(a)

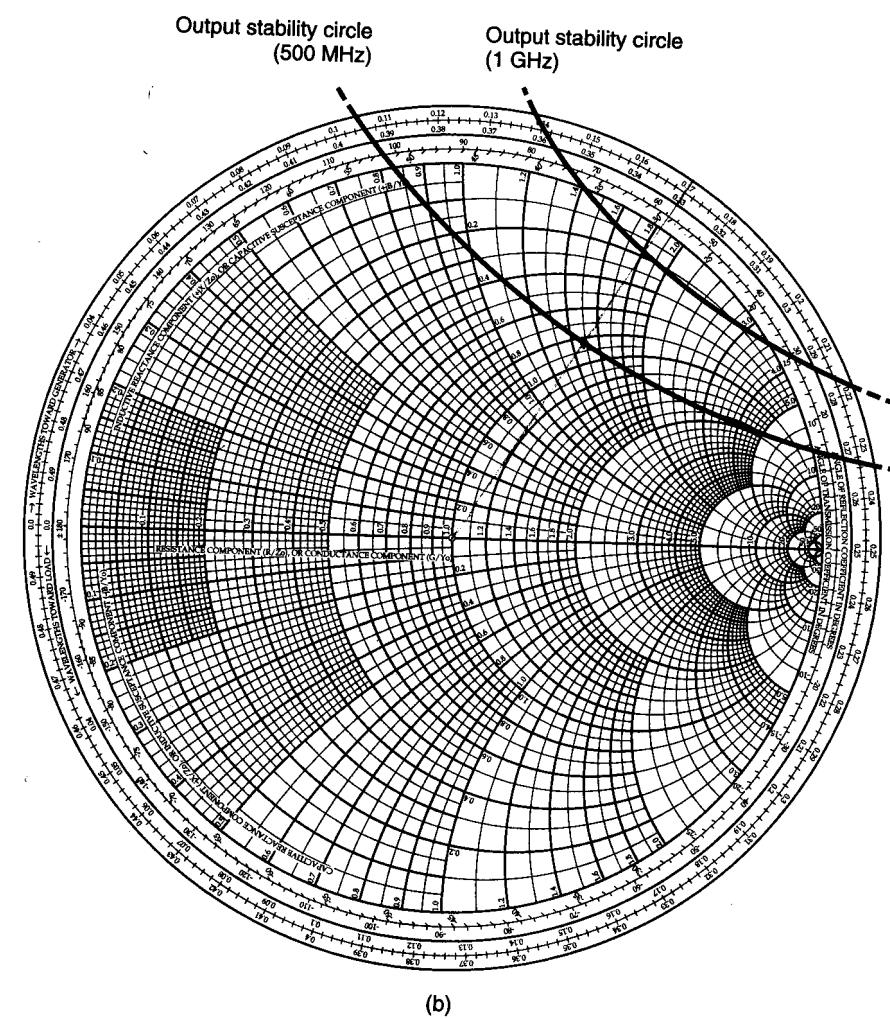
Figure 3.3.6 (a) Input stability circles for Example 3.3.1; (b) output stability circles.

The preceding inequality requires that $|S_{11}| < 1$ and $|S_{22}| < 1$ for unconditional stability of a unilateral two-port.

It is interesting to see what happens to the stability circles in the limit as $S_{12} \rightarrow 0$. The reader is referred to Problem 3.10 for this analysis.

In the potentially unstable situation illustrated in Figs. 3.3.3 and 3.3.4, the real part of the input and output impedances can be negative for some source and load reflection coefficients. In this case, selecting Γ_s and Γ_L in the stable region produces a stable operation.

Even when the selection of Γ_L and Γ_s produces $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$, the circuit can be made stable if the total input and output loop resistance in Fig. 3.3.1 is positive. In other words, the circuit is stable if

**Figure 3.3.6** Continued

and

$$\operatorname{Re}(Z_s + Z_{IN}) > 0$$

$$\operatorname{Re}(Z_L + Z_{OUT}) > 0$$

A potentially unstable transistor can be made unconditionally stable by either resistively loading the transistor or by adding negative feedback. These techniques are not recommended in narrowband amplifiers because of the resulting degradation in power gain, noise figure, and VSWRs. Narrowband

amplifier design with potentially unstable transistors is best done by the proper selection of Γ_s and Γ_L to ensure stability. On the other hand, the techniques are popular in the design of some broadband amplifiers in which the transistor is potentially unstable.

The following example illustrates how resistive loading can stabilize a potentially unstable transistor.

Example 3.3.2

The S parameters of a transistor at $f = 800$ MHz are

$$S_{11} = 0.65 \angle -95^\circ$$

$$S_{12} = 0.035 \angle 40^\circ$$

$$S_{21} = 5 \angle 115^\circ$$

$$S_{22} = 0.8 \angle -35^\circ$$

Determine the stability and show how resistive loading can stabilize the transistor.

Solution. From (3.3.16) and (3.3.17) we find that $K = 0.547$ and $\Delta = 0.504$ 249.6° . Since $K < 1$, the transistor is potentially unstable at $f = 800$ MHz.

The input and output stability circles are calculated using (3.3.7) to (3.3.10):

$$C_s = 1.79 \angle 122^\circ \quad C_L = 1.3 \angle 48^\circ$$

$$r_s = 1.04 \quad r_L = 0.45$$

Figure 3.3.7 shows the plot of the stability circles, together with the stable region.

For the input stability circle the Smith chart in Fig. 3.3.7 represents the Γ_s plane and for the output stability circle the Γ_L plane. It can be seen that a series resistor with the input of approximately 9Ω assures stability at the input (see Fig. 3.3.8). The series addition of a 9Ω resistor produces an impedance Z_s' equal to $Z_s + 9\Omega$. For any passive termination Z_s' , the real part of Z_s will be greater than 9Ω . Therefore, its associated reflection coefficient Γ_s will always be in the stable region in Fig. 3.3.7.

Also, a shunt resistor with the input of approximately $0.7/50 = 14 \text{ mS}$ (or 71.5Ω) produces stability at the input. Looking at the output stability circle, it follows that either a series resistor of approximately 29Ω or a shunt resistor of approximately 500Ω at the output produces stability at the output. The four choices of resistive loading are shown in Fig. 3.3.9. Usually, stabilizing one port of a transistor results in an unconditionally stable device.

All four choices of resistive loading affect the gain performance of the amplifier. In addition, from a practical point of view, resistive loading at the input (as shown in Figs. 3.3.9a and 3.3.9b) is not used because it produces a significant deterioration in the noise performance of the amplifier (see Chapter 4).

In some potentially unstable designs of broadband amplifiers, the shunt resistor loading at the output, as shown in Fig. 3.3.9d, produces a trade-off between gain and stability that is quite acceptable, resulting in a stable two-port with reasonable gain over a wide bandwidth (see Example 4.4.3).

For the stabilized shunt resistor configuration in Fig. 3.3.9d (i.e., with a 500Ω shunt resistor), the resulting S parameters are

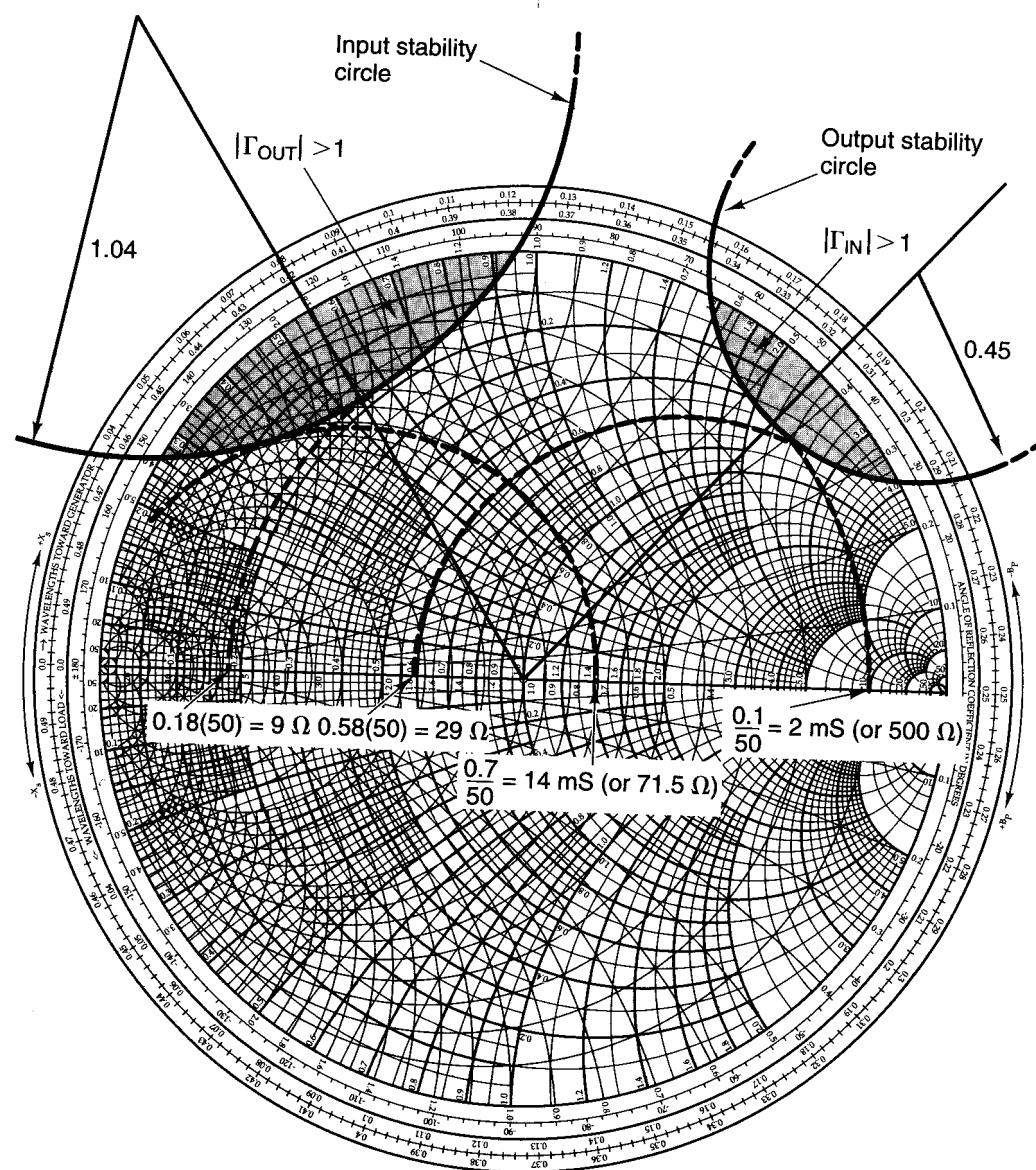


Figure 3.3.7 Input and output stability circles.

$$S_{11} = 0.65 \angle -94^\circ$$

$$S_{12} = 0.032 \angle 41.2^\circ$$

$$S_{21} = 4.62 \angle 116.2^\circ$$

$$S_{22} = 0.66 \angle -36^\circ$$

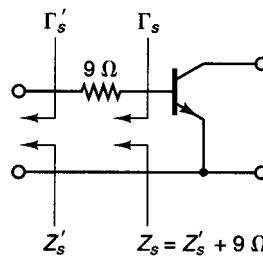


Figure 3.3.8 Series resistive loading of the transistor at the input port.

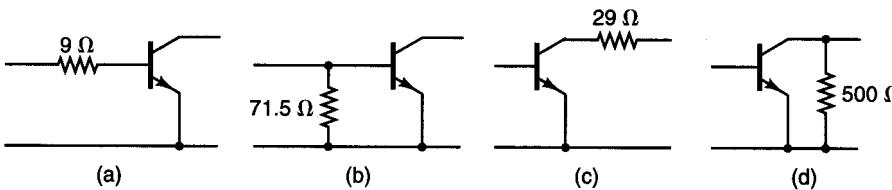


Figure 3.3.9 Four types of resistive loading to improve stability.

and from (3.3.16) and (3.3.17) $K = 1.04$ and $\Delta = 0.409|250.13^\circ|$, which show that the stabilized network in Fig. 3.3.9d is unconditionally stable at $f = 800$ MHz.

Negative feedback can be used to stabilize a transistor by neutralizing S_{12} —that is, by making $S_{12} = 0$. However, this is not commonly done. In a broadband amplifier design using a potentially unstable transistor, a common procedure is to use resistive loading to stabilize the transistor and negative feedback to provide the proper ac performance—that is, to provide constant gain and low input and output VSWR.

3.4 CONSTANT-GAIN CIRCLES: UNILATERAL CASE

A two-port network is unilateral when $S_{12} = 0$. In a unilateral transistor, $\Gamma_{\text{IN}} = S_{11}$, $\Gamma_{\text{OUT}} = S_{22}$, and the unilateral transducer power gain from (3.2.1) and (3.2.2), called G_{TU} , is given by

$$G_{TU} = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.4.1)$$

The first term in (3.4.1) depends on the S_{11} parameter of the transistor and the source reflection coefficient. The second term, $|S_{21}|^2$, depends on the transistor scattering parameter S_{21} ; and the third term depends on the S_{22} parameter of the transistor and the load reflection coefficient. We can think of (3.4.1) as being composed of three distinct and independent gain terms. Therefore, we can write (3.4.1) in the form

$$G_{TU} = G_s G_o G_L \quad (3.4.2)$$

where

$$G_s = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} \quad (3.4.3)$$

$$G_o = |S_{21}|^2 \quad (3.4.4)$$

$$G_L = \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.4.5)$$

and the microwave amplifier can be represented by three different gain (or loss) blocks, as shown in Fig. 3.4.1. The input matching network determines Γ_s and therefore the value of G_s according to (3.4.3); the transistor gain is $G_o = |S_{21}|^2$; and the output matching network determines Γ_L and therefore the value of G_L according to (3.4.5).

The terms G_s and G_L represent the gain or loss produced by the matching or mismatching of the input or output circuits, respectively. The term G_o affects the degree of matching or mismatching between Γ_s and S_{11} . Although the G_s block is made up of passive components, it can either have a gain contribution greater than unity or a loss. The reason we usually refer to G_s as a gain block is that there is an intrinsic mismatch loss between Z_o , the matching network, and S_{11} (i.e., between Γ_s and S_{11}). Therefore, decreasing the mismatch loss can be thought of as providing a gain. Similarly, the term G_L affects the output matching and can be thought of as the output gain block. The term G_o is related to the device and is equal to $|S_{21}|^2$. In terms of decibels, we can write from (3.4.2) to (3.4.5)

$$G_{TU}(\text{dB}) = G_s(\text{dB}) + G_o(\text{dB}) + G_L(\text{dB})$$

If we optimize Γ_s and Γ_L to provide maximum gain in G_s and G_L , we refer to the gain as the maximum unilateral transducer power gain, called $G_{TU,\max}$. For a unilateral unconditional stable transistor (i.e., for $|S_{11}| < 1$ and $|S_{22}| < 1$), the maximum values of G_s and G_L are obtained when (see Problem 3.15).

$$\Gamma_s = S_{11}^*$$

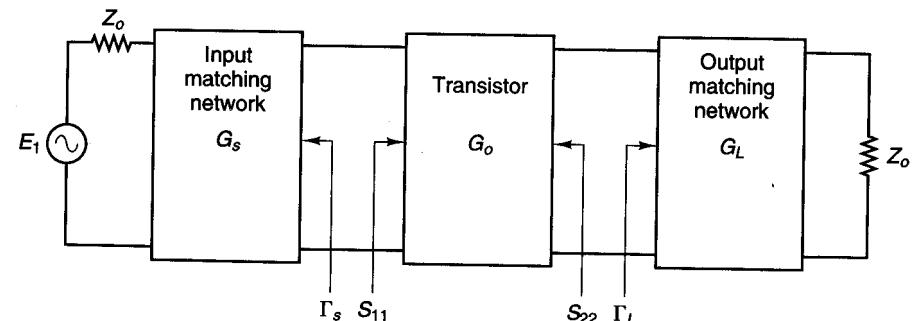


Figure 3.4.1 Unilateral transducer power gain block diagram.

and

$$\Gamma_L = S_{22}^*$$

Therefore, from (3.4.3) and (3.4.5) we obtain

$$G_{s,\max} = \frac{1}{1 - |S_{11}|^2}$$

$$G_{L,\max} = \frac{1}{1 - |S_{22}|^2}$$

and (3.4.2) gives

$$\begin{aligned} G_{TU,\max} &= G_{s,\max} G_o G_{L,\max} \\ &= \frac{1}{1 - |S_{11}|^2} |S_{21}|^2 \frac{1}{1 - |S_{22}|^2} \end{aligned} \quad (3.4.6)$$

The appropriate block diagram for (3.4.6) is shown in Fig. 3.4.2.

Observing that in the unilateral case $\Gamma_{IN} = S_{11}$ and $\Gamma_{OUT} = S_{22}$, the maximum value of G_{TU} , which occurs when $\Gamma_s = S_{11}^* = \Gamma_{IN}^*$ and $\Gamma_L = S_{22}^* = \Gamma_{OUT}^*$, is equal to the maximum value of G_p and G_A [see (3.2.3) and (3.2.4)]. That is, $G_{TU,\max} = G_{pU,\max} = G_{AU,\max}$.

The unilateral transducer power gain is given by (3.4.1) or (3.4.2), and the maximum unilateral transducer power gain, obtained when $\Gamma_s = S_{11}^*$ and $\Gamma_L = S_{22}^*$, is given by (3.4.6). The expressions for G_s and G_L in (3.4.3) and (3.4.5) are similar in form and can be written in the general form

$$G_i = \frac{1 - |\Gamma_i|^2}{|1 - S_{ii}\Gamma_i|^2} \quad (3.4.7)$$

where $i = s$ with $ii = 11$ or $i = L$ with $ii = 22$. The design for a specific gain is based on (3.4.7).

Two cases must be considered in the analysis of (3.4.7): the unconditionally stable case, where $|S_{ii}| < 1$, and the potentially unstable case, where $|S_{ii}| > 1$.

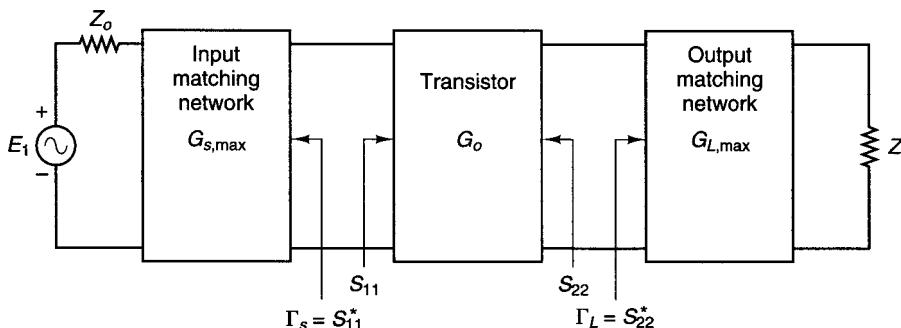


Figure 3.4.2 Maximum unilateral transducer power gain block diagram.

Unconditionally Stable Case, $|S_{ii}| < 1$

The maximum value of (3.4.7) is obtained when $\Gamma_i = S_{ii}^*$, and it is given by

$$G_{i,\max} = \frac{1}{1 - |S_{ii}|^2} \quad (3.4.8)$$

The terminations that produce $G_{i,\max}$ are called the *optimum terminations*.

From (3.4.7), G_i has a minimum value of zero when $|\Gamma_i| = 1$. Other values of Γ_i produce values of G_i between zero and $G_{i,\max}$. That is,

$$0 \leq G_i \leq G_{i,\max}$$

The values of Γ_i that produce a constant gain G_i will be shown to lie in a circle in the Smith chart. These circles are called *constant G_i circles* (i.e., for $i = s$ the circles are *constant G_s circles* and for $i = L$ the circles are *constant G_L circles*).

Define the normalized gain factor as

$$g_i = \frac{G_i}{G_{i,\max}} = G_i(1 - |S_{ii}|^2) = \frac{1 - |\Gamma_i|^2}{|1 - S_{ii}\Gamma_i|^2}(1 - |S_{ii}|^2) \quad (3.4.9)$$

such that

$$0 \leq g_i \leq 1$$

In Appendix D it is shown that the values of Γ_i that produce a constant value of g_i in (3.4.9) lie in a circle whose equation is

$$|\Gamma_i - C_{g_i}| = r_{g_i} \quad (3.4.10)$$

where the center of the circle is given by

$$C_{g_i} = \frac{g_i S_{ii}^*}{1 - |S_{ii}|^2(1 - g_i)} \quad (3.4.11)$$

and the radius is

$$r_{g_i} = \frac{\sqrt{1 - g_i}(1 - |S_{ii}|^2)}{1 - |S_{ii}|^2(1 - g_i)} \quad (3.4.12)$$

Each constant value of g_i generates a new constant G_i circle. Equations (3.4.11) and (3.4.12) can be used to generate the constant G_s circles and the constant G_L circles.

Figure 3.4.3 illustrates a constant G_i circle. The distance from the origin to the center of a constant G_i circle is given by $|C_{g_i}|$ in (3.4.11), and the angle of inclination, α_i , is equal to the phase of C_{g_i} (which is the phase of S_{ii}^*).

It is observed that when $g_i = 1$ (i.e., when $G_i = G_{i,\max}$), (3.4.12) gives $r_{g_i} = 0$ and (3.4.11) gives $C_{g_i} = S_{ii}^*$. Therefore, the constant G_i circle for maximum gain is represented by a point, located at S_{ii}^* .

In conclusion, the procedure for drawing the constant G_i circles in the Z Smith chart is as follows:

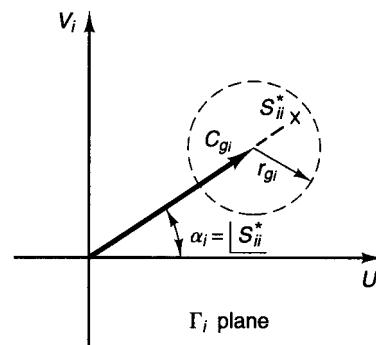


Figure 3.4.3 A constant G_i circle in the Smith chart.

1. Locate S_{ii}^* and draw a line from the origin to S_{ii}^* . At S_{ii}^* , the gain is $G_{i,\max}$ and is given by (3.4.8).
2. Determine the values of G_i , where $0 \leq G_i \leq G_{i,\max}$, for which the constant G_i circles are to be drawn, and calculate the corresponding values of $g_i = G_i/G_{i,\max}$.
3. From (3.4.11), determine the values of C_{gi} for each g_i .
4. From (3.4.12), determine the values of r_{gi} for each g_i .

The 0-dB circle ($G_i = 1$) always passes through the origin of the Smith chart. This is not a coincidence. In fact, $G_i = 1$ occurs when $\Gamma_i = 0$, and from (3.4.9)

$$g_{i,0 \text{ dB}} = 1 - |S_{ii}|^2$$

Then, from (3.4.11) and (3.4.12),

$$r_{g_i,0 \text{ dB}} = |C_{g_i,0 \text{ dB}}| = \frac{|S_{ii}|}{1 + |S_{ii}|^2}$$

which shows that the radius and the distance from the origin to the center of the 0-dB constant G_i circle are identical.

A typical set of constant G_s circles is calculated in the following example and shown in Fig. 3.4.4.

Example 3.4.1

The S parameters of a BJT measured at $V_{CE} = 10 \text{ V}$, $I_C = 30 \text{ mA}$, and $f = 1 \text{ GHz}$, in a $50\text{-}\Omega$ system, are

$$S_{11} = 0.73 \angle 175^\circ$$

$$S_{12} = 0$$

$$S_{21} = 4.45 \angle 65^\circ$$

$$S_{22} = 0.21 \angle -80^\circ$$

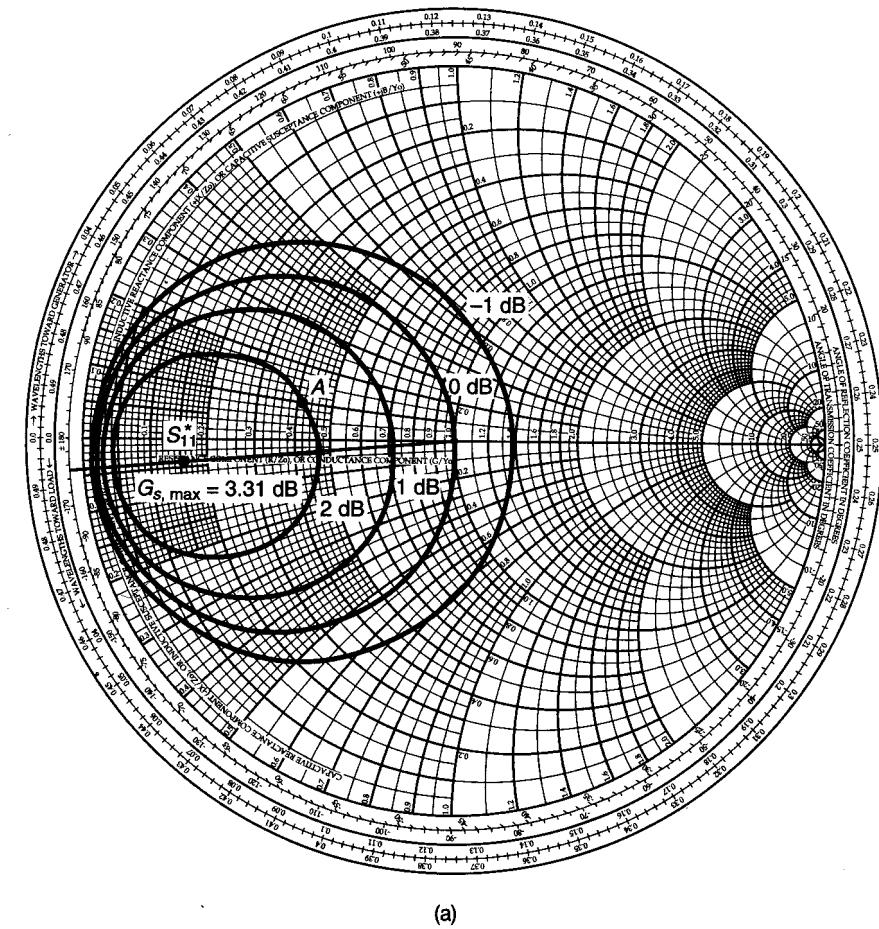


Figure 3.4.4 (a) Constant-gain circles for $G_s = 2, 1, 0$, and -1 dB ; (b) calculations of constant-gain circles.

- (a) Calculate the optimum terminations.
- (b) Calculate $G_{s,\max}$, $G_{L,\max}$, and $G_{TU,\max}$ in decibels.
- (c) Draw several G_s constant-gain circles.
- (d) Design the input matching network for $G_s = 2$ dB.

Solution. (a) The optimum terminations are

$$\Gamma_s = S_{11}^* = 0.73 \angle -175^\circ$$

and

$$\Gamma_L = S_{22}^* = 0.21 \angle 80^\circ$$

Using the Smith chart, the impedances associated with Γ_s and Γ_L are $Z_s = 50(0.152 - j0.047) = 7.6 - j2.35 \Omega$ and $Z_L = 50(0.97 + j0.43) = 48.5 + j21.5 \Omega$.

(b) From (3.4.8) we find that

$$G_{s,\max} = \frac{1}{1 - |S_{11}|^2} = 2.141 \quad \text{or} \quad 3.31 \text{ dB}$$

$$G_{L,\max} = \frac{1}{1 - |S_{22}|^2} = 1.046 \quad \text{or} \quad 0.195 \text{ dB}$$

Since

$$G_o = |S_{21}|^2 = 19.8 \quad \text{or} \quad 12.97 \text{ dB}$$

then

$$G_{TU,\max}(\text{dB}) = 3.31 + 12.97 + 0.195 = 16.47 \text{ dB}$$

(c) Since $G_{s,\max} = 3.31$ dB, constant-gain circles at 2, 1, 0, and -1 dB are drawn in Fig. 3.4.4a. The necessary calculations are given in Fig. 3.4.4b.

(d) In an amplifier design using this transistor, we observe that the output gain block provides little gain (i.e., $G_{L,\max} = 0.195$ dB); therefore, the output matching network is designed to present the optimum termination $\Gamma_L = 0.21 \angle 80^\circ$.

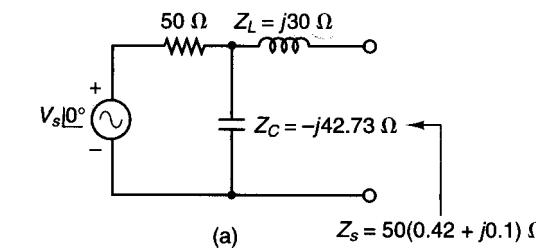
Any Γ_s along the $G_s = 2$ dB circle provides the constant gain. Selecting Γ_s at point A (i.e., $\Gamma_s = 0.413 \angle 166^\circ$ or $z_s = 0.42 + j0.1$) in Fig. 3.4.4a results in the input matching network shown in Fig. 3.4.5a. The details of the matching network design are shown in Fig. 3.4.5b. With $\Gamma_s = 0.413 \angle 166^\circ$ and $\Gamma_L = 0.21 \angle 80^\circ$, it follows that $G_s = 2$ dB and $G_L = G_{L,\max} = 0.195$ dB; hence, the transducer gain of the amplifier is G_{TU} (dB) = $2 + 12.97 + 0.195 = 15.16$ dB.

Potentially Unstable Case, $|S_{ii}| > 1$

In this case $|S_{ii}| > 1$ and it is possible for a passive termination to produce an infinite value of G_i . The infinite value of G_i in (3.4.7) is produced by the critical value of Γ_i , called $\Gamma_{i,c}$, given by

$$\Gamma_{i,c} = \frac{1}{S_{ii}} \quad (3.4.13)$$

Equation (3.4.13) basically states that the real part of the impedance associated with $\Gamma_{i,c}$ is equal to the magnitude of the negative resistance associated



$$Z_s = 50(0.42 + j0.1) \Omega$$

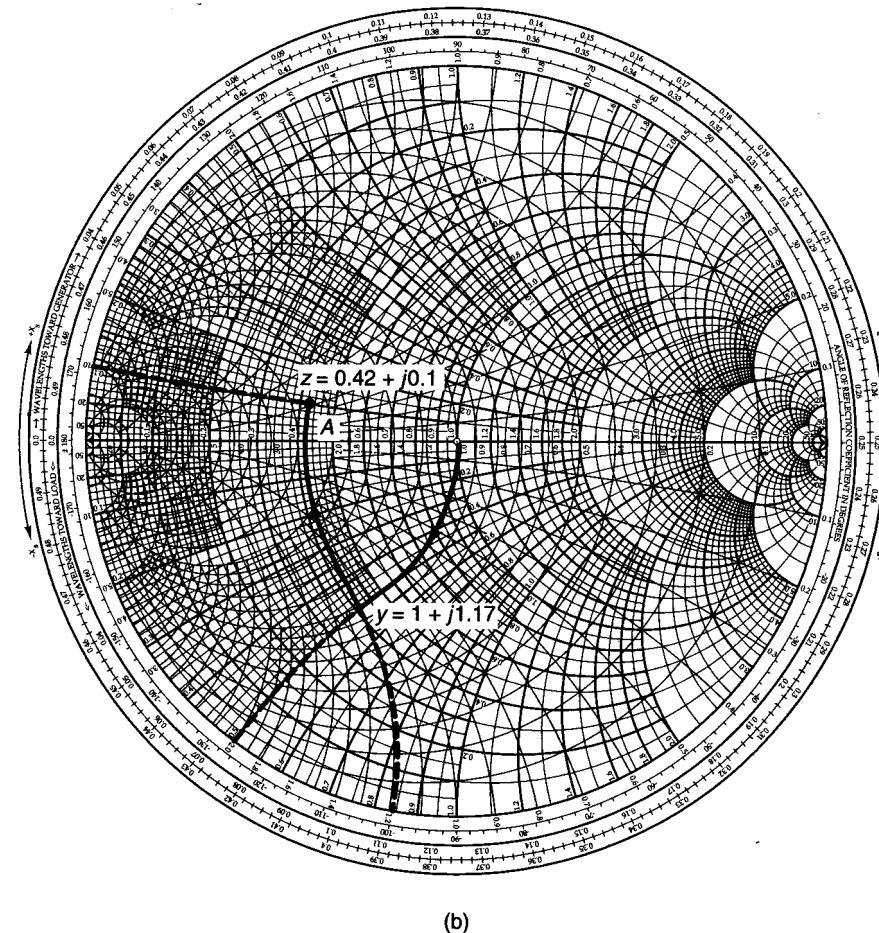


Figure 3.4.5 Input matching network for $G_s = 2$ dB.

with S_{ii} . Therefore, the total input or output loop resistance is zero, and oscillations will occur.

With g_i defined as in (3.4.9)—namely,

$$g_i = G_i(1 - |S_{ii}|^2) = \frac{1 - |\Gamma_i|^2}{|1 - S_{ii}\Gamma_i|^2} (1 - |S_{ii}|^2) \quad (3.4.14)$$

where now g_i can attain negative values because $|S_{ii}| > 1$ —the derivation of a constant G_i circle is identical to (3.4.10). Hence, with g_i given by (3.4.14), the center of a constant G_i circle is given by (3.4.11), and the radius is given by (3.4.12). The gain G_i is infinite at $\Gamma_i = \Gamma_{i,c} = 1/S_{ii}$. Since the argument of C_{gi} (i.e., $|S_{ii}^*$) is identical to the argument of $1/S_{ii}$, it follows that the centers of the circles are located along a line drawn from the origin to the point $1/S_{ii}$.

As discussed in Section 2.2, the negative resistance associated with S_{ii} , where $|S_{ii}| > 1$, can be calculated using the Smith chart by locating the point $1/S_{ii}^*$ and interpreting the resistance circles as being negative and the reactance circles as labeled.

To prevent oscillations in the input or output port Γ_i must be selected such that the real part of the termination impedance is larger than the magnitude of the negative resistance associated with the point $1/S_{ii}^*$. When a negative resistance occurs at the input, the stable region is that region where values of Γ_s produce a source impedance such that

$$\operatorname{Re}(Z_s) > |\operatorname{Re}(Z_{IN})|$$

Similarly, when a negative resistance occurs at the output, Γ_L is selected such that

$$\operatorname{Re}(Z_L) > |\operatorname{Re}(Z_{OUT})|$$

A typical construction is illustrated in Fig. 3.4.6, where the critical value of Γ_s (i.e., $\Gamma_{s,c} = 1/S_{11}$) and two constant G_s circles are shown.

Example 3.4.2

The S parameters of a GaAs FET measured at $V_{DS} = 5$ V, $I_{DS} = 10$ mA, and $f = 1$ GHz in a 50Ω system are

$$S_{11} = 2.27 \angle -120^\circ$$

$$S_{12} = 0$$

$$S_{21} = 4 \angle 50^\circ$$

$$S_{22} = 0.6 \angle -80^\circ$$

- (a) Calculate the input impedance and the optimum output termination.
- (b) Determine the unstable region in the Smith chart and construct constant-gain circles for $G_s = 5$ dB and $G_s = 3$ dB.
- (c) Design the input matching network for $G_s = 3$ dB with the greatest degree of stability.
- (d) Determine G_{TU} in decibels.

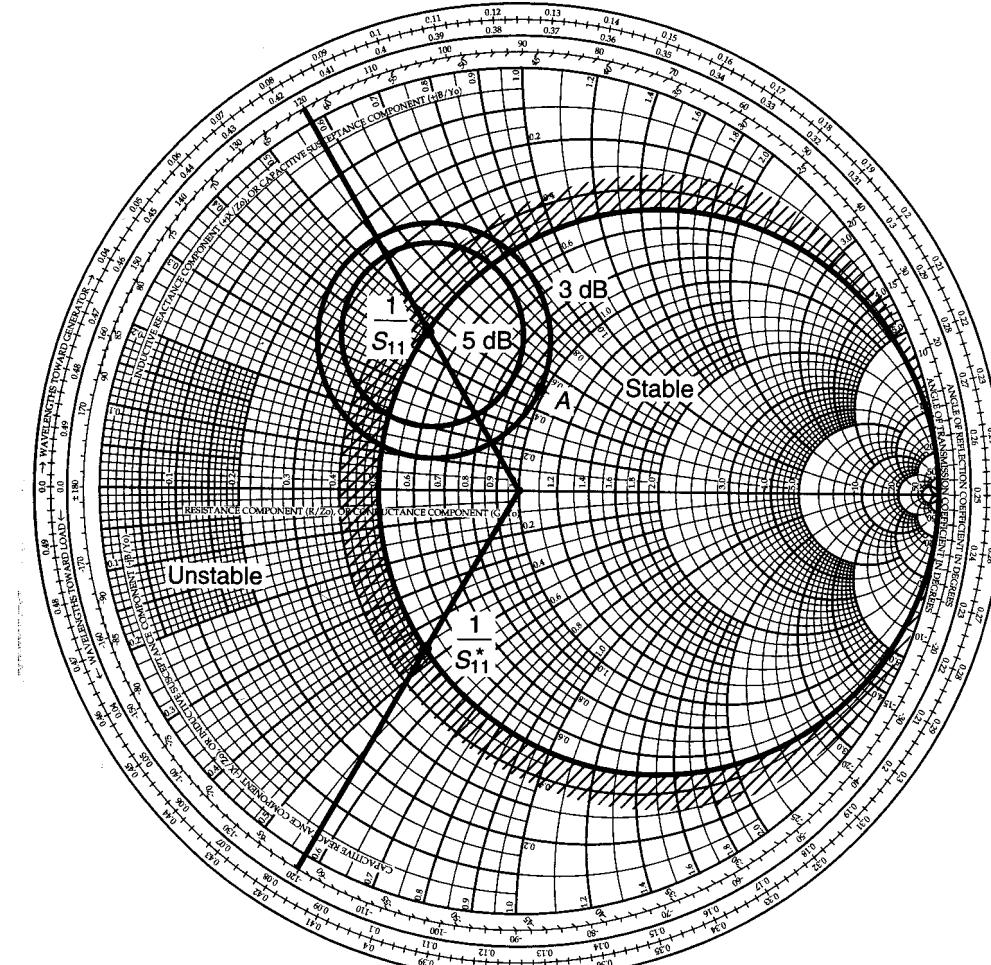


Figure 3.4.6 Stable and unstable regions and constant G_s circles for Example 3.4.2.

Solution. (a) The input impedance is obtained from the Smith chart at the point $1/S_{11}^* = 0.44 \angle -120^\circ$ (see Fig. 3.4.6)—namely,

$$Z_{IN} = 50(-0.5 - j0.46) = -25 - j23 \Omega$$

The optimum termination for G_L is

$$\Gamma_L = S_{22}^* = 0.6 \angle 80^\circ$$

The impedance associated with Γ_L is obtained from the Smith chart as $Z_L = 50(0.56 + j1.03) = 28 + j51.5 \Omega$.

(b) The unstable region is where $\operatorname{Re}(Z_s) < |\operatorname{Re}(Z_{IN})|$. The unstable region is marked in Fig. 3.4.6.

In order to construct the constant-gain circle for $G_s = 5 \text{ dB}$, we first locate the point $1/S_{11}$ in Fig. 3.4.6. Then, from (3.4.14), (3.4.11), and (3.4.12), we find that

$$g_s = 3.16[1 - (2.27)^2] = -13.123$$

$$r_{g_s} = \frac{\sqrt{1 + 13.123}}{1 - (2.27)^2(1 + 13.123)} = 0.217$$

and

$$|C_{g_s}| = \frac{-13.123(2.27)}{1 - (2.27)^2(1 + 13.123)} = 0.415$$

The $G_s = 5 \text{ dB}$ circle is drawn in Fig. 3.4.6. Similarly, for the $G_s = 3 \text{ dB}$ circle, we find that $g_s = -8.286$, $|C_{g_s}| = 0.401$, and $r_{g_s} = 0.27$.

(c) In order to obtain the greatest degree of stability, we select Γ_s on the $G_s = 3 \text{ dB}$ circle such that it has the largest positive real part. That is, Γ_s is selected at point A in Fig. 3.4.6—namely,

$$\Gamma_s = 0.245 \underline{|79^\circ|}$$

or

$$Z_s = 50(0.97 + j0.5) = 48.5 + j25 \Omega$$

Since the input loop resistance is $48.5 - 25 = 23.5 \Omega$, the input port is stable.

(d) Since $G_s = 3 \text{ dB}$,

$$G_{L,\max} = \frac{1}{1 - |S_{22}|^2} = \frac{1}{1 - (0.6)^2} = 1.562 \quad \text{or} \quad 1.94 \text{ dB}$$

and

$$G_o = |S_{21}|^2 = (4)^2 = 16 \quad \text{or} \quad 12.04 \text{ dB}$$

and the unilateral transducer gain is

$$G_{TU}(\text{dB}) = 3 + 12.04 + 1.94 = 16.98 \text{ dB}$$

3.5 UNILATERAL FIGURE OF MERIT

When S_{12} can be set equal to zero, the design procedure is much simpler. In order to determine the error involved in assuming $S_{12} = 0$, we form the magnitude ratio of G_T and G_{TU} from (2.6.13) and (3.4.1)—namely,

$$\frac{G_T}{G_{TU}} = \frac{1}{|1 - X|^2} \quad (3.5.1)$$

where

$$X = \frac{S_{12}S_{21}\Gamma_s\Gamma_L}{(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L)}$$

From (3.5.1) the ratio of the transducer power gain to the unilateral transducer power gain is bounded by

$$\frac{1}{(1 + |X|)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1 - |X|)^2}$$

When $\Gamma_s = S_{11}^*$ and $\Gamma_L = S_{22}^*$, G_{TU} has a maximum value and, in this case, the maximum error introduced when using G_{TU} is bounded by

$$\frac{1}{(1 + U)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1 - U)^2} \quad (3.5.2)$$

where

$$U = \frac{|S_{12}||S_{21}||S_{11}||S_{22}|}{(1 - |S_{11}|^2)(1 - |S_{22}|^2)} \quad (3.5.3)$$

is known as the *unilateral figure of merit*.

The value of U varies with frequency because of its dependence on the S parameters. A typical variation of U with frequency is shown in Fig. 3.5.1. In this case, the maximum value of U occurs at 100 MHz and 1 GHz and is given by $U = -15 \text{ dB}$ or $U = 0.03$. Therefore, from (3.5.2),

$$\frac{1}{(1 + 0.03)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1 - 0.03)^2}$$

or, in decibels,

$$-0.26 \text{ dB} < \frac{G_T}{G_{TU}} < 0.26 \text{ dB}$$

and the maximum error is $\pm 0.26 \text{ dB}$ at 100 MHz and 1 GHz. In some designs this error is small enough to justify the unilateral assumption.

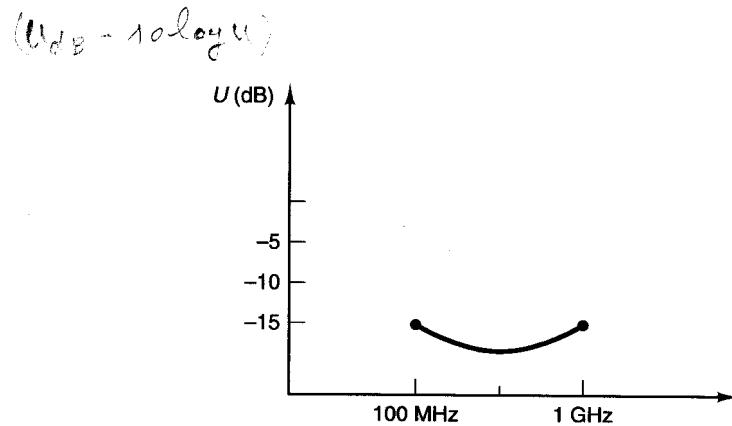


Figure 3.5.1 Frequency dependence of the unilateral figure of merit.

3.6 SIMULTANEOUS CONJUGATE MATCH: BILATERAL CASE

When $S_{12} \neq 0$ and the unilateral assumption cannot be made, the input and output reflection coefficients are given by (3.2.5) and (3.2.6), respectively. The conditions required to obtain maximum transducer power gain are

$$\Gamma_s = \Gamma_{IN}^* \quad (3.6.1)$$

and

$$\Gamma_L = \Gamma_{OUT}^* \quad (3.6.2)$$

These conditions are illustrated in Fig. 3.6.1 and are referred to as the simultaneous conjugate match conditions. When the input and output are matched, it follows that $(VSWR)_{in} = (VSWR)_{out} = 1$.

From (3.2.5), (3.2.6), (3.6.1), and (3.6.2), we can write

$$\Gamma_s^* = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \quad (3.6.3)$$

and

$$\Gamma_L^* = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \quad (3.6.4)$$

Solving (3.6.3) and (3.6.4) simultaneously gives the values Γ_s and Γ_L required for a simultaneous conjugate match. Calling these values Γ_{MS} and Γ_{ML} , we obtain

$$\Gamma_{MS} = \frac{B_1 \pm \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \quad (3.6.5)$$

$$\Gamma_{ML} = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \quad (3.6.6)$$

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 \quad (3.6.7)$$

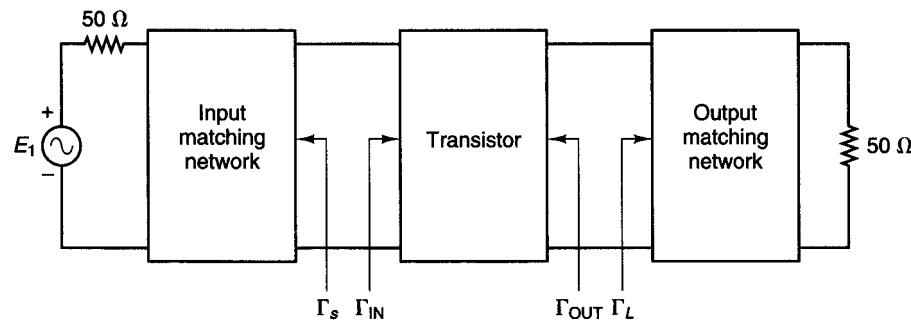


Figure 3.6.1 Simultaneous conjugate match exists when $\Gamma_s = \Gamma_{IN}^*$ and $\Gamma_L = \Gamma_{OUT}^*$.

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2 \quad (3.6.8)$$

$$C_1 = S_{11} - \Delta S_{22}^*$$

$$C_2 = S_{22} - \Delta S_{11}^*$$

In what follows, we will show that for an unconditionally stable two-port network, the solutions with a minus sign in (3.6.5) and (3.6.6) are the useful ones [3.2, 3.3].

If $|B_1/2C_1| > 1$ and $B_1 > 0$ in (3.6.5), the solution with the minus sign produces $|\Gamma_{MS}| < 1$ and the solution with the plus sign produces $|\Gamma_{MS}| > 1$. If $|B_1/2C_1| > 1$ with $B_1 < 0$ in (3.6.5), the solution with the plus sign produces $|\Gamma_{MS}| < 1$ and the solution with the minus sign produces $|\Gamma_{MS}| > 1$. Similar considerations apply to (3.6.6). The proofs for the previous statements are given in Appendix E, Section E.1.

Since it can be shown that $|B_i/2C_i|^2 > 1$ ($i = 1$ or 2) is similar to $K^2 > 1$, it follows that the condition $|B_i/2C_i| > 1$ is similar to $|K| > 1$ (see Appendix E, Section E.2). Therefore, if $|K| > 1$ with K positive, one solution of (3.6.5) and (3.6.6) has a magnitude less than 1, and the other solution has a magnitude greater than 1. In fact, for $K > 1$ and $B_i > 0$, the solutions with the minus sign have magnitudes less than 1. The analysis for $|K| > 1$ with K negative is left as an exercise (see Problem 3.20); it follows that for $K < -1$ a simultaneous conjugate match does not exist.

Associated with Γ_{MS} and Γ_{ML} are a source and a load impedance. The real parts of these impedances are positive if $|\Gamma_{MS}| < 1$ and $|\Gamma_{ML}| < 1$. From the previous considerations, we conclude that in terms of K , the condition that a two-port network can be simultaneously matched with $|\Gamma_{MS}| < 1$ and $|\Gamma_{ML}| < 1$ is

$$K > 1$$

The condition $K > 1$ is only a necessary condition for unconditional stability. Therefore, a simultaneous conjugate match having unconditional stability is possible if $K > 1$ and $|\Delta| < 1$. Since $|\Delta| < 1$ implies that $B_1 > 0$ and $B_2 > 0$, the minus signs must be used in (3.6.5) and (3.6.6) when calculating the simultaneous conjugate match for an unconditionally stable two-port network.

In what follows, any reference to a simultaneous conjugate match assumes that the two-port network is unconditionally stable. In a potentially unstable situation, the design procedure is best done in terms of G_p or G_A (see Section 3.7).

The maximum transducer power gain, under simultaneous conjugate match conditions, is obtained from (3.2.1) with $\Gamma_s = \Gamma_{IN}^* = \Gamma_{MS}$ and $\Gamma_L = \Gamma_{OUT}^* = \Gamma_{ML}$. Thus, we obtain

$$G_{T,max} = \frac{1}{1 - |\Gamma_{MS}|^2} |S_{21}|^2 \frac{1 - |\Gamma_{ML}|^2}{|1 - S_{22}\Gamma_{ML}|^2} \quad (3.6.9)$$

Substituting (3.6.5) and (3.6.6) into (3.6.9), it can be shown (see Appendix F) that $G_{T,max}$ can be expressed in the form

$$G_{T,max} = \frac{|S_{21}|}{|S_{12}|} (K - \sqrt{K^2 - 1}) \quad (3.6.10)$$

Ave. \ominus dars $|\Gamma_{MS}| < 1$ $|\Gamma_{ML}| < 1$

Since under simultaneous conjugate match conditions $G_T = G_p = G_A$, it follows that $G_{T,\max} = G_{p,\max} = G_{A,\max}$.

The maximum stable gain is defined as the value of $G_{T,\max}$ when $K = 1$ —namely,

$$G_{\text{MSG}} = \frac{|S_{21}|}{|S_{12}|} \quad (3.6.11)$$

For a potentially unstable transistor, G_{MSG} is a figure of merit. Figure 3.6.2 illustrates a typical way in which $G_{T,\max} = G_{p,\max} = G_{A,\max}$ and G_{MSG} are given by a manufacturer. At the frequencies where the transistor is unconditionally stable, $G_{A,\max}$ is calculated and plotted in Fig. 3.6.2. In Fig. 3.6.2, $G_{A,\max}$ is denoted by MAG (maximum available gain). At the frequencies where the transistor is potentially unstable, G_{MSG} is plotted in Fig. 3.6.2. From Fig. 3.6.2 it is seen that the transistor is potentially unstable below 1.5 GHz since MSG (i.e., G_{MSG}) is given; and above 2 GHz the transistor is unconditionally stable since MAG is given.

A simultaneous conjugate match does not exist for $K < 1$. However, in a potentially unstable two-port network with $K > 1$ but $|\Delta| > 1$ (which is similar to $B_1 < 0$ and $B_2 < 0$), solutions to (3.6.5) and (3.6.6) using the plus sign produce $|\Gamma_{Ms}| < 1$ and $|\Gamma_{ML}| < 1$. In such a case (i.e., for $K > 1$ and $|\Delta| > 1$), the values of Γ_{Ms} and Γ_{ML} given by (3.6.5) and (3.6.6) using the plus sign result in a minimum value of G_T , and the input and output VSWR are unity. Substituting these values of Γ_{Ms} and Γ_{ML} into (3.6.9), the minimum value of G_T is given by

$$\text{Avec } \oplus \text{ dans } \Gamma_{Ms} \text{ et } \Gamma_{ML} \quad G_{T,\min} = \frac{|S_{21}|}{|S_{12}|} (K + \sqrt{K^2 - 1}) \quad (3.6.12)$$

Recall that in a potentially unstable situation the maximum value of G_T approaches infinity as Γ_s and Γ_L approach the unstable region. Therefore, the expression (3.6.12) gives the minimum value that G_T can have when $K > 1$ and

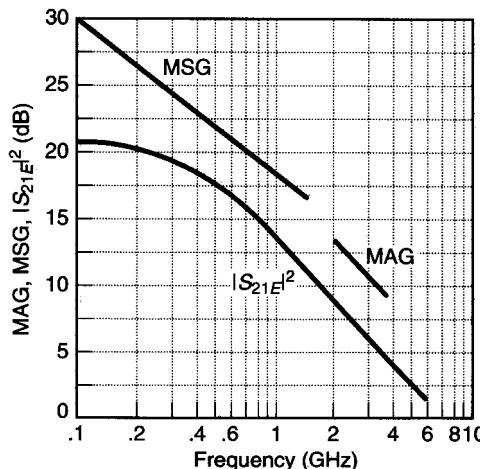


Figure 3.6.2 Typical MAG (i.e., $G_{A,\max}$), MSG (i.e., G_{MSG}), and $|S_{21E}|^2$ versus frequency at $V_{CE} = 18$ V and $I_C = 30$ mA for the HXTR-5103. (From HP Microwave and RF Designer's Catalog 1990-1991; courtesy of Hewlett Packard.)

$|\Delta| > 1$. Also, since the input and output ports are conjugate matched, it follows that $(\text{VSWR})_{\text{in}} = (\text{VSWR})_{\text{out}} = 1$.

Example 3.6.1

Design a microwave amplifier using a GaAs FET to operate $f = 6$ GHz with maximum transducer power gain. The transistor S parameters at the linear bias point, $V_{DS} = 4$ V and $I_{DS} = 0.5I_{DSS}$, are

$$\begin{aligned} S_{11} &= 0.641 \angle -171.3^\circ \\ S_{12} &= 0.057 \angle 16.3^\circ \\ S_{21} &= 2.058 \angle 28.5^\circ \\ S_{22} &= 0.572 \angle -95.7^\circ \end{aligned}$$

Solution. From (3.3.16) and (3.3.17), we obtain $K = 1.504$ and $\Delta = 0.3014 \angle 109.88^\circ$. Since $K > 1$ and $|\Delta| < 1$, the GaAs FET is unconditionally stable.

It is of interest to check if the amplifier can be considered unilateral. From (3.5.3), $U = 0.1085$, and from (3.5.2),

$$-0.89 \text{ dB} < \frac{G_T}{G_{TU}} < 1 \text{ dB}$$

The preceding inequality shows that S_{12} cannot be neglected.

The reflection coefficients for a simultaneous conjugate match are calculated from (3.6.5) and (3.6.6) (using the minus sign) as follows:

$$\begin{aligned} B_1 &= 0.9928 \\ B_2 &= 0.8255 \\ C_1 &= 0.4786 \angle -177.3^\circ \\ C_2 &= 0.3911 \angle -103.9^\circ \\ \Gamma_{Ms} &= 0.762 \angle 177.3^\circ \end{aligned}$$

and

$$\Gamma_{ML} = 0.718 \angle 103.9^\circ$$

The maximum transducer power gain, from (3.6.10), is

$$G_{T,\max} = \frac{2.058}{0.057} (1.504 - \sqrt{(1.504)^2 - 1}) = 13.74 \quad \text{or} \quad 11.38 \text{ dB}$$

The design of the matching networks using microstrip lines is illustrated in Fig. 3.6.3, where the admittances associated with Γ_{Ms} and Γ_{ML} are

$$Y_{Ms} = \frac{7.2 - j1.23}{50} = (144 - j24.6) \times 10^{-3} \text{ S}$$

and

$$Y_{ML} = \frac{0.414 - j1.19}{50} = (8.28 - j23.8) \times 10^{-3} \text{ S}$$

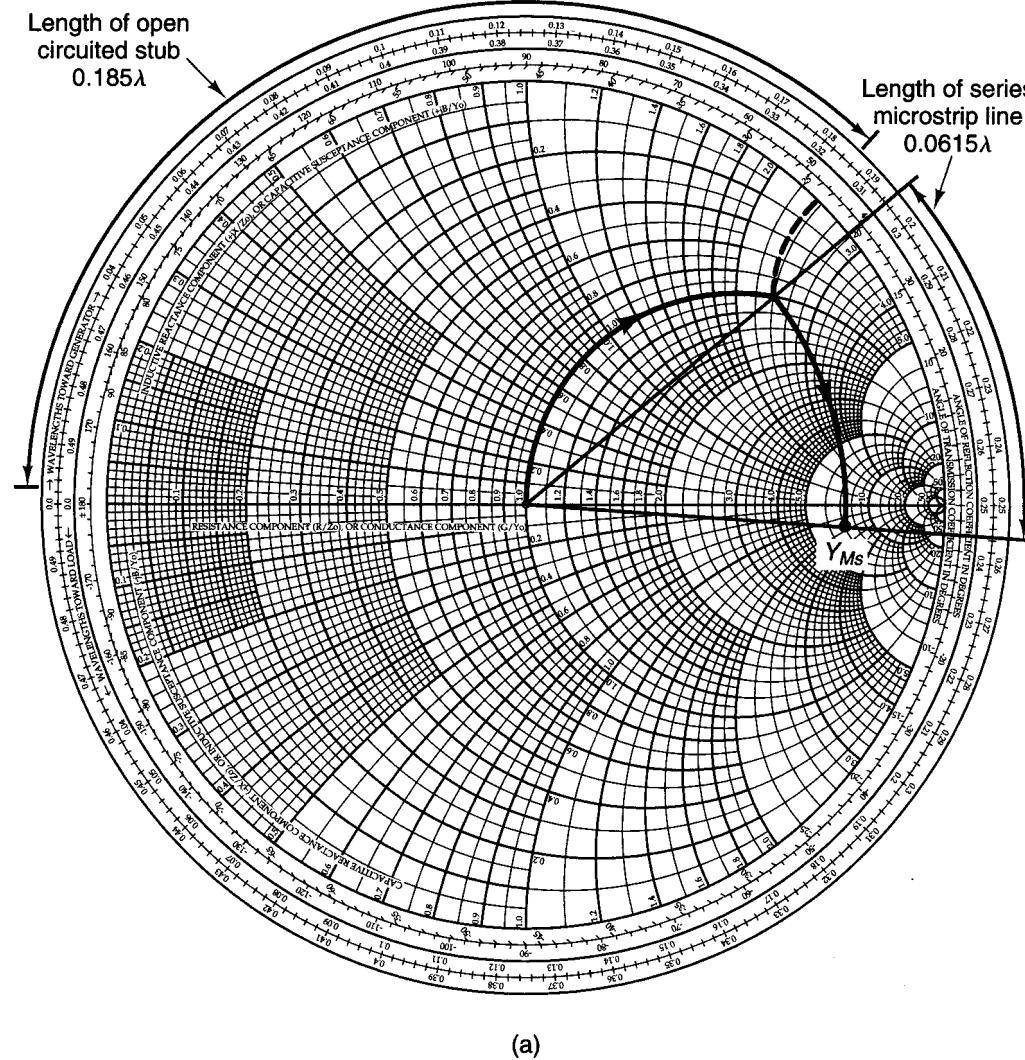


Figure 3.6.3 (a) Design of the input matching network; (b) design of the output matching network.

The input matching network can be designed with an open shunt stub of length 0.185λ and a series transmission line of length 0.0615λ . The output matching network is designed with an open shunt stub of length 0.176λ and a series transmission line of length 0.169λ .

The ac amplifier schematic is shown in Fig. 3.6.4. Using Duroid® ($\epsilon_r = 2.23$, $h = 0.7874$ mm) for the board material, we find that $W = 2.41$ mm for a characteristic impedance of 50Ω , $\epsilon_{eff} = 1.91$, and $\lambda = 0.7236\lambda_0$, where $\lambda_0 = 5$ cm at $f = 6$ GHz. The microstrip lengths at $f = 6$ GHz are

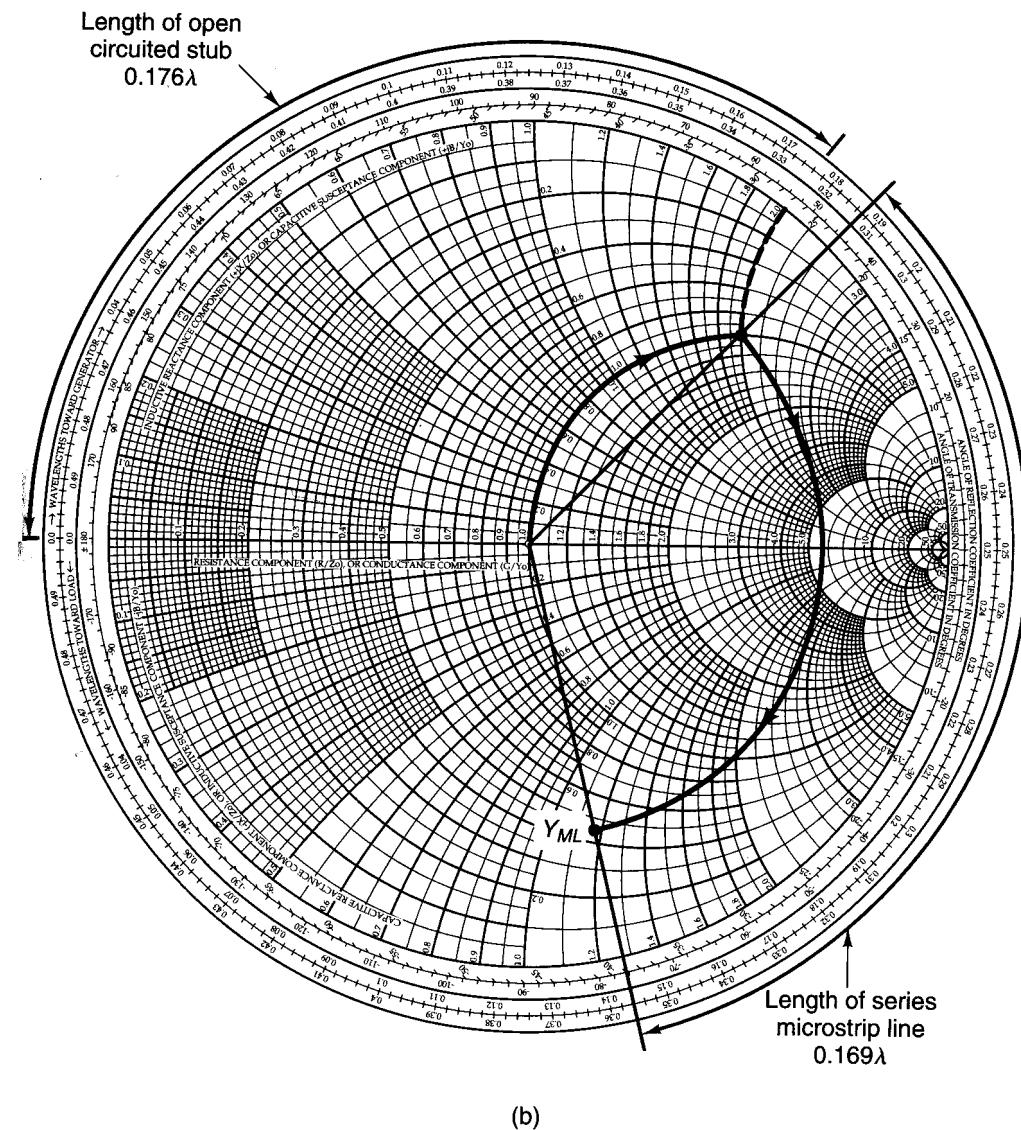


Figure 3.6.3 Continued

$$0.185\lambda = 6.70 \text{ mm}$$

$$0.0615\lambda = 2.23 \text{ mm}$$

$$0.169\lambda = 6.12 \text{ mm}$$

$$0.176\lambda = 6.37 \text{ mm}$$

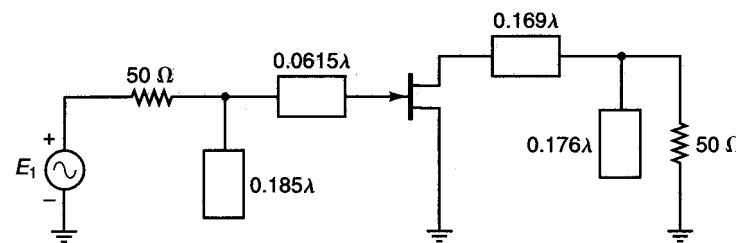


Figure 3.6.4 The ac schematic of a GaAs FET microwave amplifier. All microstrip lines have a characteristic impedance of 50 Ω.

The design for $G_{T,\max}$ with Γ_{Ms} and Γ_{ML} , at 6 GHz assures that the input and output VSWR are 1.

This example is revisited in Chapter 4 (Section 4.3), where noise considerations are included. Finally, we should point out that the stability must be checked at all frequencies, so that the reflection coefficients Γ_{Ms} and Γ_{ML} provide stable operation.

We have seen that for an unconditional stable device, the terminations Γ_{Ms} and Γ_{ML} , given by (3.6.5) and (3.6.6), will produce a simultaneous conjugate match which results in the maximum value of the transducer power gain. If the design calls for a transducer power gain different from the maximum, a constant-gain circle procedure based on (3.2.1) or (3.2.2) can be attempted. As we will see, such a procedure is not practical.

A constant-gain circle procedure based on (3.2.1) could be attempted as follows. Write (3.2.1) in the form

$$G_T = G'_s G_o G_L \quad (3.6.13)$$

where

$$G'_s = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_{IN} \Gamma_s|^2} \quad (3.6.14)$$

$$G_o = |S_{21}|^2$$

and

$$G_L = \frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2} \quad (3.6.15)$$

Then the design procedure is as follows:

- From (3.6.15), the constant G_L circles can be drawn using (3.4.9) to (3.4.12). Select the desired Γ_L for a given G_L gain. Observe that $\Gamma_s = S_{22}^*$ produces $G_{L,\max}$ but not $G_{T,\max}$.

- Calculate Γ_{IN} from (3.2.5). Observe that Γ_{IN} depends on Γ_L ; therefore, G'_s depends on G_L .
- From (3.6.14), the constant G'_s circles can be drawn using (3.4.9) to (3.4.12) (observing that Γ_{IN} replaces S_{22}). $\Gamma_s = \Gamma_{IN}^*$ produces $G'_{s,\max}$, and constant G'_s circles can be drawn for $G'_s < G'_{s,\max}$. Select the desired Γ_s for a given G'_s gain. Of course, the values of G'_s might not be satisfactory for the desired G_T . This will require the selection of another Γ_L and the procedure repeated.
- Design the matching networks.

The procedure just outlined is not recommended for a practical design since Γ_{IN} is a function of Γ_L , making the G'_s function dependent of the G_L function. Furthermore, the centers of the gain circles at $\Gamma_L = S_{22}^*$ and $\Gamma_s = \Gamma_{IN}^*$ do not give $G_{T,\max}$. In fact, the graphical approach becomes tedious because of the iterative process required for obtaining the desired gain.

As shown in the next section, the design of a microwave transistor amplifier in the unconditional stable bilateral case, for a gain different from $G_{T,\max}$, can be done using the operating power gain equation (the available power gain equation can also be used). When the transistor is unconditionally stable, a simultaneous conjugate match can be found, and the design procedure is based on $G_{T,\max}$ or (as shown in the next section) on the operating power gain. In fact, when designing for $G_{T,\max}$, which is equal to $G_{p,\max}$ and to $G_{A,\max}$, all design procedures result in $\Gamma_s = \Gamma_{Ms}$ and $\Gamma_L = \Gamma_{ML}$.

3.7 OPERATING AND AVAILABLE POWER-GAIN CIRCLES

Operating Power-Gain Circles

When S_{12} cannot be neglected, a design procedure based on the operating power gain G_p is commonly used. The operating power gain is independent of the source impedance; therefore, an operating power-gain circle procedure for both unconditionally stable and potentially unstable transistors is simple and recommended for practical designs.

Again we must consider two cases, the unconditionally stable case and the potentially unstable case.

Unconditionally stable bilateral case. To develop a design procedure with G_p , we write (3.2.3) in the form

$$G_p = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2)}{\left(1 - \frac{|S_{11} - \Delta \Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2}\right) |1 - S_{22} \Gamma_L|^2} = |S_{21}|^2 g_p \quad (3.7.1)$$

where

$$\begin{aligned} g_p &= \frac{G_p}{|S_{21}|^2} = \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2 - |S_{11} - \Delta\Gamma_L|^2} \\ &= \frac{1 - |\Gamma_L|^2}{1 - |S_{11}|^2 + |\Gamma_L|^2(|S_{22}|^2 - |\Delta|^2) - 2 \operatorname{Re}(\Gamma_L C_2)} \end{aligned} \quad (3.7.2)$$

and

$$C_2 = S_{22} - \Delta S_{11}^* \quad (3.7.3)$$

Here G_p and g_p are functions of the device S parameters and Γ_L .

In Appendix G the values of Γ_L in (3.7.2) that produce a constant value of g_p are shown to lie on a circle, known as an operating power-gain circle. The equation for an operating power-gain circle in the Γ_L plane, with g_p as a parameter, is

$$|\Gamma_L - C_p| = r_p$$

where the center of the circle C_p is located at

$$C_p = \frac{g_p C_2^*}{1 + g_p (|S_{22}|^2 - |\Delta|^2)} \quad (3.7.4)$$

and the radius of the circle is given by

$$r_p = \frac{[1 - 2K|S_{12}S_{21}|g_p + |S_{12}S_{21}|^2 g_p^2]^{1/2}}{|1 + g_p (|S_{22}|^2 - |\Delta|^2)|} \quad (3.7.5)$$

Equation (3.7.4) shows that the distance from the origin to the center of a power-gain circle is simply $|C_p|$, and the angle of inclination of the circle is C_2^* .

The maximum operating power gain occurs at the value of Γ_L when $r_p = 0$. Therefore, from (3.7.5) we can write

$$g_{p,\max}^2 |S_{12}S_{21}|^2 - 2K|S_{12}S_{21}|g_{p,\max} + 1 = 0 \quad (3.7.6)$$

where $g_{p,\max}$ is the maximum value of g_p . The solution to (3.7.6) for unconditional stability is

$$g_{p,\max} = \frac{1}{|S_{12}S_{21}|} (K - \sqrt{K^2 - 1}) \quad (3.7.7)$$

Therefore, substituting (3.7.7) into (3.7.1) gives

$$G_{p,\max} = \frac{|S_{21}|}{|S_{12}|} (K - \sqrt{K^2 - 1}) \quad (3.7.8)$$

As expected, (3.7.8) is identical to $G_{T,\max}$ in (3.6.10).

The value of Γ_L that produces $G_{p,\max}$ follows by substituting $g_p = g_{p,\max}$ in (3.7.4). This value of $\Gamma_L = C_{p,\max}$ must be equal to Γ_{ML} . That is, from (3.7.4)

$$\Gamma_{ML} = C_{p,\max} = \frac{g_{p,\max} C_2^*}{1 + g_{p,\max} (|S_{22}|^2 - |\Delta|^2)} \quad (3.7.9)$$

Substituting (3.7.7) into (3.7.9) and performing some manipulations, it follows that (3.7.9) is identical to (3.6.6), using the minus sign in (3.6.6). This derivation is presented in Appendix H.

The lowest value of g_p is zero, which corresponds to $G_p = 0$. From (3.7.1), $G_p = 0$ occurs when $|\Gamma_L| = 1$. In other words, the operating power gain is zero when all the output power is reflected from the load (i.e., when $|\Gamma_L| = 1$).

For a given G_p , Γ_L is selected from the constant operating power-gain circles. $G_{p,\max}$ results when Γ_L is selected at the distance where $g_{p,\max} = G_{p,\max}/|S_{21}|^2$. The maximum output power results when a conjugate match is selected at the input (i.e., $\Gamma_s = \Gamma_{IN}^*$). It also follows that when $\Gamma_s = \Gamma_{IN}^*$, the input power is equal to the maximum available input power. Therefore, under these circumstances the maximum transducer power gain ($G_{T,\max}$) and the operating power gain are equal, and the values of Γ_s and Γ_L that result in $G_{p,\max}$ are identical to Γ_{MS} and Γ_{ML} , respectively.

The procedure for drawing a constant operating power-gain circle in the Z Smith chart is as follows:

1. For a given G_p , the center and radius of the constant operating power-gain circle are given by (3.7.4) and (3.7.5).
2. Select the desired Γ_L .
3. For the given Γ_L , maximum output power is obtained with a conjugate match at the input—namely, with $\Gamma_s = \Gamma_{IN}^*$, where Γ_{IN} is given by (3.2.5). This value of Γ_s produces the transducer power gain $G_T = G_p$.

Example 3.7.1

Design the amplifier in Example 3.6.1 to have an operating power gain of 9 dB instead of $G_{T,\max} = G_{p,\max} = 11.38$ dB.

Solution. Since

$$|S_{21}|^2 = (2.058)^2 = 4.235 \quad \text{or} \quad 6.27 \text{ dB}$$

then

$$g_p = \frac{G_p}{|S_{21}|^2} = \frac{7.94}{4.235} = 1.875$$

From the results in Example 3.6.1, $K = 1.504$, $|\Delta| = 0.3014$, and $C_2 = 0.3911[-103.9^\circ]$. Therefore, the radius and center of the 9-dB operating power-gain circle, from (3.7.4) and (3.7.7), are $r_p = 0.431$ and $C_p = 0.508[103.9^\circ]$.

The graphical construction is shown in Fig. 3.7.1a. The 9-dB operating power-gain circle shows all loads that produce $G_p = 9$ dB. The load reflection coefficient can be selected at point A—namely, $\Gamma_L = 0.36[47.5^\circ]$. Then the required Γ_s for maximum output power is

$$\Gamma_s = \Gamma_{IN}^* = \left[S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right]^* = 0.629[175.51^\circ]$$

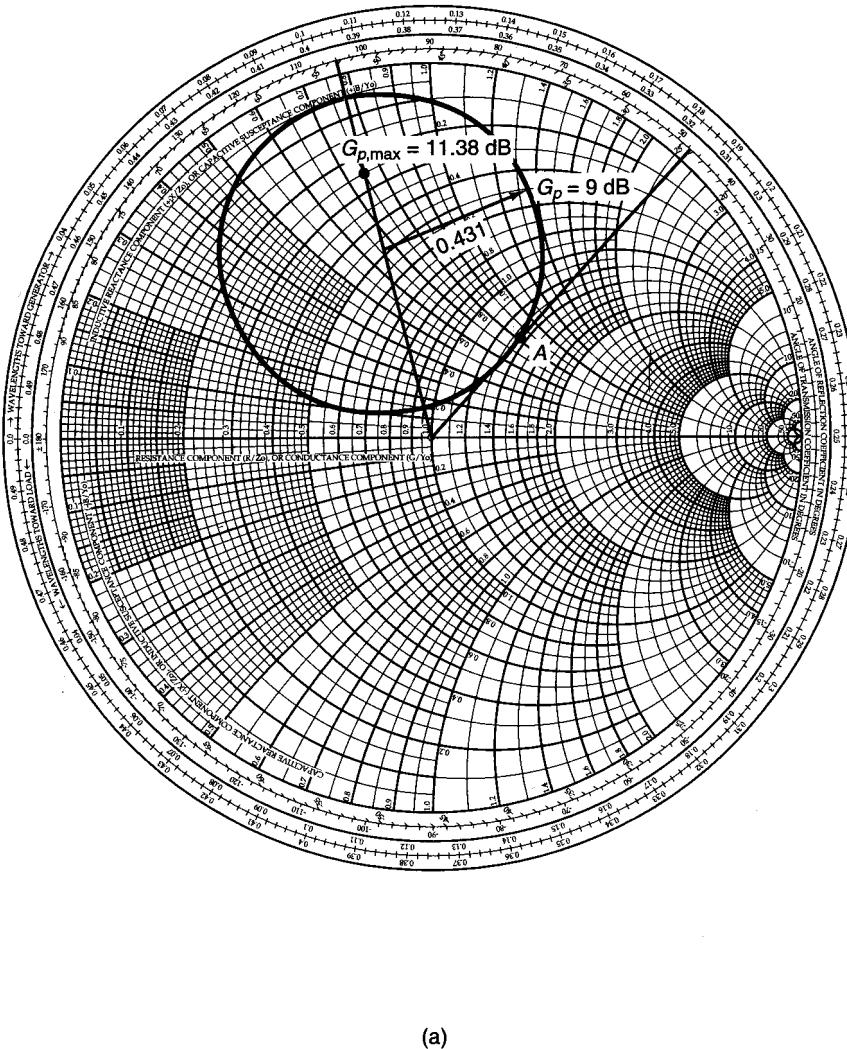


Figure 3.7.1 (a) Operating power-gain circle for $G_p = 9$ dB and location of $G_{p,\max} = 11.38$ dB; (b) the block diagram of the amplifier.

Since $\Gamma_s = \Gamma_{IN}^*$, it follows that $G_T = G_p = 9$ dB. The block diagram of the amplifier is shown in Fig. 3.7.1b.

The input VSWR is 1 since $\Gamma_s = \Gamma_{IN}^*$. The output VSWR is calculated using (2.8.4) and (2.8.6). From (3.2.6) the output reflection coefficient is $\Gamma_{OUT} = 0.67 \angle -102.66^\circ$. Using (2.8.6), we obtain

$$|\Gamma_b| = \left| \frac{\Gamma_{OUT} - \Gamma_L^*}{1 - \Gamma_{OUT}\Gamma_L} \right| = \left| \frac{0.67 \angle -102.66^\circ - 0.36 \angle -47.5^\circ}{1 - 0.67 \angle -102.66^\circ (0.36 \angle 47.5^\circ)} \right| = 0.622$$

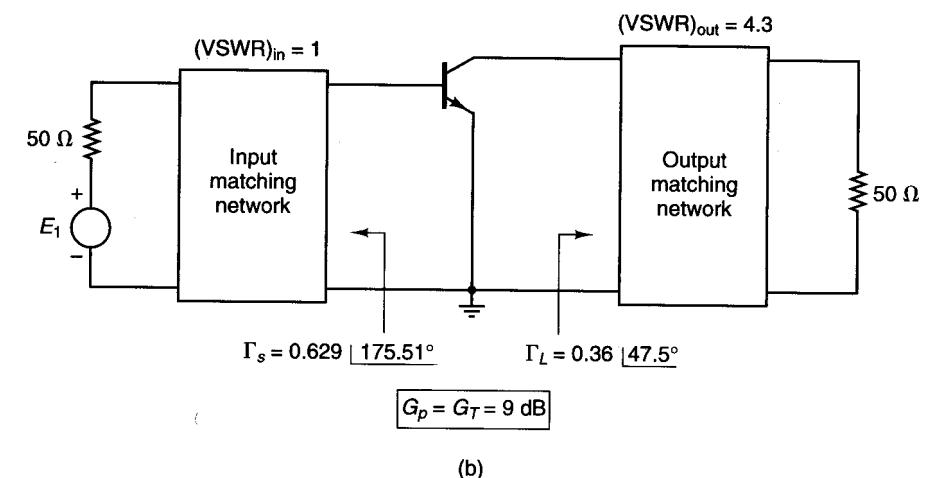


Figure 3.7.1 Continued

Therefore, the output VSWR is

$$(VSWR)_{out} = \frac{1 + 0.622}{1 - 0.622} = 4.3$$

The location of $G_{p,\max} = 11.38$ dB can be found as follows:

$$g_{p,\max} = \frac{G_{p,\max}}{|S_{21}|^2} = \frac{13.74}{(2.058)^2} = 3.24$$

$$r_{p,\max} = 0$$

and the value of Γ_L that produces $G_{p,\max}$ is

$$\Gamma_{ML} = C_{p,\max} = \frac{g_{p,\max} C_2^*}{1 + g_{p,\max} (|S_{22}|^2 - |A|^2)} = 0.718 \angle 103.9^\circ$$

This value is identical to the value Γ_{ML} found in Example 3.6.1. The associated Γ_s for maximum output power, which is Γ_{Ms} , is obtained from

$$\Gamma_{Ms} = \left[S_{11} + \frac{S_{12}S_{21}\Gamma_{ML}}{1 - S_{22}\Gamma_{ML}} \right]^* = 0.762 \angle 177.3^\circ$$

As expected, this value is identical to the value of Γ_{Ms} in Example 3.6.1.

The previous example illustrates that for a given Γ_L on a constant G_p circle, there is an associated $\Gamma_s = \Gamma_{IN}^*$ that gives $(VSWR)_{in} = 1$. Then for $\Gamma_s = \Gamma_{IN}^*$, there is a Γ_{OUT} . The mismatch at the output between Γ_L and Γ_{OUT} determines $(VSWR)_{out}$ according to (2.8.4). We can try other values of Γ_L on the constant G_p circle and calculate for each Γ_L the associated $\Gamma_s = \Gamma_{IN}^*$. Then for each Γ_s the associated Γ_{OUT} is calculated and the resulting

$(VSWR)_{out}$. This analysis can be done for a series of points on the constant G_p circle to see if a better $(VSWR)_{out}$ can be obtained. In some cases, a certain mismatch at the input might be necessary in order to obtain the desired $(VSWR)_{out}$. Design considerations involving gain and VSWRs are discussed in Section 3.8.

Potentially unstable bilateral case. With a potentially unstable transistor, the design procedure for a given G_p is as follows:

1. For a given G_p , draw the constant operating power-gain circle using (3.7.4) and (3.7.5), and also draw the output stability circle as discussed in Section 3.3 [i.e., see (3.3.7) and (3.3.8)]. Select a value of Γ_L that is in the stable region and not too close to the stability circle.
2. Calculate Γ_{IN} using (3.2.5) and determine if a conjugate match at the input is feasible. That is, draw the input stability circle as discussed in Section 3.3 [i.e., see (3.3.9) and (3.3.10)] and determine if $\Gamma_s = \Gamma_{IN}^*$ lies in the input stable region.
3. If $\Gamma_s = \Gamma_{IN}^*$ is not in the stable region or is in the stable region but very close to the input stability circle, the value of Γ_s can be selected arbitrarily or a new value of G_p can be selected. Of course, we must be careful when selecting Γ_s arbitrarily since the value of Γ_s affects the output power and the VSWR.

The values of Γ_L and Γ_s should not be too close to their respective stability circles, because oscillations might occur due to component variations that can place Γ_s and Γ_L in the unstable regions.

Since G_p can be infinite in a potentially unstable case, it is practical to keep the value of G_p below the figure of merit value G_{MSG} [given in (3.6.11)]. The design for G_p lower than G_{MSG} can be performed with good stability and practical values of the input and output VSWR. On the other hand, a design for a G_p greater than G_{MSG} usually produces values of Γ_L and Γ_s close to the unstable regions and large values of the input and output VSWR. Designs procedures involving potentially unstable transistors and VSWR considerations are discussed in Section 3.8.

In a potentially unstable situation with $K > 1$ and $|\Delta| > 1$, there is a $G_{p,min}$, just like $G_{T,min}$ in (3.6.12). In fact, in a potentially unstable situation with $K > 1$ and $|\Delta| > 1$, the other solution of (3.7.6), denoted by $g_{p,min}$, is

$$g_{p,min} = \frac{1}{|S_{12}S_{21}|} (K + \sqrt{K^2 - 1}) = \frac{G_{p,max}}{|S_{21}|^2}$$

Hence,

$$G_{p,min} = \frac{|S_{21}|}{|S_{12}|} (K + \sqrt{K^2 - 1}) \quad (3.7.10)$$

Equation (3.7.10) gives the minimum value that G_p can have inside the stable region in a potentially unstable case with $K > 1$ and $|\Delta| > 1$. The maximum value that G_p can have is infinite.

The value of Γ_L that produces $G_{p,min}$, denoted by $\Gamma_{L,min}$, is

$$\Gamma_{L,min} = \frac{g_{p,min} C_2^*}{1 + g_{p,min} (|S_{22}|^2 - |\Delta|^2)} \quad (3.7.11)$$

It also follows that $\Gamma_{L,min}$ in (3.7.11) is identical to Γ_{ML} in (3.6.6) when the plus sign is used in (3.6.6).

Example 3.7.2

The S parameters of a GaAs FET at $I_D = 50\% I_{DSS}$, $I_{DSS} = 10 \text{ mA}$, $V_{DS} = 5 \text{ V}$, and $f = 8 \text{ GHz}$ are

$$S_{11} = 0.5 \angle -180^\circ$$

$$S_{12} = 0.08 \angle 30^\circ$$

$$S_{21} = 2.5 \angle 70^\circ$$

$$S_{22} = 0.8 \angle -100^\circ$$

The transistor is potentially unstable at 8 GHz with $G_{MSG} = 14.9 \text{ dB}$. Design an amplifier with $G_p = 10 \text{ dB}$.

Solution. First we will verify that the transistor is potentially unstable at 8 GHz. From (3.3.16) and (3.3.17), we obtain $K = 0.4$ and $\Delta = 0.223 \angle 62.12^\circ$. Since $K < 1$, the GaAs FET is potentially unstable. Also, from (3.6.11) the value of G_{MSG} is $G_{MSG} = 2.5/0.08 = 31.25$ or 14.9 dB .

In order to design for $G_p = 10 \text{ dB}$ (4.9 dB less than the G_{MSG}), the 10-dB operating power-gain circle and the output stability circle must be calculated. The radius and center of the 10-dB power gain circle, from (3.7.4) and (3.7.5), are $r_p = 0.473$ and $C_p = 0.572 \angle 97.2^\circ$. The radius and center of the output stability circle, from (3.3.7) and (3.3.8), are $r_L = 0.34$ and $C_L = 1.18 \angle 97.2^\circ$.

The Smith chart in Fig. 3.7.2 shows the construction of the 10-dB operating power-gain circle and the output stability circle. Since $|S_{11}| < 1$, the stable region is the region outside the output stability circle. Γ_L is selected on the 10-dB power-gain circle at location A—namely, $\Gamma_L = 0.1 \angle 97^\circ$ or $Z_L = 50 (0.96 + j0.19) \Omega$.

For a conjugate match at the input, Γ_s is given by $\Gamma_s = \Gamma_{IN}^* = 0.52 \angle 179.32^\circ$ and we must determine if the value of Γ_s is in the stable region. The radius and center of the input stability circle, from (3.3.9) and (3.3.10), are $r_s = 1.0$ and $C_s = 1.67 \angle 171^\circ$, where the stable region is the region outside the input stability circle. Therefore, Γ_s is a stable source reflection coefficient. Since $\Gamma_s = \Gamma_{IN}^*$, it follows that $G_T = G_p = 10 \text{ dB}$ and $(VSWR)_{in} = 1$.

From (3.2.6), the output reflection coefficient is $\Gamma_{out} = 0.934 \angle -97.18^\circ$. Using (2.8.6), we obtain $|\Gamma_b| = 0.918$ and the output VSWR is

$$(VSWR)_{out} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} = \frac{1 + 0.918}{1 - 0.918} = 23.5$$

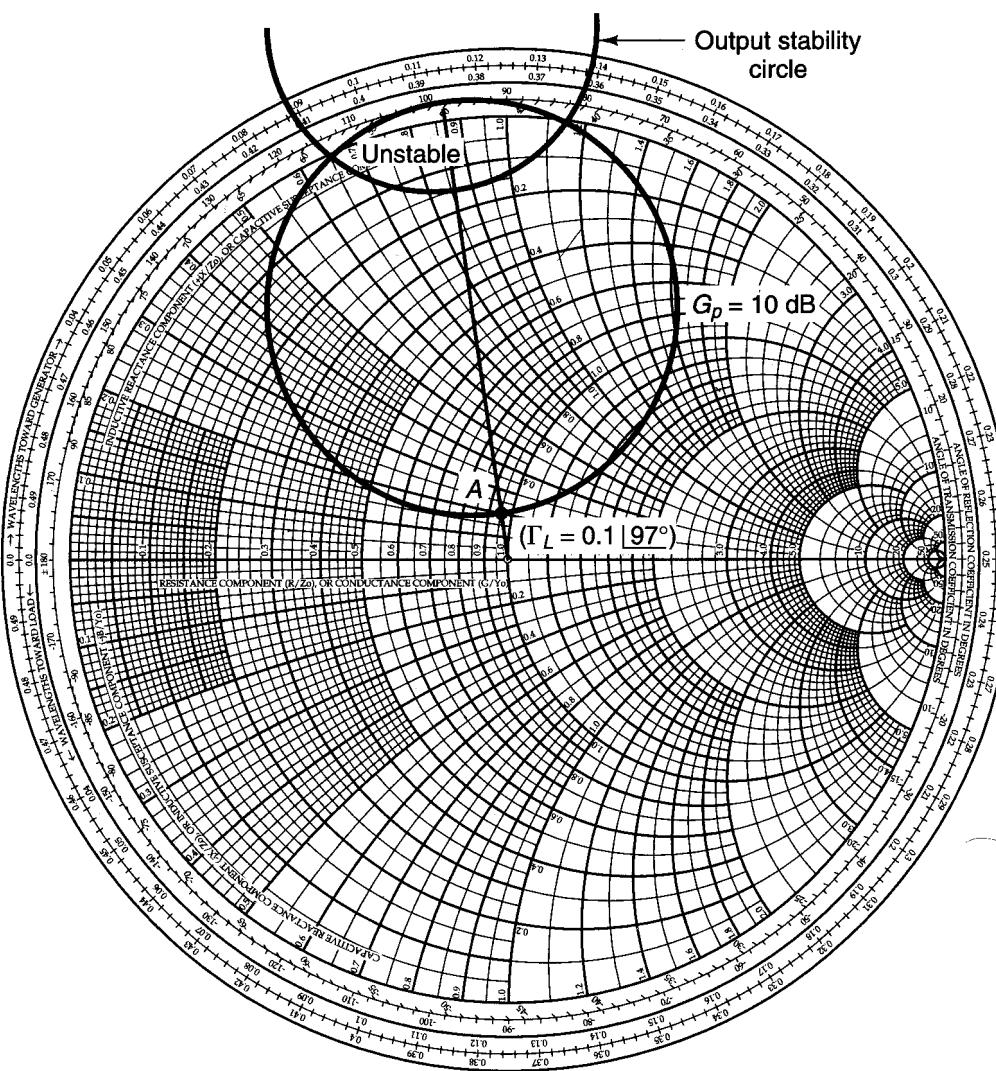


Figure 3.7.2 Power and stability circles construction for Example 3.7.2.

The previous calculations show that the output port of the amplifier is significantly mismatched in order to obtain $G_p = 10 \text{ dB}$. The output VSWR can be lowered by accepting a higher-input VSWR or by changing the amplifier gain requirements. For example, with $\Gamma_s = 0.22 | 177^\circ$ and $\Gamma_L = 0.1 | 97^\circ$, it follows that $(\text{VSWR})_{\text{in}} = 2$ and $(\text{VSWR})_{\text{out}} = 9.9$. A formal procedure involving gain and VSWR considerations is discussed in Section 3.8.

Using the values of the S parameters in Example 3.7.2, the operating power-gain circles for $G_p = 0 \text{ dB}, 10 \text{ dB}, 15 \text{ dB}, 20 \text{ dB}$, and 30 dB were calculated and plotted in Fig. 3.7.3. Observe that the $G_p = 0$ (i.e., $-\infty \text{ dB}$) circle oc-

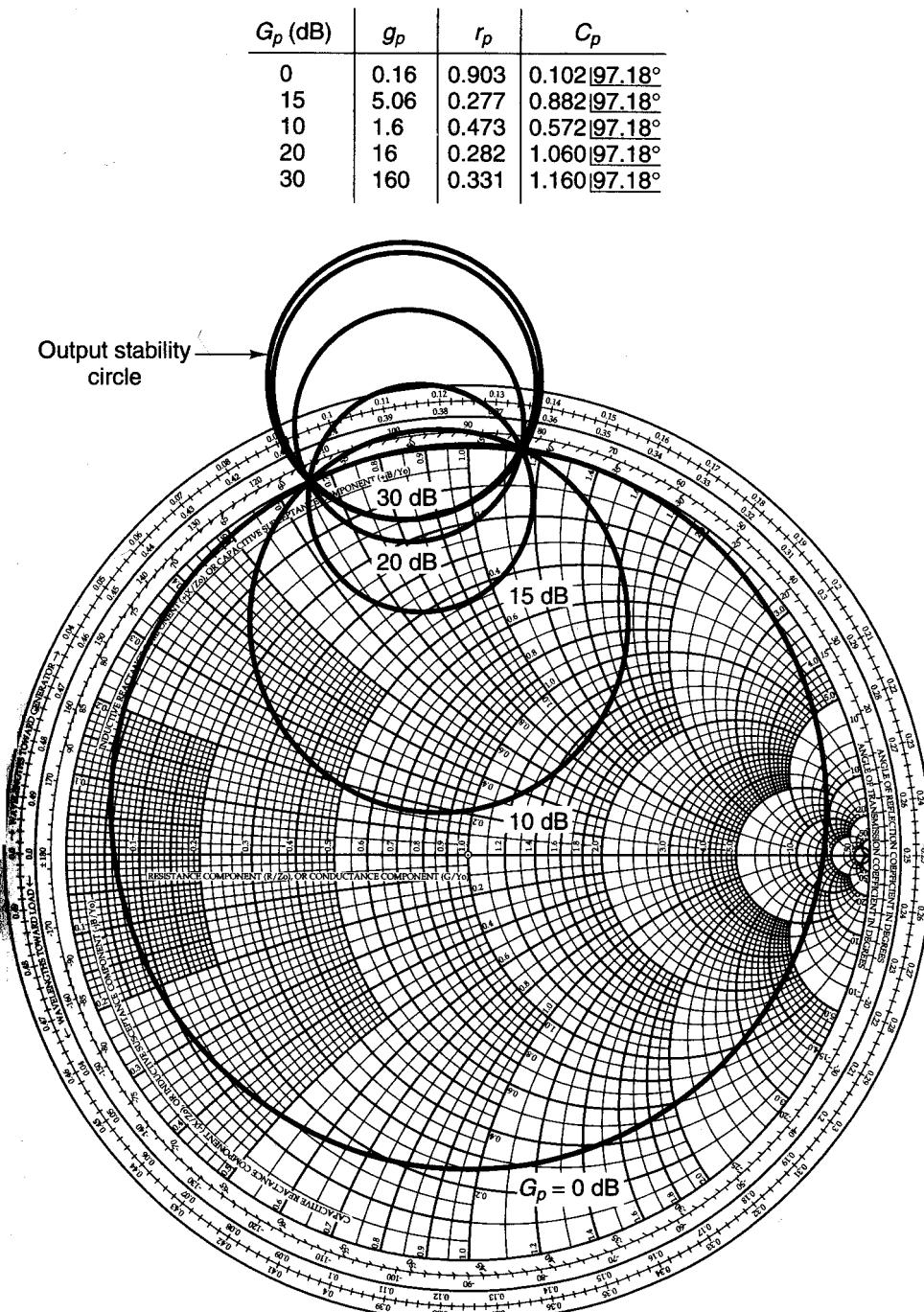


Figure 3.7.3 Typical behavior of the operating power-gain circles. The S parameters in Example 3.7.2 were used to calculate the constant gain circles.

curs when $|\Gamma_L| = 1$, and the operating power-gain circles approach the output stability circles as G_p becomes infinite.

Since G_p becomes infinite when Γ_L is on the output stability circle, it follows that the power-gain circle for $G_p = \infty$ is identical to the output stability circle. In fact, for $G_p \rightarrow \infty$ we have $g_p \rightarrow \infty$ and, taking the limit as g_p approaches infinity in (3.7.4), we obtain

$$C_p = \frac{C_2^*}{|S_{22}|^2 - |\Delta|^2}$$

and

$$r_p \approx \frac{|S_{12}S_{21}|^2 g_p^{1/2}}{|g_p|(|S_{22}|^2 - |\Delta|^2)} = \frac{|S_{12}S_{21}|}{|S_{22}|^2 - |\Delta|^2}$$

which are recognized as the equations for C_L and r_L in (3.3.8) and (3.3.7), respectively.

From the plots in Fig. 3.7.2, it is seen that the power-gain circles and the output stability circles intersect the Smith chart at the same points. This is not a coincidence, and it can be shown (see Problem 3.28) that for a potentially unstable device these circles intersect the Smith chart at the same points.

Example 3.7.3

(This example illustrates the power-gain circles for a two-port network with $K > 1$ and $|\Delta| > 1$.)

The S parameters of a two-port network at $f = 4$ GHz are

$$S_{11} = 0.7 \underline{-50^\circ}$$

$$S_{12} = 0.27 \underline{75^\circ}$$

$$S_{21} = 5 \underline{120^\circ}$$

$$S_{22} = 0.6 \underline{80^\circ}$$

Plot several operating power-gain circles ranging from $G_p = G_{p,\min}$ to $G_p = \infty$.

Solution. For this two-port network, it follows that $K = 1.202$ and $\Delta = 1.76 \underline{18.54^\circ}$ (i.e., $|\Delta| = 1.76$ is greater than 1). Since $K > 1$ and $|\Delta| > 1$ (or $B_2 < 0$), we obtain from (3.6.6), using the plus sign in (3.6.6),

$$B_1 = -1.964$$

$$B_2 = -2.224$$

$$C_1 = 0.395 \underline{97.92^\circ}$$

$$C_2 = 0.654 \underline{-121.95^\circ}$$

$$\Gamma_{L,\min} = \Gamma_{ML} = \frac{-2.224 + \sqrt{(2.224)^2 - 4(0.654)^2}}{2(0.654 \underline{-121.95^\circ})} = 0.325 \underline{-58.05^\circ}$$

Then, using (3.7.10),

$$G_{p,\min} = G_{T,\min} = \frac{5}{0.27} (1.202 + \sqrt{(1.202)^2 - 1}) = 34.61 \text{ (or } 15.39 \text{ dB)}$$

The output stability circle ($C_L = 0.239 \underline{-58.05^\circ}$ and $r_L = 0.494$) and power-gain circles for $G_{p,\min} = 15.39$ dB and $G_p = 18$ dB, 20 dB, and 30 dB were calculated and are plotted in Fig. 3.7.4a. Observe that the operating power-gain circles approach the output stability circle as G_p becomes infinite.

It is simple to show that in the unstable region G_p in decibels becomes negative, approaching $-\infty$ dB at $|\Gamma| = 1$.

For maximum power transfer to the load, the source reflection coefficient is

$$\Gamma_s = \Gamma_{IN}^* = \left(0.7 \underline{-50^\circ} + \frac{0.27 \underline{75^\circ} (5 \underline{120^\circ}) 0.325 \underline{-58.05^\circ}}{1 - 0.6 \underline{80^\circ} (0.325 \underline{-58.05^\circ})} \right)^* = 0.21 \underline{82.08^\circ}$$

This value of Γ_s is identical to the value of Γ_{Ms} in (3.6.5), using the plus sign in (3.6.5). That is,

$$\Gamma_s = \Gamma_{Ms} = 0.21 \underline{82.08^\circ}$$

The input stability circle ($C_s = 0.152 \underline{82.08^\circ}$ and $r_s = 0.518$) and Γ_{Ms} are shown in Fig. 3.7.4b.

Since a simultaneous conjugate match condition exists it follows that the input and output VSWR are equal to one.

Available Power-Gain Circles

Unconditionally stable bilateral case. The derivation of the constant available power-gain circles is similar to that of the operating power-gain circles. From (3.2.4), we write G_A in the form

$$G_A = \frac{|S_{21}|^2 (1 - |\Gamma_s|^2)}{\left(1 - \frac{|S_{22} - \Delta \Gamma_s|^2}{|1 - S_{11} \Gamma_s|} \right) |1 - S_{11} \Gamma_s|^2} = |S_{21}|^2 g_a \quad (3.7.12)$$

where

$$g_a = \frac{G_A}{|S_{21}|^2} = \frac{1 - |\Gamma_s|^2}{1 - |S_{22}|^2 + |\Gamma_s|^2 (|S_{11}|^2 - |\Delta|^2) - 2 \operatorname{Re}(\Gamma_s C_1)} \quad (3.7.13)$$

and

$$C_1 = S_{11} - \Delta S_{22}^* \quad (3.7.14)$$

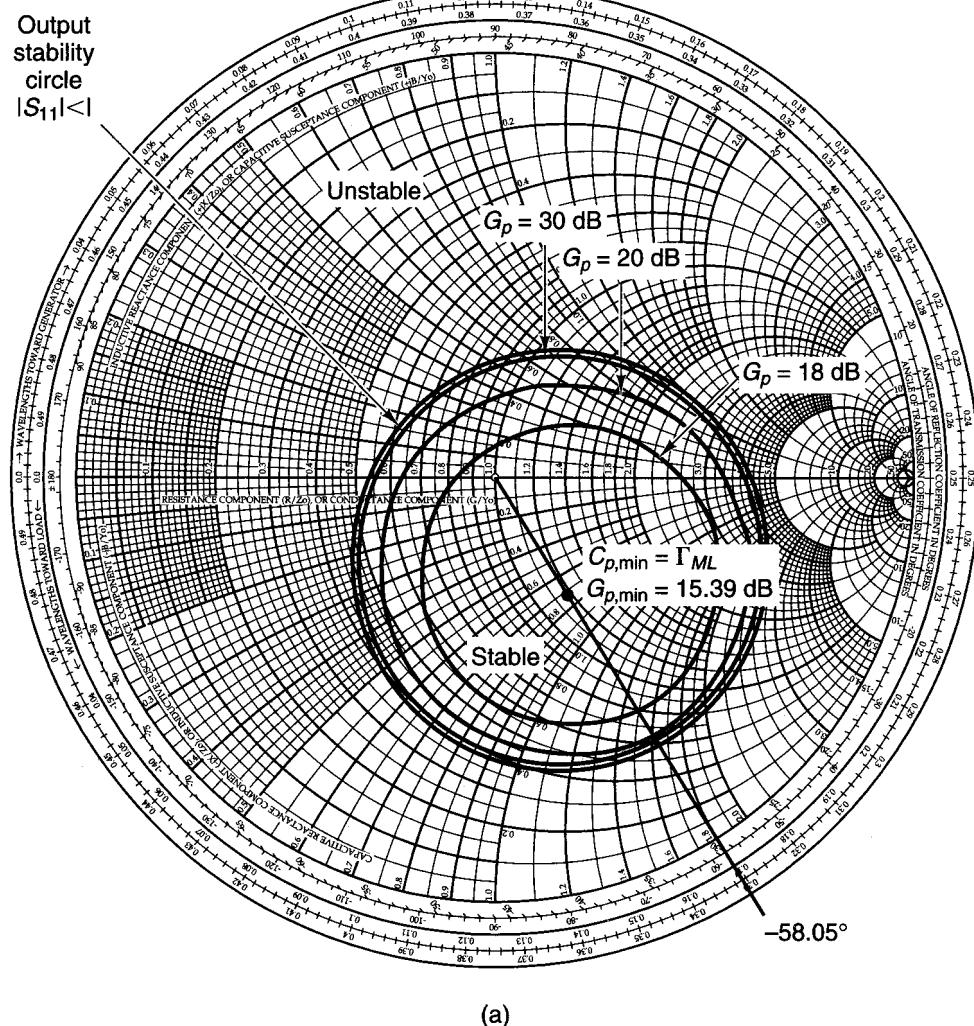
Since (3.7.12), (3.7.13), and (3.7.14) are identical in form to (3.7.1), (3.7.2), and (3.7.3), it follows that the center C_a and radius r_a of constant operating power-gain circles are given by

$$C_a = \frac{g_a C_1^*}{1 + g_a (|S_{11}|^2 - |\Delta|^2)} \quad (3.7.15)$$

and

$$r_a = \frac{[1 - 2K|S_{12}S_{21}|g_a + |S_{12}S_{21}|^2 g_a^2]^{1/2}}{|1 + g_a (|S_{11}|^2 - |\Delta|^2)|} \quad (3.7.16)$$

G_p (dB)	g_p	r_p	C_p
18	2.524	0.356	0.280 -58.05°
20	4	0.417	0.263 -58.05°
30	40	0.487	0.242 -58.05°



(a)

Figure 3.7.4 (a) Output stability circle and operating power-gain circles for Example 3.7.3; (b) input stability circle and Γ_{Ms} .

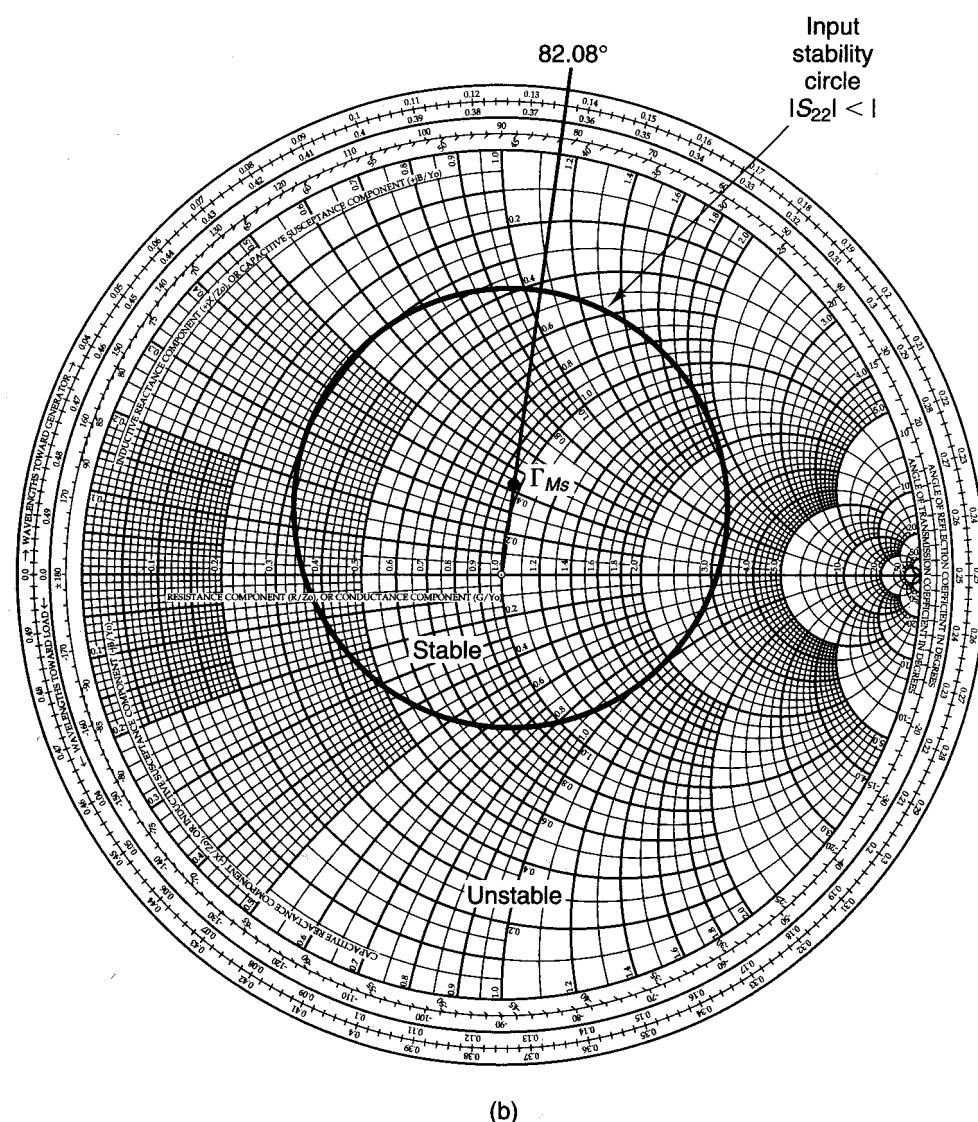


Figure 3.7.4 Continued

For a given G_A , a constant available power-gain circle can be plotted using (3.7.15) and (3.7.16). All Γ_s on this circle produce the given G_A . For the given G_A , maximum output power is obtained with $\Gamma_L = \Gamma_{OUT}^*$, where Γ_{OUT} is given by (3.2.6). This value of Γ_L produces the transducer power gain $G_T = G_A$.

Since the constant available power-gain circles and the constant noise figure circles are functions of Γ_s , they can be plotted together on the Smith chart

and the trade-offs that result between gain and noise figure can be analyzed. These concepts are studied in Chapter 4.

Potentially unstable bilateral case. With a potentially unstable transistor, the design procedure for a given G_A is as follows:

1. For a given G_A , draw the constant available gain circle using (3.7.15) and (3.7.16), and also draw the input stability circle using (3.3.9) and (3.3.10). Select a value of Γ_s that is in the stable region and not too close to the stability circle.
2. Calculate Γ_{OUT} using (3.2.6) and determine if a conjugate match at the output is feasible. That is, draw the output stability circle and determine if $\Gamma_L = \Gamma_{\text{OUT}}^*$ lies in the stable region.
3. If $\Gamma_L = \Gamma_{\text{OUT}}^*$ is not in the stable region or is in the stable region but very close to the output stability circle, the value of Γ_L can be selected arbitrarily or a new value of G_A can be selected. Of course, we must be careful when selecting Γ_L arbitrarily since the value of Γ_L affects the output power and the VSWR.

The values of Γ_L and Γ_s should not be too close to their respective stability circles, because oscillations might occur due to component variations that can place Γ_L and Γ_s in their respective unstable regions.

3.8 CONSTANT VSWR CIRCLES

The design specifications of a microwave amplifier usually include the maximum allowable values of its input VSWR and of its output VSWR. Constant input VSWR circles can be drawn on the Γ_s plane, and constant output VSWR circles can be drawn on the Γ_L plane. These circles can be used in the development of certain designs.

Consider Fig. 3.8.1, which shows a microwave amplifier. The input VSWR is given by

$$(\text{VSWR})_{\text{in}} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} \quad (3.8.1)$$

where $|\Gamma_a|$ is given by (2.8.3)—namely,

$$|\Gamma_a| = \left| \frac{\Gamma_{\text{IN}} - \Gamma_s^*}{1 - \Gamma_{\text{IN}}\Gamma_s} \right| \quad (3.8.2)$$

Equation (3.8.2) shows that for a given Γ_{IN} , the reflection coefficients Γ_s and Γ_a are related by a bilinear transformation. Hence, constant values of $|\Gamma_a|$ are obtained by values of Γ_s that lie on a circle.

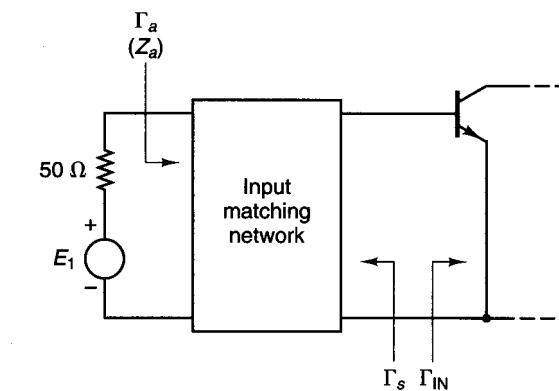


Figure 3.8.1 Input portion of a microwave amplifier.

The design for a given $(\text{VSWR})_{\text{in}}$ can be performed as follows. From (3.8.1), the value of $(\text{VSWR})_{\text{in}}$ fixes the value of $|\Gamma_a|$. Then (3.8.2) is used to plot the values of Γ_s that result in the constant value $|\Gamma_a|$. In Appendix I it is shown that from (3.8.2) a constant $|\Gamma_a|$ circle satisfies the equation

$$|\Gamma_s - C_{V_i}| = r_{V_i}$$

where the center C_{V_i} is given by

$$C_{V_i} = \frac{\Gamma_{\text{IN}}^*(1 - |\Gamma_a|^2)}{1 - |\Gamma_a\Gamma_{\text{IN}}|^2} \quad (3.8.3)$$

and the radius is

$$r_{V_i} = \frac{|\Gamma_a|(1 - |\Gamma_{\text{IN}}|^2)}{1 - |\Gamma_a\Gamma_{\text{IN}}|^2} \quad (3.8.4)$$

In an unconditional stable case, and in many potentially unstable cases, Γ_s can be selected equal to Γ_{IN}^* in order to get $(\text{VSWR})_{\text{in}} = 1$. Equations (3.8.3) and (3.8.4) demonstrate this fact. That is, from (3.8.1) with $(\text{VSWR})_{\text{in}} = 1$, it follows that $|\Gamma_a| = 0$, and from (3.8.3) and (3.8.4) we obtain

$$C_{V_i}|_{|\Gamma_a|=0} = \Gamma_{\text{IN}}^*$$

and

$$r_{V_i}|_{|\Gamma_a|=0} = 0$$

In other words, the value of $\Gamma_s = \Gamma_{\text{IN}}^*$ produces $|\Gamma_a| = 0$ and consequently $(\text{VSWR})_{\text{in}} = 1$.

Similar relations are obtained for the output VSWR. The output VSWR is given by

$$(\text{VSWR})_{\text{out}} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} \quad (3.8.5)$$

where $|\Gamma_b|$ is given by (2.8.6)—namely,

$$|\Gamma_b| = \left| \frac{\Gamma_{\text{OUT}} - \Gamma_L^*}{1 - \Gamma_{\text{OUT}}\Gamma_L} \right| \quad (3.8.6)$$

The constant $|\Gamma_b|$ circles satisfy the equation

$$|\Gamma_L - C_{V_o}| = r_{V_o}$$

where

$$C_{V_o} = \frac{\Gamma_{\text{OUT}}^*(1 - |\Gamma_b|^2)}{1 - |\Gamma_b\Gamma_{\text{OUT}}|^2} \quad (3.8.7)$$

and

$$r_{V_o} = \frac{|\Gamma_b|(1 - |\Gamma_{\text{OUT}}|^2)}{1 - |\Gamma_b\Gamma_{\text{OUT}}|^2} \quad (3.8.8)$$

Example 3.8.1

(a) The S parameters of a GaAs FET at 12 GHz, $V_{DS} = 3.5$ V, and $I_{DS} = 25$ mA are

$$S_{11} = 0.6 \angle 36^\circ$$

$$S_{12} = 0.14 \angle -85^\circ$$

$$S_{21} = 2.3 \angle -80^\circ$$

$$S_{22} = 0.15 \angle 45^\circ$$

Determine $G_{A,\max}$ and draw the constant G_A circle that is 1 dB less than $G_{A,\max}$.

(b) Select several values of Γ_s on the $G_A = G_{A,\max} - 1$ dB circle. For each Γ_s value, determine the values of Γ_L that lie on the constant $(VSWR)_{\text{out}} = 1.5$ circle, and draw the constant $(VSWR)_{\text{out}} = 1.5$ circles.

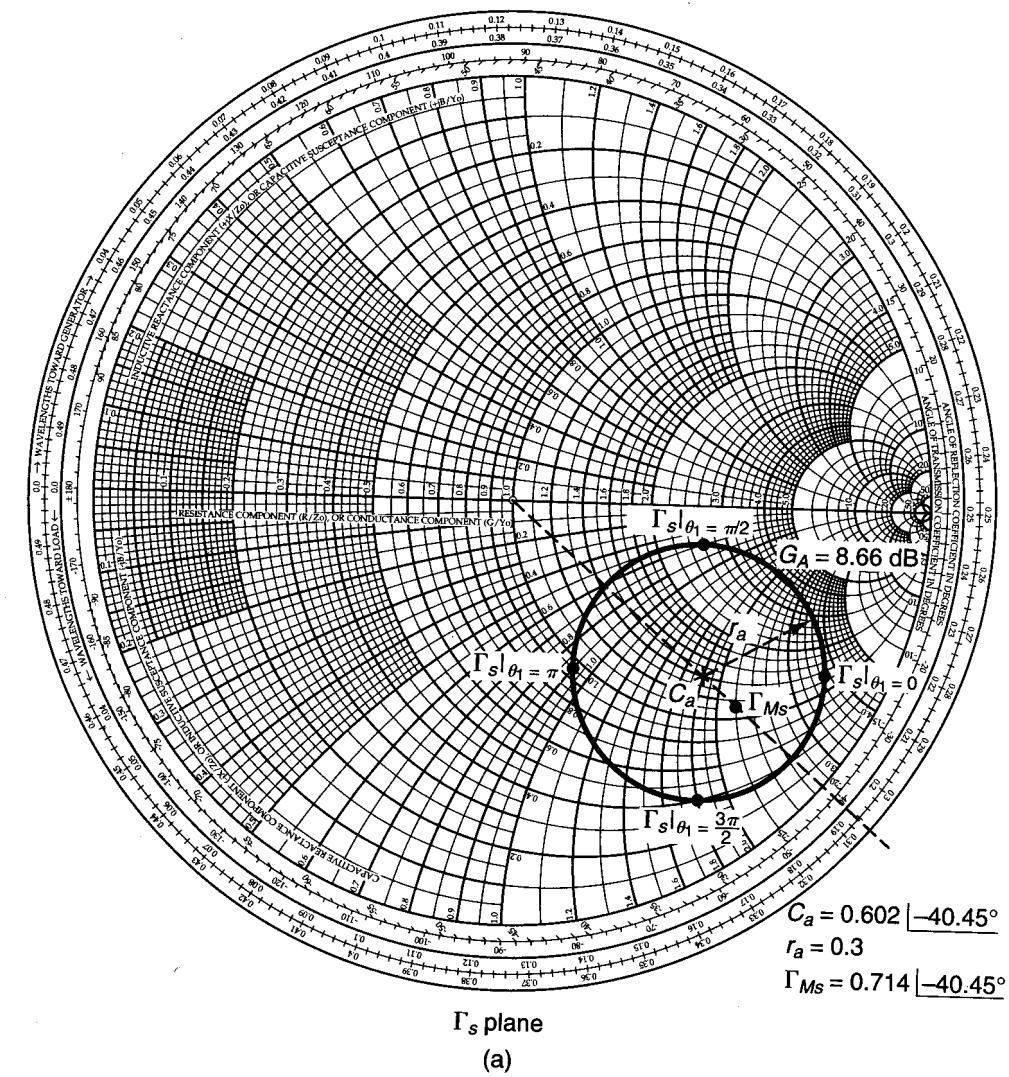
(c) Select several values of Γ_L on the $(VSWR)_{\text{out}} = 1.5$ circle. For each Γ_L value, calculate $(VSWR)_{\text{in}}$.

Solution. (a) This transistor is unconditionally stable since $K = 1.17$ and $\Delta = 0.368 \angle 27.91^\circ$. This transistor can be simultaneous conjugate matched with $\Gamma_{Ms} = 0.714 \angle -40.45^\circ$ and $\Gamma_{ML} = 0.387 \angle -129.36^\circ$, resulting in $(VSWR)_{\text{in}} = (VSWR)_{\text{out}} = 1$. From (3.6.10),

$$G_{A,\max} = \frac{|S_{21}|}{|S_{12}|} (K - \sqrt{K^2 - 1}) = \frac{2.3}{0.14} (1.17 - \sqrt{1.17^2 - 1}) = 9.24 \text{ (or } 9.66 \text{ dB)}$$

The analysis calls for a gain G_A different from $G_{A,\max}$ and the selection of Γ_s and Γ_L for certain VSWR performance. The design methods used in this example have applications in the design of low-noise amplifiers (discussed in Chapter 4). In the design of low-noise amplifiers, we commonly make trade-offs between gain, noise performance, and VSWRs.

The constant G_A circle that is 1 dB less than $G_{A,\max}$ is the circle for $G_A = 8.66$ dB. The center and radius of the $G_A = 8.66$ dB circle, from (3.7.15) and (3.7.16) with $g_a = 7.3451/(2.3)^2 = 1.3886$, are $C_a = 0.602 \angle -40.45^\circ$ and $r_a = 0.3$. The $G_A = 8.66$ dB constant gain circle is drawn in Fig. 3.8.2a.



Γ_s plane

(a)

Figure 3.8.2 (a) The $G_A = 8.66$ dB constant gain circle and Γ_{Ms} ; (b) four values of Γ_s on the $G_A = 8.66$ dB circle and calculations of $(VSWR)_{\text{in}}$ for $(VSWR)_{\text{out}} = 1$; (c) the four $(VSWR)_{\text{out}} = 1.5$ circles; (d) four values of Γ_L on the $(VSWR)_{\text{out}} = 1.5$ circles, the corresponding values of Γ_{IN} and $|\Gamma_a|$, and the resulting values of $(VSWR)_{\text{in}}$.

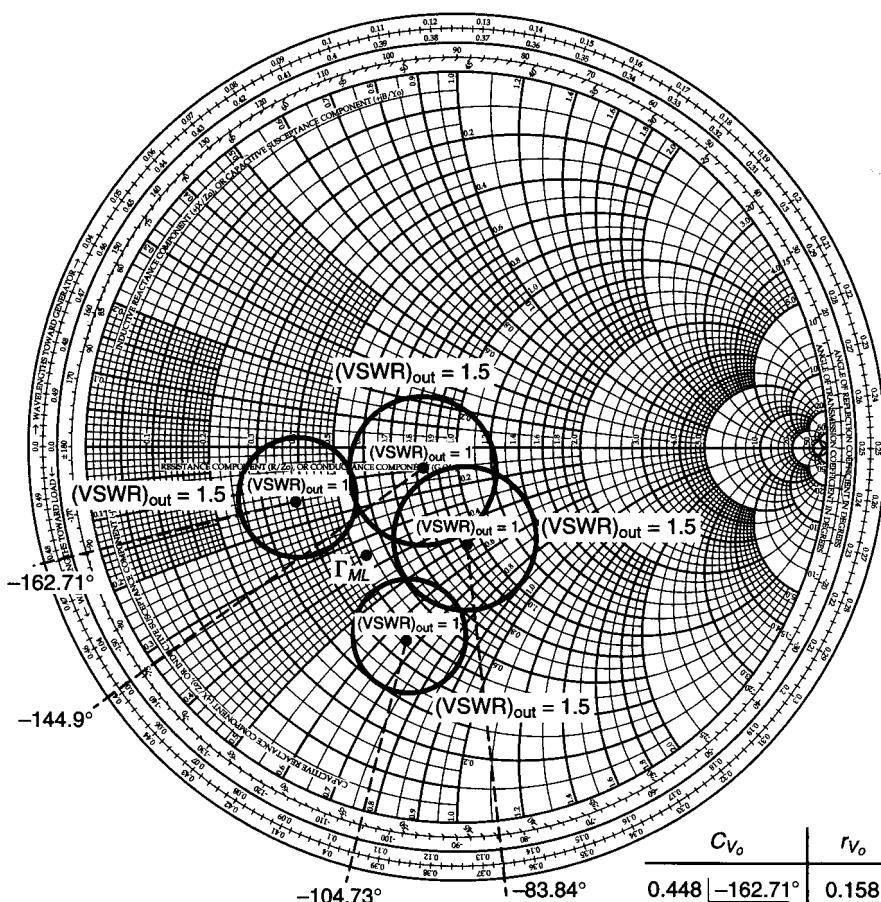
(b) The values of Γ_s on the $G_A = 8.66$ dB circle are given by

$$\Gamma_s = C_a + r_a e^{j\theta_1} = 0.602 \angle -40.45^\circ + 0.3 e^{j\theta_1}$$

Four convenient values of Γ_s on this circle are shown in Fig. 3.8.2a. They correspond to $\theta_1 = 0, \pi/2, \pi$, and $3\pi/2$. The values of Γ_s and the associated values of Γ_{OUT} are listed in

Γ_s	Γ_{OUT}	$\Gamma_L = \Gamma_{\text{OUT}}^*$	(VSWR) _{out}	Γ_{IN}	$ \Gamma_a $	(VSWR) _{in}
0.853 $[-27.25^\circ]$ for $\theta_1 = 0$	0.463 $[162.71^\circ]$	0.463 $[-162.71^\circ]$	1	0.743 $[34.62^\circ]$	0.395	2.31
0.467 $[-11.18^\circ]$ for $\theta_1 = \pi/2$	0.108 $[144.9^\circ]$	0.108 $[-144.9^\circ]$	1	0.634 $[36.71^\circ]$	0.394	2.30
0.421 $[-67.96^\circ]$ for $\theta_1 = \pi$	0.259 $[83.84^\circ]$	0.259 $[-83.84^\circ]$	1	0.630 $[43.53^\circ]$	0.394	2.30
0.829 $[-56.44^\circ]$ for $\theta_1 = 3\pi/2$	0.533 $[104.73^\circ]$	0.533 $[-104.73^\circ]$	1	0.727 $[46.85^\circ]$	0.395	2.31

(b)



(c)

Figure 3.8.2 Continued

Γ_s	Γ_{OUT}	Γ_L	Γ_{IN}	$ \Gamma_a $	(VSWR) _{in}
0.853 $[-27.25^\circ]$ for $\theta_1 = 0$	0.463 $[162.71^\circ]$	0.301 $[-153.73^\circ]$ for $\theta_2 = 0$	0.695 $[36.39^\circ]$	0.470	2.77
		0.428 $[176.66^\circ]$ for $\theta_2 = \pi/2$	0.719 $[31.29^\circ]$	0.371	2.18
		0.601 $[-167.19^\circ]$ for $\theta_2 = \pi$	0.781 $[33.11^\circ]$	0.320	1.94
		0.517 $[-145.76^\circ]$ for $\theta_2 = 3\pi/2$	0.762 $[37.91^\circ]$	0.460	2.70
0.467 $[-11.18^\circ]$ for $\theta_1 = \pi/2$	0.108 $[144.9^\circ]$	0.128 $[-27.91^\circ]$ for $\theta_2 = 0$	0.573 $[39.14^\circ]$	0.351	2.08
		0.162 $[121.62^\circ]$ for $\theta_2 = \pi/2$	0.612 $[31.30^\circ]$	0.320	1.94
		0.289 $[-168.07^\circ]$ for $\theta_2 = \pi$	0.689 $[34.54^\circ]$	0.447	2.62
		0.272 $[-108.27^\circ]$ for $\theta_2 = 3\pi/2$	0.662 $[41.79^\circ]$	0.470	2.77
0.421 $[-67.96^\circ]$ for $\theta_1 = \pi$	0.259 $[83.84^\circ]$	0.327 $[-49.23^\circ]$ for $\theta_2 = 0$	0.573 $[46.5^\circ]$	0.304	1.87
		0.066 $[-66.19^\circ]$ for $\theta_2 = \pi/2$	0.599 $[38.05^\circ]$	0.398	2.32
		0.295 $[-122.88^\circ]$ for $\theta_2 = \pi$	0.682 $[40.46^\circ]$	0.482	2.86
		0.435 $[-86.48^\circ]$ for $\theta_2 = 3\pi/2$	0.665 $[48^\circ]$	0.411	2.40
0.829 $[-56.44^\circ]$ for $\theta_1 = 3\pi/2$	0.533 $[104.73^\circ]$	0.501 $[-88.59^\circ]$ for $\theta_2 = 0$	0.684 $[49.28^\circ]$	0.390	2.28
		0.381 $[-110.25^\circ]$ for $\theta_2 = \pi/2$	0.693 $[43.43^\circ]$	0.477	2.82
		0.572 $[-118.83^\circ]$ for $\theta_2 = \pi$	0.761 $[44.22^\circ]$	0.449	2.63
		0.658 $[-101.54^\circ]$ for $\theta_2 = 3\pi/2$	0.757 $[49.53^\circ]$	0.311	1.90

(d)

Figure 3.8.2 Continued

Fig. 3.8.2b. For $\Gamma_L = \Gamma_{\text{OUT}}^*$, Fig. 3.8.2b shows the values of $(\text{VSWR})_{\text{out}}$, the corresponding values of Γ_{IN} , the values of $|\Gamma_a|$ [using (3.8.2)], and the resulting values of $(\text{VSWR})_{\text{in}}$ [using (3.8.1)]. From Fig. 3.8.2b, it is seen that with $\Gamma_L = \Gamma_{\text{OUT}}^*$, the input VSWR is approximately 2.3 for the four values of Γ_s .

For an output VSWR of 1.5, we have from (3.8.5) that $|\Gamma_b| = 0.2$. Then, using (3.8.7) and (3.8.8), the center and radius of the $(\text{VSWR})_{\text{out}} = 1.5$ circles are calculated and listed in Fig. 3.8.2c. The four $(\text{VSWR})_{\text{out}} = 1.5$ circles, as well as the four values where $\Gamma_L = \Gamma_{\text{OUT}}^*$ [i.e., $(\text{VSWR})_{\text{out}} = 1$], are drawn in Fig. 3.8.2c.

(c) The values of Γ_L on the $(\text{VSWR})_{\text{out}} = 1.5$ circles are given by

$$\Gamma_L = C_{V_o} + r_{V_o} e^{j\theta_2}$$

Four convenient values of Γ_L on the $(\text{VSWR})_{\text{out}} = 1.5$ circle correspond to $\theta_2 = 0, \pi/2, \pi$, and $3\pi/2$. In Fig. 3.8.2d, four values of Γ_L are calculated on each $(\text{VSWR})_{\text{out}} = 1.5$ circle, as well as the corresponding values of Γ_{IN} and $|\Gamma_a|$ and the resulting values of $(\text{VSWR})_{\text{in}}$. From the tabulated values, it is seen that there are several values of Γ_s and Γ_L that result in $(\text{VSWR})_{\text{in}}$ values around 1.9 with $(\text{VSWR})_{\text{out}} = 1.5$. For example, with $\Gamma_s = 0.421[-67.96^\circ]$ and $\Gamma_L = 0.327[-49.23^\circ]$, we have $(\text{VSWR})_{\text{out}} = 1.5$ and $(\text{VSWR})_{\text{in}} = 1.87$.

For further practice with this type of calculations, the reader is referred to Problem 3.30, where the analyses in this example are to be performed for the $G_A = G_{A,\text{max}} - 2$ dB circle (i.e., the $G_A = 7.66$ dB circle).

Since Γ_L and Γ_{IN} , as well as Γ_s and Γ_{OUT} , are related by bilinear transformations [see (3.2.5) and (3.2.6)], it follows that circles in the Γ_L plane map into circles in the Γ_{IN} plane, and circles in the Γ_s plane map into circles in the Γ_{OUT} plane. Specifically, these transformations are useful to map values of Γ_L on a constant G_p circle into a circle in the Γ_s plane, where $\Gamma_s = \Gamma_{\text{IN}}^*$ and values of Γ_s on a constant G_A circle into a circle in the Γ_L plane, where $\Gamma_L = \Gamma_{\text{OUT}}^*$.

In Appendix J it is shown that the values of Γ_L on a circle given by

$$|\Gamma_L - C_{oo}| = r_{oo}$$

map into a circle in the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane given by

$$|\Gamma_s - C_i| = r_i$$

where the center of the circle is given by

$$C_i = \frac{(1 - S_{22}C_{oo})(S_{11} - \Delta C_{oo})^* - r_{oo}^2 \Delta^* S_{22}}{|1 - S_{22}C_{oo}|^2 - r_{oo}^2 |S_{22}|^2} \quad (3.8.9)$$

and the radius is

$$r_i = \frac{r_{oo} |S_{12}S_{21}|}{||1 - S_{22}C_{oo}|^2 - r_{oo}^2 |S_{22}|^2|} \quad (3.8.10)$$

Similarly, a circle in the Γ_s plane given by

$$|\Gamma_s - C_{ii}| = r_{ii}$$

maps into a circle in the $\Gamma_L = \Gamma_{\text{OUT}}^*$ plane given by

$$|\Gamma_L - C_o| = r_o$$

where

$$C_o = \frac{(1 - S_{11}C_{ii})(S_{22} - \Delta C_{ii})^* - r_{ii}^2 \Delta^* S_{11}}{|1 - S_{11}C_{ii}|^2 - r_{ii}^2 |S_{11}|^2} \quad (3.8.11)$$

and

$$r_o = \frac{r_{ii} |S_{12}S_{21}|}{||1 - S_{11}C_{ii}|^2 - r_{ii}^2 |S_{11}|^2|} \quad (3.8.12)$$

Example 3.8.2

Design a microwave amplifier using a GaAs FET whose S parameters at 4 GHz, $V_{DS} = 2$ V, and $I_{DS} = 25$ mA are

$$\begin{aligned} S_{11} &= 0.55[-120^\circ] & S_{12} &= 0.14[30^\circ] \\ S_{21} &= 3.5[60^\circ] & S_{22} &= 0.2[-50^\circ] \end{aligned}$$

Analyze the trade-offs between operating power gain, stability, and VSWRs.

Solution. For this transistor, $K = 0.947$ and $\Delta = 0.521[-102.01^\circ]$. Since $K < 1$, the transistor is potentially unstable at 4 GHz. The centers and radii of the input and output stability circles are

$$\begin{aligned} C_s &= 16.47[130.7^\circ] & C_L &= 1.22[-59.25^\circ] \\ r_s &= 15.52 & r_L &= 2.12 \end{aligned}$$

The maximum stable gain is

$$G_{\text{MSG}} = \frac{|S_{21}|}{|S_{12}|} = \frac{3.5}{0.14} = 25 \text{ (or } 13.98 \text{ dB)}$$

Hence, we select a value of G_p lower than G_{MSG} . The value selected for this design is $G_p = 12$ dB, which is approximately 2 dB lower than G_{MSG} .

From (3.7.4) and (3.7.5), with $g_p = 15.849/(3.5)^2 = 1.294$, the center and radius of the $G_p = 12$ dB constant-gain circle are $C_p = 0.519[120.75^\circ]$ and $r_p = 0.639$. Figure 3.8.3a shows the $G_p = 12$ dB constant-gain circle and the output stability circle. Next we select values of Γ_L on the 12-dB gain circle that are far away from the output stability circle. The values of Γ_L on the 12-dB gain circle are given by $\Gamma_L = C_p + r_p e^{j\theta_1}$. The design procedure is described for two values of Γ_L , shown as point *a* (where $\theta_1 = 0$) and point *b* (where $\theta_1 = 3\pi/2$). At point *a* the value of Γ_L is $0.582[50.05^\circ]$, and at point *b* the value is $0.328[-143.98^\circ]$. Next we analyze the mapping of the $G_p = 12$ dB constant-gain circle onto the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane. Using (3.8.9) and (3.8.10) with $C_{oo} = C_p = 0.519[120.75^\circ]$ and $r_{oo} = r_p = 0.639$, we obtain $C_i = 0.8[130.7^\circ]$ and $r_i = 0.338$. This circle is drawn in the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane in Fig. 3.8.3b.

Several calculations are shown in Fig. 3.8.3c. Using (3.2.5), the second column shows the specific mapping of points *a* and *b* in the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane. These points are denoted by *a'* and *b'* in Fig. 3.8.3b. The third column in Fig. 3.8.3c shows that $(\text{VSWR})_{\text{in}} = 1$ for $\Gamma_s = \Gamma_{\text{IN}}^*$. Using (3.2.6), the fourth column shows the corresponding values of Γ_{out} . From (3.8.6), the magnitude of $|\Gamma_b|$ is calculated in the fifth column. Finally, from (3.8.5),

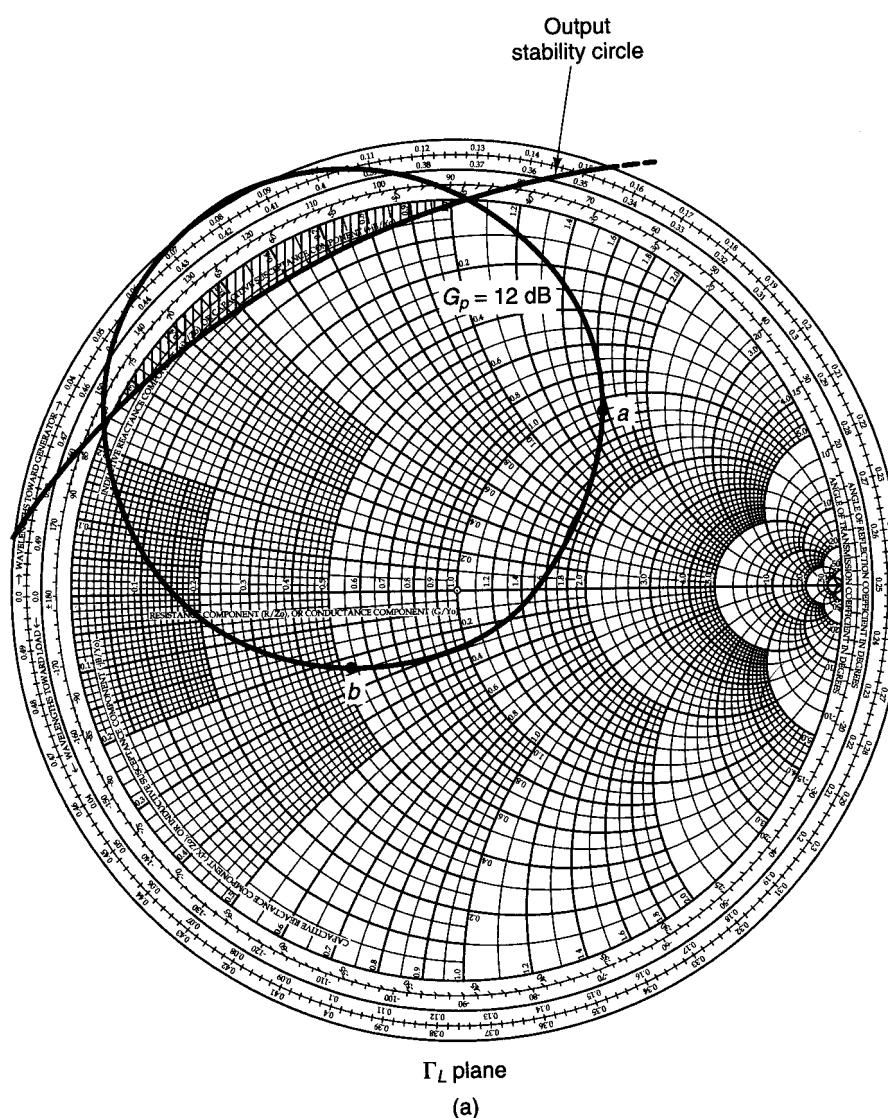


Figure 3.8.3 (a) The $G_p = 12 \text{ dB}$ constant-gain circle and the output stability circle;

(b) mapping of the $G_p = 12 \text{ dB}$ constant-gain circle onto the $\Gamma_s = \Gamma_{IN}^*$ plane, the input stability circle, and the constant $(\text{VSWR})_{in} = 1.5$ circles; (c) calculations for Γ_L values at point *a* and *b* with $\Gamma_s = \Gamma_{IN}^*$ [i.e., $(\text{VSWR})_{in} = 1$]; (d) calculations for Γ_L values at points *a* and *b* with $(\text{VSWR})_{in} = 1.5$.

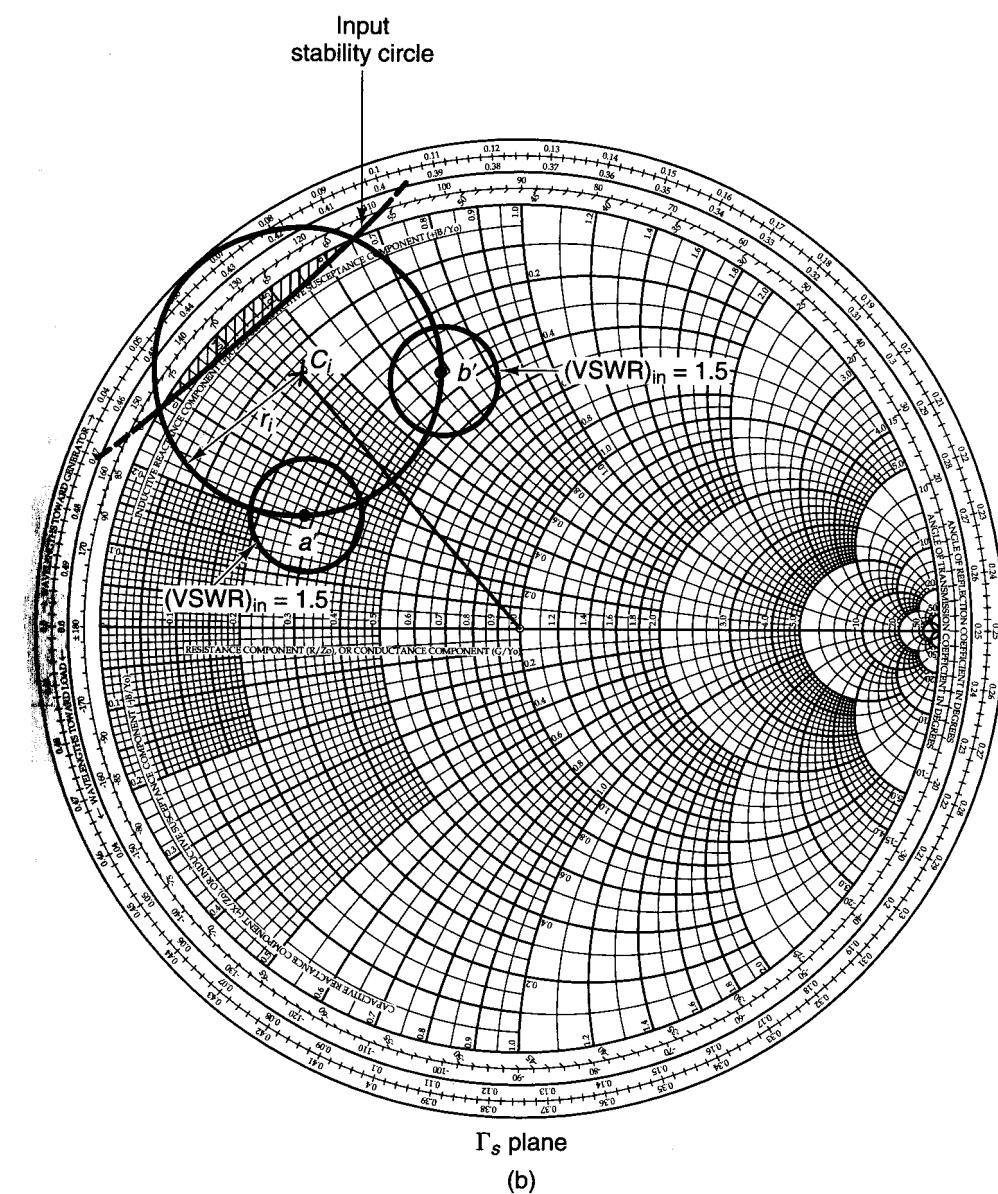


Figure 3.8.3 Continued

Γ_L	$\Gamma_s = \Gamma_{IN}^*$	(VSWR) _{in}	Γ_{OUT}	$ \Gamma_b $	(VSWR) _{out}
$0.582 50.05^\circ$ for $\theta_1 = 0$	$0.588 152.74^\circ$	1	$0.528 -85.97^\circ$	0.448	2.62
$0.328 -143.98^\circ$ for $\theta_1 = 270^\circ$	$0.625 107.15^\circ$	1	$0.399 -143.55^\circ$	0.447	2.62

(c)

Γ_L	Γ_{IN}	C_{Vi} and r_{Vi} for (VSWR) _{in} = 1.5	Γ_s	Γ_{OUT}	$ \Gamma_b $	(VSWR) _{out}
$0.582 50.05^\circ$ for $\theta_1 = 0$	$0.588 152.74^\circ$	$C_{Vi} = 0.572 152.74^\circ$ $r_{Vi} = 0.133$	$0.458 145.09^\circ$ for $\theta_2 = 0$ $0.643 142.16^\circ$ for $\theta_2 = \pi/2$ $0.692 157.78^\circ$ for $\theta_2 = \pi$ $0.525 165.77^\circ$ for $\theta_2 = 3\pi/2$	$0.410 -90.33^\circ$ $0.483 -74.36^\circ$ $0.569 -97.81^\circ$ $0.621 -80.87^\circ$	0.454 0.324 0.572 0.451	2.66 1.96 3.67 2.64
$0.328 -143.98^\circ$ for $\theta_1 = 270^\circ$	$0.625 -107.15^\circ$	$C_{Vi} = 0.61 107.15^\circ$ $r_{Vi} = 0.124$	$0.586 95.48^\circ$ for $\theta_2 = 0$ $0.730 104.28^\circ$ for $\theta_2 = \pi/2$ $0.657 117.53^\circ$ for $\theta_2 = \pi$ $0.493 111.40^\circ$ for $\theta_2 = 3\pi/2$	$0.293 -156.53^\circ$ $0.485 -156.23^\circ$ $0.315 -125.74^\circ$ $0.496 -130.80^\circ$	0.324 0.460 0.453 0.571	1.96 2.70 2.66 3.66

(d)

Figure 3.8.3 Continued

the sixth column lists the values of (VSWR)_{out}. The values of Γ_s at a' and b' are far away from the input stability circle. Hence, if (VSWR)_{out} = 2.62 is satisfactory, the design for $G_p = 12$ dB can be implemented with the values of Γ_s and Γ_L listed in Fig. 3.8.3c.

A smaller value of (VSWR)_{out} can be obtained by relaxing the input VSWR specifications. To illustrate this point, we design the amplifier with (VSWR)_{in} = 1.5. The design calculations are shown in Fig. 3.8.3d. The center and radius of the constant (VSWR)_{in} = 1.5 circle are made using (3.8.3) and (3.8.4) with $|\Gamma_a| = 0.2$. These values are listed in the third column in Fig. 3.8.3d, and the resulting (VSWR)_{in} = 1.5 circles are plotted in Fig. 3.8.3b.

The values of Γ_s on the (VSWR)_{in} = 1.5 circle are given by

$$\Gamma_s = C_{Vi} + r_{Vi}e^{j\theta_2}$$

Four convenient values on the (VSWR)_{in} = 1.5 circles are listed in the fourth column of Fig. 3.8.3d. They correspond to $\theta_2 = 0, \pi/2, \pi$, and $3\pi/2$. Using (3.2.6), Γ_{OUT} is calculated in the fifth column. From (3.8.6), $|\Gamma_b|$ is calculated in the sixth column. Finally, using (3.8.5), (VSWR)_{out} is calculated in the seventh column. From Fig. 3.8.3d, it is seen that the values associated with (VSWR)_{out} = 1.96 are far away from the input stability circle. Thus, a microwave amplifier with $G_p = 12$ dB can be designed with $\Gamma_L = 0.582 |35^\circ$ and $\Gamma_s = 0.643 |142.16^\circ$, or with $\Gamma_L = 0.328 |-143.98^\circ$ and $\Gamma_s = 0.586 |95.48^\circ$.

It is of interest to analyze the design for a value of G_p greater than G_{MSG} (e.g., $G_p = 15$ dB). From (3.7.4), with $g_p = 31.623/(3.5)^2 = 2.581$, the center and radius of the $G_p = 15$ dB constant-gain circle are $C_p = 1.8 |120.75^\circ$ and $r_p = 1.12$. Using (3.8.9) and (3.8.10), the 15-dB gain circle is mapped into a circle in the $\Gamma_s = \Gamma_{IN}^*$ plane with center and radius given by $C_i = 1.53 |130.7^\circ$ and $r_i = 0.652$.

The 15-dB gain circle is shown in Fig. 3.8.4a and its mapping in the $\Gamma_s = \Gamma_{IN}^*$ plane in Fig. 3.8.4b. Two values of Γ_L are selected on the 15-dB gain circle. These points are denoted by a and b in Fig. 3.8.4a. The mapping of the points a and b is shown in the $\Gamma_s = \Gamma_{IN}^*$ plane in Fig. 3.8.4b. The input stability circle is also shown in Fig. 3.8.4b.

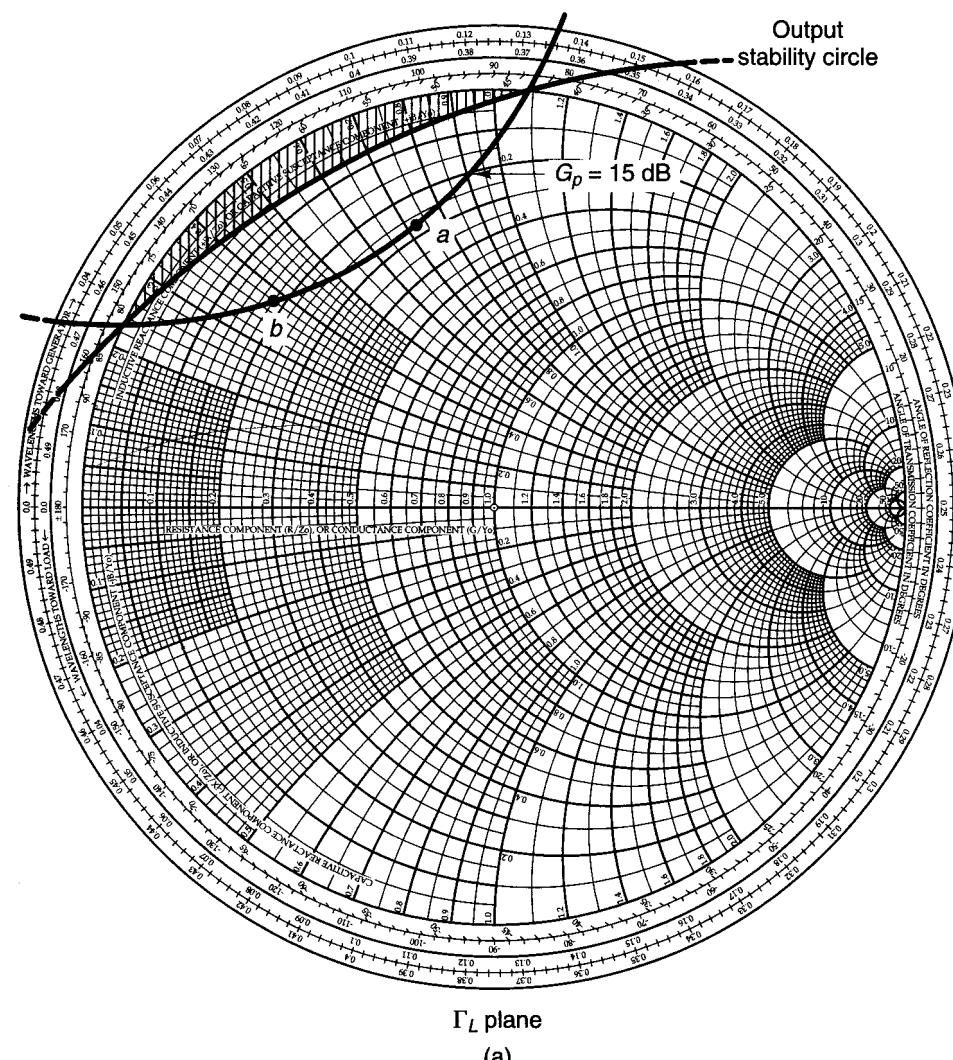
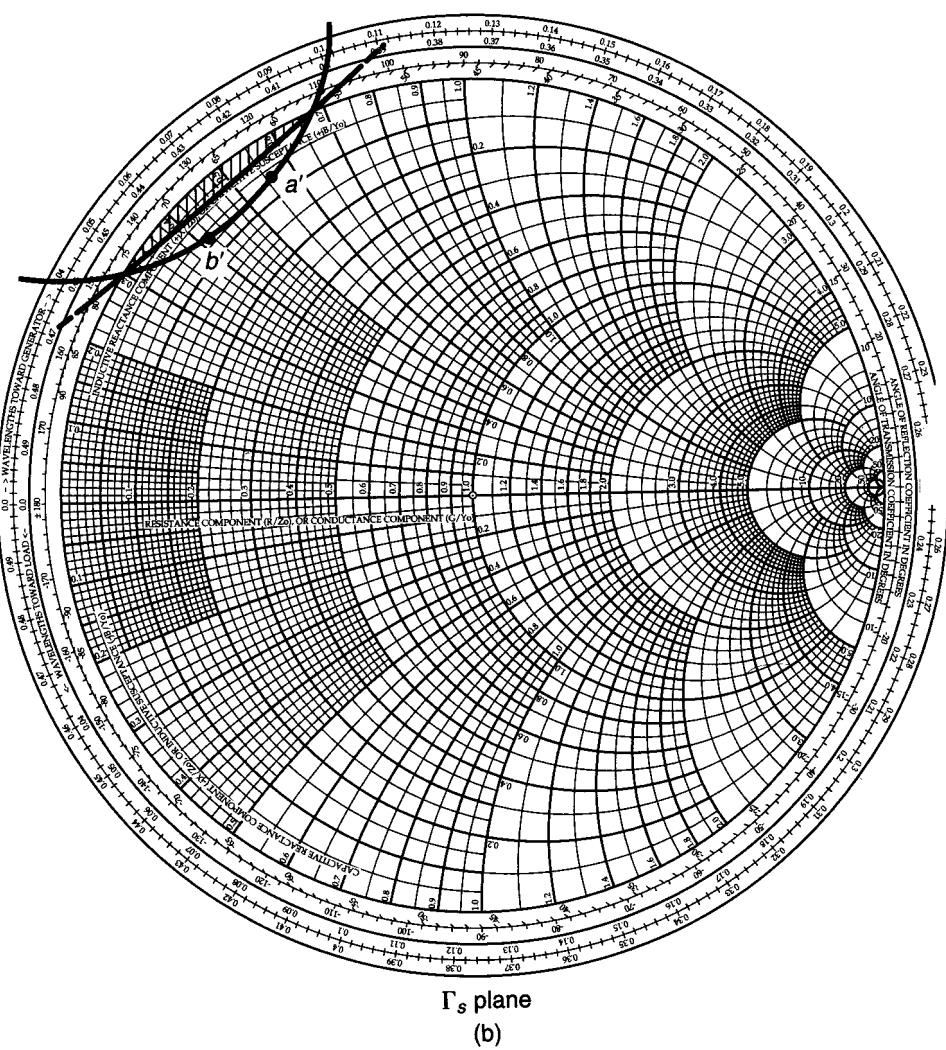


Figure 3.8.4 (a) The $G_p = 15$ dB constant-gain circle and the output stability circle; (b) mapping of the $G_p = 15$ dB constant-gain circle onto the $\Gamma_s = \Gamma_{IN}^*$ plane, and the input stability circle; (c) calculations for Γ_L values at points a and b with $\Gamma_s = \Gamma_{IN}^*$ [i.e., (VSWR)_{in} = 1].



Γ_L	$\Gamma_s = \Gamma_{IN}^*$	$(VSWR)_{in}$	Γ_{OUT}	$ \Gamma_b $	$(VSWR)_{out}$
0.717 106.22°	0.884 134.73°	1	0.878 -109.62°	0.449	2.63
0.730 137.37°	0.906 121.68°	1	0.884 -133.66°	0.453	2.65

(c)

Figure 3.8.4 Continued

plane as a' and b' in Fig. 3.8.4b. The values of Γ_L at points a and b are listed in Fig. 3.8.4c, as well as the design calculations. The third column shows that $(VSWR)_{in} = 1$ for $\Gamma_s = \Gamma_{IN}^*$. The fourth column shows the corresponding values of Γ_{OUT} , the fifth column the values of $|\Gamma_b|$, and the sixth column the values of $(VSWR)_{out}$.

An examination of Fig. 3.8.4b shows that the points a' and b' are very close to the unstable region, and the resulting $(VSWR)_{out}$ is greater than 2.63. This is typical of a design when the value of G_p is selected close to the value of G_{MSG} or greater than G_{MSG} .

Further improvements can be done by performing calculations of the type illustrated in Figs. 3.8.3c and 3.8.3d. However, a CAD program can simplify the task considerably. In Appendix "Computer-Aided Designs," Example CAD.2, it is shown that with $\Gamma_L = 0.328 | -143.98^\circ$ and $\Gamma_s = 0.612 | 95.2^\circ$ it follows that $G_p = 12$ dB, $(VSWR)_{in} = 1.5$, and $(VSWR)_{out} = 1.93$. This is the smallest $(VSWR)_{out}$ that can be obtained with Γ_s on the $(VSWR)_{in} = 1.5$ circle shown in Fig. 3.8.3b. Furthermore, a CAD optimization to minimize the VSWRs shows that with $\Gamma_L = 0.319 | -176.51^\circ$ and $\Gamma_s = 0.642 | 102.61^\circ$, it follows that $G_p = 12.7$ dB, $(VSWR)_{in} = 1.52$, and $(VSWR)_{out} = 1.58$.

3.9 DC BIAS NETWORKS

It has been said that the least considered factor in microwave transistor amplifier design is the bias network [3.4]. While considerable effort is spent in designing for a given gain, noise figure, and bandwidth, little effort is spent in the dc bias network. The cost per decibel of microwave power gain or noise figure is high, and the designer cannot sacrifice the amplifier performance by having a poor dc bias design.

The purpose of a good dc bias design is to select the proper quiescent point and hold the quiescent point constant over variations in transistor parameters and temperature. A resistor bias network can be used with good results over moderate temperature changes. However, an active bias network is usually preferred for large temperature changes.

In the discussion that follows, we first consider the dc bias design for BJTs and then the bias design of GaAs FETs.

BJT Bias Networks

At low frequencies, a bypassed emitter resistor is an important contributor to the quiescent-point stability. At microwave frequencies, the bypass capacitor, which is in parallel with the emitter resistor, can produce oscillations by making the input port unstable at some frequencies. Furthermore, an emitter resistor will degrade the noise performance of the amplifier. Therefore, in most microwave transistor amplifiers, especially in the gigahertz region, the emitter lead of the transistor is grounded.

At microwave frequencies, the transistor parameters that are affected most by temperature are I_{CBO} , h_{FE} , and V_{BE} . The conventional reverse current I_{CBO} (i.e., I_{CBO} at low frequencies) doubles every 10°C rise in temperature. That is,

$$I_{CBO,T_2} = I_{CBO,T_1} 2^{(T_2 - T_1)/10}$$

where I_{CBO,T_2} and I_{CBO,T_1} are the values of I_{CBO} at temperatures T_2 and T_1 , respectively. The temperature T_1 is usually the temperature at which the manufacturer measures I_{CBO} . This temperature is usually 25°C.

A microwave transistor has a more complicated reverse current flow. The reverse current flow of a microwave transistor is composed of two components; one is the conventional I_{CBO} and the other is a surface current, I_s , that flows across the top of the silicon lattice. The reverse current in a microwave transistor, which is referred to simply as I_{CBO} , increases at a rate much slower than the conventional I_{CBO} . A typical plot of the reverse current versus temperature for a microwave transistor is shown in Fig. 3.9.1. The conventional I_{CBO} slope is also shown in the figure for comparison.

The base-to-emitter voltage V_{BE} has a negative temperature coefficient, approximately given by

$$\frac{\Delta V_{BE}}{\Delta T} \approx -2 \times 10^{-3} \frac{V}{^{\circ}C}$$

The dc value of the current gain h_{FE} is defined as the value of the collector-to-base current at a constant value of V_{CE} . That is,

$$h_{FE} = \left. \frac{I_C}{I_B} \right|_{V_{CE}=\text{constant}}$$

The dc value of h_{FE} is typically found to increase linearly with temperature at the rate of 0.5%/°C.

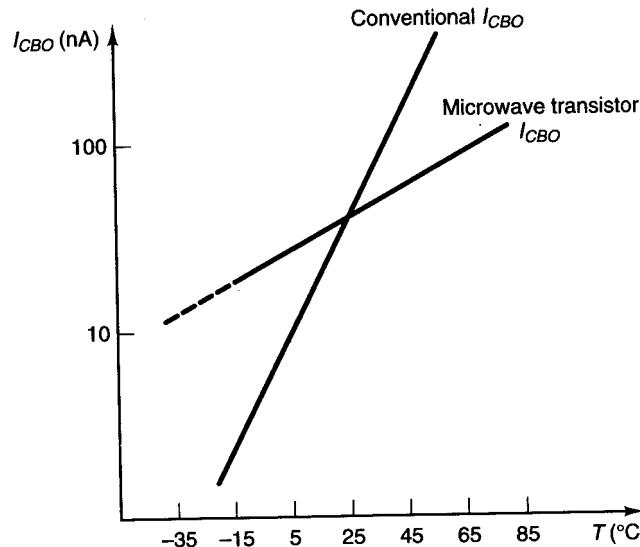


Figure 3.9.1 Typical reverse current versus temperature for a microwave transistor.

In order to find the change in collector current as a function of temperature in a dc bias network, we first find the expression for the collector current valid for any temperature. Then, observing that the temperature-sensitive parameters are I_{CBO} , h_{FE} , and V_{BE} , we can write

$$I_C = f(I_{CBO}, h_{FE}, V_{BE})$$

and

$$\Delta I_C = \left(\frac{\Delta I_C}{\Delta I_{CBO}} \right) \Big|_{\Delta V_{BE}=0} \Delta I_{CBO} + \left(\frac{\Delta I_C}{\Delta h_{FE}} \right) \Big|_{\Delta I_{CBO}=0} \Delta h_{FE} + \left(\frac{\Delta I_C}{\Delta V_{BE}} \right) \Big|_{\Delta h_{FE}=0} \Delta V_{BE} \quad (3.9.1)$$

Defining the stability factors as

$$S_i = \left. \frac{\Delta I_C}{\Delta I_{CBO}} \right|_{\Delta h_{FE}=0, \Delta V_{BE}=0}$$

$$S_{h_{FE}} = \left. \frac{\Delta I_C}{\Delta h_{FE}} \right|_{\Delta I_{CBO}=0, \Delta V_{BE}=0}$$

and

$$S_{V_{BE}} = \left. \frac{\Delta I_C}{\Delta V_{BE}} \right|_{\Delta h_{FE}=0, \Delta I_{CBO}=0}$$

we can write (3.9.1) in the form

$$\Delta I_C = S_i \Delta I_{CBO} + S_{h_{FE}} \Delta h_{FE} + S_{V_{BE}} \Delta V_{BE} \quad (3.9.2)$$

For a given dc bias network, the stability factors can be calculated and (3.9.2) can be used to predict the variations of I_C with temperature. In a design procedure, the maximum variation of I_C in a temperature range can be selected and (3.9.2) can be used to find the required stability factors. In turn, the stability factors together with the Q -point location will fix the value of the resistors in the bias network.

Two grounded-emitter dc bias networks that can be used at microwave frequencies are shown in Fig. 3.9.2. The network in Fig. 3.9.2b produces lower values of resistance and therefore is more compatible with thin- or thick-film resistor values.

Example 3.9.1

Design the dc bias network shown in Fig. 3.9.2b for $V_{CE} = 10$ V and $I_C = 10$ mA. Assume that $I_{CBO} = 0$, $V_{BE} = 0.7$ V, and $h_{FE} = 50$.

Solution. In this example we follow a procedure that results in good stability factors. Let the supply voltage V_{CC} be 20 V. The base current (I_B) is

$$I_B = \frac{I_C}{h_{FE}} = \frac{10 \times 10^{-3}}{50} = 200 \mu A$$

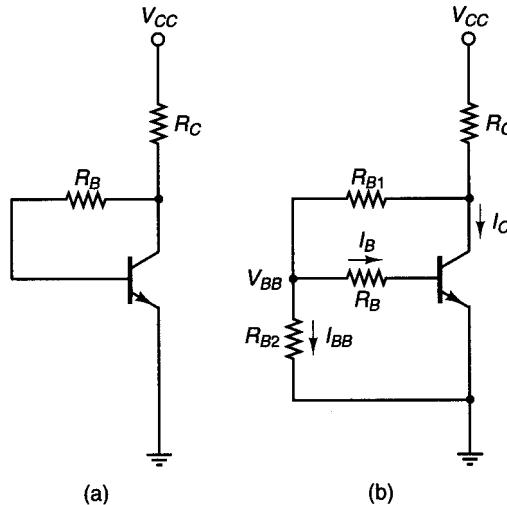


Figure 3.9.2 (a) Voltage feedback bias network; (b) voltage feedback bias network with constant-base current source.

Assuming \$V_{BB}\$ to be 2 V, we find that

$$R_B = \frac{V_{BB} - V_{BE}}{I_B} = \frac{2 - 0.7}{200 \times 10^{-6}} = 6.5 \text{ k}\Omega$$

\$R_{B2}\$ is calculated assuming that \$I_{BB} = 1\$ mA (i.e., \$I_{BB} = 5I_B\$)—namely,

$$R_{B2} = \frac{V_{BB}}{I_{BB}} = \frac{2}{1 \times 10^{-3}} = 2 \text{ k}\Omega$$

\$R_{B1}\$ is obtained from

$$R_{B1} = \frac{V_{CE} - V_{BB}}{I_{BB} + I_B} = \frac{10 - 2}{(1 + 0.2) \times 10^{-3}} = 6.66 \text{ k}\Omega$$

and \$R_C\$ is obtained from

$$R_C = \frac{V_{CC} - V_{CE}}{I_C + I_{BB} + I_B} = \frac{20 - 10}{(10 + 1 + 0.2) \times 10^{-3}} = 893 \Omega$$

The assumption \$I_{BB} \gg I_B\$ and \$V_{BB} \approx 10\% V_{CC}\$ produces good stability factors.

At the lower microwave frequencies, the dc biasing network shown in Fig. 3.9.3 with a bypassed emitter resistor can be used. The bypassed emitter resistor provides excellent stability. For this network, it is easy to show that

$$I_C = \frac{h_{FE}(V_{TH} - V_{BE})}{R_{TH} + (h_{FE} + 1)R_E} + \frac{(h_{FE} + 1)I_{CBO}(R_{TH} + R_E)}{R_{TH} + (h_{FE} + 1)R_E}$$

Figure 3.9.3 A dc bias network with a bypassed emitter resistor.

where

$$V_{TH} = \frac{V_{CC}R_2}{R_1 + R_2}$$

and

$$R_{TH} = \frac{R_1R_2}{R_1 + R_2}$$

The stability factors are

$$S_i = \frac{(h_{FE} + 1)(R_{TH} + R_E)}{R_{TH} + (h_{FE} + 1)R_E}$$

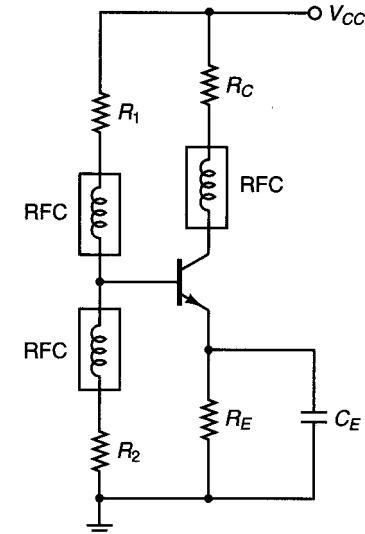
$$S_{h_{FE}} \approx \frac{I_{C1}}{h_{FE}} \frac{S_{i2}}{h_{FE,2}} \quad (3.9.3)$$

and

$$S_{V_{BE}} = \frac{-h_{FE}}{R_{TH} + (h_{FE} + 1)R_E}$$

In (3.9.3), \$\Delta h_{FE} = h_{FE,2} - h_{FE}\$ and \$S_{i2}\$ is the value of \$S_i\$ with \$h_{FE} = h_{FE,2}\$.

An active dc biasing network is shown in Fig. 3.9.4. A pnp BJT is used to stabilize the operating point of the microwave transistor. The bypass capacitors \$C_1\$ and \$C_2\$ are typically 0.01-\$\mu\$F disk capacitors. The radio frequency chokes (RFCs) are typically made of two or three turns of No. 36 enameled wire on 0.1-in. air core. The operation of the network is as follows. If \$I_{C2}\$ tends to increase, the current \$I_3\$ increases and the emitter-to-base voltage of \$Q_1\$ (\$V_{EB,1}\$) decreases. The decrease of \$V_{EB,1}\$ decreases \$I_{E1}\$, which in turn decreases \$I_{C2}\$ and \$I_{B2}\$. The decrease in \$I_{B2}\$ and \$I_{C2}\$ produces the desired bias stability.



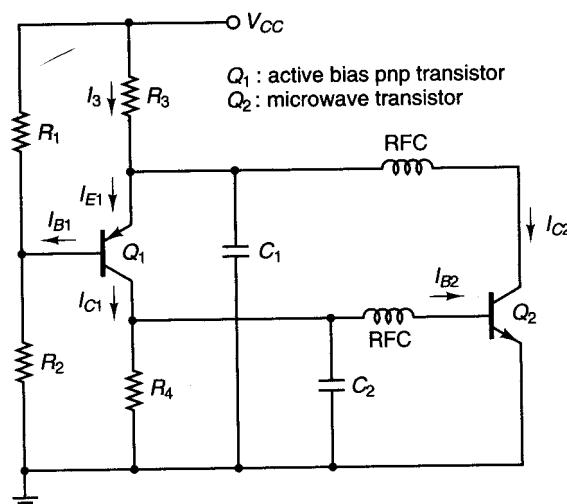


Figure 3.9.4 Active bias network for a BJT.

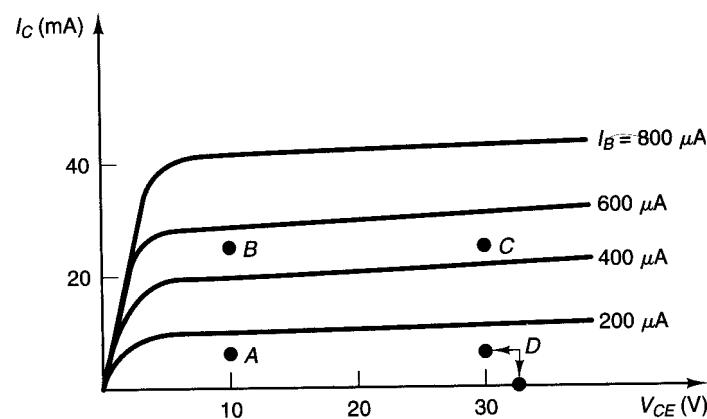


Figure 3.9.5 Selection of the dc operating point.

The selection of the dc quiescent point for a BJT depends on the particular application. For low-noise and low-power applications, the quiescent point *A* in Fig. 3.9.5 is recommended. At *A*, the BJT operates at low values of collector current. For low noise and higher power gain, the quiescent point at *B* is recommended. For high output power, in class A operation, the quiescent point at *C* is recommended. For higher output power and higher efficiency, the BJT is operated in class AB or B, using the quiescent point at *D*.

Example 3.9.2

(a) Design the bias circuit in Fig. 3.9.3 to have a quiescent point at $V_{CE} = 8\text{ V}$ and $I_C = 2\text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.7\text{ V}$.

(b) Design the active bias network in Fig. 3.9.4 to set the quiescent point of transistor Q_2 at $V_{CE2} = 8\text{ V}$ and $I_{C2} = 2\text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.7\text{ V}$.

Solution. (a) Let $V_{CC} = 15\text{ V}$ in Fig. 3.9.3. Then, with $I_C \approx I_E$, the dc load-line equation is

$$V_{CC} = V_{CE} + I_C(R_C + R_E)$$

or

$$R_C + R_E = \frac{15 - 8}{2 \times 10^{-3}} = 3.5\text{ k}\Omega \quad (3.9.4)$$

The value of R_E is usually selected using the assumption that the voltage across R_E is 10% to 20% of V_{CC} . Selecting 10% of V_{CC} , we obtain

$$R_E = \frac{10\% V_{CC}}{I_C} = \frac{0.1(15)}{2 \times 10^{-3}} = 750\Omega$$

Then, from (3.9.4),

$$R_C = 3500 - 750 = 2.75\text{ k}\Omega$$

A Thévenin's equivalent circuit between the base and ground of the transistor gives

$$V_{TH} = \frac{V_{CC} R_2}{R_2 + R_1} \quad (3.9.5)$$

and

$$R_{TH} = \frac{R_1 R_2}{R_1 + R_2} \quad (3.9.6)$$

From (3.9.5) and (3.9.6), we can solve for R_1 and R_2 in terms of V_{TH} and R_{TH} —namely,

$$R_1 = R_{TH} \frac{V_{CC}}{V_{TH}} \quad (3.9.7)$$

and

$$R_2 = \frac{R_{TH}}{1 - \frac{V_{TH}}{V_{CC}}} \quad (3.9.8)$$

For good beta stability, the value of R_{TH} is selected such that $\beta R_E = 10 R_{TH}$, or

$$R_{TH} = \frac{\beta R_E}{10} = \frac{100(750)}{10} = 7.5\text{ k}\Omega$$

The value of V_{TH} is calculated using the loop equation:

$$V_{TH} = I_B R_{TH} + 0.7 + I_E R_E = \frac{2 \times 10^{-3}}{100} (7.5 \times 10^3) + 0.7 + 2 \times 10^{-3}(750) = 2.35\text{ V}$$

From (3.9.7) and (3.9.8), it follows that

$$R_1 = (7.5 \times 10^3) \frac{15}{2.35} = 47.9\text{ k}\Omega$$

and

$$R_2 = \frac{7.5 \times 10^3}{1 - \frac{2.35}{15}} = 8.9 \text{ k}\Omega$$

(b) Let $V_{CC} = 15 \text{ V}$ in Fig. 3.9.4. The base voltage of transistor Q_2 , which is equal to the voltage across R_4 , is 0.7 V . The current I_3 is equal to the sum of I_{C1} and I_{C2} (where $I_{C2} = 2 \text{ mA}$). Designing for I_3 equal to 4 mA , it follows that

$$I_{C1} = I_3 - I_{C2} = 4 \times 10^{-3} - 2 \times 10^{-3} = 2 \text{ mA}$$

Then

$$R_4 = \frac{0.7}{I_{C1}} = \frac{0.7}{2 \times 10^{-3}} = 350 \Omega$$

Since the collector voltage of transistor Q_2 is 8 V , R_3 is calculated from

$$R_3 = \frac{V_{CC} - 8}{I_3} = \frac{15 - 8}{4 \times 10^{-3}} = 1.75 \text{ k}\Omega$$

In the active bias circuit of Fig. 3.9.4, good beta stability is obtained by letting the current in R_1 and R_2 be 20 times the base current of transistor Q_1 . That is,

$$I_{R_1} \approx I_{R_2} = 20I_{B1} = 20\left(\frac{2 \times 10^{-3}}{100}\right) = 0.4 \text{ mA}$$

Since the base voltage of transistor Q_1 is a 7.3 V , we can write

$$R_1 = \frac{V_{CC} - 7.3}{I_{R_1}} = \frac{15 - 7.3}{0.4 \times 10^{-3}} = 19.25 \text{ k}\Omega$$

and

$$R_2 = \frac{7.3}{I_{R_2}} = \frac{7.3}{0.4 \times 10^{-3}} = 18.25 \text{ k}\Omega$$

In the active bias circuit, the resistors R_2 and R_3 are implemented using a potentiometer in order to set the quiescent point of transistor Q_2 at exactly 8 V and 2 mA . An optimization and beta stability analysis of the active bias circuit in this design is performed in Appendix "Computer-Aided Designs," Example CAD.1.

GaAs FET Bias Networks

The GaAs FETs can be biased in several ways. Five basic dc network configurations for GaAs FET amplifiers are shown in Fig. 3.9.6 [3.5]. The dc bias network in Fig. 3.9.6a requires a bipolar power source, while the networks in Figs. 3.9.6b to 3.9.6e require a unipolar supply. The column "How" in Fig. 3.9.6 indicates the polarity of the sources, as well as the sequence in which the voltages must be applied to prevent transient burnout of the GaAs FET device during turn-on. For example, in the dc bias network in Fig. 3.9.6a, if the drain is biased

Figure	How	Amplifier characteristics	Power supply used
(a)	Apply V_G , then V_D	Low noise High gain High power High efficiency	Bipolar, Minimum source inductance
(b)	Apply V_S , then V_D	[same as (a)]	Positive supply
(c)	Apply V_S , then V_G	[same as (a)]	Negative supply
(d)	Apply V_D	Low noise High gain High power Lower efficiency Gain easily adjusted by varying R_S	Unipolar, incorporating R_S automatic transient protection
(e)	Apply V_G	[same as (d)]	Negative unipolar, incorporating R_S

Figure 3.9.6 Five basic dc bias networks. (From G. D. Vendelin [3.5]; reproduced with permission of *Microwaves & RF*)

positive before the gate, the transistor will operate momentarily beyond its safe operating region. Therefore, the proper turn-on sequence is: first apply a negative bias to the gate (i.e., $V_G < 0$) and then apply the drain voltage ($V_D > 0$). One method to accomplish the previous turn-on procedure is to turn both sources at the same time and to include a long RC time constant network in the V_D supply and a short RC time constant network in the negative supply V_G .

The bias networks in Figs. 3.9.6d and 3.9.6e use a source resistor. The source resistor provides automatic transient protection. However, the source resistor will degrade the noise-figure performance, and the source bypass capacitor can cause low-frequency oscillations.

The decoupling capacitors shown in Fig. 3.9.6 are sometimes shunted with zener diodes. The zener diodes provide additional protection against transients, reverse biasing, and overvoltage.

The dc bias network of a GaAs FET must provide a stable quiescent point. It is not difficult to show that the negative feedback resistor R_s decreases the effect of variations of I_D with respect to temperature and I_{DSS} .

The selection of the dc quiescent point in a GaAs FET depends on the particular application. Figure 3.9.7 shows typical GaAs FET characteristics with four quiescent points located at A, B, C, and D.

For low-noise, low-power application, the quiescent point A is recommended. At A, the FET operates at a low value of current (i.e., $I_{DS} \approx 0.15I_{DSS}$).

For low noise and higher power gain, the recommended quiescent point is at B. The bias voltage remains the same as for point A, but the drain current is increased to $I_{DS} \approx 0.9I_{DSS}$.

The GaAs FET output power level can be increased by selecting the quiescent point at C with $I_{DS} \approx 0.5I_{DSS}$. The quiescent point at C maintains class A operation. For higher efficiency, or to operate the GaAs FET in class

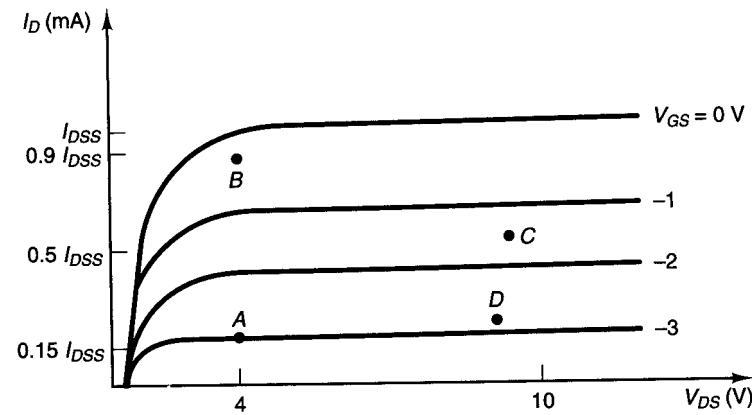


Figure 3.9.7 Typical GaAs FET characteristics and recommended quiescent points.

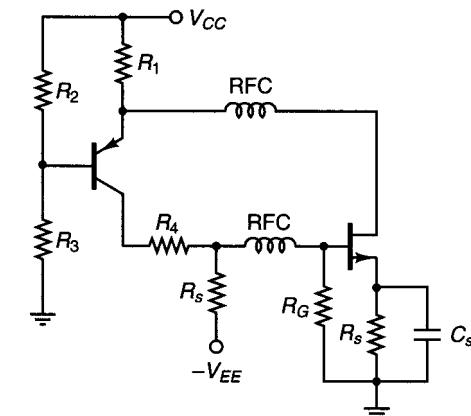


Figure 3.9.8 Active bias for a common-source GaAs FET.

AB or B, the drain-to-source current must be decreased and the quiescent point D is recommended.

An active bias network for a common-source GaAs FET is shown in Fig. 3.9.8.

PROBLEMS

- 3.1 (a)** Show that $G_T \leq G_A$ and $G_T \leq G_p$. When is the equality sign satisfied?
(b) Show that (3.2.3) can be obtained from (3.2.1) when $\Gamma_s = \Gamma_{IN}^*$, and (3.2.4) from (3.2.2) when $\Gamma_L = \Gamma_{OUT}^*$.
- 3.2 (a)** Show that the transducer power gain is given by $G_T = |S_{21}|^2$ when the source and load impedances to the transistor are equal to the reference impedance Z_0 (usually 50Ω).
(b) Determine the expression for G_p and G_A when the source and load impedances are real and equal to Z_0 .
- 3.3 (a)** A microwave amplifier diagram is shown in Fig. 3.2.2. Determine G_T , G_A , and G_p if $\Gamma_s = 0.49 \angle -150^\circ$, $\Gamma_L = 0.56 \angle 90^\circ$, and the S parameters of the transistor are

$$S_{11} = 0.54 \angle 165^\circ \quad S_{12} = 0.09 \angle 20^\circ$$

$$S_{21} = 2 \angle 30^\circ \quad S_{22} = 0.5 \angle -80^\circ$$

- (b)** Calculate P_{AVS} , P_{IN} , P_{AVN} , and P_L if $E_1 = 10 \angle 30^\circ$, $Z_1 = 50 \Omega$, and $Z_2 = 50 \Omega$.

- 3.4** The S parameters of a transistor are

$$S_{11} = 0.7 \angle 30^\circ \quad S_{12} = 0$$

$$S_{21} = 4 \angle 90^\circ \quad S_{22} = 0.5$$

The transistor is used in the amplifier shown in Fig. P3.4, where the output matching network produces $\Gamma_L = 0.5 \angle 90^\circ$. Determine the values of G_T , G_p , and G_A .

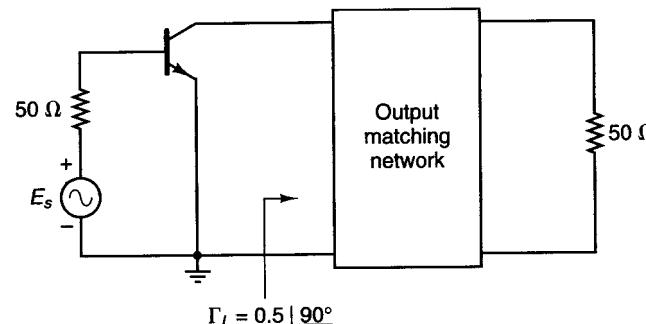


Figure P3.4

3.5 In each of the stability circle drawings shown in Fig. P3.5, indicate clearly the possible locations for a stable source reflection coefficient.

3.6 Two output stability circles are shown in Fig. P3.6. Determine the stable region for the load reflection coefficient.

3.7 The scattering parameters for three different transistors are given below. Determine the stability in each case and in a potentially unstable case, draw the input and output stability circles.

(a) $S_{11} = 0.674 | -152^\circ$
 $S_{12} = 0.075 | 6.2^\circ$
 $S_{21} = 1.74 | 36.4^\circ$
 $S_{22} = 0.6 | -92.6^\circ$

(b) $S_{11} = 0.385 | -55^\circ$
 $S_{12} = 0.045 | 90^\circ$
 $S_{21} = 2.7 | 78^\circ$
 $S_{22} = 0.89 | -26.5^\circ$

(c) $S_{11} = 0.7 | -50^\circ$
 $S_{12} = 0.27 | 75^\circ$
 $S_{21} = 5 | 120^\circ$
 $S_{22} = 0.6 | 80^\circ$

3.8 The S parameters of a GaAs FET at a given Q point are

f (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
4	0.9 -67°	0.076 43°	2.3 118°	0.68 -39°
6	0.84 -97°	0.112 24°	2.06 87°	0.6 -58°
8	0.73 -140°	0.135 -5°	2.04 53°	0.47 -85°
10	0.67 -178°	0.146 -27°	1.81 18°	0.42 -120°
14	0.63 115°	0.133 -66°	1.42 -38°	0.36 -172°

Draw the input stability circles (at each frequency) in a Smith chart and the output stability circles in another Smith chart. Indicate the unstable regions.

3.9 This problem analyzes some interesting (theoretical) cases of the stability circles. The S parameters of several two-port networks are

- (a) $S_{11} = 1/\sqrt{2}$, $S_{12} = 1/\sqrt{2} | -180^\circ$, $S_{21} = 1/\sqrt{2}$, and $S_{22} = 1/\sqrt{2}$.
- (b) $S_{11} = 1/\sqrt{2}$, $S_{12} = 1.707$, $S_{21} = 1.707$, and $S_{22} = 1/\sqrt{2}$.
- (c) $S_{11} = 1/\sqrt{2}$, $S_{12} = 0.292$, $S_{21} = 0.292$, and $S_{22} = 1/\sqrt{2}$.
- (d) $S_{11} = 1$, $S_{12} = \sqrt{2}$, $S_{21} = \sqrt{2}$, and $S_{22} = 1$.

Determine K and $|\Delta|$ and draw the input and output stability circles.

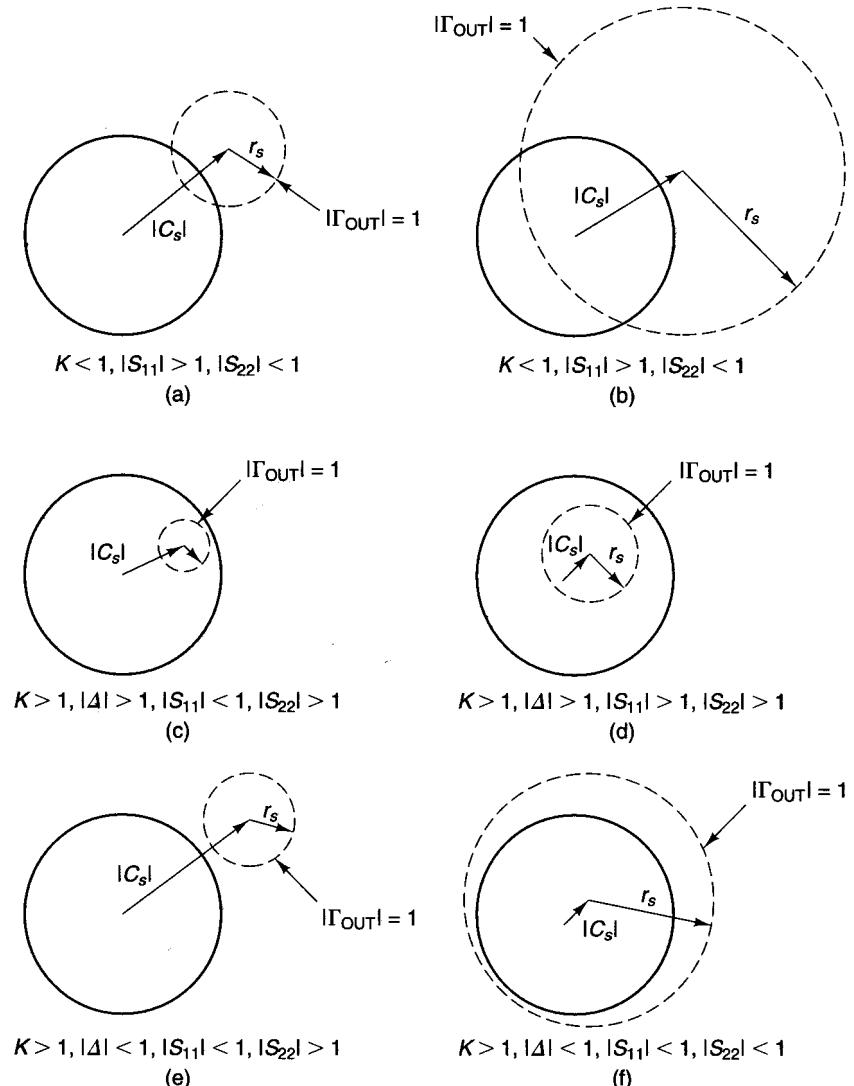


Figure P3.5

3.10 (a) Show that in the limit as S_{12} approaches zero, the center and radius of the stability circles are $C_S \approx 1/S_{22}$, $r_s \approx 0$, $C_L \approx 1/S_{11}$, and $r_L \approx 0$.

(b) The S parameters of a two-port network are

$$\begin{aligned} S_{11} &= 2 | 90^\circ & S_{12} &= 0 \\ S_{21} &= 2 & S_{22} &= 0.1 | 45^\circ \end{aligned}$$

Draw the stability circles and show the unstable regions.

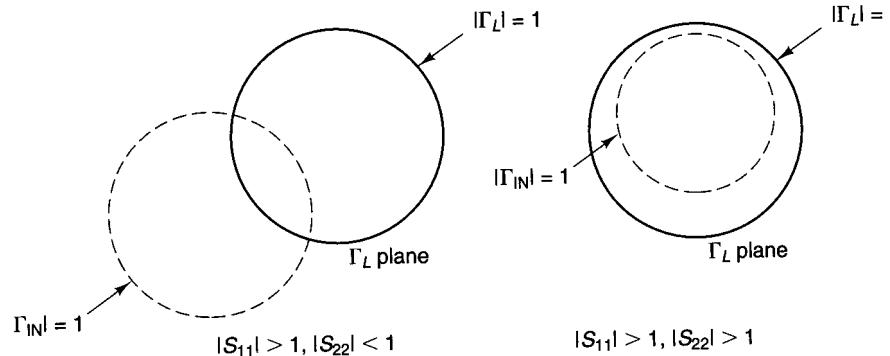


Figure P3.6

- 3.11** Show that the source stability circle does not enclose the center of the Smith chart when

$$|S_{22}| < 1 \text{ and } |\Delta| < S_{11}$$

or

$$|S_{22}| > 1 \text{ and } |\Delta| > S_{11}$$

Also show that the load stability circle does not enclose the center of Smith chart when

$$|S_{11}| < 1 \text{ and } |\Delta| < S_{22}$$

or

$$|S_{11}| > 1 \text{ and } |\Delta| > S_{22}$$

- 3.12** The conditions for unconditional stability were analyzed by considering the values in the Γ_s and Γ_L plane that result in $|\Gamma_{IN}| < 1$ and $|\Gamma_{OUT}| < 1$. An alternative approach is to consider the values in the Γ_{IN} and Γ_{OUT} plane, where $|\Gamma_{IN}| < 1$ and $|\Gamma_{OUT}| < 1$, that results in $|\Gamma_s| < 1$ and $|\Gamma_L| < 1$. Using this approach, show that the plot of the $|\Gamma_s| = 1$ circle in the Γ_{OUT} plane has radius and center given by

$$r_{OUT} = \frac{|S_{12}S_{21}|}{1 - |S_{11}|^2}$$

and

$$C_{OUT} = S_{22} + \frac{S_{12}S_{21}S_{11}^*}{1 - |S_{11}|^2} = \frac{S_{22} - \Delta S_{11}^*}{1 - |S_{11}|^2}$$

Also, the plot of the $|\Gamma_L| = 1$ circle in the Γ_{IN} plane has radius and center given by

$$r_{IN} = \frac{|S_{12}S_{21}|}{1 - |S_{22}|^2}$$

and

$$C_{IN} = S_{11} + \frac{S_{12}S_{21}S_{22}^*}{1 - |S_{22}|^2} = \frac{S_{11} - \Delta S_{22}^*}{1 - |S_{22}|^2}$$

- 3.13 (a)** Show that

$$|\Delta| \leq |S_{11}| |S_{22}| + |S_{12}S_{21}|$$

and

$$|S_{11}| |S_{22}| \leq |\Delta| + |S_{12}S_{21}|$$

Substitute these inequalities in (3.3.13) and verify that

$$(1 - |\Delta|)^2 > (|S_{11}|^2 - |S_{22}|^2)^2$$

Therefore, show that

$$B_1 B_2 > 0$$

and

$$B_1 + B_2 = 2(1 - |\Delta|^2)$$

where B_1 is given by (3.3.21) and

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2$$

- (b)** Use the previous results to show that the conditions $K > 1$ and $B_1 > 0$ are similar to $K > 1$ and $B_2 > 0$ (see Appendix C).

- (c)** Show that the condition $|\Delta| < 1$ is similar to $B_1 > 0$.

- 3.14** Show how resistive loading can stabilize a transistor whose S parameters at $f = 750$ MHz are

$$\begin{aligned} S_{11} &= 0.69 \angle -78^\circ & S_{12} &= 0.033 \angle 41.4^\circ \\ S_{21} &= 5.67 \angle 123^\circ & S_{22} &= 0.84 \angle -25^\circ \end{aligned}$$

Consider the four types of resistive loading shown in Fig. 3.3.9.

- 3.15** Prove that the maximum unilateral transducer power gain in (3.4.6) is obtained when $\Gamma_s = S_{11}^*$ and $\Gamma_L = S_{22}^*$.

- 3.16 (a)** Design a microwave transistor amplifier for $G_{TU,\max}$ using a BJT whose S parameters in a 50Ω system at $V_{CE} = 10$ V, $I_C = 20$ mA, and $f = 1$ GHz are

$$\begin{aligned} S_{11} &= 0.706 \angle -160^\circ \\ S_{12} &= 0 \\ S_{21} &= 5.01 \angle 85^\circ \\ S_{22} &= 0.508 \angle -20^\circ \end{aligned}$$

- (b)** Draw the constant-gain circles for $G_s = 2, 1, 0$, and -1 dB.

- 3.17** The scattering parameters of a GaAs FET in a 50Ω system are

$$\begin{aligned} S_{11} &= 2.3 \angle -135^\circ \\ S_{12} &= 0 \\ S_{21} &= 4 \angle 60^\circ \\ S_{22} &= 0.8 \angle -60^\circ \end{aligned}$$

- (a) Determine the unstable region in the Smith chart and construct the constant-gain circle for $G_s = 4$ dB.
 (b) Design the input matching network for $G_s = 4$ dB with the greatest degree of stability.
 (c) Draw the complete ac amplifier schematic.

3.18 A microwave amplifier is to be designed for $G_{TU,\max}$ using a transistor with

$$\begin{aligned} S_{11} &= 0.5 \angle 140^\circ & S_{12} &= 0 \\ S_{21} &= 5 \angle 45^\circ & S_{22} &= 0.6 \angle -95^\circ \end{aligned}$$

The S parameters were measured in a 50Ω system at $f = 900$ MHz, $V_{CE} = 15$ V, and $I_C = 15$ mA.

- (a) Determine $G_{TU,\max}$.
 (b) Design two different microstrip matching networks.
 (c) Draw the constant gain circle for $G_L = 1$ dB.
 (d) If the S parameters at 1 GHz are

$$\begin{aligned} S_{11} &= 0.48 \angle 137^\circ & S_{12} &= 0 \\ S_{21} &= 4.6 \angle 48^\circ & S_{22} &= 0.57 \angle -99^\circ \end{aligned}$$

calculate the gain G_T at 1 GHz for the designs in part (b).

3.19 (a) Verify the equations for a simultaneous conjugate match in (3.6.5) and (3.6.6).
 (b) Show that for small S_{12} , Γ_M and Γ_{ML} are close to S_{11}^* and S_{22}^* , respectively.

3.20 (a) Prove the identities

$$|C_1|^2 = |S_{11} - \Delta S_{22}^*|^2 = |S_{12}S_{21}|^2 + (1 - |S_{22}|^2)(|S_{11}|^2 - |\Delta|^2)$$

and

$$|C_2|^2 = |S_{22} - \Delta S_{11}^*|^2 = |S_{12}S_{21}|^2 + (1 - |S_{11}|^2)(|S_{22}|^2 - |\Delta|^2)$$

(b) Analyze the solutions to (3.6.5) and (3.6.6) when $|K| > 1$ with K negative.

3.21 Design a microwave transistor amplifier for $G_{T,\max}$ using a BJT whose S parameters in a 50Ω system at $V_{CE} = 10$ V, $I_C = 4$ mA, and $f = 750$ MHz are

$$\begin{aligned} S_{11} &= 0.277 \angle -59^\circ \\ S_{12} &= 0.078 \angle 93^\circ \\ S_{21} &= 1.92 \angle 64^\circ \\ S_{22} &= 0.848 \angle -31^\circ \end{aligned}$$

(This problem is based on a design given in Ref. [3.6].)

3.22 The output matching network shown in Fig. P3.22 was designed at 2 GHz for a simultaneous conjugate match.

- (a) Determine Γ_{ML} .
 (b) If the microstrip is alumina ($\epsilon_r = 10$) and $h = 30$ mils, determine the length of the 0.25λ line.

3.23 The matching network in Fig. P3.23 was designed for $\Gamma_{ML} = 0.718 \angle 103.9^\circ$. Determine Γ_x .

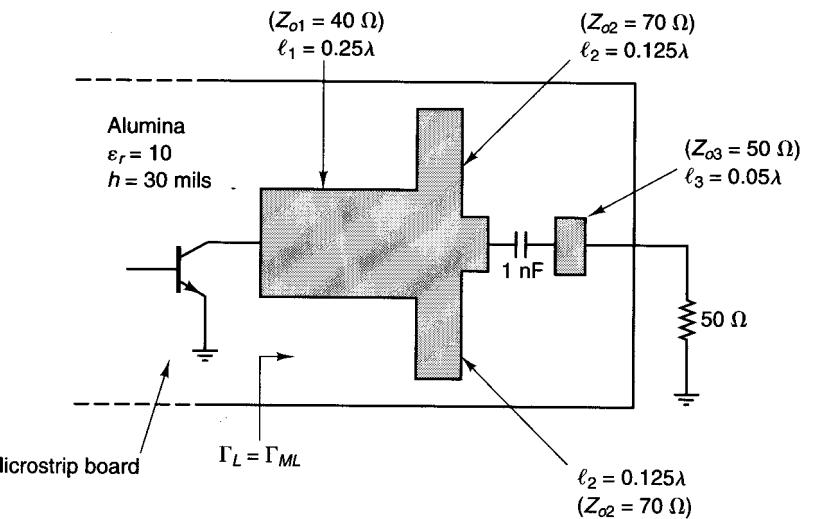


Figure P3.22

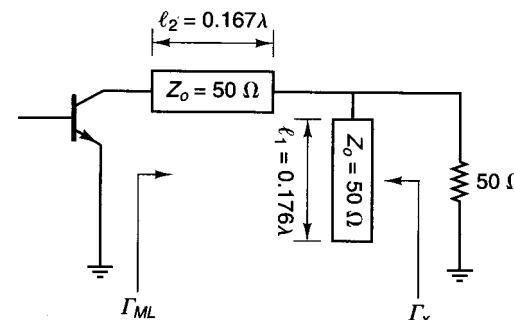


Figure P3.23

3.24 The S parameters from 3 GHz to 4 GHz of a BJT are

f (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
3	0.575 $\angle -173^\circ$	0.043 $\angle 25^\circ$	2.21 $\angle 48^\circ$	0.773 $\angle -58^\circ$
3.5	0.56 $\angle 180^\circ$	0.046 $\angle 25^\circ$	1.95 $\angle 37^\circ$	0.795 $\angle -64^\circ$
4	0.548 $\angle 173^\circ$	0.049 $\angle 24^\circ$	1.67 $\angle 29^\circ$	0.816 $\angle -71^\circ$

- (a) Can you design for a simultaneous conjugate match at 3.5 GHz?
 (b) Design an amplifier for $G_{T,\max}$ at 3.5 GHz.
 (b) Plot G_T in decibels versus frequency from 3 GHz to 4 GHz.

- 3.25** Design a microwave transistor amplifier at $f = 750$ MHz to have $G_p = 10$ dB using the BJT in Problem 3.21. For the value of Γ_L selected and with $\Gamma_s = \Gamma_{IN}^*$, determine $(VSWR)_{in}$ and $(VSWR)_{out}$.
(b) Determine the reflection coefficients for $G_{p,max}$ and show that they are identical to Γ_{Ms} and Γ_{ML} in Problem 3.21.

- 3.26** At 2 GHz, a GaAs FET has the following S parameters:

$$S_{11} = 0.7 \angle -65^\circ$$

$$S_{12} = 0.03 \angle 60^\circ$$

$$S_{21} = 3.2 \angle 110^\circ$$

$$S_{22} = 0.8 \angle -30^\circ$$

Determine the stability and design an amplifier with $G_p = 10$ dB. For the value of Γ_L selected and with $\Gamma_s = \Gamma_{IN}^*$, determine $(VSWR)_{in}$ and $(VSWR)_{out}$.

- 3.27** The S parameters of a transistor are

$$S_{11} = 0.78 \angle -102^\circ \quad S_{12} = 0.063 \angle 46^\circ$$

$$S_{21} = 2.43 \angle 84^\circ \quad S_{22} = 0.7 \angle -57^\circ$$

- (a)** Draw the $G_p = 10$ dB constant-gain circle.
(b) Select several values of Γ_L on the 10-dB gain circles. For each value of Γ_L selected, and with $\Gamma_s = \Gamma_{IN}^*$, determine $(VSWR)_{in}$ and $(VSWR)_{out}$.
(c) Draw the 15-dB, 20-dB, and 40-dB constant operating power-gain circle. Observe that as G_p becomes infinite, the power-gain circles approach the output stability circle.

- 3.28** Show that in a potentially unstable case when the operating power-gain circles and the output stability circle intersect the edge of the Smith chart, the points of intersection are identical (see Figs. 3.7.2 and 3.7.3).

- 3.29** The S parameters of a transistor are

$$S_{11} = 0.5 \angle 45^\circ \quad S_{12} = 0.4 \angle 145^\circ$$

$$S_{21} = 4 \angle 120^\circ \quad S_{22} = 0.4 \angle -40^\circ$$

- (a)** Draw the stability circles and show the stable regions.
(b) Calculate G_p if the transistor is used with $\Gamma_s = 0.2 \angle 145^\circ$ and $\Gamma_L = 0$.
(c) What is the value of the largest operating power gain that can be obtained?

- 3.30** Perform the analyses in Example 3.8.1 for the constant-gain circle $G_A = G_{A,max} - 2$ dB (i.e., $G_A = 7.66$ dB).

- 3.31** The parameters of a two-port network at 5 GHz are

$$S_{11} = 0.75 \angle -60^\circ \quad S_{12} = 0.3 \angle 70^\circ$$

$$S_{21} = 6 \angle 90^\circ \quad S_{22} = 0.5 \angle 60^\circ$$

- (a)** Verify that this two-port network is potentially unstable with $K > 1$ and $|A| > 1$.
(b) Draw the output stability circle.
(c) Calculate Γ_{ML} .
(d) Calculate $G_{p,min}$ and $G_{T,min}$.

- (e)** Draw the constant operating power-gain circles that are 2 dB, 5 dB, 10 dB, and 20 dB larger than $G_{p,min}$.
(f) Draw the input stability circle and calculate Γ_{Ms} .
(g) Determine $(VSWR)_{in}$ and $(VSWR)_{out}$ if the circuit is designed with $\Gamma_s = \Gamma_{Ms}$ and $\Gamma_L = \Gamma_{ML}$.

- 3.32** The S parameters of a GaAs FET at $f = 12$ GHz, $V_{DS} = 3.5$ V, and $I_D = 25$ mA are

$$S_{11} = 0.6 \angle 36^\circ \quad S_{12} = 0.14 \angle -85^\circ$$

$$S_{21} = 2.3 \angle -80^\circ \quad S_{22} = 0.15 \angle 45^\circ$$

- (a)** Determine $G_{p,max}$ and draw the constant G_p circle that is 1 dB less than $G_{p,max}$.
(b) Select several values of Γ_L on the $G_p = G_{p,max} - 1$ dB circle. For each Γ_L value, determine the values of Γ_s that lie on the constant $(VSWR)_{in} = 1.5$ circle, and draw the constant $(VSWR)_{in} = 1.5$ circle.
(c) Select several values of Γ_s on the $(VSWR)_{in} = 1.5$ circle. For each Γ_s value, calculate $(VSWR)_{out}$.

- 3.33 (a)** Map the $G_p = 10$ dB circle in Fig. 3.7.2 to the $\Gamma_s = \Gamma_{IN}^*$ plane.

- (b)** For $\Gamma_L = 0.1 \angle 97^\circ$, show the values of Γ_s that produce $(VSWR)_{in} = 2$.

- 3.34** Design a microwave amplifier using a BJT whose S parameters at 1 GHz are

$$S_{11} = 0.6 \angle -170^\circ \quad S_{12} = 0.03 \angle 50^\circ$$

$$S_{21} = 8 \angle 80^\circ \quad S_{22} = 0.45 \angle -30^\circ$$

Analyze the trade-offs between operating power gain, stability, and VSWRs.

- 3.35** Design a microwave amplifier using a GaAs FET whose S parameters at 6 GHz, $V_{DS} = 3$ V, and $I_{DS} = 20$ mA are

$$S_{11} = 0.6 \angle -170^\circ \quad S_{12} = 0.125 \angle -12^\circ$$

$$S_{21} = 3.1 \angle 25^\circ \quad S_{22} = 0.4 \angle -90^\circ$$

Analyze the trade-offs between available power gain, stability, and VSWRs.

- 3.36** The discrete amplifier in Fig. P3.36 was designed to operate at 500 MHz for $G_{T,max}$.

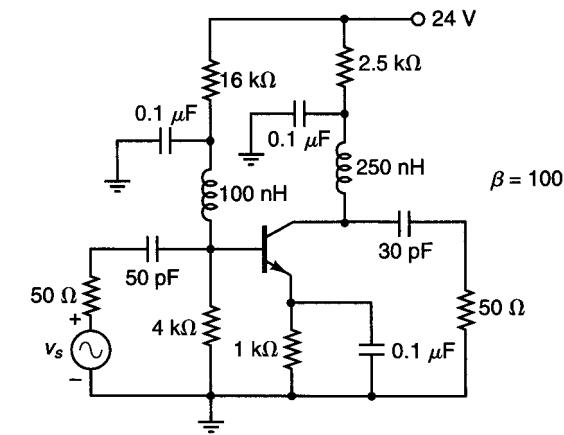


Figure P3.36

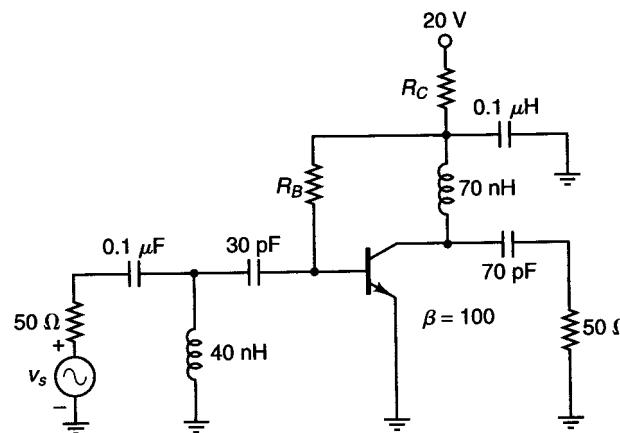


Figure P3.37

- (a) Draw the dc model.
 (b) Determine the approximate value of I_C and V_{CE} at the Q point.
 (c) Radio frequency coils could be connected in series with the $4\text{ k}\Omega$, $16\text{ k}\Omega$, and $2.5\text{ k}\Omega$ resistors. Are they necessary?
 (d) Draw the ac model.
 (e) Determine the value of Γ_s and Γ_L .
- 3.37** A 300-MHz low-noise amplifier is shown in Fig. P3.37. The dc bias circuit is recognized as the dc voltage feedback bias configuration in Fig. 3.9.2.
- (a) Draw the dc model.
 (b) Determine the values of R_L and R_B for a Q point at 10 V , 5 mA .
 (c) Determine the values of Γ_s and Γ_L .
- 3.38** Design the bias circuit in Fig. 3.9.3 to have a Q point at $V_{CE} = 10\text{ V}$ and $I_C = 10\text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.75\text{ V}$.
- 3.39** Design the active bias network in Fig. 3.9.4 to set the quiescent point of transistor Q_2 at $V_{CE2} = 10\text{ V}$ and $I_{C2} = 10\text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.75\text{ V}$.
- 3.40** (a) Derive the stability factors for the dc bias networks in Fig. 3.9.2a.
 (b) Design the circuit for a quiescent point at $V_{CE} = 10\text{ V}$ and $I_C = 10\text{ mA}$. Use $h_{FE} = 50$ and $V_{CC} = 20\text{ V}$.
 (c) What happens to the quiescent point if h_{FE} changes from 50 to 100.
- 3.41** Design the dc bias network shown in Fig. 3.9.3 for $V_{CE} = 6\text{ V}$, $I_C = 1\text{ mA}$, and $S_i = 5$. Assume that $h_{FE} = 100$ and $I_{CBO} = 1\text{ }\mu\text{A}$ at 25°C . Calculate the resulting stability factors and find what happens to the operating point if the temperature increases to 75°C .
- 3.42** Design the active bias network in Fig. 3.9.8 to set the quiescent point of GaAs FET at $V_{DS} = 3\text{ V}$ and $I_{DS} = 10\text{ mA}$. The pinchoff voltage is -3 V , and the drain saturation current is 30 mA.

- [3.1] D. Woods, "Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in Terms of S Parameters," *IEEE Transactions on Circuits and Systems*, February 1976.
- [3.2] K. Kurokawa, "Power Waves and the Scattering Matrix," *IEEE Transactions on Microwave Theory and Techniques*, March 1965.
- [3.3] G. E. Bodway, "Two Port Power Flow Analysis Using Generalized Scattering Parameters," *Microwave Journal*, May 1967.
- [3.4] "Microwave Transistor Bias Considerations," Hewlett-Packard Application Note 944-1, April 1975.
- [3.5] G. D. Vendelin, "Five Basic Bias Designs for GaAs FET Amplifiers," *Microwaves & RF*, February 1978.
- [3.6] W. H. Froehner, "Quick Amplifier Design with Scattering Parameters," *Electronics*, October 1967.

REFERENCES

- [3.1] D. Woods, "Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in Terms of S Parameters," *IEEE Transactions on Circuits and Systems*, February 1976.

MATCHING NETWORKS AND SIGNAL FLOW GRAPHS

2.1 INTRODUCTION

The analysis of transmission-line problems and of matching circuits at microwave frequencies can be cumbersome in analytical form. The Smith chart provides a very useful graphical aid to the analysis of these problems. The Smith chart is basically a plot of all passive impedances in a reflection coefficient chart of unit radius. The reading accuracy from the Smith chart is sufficient for most practical microwave transistor amplifier design problems.

Matching circuits that provide optimum performance in a microwave amplifier can be easily and quickly designed using the normalized impedance and admittance Smith chart. The Smith chart is also used to present the frequency dependence of scattering parameters and other amplifier characteristics.

The characteristics of microstrip transmission lines are presented in this chapter. The mode of propagation in a microstrip line is assumed to be quasi-transverse electromagnetic. Although radiation losses in a microstrip line can be severe, the use of a thin material, having a high dielectric constant, between the top strip conductor and the ground plane of a microstrip line reduces the radiation losses to a minimum.

Microstrip lines find extensive use as passive circuit elements and as a medium in which the complete microwave amplifier can be built. The interconnection features of the microstrip line are unsurpassed. Transistors in chip or packaged form can be easily attached to the strip conductors of the microstrip line. Some practical circuit construction techniques using microstrips are presented.

In this chapter signal flow graphs are discussed. The description of two-port networks in terms of S parameters permits the use of signal flow graphs in the analysis of microwave amplifiers. Power-gain expressions, as well as other relations, are derived in Section 2.6 using signal flow graphs. In Section 2.7 the power-gain expressions are derived using direct manipulations of the traveling wave relations.

2.2 THE SMITH CHART

The Smith chart is the representation in the reflection coefficient plane, called the Γ plane, of the relation

$$\Gamma = \frac{Z - Z_o}{Z + Z_o} \quad (2.2.1)$$

for all values of Z , such that $\text{Re}[Z] \geq 0$. Z_o is the characteristic impedance of the transmission line or a reference impedance value. Defining the normalized impedance z as

$$z = \frac{Z}{Z_o} = \frac{R + jX}{Z_o} = r + jx$$

we can write (2.2.1) in the form

$$\Gamma = \frac{z - 1}{z + 1} \quad (2.2.2)$$

Figure 2.2.1a illustrates the properties of the transformation (2.2.2) for some values of z . For example, if $Z = 50 \Omega$ and $Z_o = 50 \Omega$, then $z = 1$ and $\Gamma = 0$. That is, the point $z = 1$ in the normalized z plane maps into the origin of the Γ plane. From (2.2.2) it also follows that the point $z = 0$ maps into the point $\Gamma = -1$ (i.e., $U = -1$ and $V = 0$).

Next we consider the mapping of normalized impedances having constant real and imaginary parts. For example, for $z = 1 + jx$ the corresponding values of Γ are

$$\Gamma = \frac{jx}{2 + jx} \quad \text{or} \quad \left\{ \begin{array}{ll} \Gamma = 0 & \text{for } x = 0 \\ \Gamma = 0.447 \angle 63.43^\circ & \text{for } x = \pm 1 \\ \Gamma = 0.707 \angle 45^\circ & \text{for } x = \pm 2 \\ \text{etc.} & \end{array} \right.$$

The mapping of the various points along $z = 1 + jx$ is shown in Fig. 2.2.1a. In fact, we will show that the mapping is a circle of radius $1/2$ centered at $U = 1/2$ and $V = 0$. Since this circle represents the mapping of all points with $r = 1$, the circle is known as the constant resistance circle for $r = 1$.

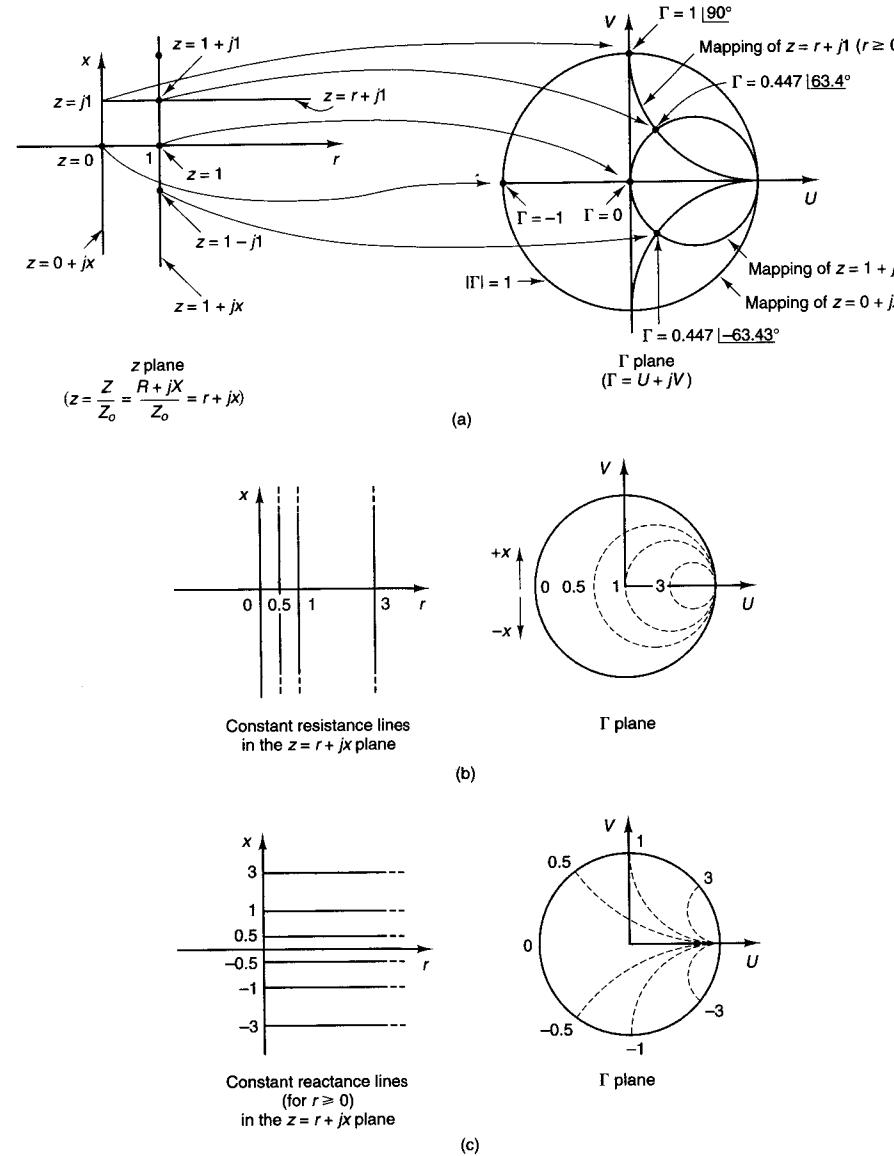


Figure 2.2.1 Development of the Smith chart.

From (2.2.2) it follows that the imaginary axis (i.e., $z = 0 + jx$) maps into the unit circle given by $|\Gamma| = 1$. Therefore, all passive impedances—that is, impedances having $r \geq 0$ —map inside the unit circle (i.e., $|\Gamma| \leq 1$) in the Γ plane.

Figure 2.2.1a also shows the mapping of $z = r + j1$ ($r \geq 0$) (i.e., a straight line having a constant imaginary value of one). For $z = r + j1$ ($r \geq 0$) it follows that the corresponding values of Γ lie along a portion of a circle inside $|\Gamma| \leq 1$

having radius 1, centered at $U = 1$ and $V = 1$. This circle is known as the constant reactance circle for $x = 1$. The portion of the circle outside the Smith chart corresponds to the mapping of $z = r + j1$ for $r < 0$ (i.e., for negative resistances).

The transformation (2.2.2) can be analyzed in general as follows. Let

$$\Gamma = U + jV = \frac{(r - 1) + jx}{(r + 1) + jx}$$

Then rationalize and separate the real and imaginary parts to obtain

$$U = \frac{r^2 - 1 + x^2}{(r + 1)^2 + x^2} \quad (2.2.3)$$

and

$$V = \frac{2x}{(r + 1)^2 + x^2} \quad (2.2.4)$$

Eliminating x from (2.2.3) and (2.2.4) results in

$$\left(U - \frac{r}{r + 1} \right)^2 + V^2 = \left(\frac{1}{r + 1} \right)^2$$

which is the equation of a family of circles centered at $U = r/(r + 1)$ and $V = 0$, with radii $1/(r + 1)$. The constant resistance circles for $r = 0, 0.5, 1$, and 3 are shown in Fig. 2.2.1b. Observe that in Fig. 2.2.1b the constant resistance circles in the Γ plane are labeled according to the constant resistance values (i.e., the value of r in the z plane).

Eliminating r from (2.2.3) and (2.2.4) results in

$$(U - 1)^2 + \left(V - \frac{1}{x} \right)^2 = \left(\frac{1}{x} \right)^2$$

which is the equation of a family of circles centered at $U = 1$ and $V = 1/x$, with radii $1/x$. The portion of constant-reactance circles for $x = -3, -1, -0.5, 0, 1, 0.5$, and 3 (with $r \geq 0$) is shown in Fig. 2.2.1c.

There is a one-to-one correspondence between points in the z plane and points in the Γ plane. The plot of the constant-resistance and constant-reactance circles for all values of z such that $\text{Re}[z] \geq 0$ in a graph is known as the *Smith chart*. The Smith chart is shown in Fig. 2.2.2. Observe that the upper half of the chart represents normalized impedances having a positive reactance (i.e., x is positive) and the lower half represents negative reactances (i.e., x is negative).

The Smith chart can also be used as an admittance chart. The appropriate transformation in this case is

$$\Gamma' = \frac{y - 1}{y + 1}$$

NAME	TITLE	DWG. NO.
SMITH CHART FORM 82-BSPR (9-66)	KAY ELECTRIC COMPANY, PINE BROOK, N.J. ©1966 PRINTED IN U.S.A.	DATE

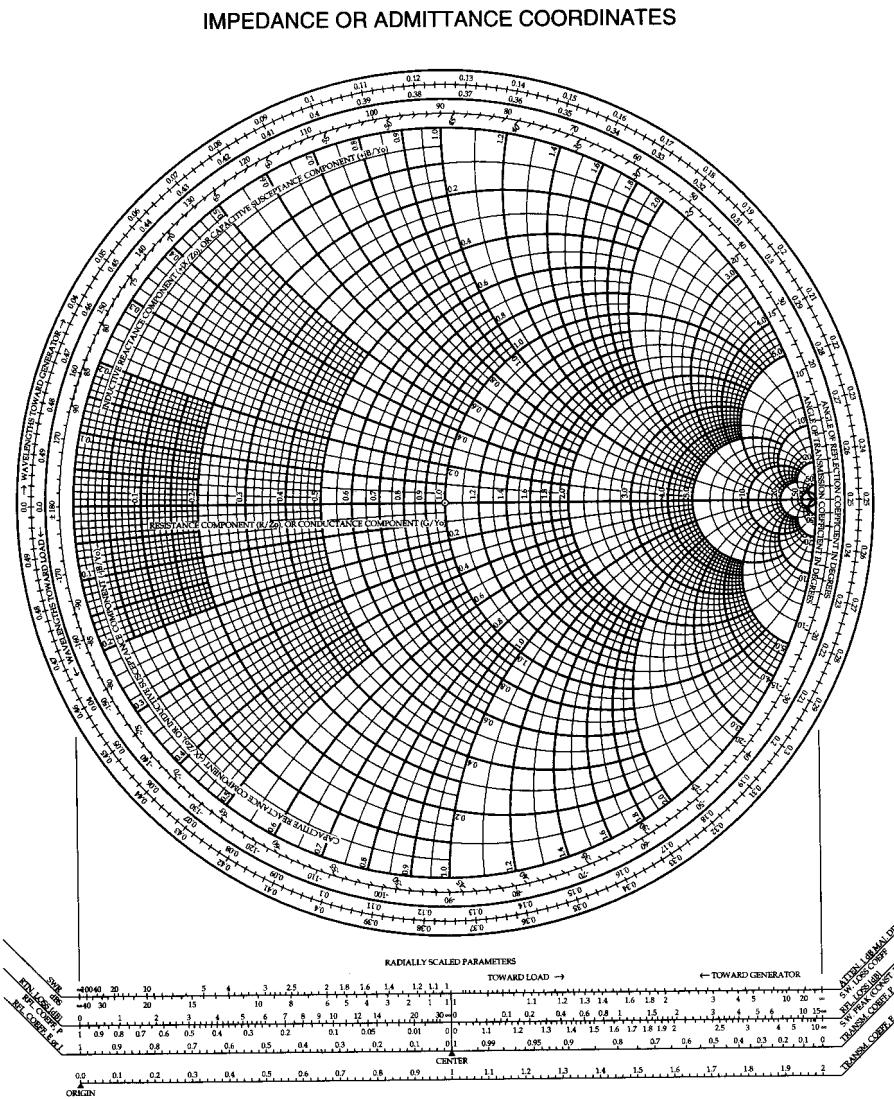


Figure 2.2.2 The Smith chart. (Reproduced with permission of Kay Electric Co., Pine Brook, N.J.)

Sec. 2.2 The Smith Chart

where the normalized admittance is $y = Y/Y_o$. Y_o is the characteristic admittance of the transmission line or a reference admittance value. Since $z = 1/y$, it follows that Γ and Γ' are related by $\Gamma' = -\Gamma$ or $\Gamma' = \Gamma^{op}$.

In the admittance chart, since

$$y = \frac{Y}{Y_e} = \frac{G + jB}{Y_e} = g + jb$$

the previous constant-resistance (r) circles become constant-conductance (g) circles and the constant-reactance (x) circles become constant-susceptance (b) circles. Observe that the upper half of the chart represents normalized admittances having a positive susceptance (i.e., b is positive) and the lower half represents negative susceptances (i.e., b is negative).

When needed for clarity, we will call a Smith chart used as an impedance chart a "Z Smith chart," and a Smith chart used as an admittance chart a "Y Smith chart."

Example 2.2.1

Locate in the Smith chart the following normalized impedances and admittances:

$$z_1 = 1 + j1, \quad z_2 = 0.4 + j0.5, \quad z_3 = 3 - j3, \quad z_4 = 0.2 - j0.6, \quad z_5 = 0$$

$$y_1 = 1 + j1, \quad y_2 = 0.4 + j0.5, \quad y_3 = 2 - j1.4, \quad y_4 = 0.5 - j0.2, \quad y_5 = \infty$$

Solution. The values of z 's and y 's are shown in Fig. 2.2.3. The Smith chart in Fig. 2.2.3a is obviously used as a Z Smith chart, and that in Fig. 2.2.3b as a Y Smith chart.

The conversion of a normalized impedance to a normalized admittance can be done easily in the Smith chart. Since from (2.2.2)

$$z = \frac{1 + \Gamma}{1 - \Gamma}$$

and

$$y = \frac{1}{z} = \frac{1 - \Pi}{1 + \Pi}$$

we observe that rotating Γ by $e^{i\pi}$ we obtain a new value of z (denoted by z') given by

$$z' = \frac{1 + \Gamma e^{j\pi}}{1 - \Gamma e^{j\pi}} = \frac{1 - \Gamma}{1 + \Gamma}$$

which is identical to the value of the admittance y . In other words, the numerical value of the impedance z' is identical to the value of the admittance $y = 1/z$.

Example 2.2.2

Find y for $z = 1 + j1$ using the Smith chart.

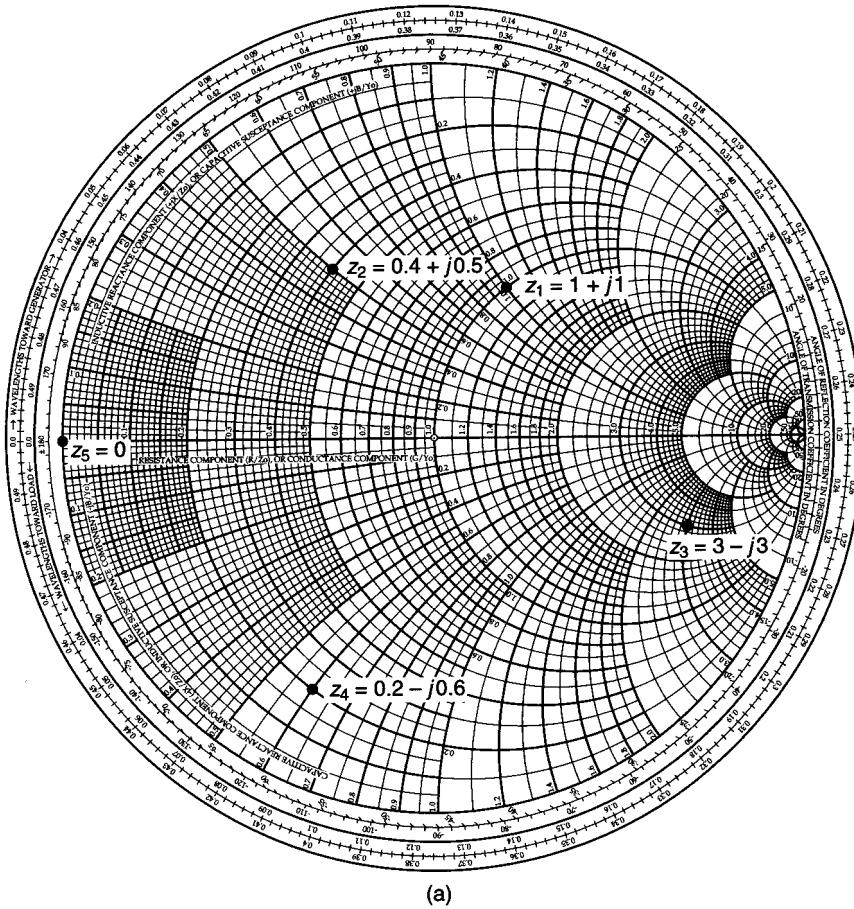


Figure 2.2.3 (a) Values of z in the Z Smith chart; (b) values of y in the Y Smith chart.

Solution. Locate the point $z = 1 + j1$ in Fig. 2.2.4. Associated with z there is a value of Γ . Rotating Γ by 180° (i.e., $e^{j\pi}$) results in the point z' , whose numerical value is that of y . The value of y is read as $0.5 - j0.5$, which of course agrees with

$$y = \frac{1}{z} = \frac{1}{1 + j1} = 0.5 - j0.5$$

Impedances having a negative real part will have a reflection coefficient whose magnitude is greater than 1. These impedances, therefore, map outside the Smith chart. Figure 2.2.5 shows a chart (known as the *compressed Smith chart*) that includes the Smith chart (i.e., $|\Gamma| \leq 1$) plus a portion of the negative impedance region.

An alternative way of handling negative resistances (i.e., $|z| > 1$) is to plot in the Smith chart $1/\Gamma^*$ and take the values of the resistance circles as being negative and the reactance circles as labeled.

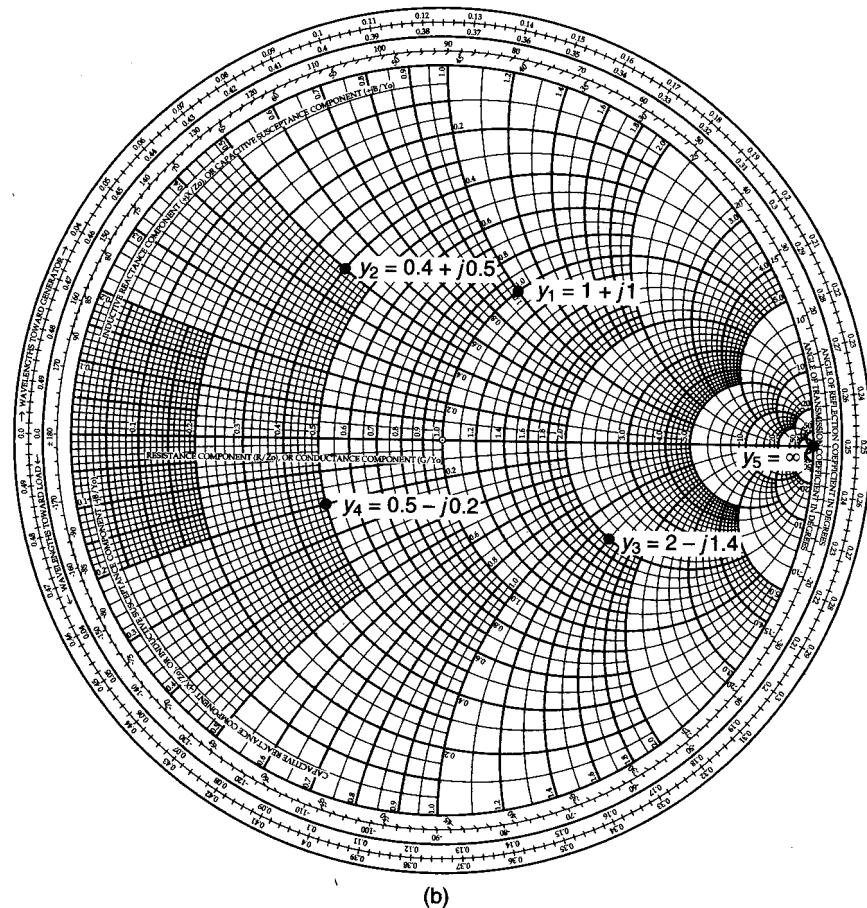


Figure 2.2.3 Continued

Example 2.2.3

Find the impedance whose reflection coefficient is $2.236|26.56^\circ$.

Solution. If we plot in the Smith chart shown in Fig. 2.2.6 the quantity

$$\frac{1}{\Gamma^*} = 0.447|26.56^\circ$$

the resulting z is $-2 + j1$. Of course, from (2.2.2),

$$\Gamma = \frac{-2 + j1 - 1}{-2 + j1 + 1} = 2.236|26.56^\circ$$

The use of the Smith chart in a transmission-line calculation follows from (1.3.35), (1.3.38), and (1.3.39). With $z = Z_L/Z_o$, we can conveniently write these equations in the form

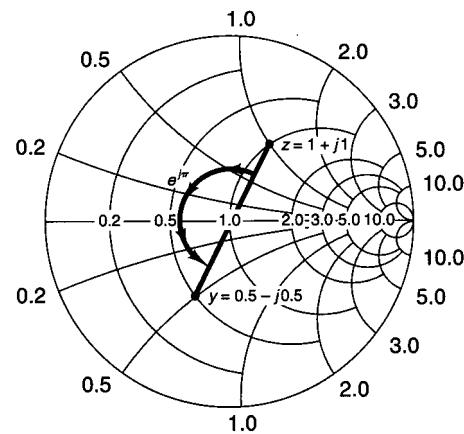


Figure 2.2.4 Conversion of z to y in the Smith chart.

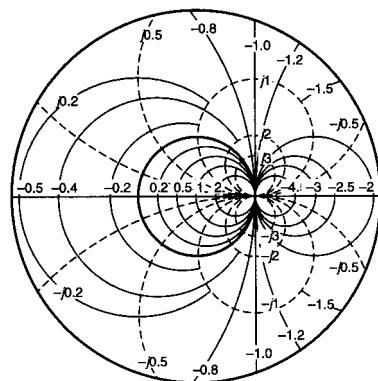


Figure 2.2.5 The compressed Smith chart. (From Ref. [1.1]; courtesy of Hewlett-Packard.)

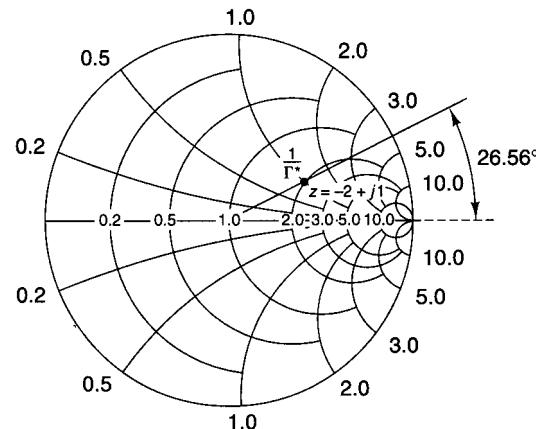


Figure 2.2.6 Negative resistances in the Smith chart.

$$\Gamma_0 = \frac{z - 1}{z + 1} \quad (2.2.5)$$

$$\Gamma_{IN}(d) = \Gamma_0 e^{-j2\beta d} \quad (2.2.6)$$

$$z_{IN}(d) = \frac{1 + \Gamma_{IN}(d)}{1 - \Gamma_{IN}(d)} \quad (2.2.7)$$

A typical transmission-line input impedance calculation involves the following steps:

1. Locate Γ_0 in the Z Smith chart for a given $z = Z_L/Z_o$ [i.e., (2.2.5)].
2. Rotate Γ_0 by $-2\beta d$ to obtain $\Gamma_{IN}(d)$ [i.e., (2.2.6)]. Observe that the rotation is along a vector of constant magnitude—namely, $|\Gamma_0| = |\Gamma_{IN}(d)|$.
3. Read the value of the normalized $z_{IN}(d)$ associated with $\Gamma_{IN}(d)$ [i.e., (2.2.7)].

Example 2.2.4

Find the input impedance, the load reflection coefficient, and the VSWR in a transmission line having an electrical length of 45° , characteristic impedance of 50Ω , and terminated in a load $Z_L = 50 + j50 \Omega$.

Solution. The transmission line is shown in Fig. 2.2.7a, where $z_L = Z_L/Z_o = 1 + j1$ and $\beta d = 2\pi d/\lambda = \pi/4$ or $d = \lambda/8 = 0.125\lambda$. In Fig. 2.2.7b, the point $z_L = 1 + j1$ is located and the vector representing Γ_0 drawn. To find Z_{IN} , we rotate along a constant $|\Gamma|$ radius a distance of -90° (i.e., $-2\beta d$)—that is, from 63.4° to -26.6° . The input impedance is read directly from the Smith chart as $z_{IN} = 2 - j1$ or $Z_{IN}(\lambda/8) = 100 - j50 \Omega$.

The previous calculations can also be made using the wavelength scales on the Smith chart. The input impedance is at a distance $d = 0.125\lambda$ from the load. From Fig. 2.2.7b, at z_L we read from the “wavelengths toward generator” scale a value of 0.162λ . This is an arbitrary value assigned to the load location. Next we add 0.125λ to obtain $0.162\lambda + 0.125\lambda = 0.287\lambda$. That is, we rotate toward the generator a distance $d = 0.125\lambda$ along a constant $|\Gamma|$ circle to reach the input of the line, which is found to be located at 0.287λ . Hence z_{IN} is read from the Smith chart to be $z = 2 - j1$ or $100 - j50 \Omega$.

The magnitude and phase of Γ_0 are read as indicated in Fig. 2.2.7b. Observe the linear scale for the magnitude of the reflection coefficient. The distance from the origin to z_L can be measured with a ruler or compass and superimposed on the linear scale. The reading of Γ_0 gives $\Gamma_0 = 0.447 |63.4^\circ|$.

Finally, the VSWR can be calculated from (1.3.44), or the distance from the origin to z_L can be measured and superimposed on the VSWR scale. The value obtained is 2.62. It can also be shown that the value of the maximum resistance in the line is numerically equal to the VSWR. This value is indicated in Fig. 2.2.7b as $VSWR = r_{max} = 2.62$.

Example 2.2.5

(a) Determine the length l of the 50Ω short-circuited transmission line shown in Fig. 2.2.8a so that the input impedance is $Z_{IN}(l) = j100\Omega$.

(b) Determine the length l of the 50Ω open-circuited transmission line shown in Fig. 2.2.8b so that the input impedance is $Z_{IN}(l) = j100\Omega$.

Solution. (a) In the short-circuited transmission line, $z_L = 0$. From Fig. 2.2.8a, the length l required to transform the load impedance $z_L = 0$ to the input impedance $z_{IN}(l) = j100/50 = j2\Omega$ is $l = 0.176\lambda$. Observe that in a short-circuited line the motion is along the edge of the chart (since $|\Gamma| = 1$ in a short-circuited line).

The length could have been calculated using (1.3.45). That is,

$$Z_{IN}(l) = j100 = j50 \tan \beta l$$

which gives $\tan \beta l = 2$ or $\beta l = 63.43^\circ = 0.352\pi$. Then

$$l = \frac{0.352\pi\lambda}{2\pi} = 0.176\lambda$$

(b) In the open-circuited transmission line, $z_L = \infty$. Therefore, from Fig. 2.2.8b the length l is 0.426λ [i.e., $(0.5\lambda - 0.25\lambda) + 0.176\lambda = 0.426\lambda$].

In many cases it is convenient to make transmission-line calculations using the Y Smith chart. The following two examples illustrate some transmission-line calculations using the Y Smith chart.

Example 2.2.6

(a) Determine the input admittance of a short-circuited transmission line having a length of $\lambda/8$ and $Y_o = 1/Z_o = 20\text{ mS}$.

(b) Determine the input admittance of an open-circuited transmission line having a length of $\lambda/8$ and $Y_o = 1/Z_o = 20\text{ mS}$.

Solution. (a) For the short-circuited line, the load admittance is $y_L = \infty$. Plotting y_L in the Y Smith chart shown in Fig. 2.2.9a and rotating along the constant gamma circle $|\Gamma| = 1$ a distance $l = \lambda/8$, we obtain $y_{IN}(l) = -j$ or

$$Y_{IN}(l) = y_{IN}(l)Y_o = -j(20 \times 10^{-3}) = -j20\text{ mS}$$

The input impedance is $Z_{IN}(l) = 1/Y_{IN}(l) = j50\Omega$.

(b) In the open-circuited line, the load admittance is $y_L = 0$. Therefore, as shown in Fig. 2.2.9b, at $l = \lambda/8$ we obtain $y_{IN}(l) = j$ or

$$Y_{IN}(l) = y_{IN}(l)Y_o = j(20 \times 10^{-3}) = j20\text{ mS}$$

The input impedance is $Z_{IN}(l) = 1/Y_{IN}(l) = -j50\Omega$.

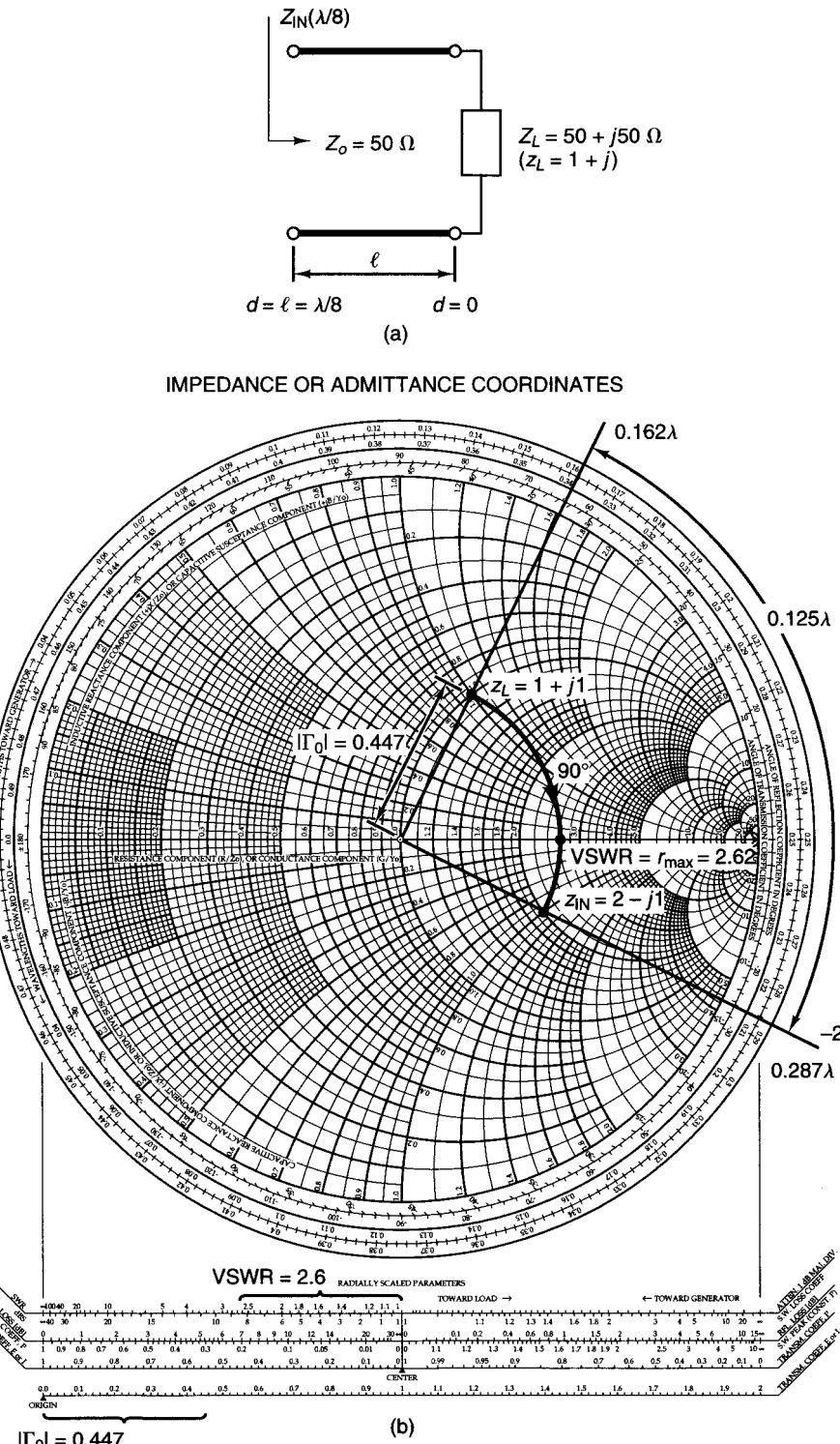


Figure 2.2.7 Typical transmission-line calculation using the Smith chart.

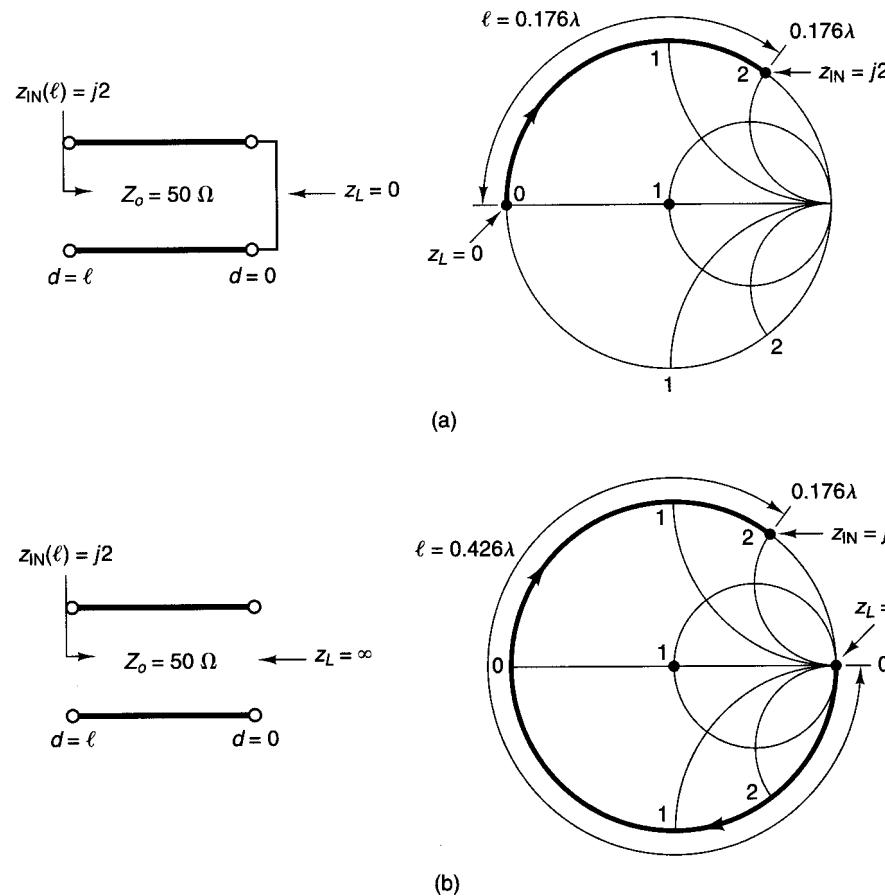


Figure 2.2.8 Circuit diagrams for Example 2.2.5.

Example 2.2.7

Solve Example 2.2.4 using a Y Smith chart.

Solution. Referring to Fig. 2.2.7a, the admittance associated with the load $z_L = Z_L/Z_o = 1 + j1$ can be obtained by rotating Γ_0 by 180° in Fig. 2.2.7b to obtain $y_L = 0.5 - j0.5$. The load y_L is shown in Fig. 2.2.10. At y_L , the “wavelengths toward generator” scale reads 0.412λ . Adding 0.125λ to 0.412λ results in the input being located at 0.537λ . Since the wavelength scale repeats every $\lambda/2$, it follows that 0.537λ is equivalent to 0.037λ . The motion from y_L to $y_{IN}(\lambda/8)$ along a constant $|\Gamma|$ circle is shown in Fig. 2.2.10. The value of $y_{IN}(l)$ is read from the Y Smith chart to be $0.4 + j0.2$, or $Y_{IN}(l) = (8 + j4) \text{ mS}$.

Comparing with Example 2.2.4, we observe that $y_{IN}(l) = 0.4 + j0.2$ is the admittance associated with $z_{IN}(l) = 1/y_{IN}(l) = 2 - j1$, and therefore $Z_{IN}(l) = 100 - j50 \Omega$, as expected.

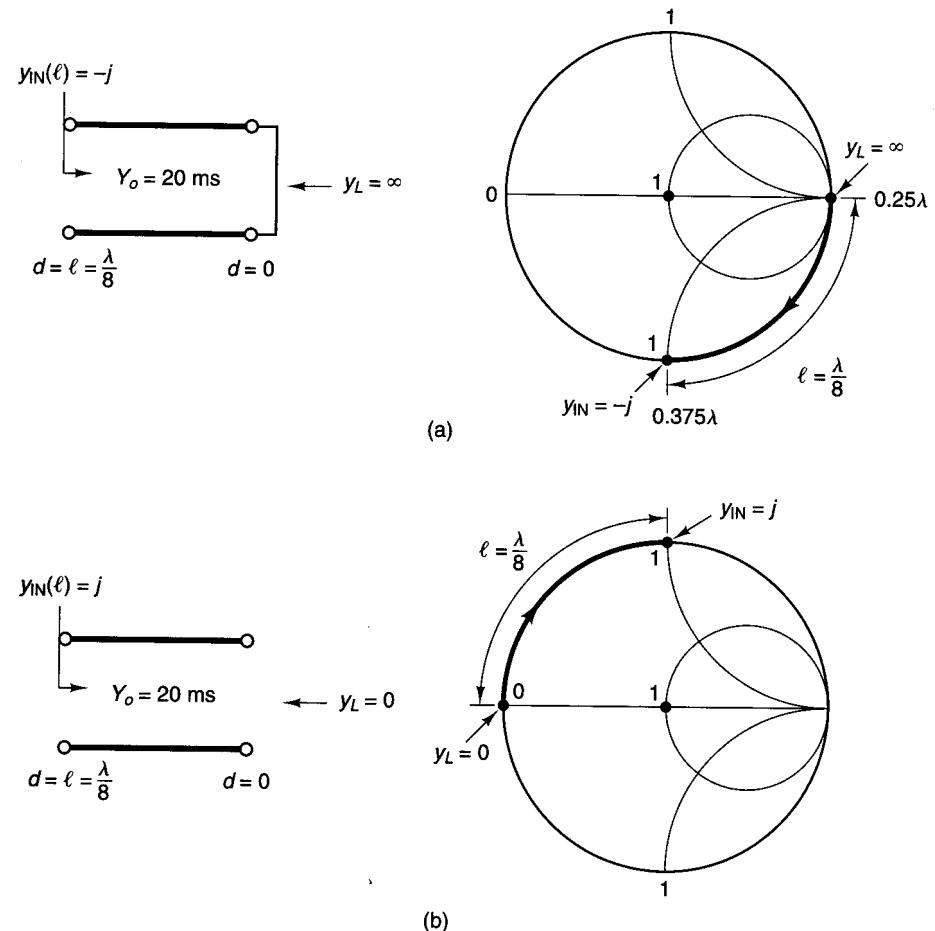


Figure 2.2.9 Circuit diagrams for Example 2.2.6.

2.3 THE NORMALIZED IMPEDANCE AND ADMITTANCE SMITH CHART

The impedance-to-admittance conversion can also be obtained by superimposing two Smith charts and rotating one of the charts by 180° . The rotated chart represents admittances and the other chart represents impedances. The superposition of the original and the rotated chart is known as the *normalized impedance and admittance coordinates Smith chart*. We will refer to this Smith chart as the *ZY Smith chart*. The *ZY Smith chart* is shown in Fig. 2.3.1, where the impedance values are shown in red and the admittance values in green. (See Fig. 2.3.1 in color on the inside cover of this book.)

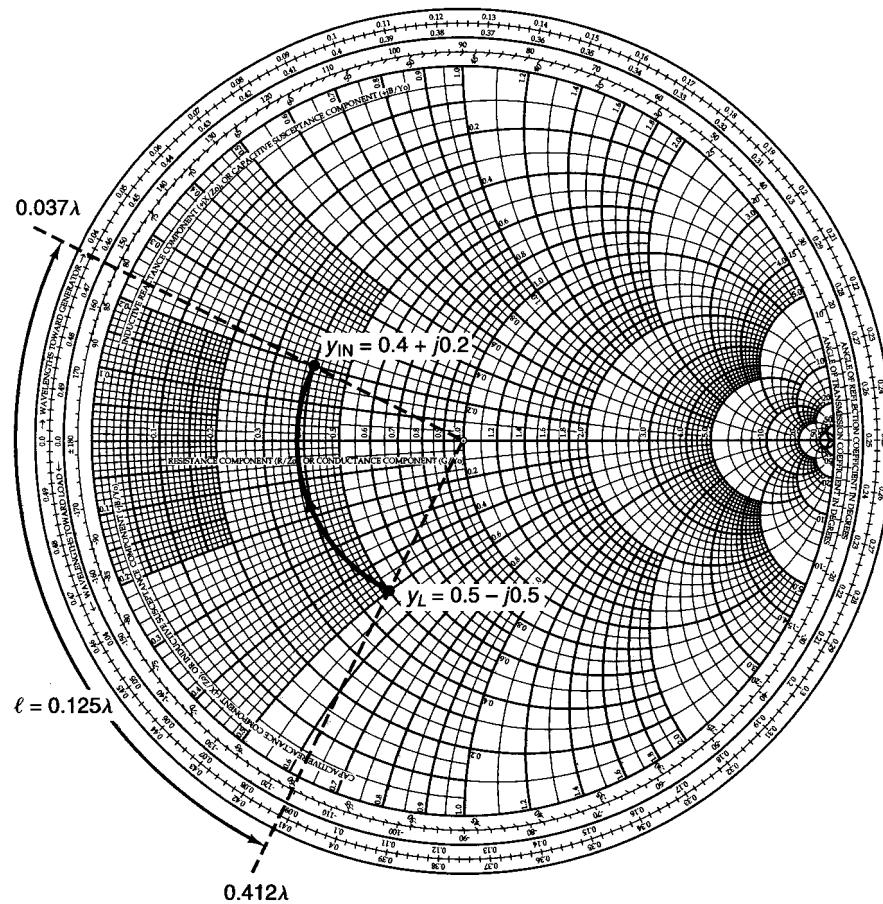


Figure 2.2.10 Solution in the Y Smith chart to the problem in Example 2.2.4.

On the left edge of the ZY Smith chart, one finds the symbols $+X_s$ and $-X_s$ to remind us that when using impedances values (red curves), the reactances are positive in the upper half of the chart and negative in the lower half. On the right edge of the ZY Smith chart, one finds $+B_p$ and $-B_p$ to remind us that when using the admittance values (i.e., green curves), the susceptances are negative in the upper half of the chart and positive in the lower half.

In the ZY Smith chart, for a given value of z the associated value of y is read directly from the admittance coordinates (shown in green), and vice versa.

Example 2.3.1

Find y for $z = 1 + j1$ using the ZY Smith chart.

NAME	TITLE	DWG. NO.
SMITH CHART FORM ZY-01-N	ANALOG INSTRUMENTS COMPANY, NEW PROVIDENCE, N.J. 07974	DATE

NORMALIZED IMPEDANCE AND ADMITTANCE COORDINATES

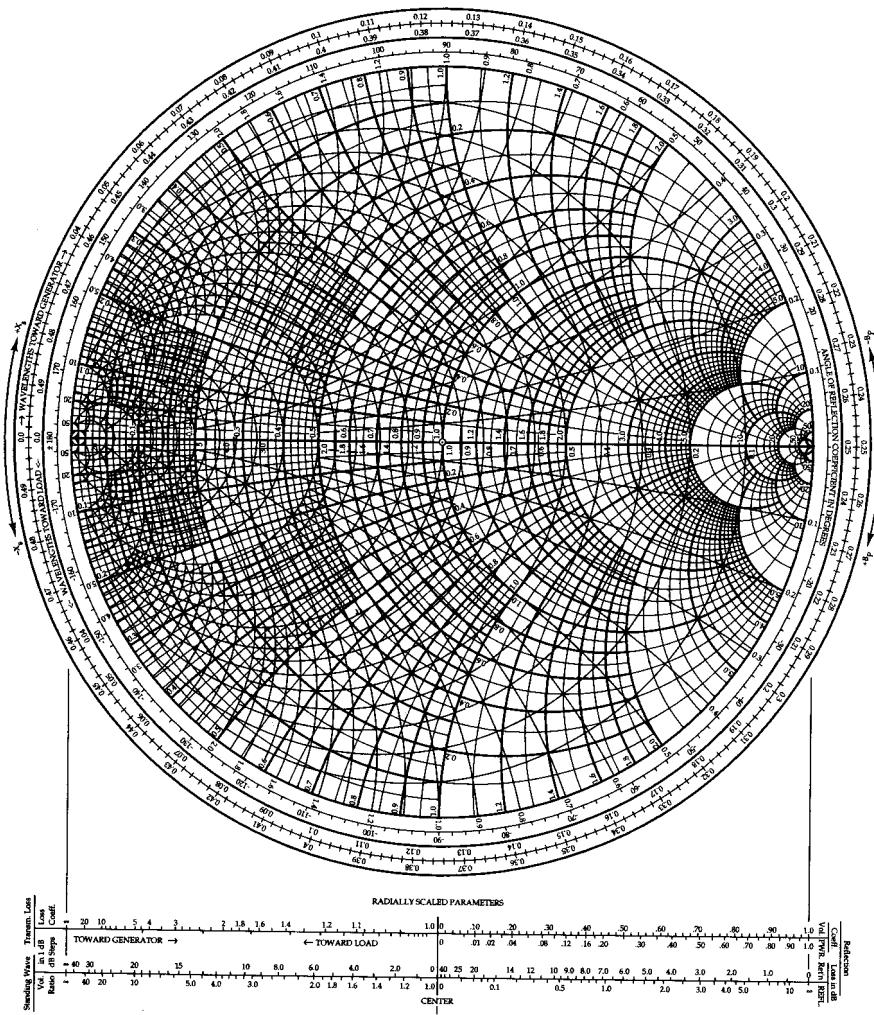


Figure 2.3.1 The normalized impedance and admittance coordinates Smith chart. (Reproduced with permission of Analog Instruments Co., New Providence, N.J.)

Solution. We can locate in the ZY Smith chart in Fig. 2.3.1 the point $z = 1 + j1$ (red curves) and read directly from the green curves the value $y = 0.5 - j0.5$.

The various Smith charts can be used to represent the frequency response of circuits, or from the frequency response of a circuit in the Smith chart an equivalent circuit model can be developed. The simplest frequency responses that can be represented in the Smith chart are those of a series RL circuit, a series RC circuit, a parallel RL circuit, and a parallel RC circuit.

In a series RL circuit, the normalized impedance is

$$z = \frac{Z}{Z_o} = \frac{R + j\omega L}{Z_o} = r + jx$$

As the frequency increases from f_a to f_b , the plot of the normalized impedance z follows a constant r circle with the reactance increasing linearly as a function of frequency, such as the typical plot for the series RL circuit shown in Fig. 2.3.2a, where $r = 0.2 \Omega$ and the reactance increases from $j0.24$ at f_a to $j0.5$ at f_b . In a series RL circuit, the reactance must change linearly with frequency.

The effect of adding a parallel capacitor to the series RL circuit is also shown in Fig. 2.3.2a. It is observed that the admittance of the capacitor (i.e., $y = jb = j\omega C/Y_o$) adds to the admittance of the series RL circuit. For the series RL circuit with a capacitor C in parallel in Fig. 2.3.2a, it is seen that at f_b the susceptance of the capacitor (i.e., $j0.9$) adds to the susceptance of the series RL circuit (i.e., $-j1.7$) to produce a series susceptance of $-j0.8$. Hence, the susceptance of the capacitor is $j0.9$ and at f_b the admittance of the circuit is $y = 0.7 - j0.8$. At f_a , the susceptance of the capacitor is such that the admittance of the circuit is $y = 2 - j1.4$.

In Fig. 2.3.2a, typical impedance plots for a series RC circuit and a series RC circuit with an inductor in parallel are also shown. In Fig. 2.3.2b, the admittance plots of parallel RL and parallel RC circuits are shown, as well as the effects of adding a series C to the parallel RL circuit and a series L to the parallel RC circuit.

Example 2.3.2

The frequency response of the normalized impedance (with $Z_o = 50 \Omega$) of a one-port network is shown in Fig. 2.3.3a as the frequency varies from 500 MHz to 1 GHz. Determine an equivalent circuit for the one-port network and the element values.

Solution. Since the frequency response follows a constant resistance circle of $r = 0.4$ and the reactance increases linearly with frequency, it follows that a series RL circuit simulates the behavior shown in Fig. 2.3.3a.

The value of R , with $Z_o = 50 \Omega$, is $R = rZ_o = 0.4(50) = 20 \Omega$. The value of L follows from

$$\frac{j\omega_b L}{Z_o} - \frac{j\omega_a L}{Z_o} = j0.4 - j0.2$$

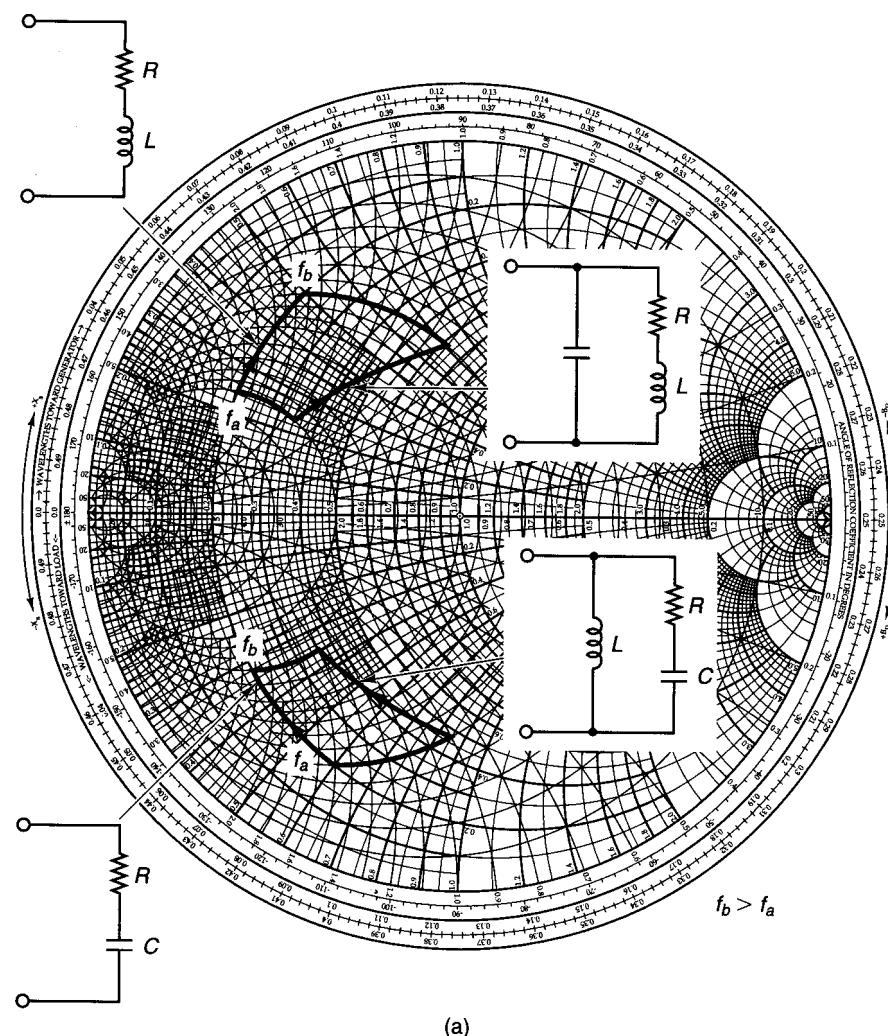


Figure 2.3.2 Characteristics of some networks in the ZY Smith chart.

or

$$L = 50 \frac{(0.4 - 0.2)}{(\omega_b - \omega_a)} = 50 \frac{0.2}{2\pi(1 \times 10^9 - 500 \times 10^6)} = 3.18 \text{ nH}$$

A typical plot of S_{11} for a transistor in the common-emitter configuration is shown in Fig. 1.10.2. It is observed that S_{11} for this transistor in chip form follows a constant-resistance circle, with a capacitive reactance at the lower frequencies and an inductive reactance at the higher frequencies. The equivalent

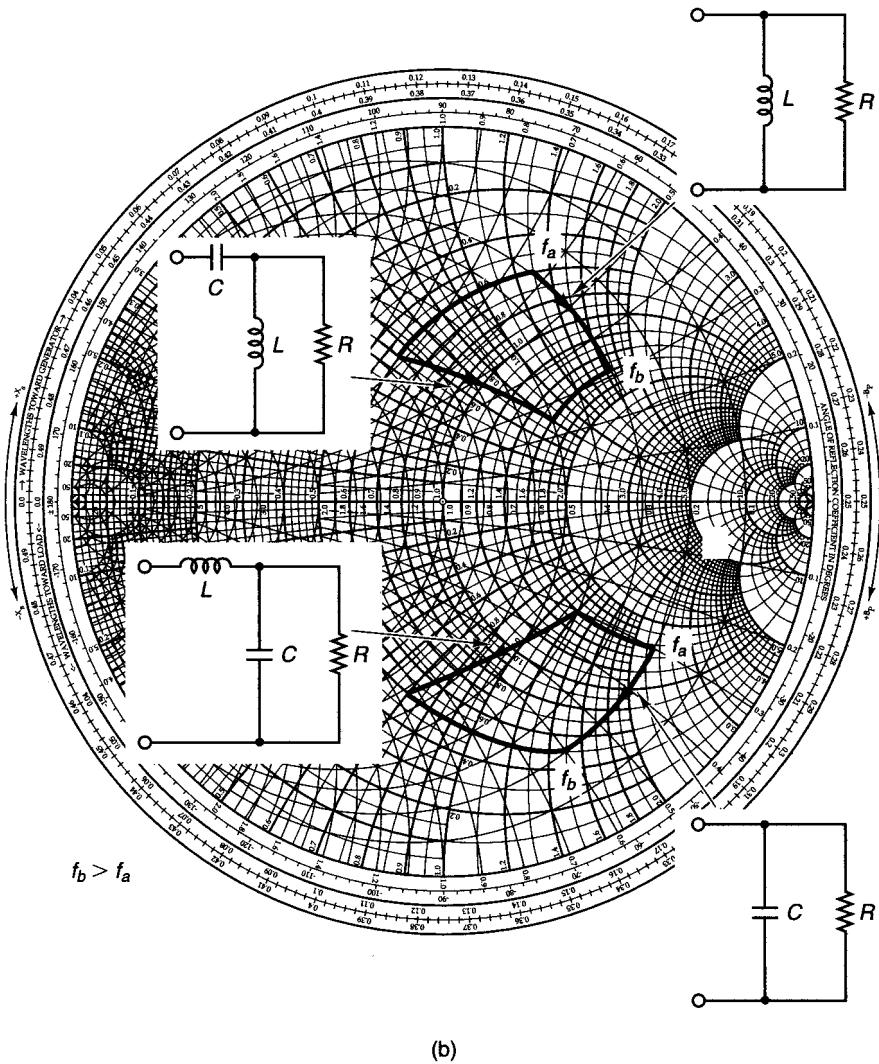


Figure 2.3.2 Continued

circuit for this transistor in chip form which exhibits the behavior of S_{11} is shown in Fig. 2.3.4a. The resistance R represents the base-to-emitter resistance plus any contact resistance. The capacitance C is due to the junction capacitance from base to emitter. The inductance L is due to the reflection properties of a transistor where the emitter resistance, when $h_{fe}(\omega)$ is complex, produces an inductive reactance across the base-to-emitter terminals.

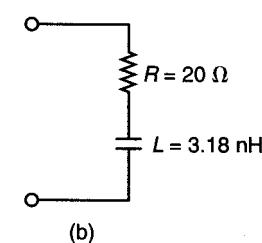
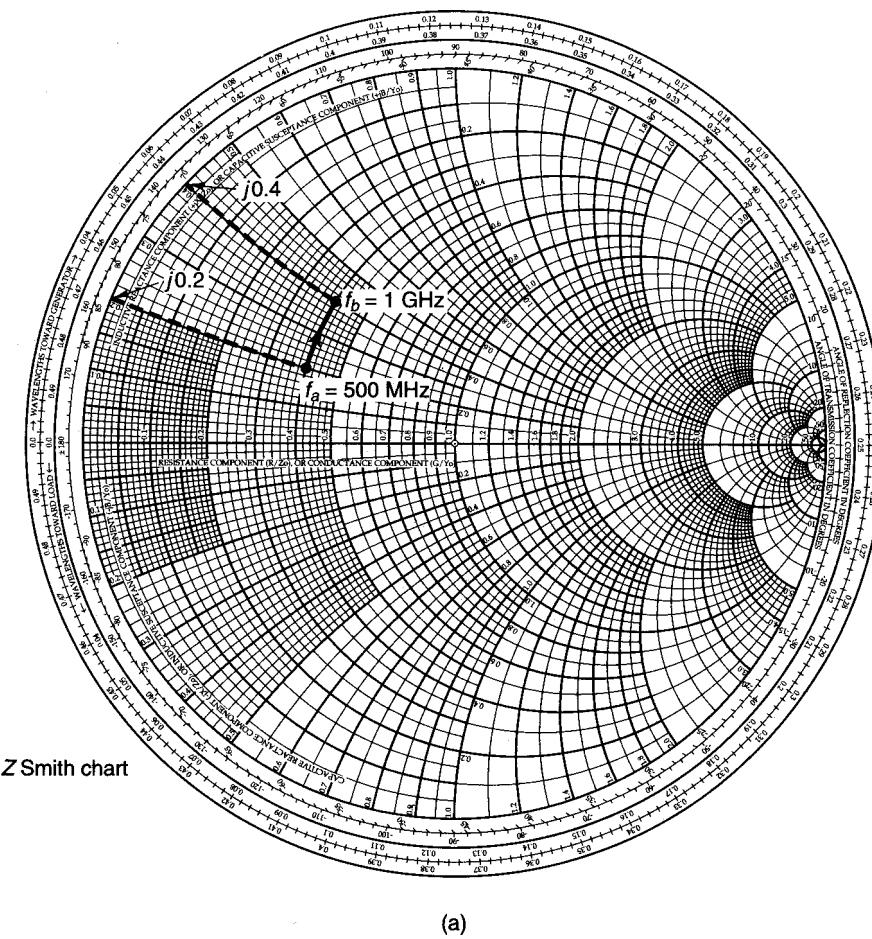


Figure 2.3.3 (a) Frequency response of a one-port network; (b) equivalent circuit of the one-port network.

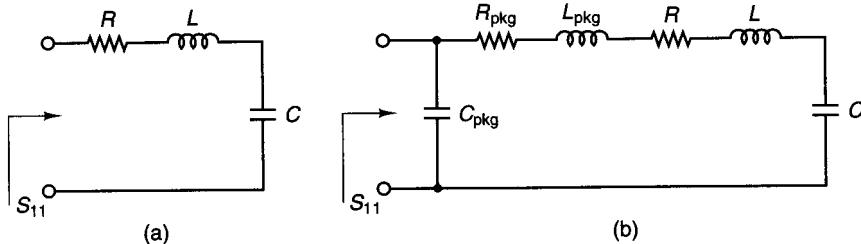


Figure 2.3.4 (a) Input equivalent circuit for a transistor in chip form; (b) input equivalent circuit for a packaged transistor.

Example 2.3.3

The frequency response of \$S_{11}\$ for a transistor chip is shown in Fig. 1.10.2, and its input equivalent network is shown in Fig. 2.3.4a. Determine the values of \$R\$, \$L\$, and \$C\$.

Solution. From Fig. 1.10.2, at \$f_a = 1\$ GHz the input impedance associated with \$S_{11}\$ is \$Z_{\text{IN}} = 50(0.2 - j0.2) = 10 - j10 \Omega\$, and at \$f_b = 10\$ GHz the input impedance is \$Z_{\text{IN}} = 50(0.2 + j0.15) = 10 + j7.5 \Omega\$. Hence, at \$\omega_a = 2\pi f_a\$ we obtain

$$10 - j10 = R + j\left(\omega_a L - \frac{1}{\omega_a C}\right) \quad (2.3.1)$$

and at \$\omega_b = 2\pi f_b\$ we obtain

$$10 + j7.5 = R + j\left(\omega_b L - \frac{1}{\omega_b C}\right) \quad (2.3.2)$$

From (2.3.1) and (2.3.2), it follows that \$R = 10 \Omega\$ and

$$-10 = 2\pi 10^9 L - \frac{1}{2\pi 10^9 C} \quad (2.3.3)$$

$$7.5 = 2\pi 10^{10} L - \frac{1}{2\pi 10^{10} C} \quad (2.3.4)$$

The simultaneous solution of (2.3.3) and (2.3.4) is \$L = 0.1024 \text{ nH}\$ and \$C = 14.95 \text{ pF}\$.

The equivalent circuit for the transistor in packaged form is a little more difficult to obtain. One equivalent circuit that will closely simulate the response of \$S_{11}\$ is shown in Fig. 2.3.4b. The resistor \$R_{\text{pkg}}\$ represents the resistance of the package, the package inductance is \$L_{\text{pkg}}\$, and the package capacitance is \$C_{\text{pkg}}\$.

2.4 IMPEDANCE MATCHING NETWORKS

The need for matching networks arises because amplifiers, in order to deliver maximum power to a load or to perform in a certain desired way, must be properly terminated at both the input and the output ports. Figure 2.4.1 illustrates a

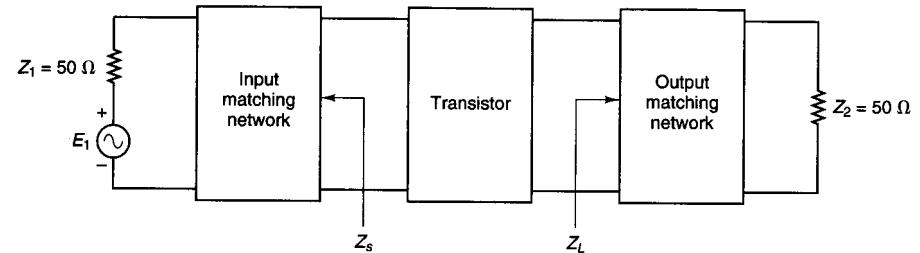


Figure 2.4.1 Block diagram of a microwave amplifier.

typical situation in which a transistor, in order to deliver maximum power to the \$50-\Omega\$ load, must have the terminations \$Z_s\$ and \$Z_L\$. The input matching network is designed to transform the generator impedance (shown as \$50 \Omega\$) to the source impedance \$Z_s\$, and the output matching network transforms the \$50-\Omega\$ termination to the load impedance \$Z_L\$.

Although many different types of matching networks can be designed, the eight Ell sections (also denoted as \$L\$ sections) shown in Fig. 2.4.2 are not only simple to design but quite practical. The matching networks are lossless in order not to dissipate any of the signal power.

The ZY Smith chart can be used conveniently in the design of matching networks. The effect of adding a series reactance element to an impedance or a parallel susceptance element to an admittance, in the ZY Smith chart, is illustrated in the following example.

Example 2.4.1

(a) Illustrate the effect of adding a series inductor \$L\$ (\$z_L = j0.8\$) to an impedance \$z\$ (\$z = 0.3 - j0.3\$) in the ZY Smith chart.

Solution. Figure 2.4.3 shows that the effect of adding a series inductance with \$z_L = j0.8\$ is to move along a constant-resistance circle from a reactance value of \$-0.3\$ to a reactance of \$0.5\$. In other words, the motion is in a clockwise direction along a constant-resistance circle.

(b) Illustrate the effect of adding a series capacitor \$C\$ (\$z_C = -j0.8\$) to an impedance \$z\$ (\$z = 0.3 - j0.3\$) in the ZY Smith chart.

Solution. Figure 2.4.4 shows that the effect of adding a series capacitor with \$z_C = -j0.8\$ is to move along a constant-resistance circle from a reactance value of \$-0.3\$ to a reactance of \$-1.1\$. In other words, the motion is in a counterclockwise direction along a constant-resistance circle.

(c) Illustrate the effect of adding a shunt inductor \$L\$ (\$y_L = -j2.4\$) to an admittance \$y\$ (\$y = 1.6 + j1.6\$) in the ZY Smith chart.

Solution. Figure 2.4.5 shows that the effect of adding a shunt inductor with \$y_L = -j2.4\$ is to move along a constant-conductance circle from a susceptance of \$1.6\$ to a

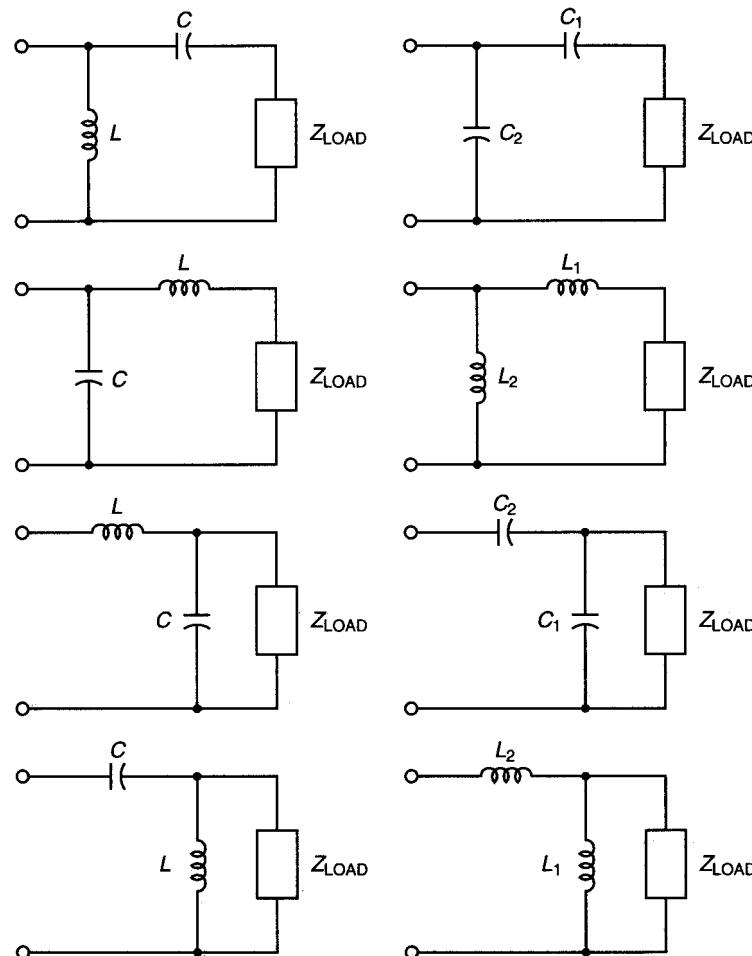


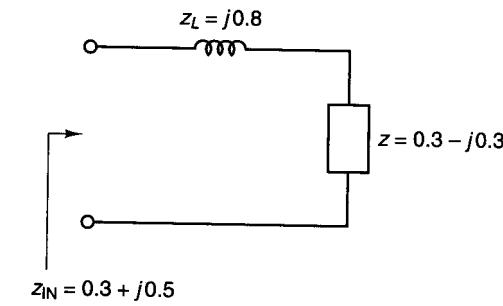
Figure 2.4.2 Matching networks.

susceptance of -0.8 . In other words, the motion is in a counterclockwise direction along a constant-conductance circle.

(d) Illustrate the effect of adding a shunt capacitor C ($y_C = j3.4$) to an admittance y ($y = 1.6 + j1.6$) in the ZY Smith chart.

Solution. Figure 2.4.6 shows that the effect of adding a shunt capacitor with $y_C = j3.4$ is to move along a constant-conductance circle from a susceptance of 1.6 to a susceptance of 5 . In other words, the motion is in a clockwise direction along a constant-conductance circle.

In conclusion, adding a series reactance produces a motion along a constant-resistance circle in the ZY Smith chart, and adding shunt susceptance



(a)

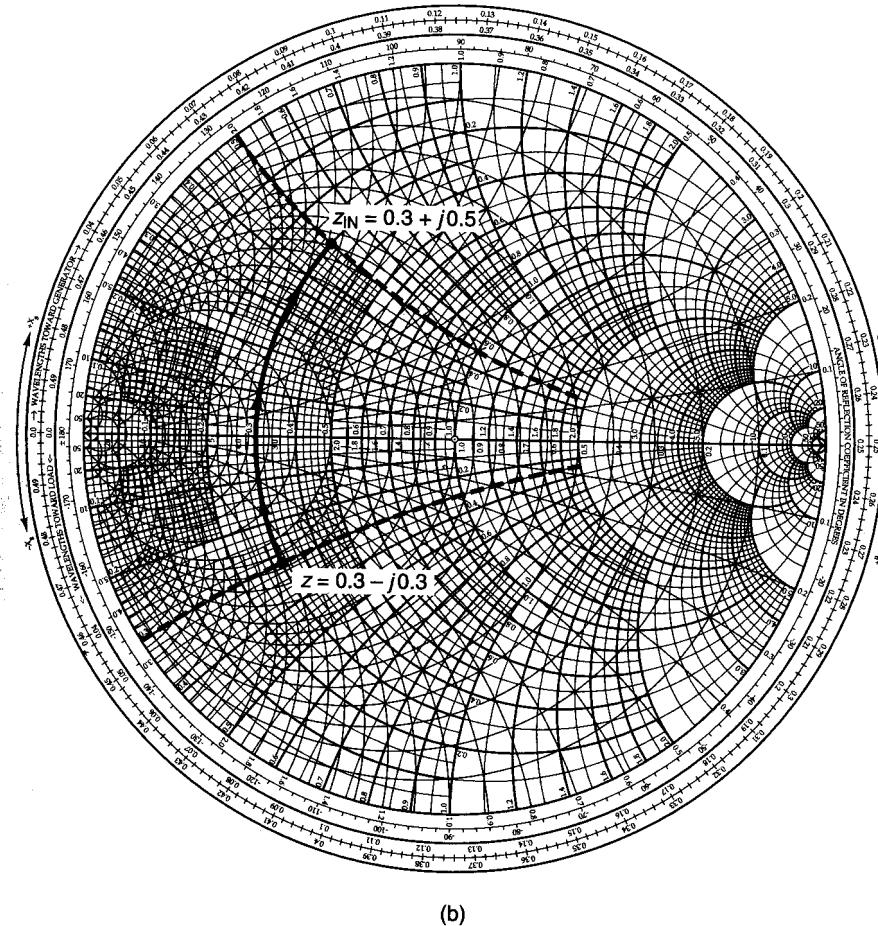
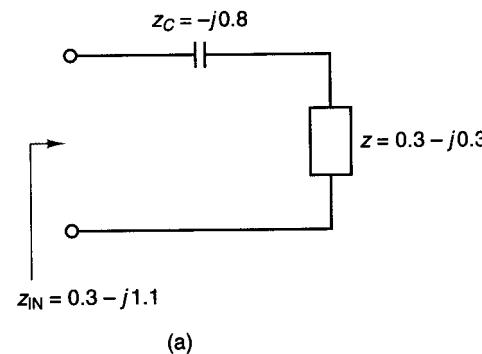
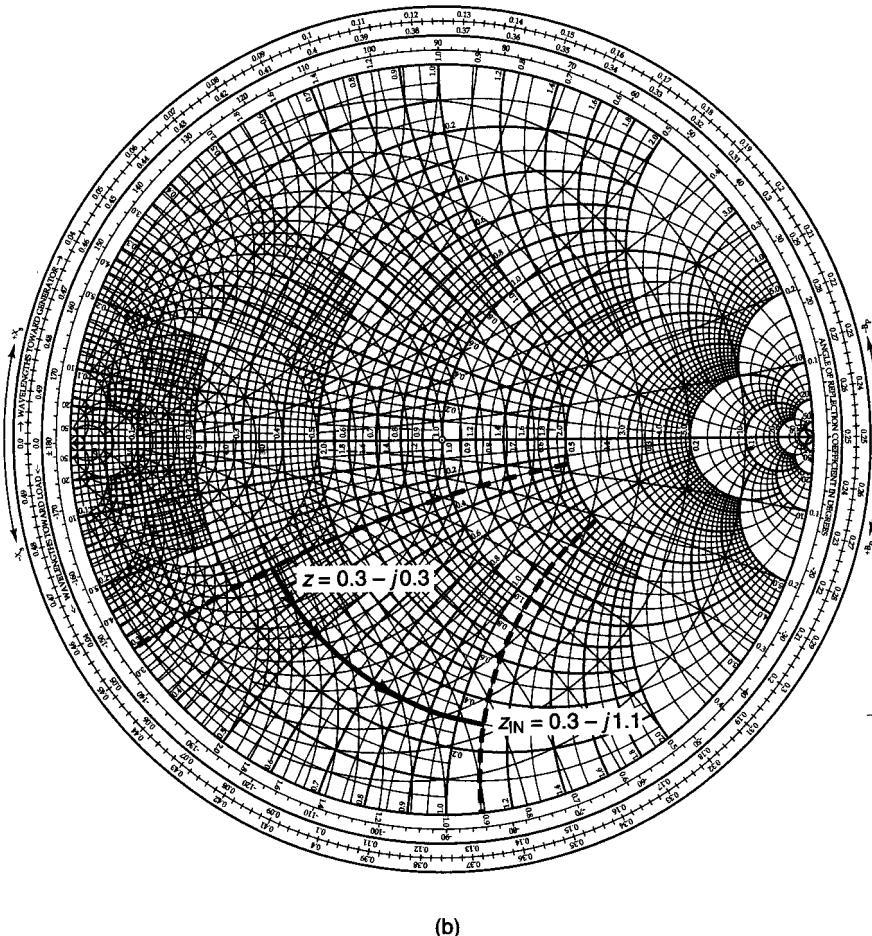


Figure 2.4.3 Effect of adding a series inductor to an impedance in the ZY Smith chart.

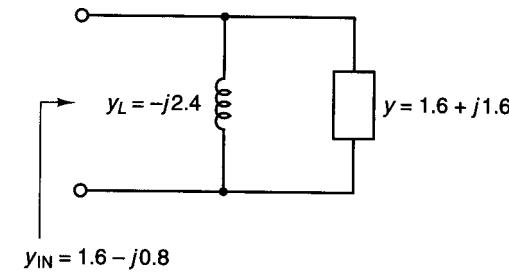


(a)

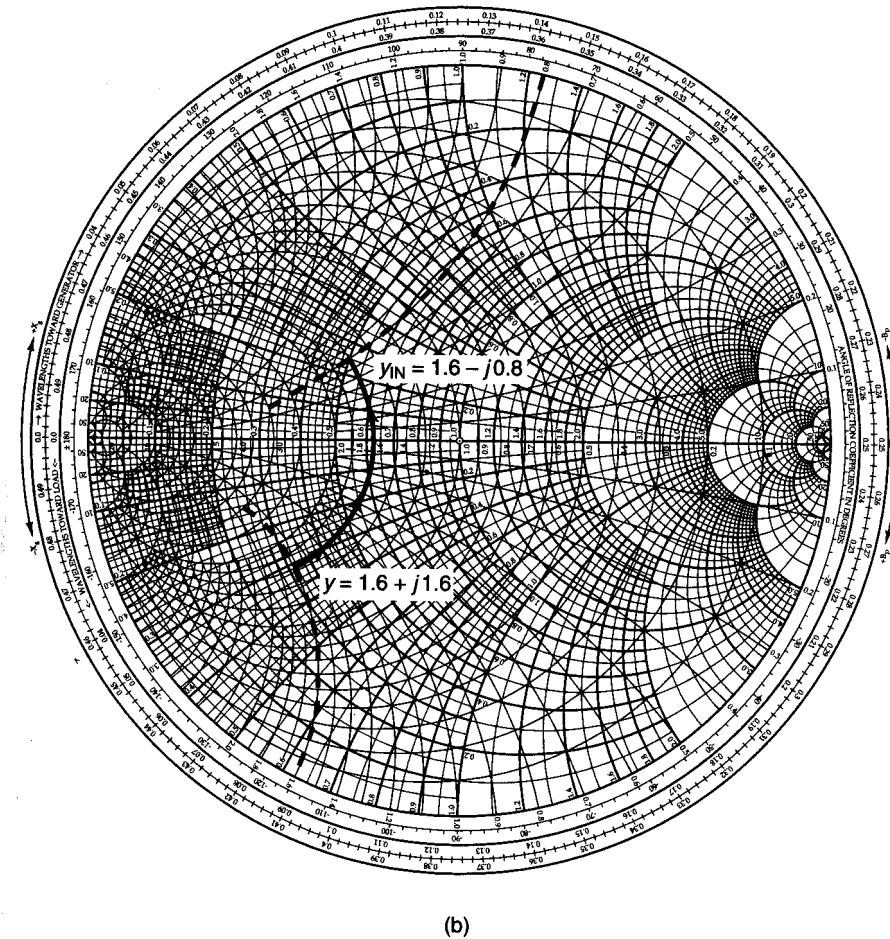


(b)

Figure 2.4.4 Effect of adding a series capacitor to an impedance in the ZY Smith chart.



(a)



(b)

Figure 2.4.5 Effect of adding a shunt inductor to an admittance in the ZY Smith chart.

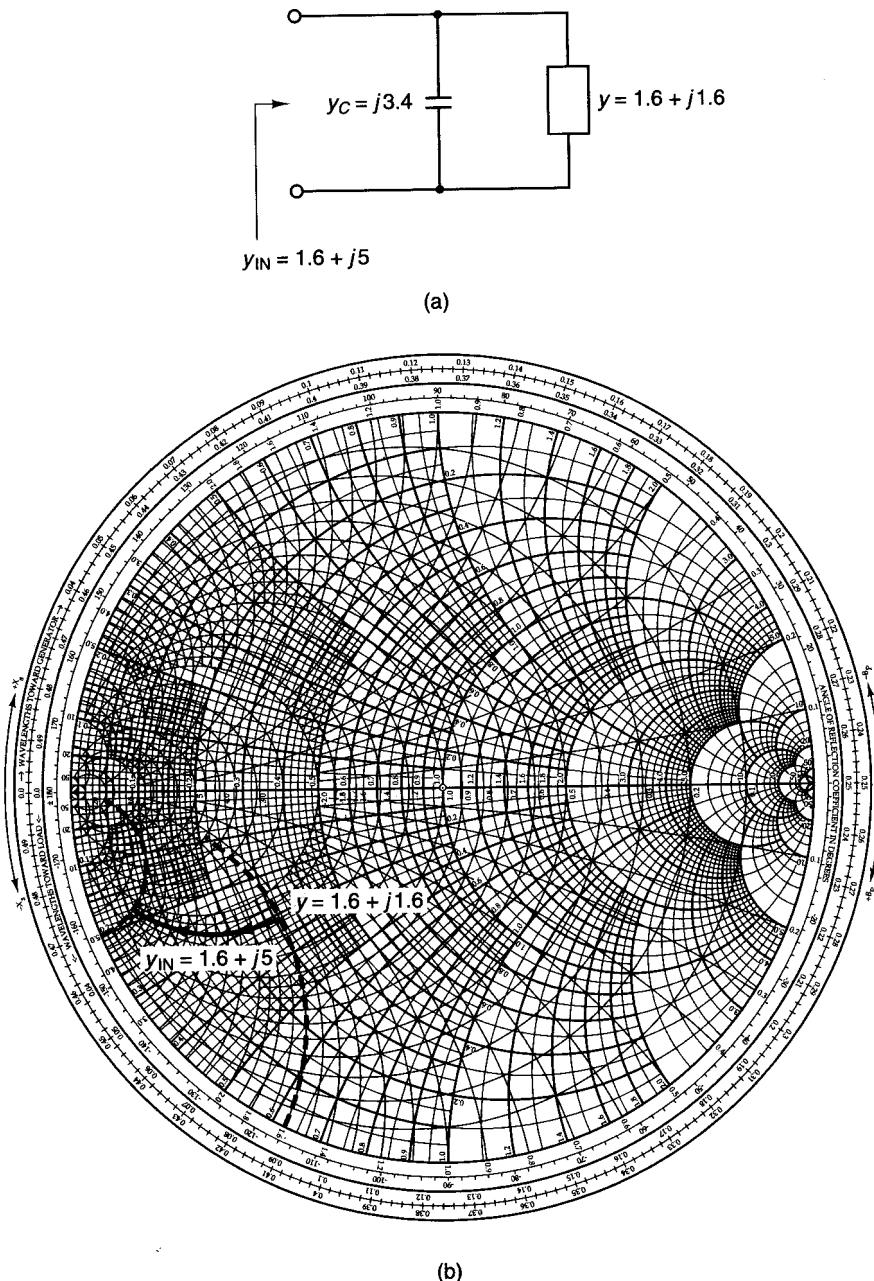


Figure 2.4.6 Effect of adding a shunt capacitor to an admittance in the ZY Smith chart.

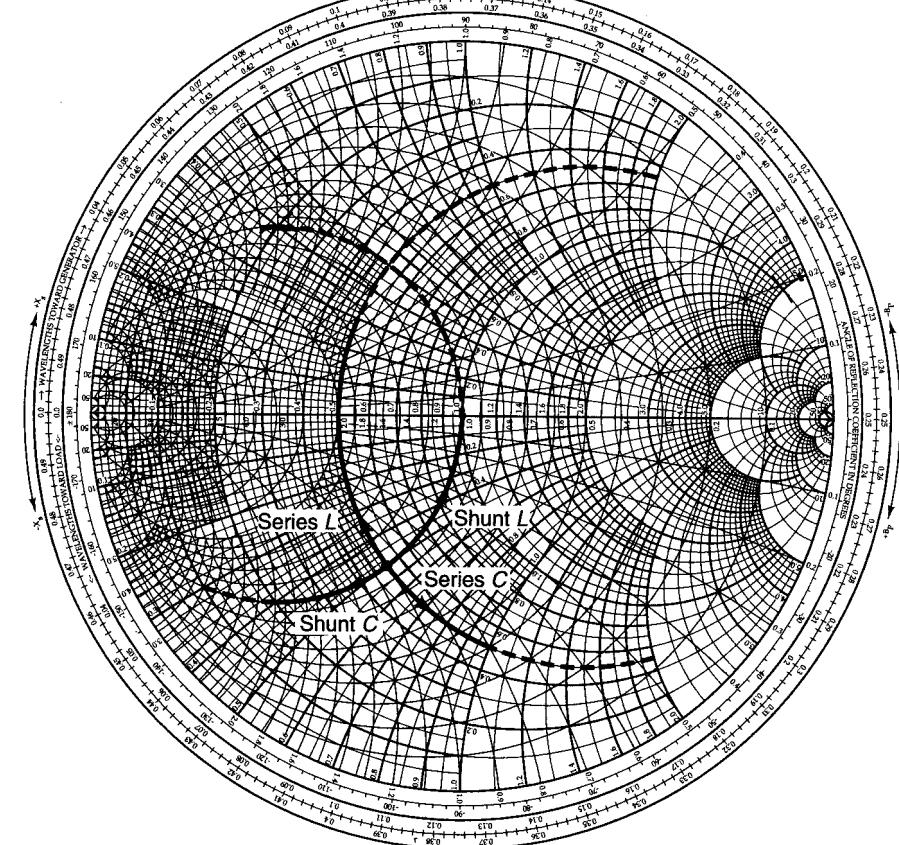


Figure 2.4.7 Effect of adding series and shunt elements in the ZY Smith chart.

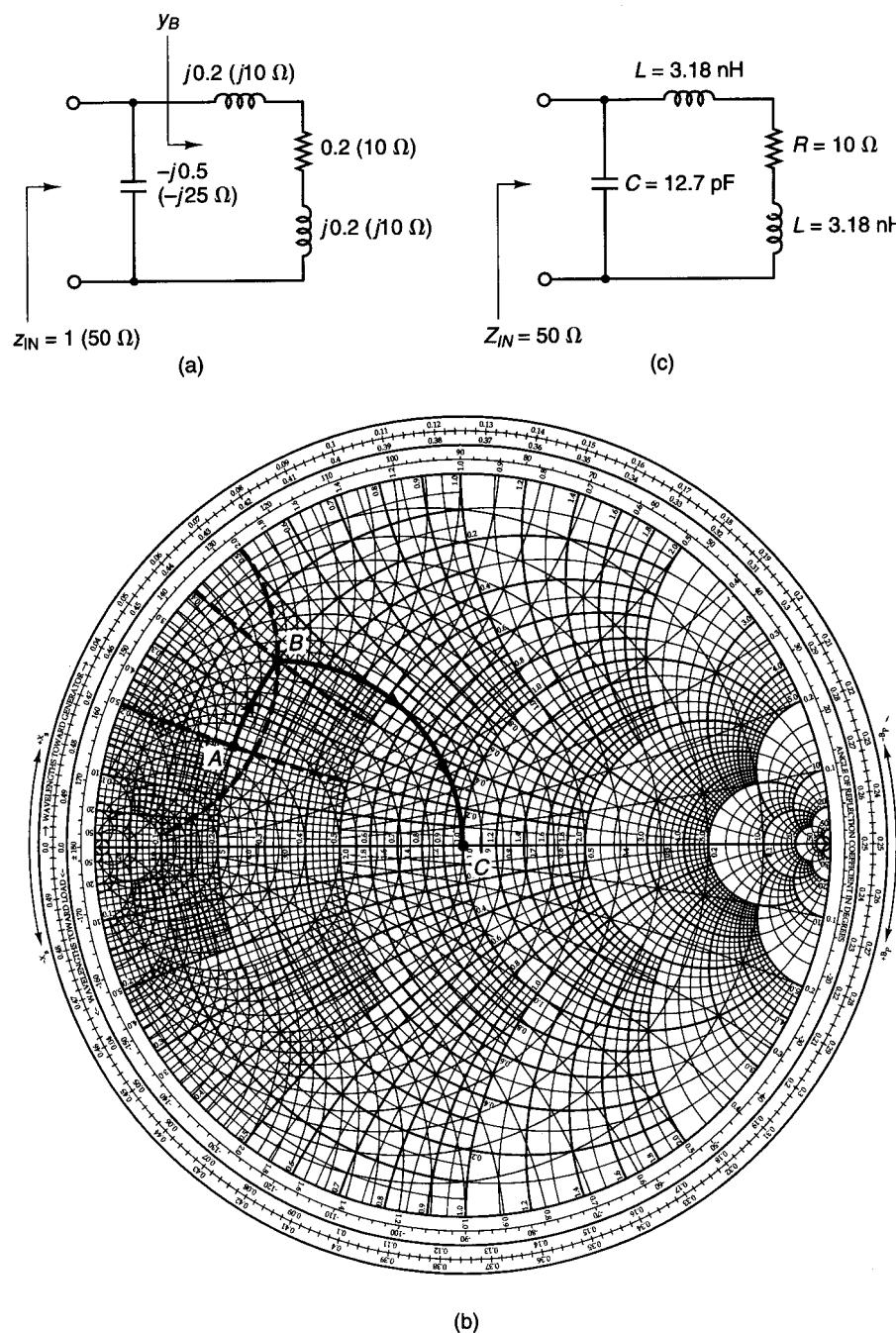
produces a motion along a constant-conductance circle in the ZY Smith chart. The four types of motions are illustrated in Fig. 2.4.7.

Designing a matching network in the ZY Smith chart consists of moving along a constant-resistance or constant-conductance circle from one value of impedance or admittance to another. Each motion along a constant-resistance or constant-conductance circle gives the value of an appropriate element. The following examples illustrate the use of the ZY Smith chart in the design of matching networks.

Example 2.4.2

A load $Z_{LOAD} = 10 + j10 \Omega$ is to be matched to a $50\text{-}\Omega$ line. Design two matching networks and specify the values of L and C at a frequency of 500 MHz.

Solution. Selecting the series L -shunt C network shown in Fig. 2.4.8a, the matching network is designed as shown in Fig. 2.4.8b. (See Figure 2.4.8b on the inside cover.)

Figure 2.4.8 Design of a series L -shunt C matching network.

The motion from point A [i.e., $z_{\text{LOAD}} = (10 + j10)/50 = 0.2 + j0.2$] to point B is along a constant-resistance circle, and we obtain for the inductor impedance $z_L = j0.4 - j0.2 = j0.2$. Observe that point B is along the unit constant-conductance circle. The admittance at point B is $y_B = 1 - j2$. The motion from point B to point C (i.e., the origin) is along a constant-conductance circle, and we obtain the capacitor admittance $y_C = 0 - (-j2) = j2$ (or $z_C = 1/j2 = -j0.5$). Therefore, at point C , $y_{\text{IN}} = z_{\text{IN}} = 1$ (or $Z_{\text{IN}} = 50 \Omega$) and the network is matched to a 50Ω line. At 500 MHz, the value of L is

$$L = \frac{10}{2\pi(500 \times 10^6)} = 3.18 \text{ nH}$$

and the value of C is

$$C = \frac{1}{25(2\pi)500 \times 10^6} = 12.74 \text{ pF}$$

The matching network at 500 MHz is shown in Fig. 2.4.8c.

The second matching network is shown in Fig. 2.4.9a and the ZY Smith chart design in Fig. 2.4.9b. (See Figure 2.4.9b on the inside cover.) The motion from A to B in Fig. 2.4.9b is along a constant-resistance circle; therefore, the impedance of the series capacitor is $z_C = -j0.4 - j0.2 = -j0.6$. The motion from B to C is along a constant-conductance circle; therefore, the admittance of the shunt inductor is $y_L = 0 - j2 = -j2$ (or $z_L = 1/-j2 = j0.5$). The design at 500 MHz is shown in Fig. 2.4.9c.

Example 2.4.3

Design the matching network shown in Fig. 2.4.10a to transform a 50Ω load at the input to an admittance $Y_{\text{OUT}} = (8 - j12) \times 10^{-3} \text{ S}$ at the output.

Solution. Figure 2.4.10b illustrates a motion in the ZY Smith chart from the origin (i.e., $z_{\text{LOAD}} = 50/50 = 1$) to $y_{\text{OUT}} = 50(8 - j12) \times 10^{-3} = 0.4 - j0.6$. (See Figure 2.4.10b on the inside cover.) The motion from A to B produces a series capacitor having an impedance of $z_C = -j1.21$. The motion from B to C produces a shunt inductor having an admittance of $y_L = -j0.6 - j0.49 = -j1.09$ (or $z_L = 1/-j1.09 = j0.917$). The matching network is shown in Fig. 2.4.10c.

Example 2.4.4

Design a matching network to transform the load $Z_{\text{LOAD}} = 100 + j100 \Omega$ to an input impedance of $Z_{\text{IN}} = 50 + j20 \Omega$.

Solution. Using a normalized value of $Z_o = 100 \Omega$, we have $z_{\text{LOAD}} = Z_{\text{LOAD}}/Z_o = 1 + j$ and $z_{\text{IN}} = Z_{\text{IN}}/Z_o = 0.5 + j0.2$. Figure 2.4.11a shows one possible solution for the matching network. The motion from A to B is produced by a shunt capacitor whose normalized admittance is $y_C = j0.86 - (-j0.5) = j1.36$ (or $z_C = 1/y_C = -j0.735$). Then, the impedance of the capacitor is

$$Z_C = 100z_C = 100(-j0.735) = -j73.5 \Omega$$

The motion from B to C requires an inductor having a normalized impedance value of $z_L = j0.2 - (-j0.87) = j1.07$. Then

$$Z_L = 100z_L = 100(j1.07) = j107 \Omega$$

The matching network is shown in Fig. 2.4.11b.

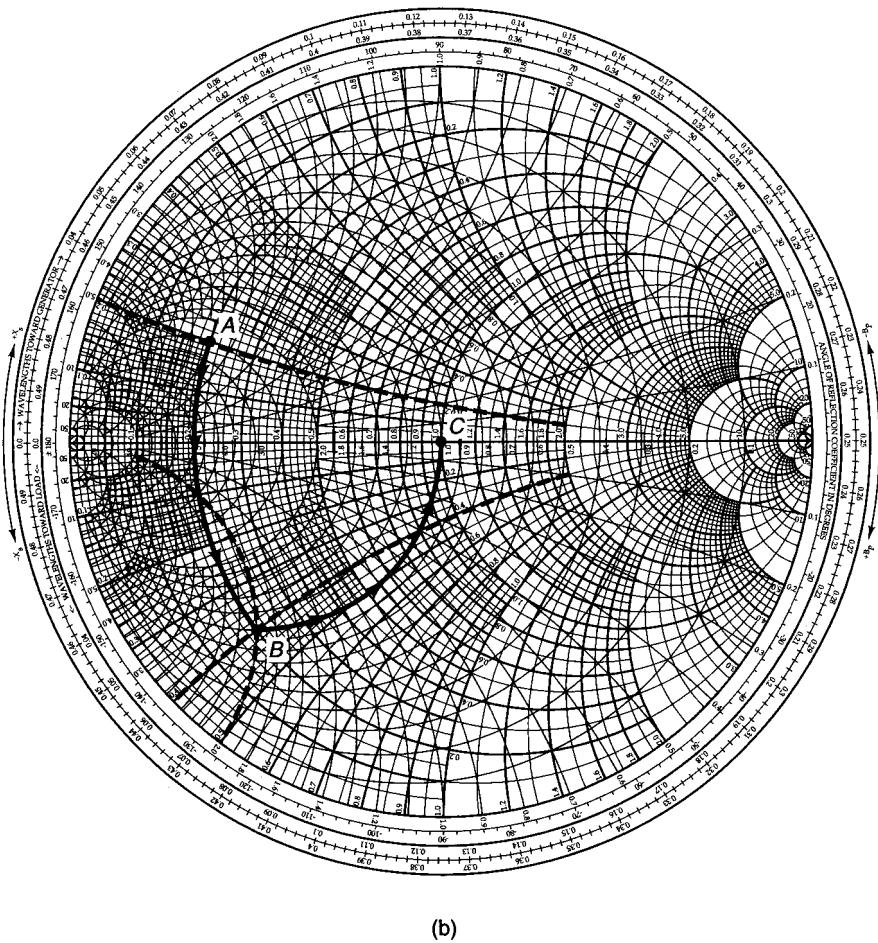
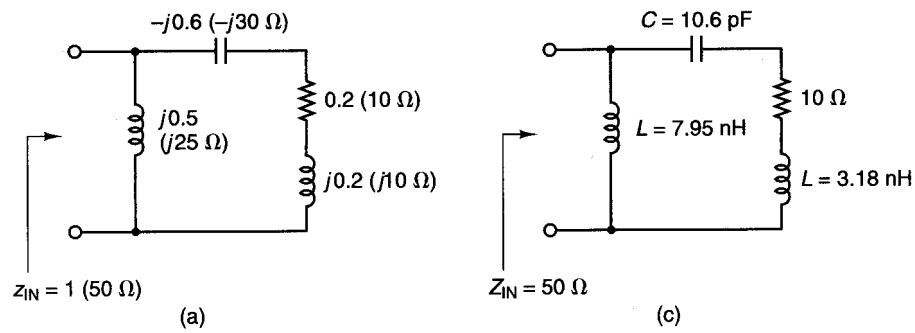


Figure 2.4.9 Design of a series C-shunt L matching network.

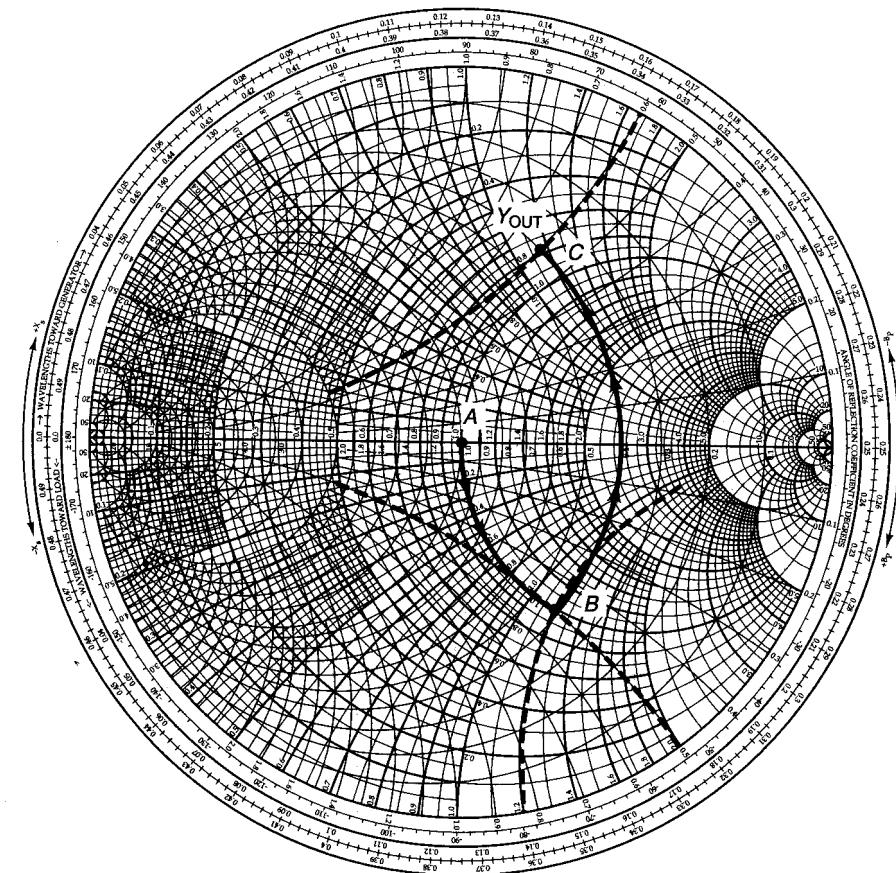
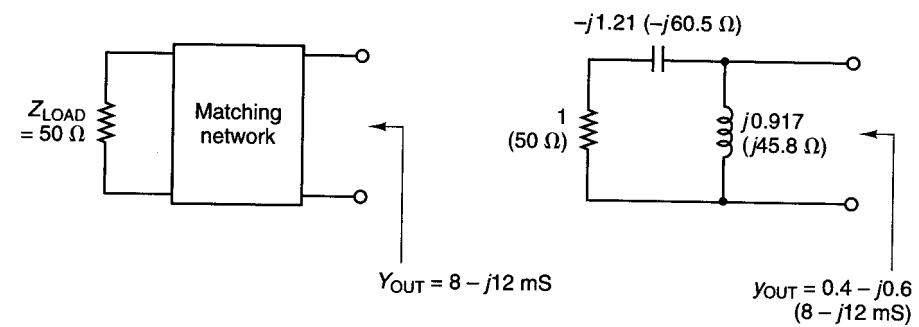


Figure 2.4.10 Matching a 50Ω load to a given Y_{OUT} using a series C-shunt L matching network.

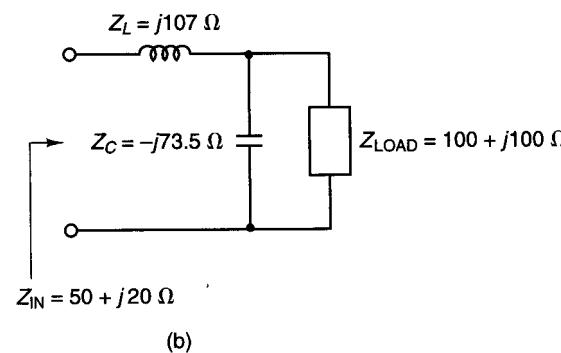
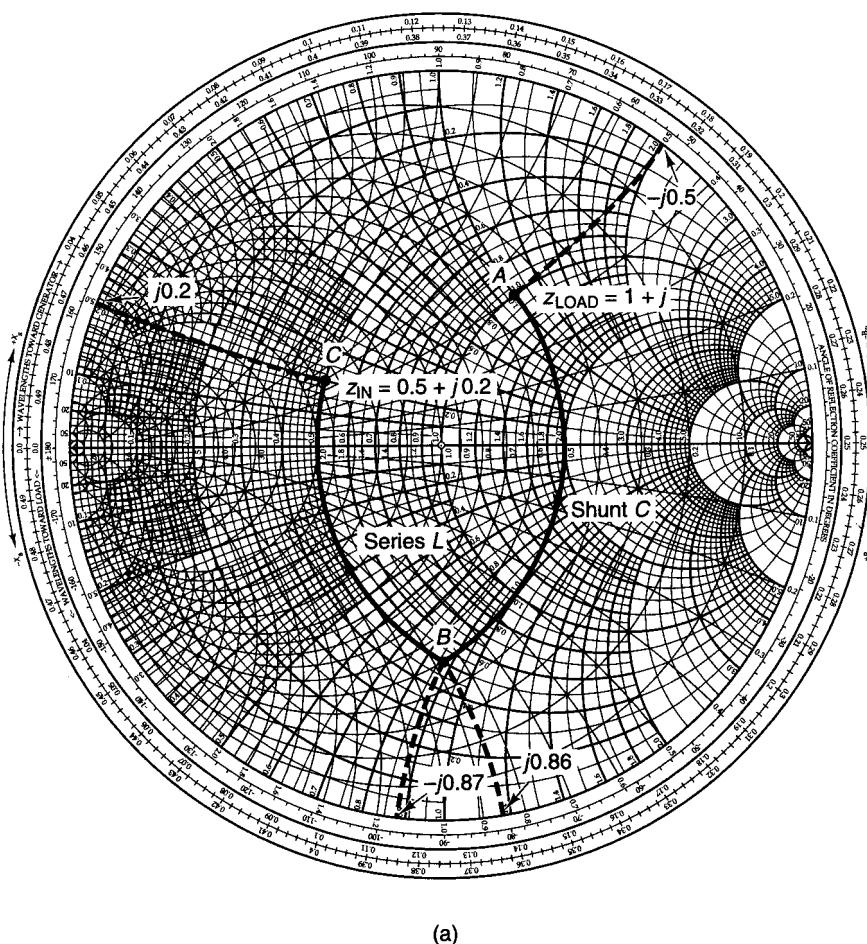


Figure 2.4.11 Matching a complex load to a given complex impedance Z_{IN} using a shunt C-series L network.

Sometimes a specific matching network cannot be used to accomplish a given match. For example, any load impedance falling in the marked region in Fig. 2.4.12a cannot be matched to 50Ω with the network in Fig. 2.4.12b because adding a series L produces motion, in a clockwise direction, away from any constant-conductance circle that passes through the origin.

Another observation regarding the Ell matching networks is that only those with an inductor and a capacitor can be used to provide a match between a resistive load and an input resistance.

In a resonant circuit, the ratio of its resonant frequency f_o to its bandwidth (BW) is known as the loaded Q of the circuit. That is,

$$Q_L = \frac{\omega_o}{BW}$$

If the bandwidth is expressed in hertz, we write

$$Q_L = \frac{f_o}{BW} \quad (2.4.1)$$

The Ell matching networks in Fig. 2.4.2 are used to provide a match at a certain frequency. The frequency response of an Ell network can be classified as either a (two-pole) low-pass filter or a high-pass filter. At each node of the Ell matching networks, there is an equivalent series input impedance, denoted by $R_s + jX_s$. Hence, a circuit node Q , denoted by Q_n , can be defined at each node as

$$Q_n = \frac{|X_s|}{R_s} \quad (2.4.2)$$

If the equivalent parallel input admittance at the node is $G_p + jB_p$, the circuit node Q can be expressed in the form

$$Q_n = \frac{|B_p|}{G_p} \quad (2.4.3)$$

For example, the normalized series input impedance of the circuit in Fig. 2.4.8a at point B (see Fig. 2.4.8b) is $0.2 + j0.4$; therefore, the circuit node Q , from (2.4.2), is $Q_n = 0.4/0.2 = 2$. Also, using the Smith chart, the equivalent parallel admittance is readily found to be $1 - j2$, producing a circuit node Q , according to (2.4.3), of $Q_n = 2/1 = 2$, which agrees with the result from (2.4.2).

The circuit in Fig. 2.4.8a is redrawn in Fig. 2.4.13a. The circuit is excited with a source voltage having a 50Ω source impedance (i.e., for a proper match at the input), and the output voltage is taken across the real part of the load impedance (i.e., the 10Ω resistor). For convenience, the two 3.18-nH inductors are combined, and the circuit can be viewed as performing a match between a 10Ω load and a 50Ω input resistance. The frequency response of the circuit

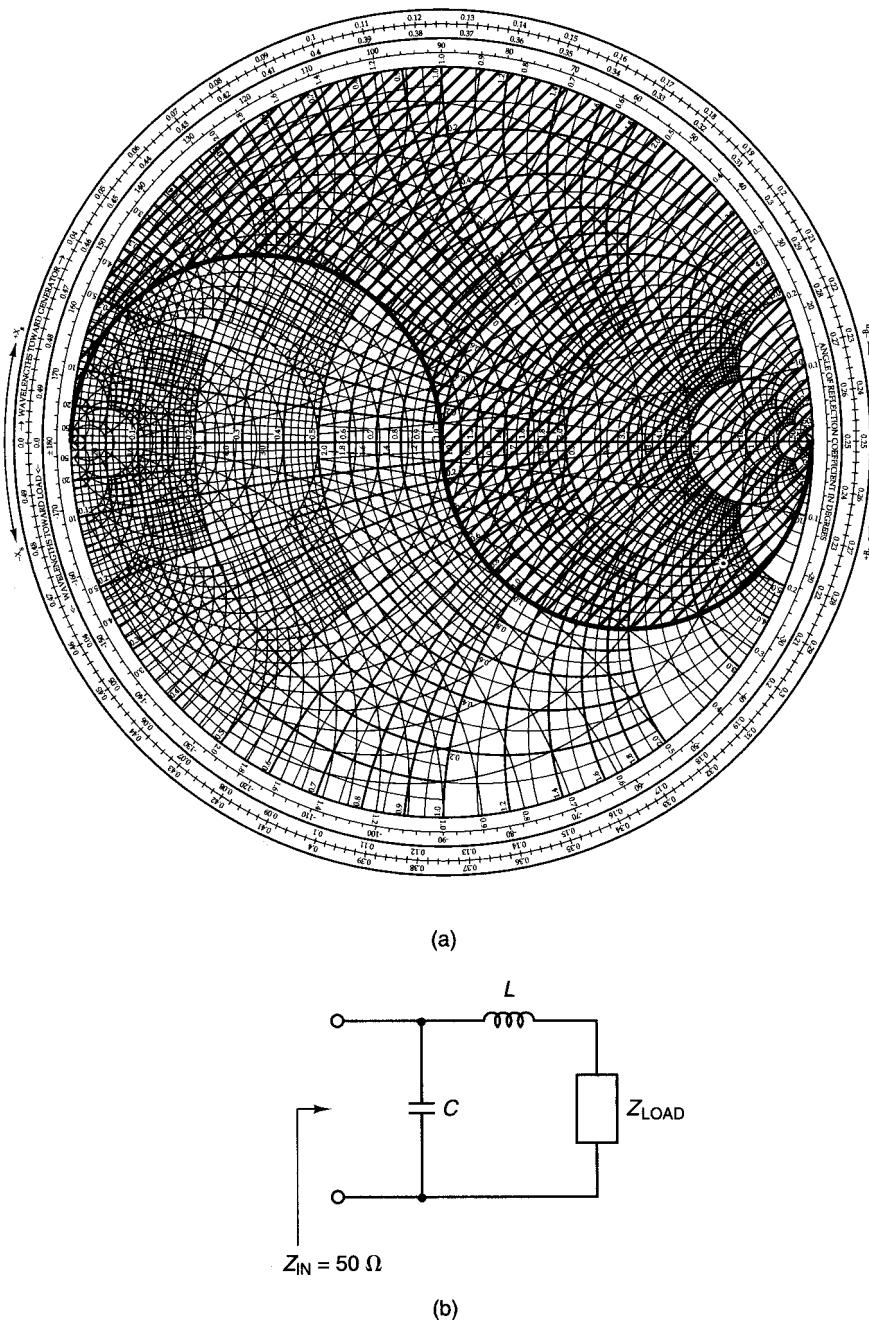


Figure 2.4.12 Forbidden region in the ZY Smith chart to match a given Z_{LOAD} to 50 Ω using a series L-shunt C matching network.

is illustrated in Fig. 2.4.13b. Figure 2.4.13b shows that a peak occurs at the frequency f_o where the match occurs, and the filter attenuates the frequencies above and below f_o .

For a narrowband range of frequencies around f_o , the filter can be viewed as a bandpass filter with a loaded Q , which can be calculated using (2.4.1). The equivalent bandpass filter is shown in Fig. 2.4.13c, as well as the frequency response. This circuit was obtained by changing $Z_B = 10 + j20 \Omega$ to an equivalent admittance (i.e., 50 Ω in parallel with $j25 \Omega$). The tuned circuit in Fig. 2.4.13c is loaded by the resistance $R_T = 50 \parallel 50 = 25 \Omega$. Hence, the loaded Q of the bandpass filter is given by

$$Q_L = \frac{\omega_o}{\text{BW}} = \omega_o R_T C = \frac{|B_C|}{G_T} = \frac{R_T}{|X_C|} = \frac{25}{25} = 1 \quad (2.4.4)$$

In Fig. 2.4.13c, since the gain at $f_o = 500$ MHz is -6 dB, it follows that the value of $|v_B|$ is $|v_s|/2$. Then, from Fig. 2.4.13a, we have

$$|v_o| = |v_B| \frac{10}{|10 + j20|} = \frac{|v_s|}{2} (0.447) = |v_s|(0.224)$$

or

$$20 \log \frac{|v_o|}{|v_s|} = -13 \text{ dB}$$

in agreement with the gain at $f_o = 500$ MHz in Fig. 2.4.13b.

The equivalent bandpass filter helps to explain the *bell-shape* response of the circuit in the neighborhood of f_o . Away from the neighborhood of f_o , the frequency responses of the circuit in Figs. 2.4.13a and 2.4.13c are quite different.

A question arises about the relation between the circuit node Q and the loaded Q associated with the response in the neighborhood of f_o . Referring to Fig. 2.4.13a, it is also observed that the circuit node Q looking toward the 50- Ω source is 2 (since the normalized admittance of the 50- Ω source and the $-j25 \Omega$ capacitor is $1 + j2$). Comparing Q_n in (2.4.3) with Q_L in (2.4.4), it follows that in the neighborhood of f_o the loaded Q of the Ell matching networks is given by

$$Q_L = \frac{Q_n}{2} \quad (2.4.5)$$

Using (2.4.5), the loaded Q of the circuit in Fig. 2.4.13a is

$$Q_L = \frac{Q_n}{2} = \frac{2}{2} = 1$$

Hence, the approximate bandwidth of the circuit in Fig. 2.4.13a is

$$\text{BW} \approx \frac{f_o}{Q_L} = \frac{500 \times 10^6}{1} = 500 \text{ MHz}$$

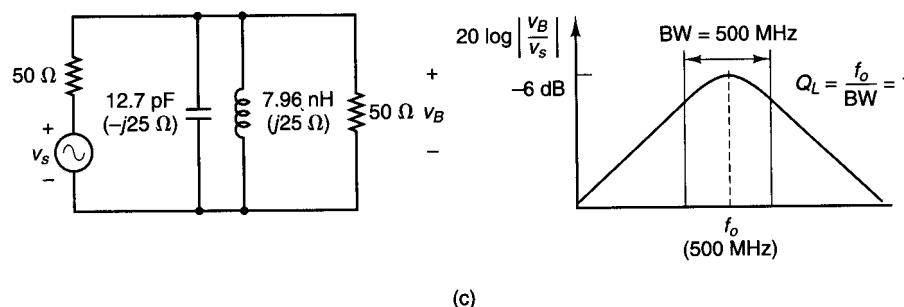
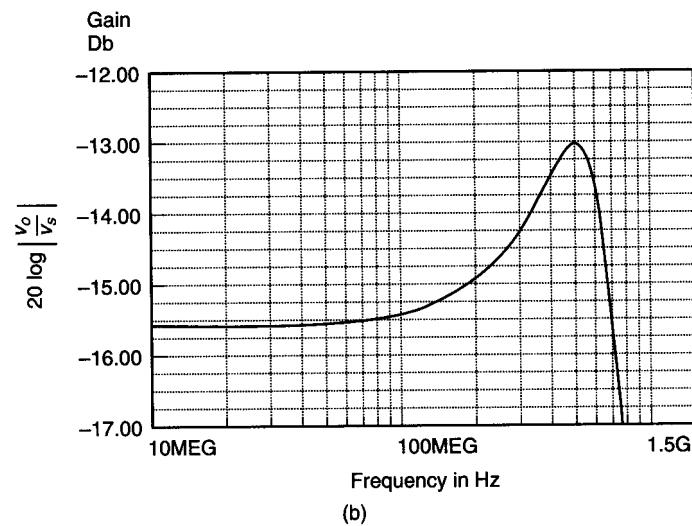
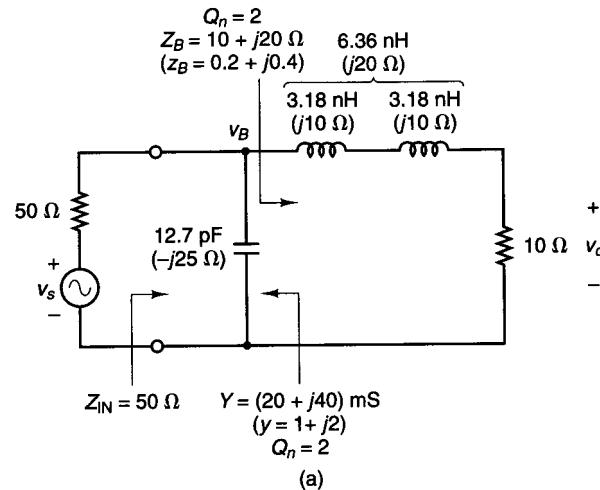


Figure 2.4.13 (a) The matching network from Fig. 2.4.8a; (b) its frequency response (the program *Microcap III* was used); (c) equivalent bandpass filter configuration in the neighborhood of f_o , and its frequency response.

Sec. 2.4 Impedance Matching Networks

The frequency response in Fig. 2.4.13b shows that the 3-dB point for $f > f_o$ occurs at $f = 720 \text{ MHz}$. For $f < f_o$, there is no 3-dB point. This is due to the low value of Q_n in Fig. 2.4.13a and, consequently, to the frequency response characteristics of the circuit for $f < f_o$. However, if we assume that the bell-shape response is symmetrical around f_o , it follows that the bandwidth is approximately

$$\text{BW} \approx 2(720 \times 10^6 - 500 \times 10^6) = 440 \text{ MHz}$$

in reasonable agreement with the predicted bandwidth.

The following example illustrates the bandpass characteristics of an Ell network when a high value of Q_n is used.

Example 2.4.5

The low-pass Ell network shown in Fig. 2.4.14a was designed to transform a 200Ω load to an input resistance of 20Ω (see Fig. 2.4.14b) at 500 MHz . Determine the loaded Q of the circuit.

Solution. From Fig. 2.4.14b, the impedance at point B is $z_B = 0.1 - j0.3$. Hence, the circuit node Q is $Q_n = 0.3/0.1 = 3$. Also, observe that the normalized impedance looking toward the source is $z_B = 0.1 + j0.3$, giving $Q_n = 3$.

From (2.4.4), it follows that Q_L is

$$Q_L = \frac{Q_n}{2} = \frac{3}{2} = 1.5$$

and the expected bandwidth is

$$\text{BW} \approx \frac{f_o}{Q_L} = \frac{500 \times 10^6}{1.5} = 333.33 \text{ MHz}$$

The frequency response of the circuit in Fig. 2.4.14c shows that $f_o = 500 \text{ MHz}$ and $\text{BW} = 650 \times 10^6 - 275 \times 10^6 = 375 \text{ MHz}$. Hence,

$$Q_L = \frac{500}{375} = 1.33$$

These values agree with the expected bandwidth and Q_L .

In conclusion, in order to obtain a high value of Q_L , the circuit node Q must be high. For the low-pass matching circuit, the attenuation characteristics for frequencies above f_o are obviously better than for frequencies below f_o . For bandpass applications, the attenuation of the harmonics above f_o is important, and the Ell filters will provide an attenuation of -12 dB/octave .

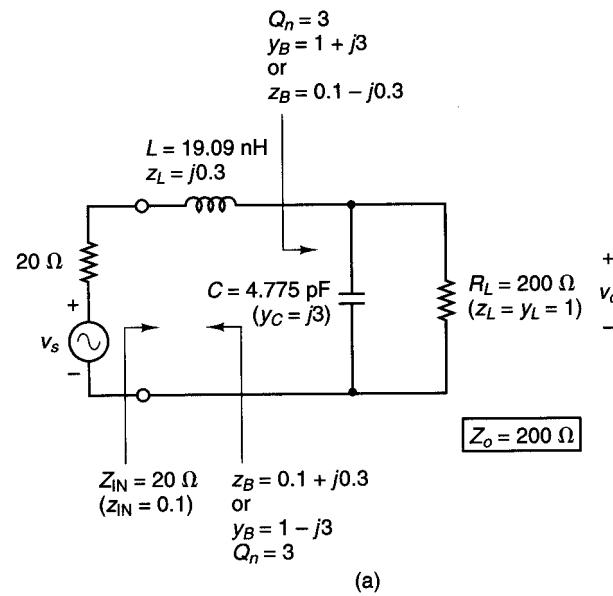


Figure 2.4.14 (a) The low-pass matching circuit for Example 2.4.5; (b) the design in the ZY Smith chart; (c) the frequency response (the program *Microcap-III* was used).

In many cases, the matching network provides a match to a complex impedance (or admittance). The frequency response of these circuits can be obtained using an excitation whose impedance (or admittance) is the complex conjugate of the impedance (or admittance) presented by the matching network. In some cases, it is convenient to represent the source excitation by its Norton's equivalent circuit. For example, in Fig. 2.4.10c the matching network was designed to present an output admittance of $Y_{\text{OUT}} = (8 - j12) \text{ mS}$. The frequency response of this circuit can be performed using a current source in parallel with a source admittance of $(8 + j12) \text{ mS}$.

When a high value of Q_L is a design consideration, the Ell matching networks in Fig. 2.4.2 might not be suitable since the Q_L cannot be controlled when matching with two elements. In fact, with the Ell matching networks we have to accept the resulting Q_n in the design. For example, in the design shown in Fig. 2.4.13a the value of Q_n is determined by the equivalent impedance at point B —namely, $z = 0.2 + j0.4$; therefore, in this circuit Q_n is fixed at the value of 2.

Higher values of Q_L than those obtained with the Ell circuits can be obtained using matching circuits with three elements. The addition of a third element to an Ell matching network results in either the lossless Tee network (also denoted as T network) or the lossless Pi network (also denoted as Π network) shown in Fig. 2.4.15.

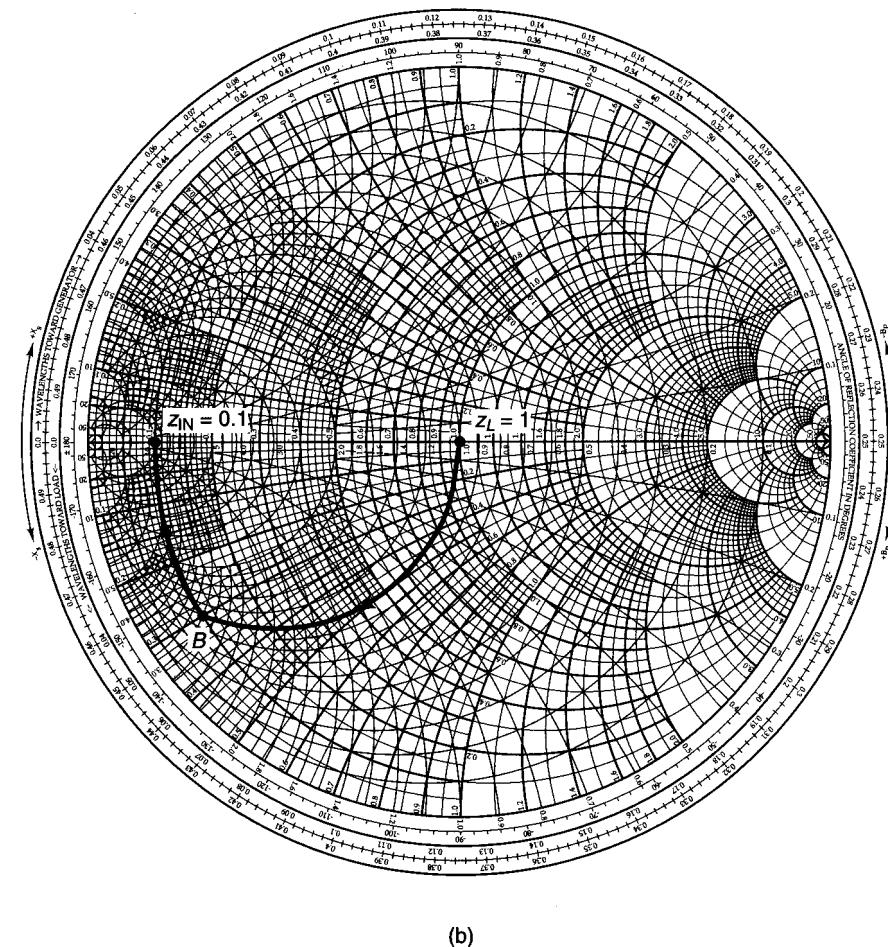


Figure 2.4.14 Continued

The addition of a third element introduces flexibility in the selection of the loaded Q , since the equivalent series impedance (or the equivalent parallel admittance) at the nodes in the circuit will determine various values of Q_n . Obviously, a high value of Q_n in the circuit will result in a high value of Q_L . However, it is not simple to exactly relate Q_n to Q_L in these circuits. The Q of a Tee or Pi network is normally taken as the highest value of Q_n in the circuit.

Several equivalent series input impedances can have the same Q_n . For example, the normalized impedances $z = 0.5 \pm j0.5$ have a Q_n of 1, as well as the impedances $z = 1 \pm j1$. Constant Q_n contours can be drawn on the Smith chart. This is illustrated in Fig. 2.4.16, where constant Q_n contours of 1, 5, and 10 are shown.

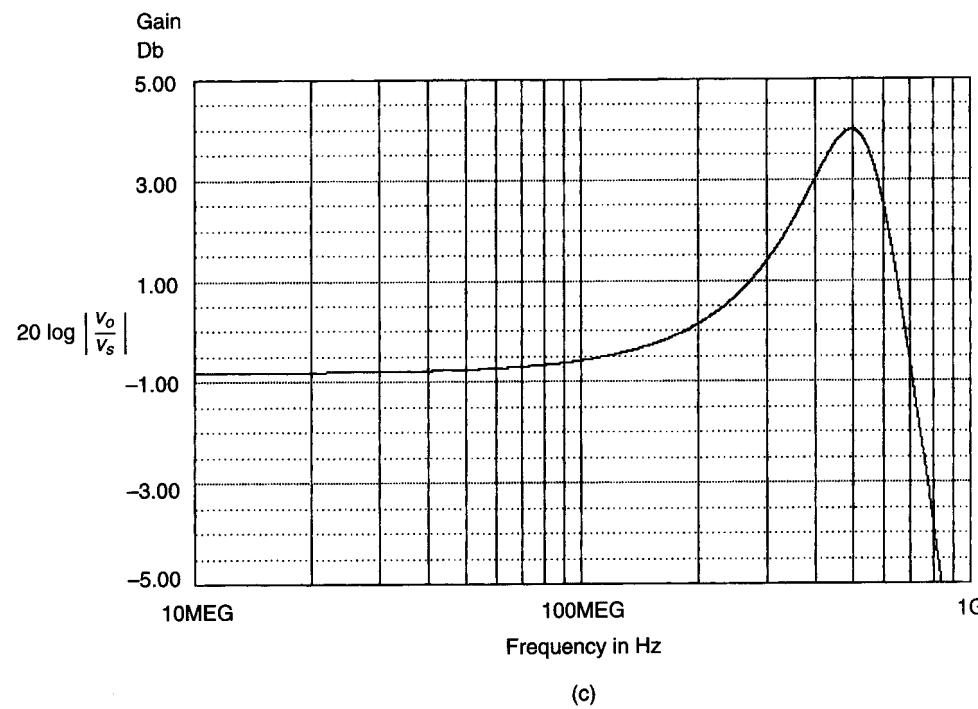


Figure 2.4.14 Continued

The upper and lower parts of the constant- Q contours can be shown to satisfy a circle equation as follows. Since

$$z = r + jx = \frac{1 + \Gamma}{1 - \Gamma} = \frac{1 - U^2 - V^2}{(1 - U)^2 + V^2} + j \frac{2U}{(1 - U)^2 + V^2}$$

then

$$Q_n = \frac{|x|}{r} = \frac{2U}{1 - U^2 - V^2}$$

which can be written as

$$U^2 + \left(V \pm \frac{1}{Q_n} \right)^2 = 1 + \frac{1}{Q_n^2} \quad (2.4.6)$$

The plus sign applies when x is positive, and the minus sign when x is negative.

Equation (2.4.6) is recognized as the equation of a circle. For $x > 0$, the center in the Γ plane is at $(0, -1/Q_n)$, and for $x < 0$ at $(0, 1/Q_n)$; the radius of the circle is

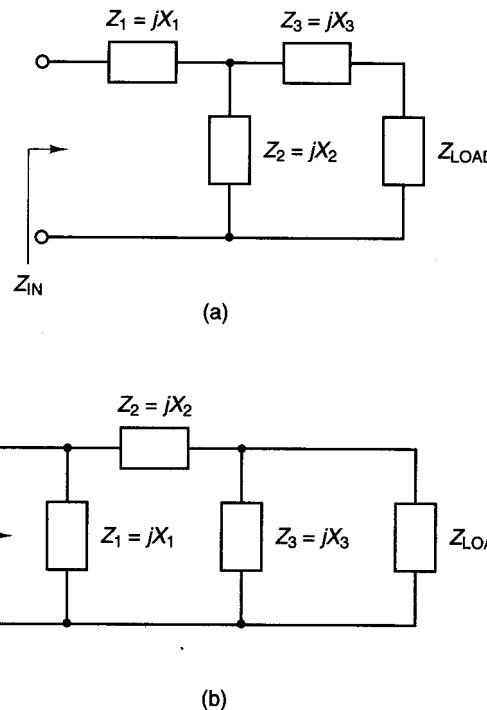


Figure 2.4.15 (a) Lossless Tee network; (b) lossless Pi network.

$$\sqrt{1 + \frac{1}{Q_n^2}}$$

For example, for the $Q_n = 5$ contour in Fig. 2.4.16, the upper and lower parts of the contour are simply one half of a circle centered at $(0, \mp 0.2)$, respectively, with radius of

$$\sqrt{1 + \frac{1}{25}} = 1.02$$

The following example illustrates the design of Tee and Pi networks for a given Q_n factor.

Example 2.4.6

Design two Tee networks to transform the load impedance $Z_{LOAD} = 50 \Omega$ to the input impedance $Z_{IN} = 10 - j15 \Omega$ with a Q_n of 5.

Solution. The design of a Tee matching network with a Q_n of 5 is illustrated in Fig. 2.4.17a, where a normalizing impedance of 50Ω was used. The motion from A to B produces a series inductor with impedance $z_{L_1} = j2$; the motion from B to C produces a

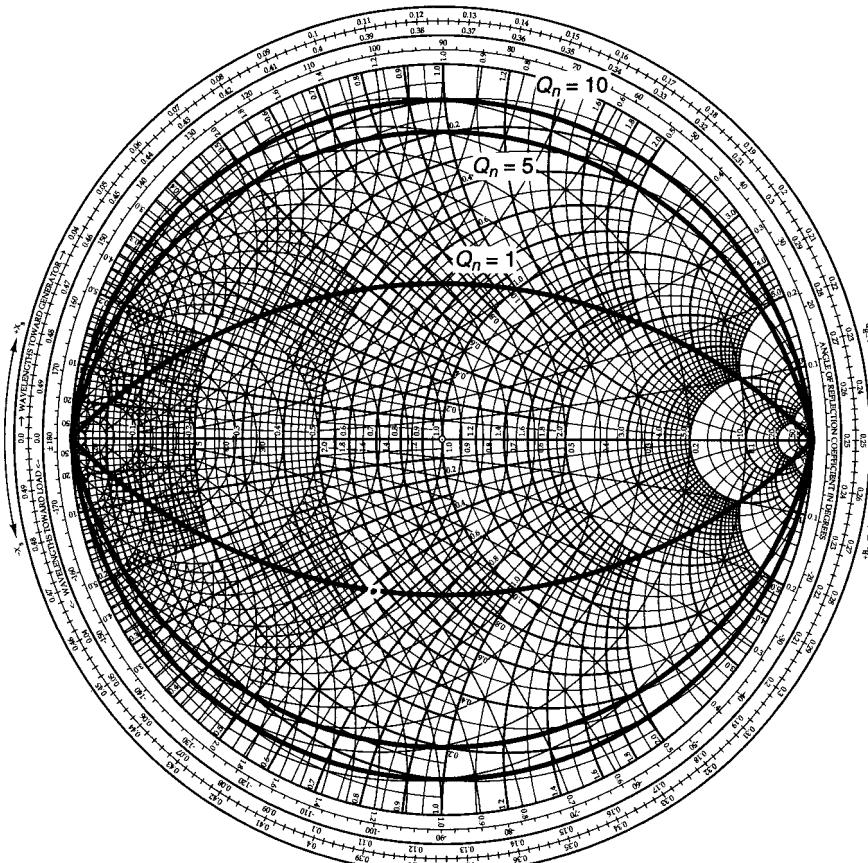


Figure 2.4.16 Constant Q_n contours for $Q_n = 1, 5$, and 10 .

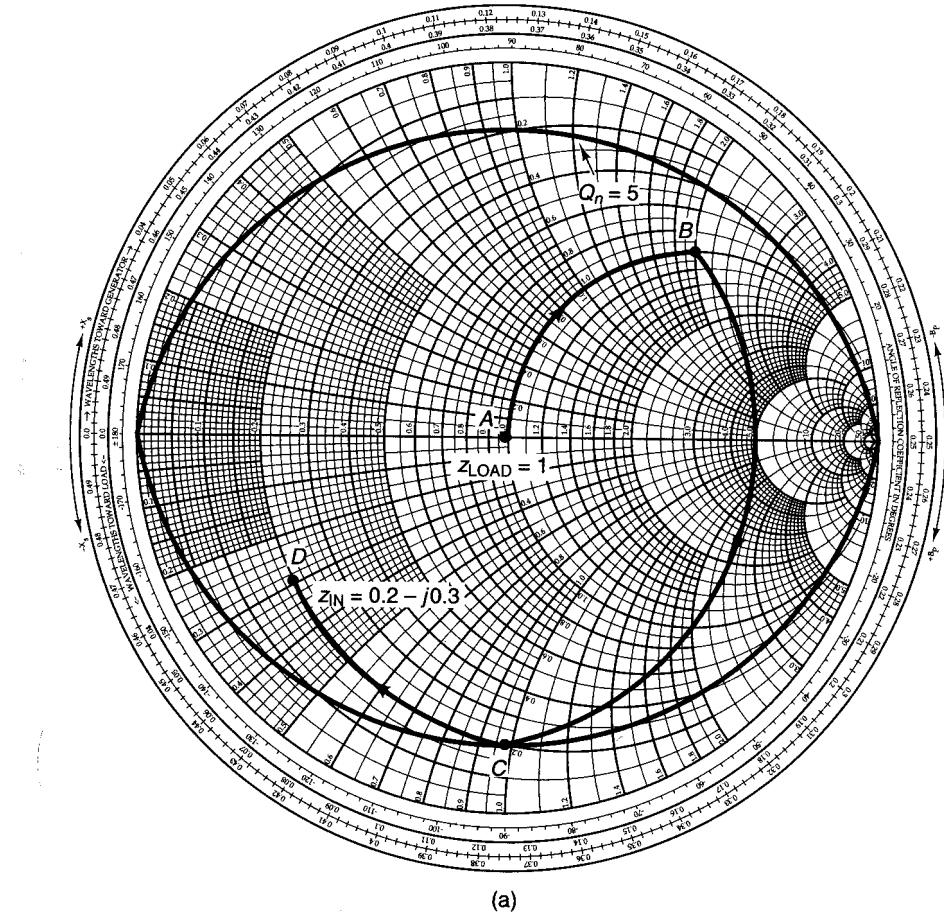
shunt capacitor with admittance $y_C = j0.96 - (-j0.4) = j1.36$ (or $z_C = 1/j1.36 = -j0.735$); and the motion from C to D produces a series inductor with impedance $z_{L_2} = -j0.3 - (-j1) = j0.7$. The impedance value at point C , which touches the $Q_n = 5$ contour, determines the Q of the network. The resulting Tee network is shown in Fig. 2.4.17b, and its frequency response at 500 MHz is shown in Fig. 2.4.17c. From Fig. 2.4.17c, the loaded Q is calculated to be

$$Q_L = \frac{f_o}{\text{BW}} = \frac{500 \times 10^6}{568 \times 10^6 - 382 \times 10^6} = 2.7$$

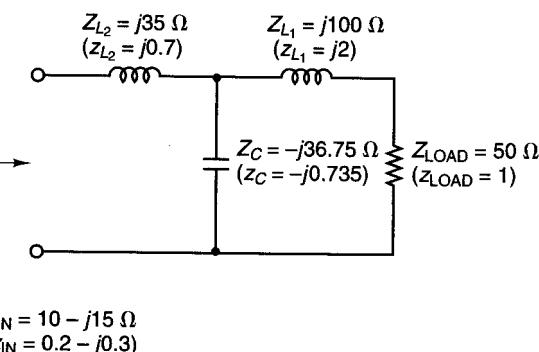
It is simple to verify that the other Tee network design, shown in Fig. 2.4.17d, transforms $Z_{\text{LOAD}} = 50 \Omega$ to $Z_{\text{IN}} = 10 - j15 \Omega$.

Example 2.4.7

Design two Pi networks to transform the load impedance $Z_{\text{LOAD}} = 50 \Omega$ to the input impedance $Z_{\text{IN}} = 150 \Omega$ with a Q_n of 5.



(a)



(b)

Figure 2.4.17 (a) Smith chart design of a Tee network for Example 2.4.6; (b) the Tee matching network; (c) the frequency response (the program *Microcap-III* was used); (d) another Tee network design.

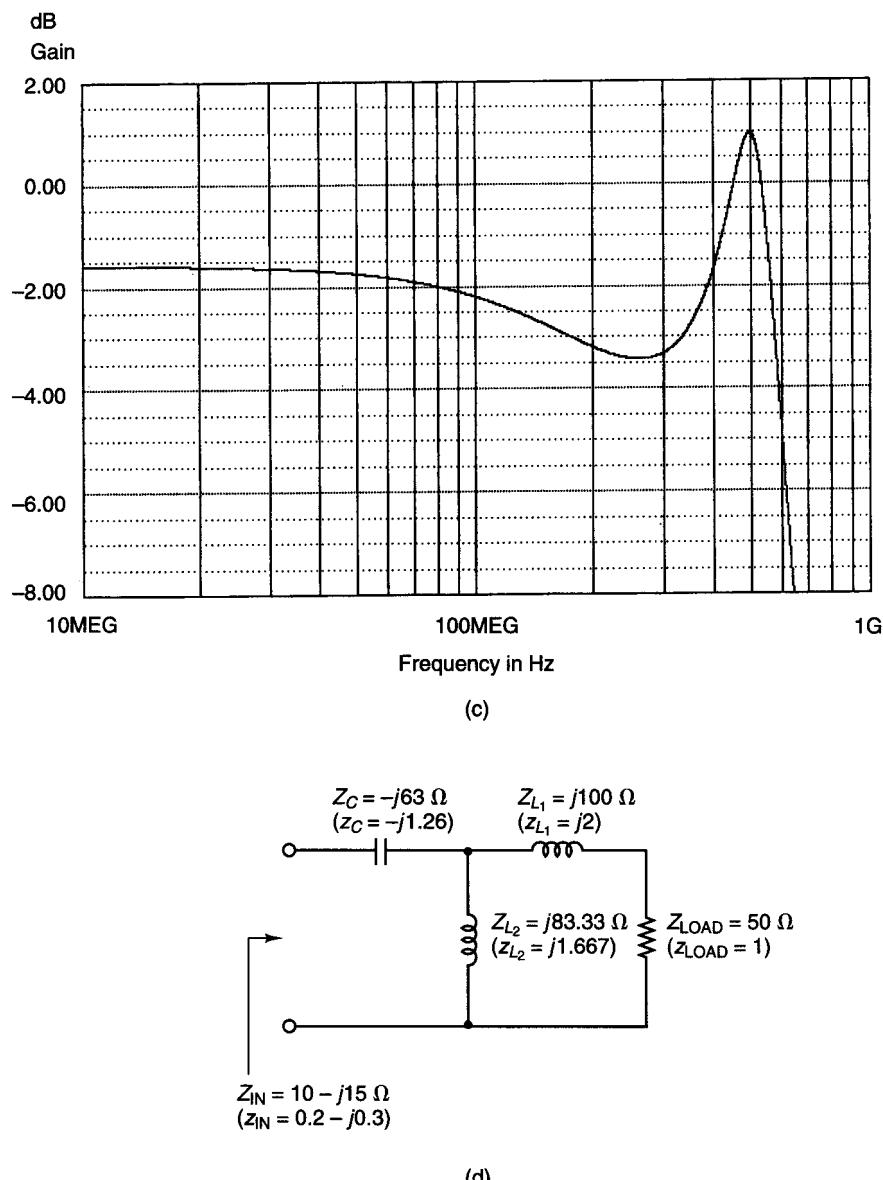


Figure 2.4.17 Continued

Solution. The design of a Pi network with a Q_n of 5 is illustrated in Fig. 2.4.18a. The motion from A to B requires a shunt inductor with admittance $y_{L_1} = -j2.9$ (or $z_{L_1} = j0.345$); the motion from B to C is produced by a series inductor with impedance $z_{L_2} = j0.55 - j0.31 = j0.24$. Point C is on the $Q_n = 5$ contour; therefore, the circuit Q is 5. Finally, the motion from C to D produces a shunt capacitor whose admittance is $y_C = 0 - (-j1.75) = j1.75$ (or $z_C = -j0.571$). The resulting Pi network is shown in Fig. 2.4.18b.

It is simple to verify that the other Pi network design, shown in Fig. 2.4.18c, transforms $Z_{LOAD} = 50 \Omega$ to $Z_{IN} = 150 \Omega$.

The design of lossless matching networks is accomplished by moving along constant-resistance and constant-conductance circles in the Smith chart. A question commonly asked is, "What happens if one moves along a constant-reactance or constant-susceptance circle in the Smith chart?". The answer is simple: A motion along a constant-reactance circle introduces a series resistance, and a motion along a constant-susceptance circle introduces a shunt resistance. The resistor is positive if the motion is along the constant resistance circle in the direction that r increases, or along the constant susceptance circle in the direction that g increases. It is important to emphasize that resistors are lossy elements and, in general, are avoided in the design of matching networks.

Example 2.4.8

A matching network is designed as shown in the ZY Smith chart in Fig. 2.4.19a to transform a 50Ω load to the input impedance $Z_{IN} = 50(1.6 - j1.2) \Omega$. Draw the matching circuit and determine the component values.

Solution. The motion from A to B results in a series inductor of value $z_L = j0.6$ (or $Z_L = 50(j0.6) = j30 \Omega$). The motion from B to C , along the constant-reactance circle $x_L = 0.6$, results in a series resistor. The impedance at B is $z_B = 1 + j0.6$, and the impedance at C is $z_C = 2.3 + j0.6$. Therefore, the value of the series resistor is $r = r_C - r_B = 2.3 - 1 = 1.3$, or $R = 50r = 50(1.3) = 65 \Omega$. Finally, the motion from C to D results in a capacitor of value $y_C = j0.3 - (-j0.1) = j0.4$ [or $z_C = -j2.5$, $Z_C = 50(-j2.5) = -j125 \Omega$]. The matching circuit is shown in Fig. 2.4.19b.

In a microwave amplifier (see Fig. 2.4.1), the input and output matching networks provide the appropriate ac impedances to the transistor. The transistor must also be biased at an appropriate quiescent point. A complete microwave amplifier contains both dc bias components and the ac matching networks. RFCs, bypass capacitors, and coupling capacitors need to be introduced so the dc bias components do not affect the ac performance of the amplifier. An example of a discrete microwave amplifier is shown in Fig. 2.4.20a. The capacitors denoted by C_B are bypass capacitors (they behave like short circuits at the frequency of operation). The RFCs behave like open circuits at the frequency of operation. The resistors R_1, R_2, R_C , and R_E set the quiescent point of the transistor. The input matching circuit consists of C_1 and L_1 , and the output matching circuit consists of C_2 and L_2 . In addition, the capacitors C_1 and C_2 act like coupling capacitors. That is, they isolate the dc bias circuit from the

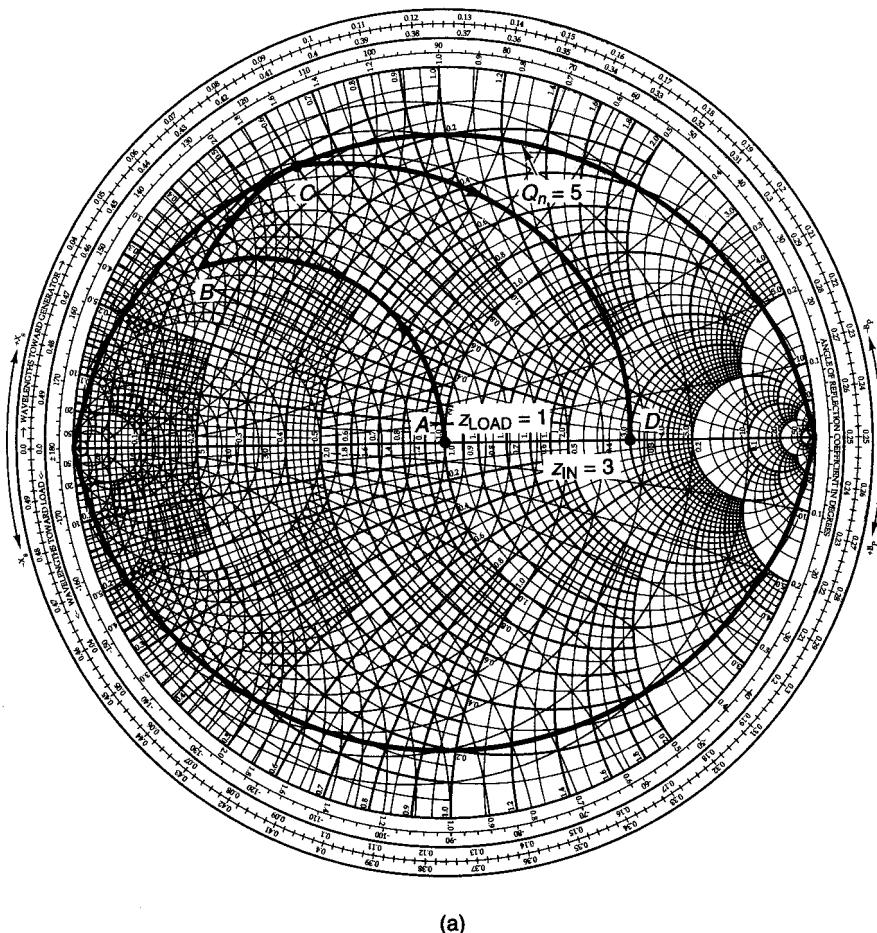


Figure 2.4.18 (a) Smith chart design of a Pi network for Example 2.4.7; (b) the Pi matching network; (c) another Pi network design.

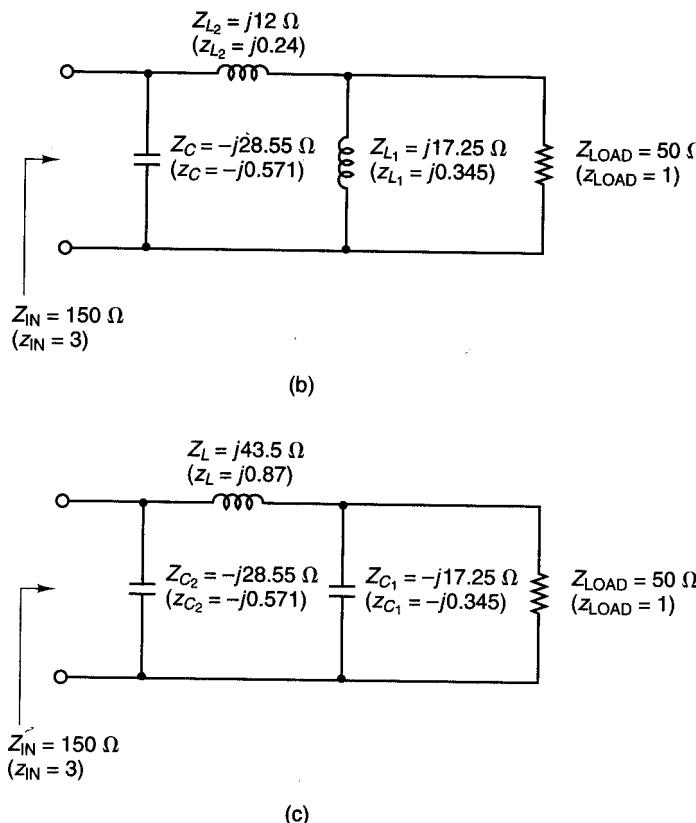
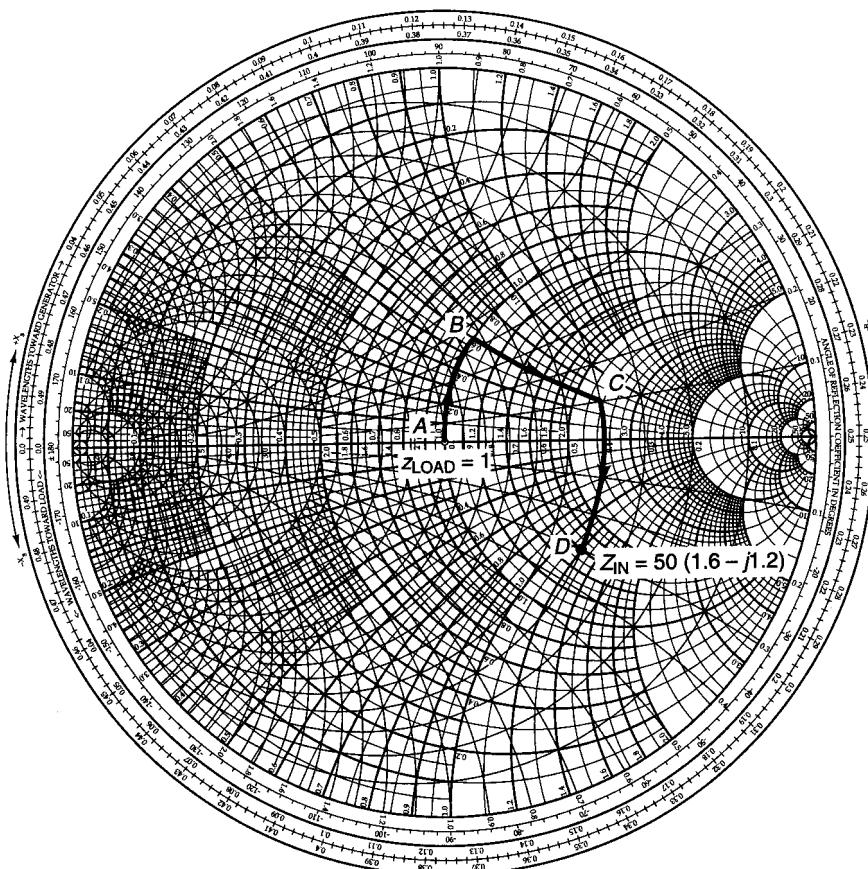
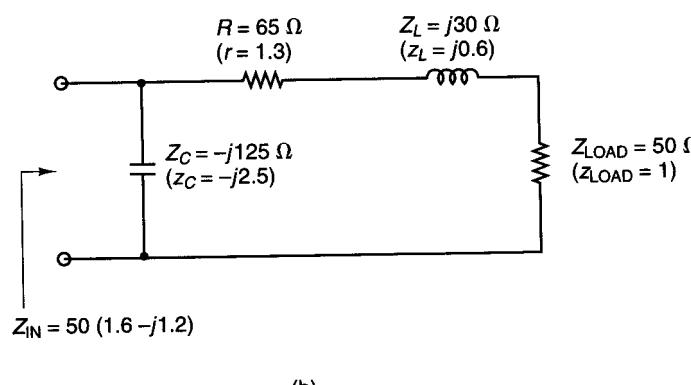


Figure 2.4.18 Continued



(a)



(b)

Figure 2.4.19 (a) Matching of $z_{LOAD} = 1$ to $z_{IN} = 1.6 - j1.2$; (b) the matching circuit.

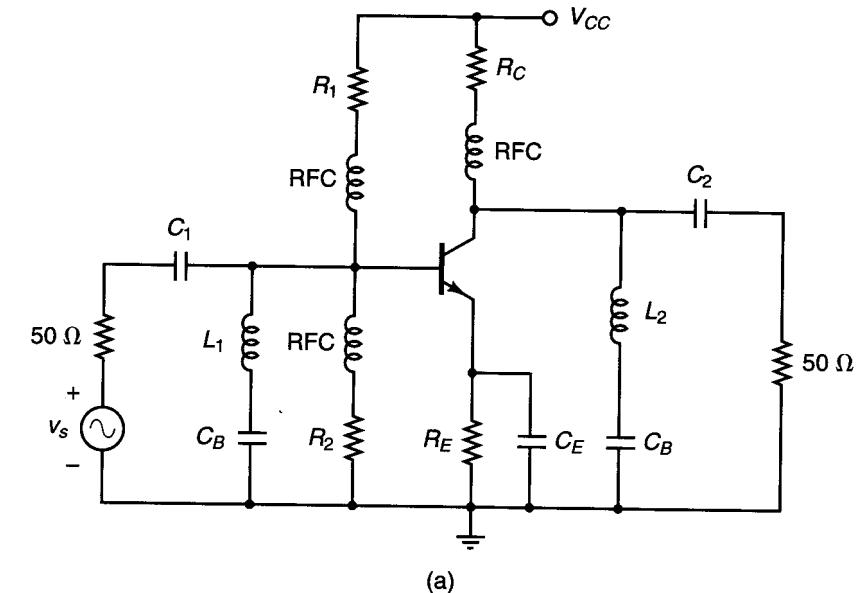


Figure 2.4.20 (a) A discrete microwave amplifier; (b) the dc model; (c) the ac model.

input source. The dc model of the amplifier is shown in Fig. 2.4.20b, and the ac model in Fig. 2.4.20c.

2.5 MICROSTRIP MATCHING NETWORKS

Microstrip lines

Microstrip lines are used extensively in building microwave transistor amplifiers because they are easily fabricated using printed-circuit techniques. Network interconnections and the placement of lumped and transistor devices

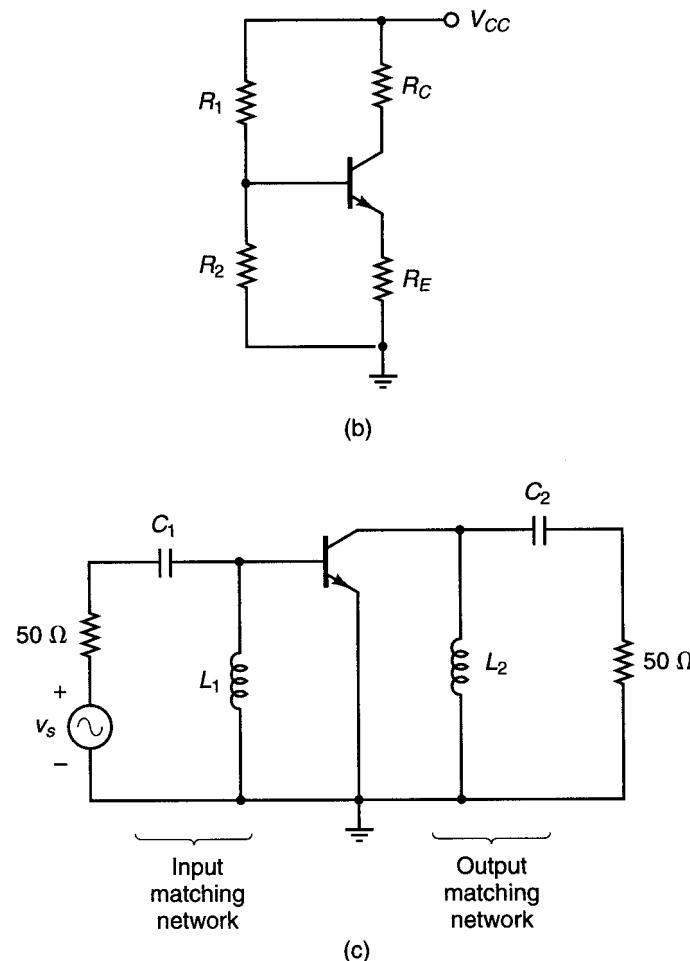


Figure 2.4.20 Continued

are easily made on its metal surface. The superior performance characteristics of the microstrip line make it one of the most important mediums of transmission in microwave transistor amplifiers and in microwave integrated-circuit technology.

A microstrip line is, by definition, a transmission line consisting of a strip conductor and a ground plane separated by a dielectric medium. Figure 2.5.1

illustrates the microstrip geometry. The dielectric material serves as a substrate and is sandwiched between the strip conductor and the ground plane. The relative dielectric constant of the substrate, ϵ_r , and ϵ are related by $\epsilon = \epsilon_r \epsilon_0$, where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m. Some typical dielectric substrates are RT/Duroid® (a trademark of Rogers Corporation, Chandler, Arizona), which is available with several values of ϵ_r (e.g., $\epsilon = 2.23\epsilon_0$, $\epsilon = 6\epsilon_0$, $\epsilon = 10.5\epsilon_0$, etc.); quartz ($\epsilon = 3.7\epsilon_0$), alumina ($\epsilon = 9\epsilon_0$); and Epsilam-10® ($\epsilon = 10\epsilon_0$).

The electromagnetic field lines in the microstrip are not contained entirely in the substrate. Therefore, the propagating mode in the microstrip is not a pure transverse electromagnetic mode (TEM mode) but a quasi-TEM. Assuming a quasi-TEM mode of propagation in the microstrip line, the phase velocity is given by

$$v_p = \frac{c}{\sqrt{\epsilon_{eff}}} \quad (2.5.1)$$

where c is the speed of light (i.e., 3×10^8 m/s) and ϵ_{eff} is the effective relative dielectric constant of the microstrip. The effective relative dielectric constant of the microstrip is related to the relative dielectric constant of the dielectric substrate and also takes into account the effect of the external electromagnetic fields (i.e., fringing effects must be considered).

Since $Z_o = \sqrt{L/C}$ and $v_p = 1/\sqrt{LC}$, the characteristic impedance of the microstrip line can be expressed in the form

$$Z_o = \frac{1}{v_p C} \quad (2.5.2)$$

where C is the capacitance per unit length of the microstrip. The wavelength in the microstrip line is given by

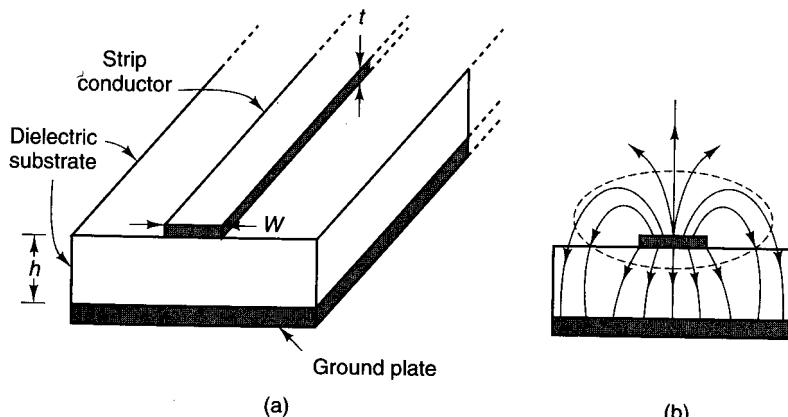


Figure 2.5.1 Microstrip geometry and field configuration. In (b), the solid lines represent electric field lines and the dashed line represents a magnetic field line.

$$\lambda = \frac{v_p}{f} = \frac{c}{f\sqrt{\epsilon_{eff}}} = \frac{\lambda_0}{\sqrt{\epsilon_{eff}}} \quad (2.5.3)$$

where λ_0 is the free-space wavelength.

As seen from (2.5.1), (2.5.2), and (2.5.3), the evaluation of v_p , Z_o , and λ in a microstrip line requires the evaluation of ϵ_{eff} and C . There are different methods for determining ϵ_{eff} and C and, of course, closed-form expressions are of great importance in microstrip-line design. The evaluation of ϵ_{eff} and C based on a quasi-TEM mode is accurate for design purposes at lower microwave frequencies. However, at higher microwave frequencies the longitudinal components of the electromagnetic fields are significant and the quasi-TEM assumption is no longer valid.

A useful set of relations for the characteristic impedance, assuming zero or negligible thickness of the strip conductor (i.e., $t/h < 0.005$), is as follows [2.1]:

For $W/h \leq 1$:

$$Z_o = \frac{60}{\sqrt{\epsilon_{eff}}} \ln \left(8 \frac{h}{W} + 0.25 \frac{W}{h} \right) \quad (2.5.4)$$

where

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[\left(1 + 12 \frac{h}{W} \right)^{-1/2} + 0.04 \left(1 - \frac{W}{h} \right)^2 \right] \quad (2.5.5)$$

For $W/h \geq 1$:

$$Z_o = \frac{120\pi/\sqrt{\epsilon_{eff}}}{W/h + 1.393 + 0.667 \ln(W/h + 1.444)} \quad (2.5.6)$$

where

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + 12 \frac{h}{W} \right)^{-1/2} \quad (2.5.7)$$

Plots of the characteristic impedance, as well as the normalized wavelength, as a function of W/h are shown in Figs. 2.5.2 and 2.5.3.

Based on the results in (2.5.3), (2.5.5), and (2.5.7) and/or in experimental data, the wavelength in the microstrip line, assuming zero or negligible thickness (i.e., $t/h \leq 0.005$) for the strip conductor, is given by the following relations [2.3]:

For $W/h \geq 0.6$:

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_r}} \left[\frac{\epsilon_r}{1 + 0.63(\epsilon_r - 1)(W/h)^{0.1255}} \right]^{1/2} \quad (2.5.8)$$

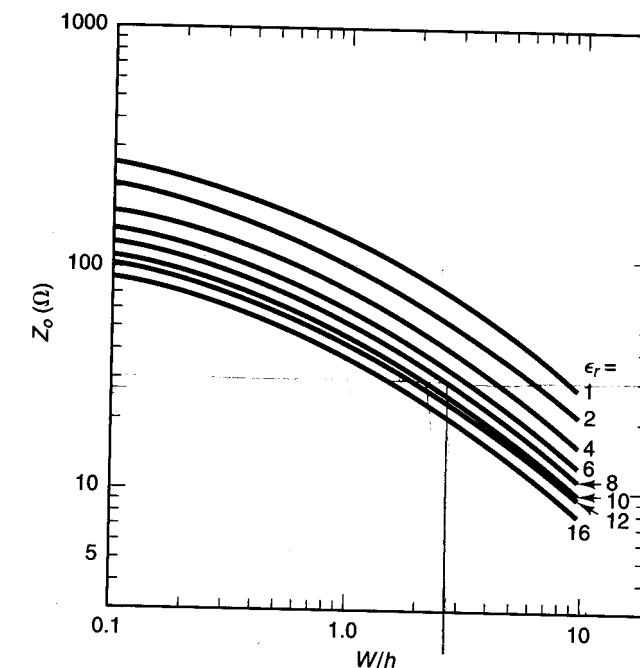


Figure 2.5.2 Characteristic impedance of the microstrip line versus W/h . (From H. Sobol [2.2]; copyright 1971, IEEE; reproduced with permission of IEEE.)

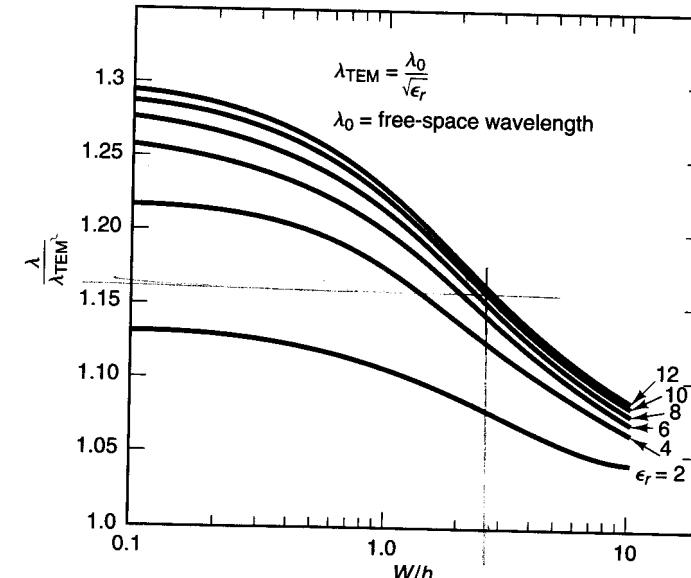


Figure 2.5.3 Normalized wavelength of the microstrip line versus W/h . (From H. Sobol [2.2]; copyright 1971, IEEE; reproduced with permission of IEEE.)

For $W/h < 0.6$:

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_r}} \left[\frac{\epsilon_r}{1 + 0.6(\epsilon_r - 1)(W/h)^{0.0297}} \right]^{1/2} \quad (2.5.9)$$

For design purposes, a set of equations relating Z_o and ϵ_r to the ratio W/h of the microstrip line is desirable. Assuming zero or negligible thickness of the strip conductor (i.e., $t/h \leq 0.005$), the expressions are as follows [2.2]:

For $W/h \leq 2$:

$$\frac{W}{h} = \frac{8e^A}{e^{2A} - 2} \quad (2.5.10)$$

For $W/h \geq 2$:

$$\frac{W}{h} = \frac{2}{\pi} \left\{ B - 1 - \ln(2B - 1) + \frac{\epsilon_r - 1}{2\epsilon_r} \left[\ln(B - 1) + 0.39 - \frac{0.61}{\epsilon_r} \right] \right\} \quad (2.5.11)$$

where

$$A = \frac{Z_o}{60} \sqrt{\frac{\epsilon_r + 1}{2}} + \frac{\epsilon_r - 1}{\epsilon_r + 1} \left(0.23 + \frac{0.11}{\epsilon_r} \right)$$

and

$$B = \frac{377\pi}{2Z_o \sqrt{\epsilon_r}}$$

The zero or negligible thickness formulas given in (2.5.4) to (2.5.11) can be modified to include the thickness of the strip conductor. The first-order effect of a strip conductor of finite thickness t is to increase the capacitance. Therefore, an approximate correction is made by replacing the strip width W by the effective width W_{eff} . The following relations for W_{eff}/h are useful when $t < h$ and $t < W/2$:

For $W/h \geq 1/2\pi$:

$$\frac{W_{\text{eff}}}{h} = \frac{W}{h} + \frac{t}{\pi h} \left(1 + \ln \frac{2h}{t} \right)$$

For $W/h \leq 1/2\pi$:

$$\frac{W_{\text{eff}}}{h} = \frac{W}{h} + \frac{t}{\pi h} \left(1 + \ln \frac{4\pi W}{t} \right)$$

The restrictions $t < h$ and $t < W/2$ are usually satisfied since for dielectric substrates a typical thickness is $t = 0.002$ in.

Various substrate materials are available for the construction of microstrip lines, with practical values of ϵ_r ranging from 2 to 10. The substrate material comes plated on both sides with copper, and an additional layer of gold

plating on top of the cooper is usually added after the circuit pattern is etched in order to prevent oxidation. Typical plating thickness of copper is from 1/2 mils to 2 mils (1 inch = 1000 mils).

The value of ϵ_r and the dielectric thickness (h) determine the width of the microstrip line for a given Z_o . These parameters also determine the speed of propagation in the line, and consequently its length. Typical dielectric thickness are 25, 30, 40, 50, and 100 mils.

Example 2.5.1

A microstrip material with $\epsilon_r = 10$ and $h = 1.016$ mm is used to build a transmission line. Determine the width for the microstrip transmission line to have a characteristic impedance of 50Ω . Also determine the wavelength and the effective relative dielectric constant of the microstrip line.

Solution. Figures 2.5.2 and 2.5.3 can be used to obtain the approximate values of W , λ , and ϵ_{ff} . From Fig. 2.5.2, with $Z_o = 50 \Omega$ and $\epsilon_r = 10$, the value of W/h is approximately 1. Hence,

$$W = h = 1.016 \text{ mm} = 0.1016 \text{ cm}$$

Sometimes W and h are expressed in mils. Since 1 inch = 1000 mils and 1 inch = 2.54 cm, it follows that the width and height in mils are

$$W = h = 0.1016 \left(\frac{1000}{2.54} \right) = 40 \text{ mils}$$

From Fig. 2.5.3, with $W/h = 1$ and $\epsilon_r = 10$, it follows that the value of $\lambda/\lambda_{\text{TEM}}$ is approximately 1.23, or

$$\lambda = 1.23\lambda_{\text{TEM}} = 1.23 \frac{\lambda_0}{\sqrt{\epsilon_r}} = \frac{1.23}{\sqrt{10}} \lambda_0 = 0.389\lambda_0$$

Hence [see (2.5.3)],

$$\epsilon_{\text{ff}} = \left(\frac{1}{0.389} \right)^2 = 6.61$$

More accurate values for W , λ , and ϵ_{ff} can be obtained using (2.5.8) to (2.5.11). From (2.5.10),

$$A = \frac{50}{60} \sqrt{\frac{10 + 1}{2}} + \frac{10 - 1}{10 + 1} \left(0.23 + \frac{0.11}{10} \right) = 2.1515$$

and

$$\frac{W}{h} = \frac{8e^{2.1515}}{e^{2(2.1515)} - 2} = 0.9563$$

Then

$$W = 0.9563(40) = 38.2 \text{ mils}$$

From (2.5.8),

$$\lambda = \frac{\lambda_0}{\sqrt{10}} \left[\frac{10}{1 + 0.63(10 - 1)(0.9563)^{0.1255}} \right]^{1/2} = 0.387\lambda_0$$

and

$$\epsilon_{ff} = \left(\frac{1}{0.387} \right)^2 = 6.68$$

The width and ϵ_{ff} of several 50-Ω microstrip lines for various values of ϵ_r and h were calculated and are tabulated in Fig. 2.5.4. For example, for a microstrip line having an alumina substrate with $h = 25$ mils and $\epsilon_r = 9.6$, it follows from Fig. 2.5.4 that $W = 24.7$ mils for $Z_o = 50 \Omega$ and $\epsilon_{ff} = 6.46$.

The formulas presented thus far are valid at frequencies where the quasi-TEM assumption can be made. When the quasi-TEM assumption is not valid, ϵ_{ff} and Z_o are functions of frequency and, therefore, the microstrip line becomes dispersive. The phase velocity of the microstrip line decreases with increasing frequency. Therefore, $\epsilon_{ff}(f)$ increases with frequency. Also, the characteristic impedance of the microstrip line increases with frequency, and it follows that the effective width $W_{eff}(f)$ decreases.

The frequency below which dispersion may be neglected is given by

$$f_o(\text{GHz}) = 0.3 \sqrt{\frac{Z_o}{h\sqrt{\epsilon_r - 1}}}$$

where h must be expressed in centimeters.

An analytical expression that shows the effect of dispersion in $\epsilon_{ff}(f)$ is [2.1]

$$\epsilon_{ff}(f) = \epsilon_r - \frac{\epsilon_r - \epsilon_{ff}}{1 + G(f/f_p)^2} \quad (\text{f in GHz})$$

	$\epsilon_r = 2.23$	4.54	6	9.6	10	30
$h = 25$ mils	$W = 76.4$ mils	46.7	37.5	24.7	23.8	6.01
	$\epsilon_{ff} = 1.91$	3.42	4.33	6.46	6.68	17.7
30	$W = 91.7$	56.1	45.0	29.7	28.5	7.21
	$\epsilon_{ff} = 1.91$	3.42	4.33	6.46	6.68	17.7
40	$W = 122.2$	74.7	60.0	39.6	38.0	9.6
	$\epsilon_{ff} = 1.91$	3.43	4.34	6.47	6.69	17.8
50	$W = 152.8$	93.4	75.0	49.4	47.6	12.0
	$\epsilon_{ff} = 1.91$	3.43	4.34	6.48	6.71	17.8
100	$W = 305.6$	186.8	150.1	98.9	95.1	24.1
	$\epsilon_{ff} = 1.91$	3.45	4.37	6.55	6.78	18.1

Figure 2.5.4 Width and ϵ_{ff} of microstrip lines for $Z_o = 50 \Omega$ for various ϵ_r and h values.

where

$$f_p = \frac{Z_o}{8\pi h} \quad (h \text{ in cm})$$

and

$$G = 0.6 + 0.009Z_o$$

Observe that when $f_p \gg f$, then $\epsilon_{ff}(f) \approx \epsilon_{ff}$. In other words, high-impedance lines on thin substrates are less dispersive.

The expression for the dispersion in Z_o is [2.1]

$$Z_o(f) = \frac{377h}{W_{eff}(f)\sqrt{\epsilon_{ff}}}$$

where

$$W_{eff}(f) = W + \frac{W_{eff}(0) - W}{1 + (f/f_p)^2}$$

and

$$W_{eff}(0) = \frac{377h}{Z_o(0)\sqrt{\epsilon_{ff}(0)}}$$

Another characteristic of the microstrip line is its attenuation. The attenuation constant is a function of the microstrip geometry, the electrical properties of the dielectric substrate and the conductors, and the frequency.

There are two types of losses in a microstrip line: a dielectric substrate loss and the ohmic skin loss in the conductors. The losses can be expressed as a loss per unit length along the microstrip line in terms of the attenuation factor α . Since the power carried by a wave traveling in the positive direction in a quasi-TEM mode is given by

$$P^+(z) = \frac{1}{2} \frac{|V^+|^2}{Z_o} e^{-2az} = P_0 e^{-2az} \quad (2.5.12)$$

where $P_0 = |V^+|^2/2Z_o$ is the power at $z = 0$. Then, from (2.5.12), we can write

$$\alpha = \frac{-dP(z)/dz}{2P(z)} = \alpha_d + \alpha_c$$

where α_d is the dielectric loss factor and α_c the conduction loss factor.

A useful set of expressions for calculating α_d is [2.1]

For a dielectric with low losses:

$$\alpha_d = 27.3 \frac{\epsilon_r}{\sqrt{\epsilon_{ff}}} \frac{\epsilon_{ff} - 1}{\epsilon_r - 1} \frac{\tan \delta}{\lambda_0} \quad \frac{\text{dB}}{\text{cm}} \quad (2.5.13)$$

where the loss tangent δ is given by

$$\tan \delta = \frac{\sigma}{\omega \epsilon}$$

For a dielectric with high losses:

$$\alpha_d = 4.34 \frac{\epsilon_{ff} - 1}{\sqrt{\epsilon_{ff}(\epsilon_r - 1)}} \left(\frac{\mu_0}{\epsilon_0} \right)^{1/2} \sigma \quad \frac{\text{dB}}{\text{cm}} \quad (2.5.14)$$

In (2.5.13) and (2.5.14), σ is the conductivity of the dielectric and $\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$.

A set of expressions for calculating α_c is [2.1]

For $W/h \rightarrow \infty$:

$$\alpha_c = \frac{8.68}{Z_o W} R_s$$

where

$$R_s = \sqrt{\frac{\pi f \mu_0}{\sigma}}$$

For $W/h \leq 1/2\pi$:

$$\alpha_c = \frac{8.68 R_s P}{2\pi Z_o h} \left[1 + \frac{h}{W_{\text{eff}}} + \frac{h}{\pi W_{\text{eff}}} \left(\ln \frac{4\pi W}{t} + \frac{t}{W} \right) \right]$$

For $1/2\pi < W/h \leq 2$:

$$\alpha_c = \frac{8.68 R_s}{2\pi Z_o h} PQ$$

For $W/h \geq 2$:

$$\alpha_c = \frac{8.68 R_s Q}{Z_o h} \left\{ \frac{W_{\text{eff}}}{h} + \frac{2}{\pi} \ln \left[2\pi e \left(\frac{W_{\text{eff}}}{2h} + 0.94 \right) \right] \right\}^{-2} \left[\frac{W_{\text{eff}}}{h} + \frac{W_{\text{eff}}/\pi h}{(W_{\text{eff}}/2h) + 0.94} \right]$$

where

$$P = 1 - \left(\frac{W_{\text{eff}}}{4h} \right)^2$$

and

$$Q = 1 + \frac{h}{W_{\text{eff}}} + \frac{h}{\pi W_{\text{eff}}} \left(\ln \frac{2h}{t} - \frac{t}{h} \right)$$

In dielectric substrates, the dielectric losses are normally smaller than conductor losses. However, dielectric losses in silicon substrates can be of the same order or larger than conductor losses.

The quality factor Q of a microstrip line is calculated from

$$Q = \frac{\beta}{2\alpha}$$

where

$$\beta = \frac{2\pi}{\lambda}$$

and α is the total loss. Therefore,

$$Q = \frac{\pi}{\lambda \alpha}$$

or in decibels we can write

$$\begin{aligned} Q &= \frac{8.686\pi}{\lambda \alpha} \quad \text{dB} \\ &= \frac{27.3}{\alpha} \quad \frac{\text{dB}}{\lambda} \end{aligned}$$

where we used the fact that 1 dB = 8.686 nepers.

A microstrip line also has radiation losses. The effect of radiation losses can be accounted for in terms of the radiation quality factor Q_r given by [2.1]

$$Q_r = \frac{Z_o}{480\pi(h/\lambda_0)F}$$

where

$$F = \frac{\epsilon_{ff}(f) + 1}{\epsilon_{ff}(f)} - \frac{(\epsilon_{ff}(f) - 1)^2}{2[\epsilon_{ff}(f)]^{2/3}} \ln \frac{\sqrt{\epsilon_{ff}(f)} + 1}{\sqrt{\epsilon_{ff}(f)} - 1}$$

is known as the *radiation factor*.

The total Q , called Q_T , of a microstrip resonator can be expressed as

$$\frac{1}{Q_T} = \frac{1}{Q_c} + \frac{1}{Q_d} + \frac{1}{Q_r}$$

where Q_d and Q_c are the quality factors of the dielectric (i.e., $Q_d = \pi/\lambda \alpha_d$) and conductor (i.e., $Q_c = \pi/\lambda \alpha_c$), respectively.

Microstrip lines can be etched to a typical accuracy of ± 1 mils for width greater than 5 mils. Circuit layouts using microstrips are easily done using a layout CAD program. For example, the Hewlett-Packard MDS program has a circuit layout capability that generates the circuit pattern from the circuit schematic. If a layout program is not available, a typical circuit layout first

involves making a scaled drawing of the circuit. A typical scale factor is 10. For example, a width of 24.7 mils in alumina microstrip with $h = 25$ mils produces a line with $Z_o = 50 \Omega$. A 1-to-1 scale drawing of a circuit using these lines would be almost impossible to obtain. Thus, it is common practice to draw the circuit in graph paper using a 10-to-1 scale. On a 10-to-1 scale, the previous $50-\Omega$ line would have a width of 247 mils, which is certainly a drawable quantity.

Electronic components come in various sizes. Some standard component size are as follows:

	Width (mils)	Length (mils)	Height (mils)
Transistors	100	120	20
Diodes	100	120	20
Chip capacitors	50	80	20
Chip coils (RFCs)	80	100	50
Chip resistors	50	80	15

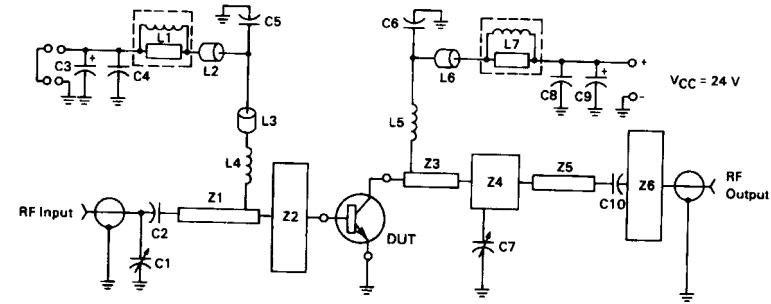
Lumped components and microstrip lines are drawn in graph paper, scaled by a factor of 10, to obtain the complete circuit layout. For example, a typical chip resistor in the scaled layout would have dimensions of $W = 500$ mils by $L = 800$ mils. Once the circuit drawing is done on a 10-to-1 scale, it is brought back to a 1-to-1 scale through photographic methods in which the reduction process is exact.

Manufacturers commonly provide test circuits for their transistors, as well as the photomaster of the test circuit (see Fig. 2.5.5).

Design of Matching Networks

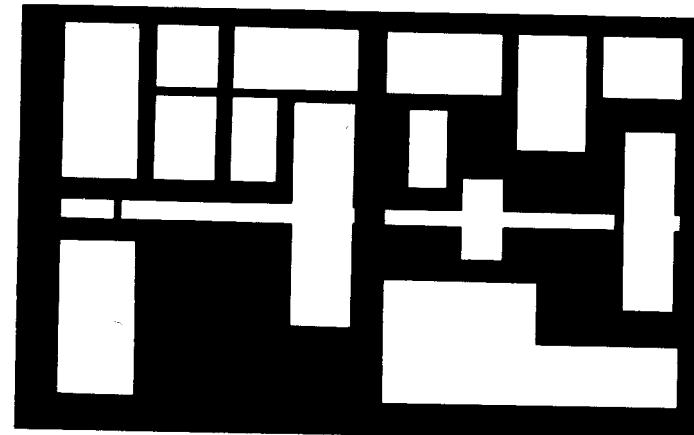
The impedance transforming properties of transmission lines can be used in the design of matching networks. A microstrip line can be used as a series transmission line, as an open-circuited stub, or as a short-circuited stub. In fact, a series microstrip line together with a short- or open-circuited shunt stub can transform a $50-\Omega$ resistor into any value of impedance.

The matching circuit configuration in which a short-circuited stub is connected in parallel with the load Z_L followed by a series microstrip transmission line is shown in Fig. 2.5.6a. The characteristic impedance of the microstrip lines is shown as 50Ω . Another way that is used to draw the schematic in Fig. 2.5.6a is shown in Fig. 2.5.6b. For comparison purposes, the same schematic drawn with $50-\Omega$ two-wire transmission lines is shown in



C1, C7 — Johanson 0.5 ~ 4.0 pF Giga-Trim
C2, C5, C6 — 91 pF Mini Underwood Micas
C3, C9 — 1.0 μ F Electrolytic
C4, C8 — 250 pF Unelco
C10 — 39 pF Mini Underwood
L1, L7 — 10 Turns Around 10 Ω 1/2 W Resistor
L2, L3, L6 — Ferrite Bead
L4, L5 — 4 Turns 26 AWG 0.1" ID
Z1, Z2, Z3, Z4, Z5, Z6 — Distributed Microstrip Elements (see photomask)
Board Material — Glass Teflon $\epsilon_r = 2.55$ $t = 0.031"$

— PHOTOMASTER FOR TEST FIXTURE



NOTE: The Printed Circuit Board shown is 75% of the original.

— 850-900 MHz TEST CIRCUIT

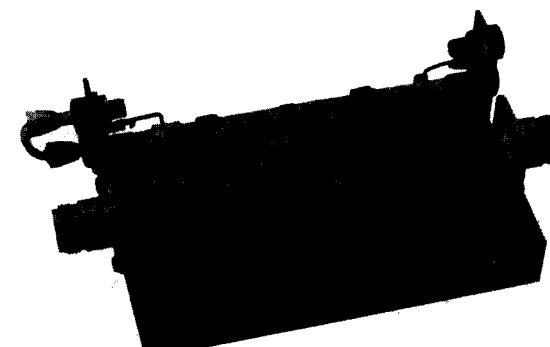


Figure 2.5.5 Test circuit, photomaster, and circuit construction for the MRF890 transistor. (From *Motorola RF Device Data*, Vol. 1, 6th edition; copyright of Motorola, used by permission.)

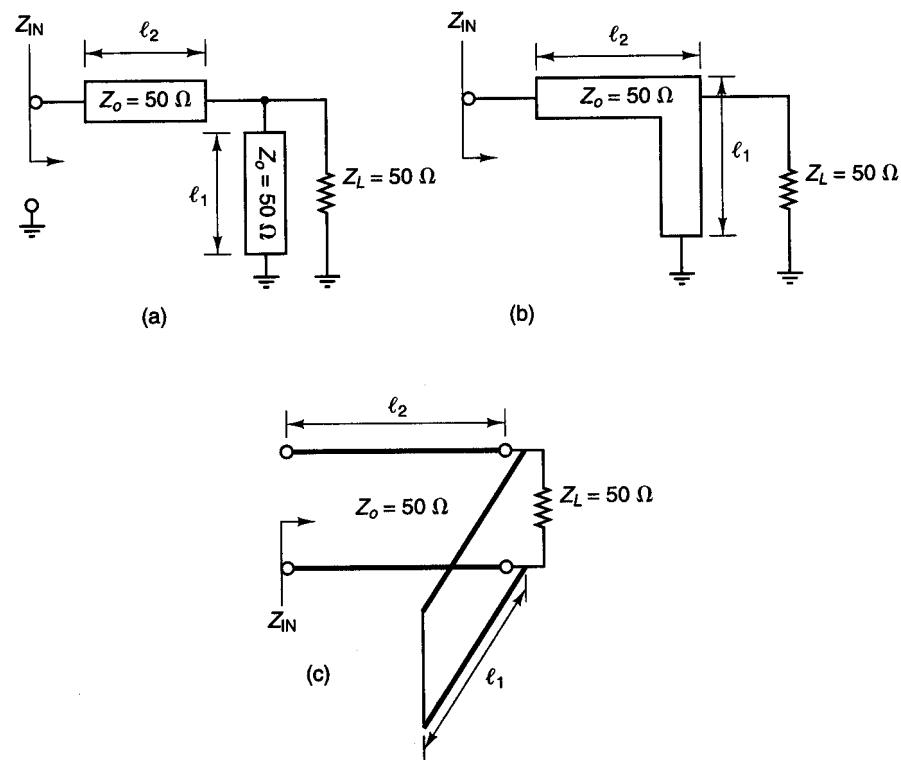


Figure 2.5.6 (a) A matching network using microstrip lines; (b) an alternative drawing; (c) schematic using two-wire transmission lines.

Fig. 2.5.6c. The shunt element in Fig. 2.5.6 could have been an open-circuited shunt stub instead of the short-circuited shunt stub shown in the figure. While the design procedure is discussed for a $50\text{-}\Omega$ load, it can be applied to an arbitrary load impedance.

The Smith chart design procedure for the matching circuit in Fig. 2.5.6a is now discussed. Since the admittance of the shunt stub adds to the load admittance, it is convenient to use the Y Smith chart to design the circuit. Figure 2.5.7a shows the matching circuit where $z_L = Z_L/Z_o = 50/50 = 1$ (or $y_L = 1/z_L = 1$). The normalized admittance of the shunt stub is written, for convenience, in the form $y_{sc} = jb_s$ ($b_s > 0$) for a capacitive susceptance, or $y_{sc} = -jb_s$ ($b_s > 0$) for an inductive susceptance. Using this notation, b_s is always positive and the length l_1 determines the value of $\pm jb_s$. The admittance y_x is given by

$$y_x = y_L + y_{sc} = 1 \pm jb_s$$

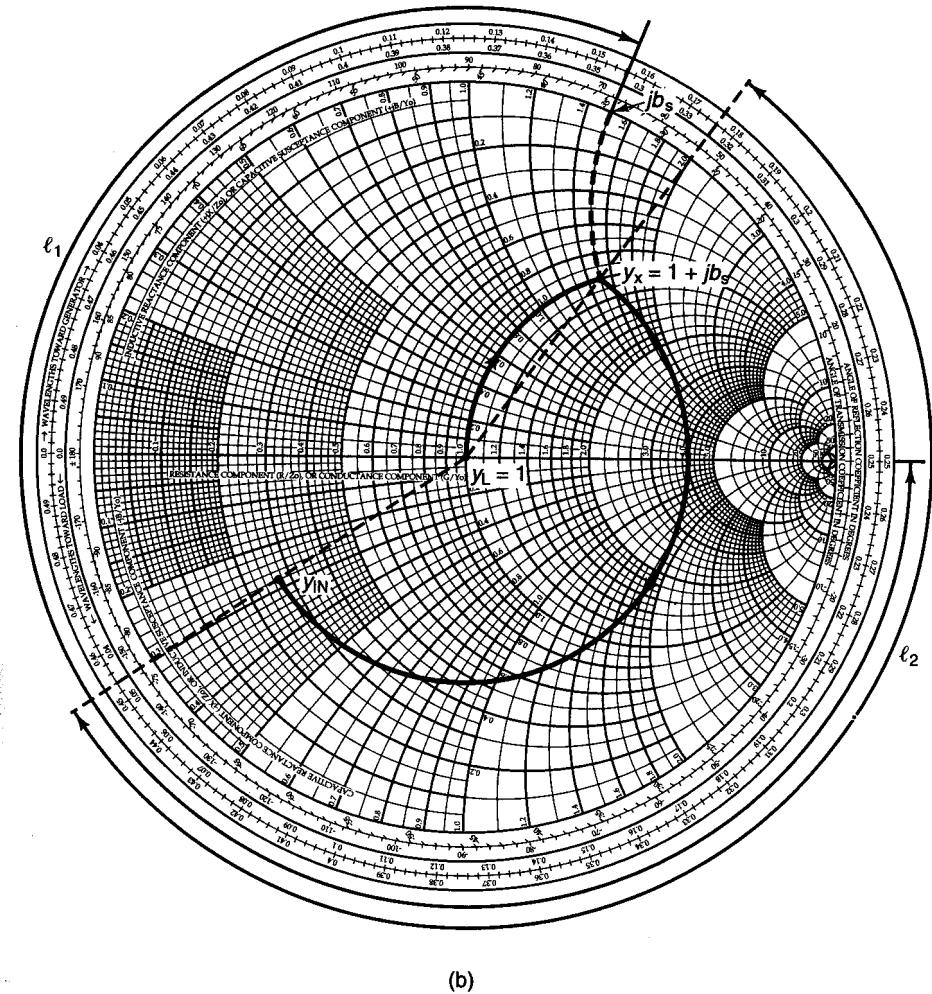
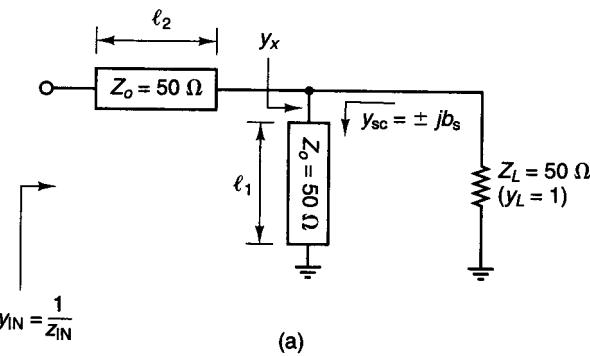


Figure 2.5.7 (a) A microstrip network to transform $y_L = 1$ to y_{IN} ; (b) the design in the Y Smith chart; with $y_{sc} = jb_s$ ($b_s > 0$); (c) the design in the Y Smith chart with $y_{sc} = -jb_s$ ($b_s > 0$).

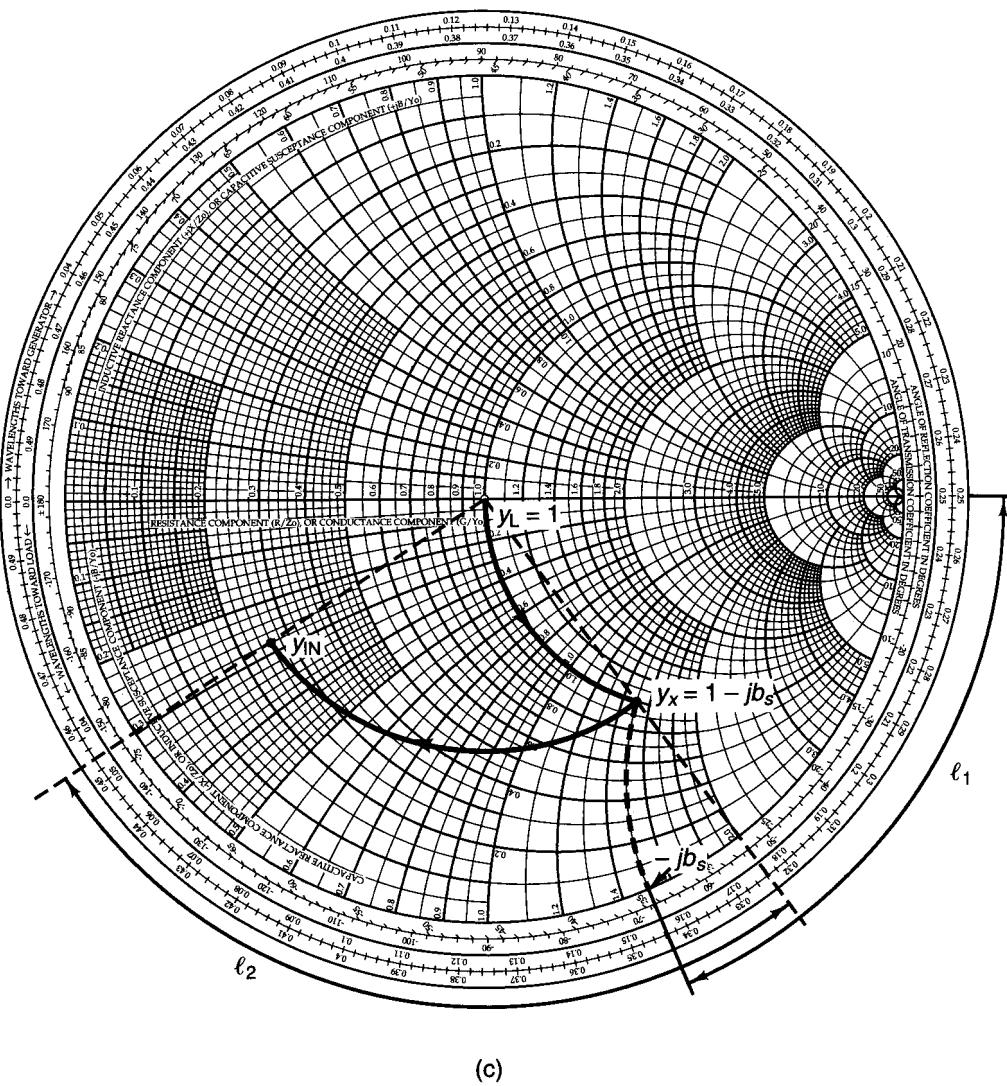


Figure 2.5.7 Continued

As shown in the Y Smith chart in Fig. 2.5.7b, the addition of $y_{sc} = jb_s$ to y_L produces a motion along the unity constant-conductance circle from $y_L = 1$ to $y_x = 1 + jb_s$. The design of the series transmission line of length l_2 is such that y_x is transformed to the admittance y_{IN} (a typical y_{IN} is shown in Fig. 2.5.7b). Consequently, the value of b_s must be selected so that y_x and y_{IN} are on a con-

stant $|\Gamma|$ circle. The readings of l_1 and l_2 in the Y Smith chart are also shown in Fig. 2.5.7b. Observe that l_1 is the length of the short-circuited stub that produces the admittance $y_{sc} = jb_s$. The value of l_1 is easily read from the edge of the Y Smith chart.

The matching design with $y_{sc} = -jb_s$ is illustrated in Fig. 2.5.7c. In this case, the shunt stub of length l_1 produces $y_{sc} = -jb_s$ and the series microstrip line of length l_2 is designed to change y_x to y_{IN} .

If an open-circuited shunt stub is used instead of the short-circuited shunt stub shown in Fig. 2.5.6, the design procedure is quite similar. In fact, the only difference is that the length l_1 in Figs. 2.5.7a and 2.5.7b is read starting from an open-circuited termination (i.e., starting from $y = 0$).

In Fig. 2.5.7a, the shunt stub simply behaves like either a shunt capacitor (see Fig. 2.5.7b with $y_{sc} = jb_s$) or a shunt inductor (see Fig. 2.5.7c with $y_{sc} = -jb_s$), changing $y_L = 1$ to the admittance $y_x = 1 \pm jb_s$. In fact, the same design procedure applies if the shunt stub is replaced by a lumped capacitor or inductor with admittance jb_s or $-jb_s$, respectively.

A practical design using the matching network topology in Fig. 2.5.6 is found in Example 2.5.2, Design 1.

The microstrip matching circuit shown in Fig. 2.5.8a can be designed to match an arbitrary load impedance Z_L to a 50Ω input impedance. This configuration resembles the one in Fig. 2.5.6. While in Fig. 2.5.6 the matching was from $Z_L = 50\Omega$ to an arbitrary Z_{IN} , in Fig. 2.5.8a the matching is from an arbitrary Z_L to $Z_{IN} = 50\Omega$. The design procedure in the Y Smith chart is shown in Fig. 2.5.8b, where the load admittance $y_L = 1/Z_L$ is transformed to $y_{IN} = 1/Z_{IN} = 1$. The length l_1 is selected so that y_L is transformed to the admittance $y_x = 1 + jb_s$ ($b_s > 0$). The motion from y_L to y_x is along a constant $|\Gamma|$ circle. Since $y_x = 1 + jb_s$, we design the short-circuited stub admittance to be $y_{sc} = -jb_s$ so that $y_{IN} = 1$. That is,

$$y_{IN} = y_x + y_{sc} = (1 + jb_s) - jb_s = 1$$

In other words, y_x must be located on the unit conductance circle at the location $y_x = 1 + jb_s$, and $y_{sc} = -jb_s$ cancels the imaginary part of y_x . The length l_2 (measured as shown in Fig. 2.5.8b) produces the short-circuited admittance $y_{sc} = -jb_s$.

The matching design with $y_{sc} = jb_s$ is illustrated in Fig. 2.5.8c. In this case, $y_x = 1 - jb_s$, so $y_{IN} = y_x + y_{sc} = (1 - jb_s) + jb_s = 1$.

Another practical microstrip matching circuit, shown in Fig. 2.5.9a, uses a series quarter-wave line with characteristic impedance Z_{o1} followed by a short-circuited shunt stub of length $\lambda/8$ or $3\lambda/8$ and characteristic impedance Z_{o2} to transform a 50Ω load to any value of input impedance. An open-circuited shunt stub can be used instead of the short-circuited shunt stub.

The design procedure for the matching circuit in Fig. 2.5.9a is as follows. Letting $Y_{IN} = G_{IN} + jB_{IN}$, where $B_{IN} > 0$, the quarter-wave transformer is used

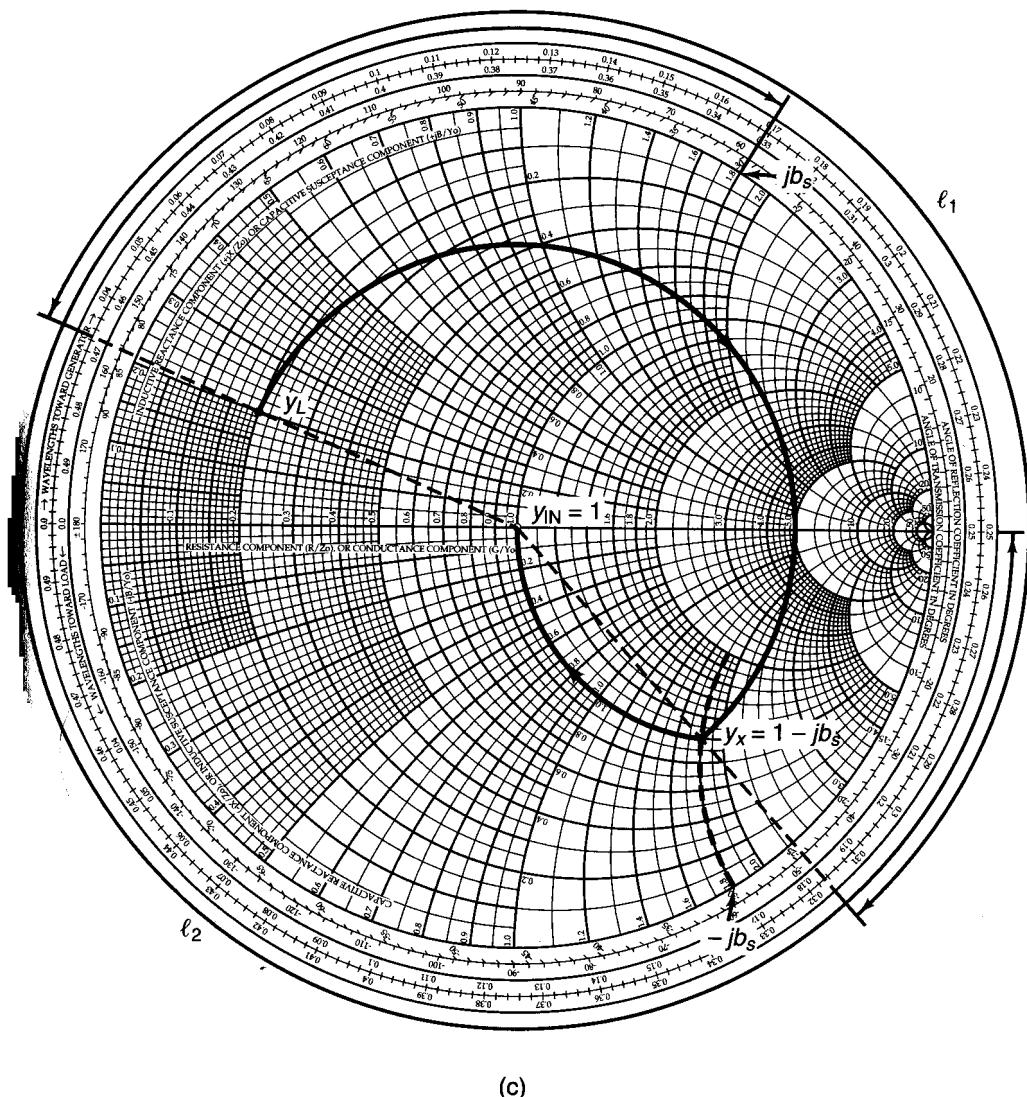
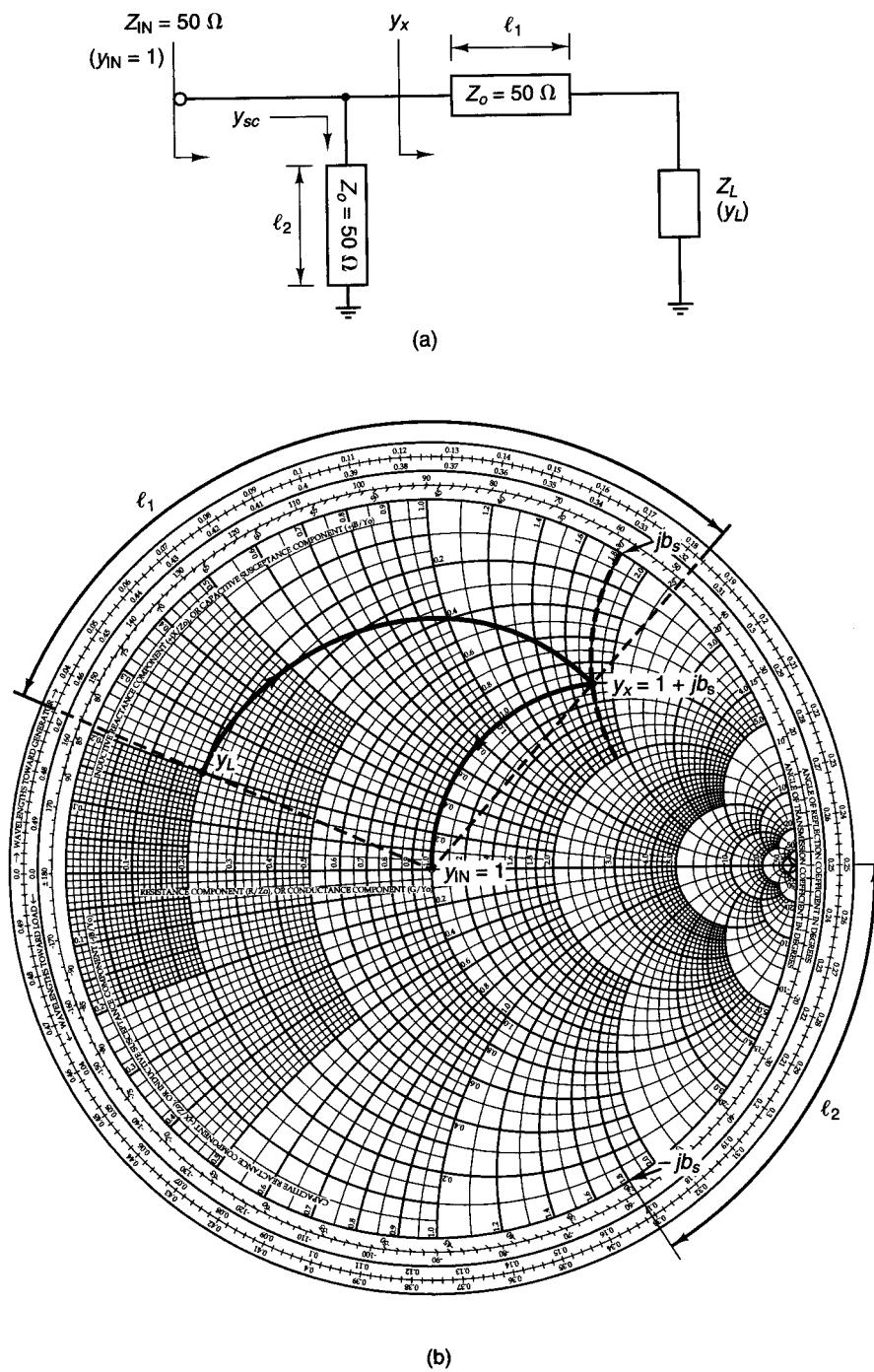


Figure 2.5.8 (a) A microstrip network to transform Z_{LOAD} to Z_{IN} ; (b) the design in the Y Smith chart with $y_{sc} = -jb_s$; (c) the design in the Y Smith chart with $y_{sc} = jb_s$.

to transform $Z_L = 50 \Omega$ to the input resistance $R_{IN} = 1/G_{IN}$. The short-circuited shunt stub is designed to produce the admittance jB_{IN} . Hence, the addition of G_{IN} and jB_{IN} produces the desired Y_{IN} .

The characteristic impedance of the quarter-wave line Z_{o1} is calculated using

$$Z_{o1} = \sqrt{Z_L R_{IN}} = \sqrt{50 R_{IN}} \quad (2.5.15)$$

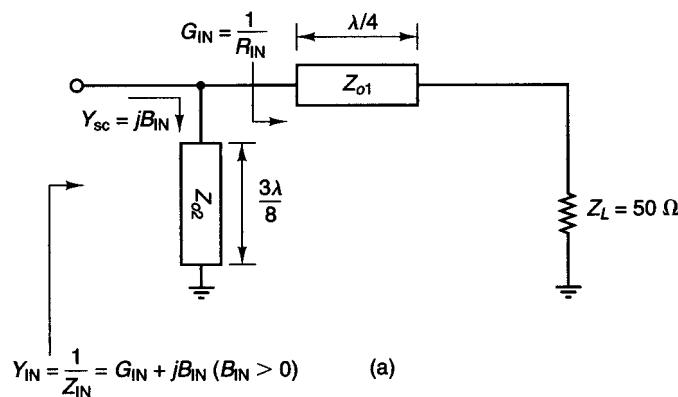


Figure 2.5.9 (a) A microstrip matching network; (b) admittance produced by a short-circuited stub of length $3\lambda/8$; (c) admittance produced by an open-circuited shunt stub of length $\lambda/8$.

Then, since a short-circuited shunt stub of length $3\lambda/8$ produces a shunt admittance of value $Y_{sc} = jY_{o2}$ (see Fig. 2.5.9b), we let jY_{o2} be equal to jB_{IN} . That is,

$$Y_{o2} = B_{IN}$$

or

$$Z_{o2} = \frac{1}{Y_{o2}} = \frac{1}{B_{IN}} \quad (2.5.16)$$

In conclusion, a characteristic impedance for the quarter-wave line Z_{o1} given by (2.5.15) produces the input conductance G_{IN} , and a characteristic impedance for the shunt stub Z_{o2} given by (2.5.16) produces the input susceptance B_{IN} . If an open-circuited shunt stub is used instead of the short-circuited shunt stub, then its length would have been $\lambda/8$ to produce $Y_{oc} = jY_{o2}$ (see Fig. 2.5.9c).

If the input admittance is given by $Y_{IN} = G_{IN} - jB_{IN}$, where $B_{IN} > 0$, as shown in Fig. 2.5.10a, Z_{o1} is calculated using (2.5.15), and a short-circuited shunt stub of length is $\lambda/8$ will produce $Y_{sc} = -jY_{o2}$ (see Fig. 2.5.10b). Therefore, letting $-jY_{o2}$ equal $-jB_{IN}$, we obtain

$$-jY_{o2} = -jB_{IN}$$

or

$$Z_{o2} = \frac{1}{Y_{o2}} = \frac{1}{B_{IN}}$$

If an open-circuited shunt stub is used instead of the short-circuited shunt stub, its length would have been $3\lambda/8$ to produce $Y_{oc} = -jY_{o2}$ (see Fig. 2.5.10c).

A practical design using the matching network topology in Fig. 2.5.9a is found in Example 2.5.2, Design 2.

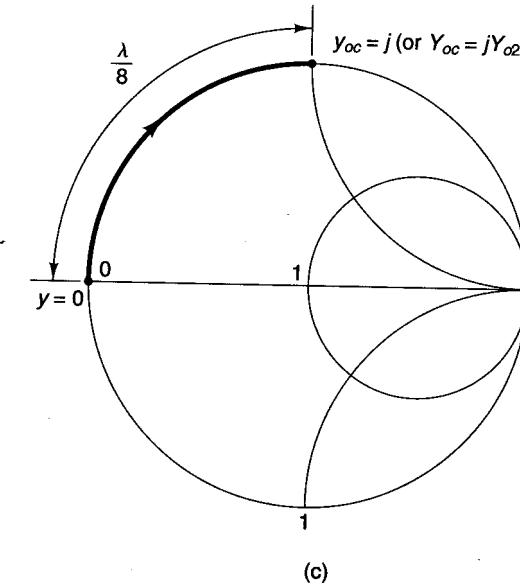
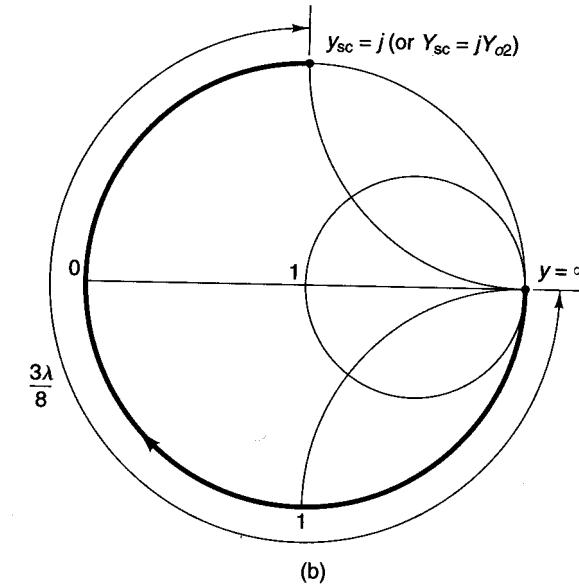


Figure 2.5.9 Continued

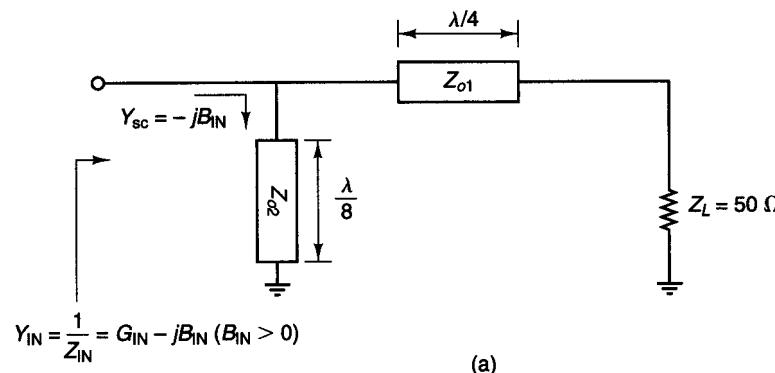


Figure 2.5.10 (a) A microstrip matching network; (b) admittance produced by a short-circuited stub of length $\lambda/8$; (c) admittance produced by an open-circuited shunt stub of length $3\lambda/8$.

The microstrip matching circuits shown in Fig. 2.5.9a and 2.5.10a can also be designed by using any practical value of Z_{o2} but setting the shunt stub length l_2 to produce the admittance of $Y_{sc} = \pm jB_{IN}$. The configuration is shown in Fig. 2.5.11.

The following examples illustrate the uses of microstrip lines in matching networks.

Example 2.5.2

Design two microstrip matching networks for the amplifier shown in Fig. 2.5.12. The normalized impedances and admittances associated with Γ_s and Γ_L can be read, to reasonable accuracy, from the ZY chart—namely,

$$y_s = \frac{1}{z_s} = \frac{1}{0.245 + j0.165} = 2.8 - j1.9$$

and

$$y_L = \frac{1}{z_L} = \frac{1}{0.325 + j0.83} = 0.4 - j1.05$$

In order to design the input matching network, we locate y_s in the Y Smith chart shown in Fig. 2.5.13a. The shortest length of microstrip line plus stub is obtained by using an open-circuited shunt stub of length 0.159λ to move from the origin (i.e., 50Ω) to point A on the Smith chart, and then using a transmission line length of 0.099λ to move from A to y_s .

Next, we locate y_L in Fig. 2.5.13b and follow a similar procedure. In this case, the shortest length of microstrip line plus stub is obtained by using a short-circuited shunt

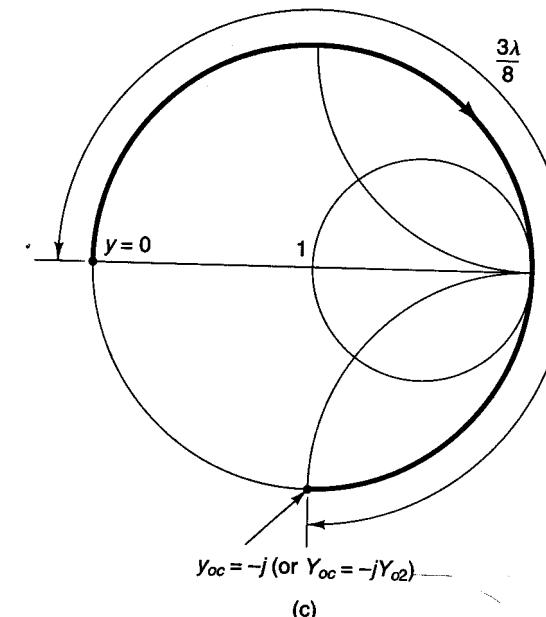
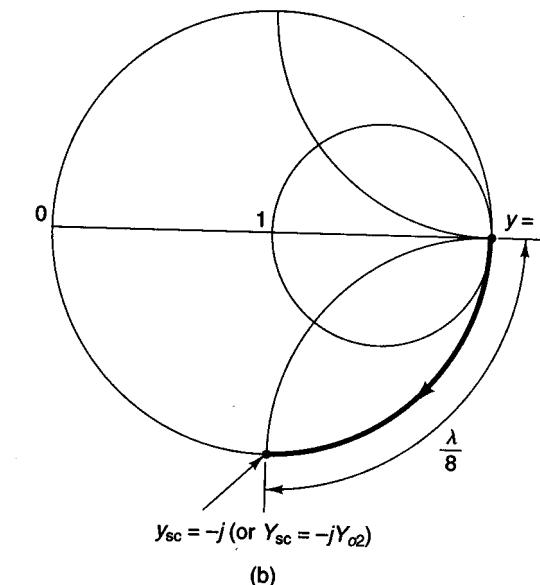


Figure 2.5.10 Continued

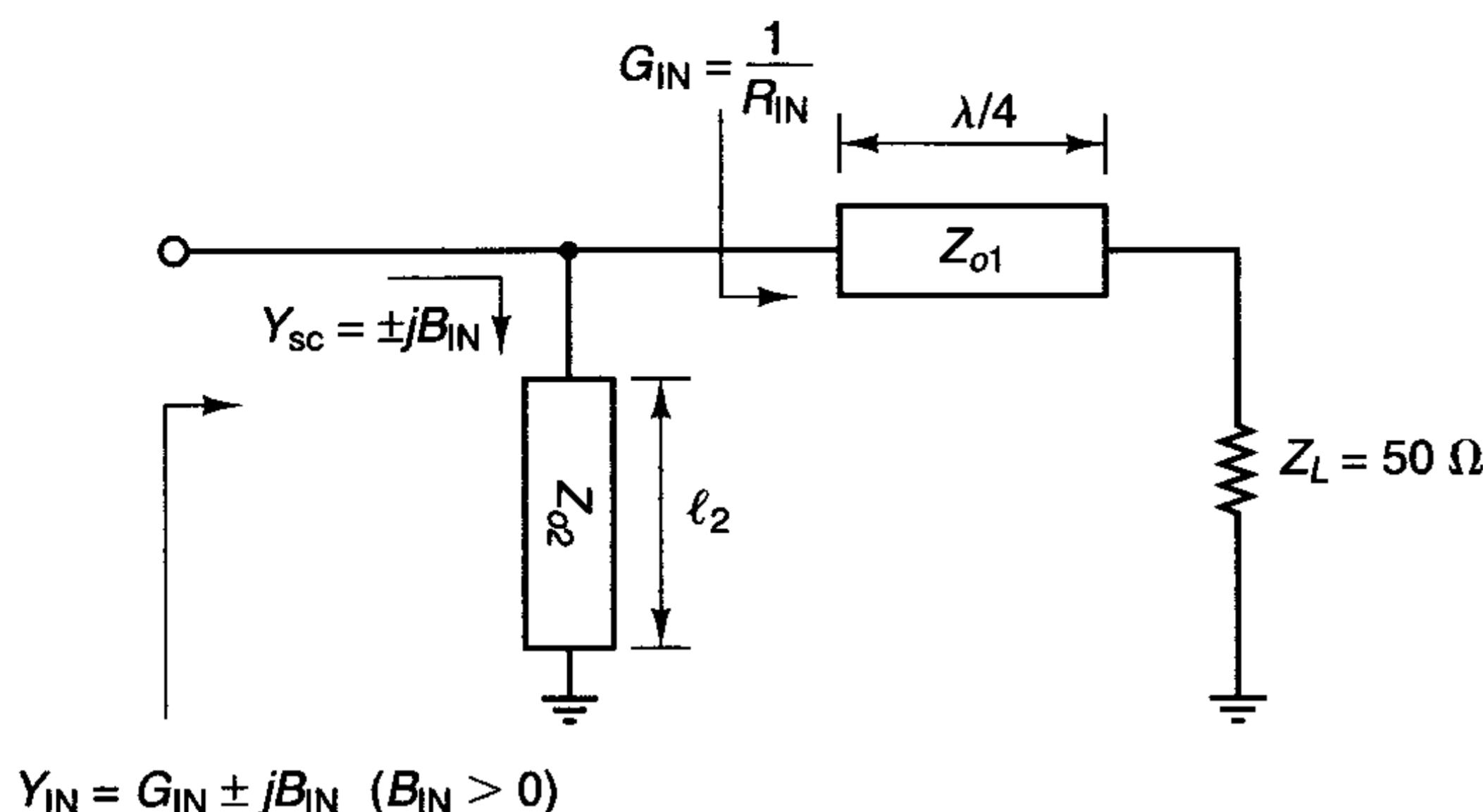


Figure 2.5.11 A microstrip matching circuit.

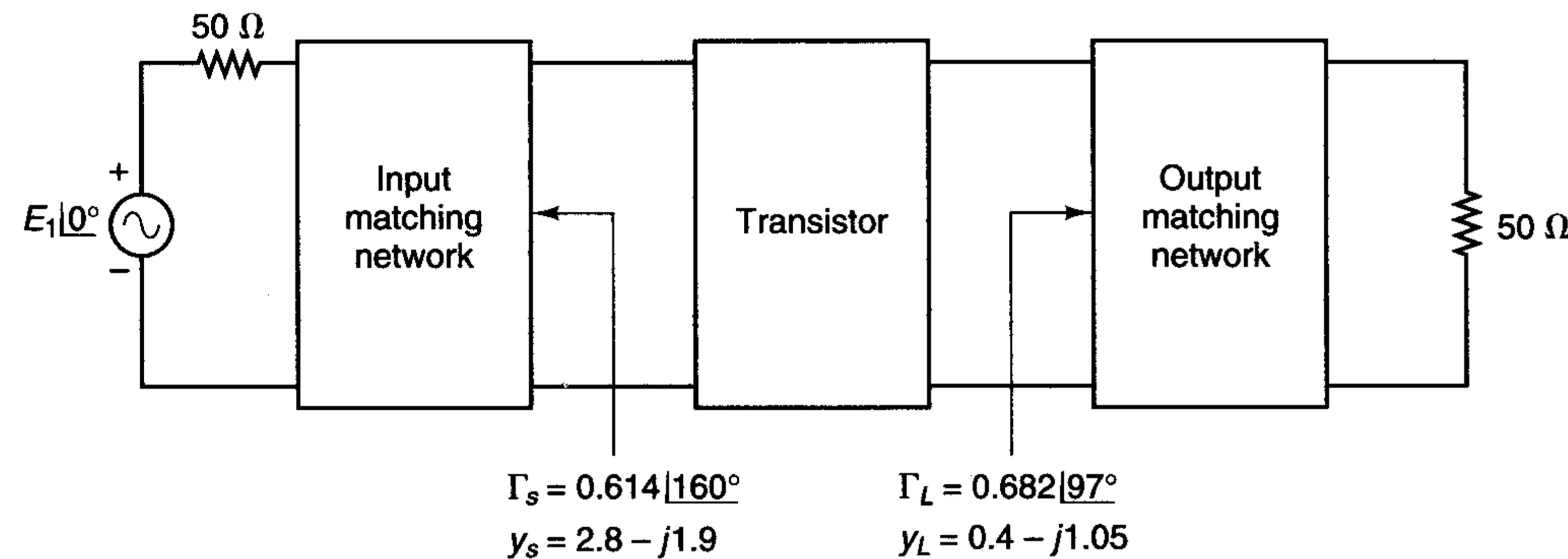


Figure 2.5.12 Amplifier block diagram.

stub of length 0.077λ to move from the origin to point B . Then a series transmission line of length 0.051λ is used to move from B to v_L .

The complete design, showing the transistor, the microstrip matching network, and the dc supply, is shown in Fig. 2.5.14. The characteristic impedance of all microstrip lines is 50Ω .

The capacitors C_A are coupling capacitors. Typical values for the chip capacitors C_A are 200 to 1000 pF, high- Q capacitor. The bypass capacitors C_B (i.e., chip capacitors, 50 to 500 pF) provide the ac short circuits for the 0.077λ and $\lambda/4$ short-circuited stubs. The $\lambda/4$ short-circuited stub, high-impedance line (denoted by $Z_o \gg$), provides the dc path for the base supply voltage. It also presents an open circuit to the ac signal at the base of the transistor. The narrowest practical line (i.e., large Z_o) should be used for the $\lambda/4$ short-circuited stub to avoid unwanted ac coupling. Typical dc bias circuits are shown in Figs. 3.9.2 and 3.9.4.

To minimize transition interaction between the shunt stubs and the series transmission lines, the shunt stubs are usually balanced along the series transmission line. A schematic of the amplifier using balanced shunt stubs is shown in Fig. 2.5.15. The

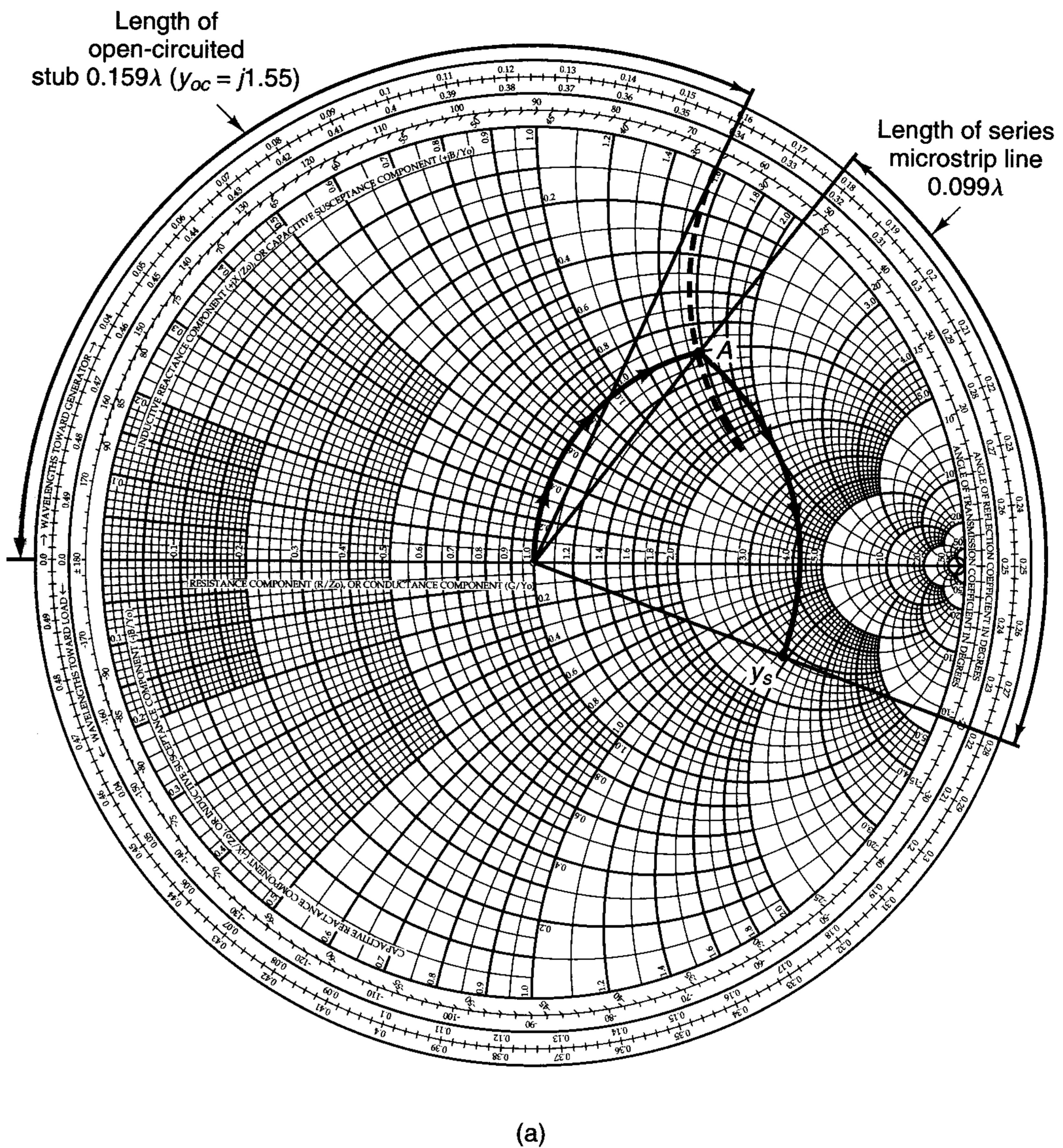


Figure 2.5.13 (a) Input matching network design; (b) output matching network design.

schematic also shows that 50- Ω lines were added on both sides of C_A to provide a soldering area.

In Fig. 2.5.15, two parallel shunt stubs must provide the same admittance as the single stub in Fig. 2.5.14. Therefore, the admittance of each side of the balanced stub must be equal to half of the total admittance. For example, each side of the input balanced shunt stubs must have an admittance of $y = j1.55/2 \equiv j0.775$. Using the Smith

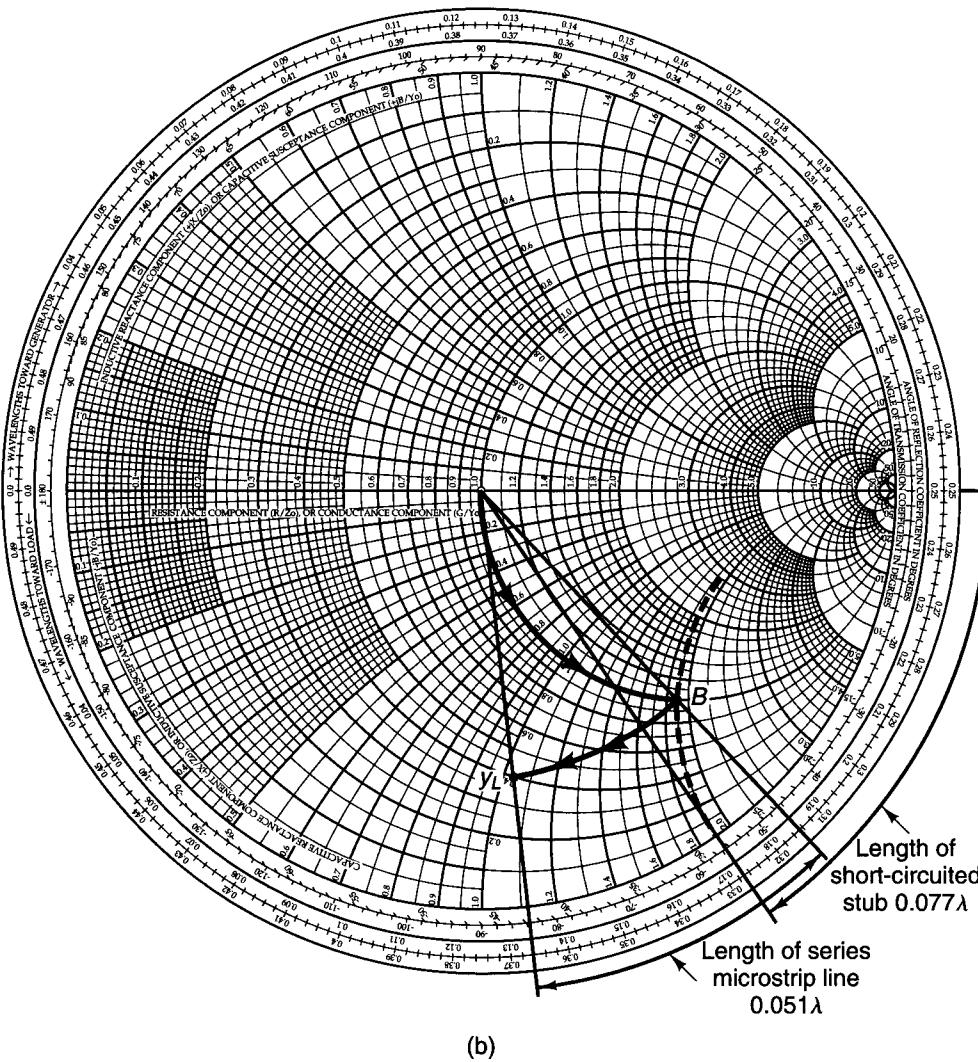
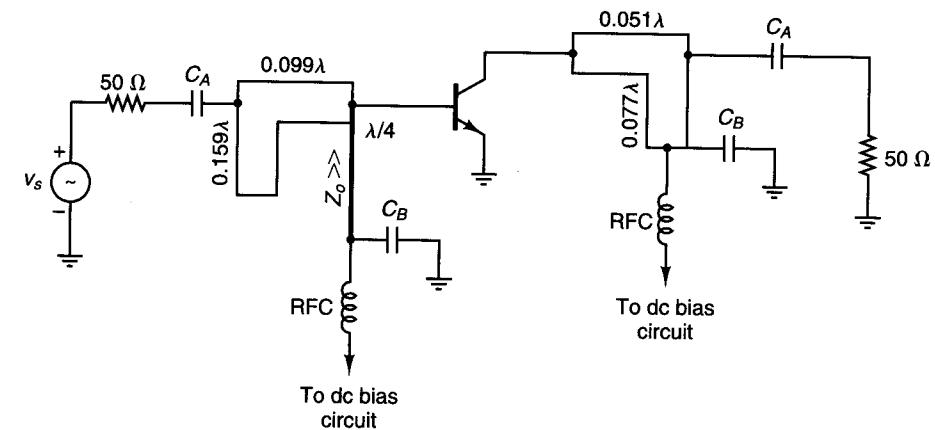
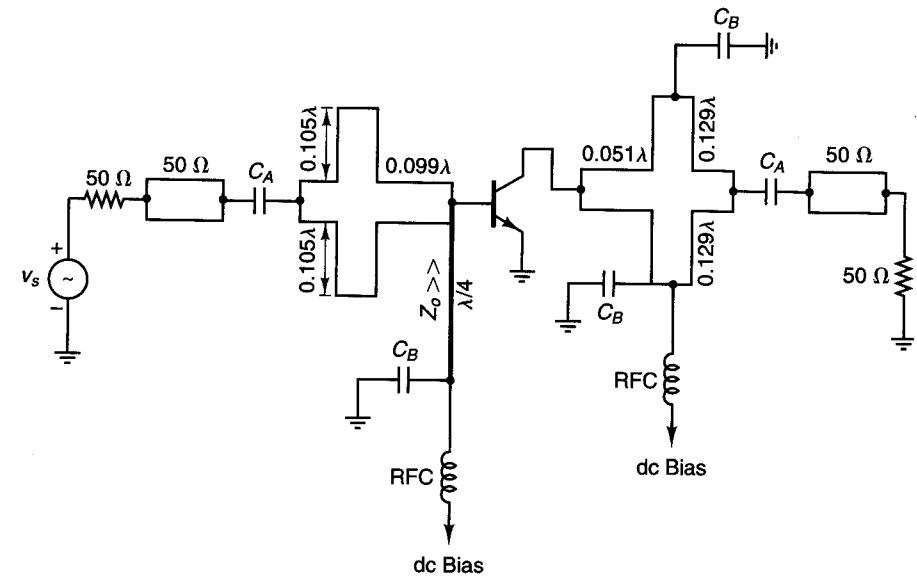


Figure 2.5.13 Continued

chart, we obtain that the length of each side must be 0.105λ . Observe that the length of the shunt stubs in Fig. 2.5.14 is not equal to the total length of the balance stubs in Fig. 2.5.15. Of course, a simple check will show that the admittance seen by the series transmission line is the same in both cases.

If we use RT/Duroid® with $\epsilon_r = 2.23$ and $h = 0.7874$ mm to build the amplifier, we find from (2.5.8) to (2.5.11) (or from Figs. 2.5.2 and 2.5.3) that a characteristic impedance of 50Ω is obtained with $W = 2.42$ mm and $\epsilon_{eff} = 1.91$. The microstrip wave-

Figure 2.5.14 Complete amplifier schematic. The characteristic impedance of the microstrip lines is 50Ω .Figure 2.5.15 Complete amplifier schematic using balanced shunt stubs. The characteristic impedance of the microstrip lines is 50Ω .

length in the 50Ω Duroid microstrip line is $\lambda = \lambda_0/\sqrt{1.91} = 0.7236\lambda_0$, where $\lambda_0 = 30\text{ cm}$ at $f = 1\text{ GHz}$. For a characteristic impedance of 100Ω in the $\lambda/4$ line, the width must be $W = 0.7\text{ mm}$. The line lengths in Fig. 2.5.15 are

$$0.105\lambda = 2.28\text{ cm}$$

$$0.099\lambda = 2.15\text{ cm}$$

$$0.051\lambda = 1.10\text{ cm}$$

$$0.129\lambda = 2.80\text{ cm}$$

$$\lambda/4 = 5.43\text{ cm}$$

Design 2: This method uses microstrip lines with different characteristic impedances, as shown in Fig. 2.5.10a. The design requires the transformation of 50Ω to $Y_s = (2.8 - j1.9)/50 = 0.056 - j0.038\text{ S}$. A quarter-wave transformer can be used to transform the source impedance of 50Ω to the resistance $1/0.056 = 17.86\Omega$. The characteristic impedance of the quarter-wave transformer is

$$Z_{o1} = \sqrt{50(17.86)} = 29.9\Omega$$

An open-circuited shunt stub can be used to obtain the admittance $-j0.038\text{ S}$. Therefore, as shown in Fig. 2.5.10c, an open-circuited shunt stub of length $3\lambda/8$ looks like a shunt inductor having the admittance $-jY_{o2}$. Equating $-jY_{o2}$ to $-j0.038\text{ S}$, we find the characteristic impedance Z_{o2} to be

$$Z_{o2} = \frac{1}{Y_{o2}} = \frac{1}{0.038} = 26.32\Omega$$

If the design is done using a short-circuited shunt stub (see Fig. 2.5.10a), its length would be $\lambda/8$ and $Z_{o2} = 26.3\Omega$.

Similarly, for the output matching network [$Y_L = (0.4 - j1.05)/50 = 0.008 - j0.021\text{ S}$], a quarter-wave line of characteristic impedance

$$Z_{o1} = \sqrt{50(125)} = 79.1\Omega$$

transforms the 50Ω load to a resistance of value $1/0.008 = 125\Omega$. An open-circuited shunt stub of length $3\lambda/8$ and characteristic impedance $Z_{o2} = 1/Y_{o2} = 1/0.021 = 47.6\Omega$ produces the required admittance of $-j0.021\text{ S}$.

The complete amplifier is shown in Fig. 2.5.16a. Figure 2.5.16b shows the amplifier using balanced shunt stubs of length $3\lambda/8$ to minimize the microstrip transition interaction. Observe that in the balance stubs the lengths were kept at $3\lambda/8$, but the characteristic impedance was doubled. For example, in Fig. 2.5.16b each half of the input balance stub must provide the admittance $-j0.038/2$ (since each half must contribute half of the total admittance). Therefore, the value of Z_{o2} for the balanced stubs at the input is $Z_{o2} = 2/0.038 = 52.6\Omega$.

Example 2.5.3

Design a microstrip matching network to transform the load $Z_L = 75 - j60\Omega$ to an input impedance of value $Z_{IN} = 15 + j30\Omega$.

Solution. In this design, let us select a Z_o different from 50Ω —for example, $Z_o = 75\Omega$. With $Z_o = 75\Omega$, the design consists of transforming a normalized load $z_L = Z_L/Z_o = 1 - j0.8$ (or $y_L = 0.61 + j0.49$) to the normalized input impedance $z_{IN} =$

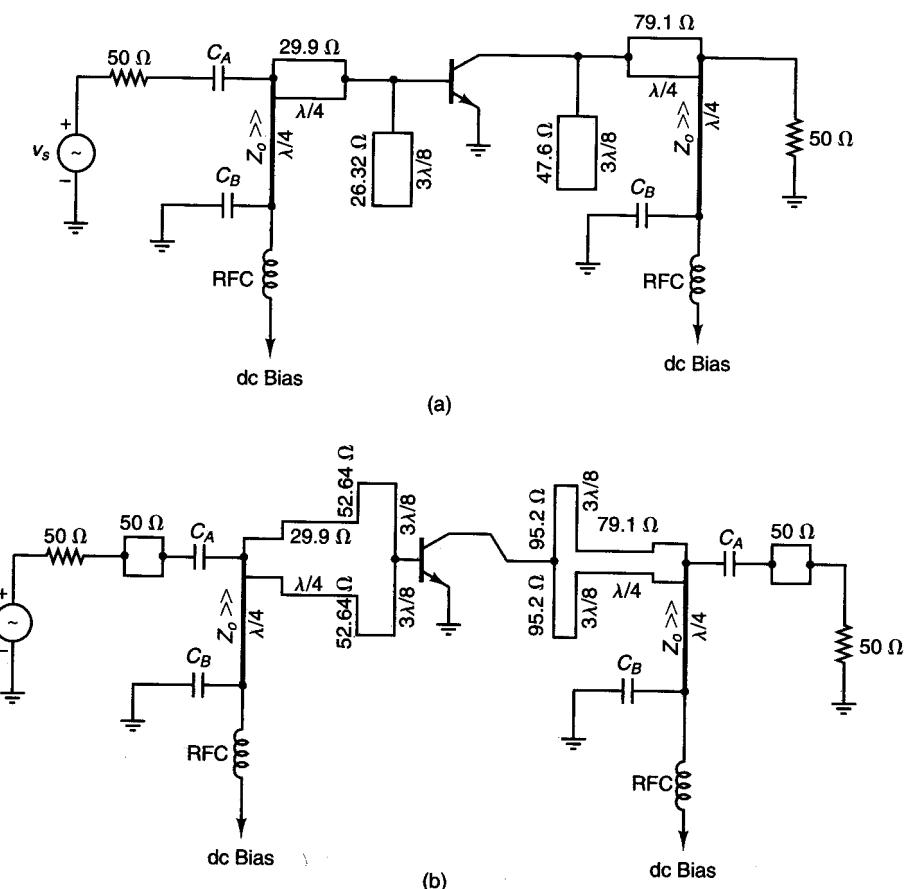


Figure 2.5.16 Matching network design using microstrip lines with different characteristic impedances.

$Z_{IN}/Z_o = 0.2 + j0.4\Omega$ (or $y_{IN} = 1 - j2$). The matching topology selected is shown in Fig. 2.5.17a, and the design in the Y Smith chart is shown in Fig. 2.5.17b. From Fig. 2.5.17b, the shunt admittance required to move from y_L to point A is $j1.5 - j0.49 = j1.01$. The admittance at point A is $y_A = 0.61 + j1.5$. An open-circuited shunt stub of length $l_1 = 0.126\lambda$ provides the admittance of $j1.01$. Then a series transmission line of length $l_2 = 0.313\lambda - 0.164\lambda = 0.149\lambda$ moves the admittance value, along a constant $|\Gamma|$ circle, from that at point A to y_{IN} .

A microstrip matching network which can be easily designed using a Z Smith chart is shown in Fig. 2.5.18a. This matching network uses a $\lambda/4$ line with characteristic impedance Z_{o1} to transform the 50Ω load ($z_L = 1$) to a resistance R_x ($r_x = R_x/50$) that lies on the constant $|\Gamma|$ circle that passes through $z_{IN} = Z_{IN}/50$. The value of Z_{o1} is given by

$$Z_{o1} = \sqrt{50R_x}$$

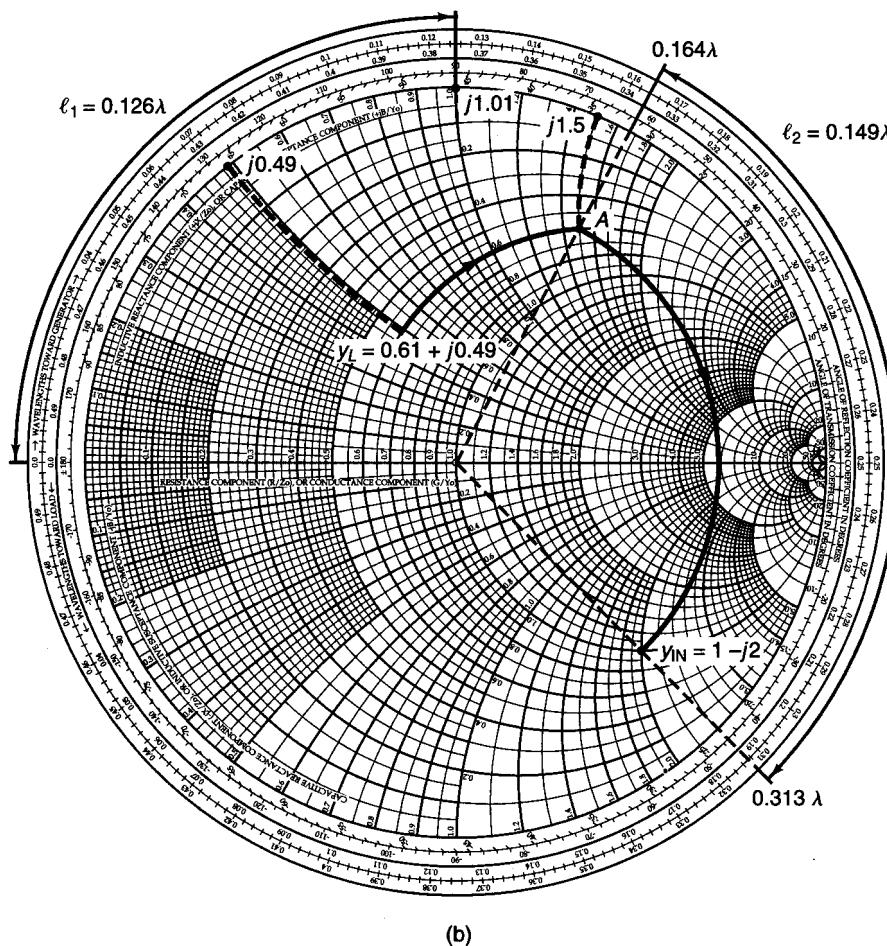
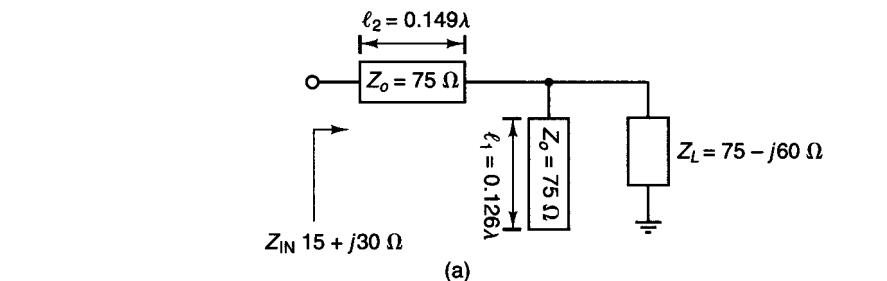


Figure 2.5.17 (a) Matching network for Example 2.5.3; (b) design in the Y Smith chart using $Z_o = 75 \Omega$.

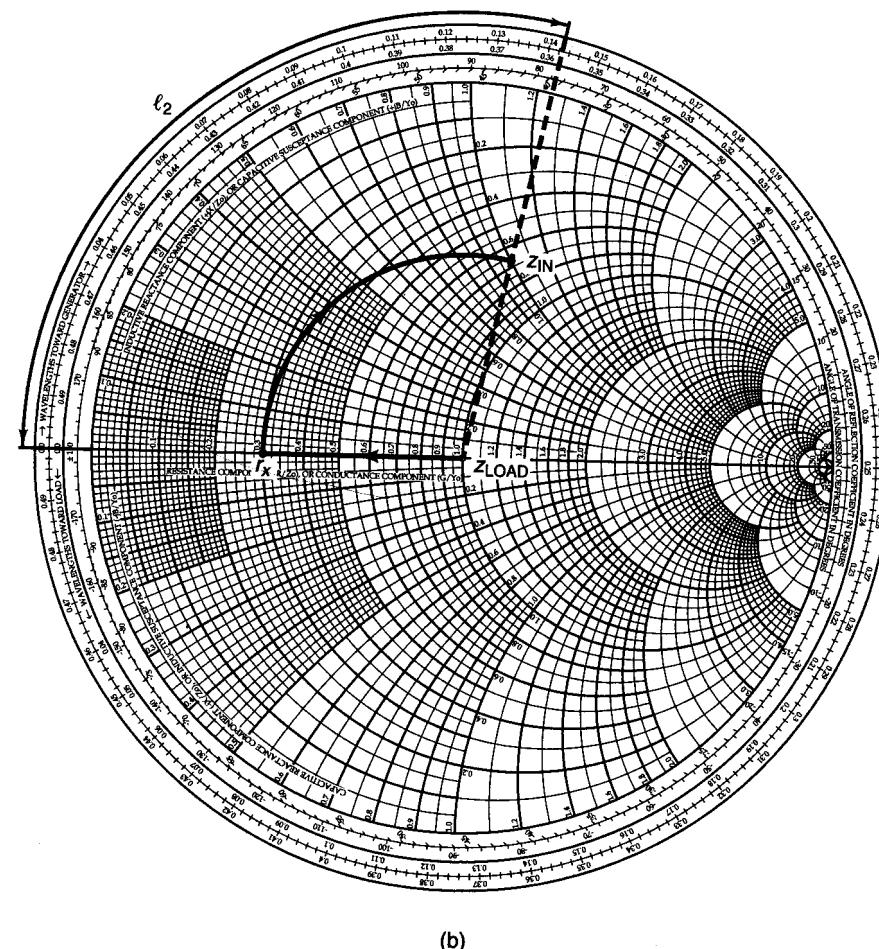
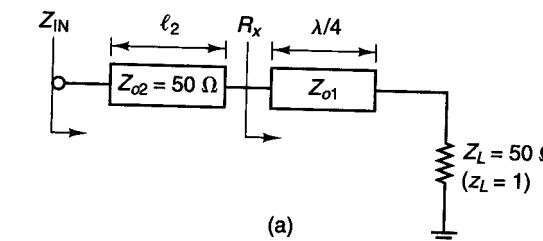


Figure 2.5.18 (a) A microstrip matching circuit; (b) design in the Z Smith chart that results in Z_{o1} smaller than 50Ω ; (c) design in the Z Smith chart that results in Z_{o1} greater than 50Ω .

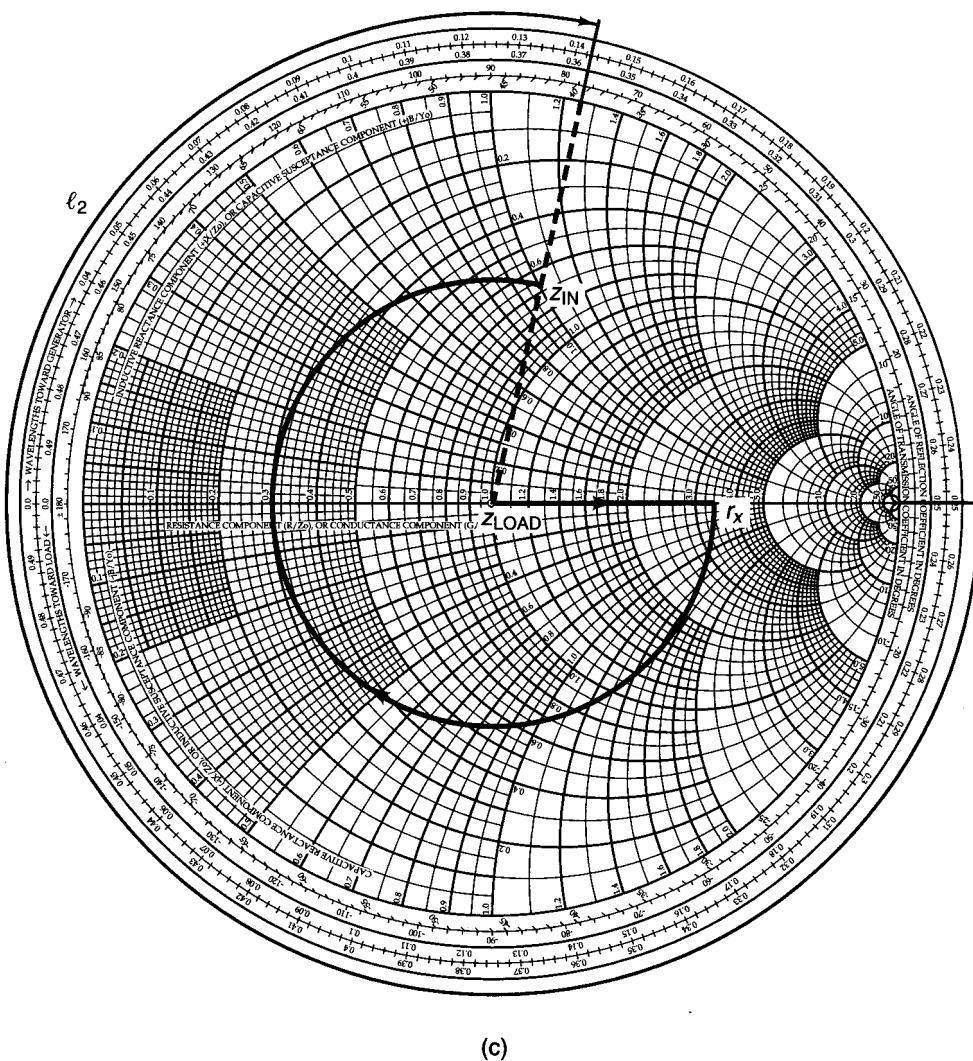


Figure 2.5.18 Continued

Then, the 50Ω line of length l_2 changes the normalized resistance r_x to the input impedance z_{IN} . The design procedure in the Z Smith chart for an arbitrary value of Z_{IN} is shown in Fig. 2.5.18b. An alternate solution is shown in Fig. 2.5.18c. The solution described in Fig. 2.5.18b produces a Z_{o1} smaller than 50Ω , while the solution in Fig. 2.5.18c produces a Z_{o1} greater than 50Ω .

Example 2.5.4

Design the microstrip matching network in Fig. 2.5.18a to transform a 50Ω load to the input impedance $Z_{IN} = 33 + j50\Omega$.

Solution. The location of $z_{IN} = Z_{IN}/50 = 0.66 + j1$ corresponds to the z_{IN} shown in Fig. 2.5.18b. From the constant $|\Gamma|$ circle through z_{IN} , we observe that $r_x = 0.3$, or $R_x = 50(0.3) = 15\Omega$. The $\lambda/4$ line is designed to transform $Z_L = 50\Omega$ to $R_x = 15\Omega$. Hence,

$$Z_{o1} = \sqrt{50(15)} = 27.4\Omega$$

Then, the 50Ω series transmission line of length $l_2 = 0.143\lambda$ produces an input impedance equal to $z_{IN} = 0.66 + j1$, or $Z_{IN} = 50(z_{IN}) = 33 + j50\Omega$.

The design of matching networks containing lumped components and microstrip transmission lines can also be done using the various Smith charts. The following example illustrates one such design.

Example 2.5.5 [cont'd. from previous page]

(a) An oscillator is designed at 2.5 GHz using the output matching topology shown in Fig. 2.5.19a. The length of the microstrips is shown for $\epsilon_{eff} = 1$ (i.e., for $v = c = 3 \times 10^{10}\text{ cm/s}$). The matching network uses a varactor diode as a voltage-variable capacitor for the control of the oscillator frequency. Determine the value of the load reflection coefficient.

(b) Specify the width, height, and length of the microstrip lines if they are constructed using an alumina substrate ($\epsilon_r = 9.6$).

Solution. (a) The wavelength in free space is

$$\lambda_0 = \frac{c}{f} = \frac{3 \times 10^{10}}{2.5 \times 10^9} = 12\text{ cm}$$

Hence, the shunt microstrip of length $l_1 = 43\text{ mm}$ (or $l_1 = 0.358\lambda_0$) has an open-circuited admittance of $y_{oc} = -j1.25$. This shunt microstrip acts like a shunt inductor. Hence, the admittance value of 50Ω in parallel with the shunt microstrip produces the admittance $y_A = 1 - j1.25$, shown as point A in the Y Smith chart in Fig. 2.5.19b.

The series microstrip of length $l_2 = 10\text{ mm}$ (or $l_2 = 0.083\lambda_0$) produces the matching from point A to point B (see Fig. 2.5.19b). This motion is along a constant $|\Gamma|$ circle. The admittance at B is $y_B = 0.41 - j0.53$. Then, the shunt capacitance of the varactor diode (3 pF or $y_C = j2.36$) produces the motion from point B to point C . At point C the admittance is $0.41 + j1.83$, which corresponds to $\Gamma_L = 0.83|-124.5^\circ|$.

In practice, the capacitance of the varactor diode is 3 pF when a specific dc voltage is applied to it. Let us assume that 3 pF occurs when the dc voltage is 4 V . A practical dc bias circuit for the varactor diode and for the transistor is shown in Fig. 2.5.19c.

(b) If the microstrip lines are built using alumina with $\epsilon_r = 9.6$, then from Fig. 2.5.4 a characteristic impedance of 50Ω can be obtained with $W = 24.7\text{ mils}$ and $h = 25\text{ mils}$.

It also follows that $\epsilon_{eff} = 6.46$ (or $\lambda = \lambda_0/\sqrt{\epsilon_{eff}} = 12/\sqrt{6.46} = 4.72\text{ cm}$) in the alumina. Hence, the length of the shunt stub, denoted by l'_1 , is

$$l'_1 = \frac{l_1}{\sqrt{\epsilon_{eff}}} = \frac{43}{\sqrt{6.46}} = 16.9\text{ mm}$$

and that of the series microstrip line, denoted by l'_2 is

$$l'_2 = \frac{l_2}{\sqrt{\epsilon_{eff}}} = \frac{10}{\sqrt{6.46}} = 3.93\text{ mm}$$

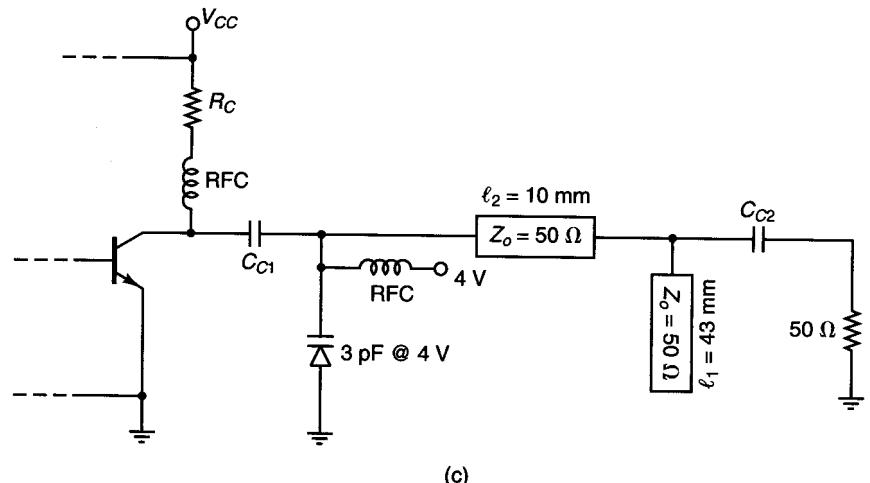
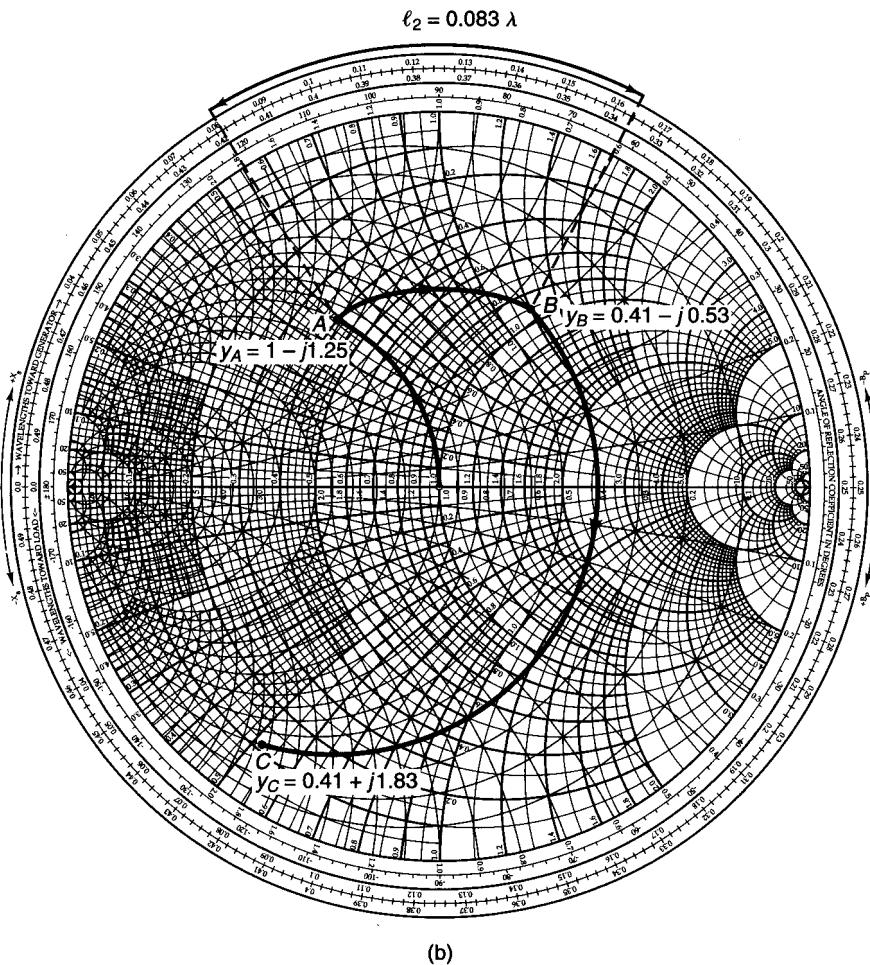
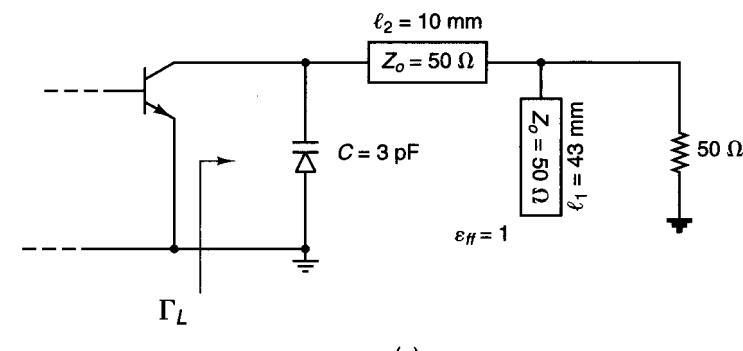


Figure 2.5.19 (a) Circuit schematic for Example 2.5.5; (b) calculation of Γ_L using the Y Smith chart; (c) an implementation of the dc bias circuit.

Example 2.5.6

Design a three-element microstrip matching network to transform a 50Ω termination to a load reflection coefficient given by $\Gamma_L = 0.48 \angle 72^\circ$.

Solution. We will solve this problem using the ZY Smith chart. The solution using this chart should be of interest to the reader (it is recommended that the reader also work this problem using the Y Smith chart). Figure 2.5.20a shows Γ_L in the ZY Smith chart, as well as the path selected for the matching network, using $Z_o = 50 \Omega$. The first element transforms the normalized admittance from point A to point B in Fig. 2.5.20a. The normalized admittance at B is $y_B = 1 - j0.82$. As shown in Fig. 2.5.20b, this can be implemented with a short-circuited shunt stub of length $l_1 = 0.141\lambda$. The second element produces the admittance $y_C = 0.5 - j0.3$ at point C. This element is implemented using a series microstrip line of length $l_2 = 0.099\lambda$. Finally, the third element changes the susceptance along a constant conductance circle of 0.5 from $-j0.3$ to $-j0.6$. This element can be implemented using a short-circuited shunt stub of length $l_3 = 0.203\lambda$ (i.e., having an admittance of $-j0.3$). The matching network is shown in Fig. 2.5.20b.

2.6 SIGNAL FLOW GRAPHS AND APPLICATIONS

A signal flow graph is a convenient technique to represent and analyze the transmission and reflection of waves in a microwave amplifier. Once the signal flow graph is developed, relations between the variables can be obtained using Mason's rule. The flow graph technique permits expressions, such as power gains and voltage gains of complex microwave amplifiers, to be derived easily. Certain rules are followed in constructing a signal flow graph:

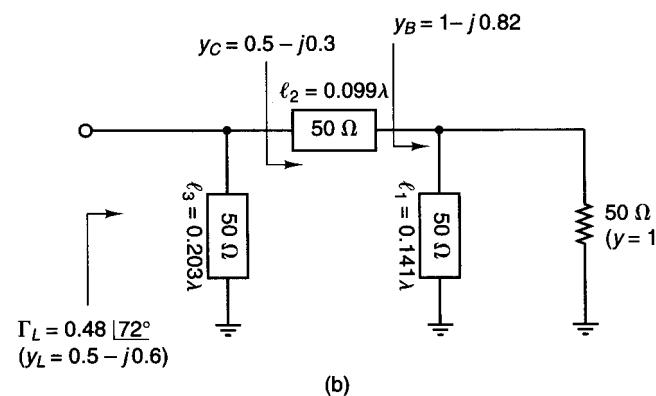
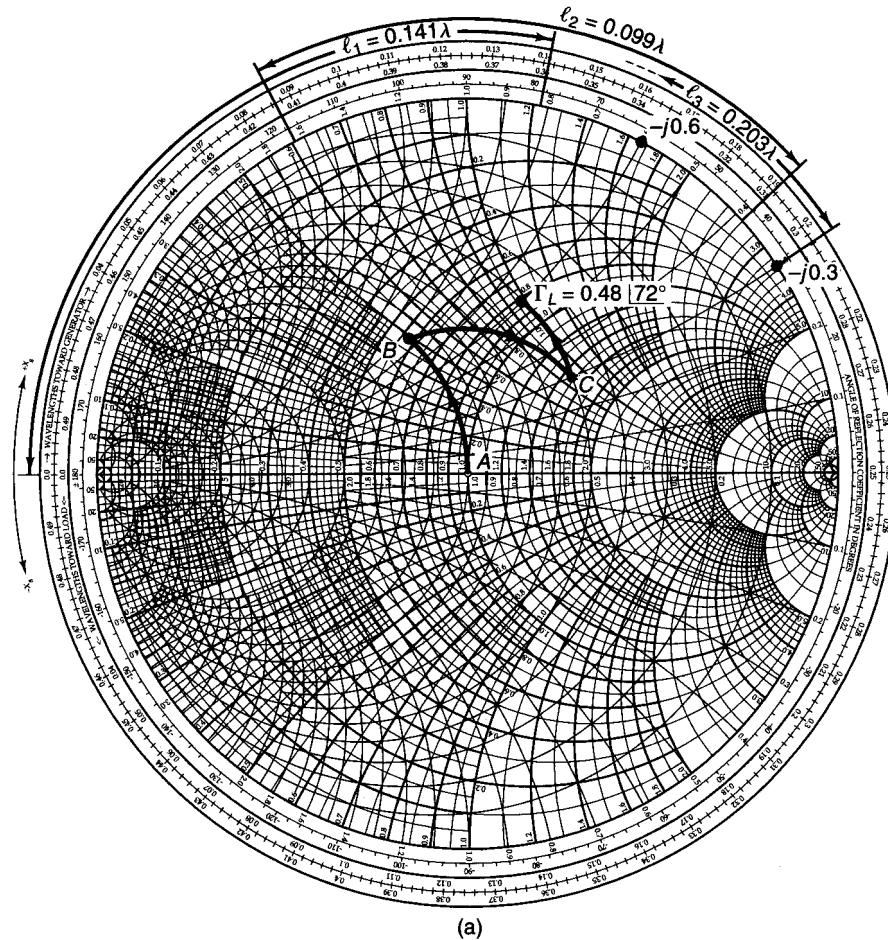


Figure 2.5.20 (a) Design in the ZY Smith chart for Example 2.5.6; (b) the matching network.

1. Each variable is designated as a node.
2. The S parameters and reflection coefficients are represented by branches.
3. Branches enter dependent variable nodes and emanate from independent variable nodes. The independent variable nodes are the incident waves, and the reflected waves are dependent variables nodes.
4. A node is equal to the sum of the branches entering it.

The signal flow graph of the S parameters of a two-port network is shown in Fig. 2.6.1. Observe that b_1 and b_2 are the dependent nodes and a_1 and a_2 the independent nodes. The complete signal flow graph of the two-port network is shown in Fig. 2.6.2.

The signal flow graph in Fig. 2.6.2 shows the relationship between the traveling waves. The incident wave a_1 at port 1 gets partly transmitted (i.e., $S_{21}a_1$) to become part of b_2 , and partly reflected (i.e., $S_{11}a_1$) to become part of b_1 . Similarly, the incident wave a_2 at port 2 gets partly transmitted (i.e., $S_{12}a_2$) to become part of b_1 and partly reflected (i.e., $S_{22}a_2$) to become part of b_2 .

In order to obtain the signal flow graph of a microwave amplifier, we need to obtain the signal flow graph of a signal generator with some internal impedance and the signal flow graph of a load impedance.

Figure 2.6.3a shows a voltage-source generator with impedance Z_s . At the terminals we can write

$$V_g = E_s + I_g Z_s \quad (2.6.1)$$

Using (1.4.1) and (1.4.2), we can express (2.6.1) in terms of traveling waves—namely,

$$V_g^+ + V_g^- = E_s + \left(\frac{V_g^+}{Z_o} - \frac{V_g^-}{Z_o} \right) Z_s$$

Solving for V_g^- , we obtain

$$b_g = b_s + \Gamma_s a_g \quad (2.6.2)$$

where

$$b_g = \frac{V_g^-}{\sqrt{Z_o}}$$

$$a_g = \frac{V_g^+}{\sqrt{Z_o}}$$

$$b_s = \frac{E_s \sqrt{Z_o}}{Z_s + Z_o}$$

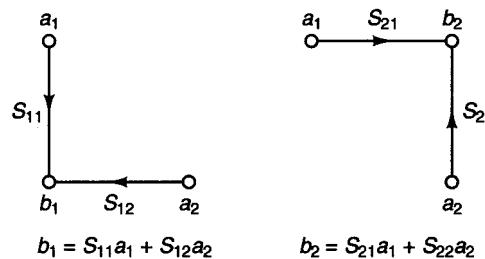


Figure 2.6.1 Signal flow graph for the scattering parameter equations.

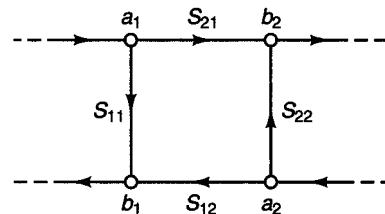


Figure 2.6.2 Signal flow graph of a two-port network.

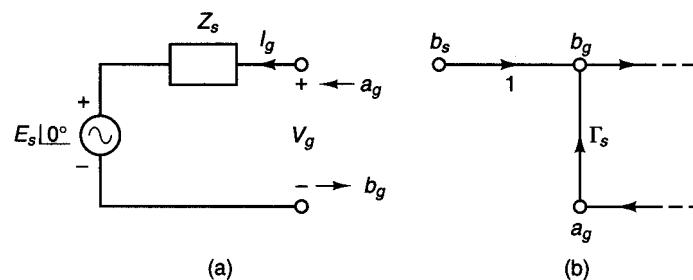


Figure 2.6.3 Signal flow graph of a voltage-source generator.

and

$$\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o}$$

From (2.6.2), the signal flow graph in Fig. 2.6.3b follows.

For the load impedance shown in Fig. 2.6.4a, we can write

$$V_L = Z_L I_L$$

In terms of traveling waves, we obtain

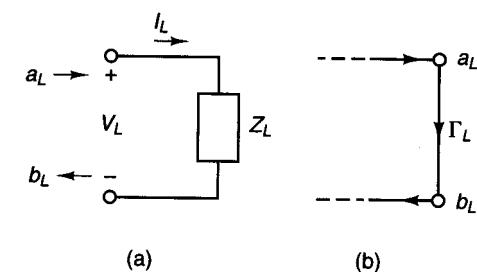


Figure 2.6.4 Signal flow graph of a load impedance.

$$V_L^+ + V_L^- = Z_L \left(\frac{V_L^+}{Z_o} - \frac{V_L^-}{Z_o} \right)$$

or

$$b_L = \Gamma_L a_L \quad (2.6.3)$$

where

$$b_L = \frac{V_L^-}{\sqrt{Z_o}}$$

$$a_L = \frac{V_L^+}{\sqrt{Z_o}}$$

and

$$\Gamma_L = \frac{Z_L - Z_o}{Z_L + Z_o}$$

The signal flow graph follows from (2.6.3) and is shown in Fig. 2.6.4b.

We can now combine the signal flow graph for the two-port network in Fig. 2.6.2 with the signal flow graphs of the signal generator (i.e., Fig. 2.6.3b) and the load (i.e., Fig. 2.6.4b). Observe that the nodes b_g , a_g , b_L , and a_L are identical to a_1 , b_1 , a_2 , and b_2 , respectively. The resulting signal flow graph of a microwave amplifier is shown in Fig. 2.6.5.

To determine the ratio or transfer function T of a dependent to an independent variable, we apply Mason's rule—namely,

$$T = \frac{P_1[1 - \sum L(1)^{(1)} + \sum L(2)^{(1)} - \dots] + P_2[1 - \sum L(1)^{(2)} + \dots] + \dots}{1 - \sum L(1) + \sum L(2) - \sum L(3) + \dots}$$

where the different terms are defined as follows.

The terms P_1 , P_2 , and so on are the different paths connecting the dependent and independent variables whose transfer function T is to be determined. A path is defined as a set of consecutive, codirectional branches along which no node is encountered more than once as we move in the graph from the independent to

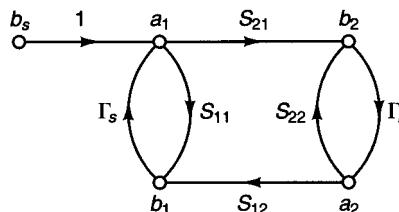


Figure 2.6.5 Signal flow graph of a microwave amplifier.

the dependent node. The value of the path is the product of all branch coefficients along the path. For example, in Fig. 2.6.5, b_s is the only independent variable. To determine the ratio b_1/b_s , we identify two paths, $P_1 = S_{11}$ and $P_2 = S_{21}\Gamma_L S_{12}$.

The term $\Sigma L(1)$ is the sum of all first-order loops. A first-order loop is defined as the product of the branches encountered in a round trip as we move from a node in the direction of the arrows back to that original node. In Fig. 2.6.5, $S_{11}\Gamma_s$, $S_{21}\Gamma_L S_{12}\Gamma_s$, and $S_{22}\Gamma_L$ are first-order loops.

The term $\Sigma L(2)$ is the sum of all second-order loops. A second-order loop is defined as the product of any two nontouching first-order loops. In Fig. 2.6.5, $S_{11}\Gamma_s$ and $S_{22}\Gamma_L$ do not touch; therefore, the product $S_{11}\Gamma_s S_{22}\Gamma_L$ is a second-order loop.

The term $\Sigma L(3)$ is the sum of all third-order loops. A third-order loop is defined as the product of three nontouching first-order loops. In Fig. 2.6.5, there are no third-order loops. Of course, the terms $\Sigma L(4)$, $\Sigma L(5)$, and so on represent fourth-, fifth-, and higher-order loops.

The terms $\Sigma L(1)^{(P)}$ is the sum of all first-order loops that do not touch the path P between the independent and dependent variables. In Fig. 2.6.5, for the path $P_1 = S_{11}$ we find that $\Sigma L(1)^{(1)} = \Gamma_L S_{22}$, and for the path $P_2 = S_{21}\Gamma_L S_{12}$ we find that $\Sigma L(1)^{(2)} = 0$.

The term $\Sigma L(2)^{(P)}$ is the sum of all second-order loops that do not touch the path P between the independent and dependent variables. In Fig. 2.6.5, we find that $\Sigma L(2)^{(P)} = 0$. Of course, $\Sigma L(3)^{(P)}$, $\Sigma L(4)^{(P)}$, and so on represent higher-order loops that do not touch the path P .

For the transfer function b_1/b_s in Fig. 2.6.5, we have found that $P_1 = S_{11}$, $P_2 = S_{21}\Gamma_L S_{12}$, $\Sigma L(1) = S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s$, $\Sigma L(2) = S_{11}\Gamma_s S_{22}\Gamma_L$, and $\Sigma L(1)^{(1)} = \Gamma_L S_{22}$. Therefore, using Mason's rule, we obtain

$$\frac{b_1}{b_s} = \frac{S_{11}(1 - \Gamma_L S_{22}) + S_{21}\Gamma_L S_{12}}{1 - (S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s) + S_{11}\Gamma_s S_{22}\Gamma_L}$$

Applications of Signal Flow Graphs

The first application of signal flow graph analysis is in the calculation of the input reflection coefficient, called Γ_{IN} , when a load is connected to the output of a two-port network. The signal flow graph is shown in Fig. 2.6.6.

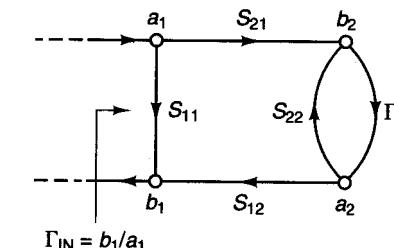


Figure 2.6.6 Signal flow graph for the input reflection coefficient Γ_{IN} .

The input reflection coefficient Γ_{IN} is defined as

$$\Gamma_{IN} = \frac{b_1}{a_1}$$

Observing that $P_1 = S_{11}$, $P_2 = S_{21}\Gamma_L S_{12}$, $\Sigma L(1) = S_{22}\Gamma_L$, and $\Sigma L(1)^{(1)} = S_{22}\Gamma_L$, we can use Mason's rule to obtain

$$\begin{aligned} \Gamma_{IN} &= \frac{S_{11}(1 - S_{22}\Gamma_L) + S_{21}\Gamma_L S_{12}}{1 - S_{22}\Gamma_L} \\ &= S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \end{aligned} \quad (2.6.4)$$

If $\Gamma_L = 0$, it follows from (2.6.4) that $\Gamma_{IN} = S_{11}$. Also, when there is no transmission from the output to the input (i.e., when $S_{12} = 0$), it follows that $\Gamma_{IN} = S_{11}$. When $S_{12} = 0$, we call the device represented by the two-port a unilateral device.

Similarly, we can calculate the output reflection coefficient $\Gamma_{OUT} = b_2/a_2$ with $b_s = 0$ from the signal flow graph shown in Fig. 2.6.7. The expression for Γ_{OUT} is

$$\Gamma_{OUT} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \quad (2.6.5)$$

Next, we use signal flow graphs to calculate power gain and voltage gain. The square of the magnitude of the incident and reflected waves represents

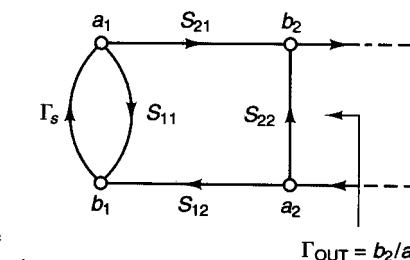


Figure 2.6.7 Signal flow graph for the output reflection coefficient Γ_{OUT} .

$$\Gamma_{OUT} = b_2/a_2|_{b_s=0}$$

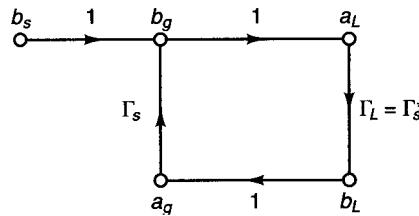


Figure 2.6.8 Signal flow graph of a voltage source connected to a conjugate matched load.

power. Therefore, the power delivered to the load in Fig. 2.6.5 is given by the difference between the incident and reflected power—namely,

$$P_L = \frac{1}{2} |b_2|^2 - \frac{1}{2} |a_2|^2 = \frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2) \quad (2.6.6)$$

The power available from a source is defined as the power delivered by the source to a conjugately matched load. Figure 2.6.8 shows the signal flow graph of a source connected to a conjugate match load (i.e., $\Gamma_L = \Gamma_s^*$). Therefore, the power available from the source, in Fig. 2.6.8, is given by

$$P_{AVS} = \frac{1}{2} |b_g|^2 - \frac{1}{2} |a_g|^2 \quad (2.6.7)$$

Observing that $b_g = b_s + b_s \Gamma_s \Gamma_s^*$ and $a_g = b_g \Gamma_s^*$, we obtain

$$b_g = \frac{b_s}{1 - |\Gamma_s|^2} \quad (2.6.8)$$

and

$$a_g = \frac{b_s \Gamma_s^*}{1 - |\Gamma_s|^2} \quad (2.6.9)$$

Substituting (2.6.8) and (2.6.9) into (2.6.7) gives

$$P_{AVS} = \frac{\frac{1}{2} |b_s|^2}{1 - |\Gamma_s|^2} \quad (2.6.10)$$

The previous results could have also been obtained as follows. Observe that the power delivered to the load Γ_L in Fig. 2.6.8 is given by

$$P_L = |a_L|^2 (1 - |\Gamma_L|^2) = \frac{\frac{1}{2} |b_s|^2 (1 - |\Gamma_L|^2)}{|1 - \Gamma_s \Gamma_L|^2}$$

Therefore, with $\Gamma_L = \Gamma_s^*$ the power delivered to the load is equal to the available power from the source, and (2.6.10) follows.

The transducer power gain, called G_T , is defined as the ratio of the power delivered to a load to the power available from the source. From (2.6.6) and (2.6.10), we obtain

$$G_T = \frac{P_L}{P_{AVS}} = \frac{|b_2|^2}{|b_s|^2} (1 - |\Gamma_L|^2) (1 - |\Gamma_s|^2) \quad (2.6.11)$$

The ratio b_2/b_s can be obtained using Mason's rule—namely,

$$\begin{aligned} \frac{b_2}{b_s} &= \frac{S_{21}}{1 - (S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s) + S_{11}\Gamma_s S_{22}\Gamma_L} \\ &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s} \end{aligned} \quad (2.6.12)$$

Substituting (2.6.12) into (2.6.11) results in

$$G_T = \frac{|S_{21}|^2 (1 - |\Gamma_s|^2) (1 - |\Gamma_L|^2)}{|(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s|^2} \quad (2.6.13)$$

The denominator of (2.6.13) can be further manipulated, and G_T can be expressed in the form

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_{IN}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.6.14)$$

or

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.6.15)$$

where Γ_{IN} and Γ_{OUT} are given by (2.6.4) and (2.6.5), respectively.

The power gain G_P is defined as the ratio of the power delivered to the load P_L to the input power to the network P_{IN} . P_L is given by (2.6.6), and the input power is given by

$$P_{IN} = \frac{1}{2} |a_1|^2 - \frac{1}{2} |b_1|^2 = \frac{1}{2} |a_1|^2 (1 - |\Gamma_{IN}|^2)$$

Hence, we can express the power gain in the form

$$G_P = \frac{P_L}{P_{IN}} = \frac{|b_2|^2 (1 - |\Gamma_L|^2)}{|a_1|^2 (1 - |\Gamma_{IN}|^2)}$$

Dividing the numerator and denominator by $|b_s|^2$ gives

$$G_P = \frac{P_L}{P_{IN}} = \frac{\frac{|b_2|^2}{|b_s|^2} (1 - |\Gamma_L|^2)}{\frac{|a_1|^2}{|b_s|^2} (1 - |\Gamma_{IN}|^2)} \quad (2.6.16)$$

The ratio b_2/b_s is given by (2.6.12), and using Mason's rule, the ratio a_1/b_s is

$$\frac{a_1}{b_s} = \frac{1 - S_{22}\Gamma_L}{1 - (S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s) + S_{11}\Gamma_s S_{22}\Gamma_L} \quad (2.6.17)$$

Substituting (2.6.12) and (2.6.17) into (2.6.16), we can express G_p in the form

$$G_p = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (2.6.18)$$

The available power gain G_A is defined as the ratio of the power available from the network P_{AVN} to the power available from the source P_{AVS} . The power available from the network is the power delivered by the network to a conjugate matched load. That is,

$$\begin{aligned} P_{AVN} &= P_L \Big|_{\Gamma_L=\Gamma_{OUT}^*} = \left[\frac{1}{2} |b_2|^2 - \frac{1}{2} |a_2|^2 \right] \Big|_{\Gamma_L=\Gamma_{OUT}^*} \\ &= \left[\frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2) \right] \Big|_{\Gamma_L=\Gamma_{OUT}^*} = \frac{1}{2} |b_2|^2 (1 - |\Gamma_{OUT}|^2) \end{aligned} \quad (2.6.19)$$

Therefore, from (2.6.19) and (2.6.10), G_A is given by

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{|b_2|^2}{|b_s|^2} (1 - |\Gamma_{OUT}|^2) (1 - |\Gamma_s|^2) \quad (2.6.20)$$

From (2.6.12), the ratio b_2/b_s with $\Gamma_L = \Gamma_{OUT}^*$ can be expressed as

$$\begin{aligned} \frac{b_2}{b_s} &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s} \\ &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - \Gamma_{OUT}\Gamma_L)} \Big|_{\Gamma_L=\Gamma_{OUT}^*} \\ &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - |\Gamma_{OUT}|^2)} \end{aligned} \quad (2.6.21)$$

Substituting (2.6.21) into (2.6.20) results in the expression

$$G_A = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{OUT}|^2} \quad (2.6.22)$$

The voltage gain of the amplifier is defined as the ratio of the output voltage to the input voltage. That is,

$$A_v = \frac{a_2 + b_2}{a_1 + b_1}$$

Dividing by b_s gives

$$A_v = \frac{a_2/b_s + b_2/b_s}{a_1/b_s + b_1/b_s}$$

Therefore, we need to calculate the ratios a_2/b_s , b_2/b_s , a_1/b_s , and b_1/b_s using Mason's rule. The expression for A_v can be shown to be

$$A_v = \frac{S_{21}(1 + \Gamma_L)}{(1 - S_{22}\Gamma_L) + S_{11}(1 - S_{22}\Gamma_L) + S_{21}\Gamma_L S_{12}} \quad (2.6.23)$$

2.7 POWER-GAIN EXPRESSIONS: ALTERNATE DERIVATIONS

In the previous section, the power-gain expressions were derived using signal flow graph theory. In this section, the power-gain expressions are derived using a direct manipulation of the S -parameters relations.

For the microwave amplifier shown in Fig. 2.7.1, the source and load reflection coefficients in a Z_o system are

$$\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o} \quad (2.7.1)$$

and

$$\Gamma_L = \frac{Z_L - Z_o}{Z_L + Z_o} \quad (2.7.2)$$

For the transistor, the input and output traveling waves measured in a Z_o system are related by

$$b_1 = S_{11}a_1 + S_{12}a_2 \quad (2.7.3)$$

and

$$b_2 = S_{21}a_1 + S_{22}a_2 \quad (2.7.4)$$

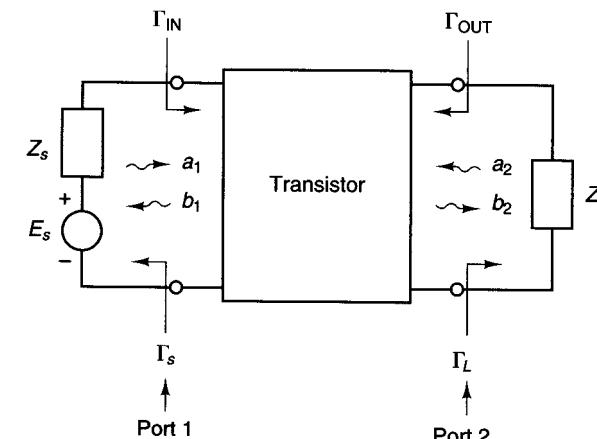


Figure 2.7.1 Block diagram of a microwave amplifier.

One should observe that the concepts of a reflection coefficient and traveling waves can be used even if there are no transmission lines at port 1 and port 2 in Fig. 2.7.1. The reflection coefficients Γ_s and Γ_L and the traveling waves a_1 , b_1 , a_2 , and b_2 are the reflection coefficients and traveling waves that would exist if transmission lines with characteristic impedances Z_o were inserted at ports 1 and 2. Alternatively, we could think of the reflection coefficients and the traveling waves as those in a transmission line of zero length and characteristic impedance Z_o connected at ports 1 and 2.

The input reflection coefficient Γ_{IN} is

$$\Gamma_{IN} = \frac{b_1}{a_1}$$

The ratio b_1/a_1 can be evaluated as follows. Since

$$a_2 = \Gamma_L b_2 \quad (2.7.5)$$

substituting (2.7.5) into (2.7.4) gives

$$b_2 = S_{21}a_1 + S_{22}\Gamma_L b_2$$

or

$$b_2 = \frac{S_{21}a_1}{1 - S_{22}\Gamma_L} \quad (2.7.6)$$

Then, substituting (2.7.5) and (2.7.6) into (2.7.3) gives

$$b_1 = S_{11}a_1 + S_{12}\Gamma_L b_2 = S_{11}a_1 + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} a_1$$

or

$$\Gamma_{IN} = \frac{b_1}{a_1} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}$$

The output reflection coefficient is defined as

$$\Gamma_{OUT} = \left. \frac{b_2}{a_2} \right|_{E_s=0}$$

With $E_s = 0$, it follows that

$$a_1 = \Gamma_s b_1 \quad (2.7.7)$$

Substituting (2.7.7) into (2.7.3) gives

$$b_1 = S_{11}\Gamma_s b_1 + S_{12}a_2$$

or

$$b_1 = \frac{S_{12}a_2}{1 - S_{11}\Gamma_s} \quad (2.7.8)$$

Then, substituting (2.7.7) and (2.7.8) into (2.7.4) gives

$$b_2 = S_{21}\Gamma_s b_1 + S_{22}a_2 = \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} a_2 + S_{22}a_2$$

Therefore, Γ_{OUT} is given by

$$\Gamma_{OUT} = \left. \frac{b_2}{a_2} \right|_{E_s=0} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s}$$

The power delivered to the input port of the transistor is

$$P_{IN} = \frac{1}{2} |a_1|^2 - \frac{1}{2} |b_1|^2 = \frac{1}{2} |a_1|^2(1 - |\Gamma_{IN}|^2) \quad (2.7.9)$$

At the input port (see Fig. 2.7.2),

$$V_1 = E_s + I_1 Z_s \quad (2.7.10)$$

Therefore, in terms of traveling waves, (2.7.10) is written as

$$a_1 = b_s + \Gamma_s b_1 \quad (2.7.11)$$

where

$$a_1 = \frac{V_1^-}{\sqrt{Z_o}}$$

$$b_1 = \frac{V_1^+}{\sqrt{Z_o}}$$

$$b_s = \frac{E_s \sqrt{Z_o}}{Z_s + Z_o}$$

and

$$\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o}$$

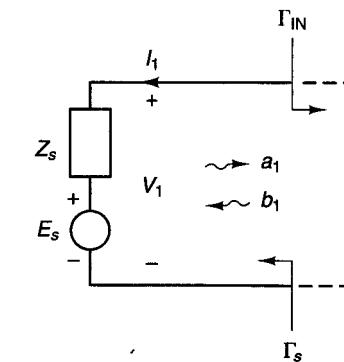


Fig. 2.7.2 The input port.

Except for a nonnomenclatural difference, (2.7.11) is identical to (2.6.2).

Since $b_1 = \Gamma_{IN}a_1$, (2.7.11) can be written as

$$a_1 = b_s + \Gamma_s \Gamma_{IN} a_1$$

or

$$a_1 = \frac{b_s}{1 - \Gamma_s \Gamma_{IN}} \quad (2.7.12)$$

Substituting (2.7.12) into (2.7.9) gives

$$P_{IN} = \frac{1}{2} |b_s|^2 \frac{1 - |\Gamma_{IN}|^2}{|1 - \Gamma_s \Gamma_{IN}|^2} \quad (2.7.13)$$

The power available from the source is equal to the input power when $\Gamma_{IN} = \Gamma_s^*$. With $\Gamma_{IN} = \Gamma_s^*$, (2.7.13) gives

$$P_{AVS} = P_{IN}|_{\Gamma_{IN}=\Gamma_s^*} = \frac{\frac{1}{2} |b_s|^2}{1 - |\Gamma_s|^2} \quad (2.7.14)$$

Observe that (2.7.14) agrees with (2.6.10).

Substituting (2.7.14) into (2.7.13), we can express P_{IN} in the form

$$P_{IN} = P_{AVS} \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_{IN}|^2)}{|1 - \Gamma_s \Gamma_{IN}|^2} \quad (2.7.15)$$

Equation (2.7.15) can be written as

$$P_{IN} = P_{AVS} M_s \quad (2.7.16)$$

where

$$M_s = \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_{IN}|^2)}{|1 - \Gamma_s \Gamma_{IN}|^2} \quad (2.7.17)$$

The factor M_s is known as the source mismatch factor (or the source mismatch loss). This factor is used to quantize what portion of P_{AVS} is delivered to the input of the transistor. Observe that if $\Gamma_{IN} = \Gamma_s^*$, (2.7.17) gives $M_s = 1$ and it follows that $P_{IN} = P_{AVS}$. This fact is expressed in the form

$$P_{IN} = P_{AVS}|_{\Gamma_{IN}=\Gamma_s^*}$$

A Thévenin's equivalent circuit at the output port of the transistor is shown in Fig. 2.7.3, where Z_{OUT} (i.e., the output impedance) is the Thévenin's impedance seen from the output terminals of the transistor. From Fig. 2.7.3, we have

$$V_L = E_{TH} - I_L Z_{OUT}$$

This expression is similar to (2.7.10) and can be expressed in terms of traveling waves as

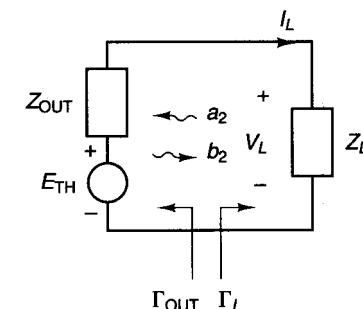


Figure 2.7.3 A Thévenin's equivalent circuit at the output port of the transistor.

$$b_2 = b_{TH} + \Gamma_{OUT} a_2 \quad (2.7.18)$$

where

$$b_2 = \frac{V_L^-}{\sqrt{Z_o}}$$

$$a_2 = \frac{V_L^+}{\sqrt{Z_o}}$$

$$b_{TH} = \frac{E_{TH} \sqrt{Z_o}}{Z_{OUT} + Z_o}$$

and

$$\Gamma_{OUT} = \frac{Z_{OUT} - Z_o}{Z_{OUT} + Z_o}$$

Since $a_2 = \Gamma_L b_2$, (2.7.18) gives

$$b_2 = b_{TH} + \Gamma_{OUT} \Gamma_L b_2$$

or

$$b_2 = \frac{b_{TH}}{1 - \Gamma_{OUT} \Gamma_L} \quad (2.7.19)$$

The power delivered to the load Z_L is

$$P_L = \frac{1}{2} |b_2|^2 - \frac{1}{2} |a_2|^2 = \frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2) \quad (2.7.20)$$

Substituting (2.7.19) into (2.7.20), we obtain

$$P_L = \frac{1}{2} |b_{TH}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT} \Gamma_L|^2} \quad (2.7.21)$$

The power available from the network P_{AVN} is equal to the power delivered to the load when $\Gamma_L = \Gamma_{OUT}^*$. With $\Gamma_L = \Gamma_{OUT}^*$, (2.7.21) gives

$$P_{AVN} = P_L |_{\Gamma_L = \Gamma_{OUT}^*} = \frac{\frac{1}{2} |b_{TH}|^2}{1 - |\Gamma_{OUT}|^2} \quad (2.7.22)$$

Substituting (2.7.22) into (2.7.21), we can express P_L in the form

$$P_L = P_{AVN} \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_{OUT}|^2)}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.7.23)$$

Equation (2.7.23) can be written in the form

$$P_L = P_{AVN} M_L \quad (2.7.24)$$

where

$$M_L = \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_{OUT}|^2)}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.7.25)$$

Observe the similarity between (2.7.24) and (2.7.16). The factor M_L is known as the load mismatch factor. This factor is used to quantize what portion of P_{AVN} is delivered to the load. For $\Gamma_L = \Gamma_{OUT}^*$, (2.7.25) gives $M_L = 1$ and it follows that $P_L = P_{AVN}$. This fact is expressed in the form

$$P_L = P_{AVN} |_{\Gamma_L = \Gamma_{OUT}^*}$$

The power gain is given by

$$G_p = \frac{P_L}{P_{IN}}$$

Therefore, using (2.7.9) and (2.7.20), we can write

$$G_p = \frac{P_L}{P_{IN}} = \frac{\frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2)}{\frac{1}{2} |a_1|^2 (1 - |\Gamma_{IN}|^2)} \quad (2.7.26)$$

Then, substituting (2.7.6) into (2.7.26), we obtain

$$G_p = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (2.7.27)$$

The transducer power gain is given by

$$G_T = \frac{P_L}{P_{AVS}} = \frac{P_L}{P_{IN} P_{AVS}} = G_p \frac{P_{IN}}{P_{AVS}} \quad (2.7.28)$$

From (2.7.16), the ratio P_{IN}/P_{AVS} is M_s ; therefore, (2.7.28) can be written as

$$G_T = G_p M_s \quad (2.7.29)$$

Substituting (2.7.17) and (2.7.27) into (2.7.29), we obtain

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_s \Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (2.7.30)$$

Manipulating the denominator, (2.7.30) can also be written in the form

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.7.31)$$

The available power gain can be expressed in the form

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{P_L}{P_{AVS}} \frac{P_{AVN}}{P_L} = \frac{G_T}{M_L}$$

Then, using (2.7.25) and (2.7.31), we can write G_A in the form

$$G_A = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{OUT}|^2} \quad (2.7.32)$$

It is of interest to investigate further the significance of the mismatch factors M_s and M_L . From (2.7.16), the factor M_s relates P_{AVS} to P_{IN} . Consider the network in Fig. 2.7.4a. For this network,

$$P_{AVS} = \frac{1}{8} \frac{|E_s|^2}{R_s}$$

and the input power is given by

$$P_{IN} = \frac{1}{2} \left| \frac{E_s}{Z_s + Z_{IN}} \right|^2 R_{IN} = \frac{1}{8} \frac{|E_s|^2}{R_s} \left(\frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} \right) = P_{AVS} \left(\frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} \right) \quad (2.7.33)$$

Comparing (2.7.33) with (2.7.16), it follows that the mismatch factor M_s is the term in parentheses in (2.7.33)—namely,

$$M_s = \frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} \quad (2.7.34)$$

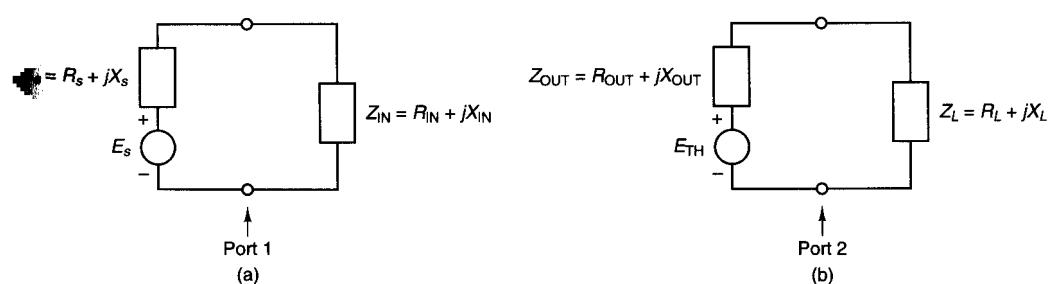


Figure 2.7.4 (a) The input port; (b) the output port.

Although it is not obvious at first sight, (2.7.34) and (2.7.17) are identical. To prove this, we use a normalizing impedance Z_o and write

$$Z_s = Z_o \frac{1 + \Gamma_s}{1 - \Gamma_s}$$

Then, the real part of Z_s is

$$R_s = \frac{1}{2} (Z_s + Z_s^*) = \frac{Z_o}{2} \left[\frac{1 + \Gamma_s}{1 - \Gamma_s} + \frac{1 + \Gamma_s^*}{1 - \Gamma_s^*} \right] = Z_o \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_s|^2} \quad (2.7.35)$$

Similarly, we express the real part of Z_{IN} as

$$R_{IN} = Z_o \frac{1 - |\Gamma_{IN}|^2}{|1 - \Gamma_{IN}|^2} \quad (2.7.36)$$

Substituting (2.7.35) and (2.7.36) into (2.7.34), and simplifying a bit, it follows that

$$M_s = \frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} = \frac{(1 - |\Gamma_{IN}|^2)(1 - |\Gamma_s|^2)}{|1 - \Gamma_s \Gamma_{IN}|^2}$$

At the output port in Fig. 2.7.4b, the factor M_L is given by

$$M_L = \frac{4R_{OUT} R_L}{|Z_{OUT} + Z_L|^2}$$

which is identical to (2.7.25).

Power-Gain Expressions in Terms of S_p Parameters

Consider the two-port network in Fig. 2.7.5, where the two-port network is characterized by its S_p parameters. In Section 1.7, the expression for G_T in terms of S_p parameters was derived [see (1.7.27)]. That is,

$$G_T = \frac{P_L}{P_{AVS}} = |S_{p21}|^2 \quad (2.7.37)$$

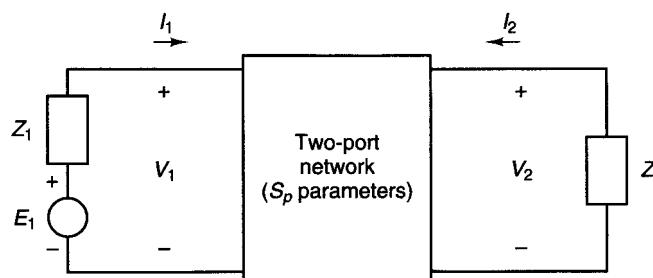


Figure 2.7.5 Two-port network characterized by its S_p parameters.

Using (2.7.28), the power gain G_p can be expressed as

$$G_p = G_T \frac{P_{AVS}}{P_{IN}} \quad (2.7.38)$$

In Fig. 2.7.5, the input power is given by

$$P_{IN} = P_{AVS}(1 - |S_{p11}|^2) \quad (2.7.39)$$

Substituting (2.7.37) and (2.7.39) into (2.7.38) gives

$$G_p = \frac{|S_{p21}|^2}{1 - |S_{p11}|^2} \quad (2.7.40)$$

Using (2.7.24), the power available from the network can be expressed as

$$G_A = \frac{G_T}{M_L} = \frac{G_T}{P_L} P_{AVN} \quad (2.7.41)$$

The power delivered to the load Z_L in Fig. 2.7.5 is given by

$$P_L = P_{AVN}(1 - |S_{p22}|^2) \quad (2.7.42)$$

Substituting (2.7.37) and (2.7.42) into (2.7.41) gives

$$G_A = \frac{|S_{p21}|^2}{1 - |S_{p22}|^2} \quad (2.7.43)$$

The power-gain expressions in terms of S_p parameters are given in (2.7.37), (2.7.40), and (2.7.43). Next, we will show that these relations are equivalent to the power-gain expressions in terms of S parameters given in (2.7.30), (2.7.27), and (2.7.32).

The derivation of the S_p parameters in terms of the S parameters is a difficult one. It consists of expressing the S parameters and S_p parameters in terms of z parameters and then eliminating the z parameters between the two relations to obtain the relation between the S_p parameters in terms of S parameters. The results are as follows:

$$S_{p11} = \frac{(1 - \Gamma_s)}{(1 - \Gamma_s^*)} \frac{(1 - \Gamma_L S_{22}) (S_{11} - \Gamma_s^*) + S_{12} S_{21} \Gamma_L}{D} \quad (2.7.44)$$

$$S_{p22} = \frac{(1 - \Gamma_L)}{(1 - \Gamma_L^*)} \frac{(1 - \Gamma_s S_{11}) (S_{22} - \Gamma_L^*) + S_{12} S_{21} \Gamma_s}{D} \quad (2.7.45)$$

$$S_{p21} = \frac{(1 - \Gamma_s)}{|1 - \Gamma_s|} \frac{|1 - \Gamma_L|}{(1 - \Gamma_L^*)} \frac{S_{21} [(1 - |\Gamma_s|^2)(1 - |\Gamma_L|^2)]^{1/2}}{D} \quad (2.7.46)$$

and

$$S_{p12} = \frac{(1 - \Gamma_L)}{|1 - \Gamma_L|} \frac{|1 - \Gamma_s|}{(1 - \Gamma_s^*)} \frac{S_{12} [(1 - |\Gamma_s|^2)(1 - |\Gamma_L|^2)]^{1/2}}{D} \quad (2.7.47)$$

where

$$D = (1 - \Gamma_s S_{11})(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s$$

Substituting (2.7.46) into (2.7.37), we obtain the expression for G_T in terms of S parameters—namely,

$$G_T = |S_{p21}|^2 = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{\text{OUT}}\Gamma_L|^2}$$

which is recognized as (2.7.30). Similarly, substituting (2.7.44) and (2.7.46) into (2.7.40), we obtain the expression for G_p in (2.7.27); and substituting (2.7.45) and (2.7.46) into (2.7.43) gives the expression for G_A in (2.7.32).

2.8 VSWR CALCULATIONS

The VSWR in a transmission line is given by (1.3.44). Since the power delivered to the load in a transmission line, excited by a source with impedance Z_o , is given by

$$P_L = P_{\text{AVS}}(1 - |\Gamma_0|^2)$$

it follows that the magnitude of the load reflection coefficient, $|\Gamma_0|$ (which is related to the VSWR), provides a measure of what portion of P_{AVS} is delivered to the load.

For example, for a VSWR = 1 it follows that $|\Gamma_0| = 0$ and $P_L = P_{\text{AVS}}$ (i.e., all the incident power is delivered to the load). For a VSWR = 1.5, it follows that $|\Gamma_0| = 0.2$ and the ratio of the incident to the reflected power from the load is $|\Gamma_0|^2 = 0.04$, or 4%. Hence, 4% of the incident power is reflected by the load (96% of the incident power is delivered to the load). For a VSWR = 2, then $|\Gamma_0| = 1/3$ and $|\Gamma_0|^2 = 0.11$, or 11%. In this case, 11% of the incident power is reflected by the load.

The values of the input and output VSWR are important to the microwave amplifier designer. For example, many microwave amplifiers require the input VSWR to be less than 1.5. In other designs, higher VSWR values must be tolerated in order to obtain other performances, such as a particular noise performance.

Figure 2.8.1 shows the input portion of a microwave amplifier. The reflection coefficient at the input of the lossless matching network, normalized to Z_o , is denoted by Γ_a . The input VSWR, denoted by $(\text{VSWR})_{\text{in}}$, is related to Γ_a by

$$(\text{VSWR})_{\text{in}} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} \quad (2.8.1)$$

where

$$\Gamma_a = \frac{Z_a - Z_o}{Z_a + Z_o}$$

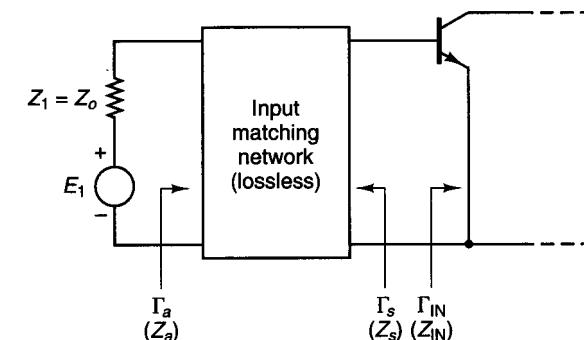


Figure 2.8.1 Input portion of a microwave amplifier.

Since the input power is given by

$$P_{\text{IN}} = P_{\text{AVS}}(1 - |\Gamma_a|^2)$$

and from (2.7.16) we have

$$P_{\text{IN}} = P_{\text{AVS}}M_s$$

it follows that

$$M_s = 1 - |\Gamma_a|^2$$

or

$$|\Gamma_a| = \sqrt{1 - M_s} \quad (2.8.2)$$

Equations (2.8.1) and (2.8.2) show that $(\text{VSWR})_{\text{in}}$ can be calculated from a knowledge of the mismatch factor M_s .

Substituting (2.7.17) into (2.8.2) and simplifying results in the following expression for $|\Gamma_a|$:

$$|\Gamma_a| = \sqrt{1 - \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_{\text{IN}}|^2)}{|1 - \Gamma_s\Gamma_{\text{IN}}|^2}} = \left| \frac{\Gamma_{\text{IN}} - \Gamma_s^*}{1 - \Gamma_{\text{IN}}\Gamma_s} \right| \quad (2.8.3)$$

This relation shows that $|\Gamma_a|$ can be calculated from a knowledge of Γ_{IN} and Γ_s .

The concept of a reflection coefficient and of a VSWR can be used in a microwave amplifier even if there are no transmission lines. For example, in Fig. 2.8.1 this can occur when the input matching network is implemented using lumped components. The reflection coefficients and $(\text{VSWR})_{\text{in}}$ in Fig. 2.8.1 can be thought of as those of a transmission line of zero length with characteristic impedance Z_o .

Similar relations can be defined for the output VSWR. The output portion of a microwave amplifier is shown in Fig. 2.8.2. For Fig. 2.8.2, we have

$$(\text{VSWR})_{\text{out}} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} \quad (2.8.4)$$

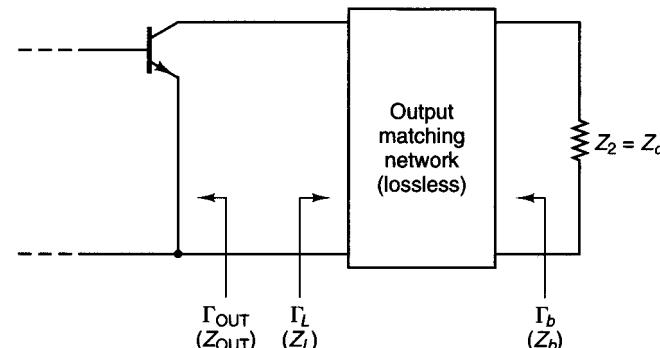


Figure 2.8.2 Output portion of a microwave amplifier.

$$\Gamma_b = \frac{Z_b - Z_o}{Z_b + Z_o}$$

$$|\Gamma_b| = \sqrt{1 - M_L} \quad (2.8.5)$$

and

$$|\Gamma_b| = \sqrt{1 - \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_{\text{OUT}}|^2)}{|1 - \Gamma_{\text{OUT}}\Gamma_L|^2}} = \left| \frac{\Gamma_{\text{OUT}} - \Gamma_L^*}{1 - \Gamma_{\text{OUT}}\Gamma_L} \right| \quad (2.8.6)$$

Example 2.8.1

(a) The input portion of the microwave amplifier in Fig. 2.5.14 is drawn in Fig. 2.8.3a. Assume that the S parameters of the transistor are such that the resulting Γ_{IN} is $0.614| -160^\circ$. Calculate $(\text{VSWR})_{\text{in}}$.

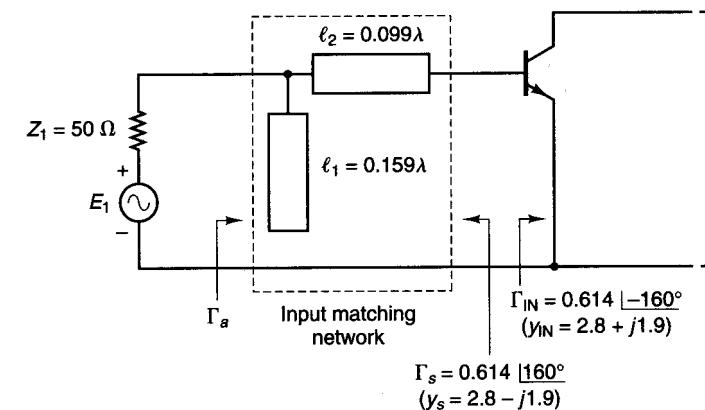
(b) Calculate $(\text{VSWR})_{\text{in}}$ in Fig. 2.8.3a if Γ_{IN} is $0.4| -145^\circ$.

Solution. (a) From (2.8.3), with $\Gamma_s = \Gamma_{\text{IN}}^* = 0.614|160^\circ$, it follows that $|\Gamma_a| = 0$. Therefore, using (2.8.1), we obtain $(\text{VSWR})_{\text{in}} = 1$.

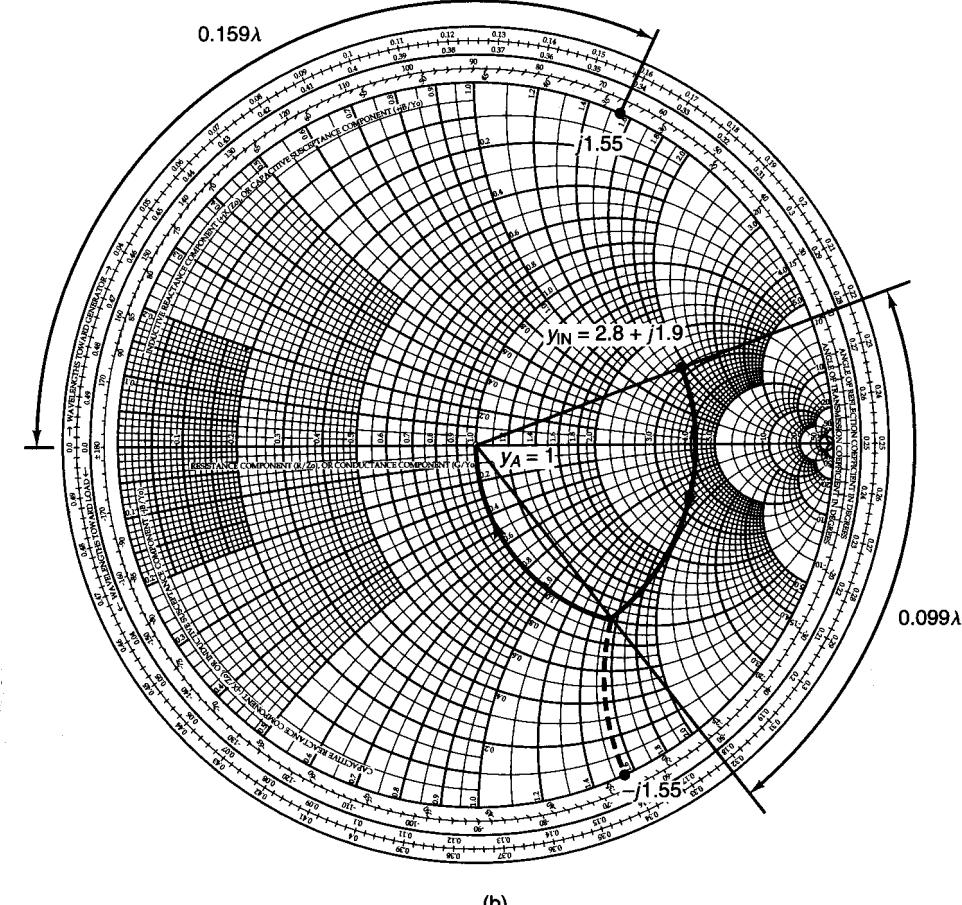
The previous calculation illustrates the fact that when $\Gamma_s = \Gamma_{\text{IN}}^*$, a conjugate matched condition exists at the input of the transistor and at the source. In other words, $Z_a = 50 \Omega$ (since $|\Gamma_a| = 0$) and the source sees a matched input impedance. This can also be verified by evaluating Z_a for the circuit in Fig. 2.8.3a. In Fig. 2.8.3b, we have plotted the normalized input admittance $y_{\text{IN}} = 1/z_{\text{IN}} = 2.8 + j1.9$ and calculated y_a to be 1, thus verifying that $Z_a = 50 \Omega$.

(b) In this case, Γ_{IN} is given as $0.4| -145^\circ$. Therefore, $\Gamma_s \neq \Gamma_{\text{IN}}^*$, so the input of the transistor is not conjugate matched for maximum power transfer. Using (2.8.3), we obtain

$$|\Gamma_a| = \left| \frac{0.4| -145^\circ - 0.614| -160^\circ}{1 - 0.4| -145^\circ(0.614|160^\circ)} \right| = 0.327$$



(a)



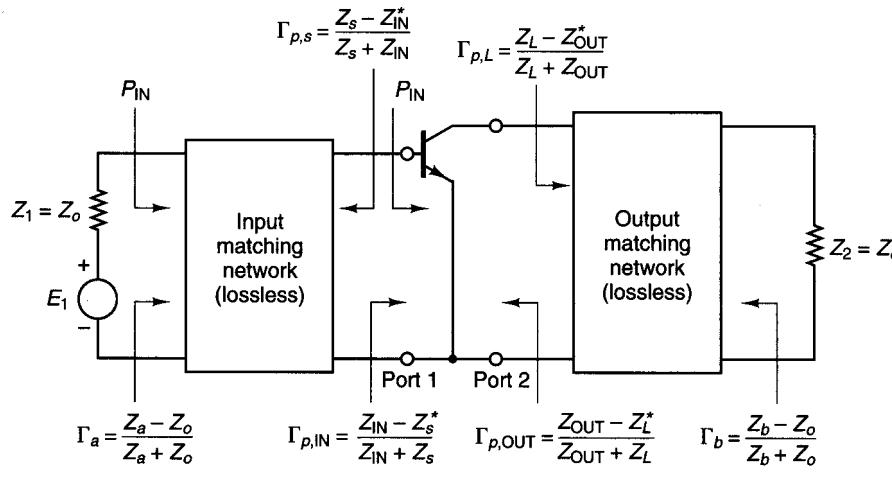
(b)

Figure 2.8.3 (a) Circuit for Example 2.8.1; (b) calculation of Z_a using the Smith chart.

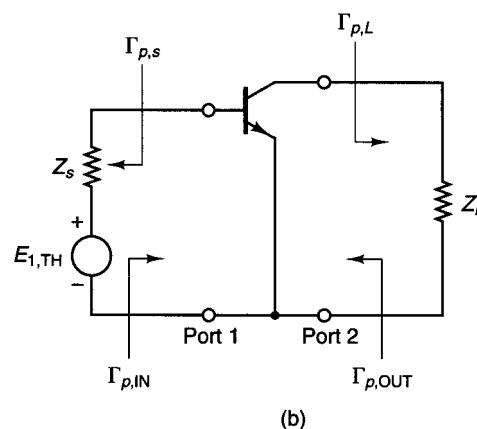
Therefore, the value of $(\text{VSWR})_{\text{in}}$ is

$$(\text{VSWR})_{\text{in}} = \frac{1 + 0.327}{1 - 0.327} = 1.97$$

Let us proceed a little further with the analysis of the VSWR. Consider Fig. 2.8.4a, where the matching networks are assumed to be lossless. The indicated coefficients Γ_a , Γ_b , $\Gamma_{p,\text{IN}}$, and $\Gamma_{p,\text{OUT}}$ are power reflection coefficients. Since the impedances Z_1 and Z_2 are real (i.e., $Z_1 = Z_2 = Z_o$), it follows that Γ_a and Γ_b are identical to the traveling-wave reflection coefficients. The equivalent circuits at the input and output ports of the transistor are shown in Fig. 2.8.4b.



(a)



(b)

Figure 2.8.4 (a) Power reflection coefficients in a microwave amplifier; (b) an equivalent circuit.

The input power at the generator input in Fig. 2.8.4a can be written as

$$P_{\text{IN}} = P_{\text{AVS}}(1 - |\Gamma_a|^2) = P_{\text{AVS}}M_s \quad (2.8.7)$$

Since the matching circuits are assumed to be lossless, at the input of the transistor we can write

$$P_{\text{IN}} = P_{\text{AVS}}(1 - |\Gamma_{p,\text{IN}}|^2) = P_{\text{AVS}}M'_s \quad (2.8.8)$$

where $\Gamma_{p,\text{IN}}$ is the power reflection coefficient at the input of the transistor—namely,

$$\Gamma_{p,\text{IN}} = \frac{Z_{\text{IN}} - Z_s^*}{Z_{\text{IN}} + Z_s} \quad (2.8.9)$$

and

$$M'_s = 1 - |\Gamma_{p,\text{IN}}|^2$$

In writing (2.8.8), we used the fact that P_{AVS} is the same at the input and output ports of the lossless matching circuit. The parameter $\Gamma_{p,\text{IN}}$ is identical to the value of S_{p11} associated with the two-port network in Fig. 2.8.4b.

Equating (2.8.7) to (2.8.8) gives

$$M_s = M'_s \quad (2.8.10)$$

or

$$|\Gamma_a| = |\Gamma_{p,\text{IN}}|$$

It also follows that $|\Gamma_{p,\text{IN}}| = |\Gamma_{p,s}|$. From the previous results, $(\text{VSWR})_{\text{in}}$ can also be expressed in terms of $\Gamma_{p,\text{IN}}$ as

$$(\text{VSWR})_{\text{in}} = \frac{1 + |\Gamma_{p,\text{IN}}|}{1 - |\Gamma_{p,\text{IN}}|} \quad (2.8.11)$$

Equation (2.8.10) shows that the mismatch factor is an invariant quantity in a lossless matching network. That is, the mismatch factor M_s at the input of the input matching network is equal to the mismatch factor M'_s at the output of the input matching network. Using (2.7.34), we can express (2.8.10) in terms of impedances as

$$\frac{4Z_oR_a}{|Z_o + Z_a|^2} = \frac{4R_sR_{\text{IN}}}{|Z_s + Z_{\text{IN}}|^2}$$

where $R_a = \text{Re}[Z_a]$, $R_s = \text{Re}[Z_s]$, and $R_{\text{IN}} = \text{Re}[Z_{\text{IN}}]$.

Similarly, at the output of the amplifier in Fig. 2.8.4a, we can write

$$M_L = M'_L$$

$$M_L = 1 - |\Gamma_b|^2 = \frac{4Z_oR_b}{|Z_o + Z_b|^2}$$

$$M'_L = 1 - |\Gamma_{p,\text{OUT}}|^2 = \frac{4R_L R_{\text{OUT}}}{|Z_L + Z_{\text{OUT}}|^2}$$

$$|\Gamma_b| = |\Gamma_{p,\text{OUT}}| = |\Gamma_{p,L}|$$

$$\Gamma_{p,\text{OUT}} = \frac{Z_{\text{OUT}} - Z_L^*}{Z_{\text{OUT}} + Z_L}$$

and

$$(\text{VSWR})_{\text{out}} = \frac{1 + |\Gamma_{p,\text{OUT}}|}{1 - |\Gamma_{p,\text{OUT}}|} \quad (2.8.12)$$

where $R_b = \text{Re}[Z_b]$, $R_L = \text{Re}[Z_L]$, and $R_{\text{OUT}} = \text{Re}[Z_{\text{OUT}}]$. The parameter $\Gamma_{p,\text{OUT}}$ is identical to the value of S_{p22} associated with the two-port network in Fig. 2.8.4b.

Example 2.8.2

- (a) Calculate the $(\text{VSWR})_{\text{in}}$ in Example 2.8.1, part (a), using (2.8.11).
- (b) Calculate the $(\text{VSWR})_{\text{in}}$ in Example 2.8.1, part (b), using (2.8.11). Also, evaluate M'_s .

Solution. (a) With $\Gamma_s = \Gamma_{\text{IN}}^*$, it follows that $Z_s = Z_{\text{IN}}^* = 12.308 + j8.297 \Omega$. Therefore, from (2.8.9) it follows that $\Gamma_{p,\text{IN}} = 0$ and from (2.8.11), $(\text{VSWR})_{\text{in}} = 1$.

(b) With $\Gamma_{\text{IN}} = 0.4|-145^\circ$ (or $Z_{\text{IN}} = 23.136 - j12.638 \Omega$), it follows from (2.8.9) that

$$\Gamma_{p,\text{IN}} = \frac{Z_{\text{IN}} - Z_s^*}{Z_{\text{IN}} + Z_s} = \frac{23.136 - j12.638 - (12.308 - j8.297)}{23.136 - j12.638 + (12.308 + j8.297)} = 0.327|-14.86^\circ$$

and

$$(\text{VSWR})_{\text{in}} = \frac{1 + 0.327}{1 - 0.327} = 1.97$$

As expected, $|\Gamma_a| = |\Gamma_{p,\text{IN}}| = |\Gamma_{p,s}| = 0.327$.

The mismatch factor M_s is

$$M_s = M'_s = 1 - |\Gamma_{p,\text{IN}}|^2 = 1 - (0.327)^2 = 0.893$$

PROBLEMS

- 2.1 (a)** Locate in the Z Smith chart the impedance $Z = 100 + j100 \Omega$ using a reference impedance of 50Ω .
- (b)** Find the value of the normalized admittance y using the Z Smith chart, and evaluate the value of Y (i.e., $Y = yY_o = y/Z_o$).
- (c)** Find y using the ZY Smith chart.
- (d)** Repeat parts (a) to (c) for the following impedances: $Z = 50 - j100 \Omega$, $Z = 25 - j25 \Omega$, $Z = j50 \Omega$, and $Z = j0 \Omega$.
- 2.2 (a)** Show that impedances having a negative real part (i.e., $z = -r + jx$) have a reflection coefficient whose magnitude is greater than 1.

Problems

- (b)** Prove that negative resistances can be handled in the Smith chart by plotting $1/\Gamma^*$ and interpreting the resistance circles as being negative and the reactance circles as marked.
- (c)** Locate in the Smith chart the impedances $Z_1 = -20 + j16 \Omega$ and $Z_2 = -200 + j25 \Omega$ and find the associated reflection coefficient. Normalize the impedances to 50Ω .
- (d)** Work the problem in part (c) in the compressed Smith chart.
- (e)** What is the value of $|\Gamma|$ on the boundary of the compressed Smith chart in Fig. 2.2.5?

- 2.3** Show that the impedance along a transmission line repeats itself at every $\lambda/2$ distance. That is,

$$Z(d) = Z\left(d + \frac{n\lambda}{2}\right), \quad n = 1, 2, 3, \dots$$

- 2.4** Show that the impedance along a transmission line can be expressed in the form

$$Z(d) = R(d) + jX(d) = |Z(d)|e^{j\theta_d}$$

where

$$R(d) = Z_o \frac{1 - |\Gamma|^2}{1 - 2|\Gamma| \cos \phi + |\Gamma|^2}$$

$$X(d) = Z_o \frac{2|\Gamma| \sin \phi}{1 - 2|\Gamma| \cos \phi + |\Gamma|^2}$$

$$|Z(d)| = Z_o \sqrt{\frac{1 + 2|\Gamma| \cos \phi + |\Gamma|^2}{1 - 2|\Gamma| \cos \phi + |\Gamma|^2}}$$

$$\theta_d = \tan^{-1} \frac{X(d)}{R(d)} = \tan^{-1} \left(\frac{2|\Gamma| \sin \phi}{1 - |\Gamma|^2} \right)$$

$$\Gamma = |\Gamma_0|e^{j\phi}, \quad \Gamma_0 = |\Gamma_0|e^{j\phi_l}, \quad \phi = \phi_l - 2\beta d$$

- 2.5** Find the input impedance, the load reflection coefficient, and the VSWR in a transmission line having an electrical length of 90° , $Z_o = 50 \Omega$, and terminated in the load $Z_L = 50 + j100 \Omega$. Work the problem in both the Z and Y Smith charts.
- 2.6 (a)** Prove that the maximum normalized resistance in a transmission line is numerically equal to the VSWR.
- (b)** Prove that the minimum normalized resistance in a transmission line is numerically equal to $1/\text{VSWR}$.
- 2.7 (a)** Determine the length l of the $50-\Omega$ short-circuited transmission line shown in Fig. 2.2.8a so that the input impedance is $Z_{\text{IN}}(l) = -j25 \Omega$.
- (b)** Determine the length l of the $50-\Omega$ open-circuited transmission line shown in Fig. 2.2.8b so that the normalized input admittance is $y_{\text{IN}}(l) = j2$.
- 2.8** The normalized admittance (with $Z_o = 50 \Omega$) of a one-port network is shown in Fig. P2.8 as the frequency varies from 500 MHz to 1 GHz. Determine an equivalent circuit for the one-port network and the element values.
- 2.9 (a)** In the series RC circuit shown in Fig. 2.3.2, the value of C is 50 pF. Determine the values of f_a and f_b .

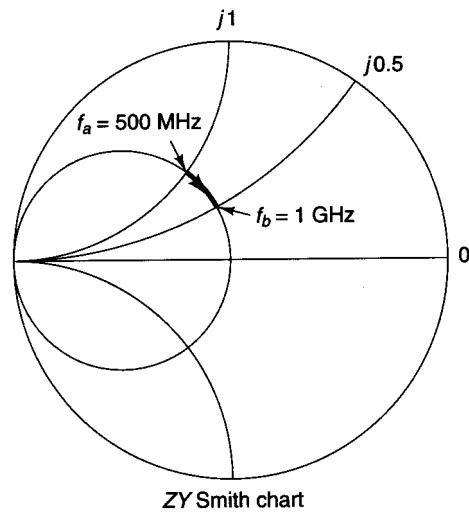


Figure P2.8

- (b) In Example 2.3.2, assume that at $f_b = 1 \text{ GHz}$ the normalized impedance is $j0.6$. At $f_a = 500 \text{ MHz}$, the normalized impedance remains at $j0.2$. Determine an equivalent circuit for the one-port network and the element values. Observe that in this problem the reactance does not increase linearly with frequency.
- 2.10 Design four different Ell matching networks to match the load $Z_{\text{LOAD}} = 10 + j40 \Omega$ to a 50Ω transmission line.
- 2.11 Design the matching network in Fig. P2.11 that provides $Y_L = (4 - j4) \times 10^{-3} \text{ S}$ to the transistor. Find the element values at 700 MHz.

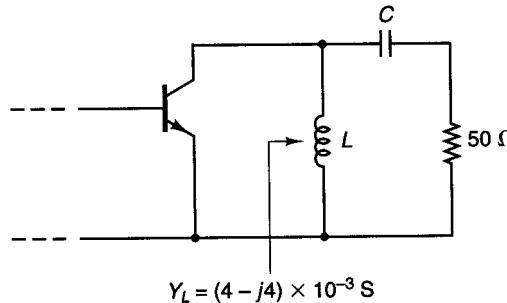


Figure P2.11

- 2.12 Two types of Ell matching networks are shown in Fig. P2.12. Select one that can match the load $Y_{\text{LOAD}} = (8 - j12) \times 10^{-3} \text{ S}$ to a 50Ω transmission line. Find the element values at $f = 1 \text{ GHz}$.

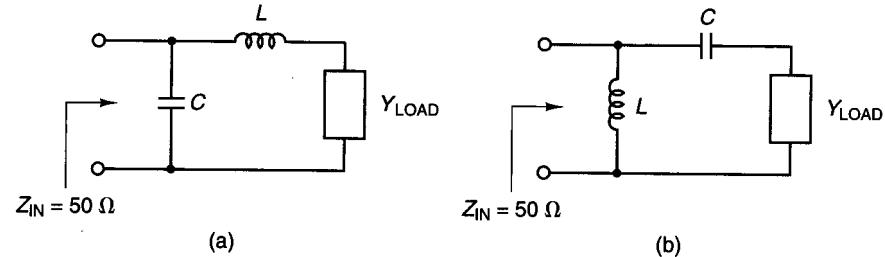


Figure P2.12

- 2.13 Show the impedance-admittance path in the ZY Smith chart for the circuit shown in Fig. P2.13 at $\omega = 10^9 \text{ rad/s}$, and evaluate Z_{IN} .

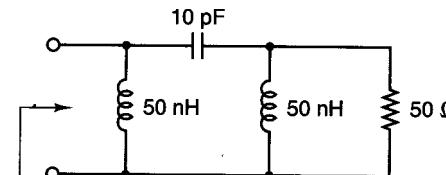


Figure P2.13

- 2.14 Design a two-element matching network to transform the load $Z_{\text{LOAD}} = 100 - j100 \Omega$ to an input impedance of $Z_{\text{IN}} = 25 + j25 \Omega$.
- 2.15 (a) Design a Tee matching network to transform $Z_{\text{LOAD}} = 50 \Omega$ to the input impedance $Z_{\text{IN}} = 20 + j20 \Omega$ with a Q of 5.
 (b) Design a Pi matching network to transform $Z_{\text{LOAD}} = 50 \Omega$ to the input impedance $Z_{\text{IN}} = 25 \Omega$ with a Q of 2.5.
- 2.16 (a) Use Figs. 2.5.2 and 2.5.3 to calculate W , λ , and ϵ_{eff} for a characteristic impedance of 50Ω using RT/Duroid® with $\epsilon_r = 2.23$ and $h = 0.7874 \text{ mm}$.
 (b) Use (2.5.8) through (2.5.11) to show that for RT/Duroid® with $\epsilon_r = 2.23$ and $h = 0.7874 \text{ mm}$, a 50Ω characteristic impedance is obtained with $W/h = 3.073$. Also, $\epsilon_{\text{eff}} = 1.91$ and $\lambda = 0.7236\lambda_0$.
- 2.17 The 10-nH inductor shown in Fig. P2.17 is to be implemented at 1 GHz using an open-circuited microstrip line with $Z_o = 50 \Omega$. The microstrip material has $\epsilon_r = 6$ and $h = 25 \text{ mils}$. Determine the width and length of the microstrip line.

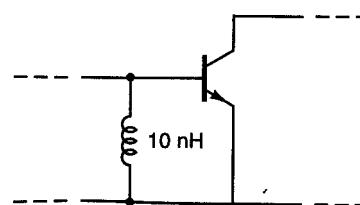


Figure P2.17

- 2.18** In the amplifier shown in Fig. 2.5.16b, calculate the width and the length of the lines at $f = 1$ GHz

- 2.19** (a) Design the matching circuit shown in Fig. P2.19 to transform the 50Ω load to the input impedance $Z_{IN} = 100 - j100 \Omega$.

- (b) What is the length l_1 if the short-circuited shunt stub is replaced by an open-circuited shunt stub?

- (c) Repeat parts (a) and (b) for $Z_{IN} = 100 + j100 \Omega$.

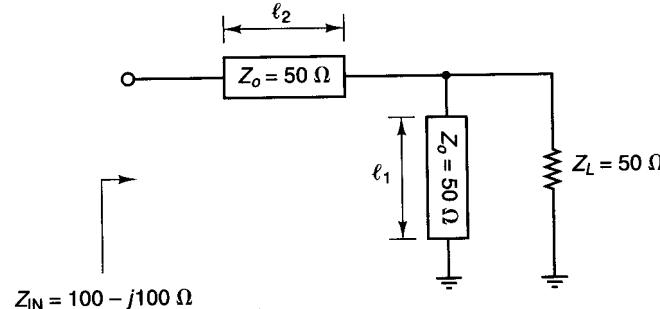


Figure P2.19

- 2.20** (a) Design a single-stub matching system (see Fig. P2.20) to match the load $Z_L = 15 + j25 \Omega$ to a 50Ω transmission line. The characteristic impedance of the short-circuited stub is 50Ω .

- (b) Design the single-stub matching system in Fig. P2.20 assuming that the characteristic impedance of the stub is 100Ω .

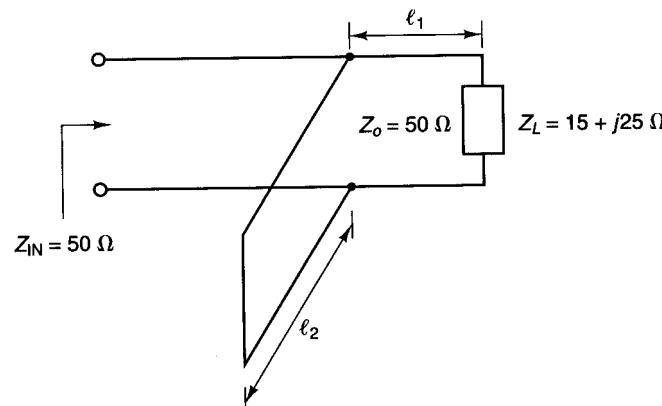


Figure P2.20

- 2.21** Design the matching circuits shown in Fig. P2.21 to transform the 50Ω load impedance to the input admittance given in the figures.

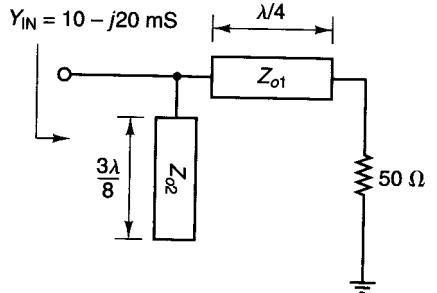
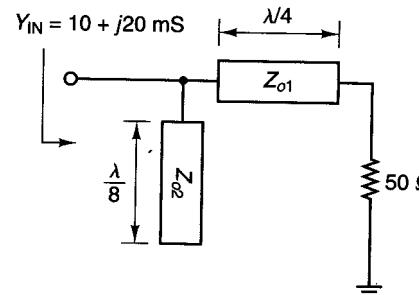
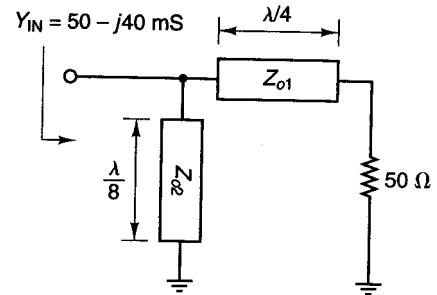
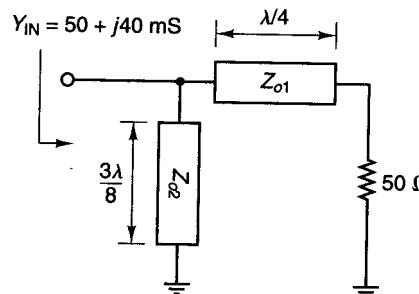


Figure P2.21

- 2.22** (a) Design the matching networks shown in Fig. P2.22 to produce the source reflection coefficient $\Gamma_s = 0.5|90^\circ|$. In Fig. P2.22b, the appropriate length for the short-circuited stub must be selected (i.e., $\lambda/8$ or $3\lambda/8$).

- (b) Design the balance form of the shunt stubs.

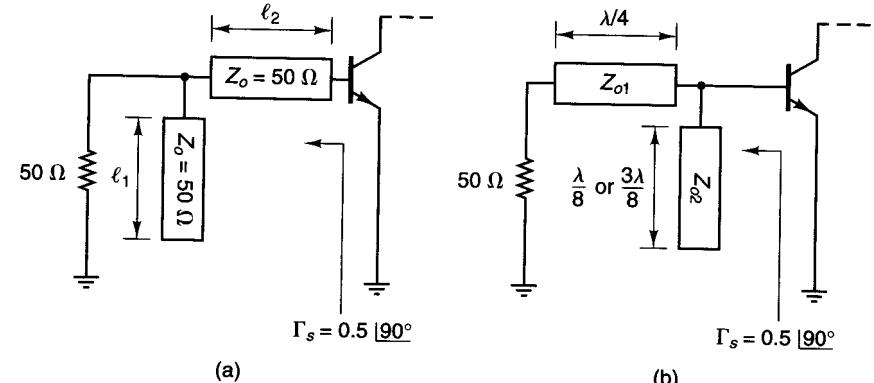


Figure P2.22

2.23 In the design shown in Fig. 2.5.16, the balance stubs of length $3\lambda/8$ were designed with characteristic impedances of 52.64Ω and 95.2Ω , respectively. Design the length of the balance stubs if the characteristic impedances are left at 26.32Ω and 47.6Ω , respectively.

2.24 (a) Design the matching network in Fig. P2.24 to produce a load reflection coefficient of $\Gamma_L = 0.4| -120^\circ$ to the transistor. The appropriate length for the balance stubs must be selected (i.e., $\lambda/8$ or $3\lambda/8$).

(b) Design the length of the balance stubs if the impedance Z_{o2} is changed to $Z_{o2}/2$.

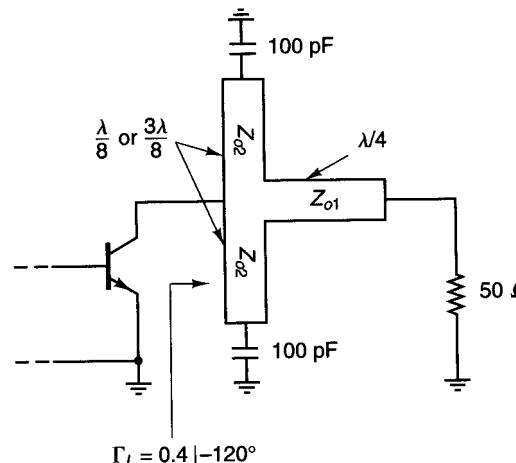


Figure P2.24

2.25 Design two microstrip matching networks for an amplifier whose reflection coefficients at $f = 800 \text{ MHz}$, in a $50\text{-}\Omega$ system, are $\Gamma_s = 0.8|160^\circ$ and $\Gamma_L = 0.7|20^\circ$. Show the diagram for the complete amplifier using balanced shunt stubs.

2.26 (a) Determine the value of Γ_s in Fig. P2.26a.

(b) Determine the value of Γ_s in Fig. P2.26b. Observe that the 100-pF capacitor is a short circuit to the ac signal.

2.27 (a) Determine the value of Γ_L in Fig. P2.27. The lengths shown are for $\epsilon_{eff} = 1$ and $f = 6 \text{ GHz}$.

(b) Show the balance form of the shunt stubs.

2.28 The input matching network shown in Fig. P2.28 was designed to obtain a certain gain. If the input reflection coefficient is $\Gamma_{IN} = 0.5|100^\circ$, determine the value of the impedance Z_A seen by the source.

2.29 Design a microstrip matching network to transform the load impedance $Z_L = 50 - j50 \Omega$ to the input impedance $Z_{IN} = 25 + j25 \Omega$ in Fig. P2.29.

2.30 Determine the value of Γ_L in Fig. P2.30.

2.31 Design a two-element matching network, as shown in the Smith chart in Fig. P2.31, that produces $\Gamma_s = 0.57|116^\circ$.

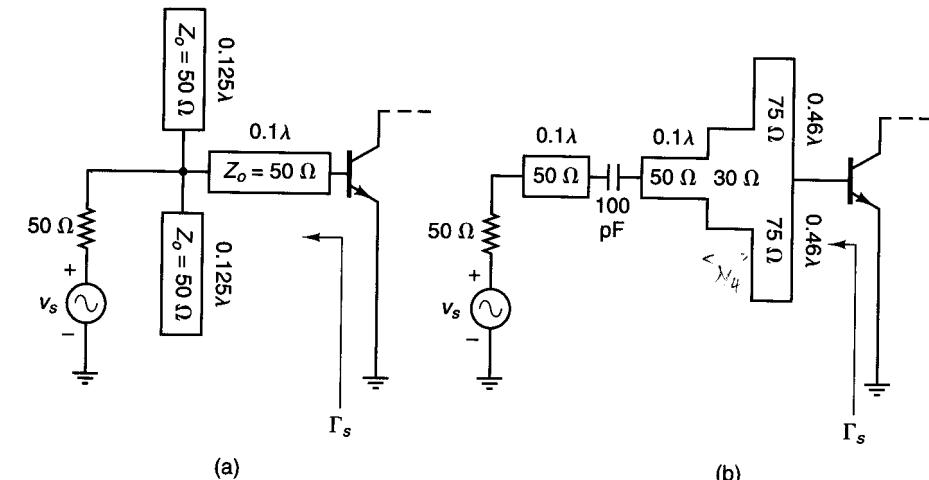


Figure P2.26

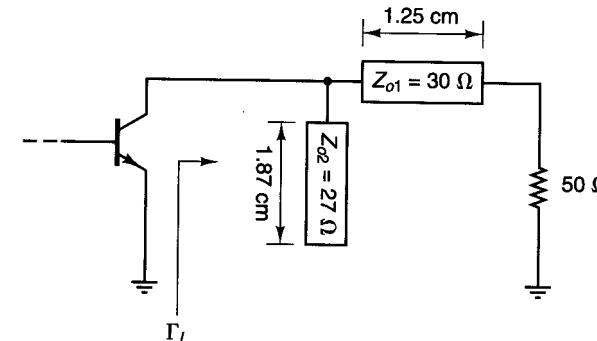


Figure P2.27

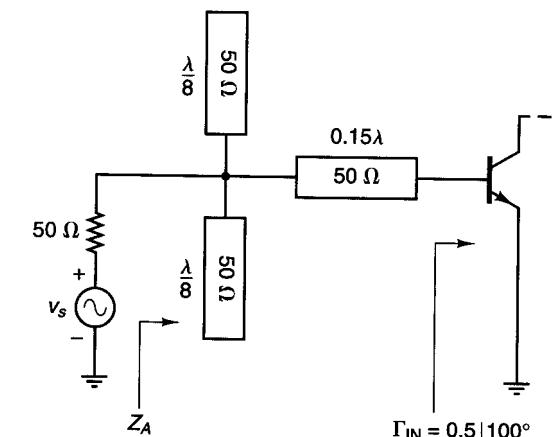


Figure P2.28

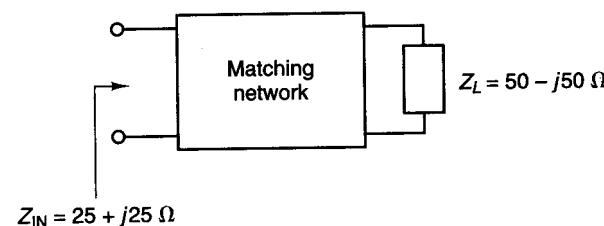


Figure P2.29

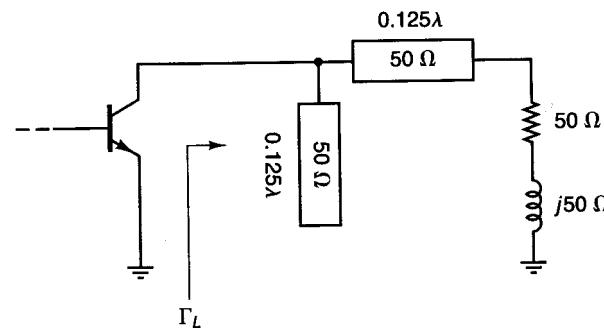


Figure P2.30

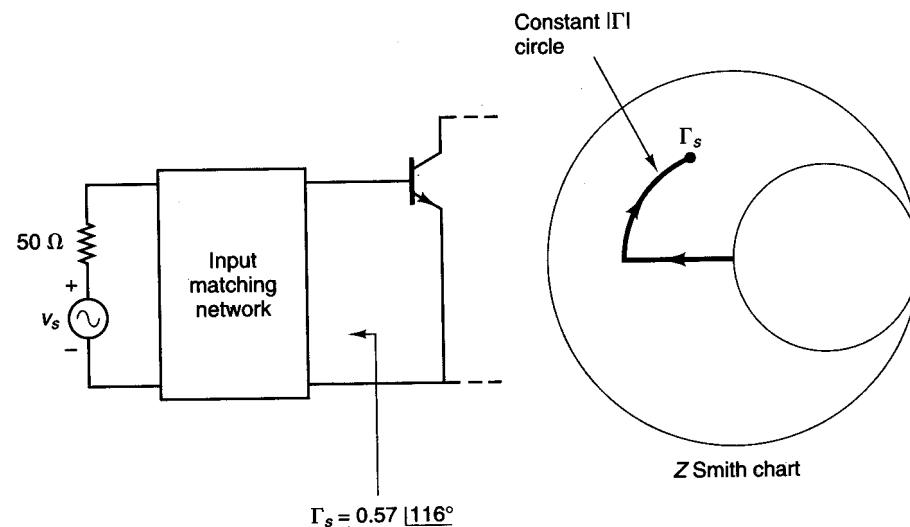


Figure P2.31

2.32 Design the matching networks in Fig. P2.32 to match the load $Z_L = 100 + j100 \Omega$ to a $50\text{-}\Omega$ transmission line.

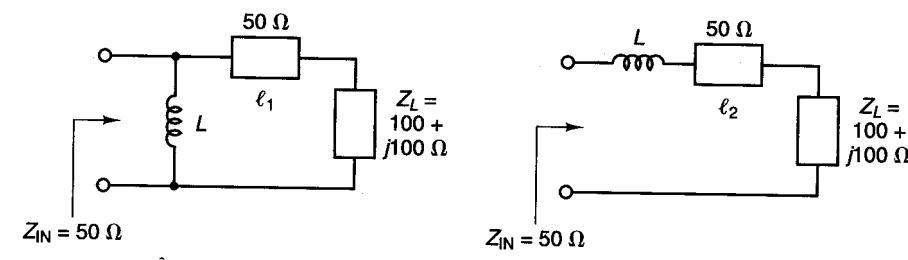


Figure P2.32

2.33 Design the matching networks in Fig. P2.33 to match a $50\text{-}\Omega$ load to the impedance Z_{IN} .

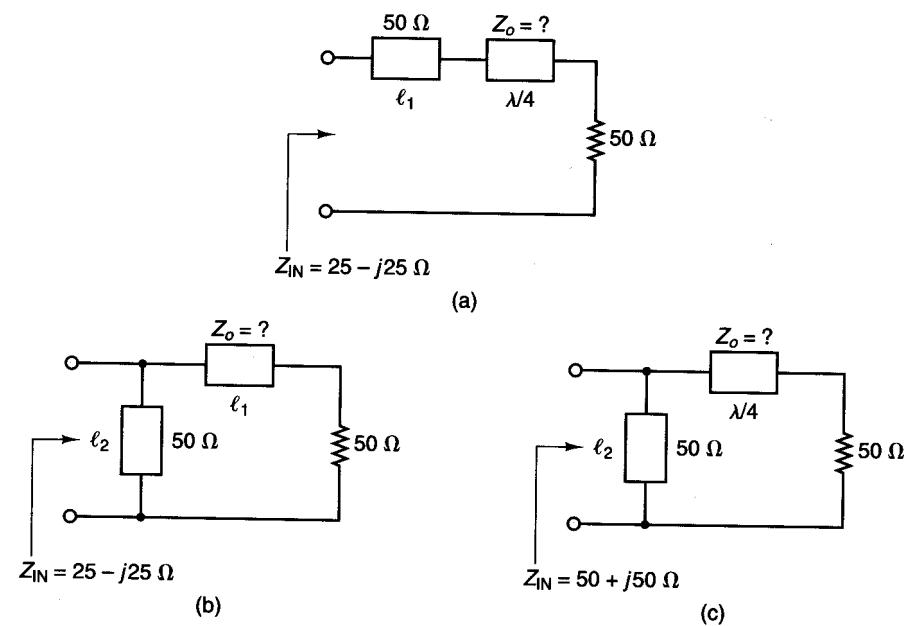


Figure P2.33

- 2.34** Verify the expressions for G_T in (2.6.14) and (2.6.15).
2.34 Verify the expression for A_v in (2.6.23).

- 2.36** (a) The output portion of a microwave amplifier is shown in Fig. P2.36. Calculate $(VSWR)_{out}$ if $\Gamma_{out} = 0.682 \angle -97^\circ$.
 (b) Verify that Z_b is 50Ω when $\Gamma_L = \Gamma_{out}^*$.
 (c) Calculate $(VSWR)_{out}$ if $\Gamma_{out} = 0.5 \angle -60^\circ$.

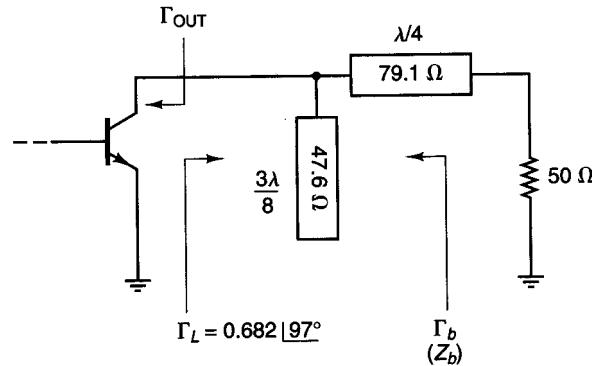


Figure P2.36

REFERENCES

- [2.1] I. J. Bahl and D. K. Trivedi, "A Designer's Guide to Microstrip Line," *Microwaves*, May 1977.
- [2.2] H. Sobol, "Applications of Integrated Circuit Technology to Microwave Frequencies," *Proceedings of the IEEE*, August 1971.
- [2.3] H. Sobol, "Extending IC Technology to Microwave Equipment," *Electronics*, March 1967.

- 2.37** (a) The input portion of a microwave amplifier is shown in Fig. P2.37. Calculate $(VSWR)_{in}$ if $\Gamma_{in} = 0.545 \angle -77.7^\circ$.
 (b) Verify that Z_a is 50Ω when $\Gamma_s = \Gamma_{in}^*$.
 (c) Calculate $(VSWR)_{in}$ if $\Gamma_{in} = 0.4 \angle 45^\circ$.

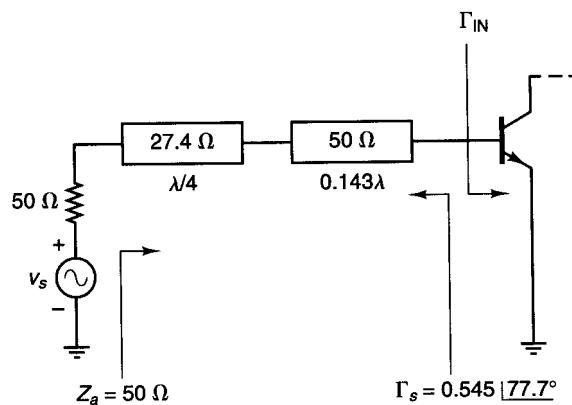


Figure P2.37

Linearization Techniques for CMOS Low Noise Amplifiers: A Tutorial

Heng Zhang, *Student Member, IEEE*, and Edgar Sánchez-Sinencio, *Life Fellow, IEEE*

(Invited Paper)

Abstract—This tutorial catalogues and analyzes previously reported CMOS low noise amplifier (LNA) linearization techniques. These techniques comprise eight categories: a) feedback; b) harmonic termination; c) optimum biasing; d) feedforward; e) derivative superposition (DS); f) IM2 injection; g) noise/distortion cancellation; and h) post-distortion. This paper also addresses broadband-LNA-linearization issues for emerging reconfigurable multiband/multistandard and wideband transceivers. Furthermore, we highlight the impact of CMOS technology scaling on linearity and outline how to design a linear LNA in a deep submicrometer process. Finally, general design guidelines for high-linearity LNAs are provided.

Index Terms—Broadband, CMOS technology scaling, derivative superposition, design guidelines, distortion canceling, feedback, feedforward, harmonic termination, IIP2, IIP3, IM2 injection, linearity, linearization, low noise amplifier (LNA), , multiband, multistandard, noise canceling, 1 dB compression point, optimum biasing, post-distortion, tutorial.

I. INTRODUCTION

THE plethora of wireless communication standards employed in a single geographic region and moreover occupying narrow frequency bands tightly constrains RF-system linearity. Furthermore, the trend in radio research is to simplify/eliminate the expensive front-end module (FEM), which demands a highly linear receiver. In particular, since the low noise amplifier (LNA) is the first block in the receiver chain, it must be sufficiently linear to suppress interference and maintain high sensitivity.

LNA linearization methods should be simple, should consume minimum power, and should preserve noise figure (NF), gain, and input matching. Many traditional linearization techniques are not feasible for LNAs. For example, resistive source degeneration and floating-gate input attenuation reduce the gain and worsen NF or input matching. Hence, LNA linearization proves significantly more challenging than that of baseband circuits [1], often requiring innovative techniques.

Growing research on reconfigurable multiband/multistandard and broadband transceivers such as ultrawideband (UWB) and

Manuscript received December 10, 2009; revised April 20, 2010; accepted June 02, 2010. Date of current version December 30, 2010. This paper was recommended by Editor-in-Chief W. A. Serdijn.

H. Zhang was with the Analog and Mixed-Signal Center, Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX 77843-3128 USA. She is now with Broadcom Corporation, Irvine, CA 92617 USA (e-mail: heng@broadcom.com).

E. Sánchez-Sinencio with the Analog and Mixed-Signal Center, Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX 77843-3128 USA (e-mail: sanchez@mail.ece.tamu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSI.2010.2055353

digital TV tuners has propelled interest in broadband LNA design. Radios in the same platform interfere with each other, and multiple channels applied simultaneously to an LNA without filtering act as in-band interferences. Consequently, broadband LNAs must maintain sufficient linearity over a wide frequency range. Emphasis on highly linear transceivers has sparked recent interest in linearizing LNAs [2]. Even though most previously reported techniques target narrowband applications and principally improve only the third-order intercept point (IIP3), we demonstrate why broadband systems require high second-order intercept point (IIP2) and 1 dB compression point (P_1 dB) as well. Because a broadband LNA is exposed to a wide frequency range, we investigate the dependence of IIP2/IIP3 on two-tone (center) frequency and frequency spacing.

Since LNAs typically have low-amplitude, high frequency inputs, the amplifier operates as a weakly nonlinear system with few relevant harmonics (typically only second and third). Thus, Volterra-series analysis [3] can capture the frequency-dependent distortion of LNAs and provide insight into how to compensate that distortion.

CMOS is the most promising technology for systems on a chip. Although MOSFETs are intrinsically more linear than bipolar transistors [4], they require higher dc current to achieve the necessary transconductance and linearity, thus linearization techniques must be employed to reduce the dc power. Deep submicrometer (DSM) technology challenges include nonlinear output conductance, mobility degradation, velocity saturation, and poly-gate depletion; which complicate CMOS LNA linearization, especially in the face of low supply voltages. We present multidimensional Taylor analysis to evaluate the effects of these nonidealities.

This paper is organized as follows. Section II analyzes previously reported CMOS LNA linearization techniques. Section III discusses new broadband-LNA-linearization issues arising in multiband/multistandard/wideband transceivers. Section IV investigates the impact of CMOS technology scaling on linearity, and provides insights into the design of linear LNAs in DSM processes. Remarks for high linearity LNA design are provided in Section V, and Section VI concludes the paper.

II. LINEARIZATION TECHNIQUES

A weakly nonlinear amplifier with input X and output Y can be approximated by the first three power series terms

$$Y = g_1 X + g_2 X^2 + g_3 X^3 \quad (1)$$

where $g_{1,2,3}$ are the linear gain and the second/third-order nonlinearity coefficients of the amplifier, respectively. The goal of linearization is to make $g_{2,3}$ small enough to be negligible, keeping only the linear term g_1 , hence $Y \approx g_1 X$. The purpose

TABLE I
OVERVIEW OF DISTORTION SOURCES AND LINEARIZATION TECHNIQUES

Distortion Sources	g_m				g_{ds}
	Intrinsic 2 nd -order	Intrinsic 3 rd -order	2 nd -order interaction	Higher order	
Linearization Methods					
Feedback	✓	✓		✓	
Harmonic termination		✓	✓		
Optimal biasing		✓			
Feedforward	✓	✓		✓	
Derivative superposition(DS)		✓			
Complementary DS	✓	✓			
Differential DS	✓	✓			
Modified DS		✓	✓		
IM2 injection		✓	✓		
Noise/distortion cancellation	✓	✓			✓
Post-distortion	✓	✓			

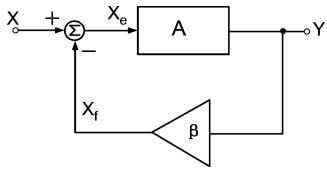


Fig. 1. Nonlinear amplifier with negative feedback.

of this paper is to discuss the main linearization techniques for LNAs.

LNA nonlinearity originates from two major sources:

- 1) nonlinear transconductance g_m , which converts linear input voltage to nonlinear output drain current; this effect is also termed “input limited”;
- 2) nonlinear output conductance g_{ds} , whose effect becomes apparent under large output voltage swing and small drain-source voltage V_{ds} (i.e., when the device operates near linear region); also referred to as “output limited.”

The MOSFET capacitances (the gate-source capacitance C_{gs} , the gate-drain capacitance C_{gd} , the drain-bulk capacitance C_{db}) are roughly linear when the transistor operates in the saturation region, and when the frequency is less than $f_T/10$ [44]. Thus, for the most part, the capacitors contribute less distortion than g_m/g_{ds} [43]; however, C_{gd} influences the linearity indirectly through feedback, which will be discussed later.

The IIP3 is degraded by both the intrinsic third-order distortion and the “second-order interaction” (caused by intrinsic second-order distortion combined with feedback), while IIP2 originates from intrinsic second-order distortion.

We categorize previously reported CMOS LNA linearization techniques into eight clusters: a) feedback; b) harmonic termination; c) optimum biasing; d) feedforward; e) derivative superposition (DS); f) IM2 injection; g) noise/distortion cancellation; and h) post-distortion. Note that DS, IM2 injection, and noise/distortion cancellation are special cases of the feedforward technique.

Table I illustrates the distortion sources and the corresponding linearization methods. Most of the reported linearization techniques focus on suppressing second- and third-order distortion of transconductance. Therefore, linearization of higher order terms (beyond third order) and output conductance still remains an open problem.

A. Feedback

Fig. 1 shows the negative feedback scheme with a nonlinear amplifier A and a linear feedback factor β , where X and Y are the input and output signals, respectively. X_f is the feedback signal, and X_e is the difference between X and X_f .

Assuming the nonlinear amplifier A can be modeled by (1), we obtain the third-order closed-loop power series for Y

$$Y = b_1 X + b_2 X^2 + b_3 X^3 \quad (2)$$

$$b_1 = \frac{g_1}{1 + T_0}$$

$$b_2 = \frac{g_2}{(1 + T_0)^3}$$

$$b_3 = \frac{1}{(1 + T_0)^4} \left(g_3 - \frac{2g_2^2}{g_1} \frac{T_0}{1 + T_0} \right). \quad (3)$$

where $b_{1,2,3}$ are the closed-loop linear gain and second/third-order nonlinearity coefficients, respectively, and $T_0 = g_1\beta$ is the linear open-loop gain. The IIP2 and IIP3 [5] of the amplifier A and the closed loop system are

$$A_{IIP2,\text{amplifier}} = \sqrt{\frac{g_1}{g_2}} \quad (4a)$$

$$A_{IIP2,\text{closeloop}} = \sqrt{\frac{|b_1|}{|b_2|}} = \sqrt{\frac{g_1}{g_2} (1 + T_o)^2} \quad (4b)$$

$$A_{IIP3,\text{amplifier}} = \sqrt{\frac{4}{3} \left| \frac{g_1}{g_3} \right|} \quad (4c)$$

$$\begin{aligned} A_{IIP3,\text{closeloop}} &= \sqrt{\frac{4}{3} \left| \frac{b_1}{b_3} \right|} \\ &= \sqrt{\frac{4 g_1}{3 g_3} \frac{(1 + T_o)^3}{\left(1 - \frac{2g_2^2}{g_1 g_3} \frac{T_o}{1+T_o} \right)}}. \end{aligned} \quad (4d)$$

Hence, negative feedback improves A_{IIP2} by a factor of $(1 + T_0)$; it also improves A_{IIP3} by a factor of $(1 + T_0)^{3/2}$ when $g_2 \approx 0$. As shown by (4d), nonzero g_2 degrades IIP3 when g_1 and g_3 have opposite signs (this is the case for typical MOSFET biases). This phenomenon is called “second-order interaction” [6]. In other words, whenever the amplifier is in feedback, the third-order nonlinearity originates from two sources:

- 1) intrinsic amplifier third-order nonlinearity;
- 2) “second-order interaction” (originated from intrinsic second-order nonlinearity of the amplifier combined with feedback).

However, feedback linearity improvement is not as effective for LNAs as for baseband circuits because:

- 1) the open loop gain T_0 cannot be large due to stringent LNA gain, noise, and power requirement;
- 2) the second-order nonlinearity contributes to the IM3 indirectly through “second-order interaction.”

To illustrate the “second-order interaction,” we use the inductively source degenerated LNA [5] as an example. Fig. 2(a) presents the circuit, and Fig. 2(b) shows its small-signal model using the notation from Fig. 1. The inductor L_s acts as a frequency-dependent feedback element with $\beta = \omega L_s$, creating a feedback path between the output current i_d and the gate-source voltage v_{in} . For simplicity, we analyze these effects with a Taylor series—for a more accurate, frequency-dependent analysis refer to the results obtained using Volterra series in [11], [14], [23].

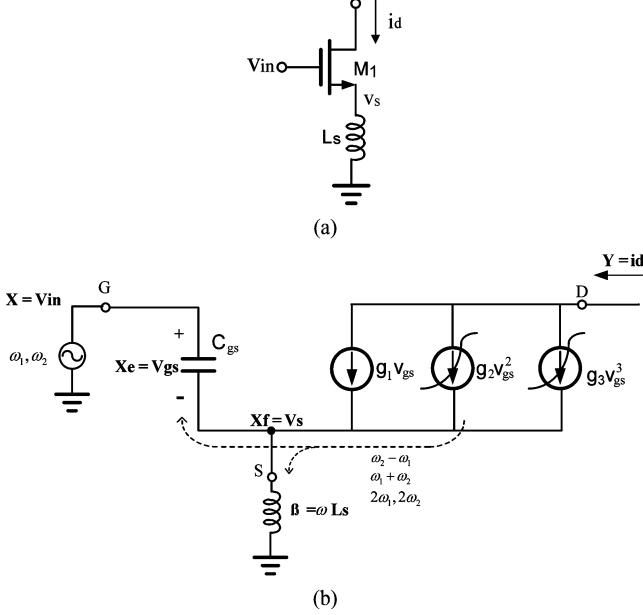


Fig. 2. (a) Inductively source-degenerated LNA. (b) Small-signal model.

First, i_d can be expressed as

$$i_d = g_1 (v_{in} - v_s) + g_2 (v_{in} - v_s)^2 + g_3 (v_{in} - v_s)^3. \quad (5)$$

where g_i is the i^{th} -order coefficient of M_1 obtained by taking the derivative of the drain-source dc current I_{DS} with respect to the gate-to-source voltage V_{GS} at the dc bias point

$$g_1 = \frac{\partial I_{DS}}{\partial V_{GS}}, \quad g_2 = \frac{1}{2!} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2}, \quad g_3 = \frac{1}{3!} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3}. \quad (6)$$

Assume input v_{in} has two frequency components ω_1 and ω_2 ; thus v_s contains components $2\omega_1$, $2\omega_2$, and $\omega_1 \pm \omega_2$ due to the second-order distortion. The product term $-2g_2 v_{in} v_s$ from $g_2 (v_{in} - v_s)^2$ generates IM3 terms $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm \omega_1$. Therefore, the intrinsic second-order nonlinearity contributes to third-order intermodulation, IM3, when a feedback mechanism is employed. Note that this “second-order interaction” problem exists even if the LNA topology is differential because the term $-2g_2 v_{in} v_s$ is an odd term and cannot be rejected by differential operation.

Though source degeneration mostly improves linearity, inductive source degeneration actually has two opposing effects on linearity:

- 1) increases A_{IIP3} by $\approx (1 + g_1 \omega L_s)^{3/2}$;
- 2) degrades A_{IIP3} due to “second-order interaction.”

Fig. 3 shows A_{IIP3} versus source-degeneration inductor L_s for two cases: input tones at 2.4 GHz, 2.41 GHz, and at 5 GHz, 5.01 GHz. Note that this simulation only includes the distortion from input transconductance, while the loading and input-

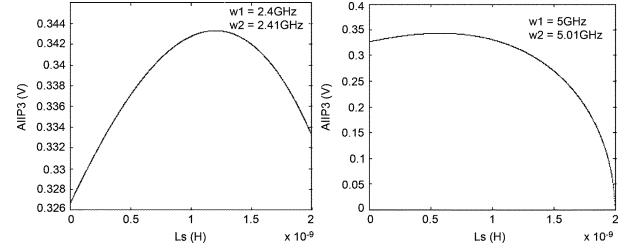


Fig. 3. A_{IIP3} versus source-degeneration inductance.

matching resonant network will also contribute to distortion in practice.

Reducing the degeneration inductance or adding a termination network such that $v_s = 0$ at the IM2 frequency can mitigate “second-order interaction;” the latter approach is called “harmonic termination.”

B. Harmonic Termination

“Harmonic termination” adds a termination network to accomplish one of two effects:

- 1) **BJT case:** sets $b_3 = 0$ in (3) with the “second-order-interaction” term;
- 2) **CMOS case:** forces a certain node voltage to zero at the IM2 frequency.

Equation (3) was obtained assuming a frequency-independent feedback factor β , which is only valid for pure resistive networks. For frequency-dependent networks as the case in Fig. 2, Volterra series [3] should be used to capture the memory effects. To obtain the third-order coefficient in the Volterra series, defined as $b_3(\omega_x, \omega_y, \omega_z)$, a three-dimensional Fourier transform is performed on the third-order impulse response $h_3(\tau_x, \tau_y, \tau_z)$ of the system. Thus, (3) becomes (7), as shown at the bottom of the page. $T(\omega) = g_1 \beta(\omega)$ is the frequency-dependent linear loop gain, which also depends on the feedback components and termination impedances Z_1 , Z_2 , and Z_3 shown in Fig. 4. The expressions $|b_3(\omega_x, \omega_y, \omega_z)|$ and $\angle b_3(\omega_x, \omega_y, \omega_z)$ give the magnitude and phase of a tone at frequency $\omega_x + \omega_y + \omega_z$ generated by third-order nonlinearity. For example, given two input tones at ω_1 & ω_2 , to get the IM3 products at $2\omega_1 - \omega_2$, choose $\omega_x = \omega_y = \omega_1$ and $\omega_z = -\omega_2$. Assuming two closely spaced tones, i.e., $\omega_1 = \omega$, $\omega_2 = -\omega - \Delta\omega$, and $\Delta\omega \approx 0$, we obtained (8), as shown at the bottom of the next page. From (8), the contribution of second-order distortion to IM3 is defined by the loop gain at subharmonic frequency $\Delta\omega$ and second-harmonic frequency 2ω , i.e., $T(\Delta\omega)$ and $T(2\omega)$. Therefore, by tuning the termination impedances at $\Delta\omega$ and/or 2ω , the amplitude and phase of the second-order interaction terms A_2 can be adjusted to cancel the intrinsic third-order distortion term g_3 , so that $b_3 \approx 0$. For narrowband applications, $\Delta\omega$ and 2ω are usually out-of-band, keeping the in-band operation unaffected, hence the “harmonic termination” technique is also called “out-of-band tuning/termination” [7], [8].

$$b_3(\omega_x, \omega_y, \omega_z) = \frac{1}{(1 + T(\omega_x + \omega_y + \omega_z))(1 + T(\omega_x))(1 + T(\omega_y))(1 + T(\omega_z))} \times \left[g_3 - \frac{2g_2^2}{3g_1} \left(\frac{T(\omega_y + \omega_z)}{1 + T(\omega_y + \omega_z)} + \frac{T(\omega_x + \omega_z)}{1 + T(\omega_x + \omega_z)} + \frac{T(\omega_x + \omega_y)}{1 + T(\omega_x + \omega_y)} \right) \right] \quad (7)$$

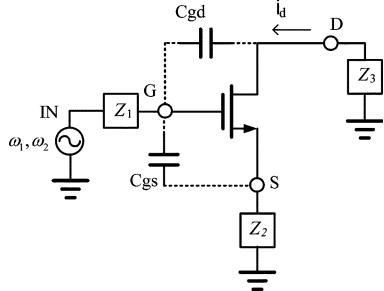


Fig. 4. Common-source LNA with termination impedances.

TABLE II
THREE INTRINSIC FEEDBACK PATHS.

Feedback Path	Path Components	
	CS-LNA	CG-LNA
Source-to-Gate	$C_{gs} + \text{source degeneration inductor } Z_2$	$C_{gs} + \text{input driving impedance}$ [10]
Drain-to-Gate	$C_{gd} + \text{output load } Z_3$	
Input-to-Gate	Input matching network Z_1	

The second-order nonlinear current can mix with the input through three intrinsic feedback paths, as listed in Table II for the common source LNA(CS-LNA) and common gate LNA(CG-LNA).

The CG-LNA inherently has less drain-to-gate feedback than the CS-LNA since its gate is ac grounded, therefore the CG-LNA usually has better linearity.

Resonant tanks can be added to optimally tune $Z_i(\Delta\omega)$ and/or $Z_i(2\omega)$ such that the second-order remixing term cancels the IM3 term. Techniques have been reported to tune the input terminal $Z_1(\Delta\omega)$ for bipolar LNAs [7]–[9]. The terminations are commonly implemented with dedicated LC networks, which provide high impedance at ω but small impedance paths to ground at $\Delta\omega$ or 2ω . However, the required inductance value is usually quite large. The low Q factors of the on-chip passive inductors limit their distortion-cancellation effectiveness and also affect noise and input matching. Furthermore, on-chip active inductors add noise and nonlinearity, so in practice off-chip inductors are employed.

Though popular in BJT LNAs, harmonic termination is less effective for CMOS LNAs [8], [12]. For a stable design, the A_2 term in (8) has a positive real part. Thus, $|\varepsilon(\Delta\omega, 2\omega)|$ can be reduced below $|g_3|$ only if g_3 is positive, which is true for a BJT, but not for a MOSFET in saturation. Therefore, both g_3 and A_2 must be reduced to improve a CMOS LNA's IIP3.

One way to reduce A_2 is to reduce both $Z_i(\Delta\omega)$ and $Z_i(2\omega)$ [13]. A cascode configuration can reduce Z_3 to $1/g_1$ [13], and

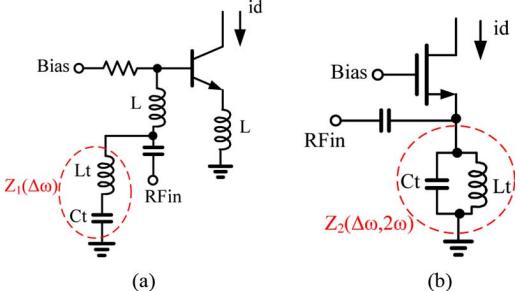


Fig. 5. Harmonic termination: (a) Common-emitter stage with low-frequency-trap network (L is the package inductance) [9]. (b) Common-gate stage with "RF current source" [13].

capacitive cross coupling in the cascode stage further reduces Z_3 to $1/(2g_1)$ [14]. Although their IIP3 improvement is not as great as that attainable from large passive LC components, it is more feasible. In [10] and [13] an LC-resonant RF current source reduces Z_2 . Fig. 5 shows some termination examples, in which L_t and C_t form low-frequency/second-harmonic trap networks $Z_1(\Delta\omega)$ [Fig. 5(a)] and $Z_2(\Delta\omega, 2\omega)$ [Fig. 5(b)].

Harmonic termination only works well in narrowband systems because the tuning network is optimized at $\Delta\omega$ and 2ω , and only works for a narrow range of two-tone spacing/center frequencies [7]. For wideband applications, $\Delta\omega$ and 2ω vary considerably, so it is difficult to tune out the termination impedance. Furthermore, $\Delta\omega$ and 2ω may fall in-band, affecting the normal operations.

In summary, to improve CMOS LNA linearity, we should ensure a small intrinsic third-order coefficient g_3 of the transistor, and relax the "second-order interaction." Adding a harmonic termination network alleviates the latter. Next, we will discuss a few techniques to reduce the third-order coefficient g_3 .

C. Optimal Biasing

Assume the main nonlinearity of a MOS transistor arises from transconductance nonlinearity, as modeled in (5). To characterize this single-transistor nonlinearity, we fixed its drain-source voltage V_{ds} , swept the gate-source voltage V_{gs} , and then took the first three derivatives of the drain-source dc current I_{ds} with respect to V_{gs} [as defined in (6)] at every dc bias point to obtain the plots in Fig. 6. While g_2 is always positive, g_3 has a sign inversion.

- Small V_{gs} : $g_3 > 0$ because the transistor operates in weak inversion, where the I_{ds} vs V_{gs} relation is exponential.
- Large V_{gs} : $g_3 < 0$ because mobility degradation/velocity saturation cause gain compression. The key idea of "optimum biasing" is to bias the transistor at the "sweet spot" $g_3 = 0$ [44].

$$b_3(\omega, \omega, -\omega - \Delta\omega) \cong \frac{1}{(1 + T(\omega))^3 (1 + T(-\omega))} \times \left[g_3 - \underbrace{\frac{2g_2^2}{3g_1} \left(\frac{2T(\Delta\omega)}{1 + T(\Delta\omega)} + \frac{T(2\omega)}{1 + T(2\omega)} \right)}_{A_2} \right] \underbrace{\varepsilon(\Delta\omega, 2\omega)}_{\varepsilon(\Delta\omega, 2\omega)}. \quad (8)$$

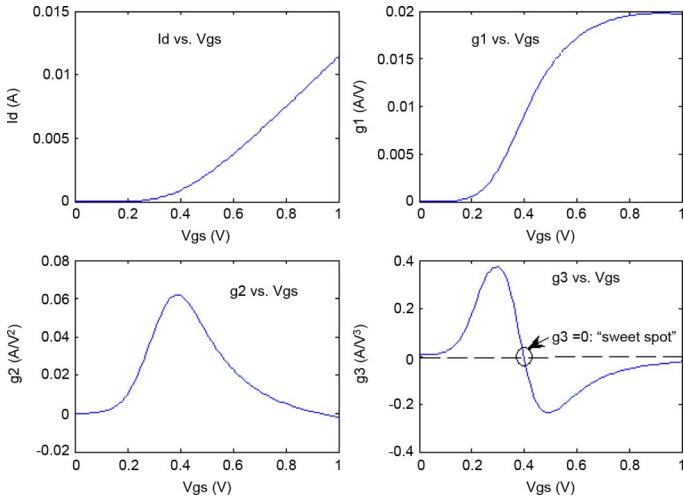


Fig. 6. NMOS transconductance characteristics (UMC 90 nm CMOS process, $W/L = 20/0.08 \mu\text{m}$, $V_{ds} = 1 \text{ V}$).

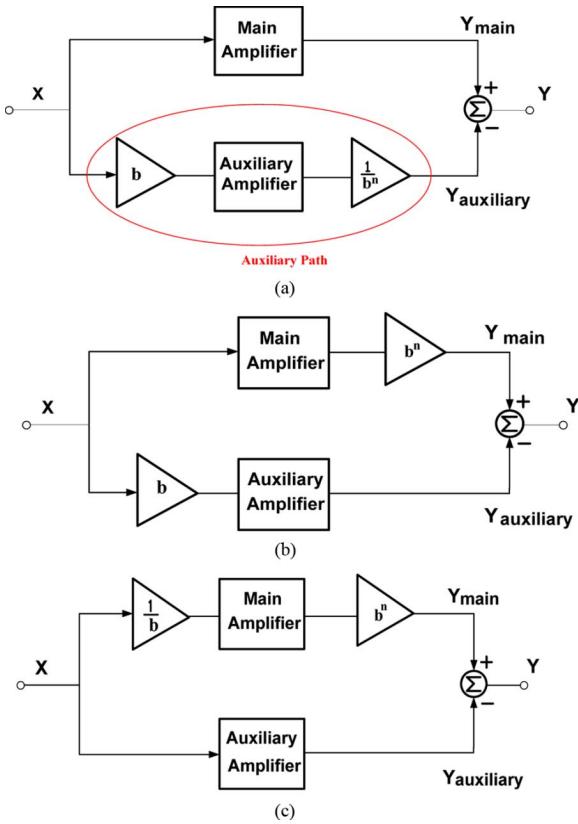


Fig. 7. Three representations of the feedforward linearization technique.

Though simple in principle, the optimal biasing technique has the following limitations.

- 1) The cancellation is sensitive to process variations (e.g., V_{th}), so we recommend constant-current or constant-gm biasing over constant-voltage biasing.
- 2) The technique is sensitive to operating point, resulting in a limited input-signal amplitude range for effective distortion cancellation.
- 3) The sweet spot shifts to a lower bias current level as the gain increases, since the output swing increases and nonlinear output conductance starts to play a role.

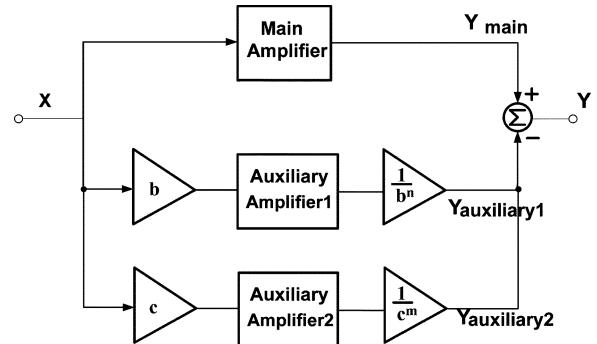


Fig. 8. Proposed dual-auxiliary-path feedforward linearization technique.

- 4) Due to the “second-order interaction,” the IIP3 peak at the “sweet spot” decreases and will finally disappear as source degeneration inductance increases.
- 5) The sweet spot is frequency-dependent, and the IIP3 peak decreases due to parasitic effects [44].
- 6) Biasing the transistor at $g_3 = 0$ restricts the input-stage transconductance, lowering gain and increasing NF.

An automatic bias circuit could mitigate some of these effects [15]; however, this “automatic” bias circuit itself is prone to process variations and requires manual tuning in practice. The bias point for optimum IIP3 is shifted from the bias for zero g_3 due to “second-order interaction.”

In summary, the “sweet spot” is a single transistor characteristic and only signifies optimum intrinsic third-order transconductance nonlinearity. Many other factors will weaken the IIP3 improvement at the “sweet spot.” Furthermore, some claim that no “sweet spot” exists in practical LNAs because of input/output networks and parasitics [44].

D. Feedforward

From (5), note that simultaneous cancellation of g_2 and g_3 with minimum effects on g_1 requires more degrees of freedom. Generating additional nonlinear currents/voltages, and subsequently summing (subtracting) them accomplishes such simultaneous cancellation. These actions constitute feedforward, as illustrated in Fig. 7(a) [16]. An auxiliary path includes a replica amplifier and signal-scaling factors b and $1/b^n$ at its input/output, respectively, to replicate the distortion in the main path. We use $n = 2$ or 3 depending on whether IM2 or IM3 is to be canceled. Note that if the amplifiers are differential, the second-order distortion is ideally zero, and $n = 3$ yields a linear output. Without loss of generality, the following discusses the single-ended case. To obtain the total output Y , we subtract the output of the auxiliary amplifier ($Y_{auxiliary}$) from that of the main amplifier (Y_{main}). Assuming $|b| > 1$, by changing the location and value of scaling factor, we propose two alternate implementations, shown in Fig. 7(b) and 7(c), respectively.

Assuming the main and auxiliary amplifiers have the same nonlinearity coefficients g_i , we have

- a) See equation (9)–(11) at the bottom of the next page.
- b) See equation (12)–(14) at the bottom of the next page.
- c) See equation (15)–(17) at the bottom of the next page.

In these equations, $g_2 = 0$ for differential amplifiers. Comparing (11), (14), and (17), the implementation in a) has a gain-attenuation factor of $(1 - 1/b^{n-1})$, thus gain is reduced by 2.5 dB with $b = 2$ and $n = 3$ as in [16]. On the other hand, the proposed implementations in b) and c) increase the gain. Note

that c)'s input attenuator 1/b worsens its NF. The implementations in Fig. 7 can only cancel one type of harmonic at a time; to reduce both second- and third-order distortion simultaneously, we need an additional degree of freedom, which we could attain with two auxiliary paths as shown in Fig. 8.

Assuming the main and auxiliary amplifiers have the same nonlinearity coefficients g_i , we have (18)–(20), at the bottom of the page. In (20), we have two equations (second and third term equals zero) and four variables (b, c, n, m), resulting in multiple solutions. A possible additional constraint is to bound the reduction in linear gain to be less than, say, 20%, and one reasonable solution set is: $b = -2, c = -3, n = 0, m = 1$. This choice causes the linear gain to double.

Note that if the amplifiers are differential, all even order harmonics are ideally zero, and the implementation in Fig. 8 can cancel both third- and fifth-order distortion.

This general feedforward technique improves linearity without relying on the amplifier's linearity characteristics; however, it has several disadvantages.

- 1) Accurate, noiseless, and highly linear scaling factors (b, c) are often not feasible. For instance, the off-chip coaxial assembly used in [16] is expensive and cannot be integrated.
- 2) The added active components introduce more noise.

3) Highly sensitive to mismatch between the main and auxiliary gain stages.

4) Power is doubled or tripled since the auxiliary amplifier is an exact copy of the main amplifier.

[17] reports an improved feedforward technique, where the auxiliary path only passes the IM3 products. Hence, its dynamic range is relaxed, resulting in only 21% power overhead. Next, we will discuss three special cases of the feedforward technique: derivative superposition, IM2 injection, and noise/distortion cancellation.

E. Derivative Superposition (DS)

The derivative superposition (DS) method [10], [11], [13], [18]–[20] is a special case of the feedforward technique. Notice that the DS method is obtained when $b = 1$ in Fig. 7 and when the main/auxiliary amplifiers are implemented with transistors operating in different regions. Fig. 9(a) depicts a dual-NMOS implementation of the DS method. $M_{A/B}$ denotes the main/auxiliary transistor, respectively, and the input matching network is omitted for simplicity.

This method is called “derivative superposition” because it adds the third derivatives (g_3) of drain current from the main and auxiliary transistors to cancel distortion. As discussed in

$$Y_{\text{main}} = g_1 X + g_2 X^2 + g_3 X^3 \quad (9)$$

$$Y_{\text{auxiliary}} = \left[g_1 (bX) + g_2 (bX)^2 + g_3 (bX)^3 \right] \frac{1}{b^n} \quad (10)$$

$$\begin{aligned} Y &= Y_{\text{main}} - Y_{\text{auxiliary}} \\ &= g_1 \left(1 - \frac{1}{b^{n-1}} \right) X + g_2 \underbrace{\left(1 - \frac{1}{b^{n-2}} \right) X^2}_{\text{Residue Distortion}} + g_3 \underbrace{\left(1 - \frac{1}{b^{n-3}} \right) X^3}_{\text{Residue Distortion}}. \end{aligned} \quad (11)$$

$$Y_{\text{main}} = (g_1 X + g_2 X^2 + g_3 X^3) b^n \quad (12)$$

$$Y_{\text{auxiliary}} = g_1 (bX) + g_2 (bX)^2 + g_3 (bX)^3 \quad (13)$$

$$\begin{aligned} Y &= g_1 b \left(b^{n-1} - 1 \right) X + \underbrace{g_2 b^2 \left(b^{n-2} - 1 \right) X^2 + g_3 b^3 \left(b^{n-3} - 1 \right) X^3}_{\text{Residue Distortion}} \end{aligned} \quad (14)$$

$$Y_{\text{main}} = \left[g_1 \frac{X}{b} + g_2 \left(\frac{X}{b} \right)^2 + g_3 \left(\frac{X}{b} \right)^3 \right] b^n \quad (15)$$

$$Y_{\text{auxiliary}} = g_1 X + g_2 X^2 + g_3 X^3 \quad (16)$$

$$\begin{aligned} Y &= g_1 \left(b^{n-1} - 1 \right) X + \underbrace{g_2 \left(b^{n-2} - 1 \right) X^2 + g_3 \left(b^{n-3} - 1 \right) X^3}_{\text{Residue Distortion}} \end{aligned} \quad (17)$$

$$Y_{\text{auxiliary1}} = \left[g_1 (bX) + g_2 (bX)^2 + g_3 (bX)^3 \right] \frac{1}{b^n} \quad (18)$$

$$Y_{\text{auxiliary2}} = \left[g_1 (cX) + g_2 (cX)^2 + g_3 (cX)^3 \right] \frac{1}{c^m} \quad (19)$$

$$\begin{aligned} Y &= Y_{\text{main}} - Y_{\text{auxiliary1}} - Y_{\text{auxiliary2}} \\ &= g_1 \underbrace{\left(1 - \frac{1}{b^{n-1}} - \frac{1}{c^{m-1}} \right)}_{\text{Linear Gain}} X + \underbrace{g_2 \left(1 - \frac{1}{b^{n-2}} - \frac{1}{c^{m-2}} \right) X^2 + g_3 \left(1 - \frac{1}{b^{n-3}} - \frac{1}{c^{m-3}} \right) X^3}_{\text{Residue Distortion}}. \end{aligned} \quad (20)$$

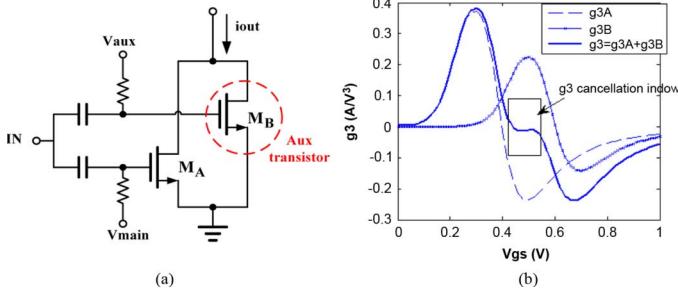


Fig. 9. (a) DS method with dual-NMOSs. (b) Third-order distortion terms of the main transistor (g_{3A}), auxiliary transistor (g_{3B}), and total output (g_3) (UMC 90 nm CMOS process, $(W/L)_{MA} = 20/0.08 \mu\text{m}$, $(W/L)_{MB} = 12/0.08 \mu\text{m}$, $V_{ds} = 1 \text{ V}$).

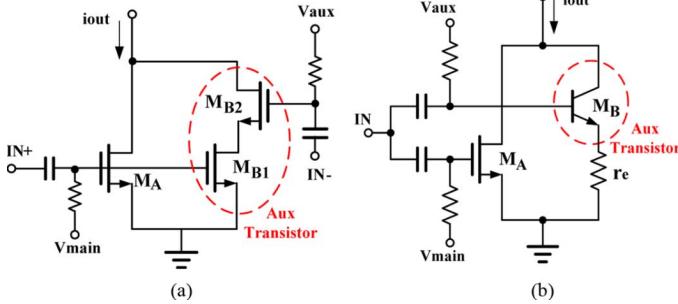


Fig. 10. DS method. (a) Additional transistor works in triode region [18]. (b) Use of a bipolar transistor [19].

Section II-C, g_3 's sign changes at the boundary of moderate and strong inversion region. Thus, proper biasing creates net zero g_3 , as shown in Fig. 9(b). Linearity is improved within a finite bias-voltage range instead of just a point. Hence the DS method is less sensitive to process variations than the optimum biasing technique. Moreover, the auxiliary path contains only one weak-inversion transistor, resulting in much smaller power consumption than the general feedforward technique. Since the DS method employs multiple transistors in parallel with their gates connected together, it is also called the “multiple gated transistor technique” (MGTR) [10], [11], [13]. Note that since positive and negative characteristic of g_3 are not symmetric, the g_3 -cancellation window is fairly narrow with only one auxiliary transistor, but the window widens with more auxiliary transistors at the cost of degraded input matching, NF, and gain [11], [20].

Fig. 10(a) and 10(b) show alternate implementations of the DS method that use a triode region [18] or bipolar [19] transistor as the auxiliary device. In Fig. 10(a), M_{B1} and M_{B2} are driven by differential input signals. M_{B1} is biased in deep triode region, and M_{B2} helps to boost the positive g_3 peak of M_{B1} to be sufficiently large to cancel the negative peak in g_3 of input transistor M_A. In Fig. 10(b), a bipolar transistor M_B provides the positive g_3 , and emitter degeneration resistor r_e reduces g_3 to match that of M_A for optimum distortion cancellation.

1) *Complementary DS*: Fig. 6 shows that the second-order term (g_2) has a positive sign for transistors working in either moderate or strong inversion region. Therefore, techniques, such as conventional DS, that improve third-order distortion usually worsen second-order distortion. The “complementary DS method” employs an NMOS/PMOS pair to improve IIP3 without hurting IIP2 [21], [28].

Fig. 11 shows the common-source and common-gate implementations. The ac current combiner in Fig. 11(b) could be either a large coupling capacitor with negligible impedance within signal bandwidth [28], or a current mirror

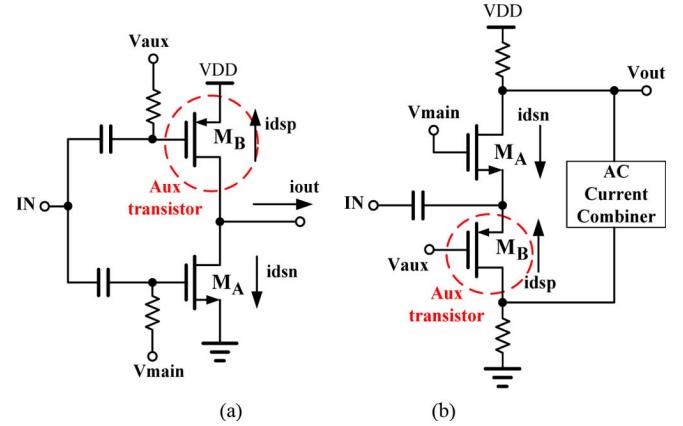


Fig. 11. (a) Complementary DS with common-source configuration [21]. (b) Complementary DS with common-gate configuration [28].

[21]. Since the ac input signal for NMOS/PMOS are out of phase, the output current is expressed as

$$i_{dsn} = g_{1A}V_{gs} + g_{2A}V_{gs}^2 + g_{3A}V_{gs}^3 \quad (21)$$

$$i_{dsp} = -g_{1B}V_{gs} + g_{2B}V_{gs}^2 - g_{3B}V_{gs}^3 \quad (22)$$

$$i_{out} = i_{dsn} - i_{dsp} = (g_{1A} + g_{1B})V_{gs} + (g_{2A} - g_{2B})V_{gs}^2 + (g_{3A} + g_{3B})V_{gs}^3 \quad (23)$$

The total transconductance increases, the IM2 term decreases because g_{2A} and g_{2B} have the same sign, and the IM3 term decreases because g_{3A} and g_{3B} have different signs. Fig. 12 compares the conventional DS and complementary DS in terms of second-order (g_2) and third-order (g_3) distortion of the output current. A cancellation window for g_3 exists in both cases at V_{gs} around 500 mV, but g_2 is maximized for conventional DS and minimized for complementary DS. Note that the g_3 cancellation window is narrower and less flat for complementary DS since PMOS and NMOS devices have different linearity characteristics, so the IIP3 improvement is not as good as that in a dual-NMOS implementation. Furthermore, as shown in (23), we can either match g_{3A} and g_{3B} for a good IIP3 while slightly canceling g_2 , or we can match g_{2A} and g_{2B} for optimum IIP2, because IIP2 and IIP3 do not share the same optimum bias. The differential DS method is essentially the same as complementary DS, which also alleviates IIP2 problem [19], [22].

“Second-order interaction” ultimately limits the IIP3 improvement at higher frequencies after the intrinsic g_3 -induced third-order distortion is canceled by the DS method. The “modified DS method” alleviates this issue [23], [24].

2) *Modified DS*: As discussed in Section II-B, three feedback paths exist for “second-order interaction”: source-to-gate, drain-to-gate, and input-to-gate. The modified DS methods [23], [24] provide an on-chip solution to minimize the source-to-gate feedback. The vector diagram in Fig. 13 graphically explains the modified-DS concept. In conventional DS, the anti-parallel g_{3A} and g_{3B} result in a zero total g_3 , but residual IM3 exists due to g_{2A} contributions (Note: here we neglect g_{2B}). In the modified DS method, g_{3B} is rotated properly such that the composite vector of g_{3A} and g_{3B} contribution is 180° out of phase with the g_{2A} contribution, yielding zero net IM3. Fig. 14(a) shows the circuit implementation of the modified DS method from [23]. Note that choice of L_2 determines the angle of g_{3B} .

Although the channel noise of weak inversion transistor M_B is negligible, its gate-induced noise is inversely proportional to drain current, and is added directly to the main transistor's (M_A) gate noise because their gates are connected together. M_B also

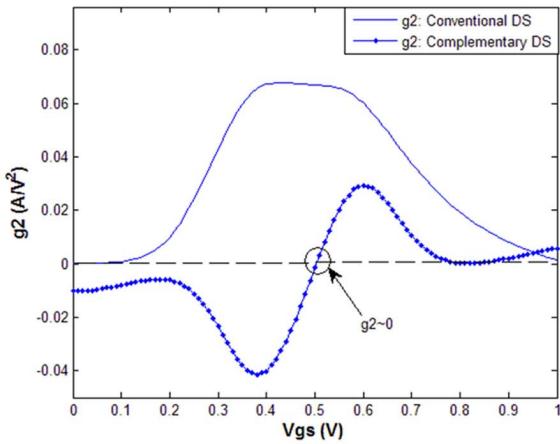


Fig. 12. Comparison of g_2 , g_3 as a function of gate bias for conventional (dual-NMOS) DS and complementary (PMOS/NMOS) DS (UMC 90 nm CMOS process, $V_{ds} = 1$ V).

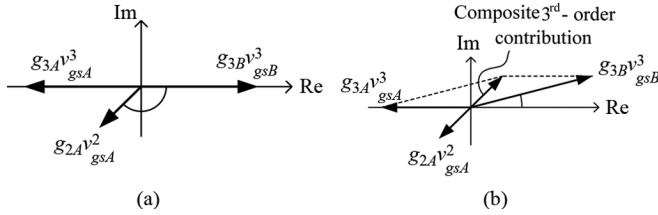


Fig. 13. Vector diagram for the distortion components of: (a) conventional DS method; (b) modified DS method [23].

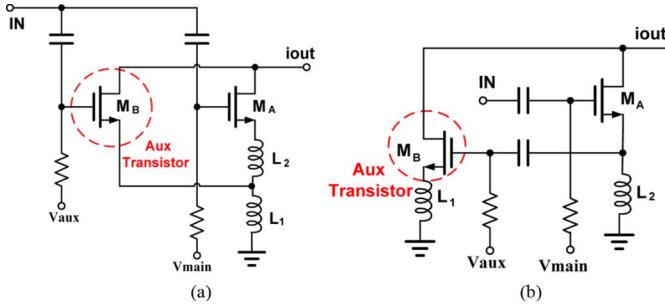


Fig. 14. Circuit implementation of modified DS method [23], [24].

affects the input impedance matching. An alternate implementation of the modified DS method reported in [24] [Fig. 14(b)] moves M_B to the source of M_A instead of directly connecting it to the input, thus minimizing the degradation in NF and input matching.

Limitations of the DS methods include the following.

- 1) The weak-inversion transistor may not operate at sufficiently high frequency.

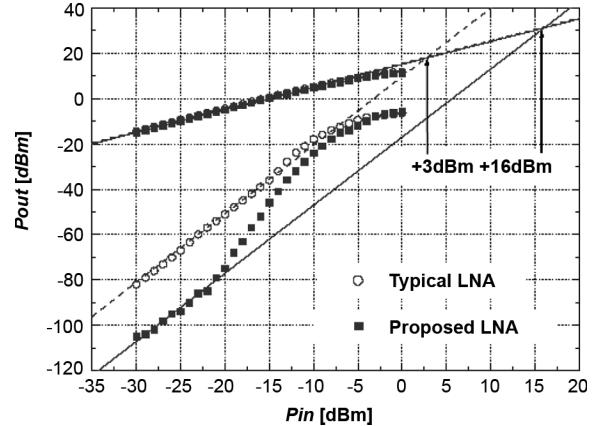


Fig. 15. Measured IIP3 of LNAs with/without DS method [18] (© 2003 IEEE).

- 2) The weak-inversion transistor cannot handle large signals or it will be turned off, resulting in a very limited distortion-cancellation range.
- 3) Weak-inversion transistor models are generally not accurate, resulting in considerable discrepancy between simulation and measurement.
- 4) Triode-region transistors' positive g_3 peaks decrease as technology scales down, thus complicating the task of matching their amplitudes with the negative peaks of g_3 in main transistors.
- 5) Matching transistors working in different regions or matching bipolar with MOS transistors is difficult if not impossible, resulting in a linearity improvement sensitive to PVT variations. Current bias with digital control bits [13] or manual adjustment is required for good results in practice.

Fig. 15 shows an example IIP3 measurement plot [18]. Although it is from a conventional DS method, similar characteristics can be observed with complementary, differential, and modified DS methods. We observe the following.

- 1) The DS method works well within the g_3 -cancellation window annotated in Figs. 9(b) and 12 ($P_{in} < -20$ dBm).
- 2) Even for inputs outside the g_3 -cancellation window, the DS method can still reduce the third-order tone below that of the conventional LNA having a main transistor with negative g_3 as long as g_3 of the auxiliary transistor is positive.
- 3) The third-order curve shows a greater-than-three slope at much smaller input amplitudes after applying the DS method, because the fifth- and higher odd-order-distortion terms contribute more appreciably after g_3 is canceled.
- 4) The DS method does not improve the compression point.

F. IM2 Injection

The IM2 injection method eliminates the explicit auxiliary path entirely by merging it with the main path to reuse the active devices and the dc current [25]. To understand the concept, we first recall (8). To reduce IM3, we should minimize the $\varepsilon(\Delta\omega, 2\omega)$ term. As previously discussed in Section II-B, making A_2 cancel g_3 is difficult for CMOS LNAs because these two terms are out-of-phase in a typical design. Furthermore, the self-generated IM2 term has too small an amplitude to suppress g_3 sufficiently.

The IM2 injection technique externally generates and injects a low-frequency IM2 component into the circuit. The injected IM2 phase is inverted with boosted amplitude for IM3 cancellation. Hence, IM2 injection could also be viewed as a smart

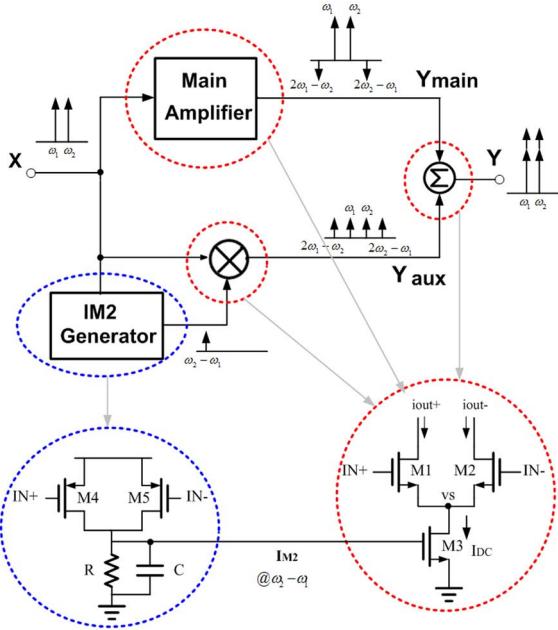


Fig. 16. Block diagram and basic cell implementation of “IM2 injection” [25].

implementation of harmonic termination. Fig. 16 illustrates the concept and basic cells. M1 and M2 are the input transistors of the LNA, and M4, M5, R, and C compose a squaring circuit to generate a low-frequency IM2 current at $\omega_2 - \omega_1$, which is then injected through M3 into the common source node v_s of the LNA. This technique utilizes second-order interaction to generate tones at $2\omega_2 - \omega_1$ and $2\omega_1 - \omega_2$ to cancel the IM3 tones arising from intrinsic third-order distortion. The design equation is

$$\underbrace{-\frac{2g_{1,M1}g_{3,M1}}{4g_{2,M1}} + \frac{3}{2}g_{2,M1}}_{\text{Main Circuit}} = \underbrace{-g_{1,M3} \times 2g_{2,M4}R}_{\text{Squaring Circuit}} \quad (24)$$

where $g_{i,Mi}$ is the i_{th} transconductance coefficient of M_i . The injected IM2 tone should be in phase with the envelope of the RF input signal. Frequency component $\omega_2 - \omega_1$ is chosen over other IM2 components ($\omega_2 + \omega_1, 2\omega_2, 2\omega_1$) because it is easier to match the phase at low frequency. Since the linear gain is added in phase, and the noise injected from the IM2 generator appears as common mode noise (suppressed by differential operation), IM2 injection circumvents gain and NF penalties.

Limitations of IM2 injection include the following.

- 1) NMOS/PMOS transistors and resistors have independent PVT variations—hence more difficult to satisfy the IM3 cancellation criteria in (24) robustly.
- 2) Since R and C in the IM2 generator introduce extra phase shift, two tone spacing must be smaller than the RC-filter cutoff frequency for negligible phase mismatch. Cancellation performance degrades as tone spacing increases.
- 3) Frequency components at $\omega_2 \pm \omega_1$ and $2\omega_{1,2}$ injected by the IM2 generator may fall into signal band and degrade the IIP2.
- 4) Noise from the IM2 generator is negligible only for differential LNAs, but would result in appreciable NF degradation for single-ended LNAs.

In short, IM2 injection applies chiefly to narrowband, differential systems with small two-tone spacing.

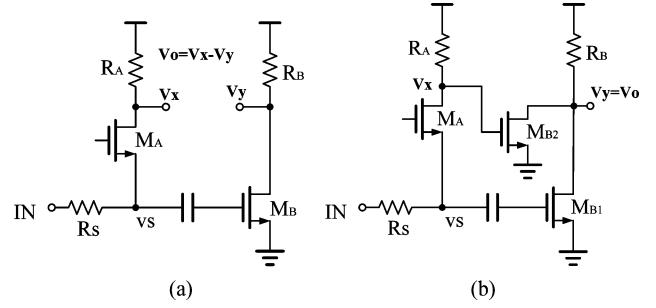


Fig. 17. Noise/distortion cancellation. (a) Differential output [27], [29]. (b) Single ended output [28].

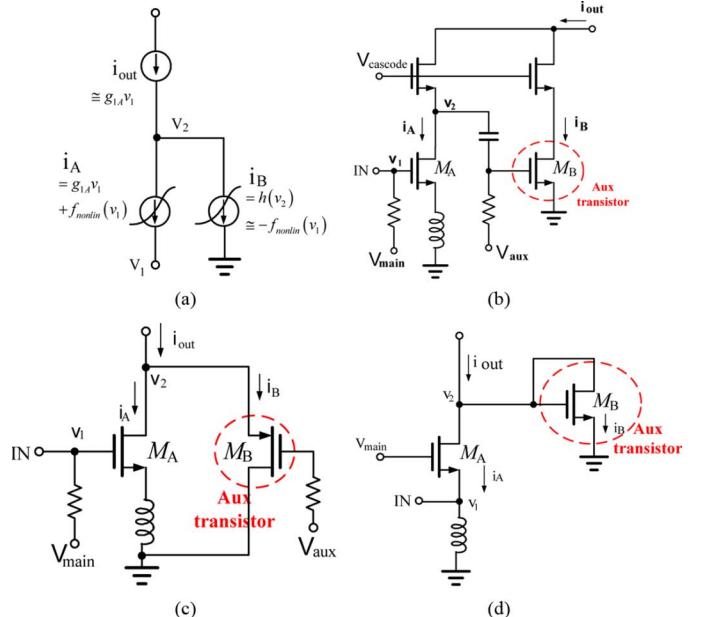


Fig. 18. Post-distortion. (a) Conceptual view. (b) Circuit implementation in [30]. (c) Circuit implementation in [31]. (d) Circuit implementation in [32].

G. Noise/Distortion Cancellation

Noise/distortion cancellation parallels CG (M_A) and CS (M_B) stages, as shown in Fig. 17 [26]–[29]. The circuit is driven by a voltage at node “IN.” The nonlinearity of M_A can be modeled as a current source between its drain and source controlled by both V_{gs} and V_{ds} . Hence, both the channel thermal noise and distortion of M_A flowing through the CG and CS paths are subtracted at the output, whereas the signal is added. Noise/distortion cancellation requires that $V_x = V_y$, i.e.,

$$g_{1,M_A}R_A = g_{1,M_B}R_B \text{ (differential output)} \quad (25a)$$

$$g_{1,M_{B1}}R_s = g_{1,M_{B2}}R_A \text{ (single – ended output).} \quad (25b)$$

Note this technique can cancel all intrinsic distortion generated by M_A , including both g_m and g_{ds} nonlinearity, while previous techniques could only compensate g_m nonlinearity.

After canceling the distortion from M_A , M_B ’s distortion dominates the residual nonlinearity, which comprises two terms: 1) M_B ’s intrinsic third-order distortion and 2) second-order interaction originating from the CG-CS cascade. Optimal biasing of M_B [27], [29] or employing complementary DS [28] could further improve the linearity.

TABLE III
PERFORMANCE COMPARISON OF SILICON-VERIFIED LINEARIZATION TECHNIQUES FOR CMOS LNAs.

Linearization Technique	Harmonic Termination [14]	Optimum biasing [15]	Feedforward [16]	Derivative Superposition [18]	Modified DS [23]	Complementary DS ***[21]	IM2 Injection **[25]	Noise/Distortion Cancellation ***[29]	Post Distortion [32]
*IIP3/ Δ IIP3	-4.4dBm/+2.5dB	+10.5dBm	5dBm/+13dB	2.7dBm/+13.4dB	2dBm/+20dB	3dBm	-10.4dBm/+10.6dB	>0dBm	5dBm/+9dB
*IIP2/ Δ IIP2	N/A	N/A	N/A	N/A	N/A	+44dBm	N/A	>+20dBm	+10dBm/+10dB
*Gain/ Δ Gain	20.4dB/+2dB	14.6dB/0dB	18dB/-2.5dB	15.3dB/-0.4dB	16dB/-0.5dB	14dB	22dB/0dB	13-15.6dB	14.3dB/-1.7dB
*NF/ Δ NF	1.92dB/0dB	1.8dB/0dB	2.6dB/+0.2dB	2.9dB/+0.1dB	1.4dB/+0.25dB	3dB	5.3dB/0dB	<3.5dB	2.7dB/+0.6dB
Power/ Δ Power	16.2mW/0%	5.4mW/0%	22.5mW/+100%	20mW/+17.5%	23.4mW/+3.4%	34.8mW	19.6mW/+0.7%	14mW	2.6mW/+1%
Supply Voltage	1.8V	2.7V	3.0V	2.5V	2.6V	2.2V	1.5V	1.2V	1.3V
Frequency	2.2GHz	880MHz	900MHz	2.2GHz	900MHz	48-1200MHz	900MHz	0.2-5.2GHz	2.5-10GHz
Process	0.35μm	0.25μm	0.35μm	0.25μm	0.25μm	0.18μm	0.18μm	65nm	0.13μm
Robustness over PVT	moderate	poor	good	moderate	moderate	moderate	moderate	good	good
Wideband?	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

H. Post-Distortion

Similar to the DS method, the post-distortion (PD) technique also utilizes an auxiliary transistor's nonlinearity to cancel that of the main device, but it is more advanced in two aspects.

- 1) The auxiliary transistor is connected to the output of main device instead of directly to the input, minimizing the impact on input matching.
- 2) All transistors operate in saturation, resulting in more robust distortion cancellation.

Fig. 18 displays a conceptual diagram of PD as well as three implementations [30]–[32]. The auxiliary transistor M_B taps voltage v_2 and replicates the nonlinear drain current of the main transistor M_A , partially canceling both second- and third-order distortion terms. The nonlinear drain currents of M_A and M_B can be modeled as

$$i_A = g_{1A}v_1 + \underbrace{g_{2A}v_1^2 + g_{3A}v_1^3}_{f_{\text{nonlin}}(v_1)} \quad (26)$$

$$i_B = g_{1B}v_2 + g_{2B}v_2^2 + g_{3B}v_2^3. \quad (27)$$

Next, suppose v_2 is related to v_1 by

$$v_2 = -b_1v_1 - b_2v_1^2 - b_3v_1^3 \quad (28)$$

where $b_1 - b_3$ are generally frequency dependent and can be extracted from simulation. In Fig. 18(a), the cascode devices were assumed to be ideal current buffers [30]. The two nonlinear currents i_A and i_B sum at node v_2 , yielding i_{out} :

$$\begin{aligned} i_{\text{out}} &= i_A + i_B = (g_{1A} - b_1g_{1B})v_1 \\ &\quad + \underbrace{(g_{2A} - b_1^2g_{2B} - b_2g_{1B})v_1^2}_{\text{2nd-order distortion}} \\ &\quad + \underbrace{(g_{3A} - b_1^3g_{3B} - g_1Bb_3 - 2g_2Bb_1b_2)v_1^3}_{\text{3rd-order distortion}}. \end{aligned} \quad (29)$$

Note that in the PD method, both the main and auxiliary transistors operate in saturation with the same $g_{1,2,3}$ polarity. Hence, M_B partially cancels the linear term as well; however, it does not substantially degrade the gain/NF because M_B is designed to be more nonlinear than M_A (i.e., $g_{2,3B}/g_{1B} \gg g_{2,3A}/g_{1A}$). Finally, note that among the three implementations, Fig. 18(b) and 18(d) probably have better performance in practical implementations, since both M_A and M_B are NMOS, which can be

matched very well in layout. In Fig. 18(c), NMOS/PMOS transistors must have commensurate nonlinearity, but are hard to match across PVT.

I. Summary

Table III compares the IIP2/IIP3 improvement and gain/NF/power penalties of the previously discussed, state-of-the-art linearization techniques. We chose only one representative reference for each technique for brevity. The best performance per row has been marked with gray color. The modified DS method achieves the best IIP3(> 20 dBm); the IM2 injection method yields minimum degradation in NF, gain, and power; and the PD method renders robust linearity improvement.

Note that transconductance linearization methods are inherently broadband, however, to apply it on wideband LNAs, we should match the delays and phases from the main and auxiliary paths, including input matching/loading network, at the desired frequency band, so that the distortion cancellation is carried out with sufficient accuracy. Most reported techniques (e.g., IM2 injection, modified DS, and harmonic termination) dealing with second-order interaction are only limited to narrowband applications.

IIP2 calibration is another linearization technique that has been extensively reported for mixers, but still remains an open problem for LNAs. The concept of IIP2 calibration is to sense and correct the dc offset with an analog or digital feedback loop [33]–[36]. Some correction approaches for mixers include adjusting the LO bias [33], the load resistor/capacitor banks [34], the current source load [35], or injecting current at the mixer output [36]. It might be possible to apply some of the methods currently employed in mixers to differential LNAs.

III. NEW ISSUES FOR WIDEBAND APPLICATIONS

Growing research on reconfigurable multiband/multistandard and broadband transceivers has increased interest in broadband LNA design. In these transceivers, hundreds of channels could enter the LNA without any prefiltering, acting as in-band interferers. Moreover, nearby radios and on-chip transmitter leakage cause increased adjacent blockers, creating severe cross modulation, intermodulation, and desensitization. Therefore, a big design challenge for broadband LNAs is to achieve high linearity over a wide frequency range, lest the SNDR at the LNA output be dominated by distortion instead of noise. Furthermore, the old textbook argument that the LNA receives a small input signal amplitude is not valid for wideband LNAs. We consider

three main concerns: IIP2, $P_{1 \text{ dB}}$, and IIP2/IIP3 vs. two-tone frequency and spacing.

A. IIP2

Most linearization methods target narrowband applications and only cancel the third-order distortion, since the second-order nonlinearity is generally out of band in narrowband system. However, for wideband receivers, many channels are present concurrently and act as in-band interferences. Thus, the second-order intermodulation products generated by certain combination of interferences are highly likely to fall into the signal band. Hence, broadband LNAs should have a good IIP2 as well as IIP3. Often, in applications like digital TV, the required IIP2/3 must be derived from a multitone test such as complex second-order distortion (CSO) and composite triple beat distortion (CTB) [21].

A fully differential LNA will improve IIP2, but requires a transformer, which is expensive for wideband systems. Other IIP2 improvement techniques include the complementary/differential DS method [21], [22], [28] and post-distortion [30]–[32]. Moreover, in DSM processes, biasing a CS-stage at the maximum gain yields a high IIP2 [29].

B. 1 dB Compression Point

The “large-signal” 1 dB compression point ($P_{1 \text{ dB}}$) has traditionally not been a major concern for LNA designers because the LNA typically has a small input signal. However, in wideband receivers, LNAs receive the accumulated power from multiple channels, which could range from -10 to 0 dBm. For example, in the A/74 standard developed by the Advanced Television Systems Committee (ATSC), many transmitters are in close spectral proximity, so the receiver is exposed to more multicarrier adjacent energy. The maximum input power (the average of multiple tones) could even exceed 0 dBm [41]. Furthermore, severe transmitter leakage, poor isolation between antennas, and single-tone blockers with large peak-to-average ratio all require a high signal-handling capability, i.e., high $P_{1 \text{ dB}}$, for the LNA to prevent desensitization, gain compression, and clipping.

IIP2/IIP3-improvement techniques typically only work over small signal ranges, and do not improve $P_{1 \text{ dB}}$ because it is a large-signal parameter. At higher input amplitudes clipping occurs, and the $P_{1 \text{ dB}}$ worsens due to limited supply voltage/dc-bias current.

$P_{1 \text{ dB}}$ -improvement techniques include

- 1) Increasing V_{dd} above nominal values to maximize the voltage headroom and performing substantial PVT simulation to guarantee breakdown/overstress will not occur.
- 2) Using low- f_T , thick-oxide transistors to handle larger voltage swings to allow even larger V_{dd} . Using such transistors degrades NF and high-frequency performance and raises cost.

Achieving high $P_{1 \text{ dB}}$ with thin-oxide devices and low supply voltages remains an open problem. Some possible approaches include

- 1) Cancel higher-order distortion, e.g., IM5 & IM7, since these become prominent at larger inputs and contribute to $P_{1 \text{ dB}}$.
- 2) Extend the effective input range of IM2/IM3 cancellation. One solution is to employ more auxiliary transistors in parallel in the DS method [11], [20]. Note that weak-inversion transistors being turned on and off at large voltage swing will add more high-order harmonic components to the circuit. A more robust solution is to combine triode and weak inversion transistors as auxiliary transistors [20].

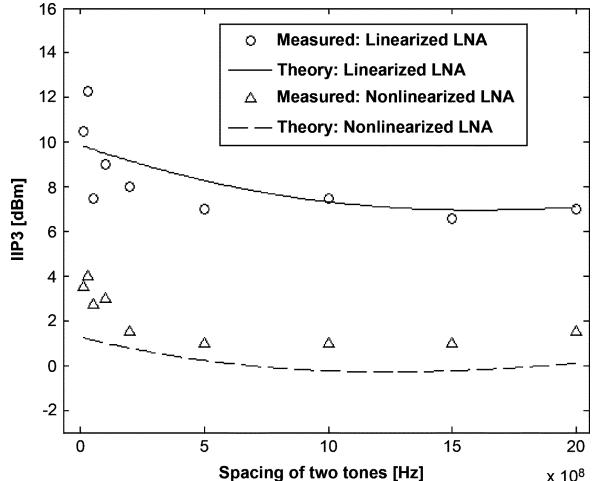


Fig. 19. Experimental and theoretical results of LNA IIP3 as a function of frequency spacing [32] (© 2009 IEEE).

- 3) Add source degeneration at the cost of extra noise.
- 4) Dynamic bias/dynamic supply [42].
- 5) Reduce the output voltage swing to relax the limitation from nonlinear output conductance. One option is to use a low-impedance load for the LNA, for example by choosing a passive mixer over an active mixer as the following stage. However, this choice requires a larger g_m stage and hence greater difficulty to linearize the transconductance.

C. IIP2/IIP3 Vs. Two Tone Frequency and Spacing

Broadband LNAs have flat gain/NF over the whole bandwidth. Likewise, IIP2/IIP3 should also be relatively flat over the signal band. Therefore, while narrowband systems typically use a specific interference frequency and a small tone spacing for the two-tone test, broadband systems require IIP2/IIP3 to be examined at various two-tone-spacing and center frequencies [32]. Fig. 19 shows an example plot.

Reactive components, such as those in the matching network, cause the frequency-dependence of IIP2/IIP3. Note that typically, this frequency-dependence is mild for operating frequencies below 1 GHz, so it is more of a concern for UWB systems (3.1–10.6 GHz) than for digital TV (54–880 MHz), for example.

IIP2 depends on two-tone-spacing. For two input-signal tones at ω_1 & ω_2 , the upper-frequency IM2 component is at $\omega_1 + \omega_2$, while the lower-frequency component is at $\omega_1 - \omega_2$. The IIP2 dependence on two-tone spacing is subtle when $\omega_1 - \omega_2$ is very small. There are two situations in which this dependence becomes more significant.

- 1) Large two-tone spacing, where larger frequency spacing yields correspondingly larger reactive effects.
- 2) Narrowband IM2 cancellation scheme. For example, in the complementary DS method with CG configuration shown in Fig. 11(b), the impedance from coupling capacitors increases with smaller two-tone spacing. Thus the ac-short condition worsens and degrades the IM2-cancellation effectiveness [21], [28].

The IIP3 dependence on two-tone spacing is mainly attributed to the “second-order interaction” as shown in (8). Therefore, the variations of $\Delta\omega$ cause the optimum point of the second-order interaction cancellation to change, resulting in worse linearity. For example, in the IM2-injection method [25] (Fig. 16), the squaring circuit experiences more phase shift at larger two-tone spacing, which degrades IIP3. In the harmonic-termination method [7], IIP3 degrades noticeably at larger $\Delta\omega$.

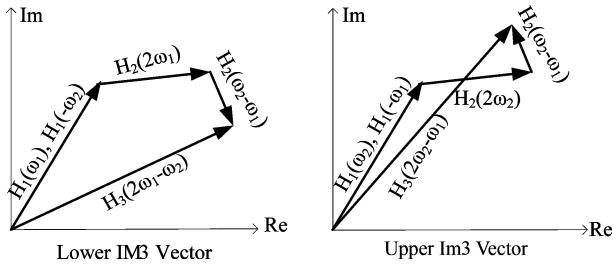


Fig. 20. Vector diagram showing the 180° out-of-phase contribution of $\omega_2 - \omega_1$ term on the upper and lower IM3 components [38].

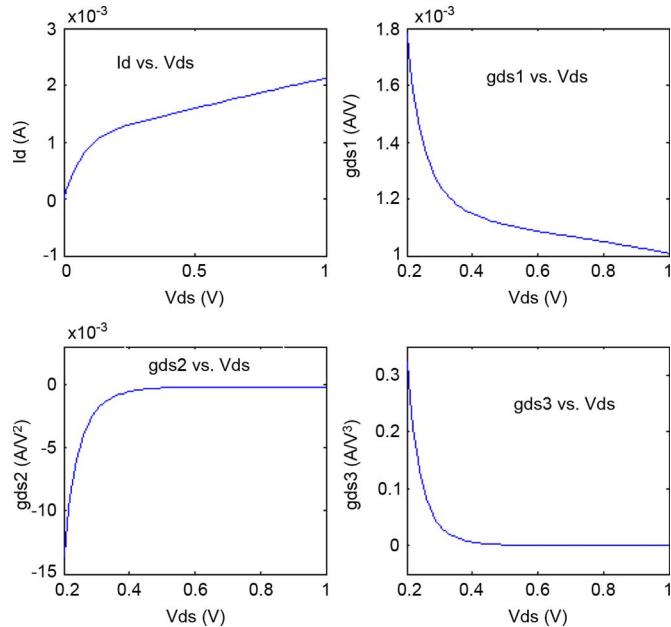


Fig. 21. NMOS output conductance nonlinearity characteristics (UMC 90 nm CMOS process, $W/L = 20/0.08 \mu m$, $V_{gs} = 0.5 V$, $V_{th} = 0.26 V$).

Another major contributor to this IIP₃ dependence is the IM3 asymmetry, also called “sideband asymmetry.” IM3 asymmetry is attributed to various types of memory effects [37]–[40], but for CMOS LNAs specifically, it is because the reactive part of the circuit impedance (e.g., termination impedance) at $\omega_2 - \omega_1$ has a 180°-out-of-phase contribution to the IM3 components at $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$. This concept is qualitatively illustrated by the vector diagram in Fig. 20 [38], where the $H_{1,2,3}$ refers to the first-, second-, and third-order Volterra-Series coefficients. The IM3 components at $(2\omega_2 - \omega_1)$ and $(2\omega_1 - \omega_2)$ have different imaginary parts (i.e., reactance), resulting in IM3 asymmetry.

Note that this IM3 asymmetry depends on bias and frequency. For very small two-tone spacing, it is hard to see any IM3 asymmetry since the reactive-impedance effect at $\Delta\omega$ is negligible; but for larger $\Delta\omega$, the reactive impedances at the second-harmonic frequency also contribute differently to the lower/upper IM3 components, which worsens the IM3 asymmetry [7] and also indicates a more obvious IIP₃ dependence on two-tone spacing. However, proper bias can reduce this IM3 asymmetry [40]. Note that in the multitone case, Adjacent channel power ratio (ACPR) asymmetry is defined correspondingly.

IV. LNA LINEARIZATION IN DSM TECHNOLOGY

A. Nonlinearity From Output Conductance g_{ds}

Distortion of MOS transistors is mainly caused by the nonlinear transconductance (g_m) and output conductance (g_{ds}). Previously published linearization techniques mainly focus on linearizing g_m , assuming that (1) drain current i_{ds} is controlled only by the gate-source voltage V_{gs} , and (2) g_{ds} nonlinearity is negligible. These assumptions are valid for small load resistance, small voltage gain, small input signal, and a drain-source voltage (V_{ds}) sufficiently large that the small-signal variation of V_{ds} does not appreciably perturb the bias point.

However, as technology scales down, the g_{ds} nonlinearity becomes more prominent. Current i_{ds} is controlled not only by V_{gs} but also the V_{ds} , which can be approximated by the two-dimensional Taylor series [29], [44]

$$\begin{aligned} i_{ds}(V_{gs}, V_{ds}) = & g_1 V_{gs} + g_2 V_{gs}^2 + g_3 V_{gs}^3 \\ & + g_{ds1} V_{ds} + g_{ds2} V_{ds}^2 + g_{ds3} V_{ds}^3 \\ & + c_{(1,1)} V_{gs} V_{ds} + c_{(2,1)} V_{gs}^2 V_{ds} \\ & + c_{(1,2)} V_{gs} V_{ds}^2 \end{aligned} \quad (30)$$

where g_i is the i^{th} -order transconductance as defined in (6); g_{dsi} represents the nonlinear output conductance effect which is proportional to the i_{ds} derivatives with respect to V_{ds} ; $c_{(m,n)}$ is the cross modulation term describing the dependence of g_i on V_{ds} or g_{dsi} on V_{gs} , as formulated in (31)

$$\begin{aligned} g_{dsi} = & \frac{1}{i!} \frac{\partial^i I_{DS}}{\partial V_{DS}^i} \\ c_{(m,n)} = & \frac{1}{m!n!} \frac{\partial^{m+n} I_{DS}}{\partial V_{GS}^m \partial V_{DS}^n} \end{aligned} \quad (31)$$

To characterize the g_{ds} nonlinearity for a single transistor, we fix its V_{gs} , and sweep the V_{ds} , by taking the first three derivatives of the drain-source dc current i_{ds} with respect to V_{ds} [as defined in (31)] at every dc bias point, we can obtain Fig. 21. It is observed that the drain current is modulated a lot by V_{ds} . g_{ds3} is large when the transistor operates at small V_{ds} ; while it decreases for large V_{ds} values.

Here we assume a negligible nonlinearity contribution from g_{mb} , otherwise three-dimensional Taylor series should be used instead.

From (30), the distortion is contributed by four parts.

- 1) g_m nonlinearity due to nonlinear $i_{ds} - V_{gs}$ relation.
- 2) g_{ds} nonlinearity from channel length modulation effect. Note that g_{ds} contributes less nonlinearity when device operates deeper into saturation region.
- 3) the dependence of g_m on V_{ds} , (partially due to the drain induced barrier lowering (DIBL) effect [29]).
- 4) the dependence of g_{ds} on V_{gs} , especially in saturation region [44].

The cross modulation effect remains fairly constant for a broad range of V_{gs} , while g_m is more linear and g_{ds} becomes more nonlinear as V_{gs} increases, V_{ds} decreases, and transistors operate close to the linear region. In [29] the V_{gs} , V_{ds} cross term $c_{1,2}$ cancels the intrinsic second-order distortion (g_2) to obtain an amplifying stage with high IIP₂. Note that when g_{ds} nonlinearity dominates (i.e., output limited), the tradeoff between gain and linearity becomes more severe.

TABLE IV
DOMINANT CONTRIBUTOR TO DISTORTION UNDER VARIOUS CONDITIONS

	g_m	g_{ds}
Small load resistance	✓	
Large load resistance		✓
Small voltage gain A_v	✓	
Large voltage gain A_v		✓
High frequency	✓	
Low frequency		✓
Saturation region	✓	

B. MOSFET Capacitance

For the most part, the capacitances of a saturation-region transistor are linear at an operating frequency less than $f_T/10$ [44]. Therefore they do not directly contribute to distortion [43]. However, if a strong blocker is present (e.g., in the order of 0dBm), C_{gs} varies significantly around the threshold voltage, and its nonlinearity becomes significant. Also, as previously mentioned, C_{gd} provides a feedback path for the “second-order interaction,” and this C_{gd} effect becomes more visible as the load impedance increases. At high frequency, C_{gd} and C_{db} reduce total output impedance and hence the output voltage swing, helping to mitigate the nonlinear g_{ds} effect. Therefore, g_m nonlinearity dominates at high frequency, while g_{ds} nonlinearity dominates at low frequency [43]. However, in those circuits where capacitive components are tuned out for a matched load, g_{ds} nonlinearity is still prominent at high frequencies. The substrate affects linearity through C_{db} with higher operating frequency, and this effect varies with different substrate-doping profiles [43]. Generally, IIP3 improves as substrate doping increases [47]. The effect from substrate leakage current can typically be neglected [45].

Table IV provides a summary: the mark “✓“ denotes whether g_m or g_{ds} dominates the distortion under the given conditions.

C. Impact of Technology Scaling On Linearity

As channel length decreases, the velocity saturation effect becomes prominent, i.e., the drain current saturates at smaller V_{ds} . Thus, the long-channel equation for drain current in saturation region needs to be modified as [5]

$$\begin{aligned} I_{ds} &= \frac{\mu C_{ox}}{2} \frac{W}{L} (V_{gs} - V_t) [(V_{gs} - V_t) \parallel (LE_{sat})] \\ &\approx \frac{\mu C_{ox}}{2} W (V_{gs} - V_t) E_{sat} \text{(for small } L\text{)} \end{aligned} \quad (32)$$

where E_{sat} is the field strength at which the carrier velocity drops to half the value extrapolated from low-field mobility. g_m becomes more linear

$$g_m = \frac{\partial I_{ds}}{\partial V_{gs}} = \frac{\mu C_{ox}}{2} W E_{sat}. \quad (33)$$

The vertical-field mobility degradation effect also helps to linearize g_m in DSM process. The long-channel equation for drain current can be modified as

$$I_{ds} = \frac{\mu C_{ox}}{2} \frac{W}{L} \frac{(V_{gs} - V_t)^2}{1 + \theta(V_{gs} - V_t)} \quad (34)$$

where $\theta \propto 1/t_{ox}$ models vertical-field mobility degradation. Equation (34) reduces to $I_{ds} \propto (V_{gs} - V_t)$ as $V_{gs} - V_t$ increases, resulting in a linear I-V curve, and g_m becomes constant with respect to bias voltage:

$$g_m \approx \frac{\mu C_{ox}}{2} \frac{W}{L} \frac{1}{\theta}. \quad (35)$$

On the other hand, g_{ds} is more nonlinear for shorter channel length, as proven by the experimental data in [43]. Furthermore,

the reduced supply/ V_{dsat} values result in the device being biased closer to the triode-saturation boundary, which worsens the g_{ds} nonlinearity. Consequently, the maximum OIP3 occurs with smaller load impedance (which mitigates the distortion contribution from nonlinear g_{ds}) and the peak IIP3 shifts to lower V_{gs} [47], since a smaller overdrive voltage allows the device to stay far away from the triode-saturation boundary while still keeping g_{ds} nonlinearity small.

The “sweet spot” in the optimal biasing technique(discussed in Section II-C) will systematically shift to higher bias-current density I_{ds}/W (i.e., larger overdrive voltage) as technology scales down [44], which means larger power is required to preserve linearity.

As oxide thickness decreases, poly-gate depletion increases, and the nonlinear gate capacitance develops strong second-order derivatives (C_{g2}) with respect to V_{gs} , which contribute to significant third-order distortion (g_3) in drain current, as shown in (36) and (37) [46]

$$I_{ds} = Qv \approx Qv_{sat} \quad (36)$$

$$\begin{aligned} g_3 &= \frac{\partial^3 I_{ds}}{\partial V_{gs}^3} \approx \frac{\partial^3 Q}{\partial V_{gs}^3} v_{sat} \approx \frac{\partial^2 C(V_{gs})}{\partial V_{gs}^2} v_{sat} \\ &= C_{g2} v_{sat} \end{aligned} \quad (37)$$

where Q is the channel charge density along the current direction, v is the carriers’ velocity, and v_{sat} is the saturated velocity for sufficiently high field. Thus, distortion increases with thinner oxides.

DIBL becomes more severe in DSM process, besides a V_{ds} -dependent g_m , DIBL also affects the linearity by changing the effective V_{th} [48]. Measured results in [48] shows that the distortion is more sensitive to DIBL effect when the drain voltage increases and the MOSFET operates in moderate region (i.e., V_{gs} is slightly higher than V_{th}).

Finally, each process has a “low frequency limit” (LFL), below which the MOSFET exhibits fairly frequency-independent linearity. LFL is closely related to the device speed and can be approximated as $f_T/5$ [48]. Therefore, it is easier to achieve IIP2/IIP3 flatness over the signal band in smaller-size technology.

In summary, as technology scales down, lower supply voltages reduce the headroom and can lead to greater nonlinearity from g_{ds} , necessitating multidimensional Taylor analysis to model the nonlinear I_{ds} . Higher-order effects such as DIBL, velocity saturation, and poly-gate depletion all affect linearity. A key challenge resides in delivering high linearity with core transistors and with a low supply voltage in the DSM processes.

V. REMARKS FOR HIGH LINEARITY LNA DESIGN

Besides applying explicit linearization techniques to the circuit, some general guidelines are helpful for designing a high-linearity LNA.

A. To Reduce g_m -Induced Distortion

From (34), the low-frequency expressions for second- and third-order intercept points A_{IIP2} and A_{IIP3} are [3], [5]:

$$A_{IIP2}^2 = \left| \frac{g_1}{g_2} \right| = |V_{dsat} (2 + \theta V_{dsat}) (1 + \theta V_{dsat})| \quad (38)$$

$$\begin{aligned} A_{IIP3}^2 &= \frac{4}{3} \left| \frac{g_1}{g_3} \right| \\ &= \left| \frac{4}{3} \frac{V_{dsat}}{\theta} (2 + \theta V_{dsat}) (1 + \theta V_{dsat})^2 \right| \end{aligned} \quad (39)$$

where $V_{dsat} = V_{gs} - V_{th}$. Equations (38) and (39) indicate that increasing V_{dsat} improves both IIP2 and IIP3. Therefore, given sufficient voltage headroom, maximizing V_{dsat} and minimizing transistor sizes helps to minimize parasitics and to linearize the circuit at the cost of increased dc current.

B. To Reduce g_{ds} -Induced Distortion

Increasing supply voltage mitigates the g_{ds} effect and also improves $P_{1\text{ dB}}$, but the voltage drop across core transistors must be ensured not to exceed the safe operation value.

Provided sufficient voltage headroom, adding cascode device allows the output impedance from transistors to be much larger than load resistor, yielding a more linear output load.

With cascode transistor, most of the output swing will show as V_{ds} variation at the cascode transistor, while the V_{ds} of input transistor remains relatively constant. Therefore, the nonlinear output conductance of the cascode transistor has more contribution to the overall distortion. It is helpful to bias the cascode transistor at smaller V_{gs} (i.e., lower overdrive voltage) to tolerate a larger swing at the drain.

If supply voltage cannot be increased, we can:

- 1) use longer channel length to reduce the channel length modulation effect (assuming speed is not an issue);
- 2) reduce the load resistance of the LNA, which may affect the design of other building blocks in the receiver.

C. To Reduce Second-Order Distortion

- 1) Biasing a CS-stage at the maximum gain yields a high IIP2 in DSM process [29].
- 2) Biasing the device for maximum f_T yields minima in the second-harmonic [48]; this intrinsic distortion cancellation results from opposite contributions of gate capacitance and g_m , as the device enters the linear region from saturation.

D. Other Tips

For inductively degenerated CS-LNAs, we can reduce Q to mitigate the “Q boosting” effect [5], provided that there is enough margin in NF and gain. Since C_{gs} has negligible effect on linearity [45], an external capacitor can be added in parallel with C_{gs} to allow more freedom for input transistor sizing. On the other hand, CG-LNAs generally provide better linearity than CS-LNAs [32] because it does not have this “Q boosting.”

Use cascode transistors whenever possible because they:

- 1) reduce second-order interaction through C_{gd} ;
- 2) reduce the voltage swing across each active device, improving reliability for DSM devices.

VI. CONCLUSIONS

We have reviewed eight categories of CMOS LNA-linearization techniques and discussed the tradeoffs among linearity, power, and PVT variations. We subsequently discussed wide-band-LNA-linearization issues for the emerging broadband transceivers, noting that IIP2 is becoming just as important as IIP3, and that improving $P_{1\text{ dB}}$ is also necessary for wide-band applications to improve high-signal-handling capability. Issues in DSM processes, such as nonlinear output conductance were examined. A key challenge resides in delivering high linearity with core transistors and low supply voltage in the DSM processes. Linearization techniques for canceling higher-order distortion terms (beyond third order), linearizing output conductance, and improving LNA $P_{1\text{ dB}}$ still remain open problems. Finally, we presented general design guidelines for high-linearity LNAs.

ACKNOWLEDGMENT

The authors wish to thank W. He, E. Pankratz, X. Chen, M. Onabajo, M. Mobarak, Dr. C. Xin, Dr. S. Krishnamurthy, and A. Amer for proofreading the manuscript. The authors also wish to thank Dr. S. Lou, J. Wardlaw, and J. Kim for helpful discussions. The authors also want to acknowledge support from the TI Wireless Group.

REFERENCES

- [1] E. Sánchez-Sinencio and J. Silva-Martinez, “CMOS transconductance amplifiers, architectures and active filters: A tutorial,” *IEE Proc.–Circuits Devices Syst.*, vol. 147, no. 1, pp. 3–12, Feb. 2000.
- [2] V. Aparin, “State-of-the-art techniques for high linearity CMOS low noise amplifiers,” presented at the IEEE RFIC Symp. Workshop WSC, Honolulu, HI, 2007.
- [3] B. H. Leung, *VLSI for Wireless Communication*. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [4] W. Sansen, “Distortion in elementary transistor circuits,” *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 46, no. 3, pp. 315–325, Mar. 1999.
- [5] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [6] S. Narayanan, “Application of Volterra series to intermodulation distortion analysis of transistor feedback amplifiers,” *IEEE Trans. Circuit Theory*, vol. CT-17, no. 4, pp. 518–527, Nov. 1970.
- Harmonic termination**
- [7] V. Aparin and C. Persico, “Effect of out-of-band terminations on intermodulation distortion in common-emitter circuits,” in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1999, vol. 3, pp. 977–980.
- [8] V. Aparin and L. E. Larson, “Linearization of monolithic LNAs using low-frequency low-impedance input termination,” in *Proc. Eur. Solid-State Circuits Conf.*, Sep. 2003, pp. 137–140.
- [9] K. L. Fong, “High-frequency analysis of linearity improvement technique of common-emitter trans-conductance stage using a low-frequency trap network,” *IEEE J. Solid-State Circuits*, vol. 35, no. 8, pp. 1249–1252, Aug. 2000.
- [10] T. W. Kim, “A common-gate amplifier with transconductance nonlinearity cancellation and its high-frequency analysis using the Volterra series,” *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 6, pp. 1461–1469, Jun. 2009.
- [11] B. Kim, J.-S. Ko, and K. Lee, “Highly linear CMOS RF MMIC amplifier using multiple gated transistors and its Volterra series analysis,” in *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, vol. 1, pp. 515–518.
- [12] J. S. Fairbanks and L. E. Larson, “Analysis of optimized input and output harmonic termination on the linearity of 5 GHz CMOS radio frequency amplifiers,” in *Radio Wireless Conf.*, Aug. 2003, pp. 293–296.
- [13] T. W. Kim, B. Kim, and K. Lee, “Highly linear receiver front-end adopting MOSFET transconductance linearization by multiple gated transistors,” *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 223–229, Jan. 2004.
- [14] X. Fan, H. Zhang, and E. Sánchez-Sinencio, “A noise reduction and linearity improvement technique for a differential cascode LNA,” *IEEE J. Solid-State Circuits*, vol. 43, no. 3, pp. 588–599, Mar. 2008.
- Optimum biasing**
- [15] V. Aparin, G. Brown, and L. E. Larson, “Linearization of CMOS LNAs via optimum gate biasing,” in *Proc. IEEE Int. Circuits Syst. Symp.*, Vancouver, BC, Canada, May 2004, vol. 4, pp. 748–751.
- Feedforward**
- [16] Y. Ding and R. Harjani, “A + 18 dBm IIP3 LNA in 0.35 μm CMOS,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2001, pp. 162–163.
- [17] E. Keehi and A. Hajimiri, “Equalization of IM3 products in wideband direct-conversion receivers,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2008, pp. 204–205.
- Derivative superposition (DS)**
- [18] Y. S. Youn, J. H. Chang, K. J. Koh, Y. J. Lee, and H. K. Yu, “A 2 GHz 16 dBm IIP3 low noise amplifier in 0.25 μm CMOS technology,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2003, pp. 452–453.
- [19] C. Xin and E. Sánchez-Sinencio, “A linearization technique for RF low noise amplifier,” in *Proc. IEEE Int. Circuits Syst. Symp.*, Vancouver, BC, Canada, May 2004, vol. IV, pp. 313–316.
- [20] H. M. Gedda, J. W. Park, and J. Silva-Martinez, “Robust derivative superposition method for linearizing broadband LNAs,” *IEE Electron. Lett.*, vol. 45, no. 9, pp. 435–436, Apr. 2009.

Complementary DS

- [21] D. Im, I. Nam, H. Kim, and K. Lee, "A wideband CMOS low noise amplifier employing noise and IM2 distortion cancellation for a digital TV tuner," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 686–698, Mar. 2009.

Differential DS

- [22] T. W. Kim and B. Kim, "A 13-dB IIP3 improved low-power CMOS RF programmable gain amplifier using differential circuit transconductance linearization for various terrestrial mobile D-TV applications," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 945–953, Apr. 2006.

Modified DS

- [23] V. Aparin and L. E. Larson, "Modified derivative superposition method for linearizing FET low-noise amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 2, pp. 571–581, Feb. 2005.
- [24] S. Ganesan, E. Sánchez-Sinencio, and J. Silva-Martinez, "A highly linear low noise amplifier," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4079–4085, Dec. 2006.

IM2 injection

- [25] S. Lou and H. C. Luong, "A linearization technique for RF receiver front-end using second-order-intermodulation injection," *IEEE J. Solid-State Circuits*, vol. 43, no. 11, pp. 2404–2412, Nov. 2008.

Noise/distortion cancellation

- [26] F. Brucolieri, E. A. M. Klumperink, and B. Nauta, "Wide-band CMOS low-noise amplifier exploiting thermal noise canceling," *IEEE J. Solid-State Circuits*, vol. 39, no. 2, pp. 275–282, Feb. 2004.
- [27] J. Jussila and P. Sivonen, "A 1.2-V highly linear balanced noise-cancelling LNA in 0.13- μ m CMOS," *IEEE J. Solid-State Circuits*, vol. 43, no. 3, pp. 579–587, Mar. 2008.
- [28] W. Chen, G. Liu, B. Zdravko, and A. M. Niknejad, "A highly linear broadband CMOS LNA employing noise and distortion cancellation," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1164–1176, May 2008.
- [29] S. C. Blaakmeer, E. A. M. Klumperink, D. M. W. Leenaerts, and B. Nauta, "Wideband balun-LNA with simultaneous output balancing, noise-canceling and distortion-canceling," *IEEE J. Solid-State Circuits*, vol. 43, no. 6, pp. 1341–1350, Jun. 2008.

Post-distortion

- [30] N. Kim, V. Aparin, K. Barnett, and C. Persico, "A cellular-band CDMA 0.25 μ m CMOS LNA linearized using active post-distortion," *IEEE J. Solid-State Circuits*, vol. 41, no. 7, pp. 1530–1534, Jul. 2006.
- [31] T.-S. Kim and B.-S. Kim, "Post-linearization of cascode CMOS LNA using folded PMOS IMD sinker," *IEEE Microw. Wireless Comp. Lett.*, vol. 16, no. 4, pp. 182–184, Apr. 2006.
- [32] H. Zhang, X. Fan, and E. Sánchez-Sinencio, "A low-power, linearized, ultra-wideband LNA design technique," *IEEE J. Solid-State Circuits*, vol. 44, no. 2, pp. 320–330, Feb. 2009.

IIP2 calibration

- [33] D. Kaczman, M. Shah, M. Alam, M. Rachedine, D. Cashen, L. Han, and A. Raghavan, "A single-chip 10-Band WCDMA/HSDPA 4-Band GSM/EDGE SAW-less CMOS receiver with DigRF 3 G interface and +90 dBm IIP2," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 718–739, Mar. 2009.
- [34] M. Hotti, J. Ryyynanen, K. Kivekas, and K. Halonen, "An IIP2 calibration technique for direct conversion receivers," in *Proc. IEEE Int. Circuits Syst. Symp.*, Vancouver, BC, Canada, May 2004, vol. 4, pp. 257–260.
- [35] W. Kim, S. Yang, Y. Moon, J. Yu, H. Shin, W. Choo, and B. Park, "IIP2 calibrator using common mode feedback circuitry," in *Proc. Eur. Solid-State Circuits Conf.*, 2005, pp. 231–234.
- [36] H. Darabi, H. Kim, J. Chiu, B. Ibrahim, and L. Serrano, "An IP2 improvement technique for zero-if down-converters," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2006, pp. 464–465.

IIP2/IIP3 dependence on two-tone spacing

- [37] W. Bosch and G. Gatti, "Measurement and simulation of memory effects in predistortion linearizers," *IEEE Trans. Microw. Theory Tech.*, vol. 37, no. 12, pp. 1885–1890, Dec. 1989.
- [38] J. F. Sevic, K. L. Burguer, and M. B. Steer, "A novel envelope-termination load-pull methods for ACPR optimization of RF/microwave power amplifiers," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Baltimore, MD, 1998, pp. 601–605.
- [39] N. B. de Carvalho and J. C. Pedro, "Two-tone IMD asymmetry in microwave power amplifiers," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Boston, MA, 2000, pp. 445–448.

- [40] N. Carvalho and J. Pedro, "A comprehensive explanation of distortion sideband asymmetries," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 9, pp. 2090–2101, Sep. 2002.

Others

- [41] N. Cowley and R. Hanrahan, "ATSC compliance and tuner design implications, video/imaging design line," *EE Times* Nov. 10, 2005 [Online]. Available: <http://www.videesignline.com/showArticle.jhtml?articleID=173601582>
- [42] S. S. Taylor and J. S. Duster, "High-linearity low noise amplifier and method," U.S. Patent 0 278 220, Nov. 13, 2008.

LNA linearization in deep-submicrometer technology

- [43] S. Kang, B. Choi, and B. Kim, "Linearity analysis of CMOS for RF application," *IEEE Trans. Microw. Theory Tech.*, vol. 51, no. 3, pp. 972–977, Mar. 2003.
- [44] B. Toole, C. Plett, and M. Cloutier, "RF circuit implications of moderate inversion enhanced linear region in MOSFETs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 2, pp. 319–328, Feb. 2004.
- [45] R. A. Baki, T. K. K. Tsang, and M. N. El-Gamal, "Distortion in RF CMOS short-channel low-noise amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 1, pp. 46–56, Jan. 2006.
- [46] C. H. Choi, Z. Yu, and R. W. Dutton, "Impact of poly-gate depletion on MOS RF linearity," *IEEE Electron Device Lett.*, vol. 24, no. 5, pp. 330–332, May 2003.
- [47] R. van Langevelde, L. F. Tiemeijer, R. J. Havens, M. J. Knitel, R. F. M. Ores, P. H. Woerlee, and D. B. M. Klaassen, "RF-distortion in deep-submicron CMOS technologies," in *IEEE IEDM Tech. Dig.*, 2000, pp. 807–810.
- [48] T. Lee and Y. Cheng, "High-frequency characterization and modeling of distortion behavior of MOSFETs for RF IC design," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1407–1414, Sep. 2004



Heng Zhang (S'07) was born in Guangzhou, China. She received the B.S. degree in electrical engineering from Peking University, Beijing, China, in 2004, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, in 2010.

In the summer and fall of 2006, she was an Analog IC Design Engineer (Co-op) with Texas Instrument, Dallas, TX, where she designed a low power ADC for hard disk applications. In summer 2007, she was with the RF and Analog Technologies Department, UMC, Sunnyvale, CA, where she researched digital calibration techniques for ADCs. Since August 2010, she has been with the Analog and Mixed Signal Group, Broadcom Corporation, Irvine, CA, working on high-speed transceivers for optical and backplane/cable applications. Her research interests include data converters and high-speed/RF circuits. Her Web site can be found at <http://www.ece.tamu.edu/~hzhang>.



Edgar Sánchez-Sinencio (S'72–M'74–SM'83–F'92–LF'10) was born in Mexico City, Mexico. He received the degree in communications and electronic engineering (Professional degree) from the National Polytechnic Institute of Mexico, Mexico City, the M.S.E.E. degree from Stanford University, Stanford, CA, and the Ph.D. degree from the University of Illinois at Champaign-Urbana, in 1966, 1970, and 1973, respectively.

He is currently the T. J. Kilby Chair Professor and Director of the Analog and Mixed-Signal Center at Texas A&M University. His research work has more than 2636 citations according to the Thomson Reuters Scientific Citation Index. He has graduated 42 M.Sc. and 34 Ph.D. students. He is a coauthor of six books on different topics, such as RF circuits, low-voltage low-power analog circuits, and neural networks. His present interests are in the area of power management, RF-communication circuits, analog and medical electronics circuit design.

Dr. Sánchez-Sinencio is a former Editor-in-Chief of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. In November 1995 he was awarded a Honoris Causa Doctorate by the National Institute for Astrophysics, Optics and Electronics, Mexico. This degree was the first honorary degree awarded for microelectronic circuit design contributions. He is a corecipient of the 1995 Guillemin-Cauer Award for his work on cellular networks. He was also the corecipient of the 1997 Darlington Award for his work on high-frequency filters. He received the IEEE Circuits and Systems Society Golden Jubilee Medal in 1999. He also received the prestigious IEEE Circuits and Systems Society 2008 Technical Achievement Award. His Web site can be found at <http://amesp02.tamu.edu/~sanchez/>.

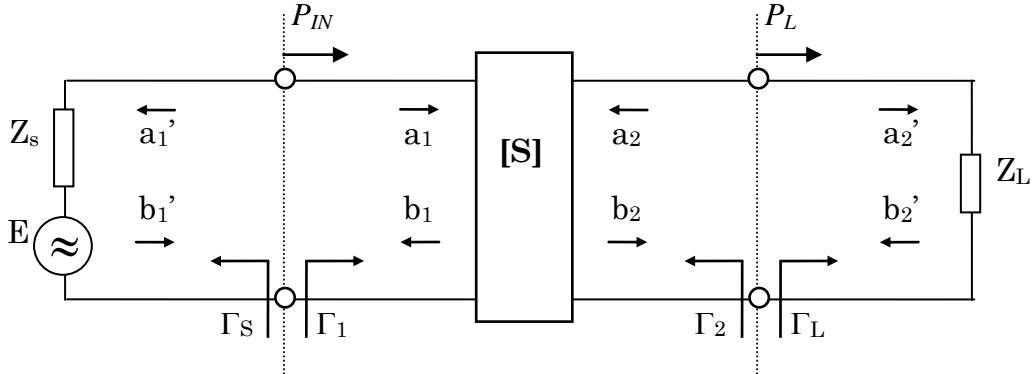


Fig. IV-16. Two-port transistor representation in terms of scattering parameters.

$$\Gamma_S = \frac{b_1}{a_1} = \frac{a_1}{b_1} \quad \Gamma_L = \frac{b_2}{a_2} = \frac{a_2}{b_2}$$

$$b_1 = S_{11}a_1 + S_{12}a_2 \quad b_2 = S_{21}a_1 + S_{22}a_2$$

STABILITY

$$\Gamma_1 = \frac{b_1}{a_1} = \frac{S_{11}a_1 + S_{12}a_2}{a_1} = S_{11} + S_{12} \frac{a_2}{a_1}$$

$$a_2 = \Gamma_L b_2 = \Gamma_L (S_{21}a_1 + S_{22}a_2) \quad \Rightarrow \quad \frac{a_2}{a_1} = \frac{\Gamma_L S_{21}}{1 - \Gamma_L S_{22}}$$

$$\text{Therefore,} \quad \Gamma_1 = S_{11} + \frac{\Gamma_L S_{12} S_{21}}{1 - \Gamma_L S_{22}}$$

$$\text{Similarly,} \quad \Gamma_2 = S_{22} + \frac{\Gamma_S S_{12} S_{21}}{1 - \Gamma_S S_{11}}$$

$$\Rightarrow \left| S_{11} + \frac{\Gamma_L S_{12} S_{21}}{1 - \Gamma_L S_{22}} \right| < 1 \quad \left| S_{22} + \frac{\Gamma_S S_{12} S_{21}}{1 - \Gamma_S S_{11}} \right| < 1$$

MATCHING

$$\Gamma_S = \Gamma_1^* = \left(S_{11} + \frac{\Gamma_L S_{12} S_{21}}{1 - \Gamma_L S_{22}} \right)^* = S_{11}^* + \frac{\Gamma_L^* S_{12}^* S_{21}^*}{1 - \Gamma_L^* S_{22}^*}$$

with $\Gamma_2 = \Gamma_L^* = S_{22} + \frac{\Gamma_S S_{12} S_{21}}{1 - \Gamma_S S_{11}}$

By substitution,

$$\Gamma_S = S_{11}^* + \frac{\left(S_{22} + \frac{\Gamma_S S_{12} S_{21}}{1 - \Gamma_S S_{11}} \right) S_{12}^* S_{21}^*}{1 - \left(S_{22} + \frac{\Gamma_S S_{12} S_{21}}{1 - \Gamma_S S_{11}} \right) S_{22}^*} = S_{11}^* + \frac{(S_{22} - S_{22} \Gamma_S S_{11} + \Gamma_S S_{12} S_{21}) S_{12}^* S_{21}^*}{1 - \Gamma_S S_{11} - (S_{22} - S_{22} \Gamma_S S_{11} + \Gamma_S S_{12} S_{21}) S_{22}^*}$$

$$\Gamma_S = S_{11}^* + \frac{(S_{22} - \Delta \Gamma_S) S_{12}^* S_{21}^*}{1 - \Gamma_S S_{11} - |S_{22}|^2 + S_{22}^* \Delta \Gamma_S}$$

with $\Delta = S_{11} S_{22} - S_{12} S_{21}$

$$\Gamma_S (1 - \Gamma_S S_{11} - |S_{22}|^2 + S_{22}^* \Delta \Gamma_S) = S_{11}^* (1 - \Gamma_S S_{11} - |S_{22}|^2 + S_{22}^* \Delta \Gamma_S) + (S_{22} - \Delta \Gamma_S) S_{12}^* S_{21}^*$$

$$\Gamma_S (1 - |S_{22}|^2) + \Gamma_S^2 (S_{22}^* \Delta - S_{11}) = \Gamma_S (S_{11}^* S_{22}^* \Delta - |S_{11}|^2 - \Delta S_{12}^* S_{21}^*) + S_{11}^* - S_{11}^* |S_{22}|^2 + S_{22} S_{12}^* S_{21}^*$$

$$\Gamma_S^2 (S_{22}^* \Delta - S_{11}) + \Gamma_S (1 - |S_{22}|^2 - S_{11}^* S_{22}^* \Delta + |S_{11}|^2 + \Delta S_{12}^* S_{21}^*) = S_{11}^* - S_{11}^* |S_{22}|^2 + S_{22} S_{12}^* S_{21}^*$$

with

$$S_{11}^* S_{22}^* \Delta - \Delta S_{12}^* S_{21}^* = |\Delta|^2$$

we obtain

$$-\Gamma_S^2 (S_{22}^* \Delta - S_{11}) - \Gamma_S (1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2) + (S_{11}^* - S_{11}^* S_{22} S_{22}^* + S_{22} S_{12}^* S_{21}^*) = 0$$

$$\Rightarrow \Gamma_S^2 (S_{11} - S_{22}^* \Delta) - \Gamma_S (1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2) + (S_{11}^* - S_{22} \Delta^*) = 0$$

$$\text{Let } B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 \quad C_1 = S_{11} - \Delta S_{22}^*$$

$$\Rightarrow \Gamma_S^2 C_1 - \Gamma_S B_1 + C_1^* = 0$$

$$\text{Two solutions : } \Gamma_S = \frac{B_1 \pm \sqrt{B_1^2 - 4|C_1|^2}}{2C_1}$$

$$\text{Similarly, let } B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2 \quad C_2 = S_{22} - \Delta S_{11}^*$$

$$\Rightarrow \Gamma_L^2 C_2 - \Gamma_L B_2 + C_2^* = 0$$

$$\text{Two solutions : } \Gamma_L = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2C_2}$$

STABILITY FACTOR

$$\text{Let } B_1^2 - 4|C_1|^2 = B_2^2 - 4|C_2|^2 = 4(K^2 - 1)|S_{12} S_{21}|^2$$

$$\Rightarrow K = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{12}| |S_{21}|}$$

a) $K \equiv 1$ it is not a practical case since

$$|\Gamma_S| = |\Gamma_L| = 1$$

b) $K \geq 1$ expanding the two solutions of Γ_S and Γ_L gives

$$\Gamma_S' = \frac{B_1 - \sqrt{B_1^2 - 4|C_1|^2}}{2C_1}; \quad \Gamma_S'' = \frac{B_1 + \sqrt{B_1^2 - 4|C_1|^2}}{2C_1}$$

$$\Gamma_L' = \frac{B_2 - \sqrt{B_2^2 - 4|C_2|^2}}{2C_2}; \quad \Gamma_L'' = \frac{B_2 + \sqrt{B_2^2 - 4|C_2|^2}}{2C_2}$$

Note that

$$|\Gamma_S'| |\Gamma_S''| = 1 \quad |\Gamma_L'| |\Gamma_L''| = 1$$

To explicit which solution has to be retained, we have to consider Δ :

- $|\Delta| < 1$ The pair of solutions is $\{\Gamma_S', \Gamma_L'\}$.

Unconditional stability and maximum gain $G_{T \max} = \left| \frac{S_{21}}{S_{12}} \right| \left(K - \sqrt{K^2 - 1} \right)$

- $|\Delta| > 1$ The pair of solutions is $\{\Gamma_S'', \Gamma_L''\}$.

Conditional stability and minimum gain $G_{T \min} = \left| \frac{S_{21}}{S_{12}} \right| \left(K + \sqrt{K^2 + 1} \right)$

c) $K \leq -1$ The two possible solutions are $\{\Gamma_S', \Gamma_L''\}$ or $\{\Gamma_S'', \Gamma_L'\}$ but they are not useful for amplifiers (*natural instability*).

d) $-1 \leq K \leq 1$ Simultaneous matching is impossible. The amplifier is conditionally stable (or potentially unstable)

\Rightarrow Stability regions. Reported into the reflection coefficient space (Smith chart), these regions are equivalent to circles, and thus are called "stability circles".

The locus of Γ_S given the critical value $\{ \Gamma_L = 1 \}$ is a circle in the complex plane (center Ω_1 and radius R_1).

The locus of Γ_L given the critical value $\{ \Gamma_S = 1 \}$ is a circle in the complex plane (center Ω_2 and radius R_2).

$$\text{In fact, let } |\Gamma_S| = 1 = \left| S_{11} + \frac{\Gamma_L S_{12} S_{21}}{1 - \Gamma_L S_{22}} \right|$$

$$\Rightarrow |S_{11} - \Gamma_L S_{11} S_{22} + \Gamma_L S_{12} S_{21}| = |1 - \Gamma_L S_{22}| \quad \text{or} \quad |S_{11} - \Gamma_L \Delta| = |1 - \Gamma_L S_{22}|$$

$$\text{Given the squares, } |S_{11} - \Gamma_L \Delta|^2 = (S_{11} - \Gamma_L \Delta)(S_{11} - \Gamma_L \Delta)^* = (1 - \Gamma_L S_{22})(1 - \Gamma_L S_{22})^*$$

$$\text{We obtain : } |\Gamma_L|^2 - \frac{2}{|S_{22}|^2 - |\Delta|^2} \cdot \text{Re} [(S_{22} - \Delta S_{11}^*) \Gamma_L] = \frac{|S_{11}|^2 - 1}{|S_{22}|^2 - |\Delta|^2}$$

$$\text{with } R_2^2 = \frac{|S_{11}|^2 - 1}{|S_{22}|^2 - |\Delta|^2} + \frac{(S_{22} - \Delta S_{11}^*)(S_{22} - \Delta S_{11}^*)^*}{(|S_{22}|^2 - |\Delta|^2)^2}$$

$$\Rightarrow \Omega_2 = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \quad R_2 = \frac{|S_{12} S_{21}|}{|S_{22}|^2 - |\Delta|^2}$$

$$\text{Similarly, } |\Gamma_S|^2 - \frac{2}{|S_{11}|^2 - |\Delta|^2} \cdot \text{Re} [(S_{11} - \Delta S_{22}^*) \Gamma_S] = \frac{|S_{22}|^2 - 1}{|S_{11}|^2 - |\Delta|^2}$$

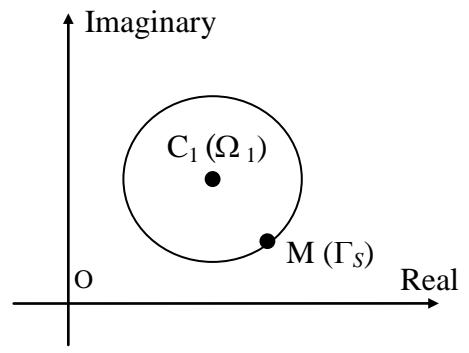
$$\Rightarrow \Omega_1 = \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \quad R_1 = \frac{|S_{21} S_{12}|}{|S_{11}|^2 - |\Delta|^2}$$

→ **Equation of a circle in the complex plane:**

$$\overline{OM} = \overline{OC_1} + \overline{C_1M} \text{ or in magnitude :}$$

$$|\overline{C_1M}|^2 = |\overline{OM} - \overline{OC_1}|^2 \rightarrow R_1^2 = (\Gamma_S - \Omega_1)(\Gamma_S - \Omega_1)^*$$

$$\Rightarrow |\Gamma_S|^2 - 2 \operatorname{Re}(\Gamma_S \Omega_1^*) = R_1^2 - |\Omega_1|^2$$



$$\Rightarrow |\Gamma_S|^2 - \frac{2}{|S_{11}|^2 - |\Delta|^2} \cdot \operatorname{Re} [(S_{11} - \Delta S_{22}^*) \Gamma_S] = \frac{|S_{22}|^2 - 1}{|S_{11}|^2 - |\Delta|^2}$$

$$\Omega_1 = \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \quad R_1 = \frac{|S_{21} S_{12}|}{|S_{11}|^2 - |\Delta|^2}$$

$$\Rightarrow |\Gamma_L|^2 - \frac{2}{|S_{22}|^2 - |\Delta|^2} \cdot \operatorname{Re} [(S_{22} - \Delta S_{11}^*) \Gamma_L] = \frac{|S_{11}|^2 - 1}{|S_{22}|^2 - |\Delta|^2}$$

$$\Omega_2 = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \quad R_2 = \frac{|S_{12} S_{21}|}{|S_{22}|^2 - |\Delta|^2}$$

3

MICROWAVE TRANSISTOR AMPLIFIER DESIGN

3.1 INTRODUCTION

This chapter develops some basic principles used in the analysis and design of microwave transistor amplifiers. Based on the S parameters of the transistor and certain performance requirements, a systematic procedure is developed for the design of a microwave transistor amplifier.

The most important design considerations in a microwave transistor amplifier are stability, power gain, bandwidth, noise, and dc requirements. This chapter deals mainly with the problems of stability and power gain in narrow-band amplifiers. VSWR considerations are also discussed. Low-noise amplifiers, broadband amplifiers, and power amplifiers are discussed in Chapter 4.

A design usually starts with a set of specifications and the selection of the proper transistor. Then a systematic mathematical solution, aided by graphical methods, is developed to determine the transistor loading (i.e., the source and load reflection coefficients) for a particular stability and gain criteria. An unconditionally stable transistor will not oscillate with any passive termination. On the other hand, a design using a potentially unstable transistor requires some analysis and careful considerations so that the passive terminations produce a stable amplifier.

Design procedures for both unilateral and bilateral transistors, based on stability and gain requirements, are described. Both passive and active dc bias networks for BJTs and GaAs FETs are analyzed. It is important to select the correct dc operating point and the proper dc network topology in order to obtain the desired ac performance.

3.2 POWER GAIN EQUATIONS

Several power gain equations appear in the literature and are used in the design of microwave amplifiers. Figure 3.2.1 illustrates a microwave amplifier signal flow graph and the different powers used in gain equations. The transducer power gain G_T , the power gain G_p (also called the *operating power gain*), and the available power gain G_A are defined as follows:

$$G_T = \frac{P_L}{P_{AVS}} = \frac{\text{power delivered to the load}}{\text{power available from the source}}$$

$$G_p = \frac{P_L}{P_{IN}} = \frac{\text{power delivered to the load}}{\text{power input to the network}}$$

and

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{\text{power available from the network}}{\text{power available from the source}}$$

The expressions for G_T , G_p , and G_A were already derived in (2.6.14), (2.6.15), (2.6.18), and (2.6.22)—namely,

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_{IN}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.2.1)$$

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (3.2.2)$$

$$G_p = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.2.3)$$

$$G_A = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{OUT}|^2} \quad (3.2.4)$$

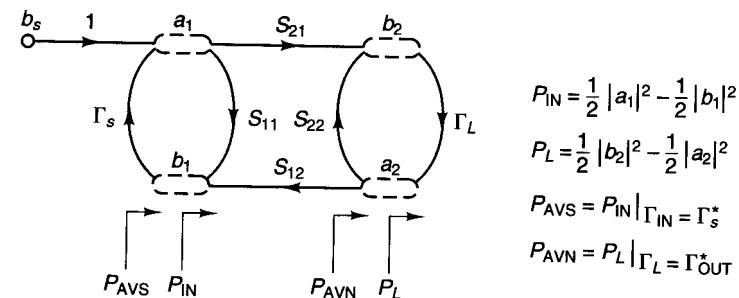


Figure 3.2.1 Different power definitions.

$$\Gamma_{IN} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \quad (3.2.5)$$

$$\Gamma_{OUT} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \quad (3.2.6)$$

G_T is a function of Γ_s , Γ_L , and the S parameters of the transistor (i.e., $G_T = f(\Gamma_s, \Gamma_L, [S])$), $G_p = f(\Gamma_L, [S])$, and $G_A = f(\Gamma_s, [S])$.

In terms of the amplifier shown in Fig. 3.2.2, the input matching network transforms the generator impedance Z_1 (usually 50Ω) to the impedance Z_s , or in other words to the source reflection coefficient Γ_s . The output matching network transforms the impedance Z_2 (usually 50Ω) to the load impedance Z_L or to the load reflection coefficient Γ_L . Observe that we use the nomenclature Z_s and Γ_s for the source impedance and source reflection coefficient at the input port of the transistor, and Z_L and Γ_L for the load impedance and load reflection coefficient at the output port of the transistor.

The values of Γ_s , Γ_L , and the S parameters of the transistor determine the gains of the amplifier according to (3.2.1) through (3.2.4).

The passive matching networks produce values of Γ_s and Γ_L such that $|\Gamma_s| < 1$ and $|\Gamma_L| < 1$. In other words, the resistive part associated with Z_s and Z_L is positive. However, from (3.2.5) and (3.2.6) it is possible that for certain values of the S parameters (where $|\Gamma_s| < 1$ and $|\Gamma_L| < 1$) that $|\Gamma_{IN}| > 1$ and/or $|\Gamma_{OUT}| > 1$. When $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$, the input or output ports of the transistor present a negative resistance and oscillations can occur. Obviously, this is a situation that we must avoid in amplifier design.

Example 3.2.1

(a) The input and output matching networks in Fig. 3.2.2 are designed to produce $\Gamma_s = 0.5|120^\circ$ and $\Gamma_L = 0.4|90^\circ$. Determine G_T , G_A , and G_p if the S parameters of the transistor are

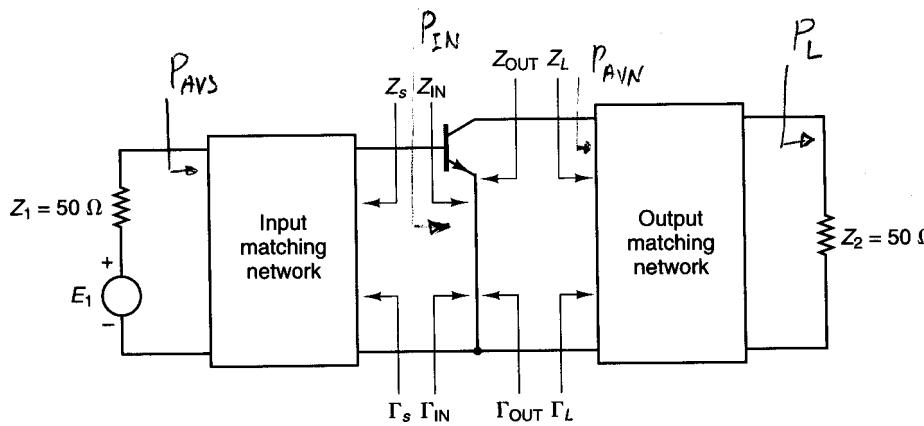


Figure 3.2.2 A microwave amplifier diagram.

$$S_{11} = 0.6| -160^\circ$$

$$S_{12} = 0.045| 16^\circ$$

$$S_{21} = 2.5| 30^\circ$$

$$S_{22} = 0.5| -90^\circ$$

(b) Calculate P_{AVS} , P_{IN} , P_{AVN} , and P_L in Fig. 3.2.2 if $E_1 = 10|0^\circ$, $Z_1 = 50 \Omega$, and $\Gamma_2 = 50 \Omega$.

(c) Calculate $(VSWR)_{in}$ and $(VSWR)_{out}$.

Solution. (a) From (3.2.5) and (3.2.6) we obtain

$$\Gamma_{IN} = 0.6| -160^\circ + \frac{0.045| 16^\circ(2.5| 30^\circ)0.4| 90^\circ}{1 - 0.5| -90^\circ(0.4| 90^\circ)} = 0.627| -164.6^\circ$$

and

$$\Gamma_{OUT} = 0.5| -90^\circ + \frac{0.045| 16^\circ(2.5| 30^\circ)0.5| 120^\circ}{1 - 0.6| -160^\circ(0.5| 120^\circ)} = 0.471| -97.63^\circ$$

Then, from (3.2.1), (3.2.3), and (3.2.4) it readily follows that

$$G_T = \frac{1 - (0.5)^2}{|1 - 0.627| -164.6^\circ(0.5| 120^\circ)|^2} \frac{(2.5)^2}{|1 - 0.5| -90^\circ(0.4| 90^\circ)|^2} = 9.43 \quad (\text{or } 9.75 \text{ dB})$$

$$G_p = \frac{1}{1 - (0.627)^2} \frac{(2.5)^2}{|1 - 0.5| -90^\circ(0.4| 90^\circ)|^2} = 13.51 \quad (\text{or } 11.31 \text{ dB})$$

and

$$G_A = \frac{1 - (0.5)^2}{|1 - 0.6| -160^\circ(0.5| 120^\circ)|^2} \frac{(2.5)^2}{|1 - (0.471)|^2} = 9.55 \quad (\text{or } 9.8 \text{ dB})$$

Since the values of $G_p = P_L/P_{IN}$ and $G_T = P_L/P_{AVS}$ differ by 1.56 dB (i.e., $11.31 \text{ dB} - 9.75 \text{ dB} = 1.56 \text{ dB}$), it follows that the input power is less than the power available from the source. In fact, recall (2.7.16) and (2.7.29)—namely,

$$P_{IN} = P_{AVS} M_s \quad (3.2.7)$$

and

$$G_T = G_p M_s \quad (3.2.8)$$

where from (2.7.17) the source mismatch factor is

$$M_s = \frac{[1 - (0.5)^2][1 - (0.627)^2]}{|1 - 0.5| -160^\circ(0.627| -164.6^\circ)|^2} = 0.6983 \quad (\text{or } -1.56 \text{ dB})$$

Observe how G_T and G_p are related by (3.2.8). That is, $9.43 = 13.51(0.6983)$ or $9.75 \text{ dB} = 11.31 \text{ dB} - 1.56 \text{ dB}$.

Also, since $G_A = P_{AVN}/P_{AVS}$ and $G_T = P_L/P_{AVS}$ are fairly close (i.e., 9.8 dB and 9.75 dB, respectively), the power delivered to the load is close to the power available from the network. In fact,

$$P_L = P_{AVN}M_L \quad (3.2.9)$$

and

$$G_T = G_A M_L \quad (3.2.10)$$

where from (2.7.25) the load mismatch factor is

$$M_L = \frac{[1 - (0.4)^2][1 - (0.471)^2]}{|1 - 0.471| - 97.63^\circ(0.4|90^\circ|)} = 0.9874 \quad (\text{or } -0.055 \text{ dB})$$

Observe how G_T and G_A are related by (3.2.10). That is, $9.43 = 9.55(0.9874)$ or 9.75 dB = 9.8 dB - 0.055 dB.

(b) The power available from the source is

$$P_{AVS} = \frac{E_1^2}{8 \operatorname{Re}[Z_1]} = \frac{10^2}{8(50)} = 0.25 \text{ W}$$

Then, using (3.2.7) the input power is

$$P_{IN} = P_{AVS}M_s = 0.25(0.6983) = 0.1745 \text{ W}$$

The power delivered to the load can be calculated using the definition of G_T . That is,

$$P_L = G_T P_{AVS} = 9.43(0.25) = 2.358 \text{ W}$$

It can also be calculated using the definition of G_P . That is,

$$P_L = G_P P_{IN} = 13.51(0.1745) = 2.358 \text{ W}$$

The power available from the network can be calculated using (3.2.9). That is,

$$P_{AVN} = \frac{P_L}{M_L} = \frac{2.358}{0.9874} = 2.39 \text{ W}$$

It can also be calculated using the definition of G_A . That is,

$$P_{AVN} = G_A P_{AVS} = 9.55(0.25) = 2.39 \text{ W}$$

(c) Since the mismatch factor M_s has been calculated in part (a), it is simple to use (2.8.1) and (2.8.2) to calculate $(VSWR)_{in}$. That is,

$$(VSWR)_{in} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} = \frac{1 + \sqrt{1 - M_s}}{1 - \sqrt{1 - M_s}} = \frac{1 + \sqrt{1 - 0.6983}}{1 - \sqrt{1 - 0.6983}} = 3.44$$

Similarly, using (2.8.4) and (2.8.5) we obtain

$$(VSWR)_{out} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} = \frac{1 + \sqrt{1 - M_L}}{1 - \sqrt{1 - M_L}} = \frac{1 + \sqrt{1 - 0.9874}}{1 - \sqrt{1 - 0.9874}} = 1.25$$

3.3 STABILITY CONSIDERATIONS

The stability of an amplifier, or its resistance to oscillate, is a very important consideration in a design and can be determined from the S parameters, the matching networks, and the terminations. In a two-port network, oscillations are possible when either the input or output port presents a negative resistance. This occurs when $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$, which for a unilateral device occurs when $|S_{11}| > 1$ or $|S_{22}| > 1$. For example, a unilateral transistor is a transistor where $S_{12} = 0$ (or its effect so small that it can be set equal to zero). If $S_{12} = 0$, it follows from (3.2.5) and (3.2.6) that $|\Gamma_{IN}| = |S_{11}|$ and $|\Gamma_{OUT}| = |S_{22}|$. Hence, if $|S_{11}| > 1$ the transistor presents a negative resistance at the input, and if $|S_{22}| > 1$ the transistor presents a negative resistance at the output.

The two-port network shown in Fig. 3.3.1 is said to be unconditionally stable at a given frequency if the real parts of Z_{IN} and Z_{OUT} are greater than zero for all passive load and source impedances. If the two-port is not unconditionally stable, it is potentially unstable. That is, some passive load and source terminations can produce input and output impedances having a negative real part.

In terms of reflection coefficients, the conditions for unconditional stability at a given frequency are

$$|\Gamma_s| < 1 \quad (3.3.1)$$

$$|\Gamma_L| < 1 \quad (3.3.2)$$

$$|\Gamma_{IN}| = \left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| < 1 \quad (3.3.3)$$

and

$$|\Gamma_{OUT}| = \left| S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \right| < 1 \quad (3.3.4)$$

where, of course, all coefficients are normalized to the same characteristic impedance Z_o .

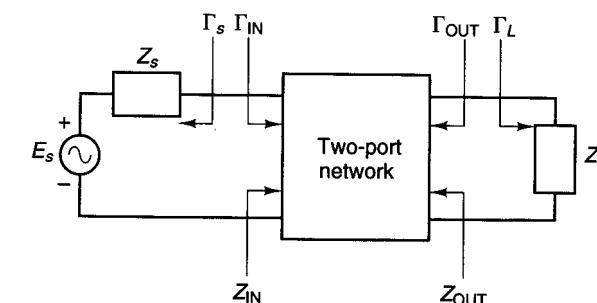


Figure 3.3.1 Stability of two-port networks.

Equations (3.3.1) and (3.3.2) state that the source and load are passive, while (3.3.3) and (3.3.4) state that the input and output impedances must also be passive (i.e., no negative resistance is associated with their real parts).

The solutions of (3.3.1) to (3.3.4) give the required conditions for the two-port network to be unconditionally stable. However, before we discuss the intricacies of the necessary and sufficient conditions for unconditional stability, a graphical analysis of (3.3.1) to (3.3.4) is presented. The graphical analysis is especially useful in the analysis of potentially unstable transistors.

When the two-port in Fig. 3.3.1 is potentially unstable, there may be values of Γ_s and Γ_L (i.e., source and load impedances) for which the real parts of Z_{IN} and Z_{OUT} are positive. These values of Γ_s and Γ_L (i.e., regions in the Smith chart) can be determined using the following graphical procedure.

First, the regions where values of Γ_L and Γ_s produce $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$ are determined, respectively. Setting the magnitude of (3.3.3) and (3.3.4) equal to 1 and solving for the values of Γ_L and Γ_s shows that the solutions for Γ_L and Γ_s lie on circles (called *stability circles*) whose equations are given by

$$\left| \Gamma_L - \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right| \quad (3.3.5)$$

and

$$\left| \Gamma_s - \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12}S_{21}}{|S_{11}|^2 - |\Delta|^2} \right| \quad (3.3.6)$$

where

$$\Delta = S_{11}S_{22} - S_{12}S_{21}$$

The derivations of (3.3.5) and (3.3.6) are given in Appendix A, Section A.2. For completeness, the reader can find in Appendix A, Section A.1, a review of circle equations and bilinear transformations.

The radii and centers of the circles where $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$ in the Γ_L plane and Γ_s plane, respectively, are obtained from (3.3.5) and (3.3.6), namely

Γ_L values for $|\Gamma_{IN}| = 1$ (Output Stability Circle):

$$r_L = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right| \quad (\text{radius}) \quad (3.3.7)$$

$$C_L = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \quad (\text{center}) \quad (3.3.8)$$

Γ_s values for $|\Gamma_{OUT}| = 1$ (Input Stability Circle):

$$r_s = \left| \frac{S_{12}S_{21}}{|S_{11}|^2 - |\Delta|^2} \right| \quad (\text{radius}) \quad (3.3.9)$$

$$C_s = \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \quad (\text{center}) \quad (3.3.10)$$

With the S parameters of a two-port device at one frequency, the expressions (3.3.7) to (3.3.10) can be calculated and plotted on a Smith chart, and the set of values of Γ_L and Γ_s that produce $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$ can be easily observed. Figure 3.3.2 illustrates the graphical construction of the stability circles where $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$. On one side of the stability circle boundary, in the Γ_L plane, we will have $|\Gamma_{IN}| < 1$ and on the other side $|\Gamma_{IN}| > 1$. Similarly, in the Γ_s plane on one side of the stability circle boundary, we will have $|\Gamma_{OUT}| < 1$ and on the other side $|\Gamma_{OUT}| > 1$.

Next we need to determine which area in the Smith chart represents the stable region—in other words, the region where values of Γ_L (where $|\Gamma_L| < 1$) produce $|\Gamma_{IN}| < 1$ and where values of Γ_s (where $|\Gamma_s| < 1$) produce $|\Gamma_{OUT}| < 1$. To this end, we observe that if $Z_L = Z_o$, then $\Gamma_L = 0$ and from (3.2.5) $|\Gamma_{IN}| = |S_{11}|$. If the magnitude of S_{11} is less than 1, then $|\Gamma_{IN}| < 1$ when $\Gamma_L = 0$. That is, the center of the Smith chart in Fig. 3.3.2a represents a stable operating point, because for $\Gamma_L = 0$ it follows that $|\Gamma_{IN}| < 1$. On the other hand, if $|S_{11}| > 1$ when $Z_L = Z_o$, then $|\Gamma_{IN}| > 1$ when $\Gamma_L = 0$ and the center of the Smith chart represents an unstable operating point. Figure 3.3.3 illustrates the two cases discussed. The shaded area represents the values of Γ_L that produce a stable operation. Similarly, Fig. 3.3.4 illustrates stable and unstable regions for Γ_s .

For unconditional stability any passive load or source in the network must produce a stable condition. From a graphical point of view, for $|S_{11}| < 1$ and $|S_{22}| < 1$, we want the stability circles shown in Figs. 3.3.3a and 3.3.4a to fall completely outside (or to completely enclose) the Smith chart. The case in which the stability circles fall completely outside the Smith chart is illustrated in Fig. 3.3.5. Therefore, the conditions for unconditional stability for all passive sources and loads can be expressed in the form

$$|C_L| - r_L > 1 \quad \text{for } |S_{11}| < 1 \quad (3.3.11)$$

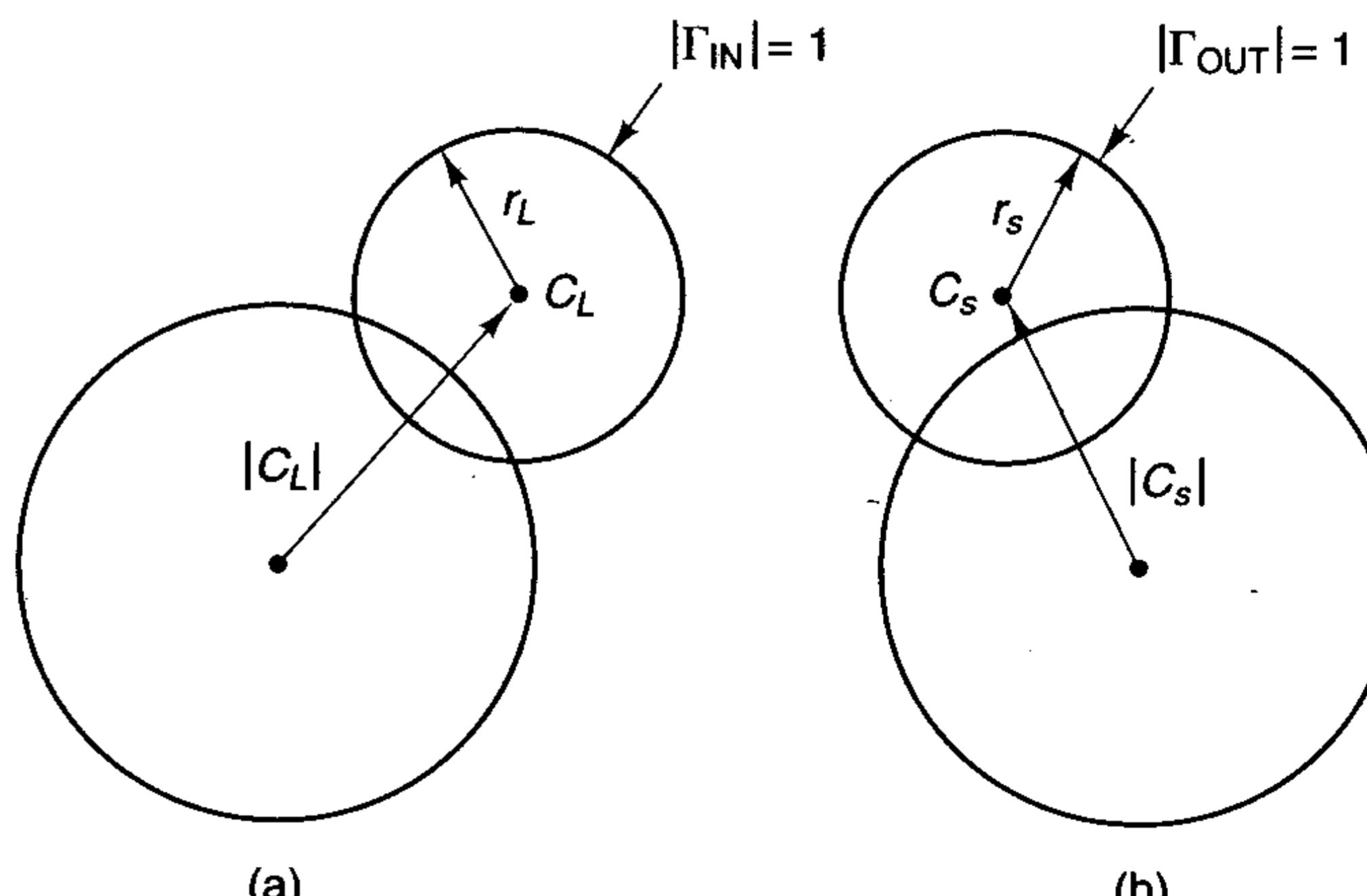


Figure 3.3.2 Stability circle construction in the Smith chart: (a) Γ_L plane; (b) Γ_s plane.

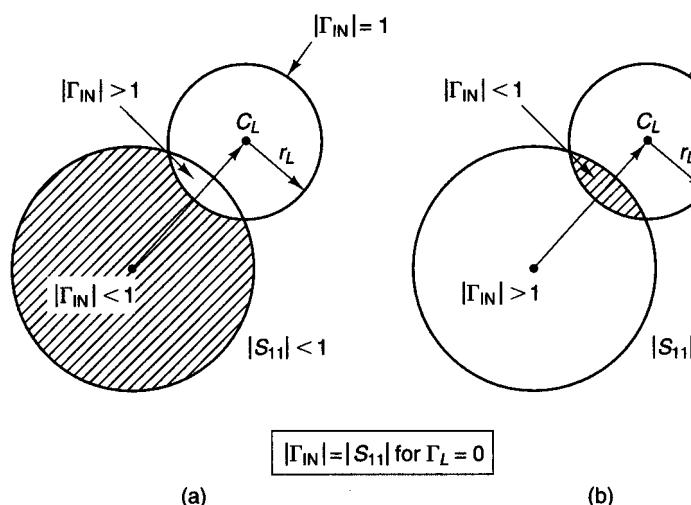


Figure 3.3.3 Smith chart illustrating stable and unstable regions in the Γ_L plane.

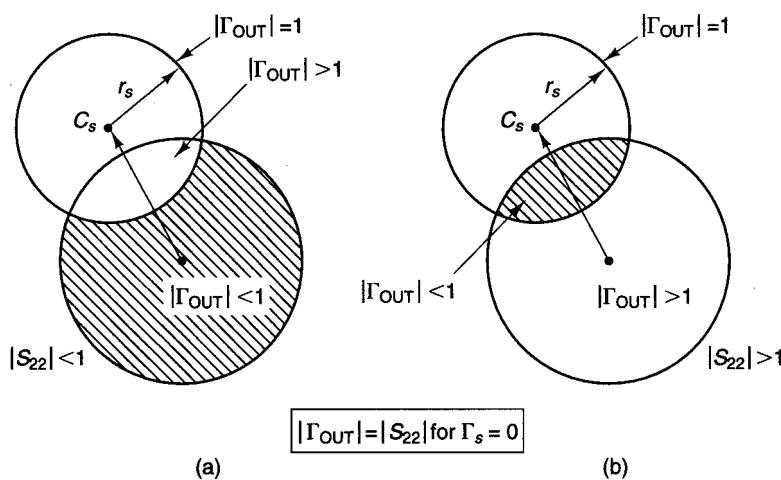


Figure 3.3.4 Smith chart illustrating stable and unstable regions in the Γ_s plane.

and

$$|C_s| - r_s > 1 \quad \text{for } |S_{22}| < 1 \quad (3.3.12)$$

If either $|S_{11}| > 1$ or $|S_{22}| > 1$, the network cannot be unconditionally stable because the termination $\Gamma_L = 0$ or $\Gamma_s = 0$ [see (3.3.3) and (3.3.4)] will produce $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$.

We now return to the necessary and sufficient conditions for a two-port to be unconditionally stable. A straightforward but somewhat lengthy manipu-

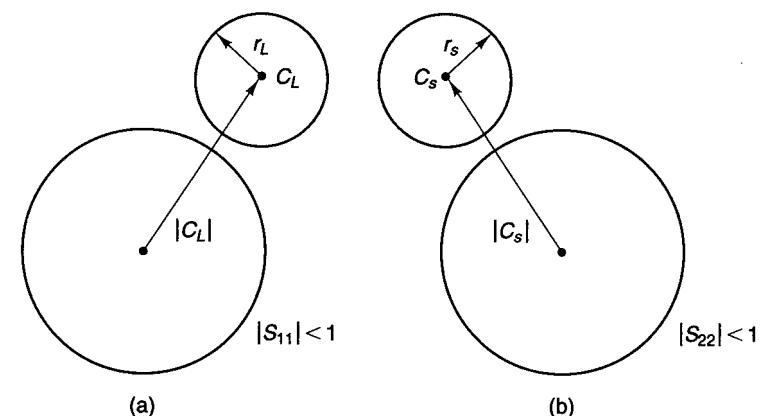


Figure 3.3.5 Conditions for unconditional stability: (a) Γ_L plane; (b) Γ_s plane.

lation of (3.3.1) to (3.3.4) results in the following necessary and sufficient conditions for unconditional stability (see Appendix B):

$$K > 1 \quad (3.3.13)$$

and

$$1 - |S_{11}|^2 > |S_{12}S_{21}| \quad (3.3.14)$$

$$1 - |S_{22}|^2 > |S_{12}S_{21}| \quad (3.3.15)$$

where

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} \quad (3.3.16)$$

and

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \quad (3.3.17)$$

There are other ways of expressing the necessary and sufficient conditions for unconditional stability [3.1]. Adding (3.3.14) and (3.3.15) gives

$$2 - |S_{11}|^2 - |S_{22}|^2 > 2|S_{12}S_{21}| \quad (3.3.18)$$

Since

$$|\Delta| = |S_{11}S_{22} - S_{12}S_{21}| \leq |S_{11}S_{22}| + |S_{12}S_{21}|$$

we use (3.3.18) to obtain

$$|\Delta| < |S_{11}S_{22}| + 1 - \frac{1}{2}|S_{11}|^2 - \frac{1}{2}|S_{22}|^2$$

$$|\Delta| < 1 - \frac{1}{2}(|S_{11}| - |S_{22}|)^2$$

or simply

$$|\Delta| < 1$$

Hence, a convenient way of expressing the necessary and sufficient conditions for unconditional stability is

$$K > 1 \quad (3.3.19)$$

and

$$|\Delta| < 1 \quad (3.3.20)$$

In this textbook we use (3.3.19) and (3.3.20) to test for unconditional stability.

Another way that the necessary and sufficient conditions for unconditional stability are found in the literature is (see Appendix C)

$$K > 1$$

and

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 > 0 \quad (3.3.21)$$

From a theoretical point of view, a two-port network can have any value of K and $|\Delta|$. From a practical point of view, most microwave transistors produced by manufacturers are either unconditionally stable or potentially unstable with $K < 1$ and $|\Delta| < 1$. In fact, in potentially unstable transistors most practical values of K are such that $0 < K < 1$. These potentially unstable transistors have source and load stability circles that intersect the boundary of the Smith chart (e.g., see Figs. 3.3.3a and 3.3.4a).

Negative values of K in the range $-1 < K < 0$ result in most of the Smith chart being unstable. Some transistor configurations (e.g., some CB configurations) used in oscillator designs are potentially unstable with negative values of K .

Example 3.3.1

The S parameters of a BJT at $V_{CE} = 15$ V and $I_C = 15$ mA at $f = 500$ MHz, 1 GHz, 2 GHz, and 4 GHz are as follows:

f (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
0.5	$0.761[-151^\circ]$	$0.025[31^\circ]$	$11.84[102^\circ]$	$0.429[-35^\circ]$
1	$0.770[-166^\circ]$	$0.029[35^\circ]$	$6.11[89^\circ]$	$0.365[-34^\circ]$
2	$0.760[-174^\circ]$	$0.040[44^\circ]$	$3.06[74^\circ]$	$0.364[-43^\circ]$
4	$0.756[-179^\circ]$	$0.064[48^\circ]$	$1.53[53^\circ]$	$0.423[-66^\circ]$

Determine the stability. If the transistor is potentially unstable at a given frequency, draw the input and output stability circles.

Solution. At $f = 500$ MHz it follows from (3.3.16) and (3.3.17) that $K = 0.482$ and $|\Delta| = 0.221[-123^\circ]$. Therefore, for this transistor at 500 MHz we have $K < 1$ and $|\Delta| < 1$. Since $K < 1$, the transistor is potentially unstable. Using (3.3.10), the center of the input stability circle is

$$C_s = \frac{(0.761[-151^\circ] - 0.221[-123^\circ](0.429[35^\circ]))^*}{(0.761)^2 - (0.221)^2} = 1.36[157.6^\circ]$$

and from (3.3.9) the radius is

$$r_s = \frac{|0.025[31^\circ](11.84[102^\circ])|}{(0.761)^2 - (0.221)^2} = 0.558$$

Similarly, from (3.3.8) and (3.3.7) the center and radius of the output stability circles are $C_L = 2.8[57.86^\circ]$ and $r_L = 2.18$.

At $f = 1$ GHz, we find that $K = 0.857$ and $|\Delta| = 0.173[-162.9^\circ]$. Therefore, since $K < 1$, the transistor is potentially unstable at 1 GHz. The center and radius of the input stability circle at 1 GHz are $C_s = 1.28[169^\circ]$ and $r_s = 0.315$; and for the output stability circle $C_L = 2.62[51.3^\circ]$ and $r_L = 1.71$.

At $f = 2$ GHz, we find that $K = 1.31$ and $|\Delta| = 0.174[160^\circ]$. Since $K > 1$ and $|\Delta| < 1$, it follows from (3.3.19) and (3.3.20) that the transistor is unconditionally stable at 2 GHz.

At $f = 4$ GHz, we find that $K = 1.535$ and $|\Delta| = 0.226[121^\circ]$. Therefore, since $K > 1$ and $|\Delta| < 1$, the transistor is unconditionally stable at 4 GHz.

The stability circles are plotted in Fig. 3.3.6 at $f = 500$ MHz and $f = 1$ GHz.

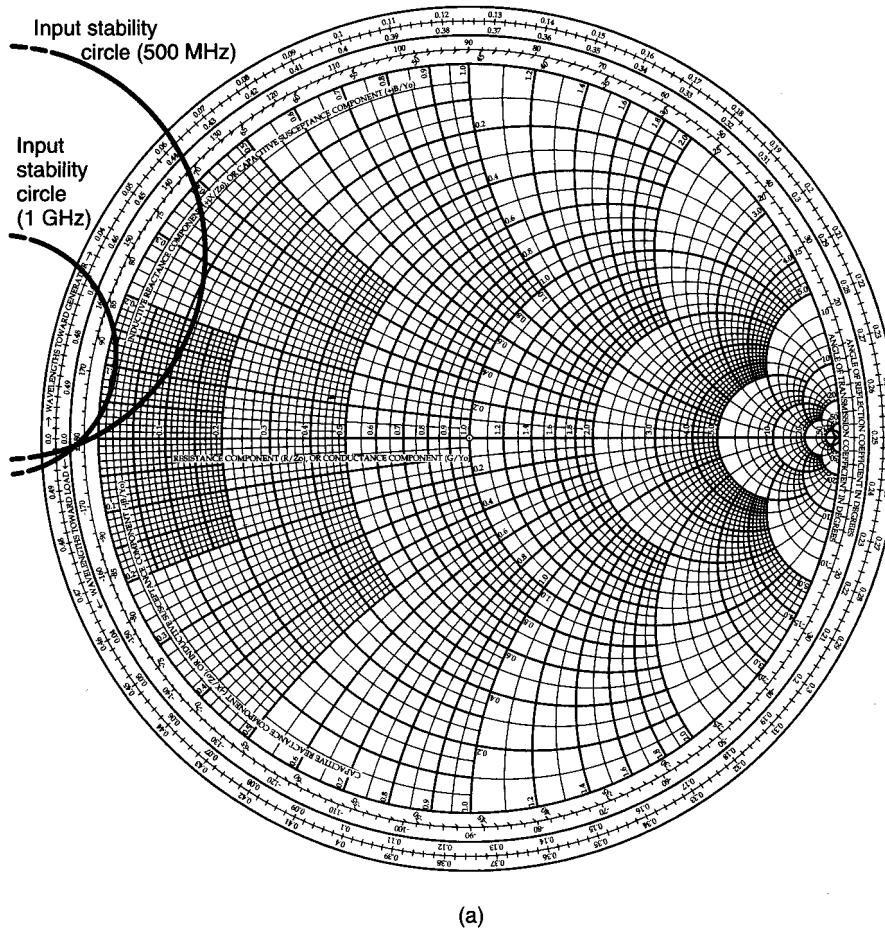
Manufacturers do not fabricate transistors with $K > 1$ and $|\Delta| > 1$. However, the addition of certain feedback networks or certain terminations to a transistor can create a potentially unstable two-port network with $K > 1$ and $|\Delta| > 1$. For example, a two-port network whose S parameters are $S_{11} = 0.75[-60^\circ]$, $S_{21} = 6[90^\circ]$, $S_{22} = 0.5[60^\circ]$, and $S_{12} = 0.3[70^\circ]$ has $K = 1.344$ and $|\Delta| = 2.156$. This two-port network has $K > 1$; however, it is potentially unstable because $|\Delta| > 1$. In fact, the input and output stability circles, from (3.3.7) to (3.3.10), are located at $C_s = 0.1[107.4^\circ]$, $r_s = 0.44$, $C_L = 0.26[-36.3^\circ]$, and $r_L = 0.41$, which can be drawn inside the Smith chart to show the unstable regions. The evaluation of B_1 in (3.3.21) gives $B_1 = -3.34$. Hence, when $|\Delta| > 1$ it follows that $B_1 < 0$. This shows that when the condition (3.3.20) for stability is not satisfied, neither is the condition (3.3.21).

A two-port network is said to be unilateral when $S_{12} = 0$. In a unilateral two-port network, $\Gamma_{IN} = S_{11}$ and $\Gamma_{OUT} = S_{22}$. Hence, we have unconditional stability if $|S_{11}| < 1$ and $|S_{22}| < 1$ for all passive source and load terminations. In fact, from (3.3.16) and (3.3.17), with $S_{12} = 0$ we have $K = \infty$ and $|\Delta| = S_{11}S_{22}$, and it follows from (3.3.19) that

$$1 - |S_{11}|^2 - |S_{22}|^2 + |S_{11}S_{22}|^2 > 0$$

or

$$(1 - |S_{11}|^2)(1 - |S_{22}|^2) > 0$$



(a)

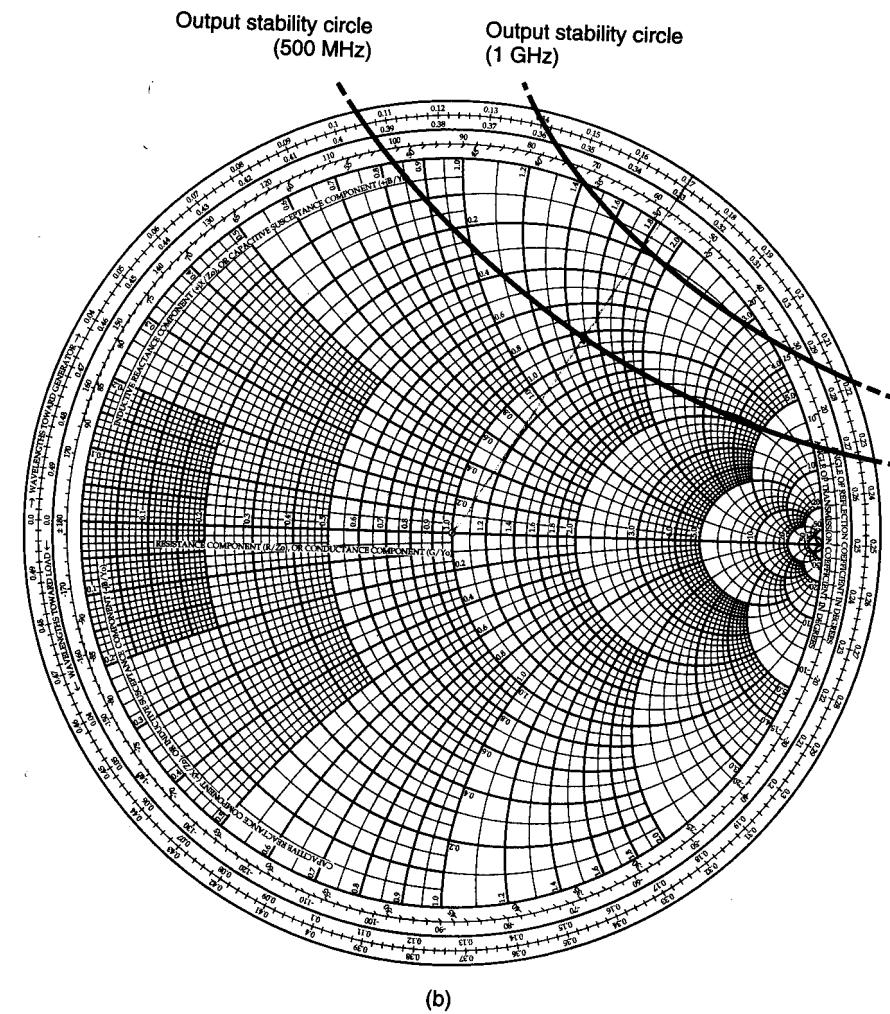
Figure 3.3.6 (a) Input stability circles for Example 3.3.1; (b) output stability circles.

The preceding inequality requires that $|S_{11}| < 1$ and $|S_{22}| < 1$ for unconditional stability of a unilateral two-port.

It is interesting to see what happens to the stability circles in the limit as $S_{12} \rightarrow 0$. The reader is referred to Problem 3.10 for this analysis.

In the potentially unstable situation illustrated in Figs. 3.3.3 and 3.3.4, the real part of the input and output impedances can be negative for some source and load reflection coefficients. In this case, selecting Γ_s and Γ_L in the stable region produces a stable operation.

Even when the selection of Γ_L and Γ_s produces $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$, the circuit can be made stable if the total input and output loop resistance in Fig. 3.3.1 is positive. In other words, the circuit is stable if

**Figure 3.3.6** Continued

and

$$\operatorname{Re}(Z_s + Z_{IN}) > 0$$

$$\operatorname{Re}(Z_L + Z_{OUT}) > 0$$

A potentially unstable transistor can be made unconditionally stable by either resistively loading the transistor or by adding negative feedback. These techniques are not recommended in narrowband amplifiers because of the resulting degradation in power gain, noise figure, and VSWRs. Narrowband

amplifier design with potentially unstable transistors is best done by the proper selection of Γ_s and Γ_L to ensure stability. On the other hand, the techniques are popular in the design of some broadband amplifiers in which the transistor is potentially unstable.

The following example illustrates how resistive loading can stabilize a potentially unstable transistor.

Example 3.3.2

The S parameters of a transistor at $f = 800$ MHz are

$$S_{11} = 0.65 \angle -95^\circ$$

$$S_{12} = 0.035 \angle 40^\circ$$

$$S_{21} = 5 \angle 115^\circ$$

$$S_{22} = 0.8 \angle -35^\circ$$

Determine the stability and show how resistive loading can stabilize the transistor.

Solution. From (3.3.16) and (3.3.17) we find that $K = 0.547$ and $\Delta = 0.504$ 249.6° . Since $K < 1$, the transistor is potentially unstable at $f = 800$ MHz.

The input and output stability circles are calculated using (3.3.7) to (3.3.10):

$$C_s = 1.79 \angle 122^\circ \quad C_L = 1.3 \angle 48^\circ$$

$$r_s = 1.04 \quad r_L = 0.45$$

Figure 3.3.7 shows the plot of the stability circles, together with the stable region.

For the input stability circle the Smith chart in Fig. 3.3.7 represents the Γ_s plane and for the output stability circle the Γ_L plane. It can be seen that a series resistor with the input of approximately 9Ω assures stability at the input (see Fig. 3.3.8). The series addition of a 9Ω resistor produces an impedance Z_s' equal to $Z_s + 9\Omega$. For any passive termination Z_s' , the real part of Z_s will be greater than 9Ω . Therefore, its associated reflection coefficient Γ_s will always be in the stable region in Fig. 3.3.7.

Also, a shunt resistor with the input of approximately $0.7/50 = 14 \text{ mS}$ (or 71.5Ω) produces stability at the input. Looking at the output stability circle, it follows that either a series resistor of approximately 29Ω or a shunt resistor of approximately 500Ω at the output produces stability at the output. The four choices of resistive loading are shown in Fig. 3.3.9. Usually, stabilizing one port of a transistor results in an unconditionally stable device.

All four choices of resistive loading affect the gain performance of the amplifier. In addition, from a practical point of view, resistive loading at the input (as shown in Figs. 3.3.9a and 3.3.9b) is not used because it produces a significant deterioration in the noise performance of the amplifier (see Chapter 4).

In some potentially unstable designs of broadband amplifiers, the shunt resistor loading at the output, as shown in Fig. 3.3.9d, produces a trade-off between gain and stability that is quite acceptable, resulting in a stable two-port with reasonable gain over a wide bandwidth (see Example 4.4.3).

For the stabilized shunt resistor configuration in Fig. 3.3.9d (i.e., with a 500Ω shunt resistor), the resulting S parameters are

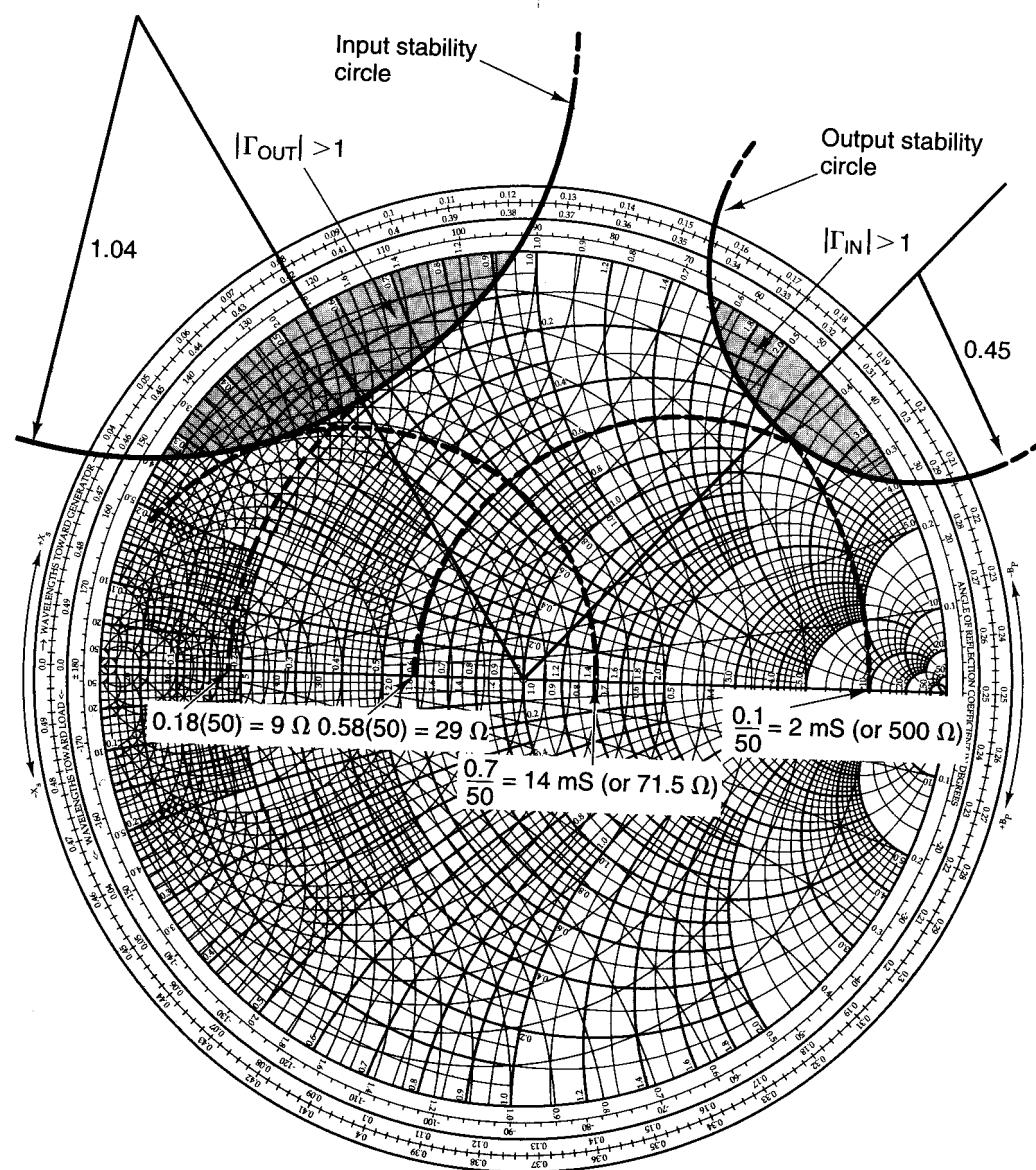


Figure 3.3.7 Input and output stability circles.

$$S_{11} = 0.65 \angle -94^\circ$$

$$S_{12} = 0.032 \angle 41.2^\circ$$

$$S_{21} = 4.62 \angle 116.2^\circ$$

$$S_{22} = 0.66 \angle -36^\circ$$

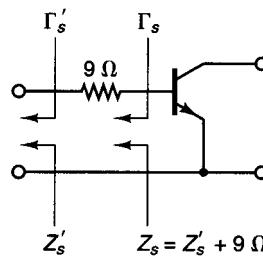


Figure 3.3.8 Series resistive loading of the transistor at the input port.

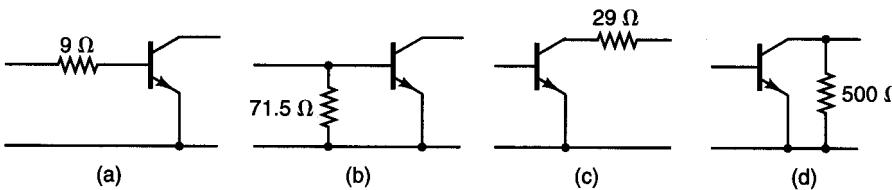


Figure 3.3.9 Four types of resistive loading to improve stability.

and from (3.3.16) and (3.3.17) $K = 1.04$ and $\Delta = 0.409|250.13^\circ|$, which show that the stabilized network in Fig. 3.3.9d is unconditionally stable at $f = 800$ MHz.

Negative feedback can be used to stabilize a transistor by neutralizing S_{12} —that is, by making $S_{12} = 0$. However, this is not commonly done. In a broadband amplifier design using a potentially unstable transistor, a common procedure is to use resistive loading to stabilize the transistor and negative feedback to provide the proper ac performance—that is, to provide constant gain and low input and output VSWR.

3.4 CONSTANT-GAIN CIRCLES: UNILATERAL CASE

A two-port network is unilateral when $S_{12} = 0$. In a unilateral transistor, $\Gamma_{\text{IN}} = S_{11}$, $\Gamma_{\text{OUT}} = S_{22}$, and the unilateral transducer power gain from (3.2.1) and (3.2.2), called G_{TU} , is given by

$$G_{TU} = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.4.1)$$

The first term in (3.4.1) depends on the S_{11} parameter of the transistor and the source reflection coefficient. The second term, $|S_{21}|^2$, depends on the transistor scattering parameter S_{21} ; and the third term depends on the S_{22} parameter of the transistor and the load reflection coefficient. We can think of (3.4.1) as being composed of three distinct and independent gain terms. Therefore, we can write (3.4.1) in the form

$$G_{TU} = G_s G_o G_L \quad (3.4.2)$$

where

$$G_s = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} \quad (3.4.3)$$

$$G_o = |S_{21}|^2 \quad (3.4.4)$$

$$G_L = \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (3.4.5)$$

and the microwave amplifier can be represented by three different gain (or loss) blocks, as shown in Fig. 3.4.1. The input matching network determines Γ_s and therefore the value of G_s according to (3.4.3); the transistor gain is $G_o = |S_{21}|^2$; and the output matching network determines Γ_L and therefore the value of G_L according to (3.4.5).

The terms G_s and G_L represent the gain or loss produced by the matching or mismatching of the input or output circuits, respectively. The term G_o affects the degree of matching or mismatching between Γ_s and S_{11} . Although the G_s block is made up of passive components, it can either have a gain contribution greater than unity or a loss. The reason we usually refer to G_s as a gain block is that there is an intrinsic mismatch loss between Z_o , the matching network, and S_{11} (i.e., between Γ_s and S_{11}). Therefore, decreasing the mismatch loss can be thought of as providing a gain. Similarly, the term G_L affects the output matching and can be thought of as the output gain block. The term G_o is related to the device and is equal to $|S_{21}|^2$. In terms of decibels, we can write from (3.4.2) to (3.4.5)

$$G_{TU}(\text{dB}) = G_s(\text{dB}) + G_o(\text{dB}) + G_L(\text{dB})$$

If we optimize Γ_s and Γ_L to provide maximum gain in G_s and G_L , we refer to the gain as the maximum unilateral transducer power gain, called $G_{TU,\max}$. For a unilateral unconditional stable transistor (i.e., for $|S_{11}| < 1$ and $|S_{22}| < 1$), the maximum values of G_s and G_L are obtained when (see Problem 3.15).

$$\Gamma_s = S_{11}^*$$

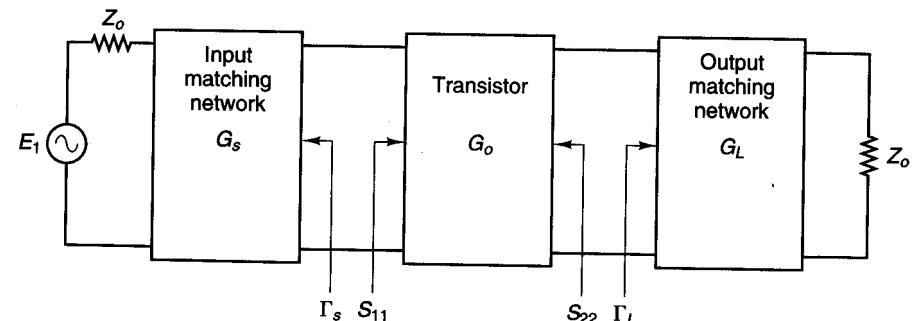


Figure 3.4.1 Unilateral transducer power gain block diagram.

and

$$\Gamma_L = S_{22}^*$$

Therefore, from (3.4.3) and (3.4.5) we obtain

$$G_{s,\max} = \frac{1}{1 - |S_{11}|^2}$$

$$G_{L,\max} = \frac{1}{1 - |S_{22}|^2}$$

and (3.4.2) gives

$$\begin{aligned} G_{TU,\max} &= G_{s,\max} G_o G_{L,\max} \\ &= \frac{1}{1 - |S_{11}|^2} |S_{21}|^2 \frac{1}{1 - |S_{22}|^2} \end{aligned} \quad (3.4.6)$$

The appropriate block diagram for (3.4.6) is shown in Fig. 3.4.2.

Observing that in the unilateral case $\Gamma_{IN} = S_{11}$ and $\Gamma_{OUT} = S_{22}$, the maximum value of G_{TU} , which occurs when $\Gamma_s = S_{11}^* = \Gamma_{IN}^*$ and $\Gamma_L = S_{22}^* = \Gamma_{OUT}^*$, is equal to the maximum value of G_p and G_A [see (3.2.3) and (3.2.4)]. That is, $G_{TU,\max} = G_{pU,\max} = G_{AU,\max}$.

The unilateral transducer power gain is given by (3.4.1) or (3.4.2), and the maximum unilateral transducer power gain, obtained when $\Gamma_s = S_{11}^*$ and $\Gamma_L = S_{22}^*$, is given by (3.4.6). The expressions for G_s and G_L in (3.4.3) and (3.4.5) are similar in form and can be written in the general form

$$G_i = \frac{1 - |\Gamma_i|^2}{|1 - S_{ii}\Gamma_i|^2} \quad (3.4.7)$$

where $i = s$ with $ii = 11$ or $i = L$ with $ii = 22$. The design for a specific gain is based on (3.4.7).

Two cases must be considered in the analysis of (3.4.7): the unconditionally stable case, where $|S_{ii}| < 1$, and the potentially unstable case, where $|S_{ii}| > 1$.

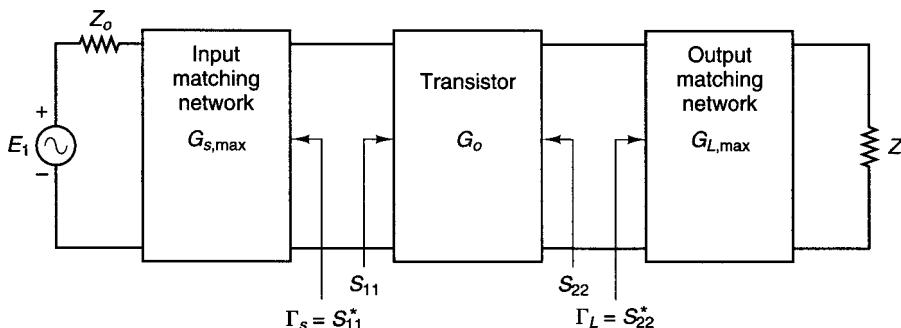


Figure 3.4.2 Maximum unilateral transducer power gain block diagram.

Unconditionally Stable Case, $|S_{ii}| < 1$

The maximum value of (3.4.7) is obtained when $\Gamma_i = S_{ii}^*$, and it is given by

$$G_{i,\max} = \frac{1}{1 - |S_{ii}|^2} \quad (3.4.8)$$

The terminations that produce $G_{i,\max}$ are called the *optimum terminations*.

From (3.4.7), G_i has a minimum value of zero when $|\Gamma_i| = 1$. Other values of Γ_i produce values of G_i between zero and $G_{i,\max}$. That is,

$$0 \leq G_i \leq G_{i,\max}$$

The values of Γ_i that produce a constant gain G_i will be shown to lie in a circle in the Smith chart. These circles are called *constant G_i circles* (i.e., for $i = s$ the circles are *constant G_s circles* and for $i = L$ the circles are *constant G_L circles*).

Define the normalized gain factor as

$$g_i = \frac{G_i}{G_{i,\max}} = G_i(1 - |S_{ii}|^2) = \frac{1 - |\Gamma_i|^2}{|1 - S_{ii}\Gamma_i|^2}(1 - |S_{ii}|^2) \quad (3.4.9)$$

such that

$$0 \leq g_i \leq 1$$

In Appendix D it is shown that the values of Γ_i that produce a constant value of g_i in (3.4.9) lie in a circle whose equation is

$$|\Gamma_i - C_{g_i}| = r_{g_i} \quad (3.4.10)$$

where the center of the circle is given by

$$C_{g_i} = \frac{g_i S_{ii}^*}{1 - |S_{ii}|^2(1 - g_i)} \quad (3.4.11)$$

and the radius is

$$r_{g_i} = \frac{\sqrt{1 - g_i}(1 - |S_{ii}|^2)}{1 - |S_{ii}|^2(1 - g_i)} \quad (3.4.12)$$

Each constant value of g_i generates a new constant G_i circle. Equations (3.4.11) and (3.4.12) can be used to generate the constant G_s circles and the constant G_L circles.

Figure 3.4.3 illustrates a constant G_i circle. The distance from the origin to the center of a constant G_i circle is given by $|C_{g_i}|$ in (3.4.11), and the angle of inclination, α_i , is equal to the phase of C_{g_i} (which is the phase of S_{ii}^*).

It is observed that when $g_i = 1$ (i.e., when $G_i = G_{i,\max}$), (3.4.12) gives $r_{g_i} = 0$ and (3.4.11) gives $C_{g_i} = S_{ii}^*$. Therefore, the constant G_i circle for maximum gain is represented by a point, located at S_{ii}^* .

In conclusion, the procedure for drawing the constant G_i circles in the Z Smith chart is as follows:

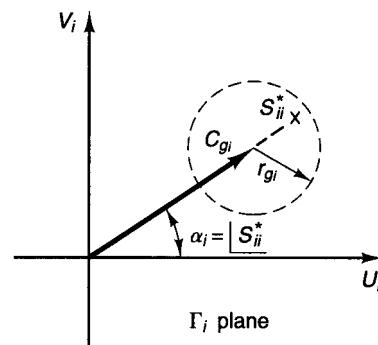


Figure 3.4.3 A constant G_i circle in the Smith chart.

1. Locate S_{ii}^* and draw a line from the origin to S_{ii}^* . At S_{ii}^* , the gain is $G_{i,\max}$ and is given by (3.4.8).
2. Determine the values of G_i , where $0 \leq G_i \leq G_{i,\max}$, for which the constant G_i circles are to be drawn, and calculate the corresponding values of $g_i = G_i/G_{i,\max}$.
3. From (3.4.11), determine the values of C_{gi} for each g_i .
4. From (3.4.12), determine the values of r_{gi} for each g_i .

The 0-dB circle ($G_i = 1$) always passes through the origin of the Smith chart. This is not a coincidence. In fact, $G_i = 1$ occurs when $\Gamma_i = 0$, and from (3.4.9)

$$g_{i,0 \text{ dB}} = 1 - |S_{ii}|^2$$

Then, from (3.4.11) and (3.4.12),

$$r_{g_i,0 \text{ dB}} = |C_{g_i,0 \text{ dB}}| = \frac{|S_{ii}|}{1 + |S_{ii}|^2}$$

which shows that the radius and the distance from the origin to the center of the 0-dB constant G_i circle are identical.

A typical set of constant G_s circles is calculated in the following example and shown in Fig. 3.4.4.

Example 3.4.1

The S parameters of a BJT measured at $V_{CE} = 10 \text{ V}$, $I_C = 30 \text{ mA}$, and $f = 1 \text{ GHz}$, in a $50\text{-}\Omega$ system, are

$$S_{11} = 0.73 \angle 175^\circ$$

$$S_{12} = 0$$

$$S_{21} = 4.45 \angle 65^\circ$$

$$S_{22} = 0.21 \angle -80^\circ$$

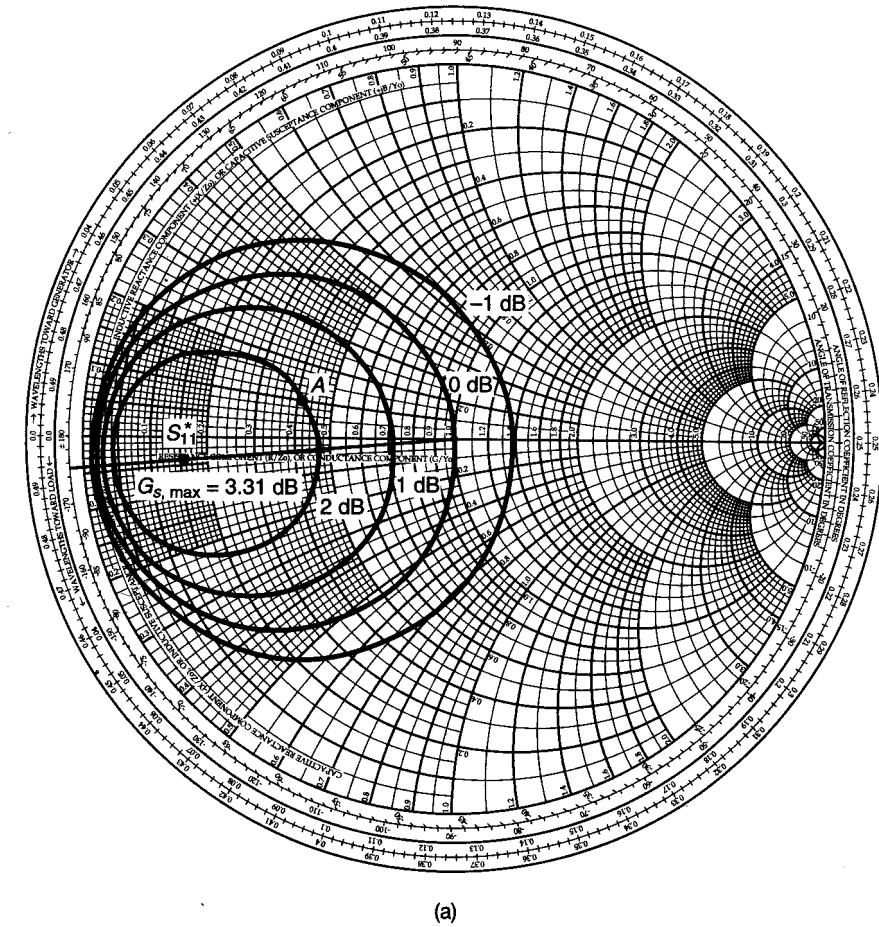


Figure 3.4.4 (a) Constant-gain circles for $G_s = 2, 1, 0$, and -1 dB ; (b) calculations of constant-gain circles.

- (a) Calculate the optimum terminations.
- (b) Calculate $G_{s,\max}$, $G_{L,\max}$, and $G_{TU,\max}$ in decibels.
- (c) Draw several G_s constant-gain circles.
- (d) Design the input matching network for $G_s = 2$ dB.

Solution. (a) The optimum terminations are

$$\Gamma_s = S_{11}^* = 0.73 \angle -175^\circ$$

and

$$\Gamma_L = S_{22}^* = 0.21 \angle 80^\circ$$

Using the Smith chart, the impedances associated with Γ_s and Γ_L are $Z_s = 50(0.152 - j0.047) = 7.6 - j2.35 \Omega$ and $Z_L = 50(0.97 + j0.43) = 48.5 + j21.5 \Omega$.

(b) From (3.4.8) we find that

$$G_{s,\max} = \frac{1}{1 - |S_{11}|^2} = 2.141 \quad \text{or} \quad 3.31 \text{ dB}$$

$$G_{L,\max} = \frac{1}{1 - |S_{22}|^2} = 1.046 \quad \text{or} \quad 0.195 \text{ dB}$$

Since

$$G_o = |S_{21}|^2 = 19.8 \quad \text{or} \quad 12.97 \text{ dB}$$

then

$$G_{TU,\max}(\text{dB}) = 3.31 + 12.97 + 0.195 = 16.47 \text{ dB}$$

(c) Since $G_{s,\max} = 3.31$ dB, constant-gain circles at 2, 1, 0, and -1 dB are drawn in Fig. 3.4.4a. The necessary calculations are given in Fig. 3.4.4b.

(d) In an amplifier design using this transistor, we observe that the output gain block provides little gain (i.e., $G_{L,\max} = 0.195$ dB); therefore, the output matching network is designed to present the optimum termination $\Gamma_L = 0.21 \angle 80^\circ$.

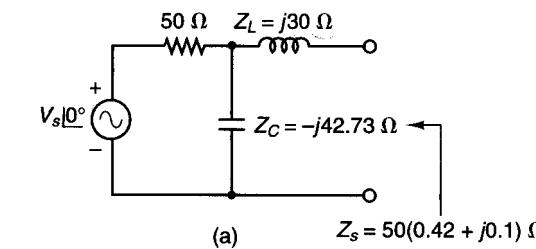
Any Γ_s along the $G_s = 2$ dB circle provides the constant gain. Selecting Γ_s at point A (i.e., $\Gamma_s = 0.413 \angle 166^\circ$ or $z_s = 0.42 + j0.1$) in Fig. 3.4.4a results in the input matching network shown in Fig. 3.4.5a. The details of the matching network design are shown in Fig. 3.4.5b. With $\Gamma_s = 0.413 \angle 166^\circ$ and $\Gamma_L = 0.21 \angle 80^\circ$, it follows that $G_s = 2$ dB and $G_L = G_{L,\max} = 0.195$ dB; hence, the transducer gain of the amplifier is G_{TU} (dB) = $2 + 12.97 + 0.195 = 15.16$ dB.

Potentially Unstable Case, $|S_{ii}| > 1$

In this case $|S_{ii}| > 1$ and it is possible for a passive termination to produce an infinite value of G_i . The infinite value of G_i in (3.4.7) is produced by the critical value of Γ_i , called $\Gamma_{i,c}$, given by

$$\Gamma_{i,c} = \frac{1}{S_{ii}} \quad (3.4.13)$$

Equation (3.4.13) basically states that the real part of the impedance associated with $\Gamma_{i,c}$ is equal to the magnitude of the negative resistance associated



$$Z_s = 50(0.42 + j0.1) \Omega$$

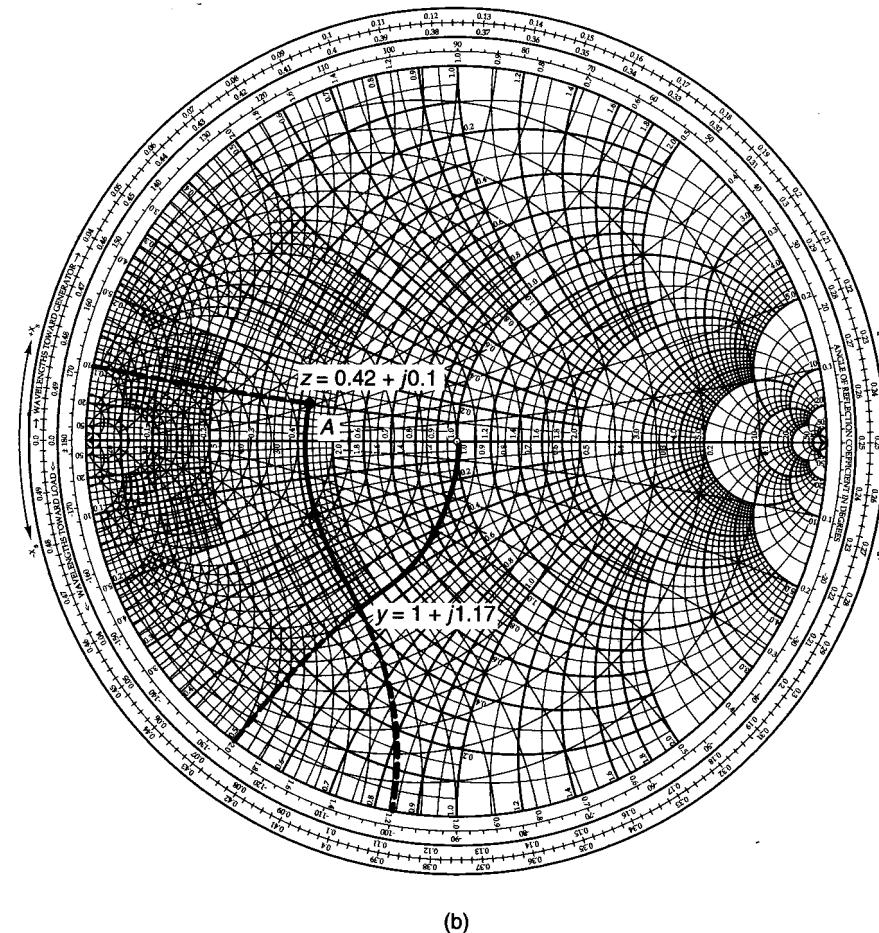


Figure 3.4.5 Input matching network for $G_s = 2$ dB.

with S_{ii} . Therefore, the total input or output loop resistance is zero, and oscillations will occur.

With g_i defined as in (3.4.9)—namely,

$$g_i = G_i(1 - |S_{ii}|^2) = \frac{1 - |\Gamma_i|^2}{|1 - S_{ii}\Gamma_i|^2} (1 - |S_{ii}|^2) \quad (3.4.14)$$

where now g_i can attain negative values because $|S_{ii}| > 1$ —the derivation of a constant G_i circle is identical to (3.4.10). Hence, with g_i given by (3.4.14), the center of a constant G_i circle is given by (3.4.11), and the radius is given by (3.4.12). The gain G_i is infinite at $\Gamma_i = \Gamma_{i,c} = 1/S_{ii}$. Since the argument of C_{gi} (i.e., $|S_{ii}^*$) is identical to the argument of $1/S_{ii}$, it follows that the centers of the circles are located along a line drawn from the origin to the point $1/S_{ii}$.

As discussed in Section 2.2, the negative resistance associated with S_{ii} , where $|S_{ii}| > 1$, can be calculated using the Smith chart by locating the point $1/S_{ii}^*$ and interpreting the resistance circles as being negative and the reactance circles as labeled.

To prevent oscillations in the input or output port Γ_i must be selected such that the real part of the termination impedance is larger than the magnitude of the negative resistance associated with the point $1/S_{ii}^*$. When a negative resistance occurs at the input, the stable region is that region where values of Γ_s produce a source impedance such that

$$\operatorname{Re}(Z_s) > |\operatorname{Re}(Z_{IN})|$$

Similarly, when a negative resistance occurs at the output, Γ_L is selected such that

$$\operatorname{Re}(Z_L) > |\operatorname{Re}(Z_{OUT})|$$

A typical construction is illustrated in Fig. 3.4.6, where the critical value of Γ_s (i.e., $\Gamma_{s,c} = 1/S_{11}$) and two constant G_s circles are shown.

Example 3.4.2

The S parameters of a GaAs FET measured at $V_{DS} = 5$ V, $I_{DS} = 10$ mA, and $f = 1$ GHz in a 50Ω system are

$$S_{11} = 2.27 \angle -120^\circ$$

$$S_{12} = 0$$

$$S_{21} = 4 \angle 50^\circ$$

$$S_{22} = 0.6 \angle -80^\circ$$

- (a) Calculate the input impedance and the optimum output termination.
- (b) Determine the unstable region in the Smith chart and construct constant-gain circles for $G_s = 5$ dB and $G_s = 3$ dB.
- (c) Design the input matching network for $G_s = 3$ dB with the greatest degree of stability.
- (d) Determine G_{TU} in decibels.

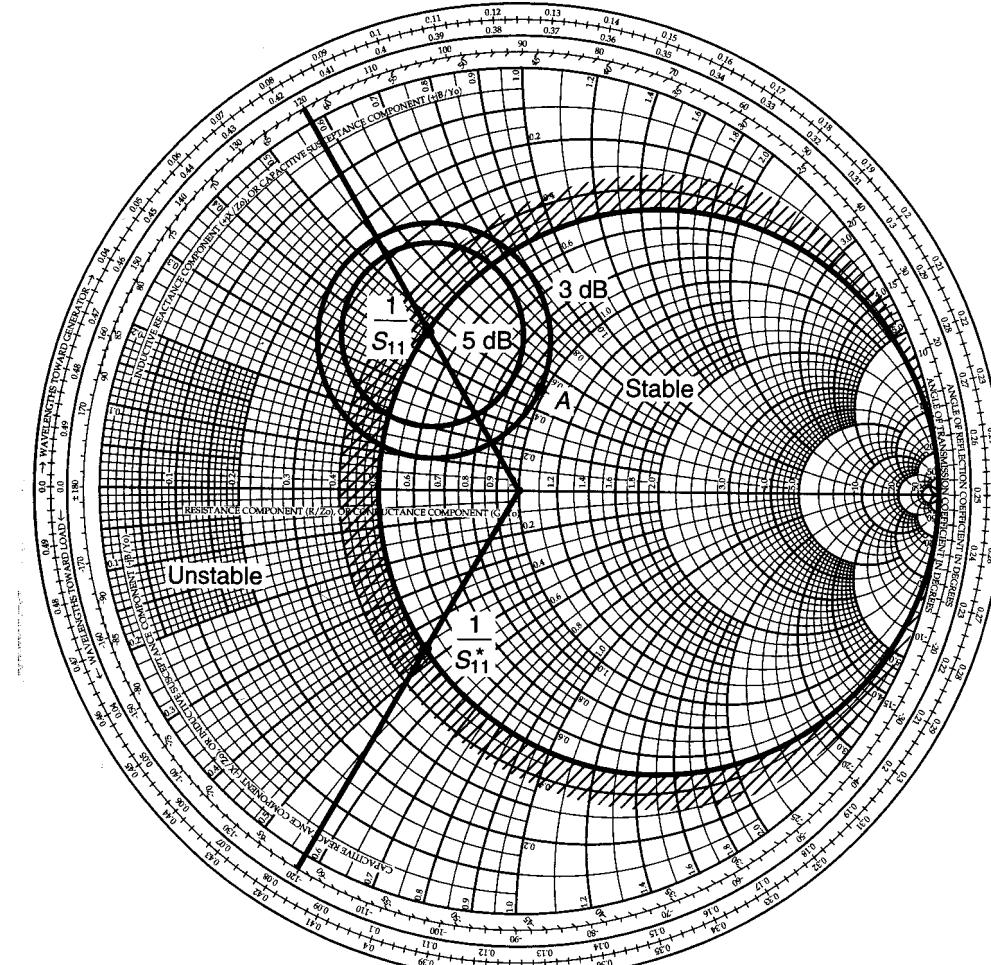


Figure 3.4.6 Stable and unstable regions and constant G_s circles for Example 3.4.2.

Solution. (a) The input impedance is obtained from the Smith chart at the point $1/S_{11}^* = 0.44 \angle -120^\circ$ (see Fig. 3.4.6)—namely,

$$Z_{IN} = 50(-0.5 - j0.46) = -25 - j23 \Omega$$

The optimum termination for G_L is

$$\Gamma_L = S_{22}^* = 0.6 \angle 80^\circ$$

The impedance associated with Γ_L is obtained from the Smith chart as $Z_L = 50(0.56 + j1.03) = 28 + j51.5 \Omega$.

(b) The unstable region is where $\text{Re}(Z_s) < |\text{Re}(Z_{IN})|$. The unstable region is marked in Fig. 3.4.6.

In order to construct the constant-gain circle for $G_s = 5 \text{ dB}$, we first locate the point $1/S_{11}$ in Fig. 3.4.6. Then, from (3.4.14), (3.4.11), and (3.4.12), we find that

$$g_s = 3.16[1 - (2.27)^2] = -13.123$$

$$r_{g_s} = \frac{\sqrt{1 + 13.123}}{1 - (2.27)^2(1 + 13.123)} = 0.217$$

and

$$|C_{g_s}| = \frac{-13.123(2.27)}{1 - (2.27)^2(1 + 13.123)} = 0.415$$

The $G_s = 5 \text{ dB}$ circle is drawn in Fig. 3.4.6. Similarly, for the $G_s = 3 \text{ dB}$ circle, we find that $g_s = -8.286$, $|C_{g_s}| = 0.401$, and $r_{g_s} = 0.27$.

(c) In order to obtain the greatest degree of stability, we select Γ_s on the $G_s = 3 \text{ dB}$ circle such that it has the largest positive real part. That is, Γ_s is selected at point A in Fig. 3.4.6—namely,

$$\Gamma_s = 0.245 \underline{|79^\circ|}$$

or

$$Z_s = 50(0.97 + j0.5) = 48.5 + j25 \Omega$$

Since the input loop resistance is $48.5 - 25 = 23.5 \Omega$, the input port is stable.

(d) Since $G_s = 3 \text{ dB}$,

$$G_{L,\max} = \frac{1}{1 - |S_{22}|^2} = \frac{1}{1 - (0.6)^2} = 1.562 \quad \text{or} \quad 1.94 \text{ dB}$$

and

$$G_o = |S_{21}|^2 = (4)^2 = 16 \quad \text{or} \quad 12.04 \text{ dB}$$

and the unilateral transducer gain is

$$G_{TU}(\text{dB}) = 3 + 12.04 + 1.94 = 16.98 \text{ dB}$$

3.5 UNILATERAL FIGURE OF MERIT

When S_{12} can be set equal to zero, the design procedure is much simpler. In order to determine the error involved in assuming $S_{12} = 0$, we form the magnitude ratio of G_T and G_{TU} from (2.6.13) and (3.4.1)—namely,

$$\frac{G_T}{G_{TU}} = \frac{1}{|1 - X|^2} \quad (3.5.1)$$

where

$$X = \frac{S_{12}S_{21}\Gamma_s\Gamma_L}{(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L)}$$

From (3.5.1) the ratio of the transducer power gain to the unilateral transducer power gain is bounded by

$$\frac{1}{(1 + |X|)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1 - |X|)^2}$$

When $\Gamma_s = S_{11}^*$ and $\Gamma_L = S_{22}^*$, G_{TU} has a maximum value and, in this case, the maximum error introduced when using G_{TU} is bounded by

$$\frac{1}{(1 + U)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1 - U)^2} \quad (3.5.2)$$

where

$$U = \frac{|S_{12}||S_{21}||S_{11}||S_{22}|}{(1 - |S_{11}|^2)(1 - |S_{22}|^2)} \quad (3.5.3)$$

is known as the *unilateral figure of merit*.

The value of U varies with frequency because of its dependence on the S parameters. A typical variation of U with frequency is shown in Fig. 3.5.1. In this case, the maximum value of U occurs at 100 MHz and 1 GHz and is given by $U = -15 \text{ dB}$ or $U = 0.03$. Therefore, from (3.5.2),

$$\frac{1}{(1 + 0.03)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1 - 0.03)^2}$$

or, in decibels,

$$-0.26 \text{ dB} < \frac{G_T}{G_{TU}} < 0.26 \text{ dB}$$

and the maximum error is $\pm 0.26 \text{ dB}$ at 100 MHz and 1 GHz. In some designs this error is small enough to justify the unilateral assumption.

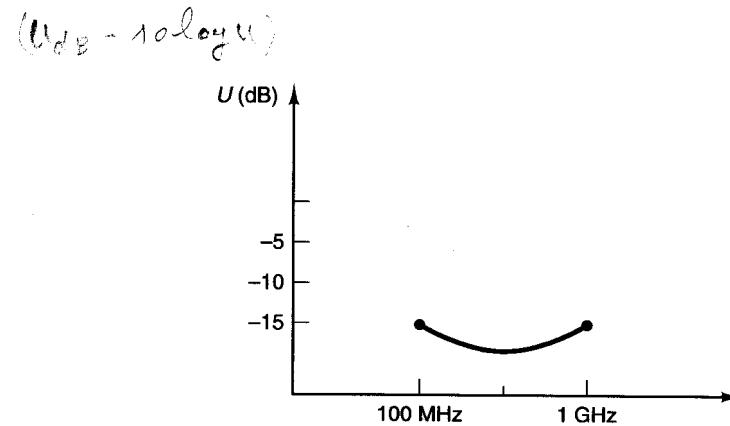


Figure 3.5.1 Frequency dependence of the unilateral figure of merit.

3.6 SIMULTANEOUS CONJUGATE MATCH: BILATERAL CASE

When $S_{12} \neq 0$ and the unilateral assumption cannot be made, the input and output reflection coefficients are given by (3.2.5) and (3.2.6), respectively. The conditions required to obtain maximum transducer power gain are

$$\Gamma_s = \Gamma_{IN}^* \quad (3.6.1)$$

and

$$\Gamma_L = \Gamma_{OUT}^* \quad (3.6.2)$$

These conditions are illustrated in Fig. 3.6.1 and are referred to as the simultaneous conjugate match conditions. When the input and output are matched, it follows that $(VSWR)_{in} = (VSWR)_{out} = 1$.

From (3.2.5), (3.2.6), (3.6.1), and (3.6.2), we can write

$$\Gamma_s^* = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \quad (3.6.3)$$

and

$$\Gamma_L^* = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \quad (3.6.4)$$

Solving (3.6.3) and (3.6.4) simultaneously gives the values Γ_s and Γ_L required for a simultaneous conjugate match. Calling these values Γ_{MS} and Γ_{ML} , we obtain

$$\Gamma_{MS} = \frac{B_1 \pm \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \quad (3.6.5)$$

$$\Gamma_{ML} = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \quad (3.6.6)$$

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 \quad (3.6.7)$$

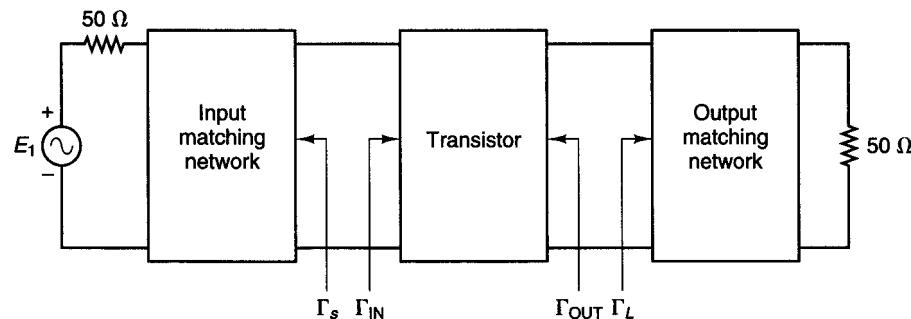


Figure 3.6.1 Simultaneous conjugate match exists when $\Gamma_s = \Gamma_{IN}^*$ and $\Gamma_L = \Gamma_{OUT}^*$.

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2 \quad (3.6.8)$$

$$C_1 = S_{11} - \Delta S_{22}^*$$

$$C_2 = S_{22} - \Delta S_{11}^*$$

In what follows, we will show that for an unconditionally stable two-port network, the solutions with a minus sign in (3.6.5) and (3.6.6) are the useful ones [3.2, 3.3].

If $|B_1/2C_1| > 1$ and $B_1 > 0$ in (3.6.5), the solution with the minus sign produces $|\Gamma_{MS}| < 1$ and the solution with the plus sign produces $|\Gamma_{MS}| > 1$. If $|B_1/2C_1| > 1$ with $B_1 < 0$ in (3.6.5), the solution with the plus sign produces $|\Gamma_{MS}| < 1$ and the solution with the minus sign produces $|\Gamma_{MS}| > 1$. Similar considerations apply to (3.6.6). The proofs for the previous statements are given in Appendix E, Section E.1.

Since it can be shown that $|B_i/2C_i|^2 > 1$ ($i = 1$ or 2) is similar to $K^2 > 1$, it follows that the condition $|B_i/2C_i| > 1$ is similar to $|K| > 1$ (see Appendix E, Section E.2). Therefore, if $|K| > 1$ with K positive, one solution of (3.6.5) and (3.6.6) has a magnitude less than 1, and the other solution has a magnitude greater than 1. In fact, for $K > 1$ and $B_i > 0$, the solutions with the minus sign have magnitudes less than 1. The analysis for $|K| > 1$ with K negative is left as an exercise (see Problem 3.20); it follows that for $K < -1$ a simultaneous conjugate match does not exist.

Associated with Γ_{MS} and Γ_{ML} are a source and a load impedance. The real parts of these impedances are positive if $|\Gamma_{MS}| < 1$ and $|\Gamma_{ML}| < 1$. From the previous considerations, we conclude that in terms of K , the condition that a two-port network can be simultaneously matched with $|\Gamma_{MS}| < 1$ and $|\Gamma_{ML}| < 1$ is

$$K > 1$$

The condition $K > 1$ is only a necessary condition for unconditional stability. Therefore, a simultaneous conjugate match having unconditional stability is possible if $K > 1$ and $|\Delta| < 1$. Since $|\Delta| < 1$ implies that $B_1 > 0$ and $B_2 > 0$, the minus signs must be used in (3.6.5) and (3.6.6) when calculating the simultaneous conjugate match for an unconditionally stable two-port network.

In what follows, any reference to a simultaneous conjugate match assumes that the two-port network is unconditionally stable. In a potentially unstable situation, the design procedure is best done in terms of G_p or G_A (see Section 3.7).

The maximum transducer power gain, under simultaneous conjugate match conditions, is obtained from (3.2.1) with $\Gamma_s = \Gamma_{IN}^* = \Gamma_{MS}$ and $\Gamma_L = \Gamma_{OUT}^* = \Gamma_{ML}$. Thus, we obtain

$$G_{T,max} = \frac{1}{1 - |\Gamma_{MS}|^2} |S_{21}|^2 \frac{1 - |\Gamma_{ML}|^2}{|1 - S_{22}\Gamma_{ML}|^2} \quad (3.6.9)$$

Substituting (3.6.5) and (3.6.6) into (3.6.9), it can be shown (see Appendix F) that $G_{T,max}$ can be expressed in the form

$$G_{T,max} = \frac{|S_{21}|}{|S_{12}|} (K - \sqrt{K^2 - 1}) \quad (3.6.10)$$

Ave. \ominus dars $\Gamma_{MS} + \Gamma_{ML}$

Since under simultaneous conjugate match conditions $G_T = G_p = G_A$, it follows that $G_{T,\max} = G_{p,\max} = G_{A,\max}$.

The maximum stable gain is defined as the value of $G_{T,\max}$ when $K = 1$ —namely,

$$G_{\text{MSG}} = \frac{|S_{21}|}{|S_{12}|} \quad (3.6.11)$$

For a potentially unstable transistor, G_{MSG} is a figure of merit. Figure 3.6.2 illustrates a typical way in which $G_{T,\max} = G_{p,\max} = G_{A,\max}$ and G_{MSG} are given by a manufacturer. At the frequencies where the transistor is unconditionally stable, $G_{A,\max}$ is calculated and plotted in Fig. 3.6.2. In Fig. 3.6.2, $G_{A,\max}$ is denoted by MAG (maximum available gain). At the frequencies where the transistor is potentially unstable, G_{MSG} is plotted in Fig. 3.6.2. From Fig. 3.6.2 it is seen that the transistor is potentially unstable below 1.5 GHz since MSG (i.e., G_{MSG}) is given; and above 2 GHz the transistor is unconditionally stable since MAG is given.

A simultaneous conjugate match does not exist for $K < 1$. However, in a potentially unstable two-port network with $K > 1$ but $|\Delta| > 1$ (which is similar to $B_1 < 0$ and $B_2 < 0$), solutions to (3.6.5) and (3.6.6) using the plus sign produce $|\Gamma_{Ms}| < 1$ and $|\Gamma_{ML}| < 1$. In such a case (i.e., for $K > 1$ and $|\Delta| > 1$), the values of Γ_{Ms} and Γ_{ML} given by (3.6.5) and (3.6.6) using the plus sign result in a minimum value of G_T , and the input and output VSWR are unity. Substituting these values of Γ_{Ms} and Γ_{ML} into (3.6.9), the minimum value of G_T is given by

$$\text{Avec } \oplus \text{ dans } \Gamma_{Ms} \text{ et } \Gamma_{ML} \quad G_{T,\min} = \frac{|S_{21}|}{|S_{12}|} (K + \sqrt{K^2 - 1}) \quad (3.6.12)$$

Recall that in a potentially unstable situation the maximum value of G_T approaches infinity as Γ_s and Γ_L approach the unstable region. Therefore, the expression (3.6.12) gives the minimum value that G_T can have when $K > 1$ and

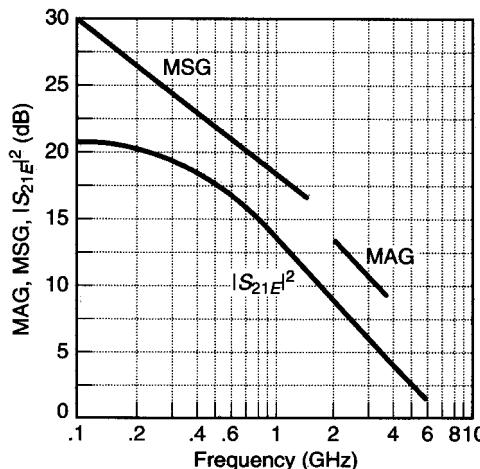


Figure 3.6.2 Typical MAG (i.e., $G_{A,\max}$), MSG (i.e., G_{MSG}), and $|S_{21E}|^2$ versus frequency at $V_{CE} = 18$ V and $I_C = 30$ mA for the HXTR-5103. (From HP Microwave and RF Designer's Catalog 1990-1991; courtesy of Hewlett Packard.)

$|\Delta| > 1$. Also, since the input and output ports are conjugate matched, it follows that $(\text{VSWR})_{\text{in}} = (\text{VSWR})_{\text{out}} = 1$.

Example 3.6.1

Design a microwave amplifier using a GaAs FET to operate $f = 6$ GHz with maximum transducer power gain. The transistor S parameters at the linear bias point, $V_{DS} = 4$ V and $I_{DS} = 0.5I_{DSS}$, are

$$S_{11} = 0.641 \angle -171.3^\circ$$

$$S_{12} = 0.057 \angle 16.3^\circ$$

$$S_{21} = 2.058 \angle 28.5^\circ$$

$$S_{22} = 0.572 \angle -95.7^\circ$$

Solution. From (3.3.16) and (3.3.17), we obtain $K = 1.504$ and $\Delta = 0.3014 \angle 109.88^\circ$. Since $K > 1$ and $|\Delta| < 1$, the GaAs FET is unconditionally stable.

It is of interest to check if the amplifier can be considered unilateral. From (3.5.3), $U = 0.1085$, and from (3.5.2),

$$-0.89 \text{ dB} < \frac{G_T}{G_{TU}} < 1 \text{ dB}$$

The preceding inequality shows that S_{12} cannot be neglected.

The reflection coefficients for a simultaneous conjugate match are calculated from (3.6.5) and (3.6.6) (using the minus sign) as follows:

$$B_1 = 0.9928$$

$$B_2 = 0.8255$$

$$C_1 = 0.4786 \angle -177.3^\circ$$

$$C_2 = 0.3911 \angle -103.9^\circ$$

$$\Gamma_{Ms} = 0.762 \angle 177.3^\circ$$

and

$$\Gamma_{ML} = 0.718 \angle 103.9^\circ$$

The maximum transducer power gain, from (3.6.10), is

$$G_{T,\max} = \frac{2.058}{0.057} (1.504 - \sqrt{(1.504)^2 - 1}) = 13.74 \quad \text{or} \quad 11.38 \text{ dB}$$

The design of the matching networks using microstrip lines is illustrated in Fig. 3.6.3, where the admittances associated with Γ_{Ms} and Γ_{ML} are

$$Y_{Ms} = \frac{7.2 - j1.23}{50} = (144 - j24.6) \times 10^{-3} \text{ S}$$

and

$$Y_{ML} = \frac{0.414 - j1.19}{50} = (8.28 - j23.8) \times 10^{-3} \text{ S}$$

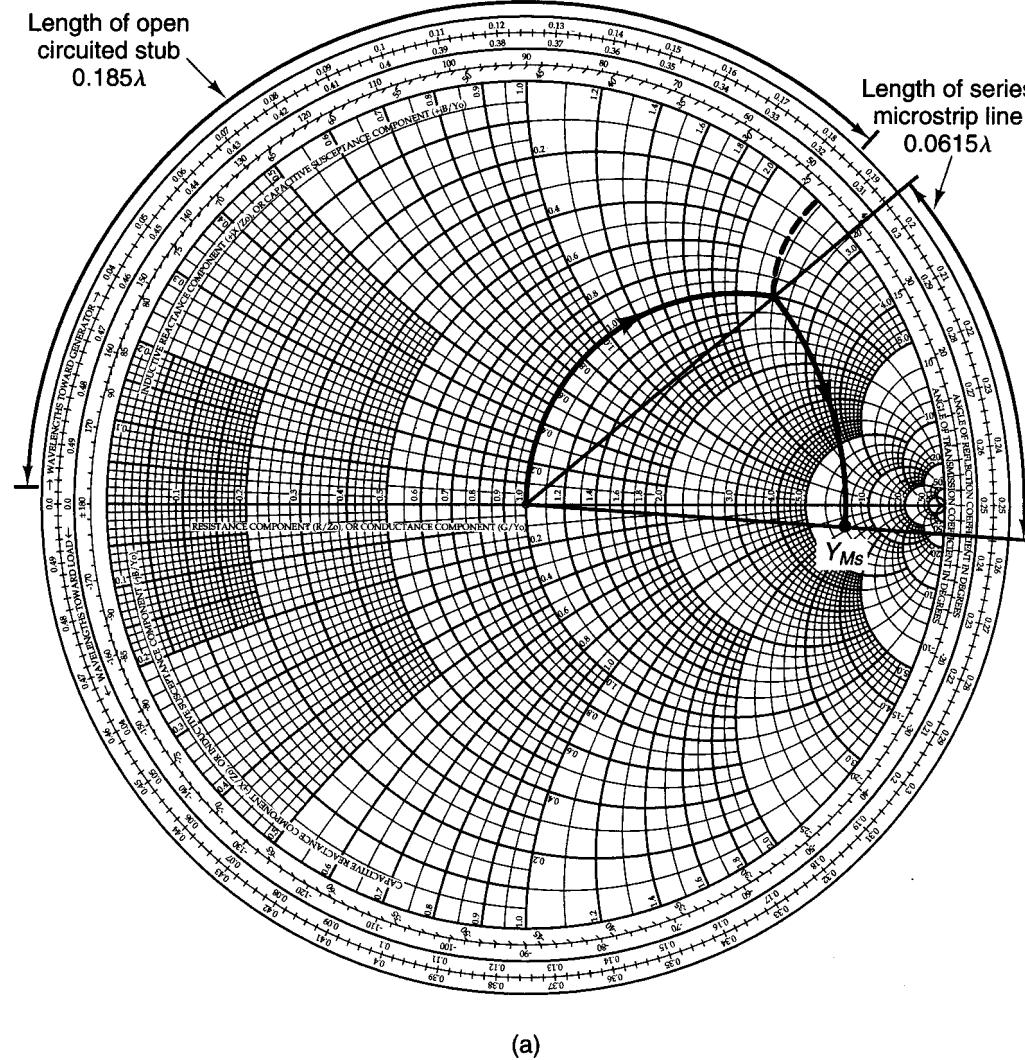


Figure 3.6.3 (a) Design of the input matching network; (b) design of the output matching network.

The input matching network can be designed with an open shunt stub of length 0.185λ and a series transmission line of length 0.0615λ . The output matching network is designed with an open shunt stub of length 0.176λ and a series transmission line of length 0.169λ .

The ac amplifier schematic is shown in Fig. 3.6.4. Using Duroid® ($\epsilon_r = 2.23$, $h = 0.7874$ mm) for the board material, we find that $W = 2.41$ mm for a characteristic impedance of 50Ω , $\epsilon_{eff} = 1.91$, and $\lambda = 0.7236\lambda_0$, where $\lambda_0 = 5$ cm at $f = 6$ GHz. The microstrip lengths at $f = 6$ GHz are

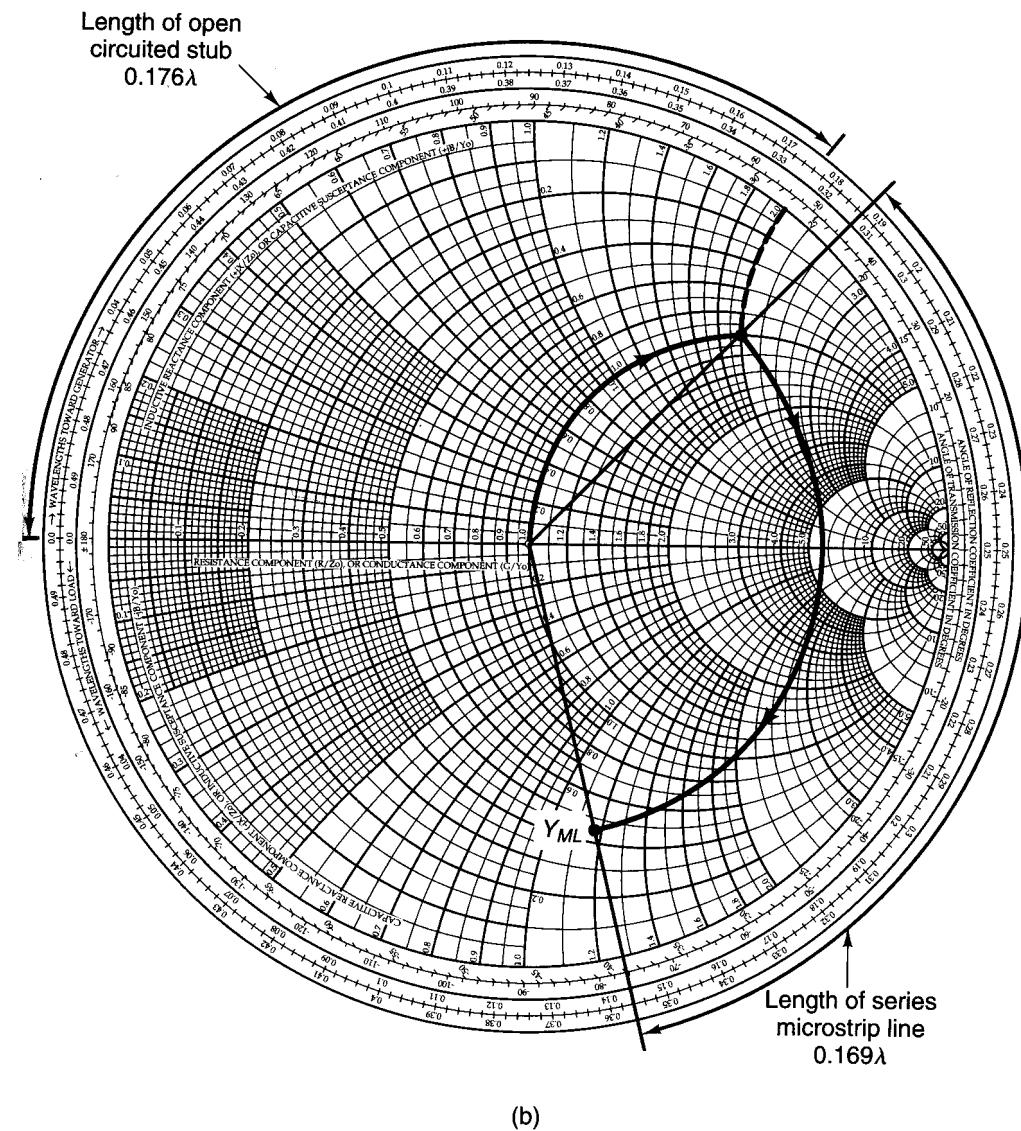


Figure 3.6.3 Continued

$$0.185\lambda = 6.70 \text{ mm}$$

$$0.0615\lambda = 2.23 \text{ mm}$$

$$0.169\lambda = 6.12 \text{ mm}$$

$$0.176\lambda = 6.37 \text{ mm}$$

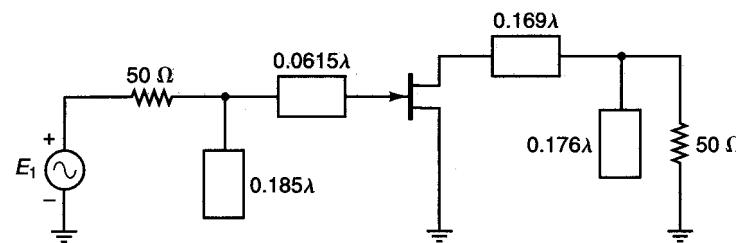


Figure 3.6.4 The ac schematic of a GaAs FET microwave amplifier. All microstrip lines have a characteristic impedance of 50 Ω.

The design for $G_{T,\max}$ with Γ_{Ms} and Γ_{ML} , at 6 GHz assures that the input and output VSWR are 1.

This example is revisited in Chapter 4 (Section 4.3), where noise considerations are included. Finally, we should point out that the stability must be checked at all frequencies, so that the reflection coefficients Γ_{Ms} and Γ_{ML} provide stable operation.

We have seen that for an unconditional stable device, the terminations Γ_{Ms} and Γ_{ML} , given by (3.6.5) and (3.6.6), will produce a simultaneous conjugate match which results in the maximum value of the transducer power gain. If the design calls for a transducer power gain different from the maximum, a constant-gain circle procedure based on (3.2.1) or (3.2.2) can be attempted. As we will see, such a procedure is not practical.

A constant-gain circle procedure based on (3.2.1) could be attempted as follows. Write (3.2.1) in the form

$$G_T = G'_s G_o G_L \quad (3.6.13)$$

where

$$G'_s = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_{IN} \Gamma_s|^2} \quad (3.6.14)$$

$$G_o = |S_{21}|^2$$

and

$$G_L = \frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2} \quad (3.6.15)$$

Then the design procedure is as follows:

- From (3.6.15), the constant G_L circles can be drawn using (3.4.9) to (3.4.12). Select the desired Γ_L for a given G_L gain. Observe that $\Gamma_s = S_{22}^*$ produces $G_{L,\max}$ but not $G_{T,\max}$.

- Calculate Γ_{IN} from (3.2.5). Observe that Γ_{IN} depends on Γ_L ; therefore, G'_s depends on G_L .
- From (3.6.14), the constant G'_s circles can be drawn using (3.4.9) to (3.4.12) (observing that Γ_{IN} replaces S_{22}). $\Gamma_s = \Gamma_{IN}^*$ produces $G'_{s,\max}$, and constant G'_s circles can be drawn for $G'_s < G'_{s,\max}$. Select the desired Γ_s for a given G'_s gain. Of course, the values of G'_s might not be satisfactory for the desired G_T . This will require the selection of another Γ_L and the procedure repeated.
- Design the matching networks.

The procedure just outlined is not recommended for a practical design since Γ_{IN} is a function of Γ_L , making the G'_s function dependent of the G_L function. Furthermore, the centers of the gain circles at $\Gamma_L = S_{22}^*$ and $\Gamma_s = \Gamma_{IN}^*$ do not give $G_{T,\max}$. In fact, the graphical approach becomes tedious because of the iterative process required for obtaining the desired gain.

As shown in the next section, the design of a microwave transistor amplifier in the unconditional stable bilateral case, for a gain different from $G_{T,\max}$, can be done using the operating power gain equation (the available power gain equation can also be used). When the transistor is unconditionally stable, a simultaneous conjugate match can be found, and the design procedure is based on $G_{T,\max}$ or (as shown in the next section) on the operating power gain. In fact, when designing for $G_{T,\max}$, which is equal to $G_{p,\max}$ and to $G_{A,\max}$, all design procedures result in $\Gamma_s = \Gamma_{Ms}$ and $\Gamma_L = \Gamma_{ML}$.

3.7 OPERATING AND AVAILABLE POWER-GAIN CIRCLES

Operating Power-Gain Circles

When S_{12} cannot be neglected, a design procedure based on the operating power gain G_p is commonly used. The operating power gain is independent of the source impedance; therefore, an operating power-gain circle procedure for both unconditionally stable and potentially unstable transistors is simple and recommended for practical designs.

Again we must consider two cases, the unconditionally stable case and the potentially unstable case.

Unconditionally stable bilateral case. To develop a design procedure with G_p , we write (3.2.3) in the form

$$G_p = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2)}{\left(1 - \frac{|S_{11} - \Delta \Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2}\right) |1 - S_{22} \Gamma_L|^2} = |S_{21}|^2 g_p \quad (3.7.1)$$

where

$$\begin{aligned} g_p &= \frac{G_p}{|S_{21}|^2} = \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2 - |S_{11} - \Delta\Gamma_L|^2} \\ &= \frac{1 - |\Gamma_L|^2}{1 - |S_{11}|^2 + |\Gamma_L|^2(|S_{22}|^2 - |\Delta|^2) - 2 \operatorname{Re}(\Gamma_L C_2)} \end{aligned} \quad (3.7.2)$$

and

$$C_2 = S_{22} - \Delta S_{11}^* \quad (3.7.3)$$

Here G_p and g_p are functions of the device S parameters and Γ_L .

In Appendix G the values of Γ_L in (3.7.2) that produce a constant value of g_p are shown to lie on a circle, known as an operating power-gain circle. The equation for an operating power-gain circle in the Γ_L plane, with g_p as a parameter, is

$$|\Gamma_L - C_p| = r_p$$

where the center of the circle C_p is located at

$$C_p = \frac{g_p C_2^*}{1 + g_p (|S_{22}|^2 - |\Delta|^2)} \quad (3.7.4)$$

and the radius of the circle is given by

$$r_p = \frac{[1 - 2K|S_{12}S_{21}|g_p + |S_{12}S_{21}|^2 g_p^2]^{1/2}}{|1 + g_p (|S_{22}|^2 - |\Delta|^2)|} \quad (3.7.5)$$

Equation (3.7.4) shows that the distance from the origin to the center of a power-gain circle is simply $|C_p|$, and the angle of inclination of the circle is C_2^* .

The maximum operating power gain occurs at the value of Γ_L when $r_p = 0$. Therefore, from (3.7.5) we can write

$$g_{p,\max}^2 |S_{12}S_{21}|^2 - 2K|S_{12}S_{21}|g_{p,\max} + 1 = 0 \quad (3.7.6)$$

where $g_{p,\max}$ is the maximum value of g_p . The solution to (3.7.6) for unconditional stability is

$$g_{p,\max} = \frac{1}{|S_{12}S_{21}|} (K - \sqrt{K^2 - 1}) \quad (3.7.7)$$

Therefore, substituting (3.7.7) into (3.7.1) gives

$$G_{p,\max} = \frac{|S_{21}|}{|S_{12}|} (K - \sqrt{K^2 - 1}) \quad (3.7.8)$$

As expected, (3.7.8) is identical to $G_{T,\max}$ in (3.6.10).

The value of Γ_L that produces $G_{p,\max}$ follows by substituting $g_p = g_{p,\max}$ in (3.7.4). This value of $\Gamma_L = C_{p,\max}$ must be equal to Γ_{ML} . That is, from (3.7.4)

$$\Gamma_{ML} = C_{p,\max} = \frac{g_{p,\max} C_2^*}{1 + g_{p,\max} (|S_{22}|^2 - |\Delta|^2)} \quad (3.7.9)$$

Substituting (3.7.7) into (3.7.9) and performing some manipulations, it follows that (3.7.9) is identical to (3.6.6), using the minus sign in (3.6.6). This derivation is presented in Appendix H.

The lowest value of g_p is zero, which corresponds to $G_p = 0$. From (3.7.1), $G_p = 0$ occurs when $|\Gamma_L| = 1$. In other words, the operating power gain is zero when all the output power is reflected from the load (i.e., when $|\Gamma_L| = 1$).

For a given G_p , Γ_L is selected from the constant operating power-gain circles. $G_{p,\max}$ results when Γ_L is selected at the distance where $g_{p,\max} = G_{p,\max}/|S_{21}|^2$. The maximum output power results when a conjugate match is selected at the input (i.e., $\Gamma_s = \Gamma_{IN}^*$). It also follows that when $\Gamma_s = \Gamma_{IN}^*$, the input power is equal to the maximum available input power. Therefore, under these circumstances the maximum transducer power gain ($G_{T,\max}$) and the operating power gain are equal, and the values of Γ_s and Γ_L that result in $G_{p,\max}$ are identical to Γ_{MS} and Γ_{ML} , respectively.

The procedure for drawing a constant operating power-gain circle in the Z Smith chart is as follows:

1. For a given G_p , the center and radius of the constant operating power-gain circle are given by (3.7.4) and (3.7.5).
2. Select the desired Γ_L .
3. For the given Γ_L , maximum output power is obtained with a conjugate match at the input—namely, with $\Gamma_s = \Gamma_{IN}^*$, where Γ_{IN} is given by (3.2.5). This value of Γ_s produces the transducer power gain $G_T = G_p$.

Example 3.7.1

Design the amplifier in Example 3.6.1 to have an operating power gain of 9 dB instead of $G_{T,\max} = G_{p,\max} = 11.38$ dB.

Solution. Since

$$|S_{21}|^2 = (2.058)^2 = 4.235 \quad \text{or} \quad 6.27 \text{ dB}$$

then

$$g_p = \frac{G_p}{|S_{21}|^2} = \frac{7.94}{4.235} = 1.875$$

From the results in Example 3.6.1, $K = 1.504$, $|\Delta| = 0.3014$, and $C_2 = 0.3911[-103.9^\circ]$. Therefore, the radius and center of the 9-dB operating power-gain circle, from (3.7.4) and (3.7.4), are $r_p = 0.431$ and $C_p = 0.508[103.9^\circ]$.

The graphical construction is shown in Fig. 3.7.1a. The 9-dB operating power-gain circle shows all loads that produce $G_p = 9$ dB. The load reflection coefficient can be selected at point A—namely, $\Gamma_L = 0.36[47.5^\circ]$. Then the required Γ_s for maximum output power is

$$\Gamma_s = \Gamma_{IN}^* = \left[S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right]^* = 0.629[175.51^\circ]$$

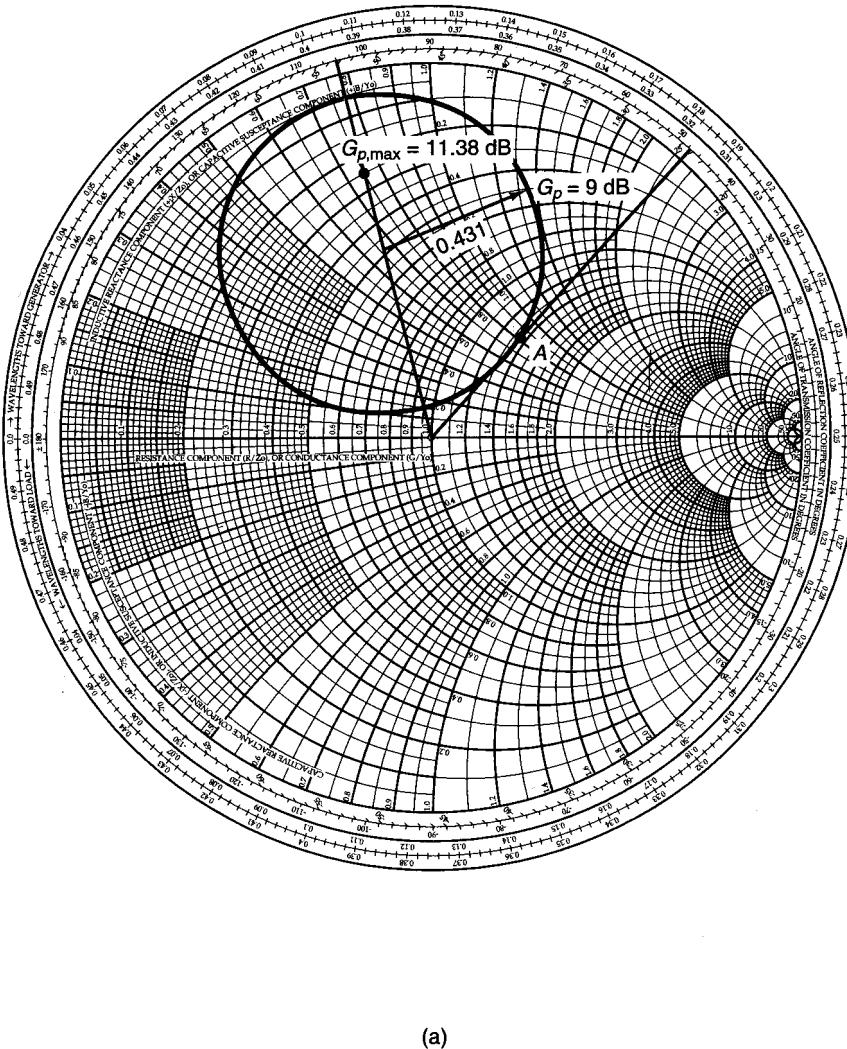


Figure 3.7.1 (a) Operating power-gain circle for $G_p = 9$ dB and location of $G_{p,\max} = 11.38$ dB; (b) the block diagram of the amplifier.

Since $\Gamma_s = \Gamma_{IN}^*$, it follows that $G_T = G_p = 9$ dB. The block diagram of the amplifier is shown in Fig. 3.7.1b.

The input VSWR is 1 since $\Gamma_s = \Gamma_{IN}^*$. The output VSWR is calculated using (2.8.4) and (2.8.6). From (3.2.6) the output reflection coefficient is $\Gamma_{OUT} = 0.67 \angle -102.66^\circ$. Using (2.8.6), we obtain

$$|\Gamma_b| = \left| \frac{\Gamma_{OUT} - \Gamma_L^*}{1 - \Gamma_{OUT}\Gamma_L} \right| = \left| \frac{0.67 \angle -102.66^\circ - 0.36 \angle -47.5^\circ}{1 - 0.67 \angle -102.66^\circ (0.36 \angle 47.5^\circ)} \right| = 0.622$$

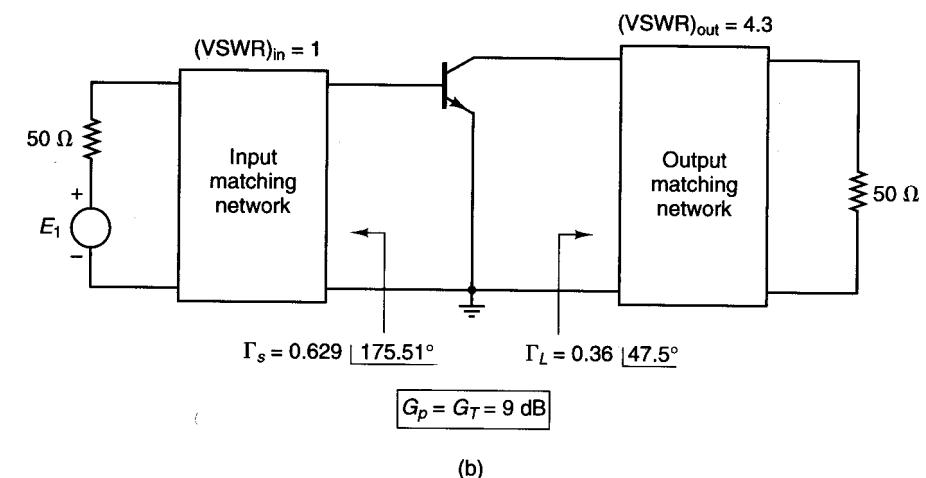


Figure 3.7.1 Continued

Therefore, the output VSWR is

$$(VSWR)_{out} = \frac{1 + 0.622}{1 - 0.622} = 4.3$$

The location of $G_{p,\max} = 11.38$ dB can be found as follows:

$$\begin{aligned} g_{p,\max} &= \frac{G_{p,\max}}{|S_{21}|^2} = \frac{13.74}{(2.058)^2} = 3.24 \\ r_{p,\max} &= 0 \end{aligned}$$

and the value of Γ_L that produces $G_{p,\max}$ is

$$\Gamma_{ML} = C_{p,\max} = \frac{g_{p,\max} C_2^*}{1 + g_{p,\max} (|S_{22}|^2 - |A|^2)} = 0.718 \angle 103.9^\circ$$

This value is identical to the value Γ_{ML} found in Example 3.6.1. The associated Γ_s for maximum output power, which is Γ_{Ms} , is obtained from

$$\Gamma_{Ms} = \left[S_{11} + \frac{S_{12}S_{21}\Gamma_{ML}}{1 - S_{22}\Gamma_{ML}} \right]^* = 0.762 \angle 177.3^\circ$$

As expected, this value is identical to the value of Γ_{Ms} in Example 3.6.1.

The previous example illustrates that for a given Γ_L on a constant G_p circle, there is an associated $\Gamma_s = \Gamma_{IN}^*$ that gives $(VSWR)_{in} = 1$. Then for $\Gamma_s = \Gamma_{IN}^*$, there is a Γ_{OUT} . The mismatch at the output between Γ_L and Γ_{OUT} determines $(VSWR)_{out}$ according to (2.8.4). We can try other values of Γ_L on the constant G_p circle and calculate for each Γ_L the associated $\Gamma_s = \Gamma_{IN}^*$. Then for each Γ_s the associated Γ_{OUT} is calculated and the resulting

$(VSWR)_{out}$. This analysis can be done for a series of points on the constant G_p circle to see if a better $(VSWR)_{out}$ can be obtained. In some cases, a certain mismatch at the input might be necessary in order to obtain the desired $(VSWR)_{out}$. Design considerations involving gain and VSWRs are discussed in Section 3.8.

Potentially unstable bilateral case. With a potentially unstable transistor, the design procedure for a given G_p is as follows:

1. For a given G_p , draw the constant operating power-gain circle using (3.7.4) and (3.7.5), and also draw the output stability circle as discussed in Section 3.3 [i.e., see (3.3.7) and (3.3.8)]. Select a value of Γ_L that is in the stable region and not too close to the stability circle.
2. Calculate Γ_{IN} using (3.2.5) and determine if a conjugate match at the input is feasible. That is, draw the input stability circle as discussed in Section 3.3 [i.e., see (3.3.9) and (3.3.10)] and determine if $\Gamma_s = \Gamma_{IN}^*$ lies in the input stable region.
3. If $\Gamma_s = \Gamma_{IN}^*$ is not in the stable region or is in the stable region but very close to the input stability circle, the value of Γ_s can be selected arbitrarily or a new value of G_p can be selected. Of course, we must be careful when selecting Γ_s arbitrarily since the value of Γ_s affects the output power and the VSWR.

The values of Γ_L and Γ_s should not be too close to their respective stability circles, because oscillations might occur due to component variations that can place Γ_s and Γ_L in the unstable regions.

Since G_p can be infinite in a potentially unstable case, it is practical to keep the value of G_p below the figure of merit value G_{MSG} [given in (3.6.11)]. The design for G_p lower than G_{MSG} can be performed with good stability and practical values of the input and output VSWR. On the other hand, a design for a G_p greater than G_{MSG} usually produces values of Γ_L and Γ_s close to the unstable regions and large values of the input and output VSWR. Designs procedures involving potentially unstable transistors and VSWR considerations are discussed in Section 3.8.

In a potentially unstable situation with $K > 1$ and $|\Delta| > 1$, there is a $G_{p,min}$, just like $G_{T,min}$ in (3.6.12). In fact, in a potentially unstable situation with $K > 1$ and $|\Delta| > 1$, the other solution of (3.7.6), denoted by $g_{p,min}$, is

$$g_{p,min} = \frac{1}{|S_{12}S_{21}|} (K + \sqrt{K^2 - 1}) = \frac{G_{p,max}}{|S_{21}|^2}$$

Hence,

$$G_{p,min} = \frac{|S_{21}|}{|S_{12}|} (K + \sqrt{K^2 - 1}) \quad (3.7.10)$$

Equation (3.7.10) gives the minimum value that G_p can have inside the stable region in a potentially unstable case with $K > 1$ and $|\Delta| > 1$. The maximum value that G_p can have is infinite.

The value of Γ_L that produces $G_{p,min}$, denoted by $\Gamma_{L,min}$, is

$$\Gamma_{L,min} = \frac{g_{p,min} C_2^*}{1 + g_{p,min} (|S_{22}|^2 - |\Delta|^2)} \quad (3.7.11)$$

It also follows that $\Gamma_{L,min}$ in (3.7.11) is identical to Γ_{ML} in (3.6.6) when the plus sign is used in (3.6.6).

Example 3.7.2

The S parameters of a GaAs FET at $I_D = 50\% I_{DSS}$, $I_{DSS} = 10 \text{ mA}$, $V_{DS} = 5 \text{ V}$, and $f = 8 \text{ GHz}$ are

$$S_{11} = 0.5 \angle -180^\circ$$

$$S_{12} = 0.08 \angle 30^\circ$$

$$S_{21} = 2.5 \angle 70^\circ$$

$$S_{22} = 0.8 \angle -100^\circ$$

The transistor is potentially unstable at 8 GHz with $G_{MSG} = 14.9 \text{ dB}$. Design an amplifier with $G_p = 10 \text{ dB}$.

Solution. First we will verify that the transistor is potentially unstable at 8 GHz. From (3.3.16) and (3.3.17), we obtain $K = 0.4$ and $\Delta = 0.223 \angle 62.12^\circ$. Since $K < 1$, the GaAs FET is potentially unstable. Also, from (3.6.11) the value of G_{MSG} is $G_{MSG} = 2.5/0.08 = 31.25$ or 14.9 dB .

In order to design for $G_p = 10 \text{ dB}$ (4.9 dB less than the G_{MSG}), the 10-dB operating power-gain circle and the output stability circle must be calculated. The radius and center of the 10-dB power gain circle, from (3.7.4) and (3.7.5), are $r_p = 0.473$ and $C_p = 0.572 \angle 97.2^\circ$. The radius and center of the output stability circle, from (3.3.7) and (3.3.8), are $r_L = 0.34$ and $C_L = 1.18 \angle 97.2^\circ$.

The Smith chart in Fig. 3.7.2 shows the construction of the 10-dB operating power-gain circle and the output stability circle. Since $|S_{11}| < 1$, the stable region is the region outside the output stability circle. Γ_L is selected on the 10-dB power-gain circle at location A—namely, $\Gamma_L = 0.1 \angle 97^\circ$ or $Z_L = 50 (0.96 + j0.19) \Omega$.

For a conjugate match at the input, Γ_s is given by $\Gamma_s = \Gamma_{IN}^* = 0.52 \angle 179.32^\circ$ and we must determine if the value of Γ_s is in the stable region. The radius and center of the input stability circle, from (3.3.9) and (3.3.10), are $r_s = 1.0$ and $C_s = 1.67 \angle 171^\circ$, where the stable region is the region outside the input stability circle. Therefore, Γ_s is a stable source reflection coefficient. Since $\Gamma_s = \Gamma_{IN}^*$, it follows that $G_T = G_p = 10 \text{ dB}$ and $(VSWR)_{in} = 1$.

From (3.2.6), the output reflection coefficient is $\Gamma_{out} = 0.934 \angle -97.18^\circ$. Using (2.8.6), we obtain $|\Gamma_b| = 0.918$ and the output VSWR is

$$(VSWR)_{out} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} = \frac{1 + 0.918}{1 - 0.918} = 23.5$$

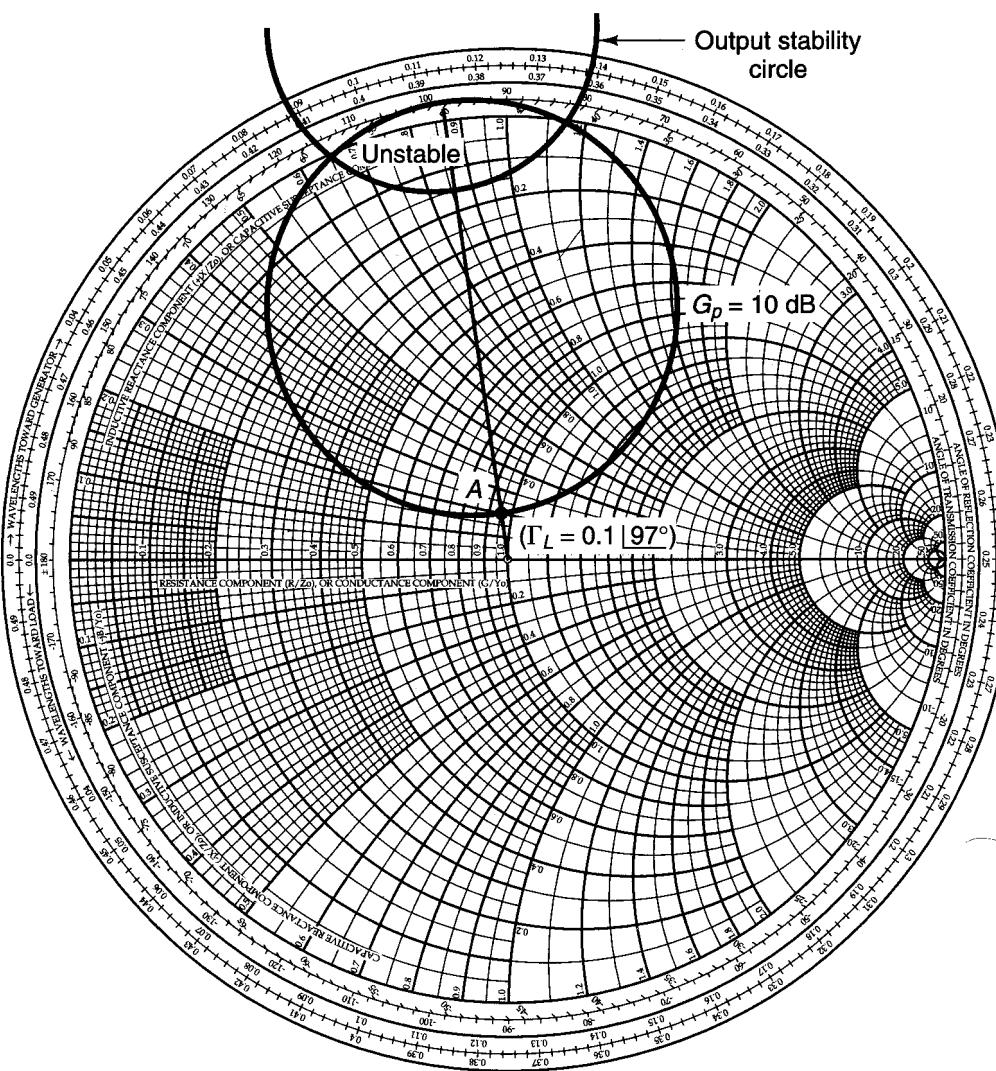


Figure 3.7.2 Power and stability circles construction for Example 3.7.2.

The previous calculations show that the output port of the amplifier is significantly mismatched in order to obtain $G_p = 10 \text{ dB}$. The output VSWR can be lowered by accepting a higher-input VSWR or by changing the amplifier gain requirements. For example, with $\Gamma_s = 0.22 | 177^\circ$ and $\Gamma_L = 0.1 | 97^\circ$, it follows that $(\text{VSWR})_{\text{in}} = 2$ and $(\text{VSWR})_{\text{out}} = 9.9$. A formal procedure involving gain and VSWR considerations is discussed in Section 3.8.

Using the values of the S parameters in Example 3.7.2, the operating power-gain circles for $G_p = 0 \text{ dB}$, 10 dB , 15 dB , 20 dB , and 30 dB were calculated and plotted in Fig. 3.7.3. Observe that the $G_p = 0$ (i.e., $-\infty \text{ dB}$) circle oc-

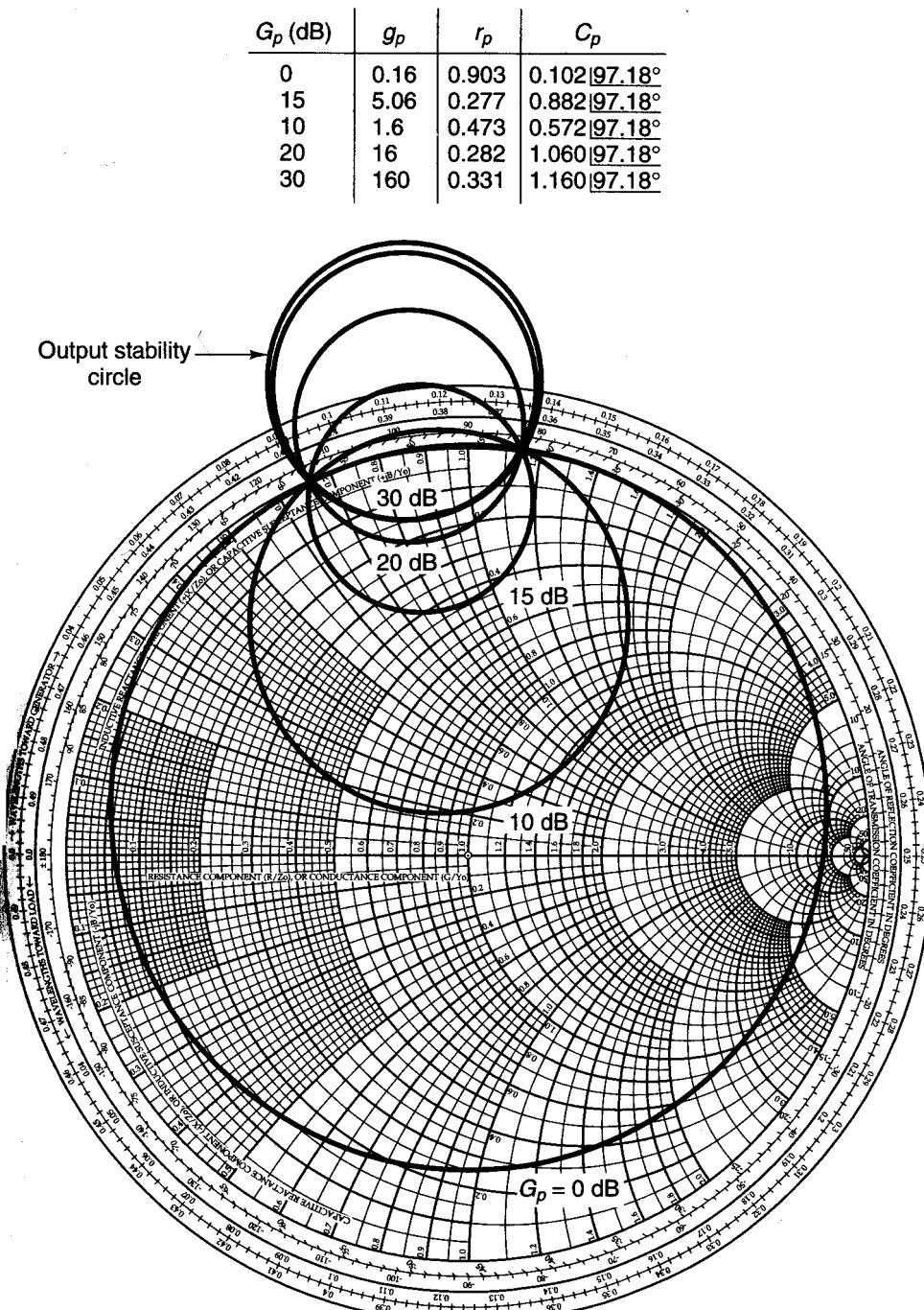


Figure 3.7.3 Typical behavior of the operating power-gain circles. The S parameters in Example 3.7.2 were used to calculate the constant gain circles.

curs when $|\Gamma_L| = 1$, and the operating power-gain circles approach the output stability circles as G_p becomes infinite.

Since G_p becomes infinite when Γ_L is on the output stability circle, it follows that the power-gain circle for $G_p = \infty$ is identical to the output stability circle. In fact, for $G_p \rightarrow \infty$ we have $g_p \rightarrow \infty$ and, taking the limit as g_p approaches infinity in (3.7.4), we obtain

$$C_p = \frac{C_2^*}{|S_{22}|^2 - |\Delta|^2}$$

and

$$r_p \approx \frac{|S_{12}S_{21}|^2 g_p^{1/2}}{|g_p|(|S_{22}|^2 - |\Delta|^2)} = \frac{|S_{12}S_{21}|}{|S_{22}|^2 - |\Delta|^2}$$

which are recognized as the equations for C_L and r_L in (3.3.8) and (3.3.7), respectively.

From the plots in Fig. 3.7.2, it is seen that the power-gain circles and the output stability circles intersect the Smith chart at the same points. This is not a coincidence, and it can be shown (see Problem 3.28) that for a potentially unstable device these circles intersect the Smith chart at the same points.

Example 3.7.3

(This example illustrates the power-gain circles for a two-port network with $K > 1$ and $|\Delta| > 1$.)

The S parameters of a two-port network at $f = 4$ GHz are

$$S_{11} = 0.7 \underline{-50^\circ}$$

$$S_{12} = 0.27 \underline{75^\circ}$$

$$S_{21} = 5 \underline{120^\circ}$$

$$S_{22} = 0.6 \underline{80^\circ}$$

Plot several operating power-gain circles ranging from $G_p = G_{p,\min}$ to $G_p = \infty$.

Solution. For this two-port network, it follows that $K = 1.202$ and $\Delta = 1.76 \underline{18.54^\circ}$ (i.e., $|\Delta| = 1.76$ is greater than 1). Since $K > 1$ and $|\Delta| > 1$ (or $B_2 < 0$), we obtain from (3.6.6), using the plus sign in (3.6.6),

$$B_1 = -1.964$$

$$B_2 = -2.224$$

$$C_1 = 0.395 \underline{97.92^\circ}$$

$$C_2 = 0.654 \underline{-121.95^\circ}$$

$$\Gamma_{L,\min} = \Gamma_{ML} = \frac{-2.224 + \sqrt{(2.224)^2 - 4(0.654)^2}}{2(0.654 \underline{-121.95^\circ})} = 0.325 \underline{-58.05^\circ}$$

Then, using (3.7.10),

$$G_{p,\min} = G_{T,\min} = \frac{5}{0.27} (1.202 + \sqrt{(1.202)^2 - 1}) = 34.61 \text{ (or } 15.39 \text{ dB)}$$

The output stability circle ($C_L = 0.239 \underline{-58.05^\circ}$ and $r_L = 0.494$) and power-gain circles for $G_{p,\min} = 15.39$ dB and $G_p = 18$ dB, 20 dB, and 30 dB were calculated and are plotted in Fig. 3.7.4a. Observe that the operating power-gain circles approach the output stability circle as G_p becomes infinite.

It is simple to show that in the unstable region G_p in decibels becomes negative, approaching $-\infty$ dB at $|\Gamma| = 1$.

For maximum power transfer to the load, the source reflection coefficient is

$$\Gamma_s = \Gamma_{IN}^* = \left(0.7 \underline{-50^\circ} + \frac{0.27 \underline{75^\circ} (5 \underline{120^\circ}) 0.325 \underline{-58.05^\circ}}{1 - 0.6 \underline{80^\circ} (0.325 \underline{-58.05^\circ})} \right)^* = 0.21 \underline{82.08^\circ}$$

This value of Γ_s is identical to the value of Γ_{Ms} in (3.6.5), using the plus sign in (3.6.5). That is,

$$\Gamma_s = \Gamma_{Ms} = 0.21 \underline{82.08^\circ}$$

The input stability circle ($C_s = 0.152 \underline{82.08^\circ}$ and $r_s = 0.518$) and Γ_{Ms} are shown in Fig. 3.7.4b.

Since a simultaneous conjugate match condition exists it follows that the input and output VSWR are equal to one.

Available Power-Gain Circles

Unconditionally stable bilateral case. The derivation of the constant available power-gain circles is similar to that of the operating power-gain circles. From (3.2.4), we write G_A in the form

$$G_A = \frac{|S_{21}|^2 (1 - |\Gamma_s|^2)}{\left(1 - \frac{|S_{22} - \Delta \Gamma_s|^2}{|1 - S_{11} \Gamma_s|} \right) |1 - S_{11} \Gamma_s|^2} = |S_{21}|^2 g_a \quad (3.7.12)$$

where

$$g_a = \frac{G_A}{|S_{21}|^2} = \frac{1 - |\Gamma_s|^2}{1 - |S_{22}|^2 + |\Gamma_s|^2 (|S_{11}|^2 - |\Delta|^2) - 2 \operatorname{Re}(\Gamma_s C_1)} \quad (3.7.13)$$

and

$$C_1 = S_{11} - \Delta S_{22}^* \quad (3.7.14)$$

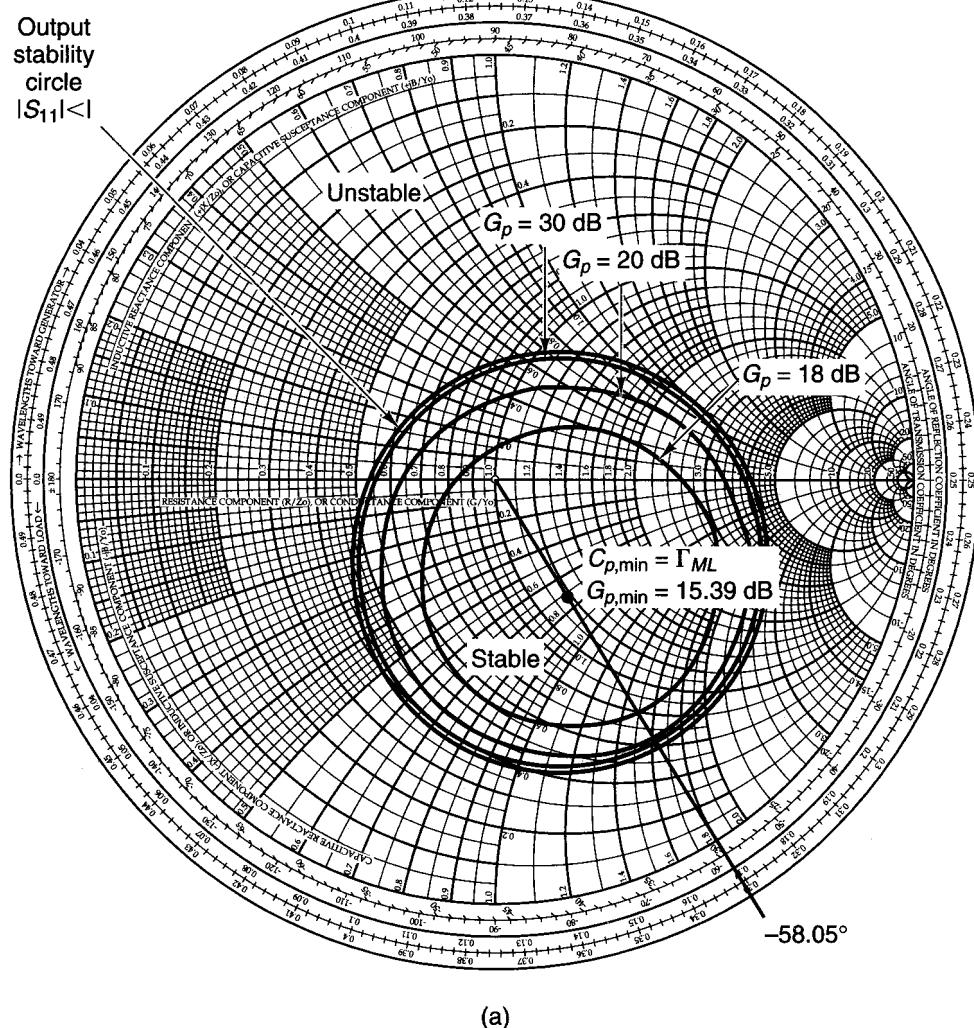
Since (3.7.12), (3.7.13), and (3.7.14) are identical in form to (3.7.1), (3.7.2), and (3.7.3), it follows that the center C_a and radius r_a of constant operating power-gain circles are given by

$$C_a = \frac{g_a C_1^*}{1 + g_a (|S_{11}|^2 - |\Delta|^2)} \quad (3.7.15)$$

and

$$r_a = \frac{[1 - 2K|S_{12}S_{21}|g_a + |S_{12}S_{21}|^2 g_a^2]^{1/2}}{|1 + g_a (|S_{11}|^2 - |\Delta|^2)|} \quad (3.7.16)$$

G_p (dB)	g_p	r_p	C_p
18	2.524	0.356	0.280 -58.05°
20	4	0.417	0.263 -58.05°
30	40	0.487	0.242 -58.05°



(a)

Figure 3.7.4 (a) Output stability circle and operating power-gain circles for Example 3.7.3; (b) input stability circle and Γ_{Ms} .

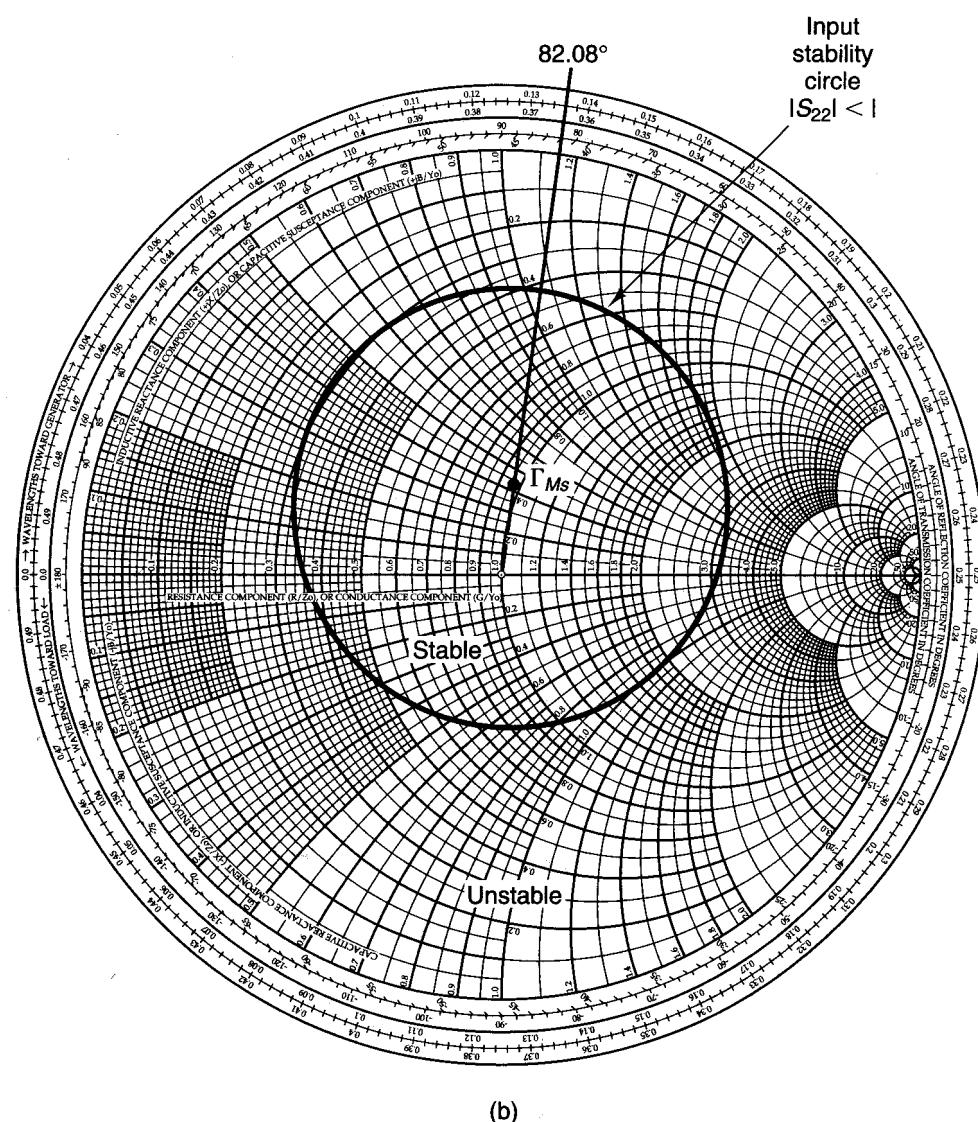


Figure 3.7.4 Continued

For a given G_A , a constant available power-gain circle can be plotted using (3.7.15) and (3.7.16). All Γ_s on this circle produce the given G_A . For the given G_A , maximum output power is obtained with $\Gamma_L = \Gamma_{OUT}^*$, where Γ_{OUT} is given by (3.2.6). This value of Γ_L produces the transducer power gain $G_T = G_A$.

Since the constant available power-gain circles and the constant noise figure circles are functions of Γ_s , they can be plotted together on the Smith chart

and the trade-offs that result between gain and noise figure can be analyzed. These concepts are studied in Chapter 4.

Potentially unstable bilateral case. With a potentially unstable transistor, the design procedure for a given G_A is as follows:

1. For a given G_A , draw the constant available gain circle using (3.7.15) and (3.7.16), and also draw the input stability circle using (3.3.9) and (3.3.10). Select a value of Γ_s that is in the stable region and not too close to the stability circle.
2. Calculate Γ_{OUT} using (3.2.6) and determine if a conjugate match at the output is feasible. That is, draw the output stability circle and determine if $\Gamma_L = \Gamma_{\text{OUT}}^*$ lies in the stable region.
3. If $\Gamma_L = \Gamma_{\text{OUT}}^*$ is not in the stable region or is in the stable region but very close to the output stability circle, the value of Γ_L can be selected arbitrarily or a new value of G_A can be selected. Of course, we must be careful when selecting Γ_L arbitrarily since the value of Γ_L affects the output power and the VSWR.

The values of Γ_L and Γ_s should not be too close to their respective stability circles, because oscillations might occur due to component variations that can place Γ_L and Γ_s in their respective unstable regions.

3.8 CONSTANT VSWR CIRCLES

The design specifications of a microwave amplifier usually include the maximum allowable values of its input VSWR and of its output VSWR. Constant input VSWR circles can be drawn on the Γ_s plane, and constant output VSWR circles can be drawn on the Γ_L plane. These circles can be used in the development of certain designs.

Consider Fig. 3.8.1, which shows a microwave amplifier. The input VSWR is given by

$$(\text{VSWR})_{\text{in}} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} \quad (3.8.1)$$

where $|\Gamma_a|$ is given by (2.8.3)—namely,

$$|\Gamma_a| = \left| \frac{\Gamma_{\text{IN}} - \Gamma_s^*}{1 - \Gamma_{\text{IN}}\Gamma_s} \right| \quad (3.8.2)$$

Equation (3.8.2) shows that for a given Γ_{IN} , the reflection coefficients Γ_s and Γ_a are related by a bilinear transformation. Hence, constant values of $|\Gamma_a|$ are obtained by values of Γ_s that lie on a circle.

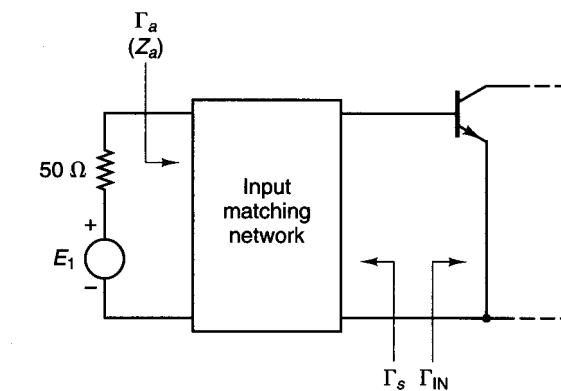


Figure 3.8.1 Input portion of a microwave amplifier.

The design for a given $(\text{VSWR})_{\text{in}}$ can be performed as follows. From (3.8.1), the value of $(\text{VSWR})_{\text{in}}$ fixes the value of $|\Gamma_a|$. Then (3.8.2) is used to plot the values of Γ_s that result in the constant value $|\Gamma_a|$. In Appendix I it is shown that from (3.8.2) a constant $|\Gamma_a|$ circle satisfies the equation

$$|\Gamma_s - C_{V_i}| = r_{V_i}$$

where the center C_{V_i} is given by

$$C_{V_i} = \frac{\Gamma_{\text{IN}}^*(1 - |\Gamma_a|^2)}{1 - |\Gamma_a\Gamma_{\text{IN}}|^2} \quad (3.8.3)$$

and the radius is

$$r_{V_i} = \frac{|\Gamma_a|(1 - |\Gamma_{\text{IN}}|^2)}{1 - |\Gamma_a\Gamma_{\text{IN}}|^2} \quad (3.8.4)$$

In an unconditional stable case, and in many potentially unstable cases, Γ_s can be selected equal to Γ_{IN}^* in order to get $(\text{VSWR})_{\text{in}} = 1$. Equations (3.8.3) and (3.8.4) demonstrate this fact. That is, from (3.8.1) with $(\text{VSWR})_{\text{in}} = 1$, it follows that $|\Gamma_a| = 0$, and from (3.8.3) and (3.8.4) we obtain

$$C_{V_i}|_{|\Gamma_a|=0} = \Gamma_{\text{IN}}^*$$

and

$$r_{V_i}|_{|\Gamma_a|=0} = 0$$

In other words, the value of $\Gamma_s = \Gamma_{\text{IN}}^*$ produces $|\Gamma_a| = 0$ and consequently $(\text{VSWR})_{\text{in}} = 1$.

Similar relations are obtained for the output VSWR. The output VSWR is given by

$$(\text{VSWR})_{\text{out}} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} \quad (3.8.5)$$

where $|\Gamma_b|$ is given by (2.8.6)—namely,

$$|\Gamma_b| = \left| \frac{\Gamma_{\text{OUT}} - \Gamma_L^*}{1 - \Gamma_{\text{OUT}}\Gamma_L} \right| \quad (3.8.6)$$

The constant $|\Gamma_b|$ circles satisfy the equation

$$|\Gamma_L - C_{V_o}| = r_{V_o}$$

where

$$C_{V_o} = \frac{\Gamma_{\text{OUT}}^*(1 - |\Gamma_b|^2)}{1 - |\Gamma_b\Gamma_{\text{OUT}}|^2} \quad (3.8.7)$$

and

$$r_{V_o} = \frac{|\Gamma_b|(1 - |\Gamma_{\text{OUT}}|^2)}{1 - |\Gamma_b\Gamma_{\text{OUT}}|^2} \quad (3.8.8)$$

Example 3.8.1

(a) The S parameters of a GaAs FET at 12 GHz, $V_{DS} = 3.5$ V, and $I_{DS} = 25$ mA are

$$S_{11} = 0.6 \angle 36^\circ$$

$$S_{12} = 0.14 \angle -85^\circ$$

$$S_{21} = 2.3 \angle -80^\circ$$

$$S_{22} = 0.15 \angle 45^\circ$$

Determine $G_{A,\max}$ and draw the constant G_A circle that is 1 dB less than $G_{A,\max}$.

(b) Select several values of Γ_s on the $G_A = G_{A,\max} - 1$ dB circle. For each Γ_s value, determine the values of Γ_L that lie on the constant $(VSWR)_{\text{out}} = 1.5$ circle, and draw the constant $(VSWR)_{\text{out}} = 1.5$ circles.

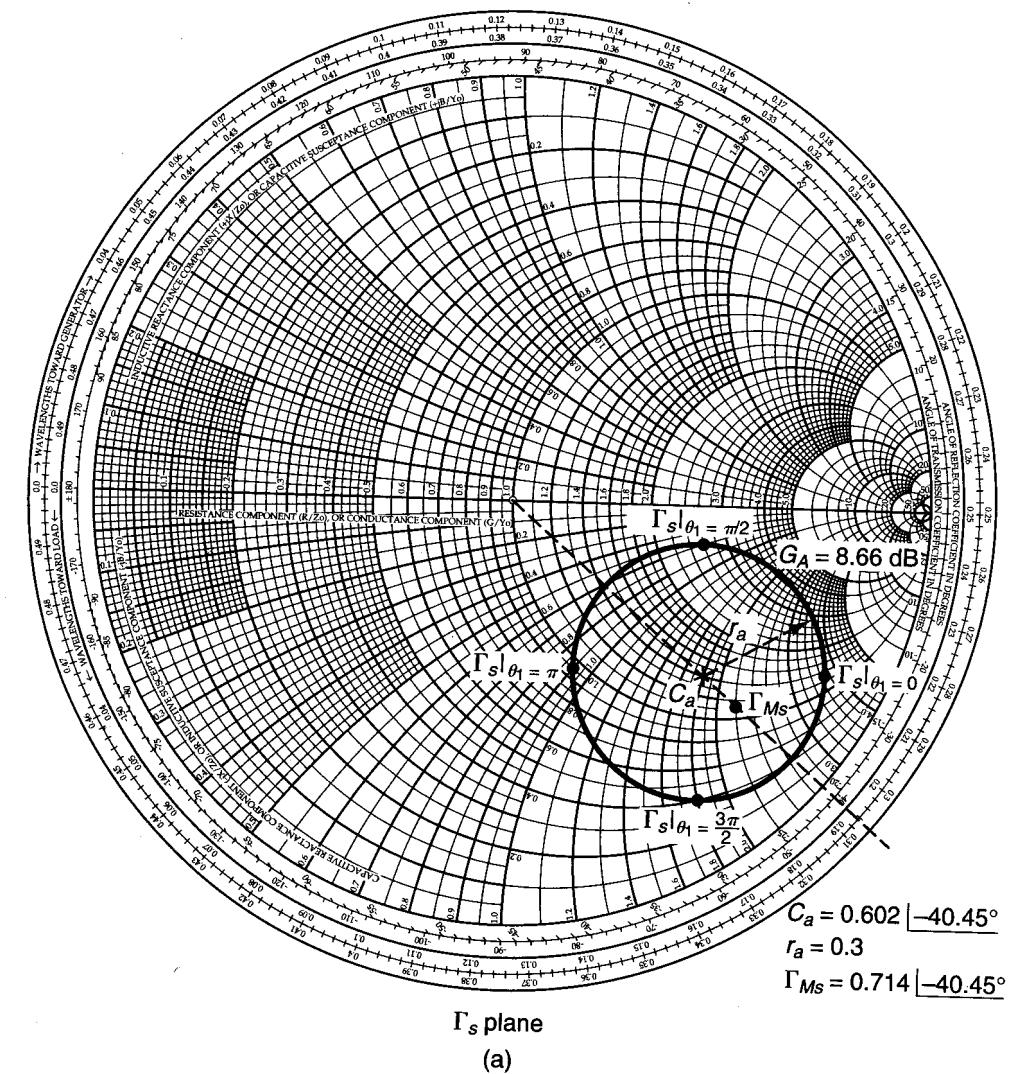
(c) Select several values of Γ_L on the $(VSWR)_{\text{out}} = 1.5$ circle. For each Γ_L value, calculate $(VSWR)_{\text{in}}$.

Solution. (a) This transistor is unconditionally stable since $K = 1.17$ and $\Delta = 0.368 \angle 27.91^\circ$. This transistor can be simultaneous conjugate matched with $\Gamma_{Ms} = 0.714 \angle -40.45^\circ$ and $\Gamma_{ML} = 0.387 \angle -129.36^\circ$, resulting in $(VSWR)_{\text{in}} = (VSWR)_{\text{out}} = 1$. From (3.6.10),

$$G_{A,\max} = \frac{|S_{21}|}{|S_{12}|} (K - \sqrt{K^2 - 1}) = \frac{2.3}{0.14} (1.17 - \sqrt{1.17^2 - 1}) = 9.24 \text{ (or } 9.66 \text{ dB)}$$

The analysis calls for a gain G_A different from $G_{A,\max}$ and the selection of Γ_s and Γ_L for certain VSWR performance. The design methods used in this example have applications in the design of low-noise amplifiers (discussed in Chapter 4). In the design of low-noise amplifiers, we commonly make trade-offs between gain, noise performance, and VSWRs.

The constant G_A circle that is 1 dB less than $G_{A,\max}$ is the circle for $G_A = 8.66$ dB. The center and radius of the $G_A = 8.66$ dB circle, from (3.7.15) and (3.7.16) with $g_a = 7.3451/(2.3)^2 = 1.3886$, are $C_a = 0.602 \angle -40.45^\circ$ and $r_a = 0.3$. The $G_A = 8.66$ dB constant gain circle is drawn in Fig. 3.8.2a.



Γ_s plane

(a)

Figure 3.8.2 (a) The $G_A = 8.66$ dB constant gain circle and Γ_{Ms} ; (b) four values of Γ_s on the $G_A = 8.66$ dB circle and calculations of $(VSWR)_{\text{in}}$ for $(VSWR)_{\text{out}} = 1$; (c) the four $(VSWR)_{\text{out}} = 1.5$ circles; (d) four values of Γ_L on the $(VSWR)_{\text{out}} = 1.5$ circles, the corresponding values of Γ_{IN} and $|\Gamma_a|$, and the resulting values of $(VSWR)_{\text{in}}$.

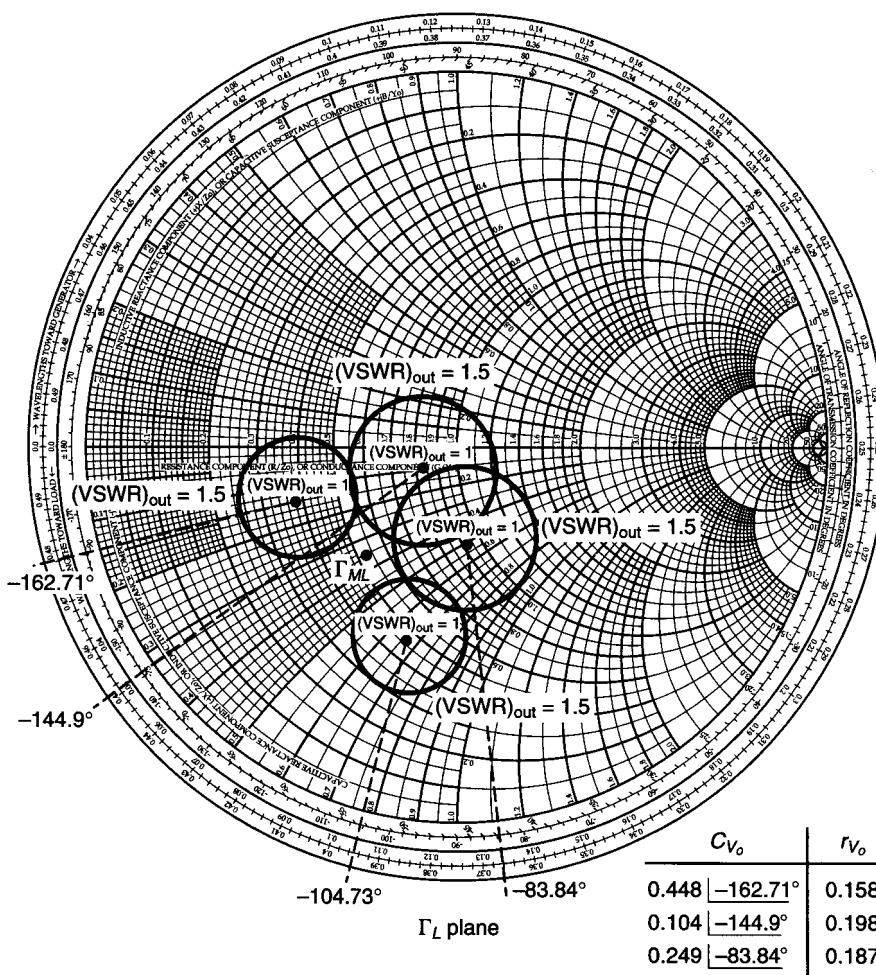
(b) The values of Γ_s on the $G_A = 8.66$ dB circle are given by

$$\Gamma_s = C_a + r_a e^{j\theta_1} = 0.602 \angle -40.45^\circ + 0.3 e^{j\theta_1}$$

Four convenient values of Γ_s on this circle are shown in Fig. 3.8.2a. They correspond to $\theta_1 = 0, \pi/2, \pi$, and $3\pi/2$. The values of Γ_s and the associated values of Γ_{OUT} are listed in

Γ_s	Γ_{OUT}	$\Gamma_L = \Gamma_{\text{OUT}}^*$	(VSWR) _{out}	Γ_{IN}	$ \Gamma_a $	(VSWR) _{in}
0.853 $[-27.25^\circ]$ for $\theta_1 = 0$	0.463 $[162.71^\circ]$	0.463 $[-162.71^\circ]$	1	0.743 $[34.62^\circ]$	0.395	2.31
0.467 $[-11.18^\circ]$ for $\theta_1 = \pi/2$	0.108 $[144.9^\circ]$	0.108 $[-144.9^\circ]$	1	0.634 $[36.71^\circ]$	0.394	2.30
0.421 $[-67.96^\circ]$ for $\theta_1 = \pi$	0.259 $[83.84^\circ]$	0.259 $[-83.84^\circ]$	1	0.630 $[43.53^\circ]$	0.394	2.30
0.829 $[-56.44^\circ]$ for $\theta_1 = 3\pi/2$	0.533 $[104.73^\circ]$	0.533 $[-104.73^\circ]$	1	0.727 $[46.85^\circ]$	0.395	2.31

(b)



(c)

Figure 3.8.2 Continued

Γ_s	Γ_{OUT}	Γ_L	Γ_{IN}	$ \Gamma_a $	(VSWR) _{in}
0.853 $[-27.25^\circ]$ for $\theta_1 = 0$	0.463 $[162.71^\circ]$	0.301 $[-153.73^\circ]$ for $\theta_2 = 0$	0.695 $[36.39^\circ]$	0.470	2.77
		0.428 $[176.66^\circ]$ for $\theta_2 = \pi/2$	0.719 $[31.29^\circ]$	0.371	2.18
		0.601 $[-167.19^\circ]$ for $\theta_2 = \pi$	0.781 $[33.11^\circ]$	0.320	1.94
		0.517 $[-145.76^\circ]$ for $\theta_2 = 3\pi/2$	0.762 $[37.91^\circ]$	0.460	2.70
0.467 $[-11.18^\circ]$ for $\theta_1 = \pi/2$	0.108 $[144.9^\circ]$	0.128 $[-27.91^\circ]$ for $\theta_2 = 0$	0.573 $[39.14^\circ]$	0.351	2.08
		0.162 $[121.62^\circ]$ for $\theta_2 = \pi/2$	0.612 $[31.30^\circ]$	0.320	1.94
		0.289 $[-168.07^\circ]$ for $\theta_2 = \pi$	0.689 $[34.54^\circ]$	0.447	2.62
		0.272 $[-108.27^\circ]$ for $\theta_2 = 3\pi/2$	0.662 $[41.79^\circ]$	0.470	2.77
0.421 $[-67.96^\circ]$ for $\theta_1 = \pi$	0.259 $[83.84^\circ]$	0.327 $[-49.23^\circ]$ for $\theta_2 = 0$	0.573 $[46.5^\circ]$	0.304	1.87
		0.066 $[-66.19^\circ]$ for $\theta_2 = \pi/2$	0.599 $[38.05^\circ]$	0.398	2.32
		0.295 $[-122.88^\circ]$ for $\theta_2 = \pi$	0.682 $[40.46^\circ]$	0.482	2.86
		0.435 $[-86.48^\circ]$ for $\theta_2 = 3\pi/2$	0.665 $[48^\circ]$	0.411	2.40
0.829 $[-56.44^\circ]$ for $\theta_1 = 3\pi/2$	0.533 $[104.73^\circ]$	0.501 $[-88.59^\circ]$ for $\theta_2 = 0$	0.684 $[49.28^\circ]$	0.390	2.28
		0.381 $[-110.25^\circ]$ for $\theta_2 = \pi/2$	0.693 $[43.43^\circ]$	0.477	2.82
		0.572 $[-118.83^\circ]$ for $\theta_2 = \pi$	0.761 $[44.22^\circ]$	0.449	2.63
		0.658 $[-101.54^\circ]$ for $\theta_2 = 3\pi/2$	0.757 $[49.53^\circ]$	0.311	1.90

(d)

Figure 3.8.2 Continued

Fig. 3.8.2b. For $\Gamma_L = \Gamma_{\text{OUT}}^*$, Fig. 3.8.2b shows the values of $(\text{VSWR})_{\text{out}}$, the corresponding values of Γ_{IN} , the values of $|\Gamma_a|$ [using (3.8.2)], and the resulting values of $(\text{VSWR})_{\text{in}}$ [using (3.8.1)]. From Fig. 3.8.2b, it is seen that with $\Gamma_L = \Gamma_{\text{OUT}}^*$, the input VSWR is approximately 2.3 for the four values of Γ_s .

For an output VSWR of 1.5, we have from (3.8.5) that $|\Gamma_b| = 0.2$. Then, using (3.8.7) and (3.8.8), the center and radius of the $(\text{VSWR})_{\text{out}} = 1.5$ circles are calculated and listed in Fig. 3.8.2c. The four $(\text{VSWR})_{\text{out}} = 1.5$ circles, as well as the four values where $\Gamma_L = \Gamma_{\text{OUT}}^*$ [i.e., $(\text{VSWR})_{\text{out}} = 1$], are drawn in Fig. 3.8.2c.

(c) The values of Γ_L on the $(\text{VSWR})_{\text{out}} = 1.5$ circles are given by

$$\Gamma_L = C_{V_o} + r_{V_o} e^{j\theta_2}$$

Four convenient values of Γ_L on the $(\text{VSWR})_{\text{out}} = 1.5$ circle correspond to $\theta_2 = 0, \pi/2, \pi$, and $3\pi/2$. In Fig. 3.8.2d, four values of Γ_L are calculated on each $(\text{VSWR})_{\text{out}} = 1.5$ circle, as well as the corresponding values of Γ_{IN} and $|\Gamma_a|$ and the resulting values of $(\text{VSWR})_{\text{in}}$. From the tabulated values, it is seen that there are several values of Γ_s and Γ_L that result in $(\text{VSWR})_{\text{in}}$ values around 1.9 with $(\text{VSWR})_{\text{out}} = 1.5$. For example, with $\Gamma_s = 0.421[-67.96^\circ]$ and $\Gamma_L = 0.327[-49.23^\circ]$, we have $(\text{VSWR})_{\text{out}} = 1.5$ and $(\text{VSWR})_{\text{in}} = 1.87$.

For further practice with this type of calculations, the reader is referred to Problem 3.30, where the analyses in this example are to be performed for the $G_A = G_{A,\text{max}} - 2$ dB circle (i.e., the $G_A = 7.66$ dB circle).

Since Γ_L and Γ_{IN} , as well as Γ_s and Γ_{OUT} , are related by bilinear transformations [see (3.2.5) and (3.2.6)], it follows that circles in the Γ_L plane map into circles in the Γ_{IN} plane, and circles in the Γ_s plane map into circles in the Γ_{OUT} plane. Specifically, these transformations are useful to map values of Γ_L on a constant G_p circle into a circle in the Γ_s plane, where $\Gamma_s = \Gamma_{\text{IN}}^*$ and values of Γ_s on a constant G_A circle into a circle in the Γ_L plane, where $\Gamma_L = \Gamma_{\text{OUT}}^*$.

In Appendix J it is shown that the values of Γ_L on a circle given by

$$|\Gamma_L - C_{oo}| = r_{oo}$$

map into a circle in the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane given by

$$|\Gamma_s - C_i| = r_i$$

where the center of the circle is given by

$$C_i = \frac{(1 - S_{22}C_{oo})(S_{11} - \Delta C_{oo})^* - r_{oo}^2 \Delta^* S_{22}}{|1 - S_{22}C_{oo}|^2 - r_{oo}^2 |S_{22}|^2} \quad (3.8.9)$$

and the radius is

$$r_i = \frac{r_{oo} |S_{12}S_{21}|}{||1 - S_{22}C_{oo}|^2 - r_{oo}^2 |S_{22}|^2|} \quad (3.8.10)$$

Similarly, a circle in the Γ_s plane given by

$$|\Gamma_s - C_{ii}| = r_{ii}$$

maps into a circle in the $\Gamma_L = \Gamma_{\text{OUT}}^*$ plane given by

$$|\Gamma_L - C_o| = r_o$$

where

$$C_o = \frac{(1 - S_{11}C_{ii})(S_{22} - \Delta C_{ii})^* - r_{ii}^2 \Delta^* S_{11}}{|1 - S_{11}C_{ii}|^2 - r_{ii}^2 |S_{11}|^2} \quad (3.8.11)$$

and

$$r_o = \frac{r_{ii} |S_{12}S_{21}|}{||1 - S_{11}C_{ii}|^2 - r_{ii}^2 |S_{11}|^2|} \quad (3.8.12)$$

Example 3.8.2

Design a microwave amplifier using a GaAs FET whose S parameters at 4 GHz, $V_{DS} = 2$ V, and $I_{DS} = 25$ mA are

$$\begin{aligned} S_{11} &= 0.55[-120^\circ] & S_{12} &= 0.14[30^\circ] \\ S_{21} &= 3.5[60^\circ] & S_{22} &= 0.2[-50^\circ] \end{aligned}$$

Analyze the trade-offs between operating power gain, stability, and VSWRs.

Solution. For this transistor, $K = 0.947$ and $\Delta = 0.521[-102.01^\circ]$. Since $K < 1$, the transistor is potentially unstable at 4 GHz. The centers and radii of the input and output stability circles are

$$\begin{aligned} C_s &= 16.47[130.7^\circ] & C_L &= 1.22[-59.25^\circ] \\ r_s &= 15.52 & r_L &= 2.12 \end{aligned}$$

The maximum stable gain is

$$G_{\text{MSG}} = \frac{|S_{21}|}{|S_{12}|} = \frac{3.5}{0.14} = 25 \text{ (or } 13.98 \text{ dB)}$$

Hence, we select a value of G_p lower than G_{MSG} . The value selected for this design is $G_p = 12$ dB, which is approximately 2 dB lower than G_{MSG} .

From (3.7.4) and (3.7.5), with $g_p = 15.849/(3.5)^2 = 1.294$, the center and radius of the $G_p = 12$ dB constant-gain circle are $C_p = 0.519[120.75^\circ]$ and $r_p = 0.639$. Figure 3.8.3a shows the $G_p = 12$ dB constant-gain circle and the output stability circle. Next we select values of Γ_L on the 12-dB gain circle that are far away from the output stability circle. The values of Γ_L on the 12-dB gain circle are given by $\Gamma_L = C_p + r_p e^{j\theta_1}$. The design procedure is described for two values of Γ_L , shown as point *a* (where $\theta_1 = 0$) and point *b* (where $\theta_1 = 3\pi/2$). At point *a* the value of Γ_L is $0.582[50.05^\circ]$, and at point *b* the value is $0.328[-143.98^\circ]$. Next we analyze the mapping of the $G_p = 12$ dB constant-gain circle onto the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane. Using (3.8.9) and (3.8.10) with $C_{oo} = C_p = 0.519[120.75^\circ]$ and $r_{oo} = r_p = 0.639$, we obtain $C_i = 0.8[130.7^\circ]$ and $r_i = 0.338$. This circle is drawn in the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane in Fig. 3.8.3b.

Several calculations are shown in Fig. 3.8.3c. Using (3.2.5), the second column shows the specific mapping of points *a* and *b* in the $\Gamma_s = \Gamma_{\text{IN}}^*$ plane. These points are denoted by *a'* and *b'* in Fig. 3.8.3b. The third column in Fig. 3.8.3c shows that $(\text{VSWR})_{\text{in}} = 1$ for $\Gamma_s = \Gamma_{\text{IN}}^*$. Using (3.2.6), the fourth column shows the corresponding values of Γ_{out} . From (3.8.6), the magnitude of $|\Gamma_b|$ is calculated in the fifth column. Finally, from (3.8.5),

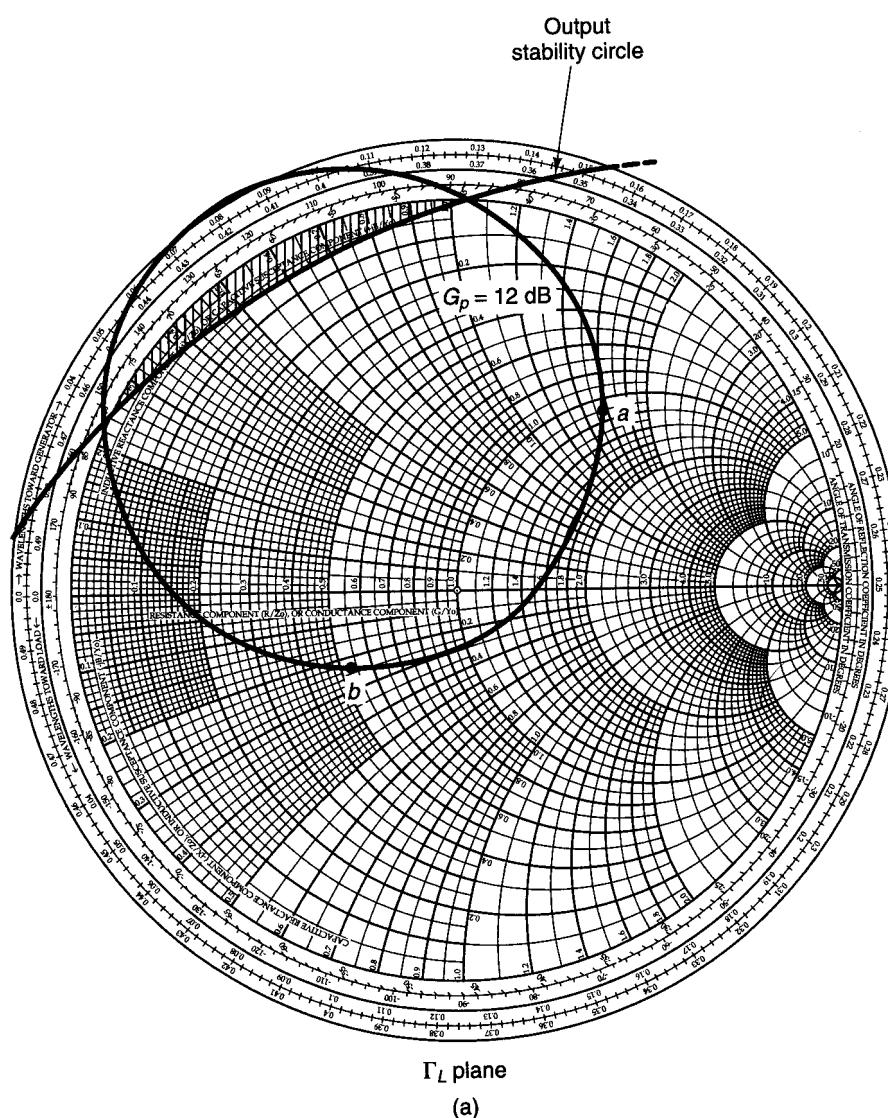


Figure 3.8.3 (a) The $G_p = 12 \text{ dB}$ constant-gain circle and the output stability circle; (b) mapping of the $G_p = 12 \text{ dB}$ constant-gain circle onto the $\Gamma_s = \Gamma_{IN}^*$ plane, the input stability circle, and the constant $(\text{VSWR})_{in} = 1.5$ circles; (c) calculations for Γ_L values at point a and b with $\Gamma_s = \Gamma_{IN}^*$ [i.e., $(\text{VSWR})_{in} = 1$]; (d) calculations for Γ_L values at points a and b with $(\text{VSWR})_{in} = 1.5$.

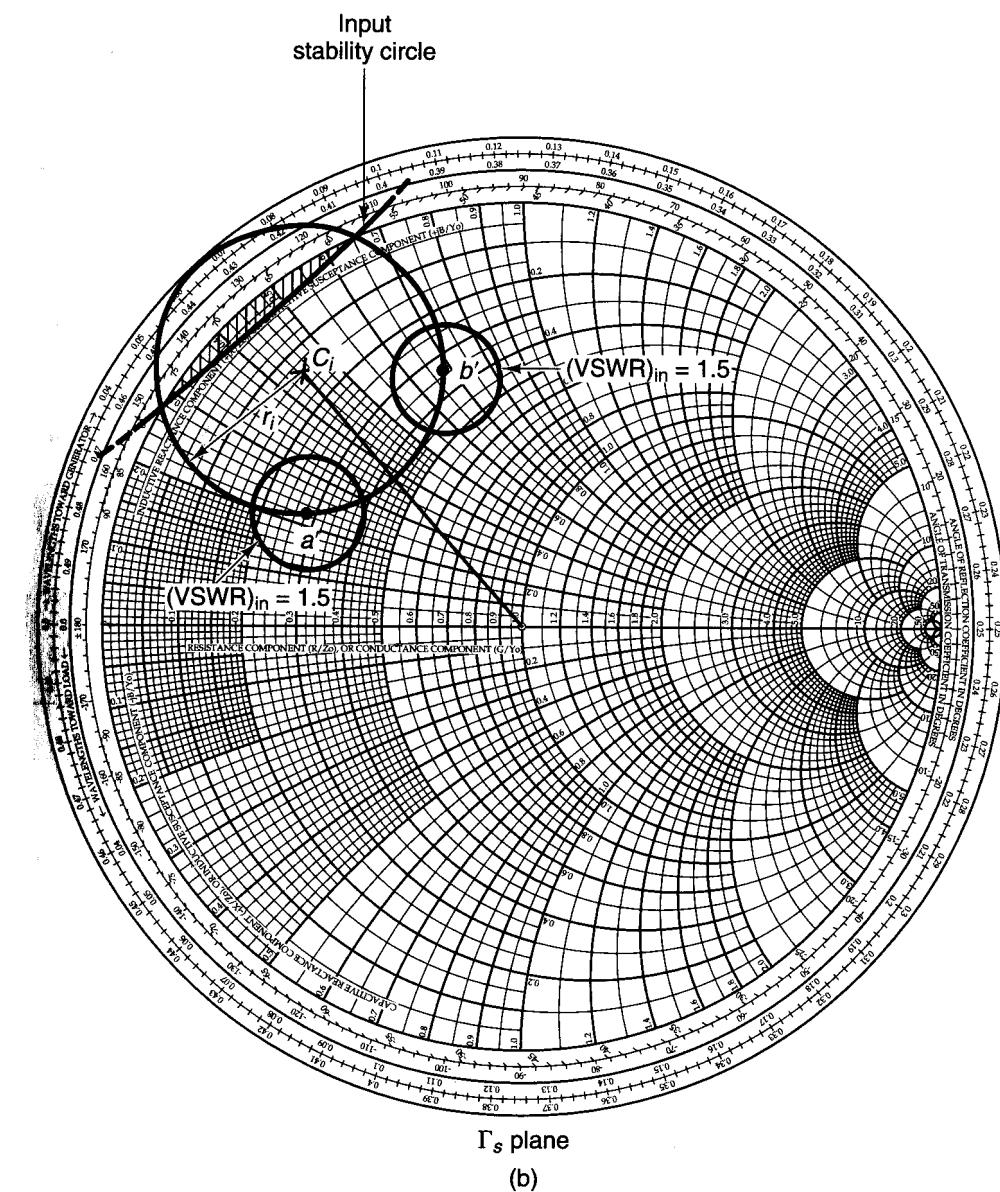


Figure 3.8.3 Continued

Γ_L	$\Gamma_s = \Gamma_{IN}^*$	(VSWR) _{in}	Γ_{OUT}	$ \Gamma_b $	(VSWR) _{out}
$0.582 50.05^\circ$ for $\theta_1 = 0$	$0.588 152.74^\circ$	1	$0.528 -85.97^\circ$	0.448	2.62
$0.328 -143.98^\circ$ for $\theta_1 = 270^\circ$	$0.625 107.15^\circ$	1	$0.399 -143.55^\circ$	0.447	2.62

(c)

Γ_L	Γ_{IN}	C_{Vi} and r_{Vi} for (VSWR) _{in} = 1.5	Γ_s	Γ_{OUT}	$ \Gamma_b $	(VSWR) _{out}
$0.582 50.05^\circ$ for $\theta_1 = 0$	$0.588 152.74^\circ$	$C_{Vi} = 0.572 152.74^\circ$ $r_{Vi} = 0.133$	$0.458 145.09^\circ$ for $\theta_2 = 0$ $0.643 142.16^\circ$ for $\theta_2 = \pi/2$ $0.692 157.78^\circ$ for $\theta_2 = \pi$ $0.525 165.77^\circ$ for $\theta_2 = 3\pi/2$	$0.410 -90.33^\circ$ $0.483 -74.36^\circ$ $0.569 -97.81^\circ$ $0.621 -80.87^\circ$	0.454 0.324 0.572 0.451	2.66 1.96 3.67 2.64
$0.328 -143.98^\circ$ for $\theta_1 = 270^\circ$	$0.625 -107.15^\circ$	$C_{Vi} = 0.61 107.15^\circ$ $r_{Vi} = 0.124$	$0.586 95.48^\circ$ for $\theta_2 = 0$ $0.730 104.28^\circ$ for $\theta_2 = \pi/2$ $0.657 117.53^\circ$ for $\theta_2 = \pi$ $0.493 111.40^\circ$ for $\theta_2 = 3\pi/2$	$0.293 -156.53^\circ$ $0.485 -156.23^\circ$ $0.315 -125.74^\circ$ $0.496 -130.80^\circ$	0.324 0.460 0.453 0.571	1.96 2.70 2.66 3.66

(d)

Figure 3.8.3 Continued

the sixth column lists the values of (VSWR)_{out}. The values of Γ_s at a' and b' are far away from the input stability circle. Hence, if (VSWR)_{out} = 2.62 is satisfactory, the design for $G_p = 12$ dB can be implemented with the values of Γ_s and Γ_L listed in Fig. 3.8.3c.

A smaller value of (VSWR)_{out} can be obtained by relaxing the input VSWR specifications. To illustrate this point, we design the amplifier with (VSWR)_{in} = 1.5. The design calculations are shown in Fig. 3.8.3d. The center and radius of the constant (VSWR)_{in} = 1.5 circle are made using (3.8.3) and (3.8.4) with $|\Gamma_a| = 0.2$. These values are listed in the third column in Fig. 3.8.3d, and the resulting (VSWR)_{in} = 1.5 circles are plotted in Fig. 3.8.3b.

The values of Γ_s on the (VSWR)_{in} = 1.5 circle are given by

$$\Gamma_s = C_{Vi} + r_{Vi}e^{j\theta_2}$$

Four convenient values on the (VSWR)_{in} = 1.5 circles are listed in the fourth column of Fig. 3.8.3d. They correspond to $\theta_2 = 0, \pi/2, \pi$, and $3\pi/2$. Using (3.2.6), Γ_{OUT} is calculated in the fifth column. From (3.8.6), $|\Gamma_b|$ is calculated in the sixth column. Finally, using (3.8.5), (VSWR)_{out} is calculated in the seventh column. From Fig. 3.8.3d, it is seen that the values associated with (VSWR)_{out} = 1.96 are far away from the input stability circle. Thus, a microwave amplifier with $G_p = 12$ dB can be designed with $\Gamma_L = 0.582 |35^\circ$ and $\Gamma_s = 0.643 |142.16^\circ$, or with $\Gamma_L = 0.328 |-143.98^\circ$ and $\Gamma_s = 0.586 |95.48^\circ$.

It is of interest to analyze the design for a value of G_p greater than G_{MSG} (e.g., $G_p = 15$ dB). From (3.7.4), with $g_p = 31.623/(3.5)^2 = 2.581$, the center and radius of the $G_p = 15$ dB constant-gain circle are $C_p = 1.8 |120.75^\circ$ and $r_p = 1.12$. Using (3.8.9) and (3.8.10), the 15-dB gain circle is mapped into a circle in the $\Gamma_s = \Gamma_{IN}^*$ plane with center and radius given by $C_i = 1.53 |130.7^\circ$ and $r_i = 0.652$.

The 15-dB gain circle is shown in Fig. 3.8.4a and its mapping in the $\Gamma_s = \Gamma_{IN}^*$ plane in Fig. 3.8.4b. Two values of Γ_L are selected on the 15-dB gain circle. These points are denoted by a and b in Fig. 3.8.4a. The mapping of the points a and b is shown in the $\Gamma_s = \Gamma_{IN}^*$

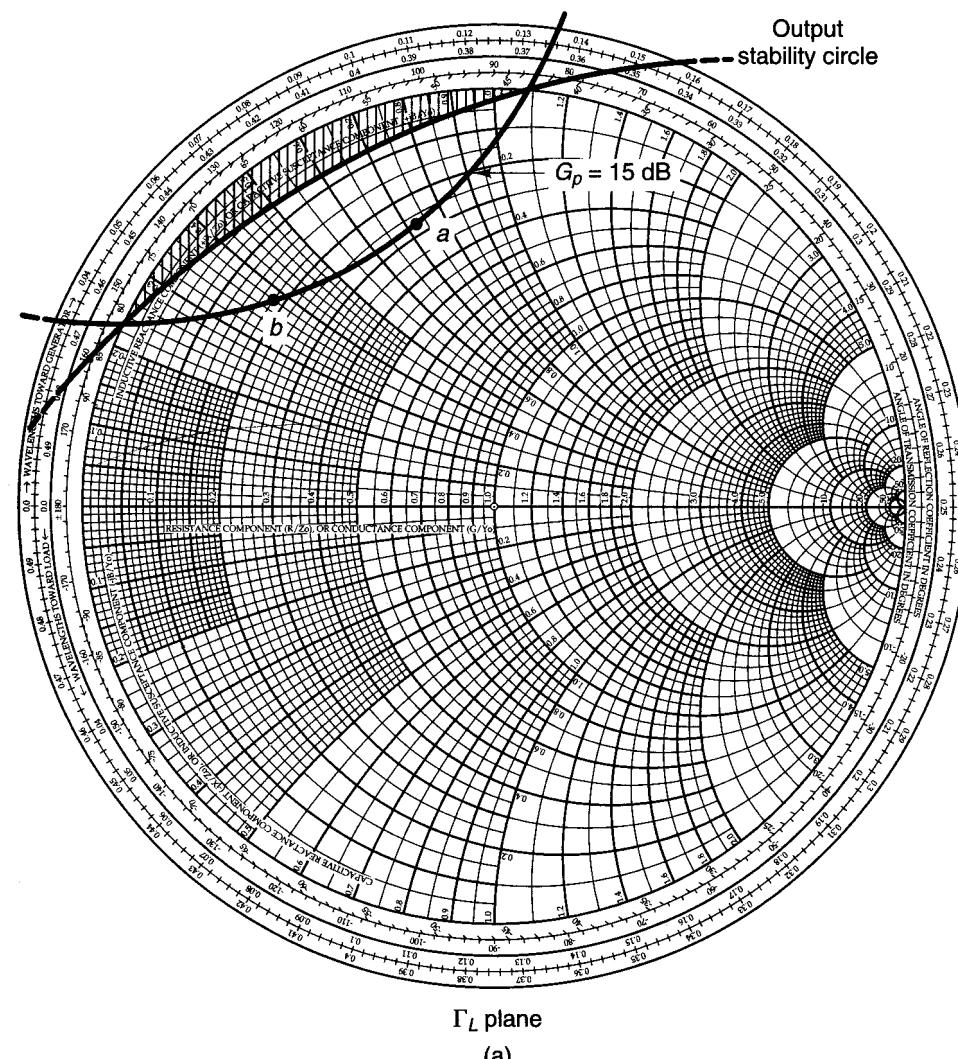
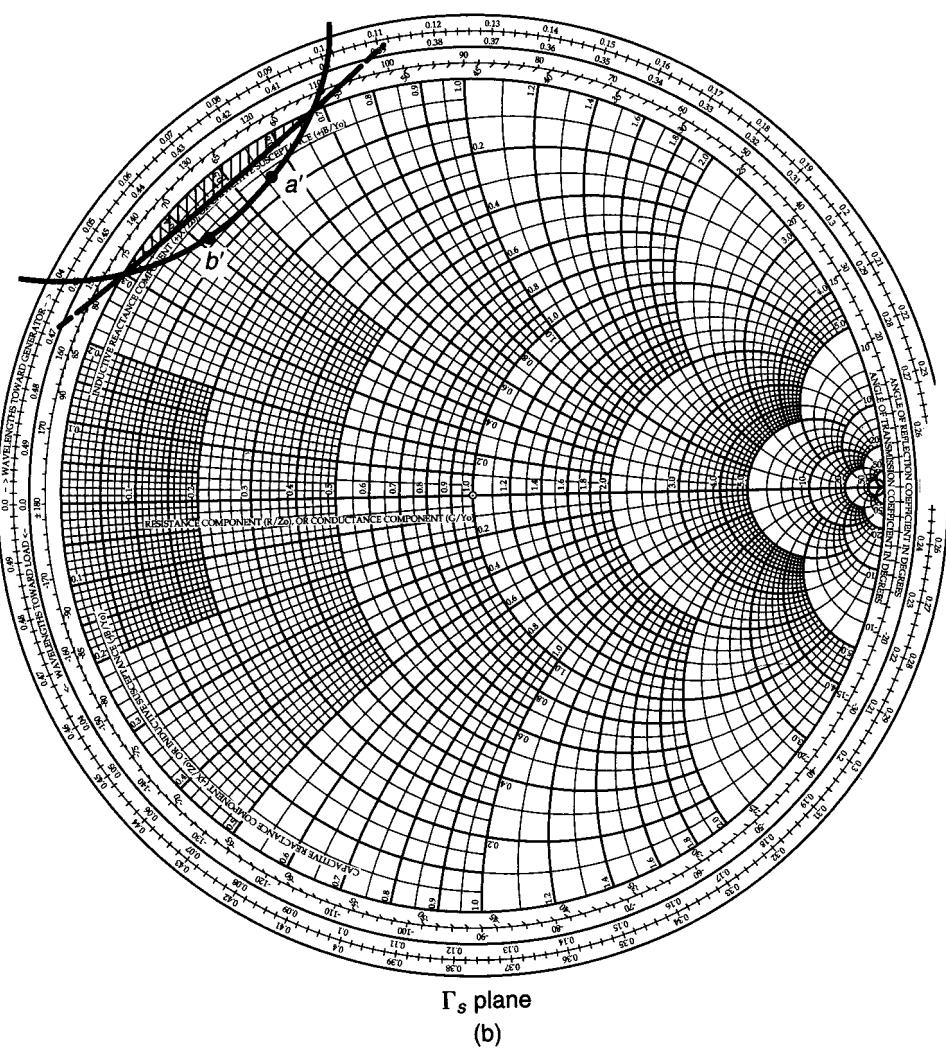


Figure 3.8.4 (a) The $G_p = 15$ dB constant-gain circle and the output stability circle; (b) mapping of the $G_p = 15$ dB constant-gain circle onto the $\Gamma_s = \Gamma_{IN}^*$ plane, and the input stability circle; (c) calculations for Γ_L values at points a and b with $\Gamma_s = \Gamma_{IN}^*$ [i.e., (VSWR)_{in} = 1].



Γ_L	$\Gamma_s = \Gamma_{IN}^*$	$(VSWR)_{in}$	Γ_{OUT}	$ \Gamma_b $	$(VSWR)_{out}$
0.717 106.22°	0.884 134.73°	1	0.878 -109.62°	0.449	2.63
0.730 137.37°	0.906 121.68°	1	0.884 -133.66°	0.453	2.65

(c)

Figure 3.8.4 Continued

plane as a' and b' in Fig. 3.8.4b. The values of Γ_L at points a and b are listed in Fig. 3.8.4c, as well as the design calculations. The third column shows that $(VSWR)_{in} = 1$ for $\Gamma_s = \Gamma_{IN}^*$. The fourth column shows the corresponding values of Γ_{OUT} , the fifth column the values of $|\Gamma_b|$, and the sixth column the values of $(VSWR)_{out}$.

An examination of Fig. 3.8.4b shows that the points a' and b' are very close to the unstable region, and the resulting $(VSWR)_{out}$ is greater than 2.63. This is typical of a design when the value of G_p is selected close to the value of G_{MSG} or greater than G_{MSG} .

Further improvements can be done by performing calculations of the type illustrated in Figs. 3.8.3c and 3.8.3d. However, a CAD program can simplify the task considerably. In Appendix "Computer-Aided Designs," Example CAD.2, it is shown that with $\Gamma_L = 0.328 | -143.98^\circ$ and $\Gamma_s = 0.612 | 95.2^\circ$ it follows that $G_p = 12$ dB, $(VSWR)_{in} = 1.5$, and $(VSWR)_{out} = 1.93$. This is the smallest $(VSWR)_{out}$ that can be obtained with Γ_s on the $(VSWR)_{in} = 1.5$ circle shown in Fig. 3.8.3b. Furthermore, a CAD optimization to minimize the VSWRs shows that with $\Gamma_L = 0.319 | -176.51^\circ$ and $\Gamma_s = 0.642 | 102.61^\circ$, it follows that $G_p = 12.7$ dB, $(VSWR)_{in} = 1.52$, and $(VSWR)_{out} = 1.58$.

3.9 DC BIAS NETWORKS

It has been said that the least considered factor in microwave transistor amplifier design is the bias network [3.4]. While considerable effort is spent in designing for a given gain, noise figure, and bandwidth, little effort is spent in the dc bias network. The cost per decibel of microwave power gain or noise figure is high, and the designer cannot sacrifice the amplifier performance by having a poor dc bias design.

The purpose of a good dc bias design is to select the proper quiescent point and hold the quiescent point constant over variations in transistor parameters and temperature. A resistor bias network can be used with good results over moderate temperature changes. However, an active bias network is usually preferred for large temperature changes.

In the discussion that follows, we first consider the dc bias design for BJTs and then the bias design of GaAs FETs.

BJT Bias Networks

At low frequencies, a bypassed emitter resistor is an important contributor to the quiescent-point stability. At microwave frequencies, the bypass capacitor, which is in parallel with the emitter resistor, can produce oscillations by making the input port unstable at some frequencies. Furthermore, an emitter resistor will degrade the noise performance of the amplifier. Therefore, in most microwave transistor amplifiers, especially in the gigahertz region, the emitter lead of the transistor is grounded.

At microwave frequencies, the transistor parameters that are affected most by temperature are I_{CBO} , h_{FE} , and V_{BE} . The conventional reverse current I_{CBO} (i.e., I_{CBO} at low frequencies) doubles every 10°C rise in temperature. That is,

$$I_{CBO,T_2} = I_{CBO,T_1} 2^{(T_2 - T_1)/10}$$

where I_{CBO,T_2} and I_{CBO,T_1} are the values of I_{CBO} at temperatures T_2 and T_1 , respectively. The temperature T_1 is usually the temperature at which the manufacturer measures I_{CBO} . This temperature is usually 25°C.

A microwave transistor has a more complicated reverse current flow. The reverse current flow of a microwave transistor is composed of two components; one is the conventional I_{CBO} and the other is a surface current, I_s , that flows across the top of the silicon lattice. The reverse current in a microwave transistor, which is referred to simply as I_{CBO} , increases at a rate much slower than the conventional I_{CBO} . A typical plot of the reverse current versus temperature for a microwave transistor is shown in Fig. 3.9.1. The conventional I_{CBO} slope is also shown in the figure for comparison.

The base-to-emitter voltage V_{BE} has a negative temperature coefficient, approximately given by

$$\frac{\Delta V_{BE}}{\Delta T} \approx -2 \times 10^{-3} \frac{V}{^{\circ}C}$$

The dc value of the current gain h_{FE} is defined as the value of the collector-to-base current at a constant value of V_{CE} . That is,

$$h_{FE} = \left. \frac{I_C}{I_B} \right|_{V_{CE}=\text{constant}}$$

The dc value of h_{FE} is typically found to increase linearly with temperature at the rate of 0.5%/°C.

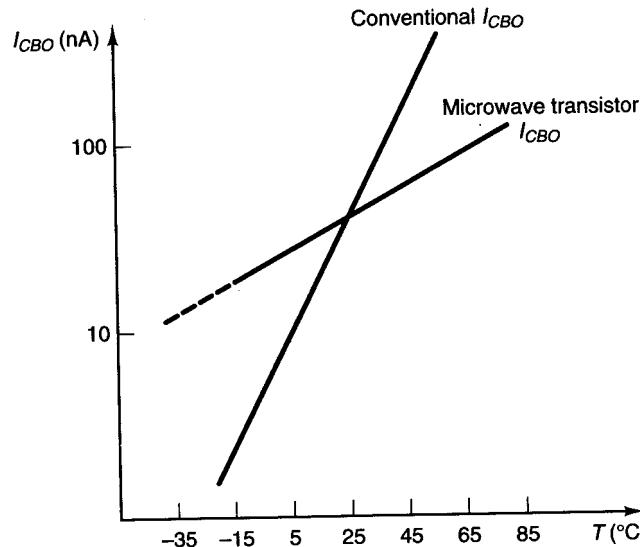


Figure 3.9.1 Typical reverse current versus temperature for a microwave transistor.

In order to find the change in collector current as a function of temperature in a dc bias network, we first find the expression for the collector current valid for any temperature. Then, observing that the temperature-sensitive parameters are I_{CBO} , h_{FE} , and V_{BE} , we can write

$$I_C = f(I_{CBO}, h_{FE}, V_{BE})$$

and

$$\Delta I_C = \left(\frac{\Delta I_C}{\Delta I_{CBO}} \right) \Big|_{\Delta V_{BE}=0} \Delta I_{CBO} + \left(\frac{\Delta I_C}{\Delta h_{FE}} \right) \Big|_{\Delta I_{CBO}=0} \Delta h_{FE} + \left(\frac{\Delta I_C}{\Delta V_{BE}} \right) \Big|_{\Delta h_{FE}=0} \Delta V_{BE} \quad (3.9.1)$$

Defining the stability factors as

$$S_i = \left. \frac{\Delta I_C}{\Delta I_{CBO}} \right|_{\Delta h_{FE}=0, \Delta V_{BE}=0}$$

$$S_{h_{FE}} = \left. \frac{\Delta I_C}{\Delta h_{FE}} \right|_{\Delta I_{CBO}=0, \Delta V_{BE}=0}$$

and

$$S_{V_{BE}} = \left. \frac{\Delta I_C}{\Delta V_{BE}} \right|_{\Delta h_{FE}=0, \Delta I_{CBO}=0}$$

we can write (3.9.1) in the form

$$\Delta I_C = S_i \Delta I_{CBO} + S_{h_{FE}} \Delta h_{FE} + S_{V_{BE}} \Delta V_{BE} \quad (3.9.2)$$

For a given dc bias network, the stability factors can be calculated and (3.9.2) can be used to predict the variations of I_C with temperature. In a design procedure, the maximum variation of I_C in a temperature range can be selected and (3.9.2) can be used to find the required stability factors. In turn, the stability factors together with the Q -point location will fix the value of the resistors in the bias network.

Two grounded-emitter dc bias networks that can be used at microwave frequencies are shown in Fig. 3.9.2. The network in Fig. 3.9.2b produces lower values of resistance and therefore is more compatible with thin- or thick-film resistor values.

Example 3.9.1

Design the dc bias network shown in Fig. 3.9.2b for $V_{CE} = 10$ V and $I_C = 10$ mA. Assume that $I_{CBO} = 0$, $V_{BE} = 0.7$ V, and $h_{FE} = 50$.

Solution. In this example we follow a procedure that results in good stability factors. Let the supply voltage V_{CC} be 20 V. The base current (I_B) is

$$I_B = \frac{I_C}{h_{FE}} = \frac{10 \times 10^{-3}}{50} = 200 \mu A$$

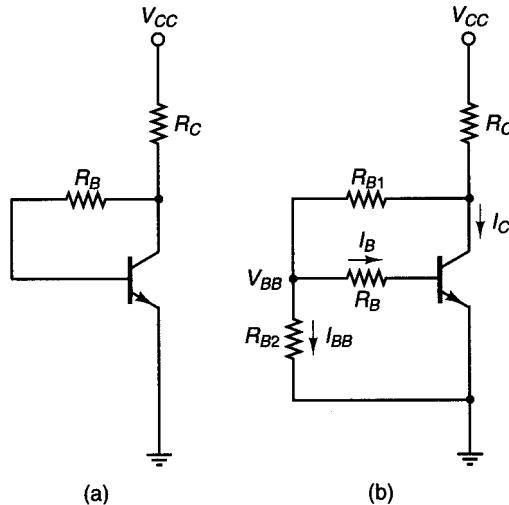


Figure 3.9.2 (a) Voltage feedback bias network; (b) voltage feedback bias network with constant-base current source.

Assuming \$V_{BB}\$ to be 2 V, we find that

$$R_B = \frac{V_{BB} - V_{BE}}{I_B} = \frac{2 - 0.7}{200 \times 10^{-6}} = 6.5 \text{ k}\Omega$$

\$R_{B2}\$ is calculated assuming that \$I_{BB} = 1\$ mA (i.e., \$I_{BB} = 5I_B\$)—namely,

$$R_{B2} = \frac{V_{BB}}{I_{BB}} = \frac{2}{1 \times 10^{-3}} = 2 \text{ k}\Omega$$

\$R_{B1}\$ is obtained from

$$R_{B1} = \frac{V_{CE} - V_{BB}}{I_{BB} + I_B} = \frac{10 - 2}{(1 + 0.2) \times 10^{-3}} = 6.66 \text{ k}\Omega$$

and \$R_C\$ is obtained from

$$R_C = \frac{V_{CC} - V_{CE}}{I_C + I_{BB} + I_B} = \frac{20 - 10}{(10 + 1 + 0.2) \times 10^{-3}} = 893 \Omega$$

The assumption \$I_{BB} \gg I_B\$ and \$V_{BB} \approx 10\% V_{CC}\$ produces good stability factors.

At the lower microwave frequencies, the dc biasing network shown in Fig. 3.9.3 with a bypassed emitter resistor can be used. The bypassed emitter resistor provides excellent stability. For this network, it is easy to show that

$$I_C = \frac{h_{FE}(V_{TH} - V_{BE})}{R_{TH} + (h_{FE} + 1)R_E} + \frac{(h_{FE} + 1)I_{CBO}(R_{TH} + R_E)}{R_{TH} + (h_{FE} + 1)R_E}$$

Figure 3.9.3 A dc bias network with a bypassed emitter resistor.

where

$$V_{TH} = \frac{V_{CC}R_2}{R_1 + R_2}$$

and

$$R_{TH} = \frac{R_1R_2}{R_1 + R_2}$$

The stability factors are

$$S_i = \frac{(h_{FE} + 1)(R_{TH} + R_E)}{R_{TH} + (h_{FE} + 1)R_E}$$

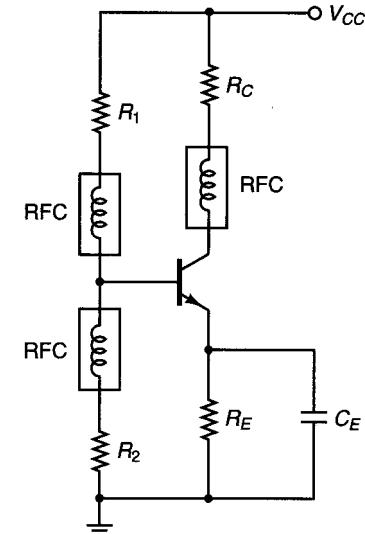
$$S_{h_{FE}} \approx \frac{I_{C1}}{h_{FE}} \frac{S_{i2}}{h_{FE,2}} \quad (3.9.3)$$

and

$$S_{V_{BE}} = \frac{-h_{FE}}{R_{TH} + (h_{FE} + 1)R_E}$$

In (3.9.3), \$\Delta h_{FE} = h_{FE,2} - h_{FE}\$ and \$S_{i2}\$ is the value of \$S_i\$ with \$h_{FE} = h_{FE,2}\$.

An active dc biasing network is shown in Fig. 3.9.4. A pnp BJT is used to stabilize the operating point of the microwave transistor. The bypass capacitors \$C_1\$ and \$C_2\$ are typically 0.01-\$\mu\$F disk capacitors. The radio frequency chokes (RFCs) are typically made of two or three turns of No. 36 enameled wire on 0.1-in. air core. The operation of the network is as follows. If \$I_{C2}\$ tends to increase, the current \$I_3\$ increases and the emitter-to-base voltage of \$Q_1\$ (\$V_{EB,1}\$) decreases. The decrease of \$V_{EB,1}\$ decreases \$I_{E1}\$, which in turn decreases \$I_{C2}\$ and \$I_{B2}\$. The decrease in \$I_{B2}\$ and \$I_{C2}\$ produces the desired bias stability.



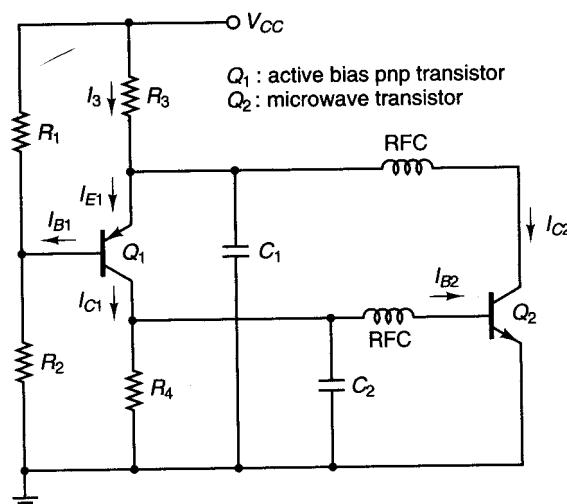


Figure 3.9.4 Active bias network for a BJT.

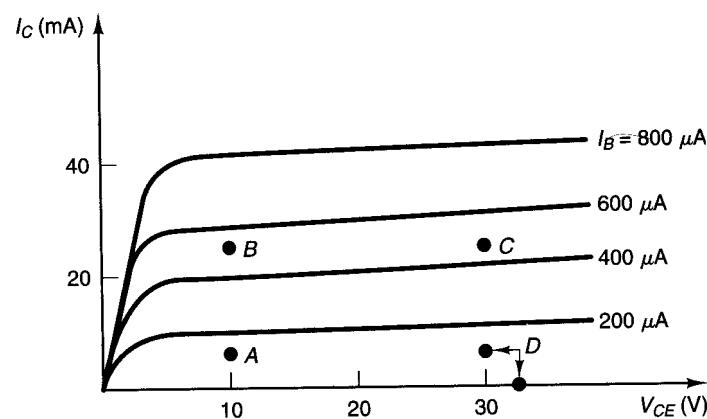


Figure 3.9.5 Selection of the dc operating point.

The selection of the dc quiescent point for a BJT depends on the particular application. For low-noise and low-power applications, the quiescent point A in Fig. 3.9.5 is recommended. At A , the BJT operates at low values of collector current. For low noise and higher power gain, the quiescent point at B is recommended. For high output power, in class A operation, the quiescent point at C is recommended. For higher output power and higher efficiency, the BJT is operated in class AB or B, using the quiescent point at D .

Example 3.9.2

(a) Design the bias circuit in Fig. 3.9.3 to have a quiescent point at $V_{CE} = 8 \text{ V}$ and $I_C = 2 \text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.7 \text{ V}$.

(b) Design the active bias network in Fig. 3.9.4 to set the quiescent point of transistor Q_2 at $V_{CE2} = 8 \text{ V}$ and $I_{C2} = 2 \text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.7 \text{ V}$.

Solution. (a) Let $V_{CC} = 15 \text{ V}$ in Fig. 3.9.3. Then, with $I_C \approx I_E$, the dc load-line equation is

$$V_{CC} = V_{CE} + I_C(R_C + R_E)$$

or

$$R_C + R_E = \frac{15 - 8}{2 \times 10^{-3}} = 3.5 \text{ k}\Omega \quad (3.9.4)$$

The value of R_E is usually selected using the assumption that the voltage across R_E is 10% to 20% of V_{CC} . Selecting 10% of V_{CC} , we obtain

$$R_E = \frac{10\% V_{CC}}{I_C} = \frac{0.1(15)}{2 \times 10^{-3}} = 750 \Omega$$

Then, from (3.9.4),

$$R_C = 3500 - 750 = 2.75 \text{ k}\Omega$$

A Thévenin's equivalent circuit between the base and ground of the transistor gives

$$V_{TH} = \frac{V_{CC} R_2}{R_2 + R_1} \quad (3.9.5)$$

and

$$R_{TH} = \frac{R_1 R_2}{R_1 + R_2} \quad (3.9.6)$$

From (3.9.5) and (3.9.6), we can solve for R_1 and R_2 in terms of V_{TH} and R_{TH} —namely,

$$R_1 = R_{TH} \frac{V_{CC}}{V_{TH}} \quad (3.9.7)$$

and

$$R_2 = \frac{R_{TH}}{1 - \frac{V_{TH}}{V_{CC}}} \quad (3.9.8)$$

For good beta stability, the value of R_{TH} is selected such that $\beta R_E = 10 R_{TH}$, or

$$R_{TH} = \frac{\beta R_E}{10} = \frac{100(750)}{10} = 7.5 \text{ k}\Omega$$

The value of V_{TH} is calculated using the loop equation:

$$V_{TH} = I_B R_{TH} + 0.7 + I_E R_E = \frac{2 \times 10^{-3}}{100} (7.5 \times 10^3) + 0.7 + 2 \times 10^{-3}(750) = 2.35 \text{ V}$$

From (3.9.7) and (3.9.8), it follows that

$$R_1 = (7.5 \times 10^3) \frac{15}{2.35} = 47.9 \text{ k}\Omega$$

and

$$R_2 = \frac{7.5 \times 10^3}{1 - \frac{2.35}{15}} = 8.9 \text{ k}\Omega$$

(b) Let $V_{CC} = 15 \text{ V}$ in Fig. 3.9.4. The base voltage of transistor Q_2 , which is equal to the voltage across R_4 , is 0.7 V . The current I_3 is equal to the sum of I_{C1} and I_{C2} (where $I_{C2} = 2 \text{ mA}$). Designing for I_3 equal to 4 mA , it follows that

$$I_{C1} = I_3 - I_{C2} = 4 \times 10^{-3} - 2 \times 10^{-3} = 2 \text{ mA}$$

Then

$$R_4 = \frac{0.7}{I_{C1}} = \frac{0.7}{2 \times 10^{-3}} = 350 \Omega$$

Since the collector voltage of transistor Q_2 is 8 V , R_3 is calculated from

$$R_3 = \frac{V_{CC} - 8}{I_3} = \frac{15 - 8}{4 \times 10^{-3}} = 1.75 \text{ k}\Omega$$

In the active bias circuit of Fig. 3.9.4, good beta stability is obtained by letting the current in R_1 and R_2 be 20 times the base current of transistor Q_1 . That is,

$$I_{R_1} \approx I_{R_2} = 20I_{B1} = 20\left(\frac{2 \times 10^{-3}}{100}\right) = 0.4 \text{ mA}$$

Since the base voltage of transistor Q_1 is a 7.3 V , we can write

$$R_1 = \frac{V_{CC} - 7.3}{I_{R_1}} = \frac{15 - 7.3}{0.4 \times 10^{-3}} = 19.25 \text{ k}\Omega$$

and

$$R_2 = \frac{7.3}{I_{R_2}} = \frac{7.3}{0.4 \times 10^{-3}} = 18.25 \text{ k}\Omega$$

In the active bias circuit, the resistors R_2 and R_3 are implemented using a potentiometer in order to set the quiescent point of transistor Q_2 at exactly 8 V and 2 mA . An optimization and beta stability analysis of the active bias circuit in this design is performed in Appendix "Computer-Aided Designs," Example CAD.1.

GaAs FET Bias Networks

The GaAs FETs can be biased in several ways. Five basic dc network configurations for GaAs FET amplifiers are shown in Fig. 3.9.6 [3.5]. The dc bias network in Fig. 3.9.6a requires a bipolar power source, while the networks in Figs. 3.9.6b to 3.9.6e require a unipolar supply. The column "How" in Fig. 3.9.6 indicates the polarity of the sources, as well as the sequence in which the voltages must be applied to prevent transient burnout of the GaAs FET device during turn-on. For example, in the dc bias network in Fig. 3.9.6a, if the drain is biased

Figure	How	Amplifier characteristics	Power supply used
(a)	Apply V_G , then V_D	Low noise High gain High power High efficiency	Bipolar, Minimum source inductance
(b)	Apply V_S , then V_D	[same as (a)]	Positive supply
(c)	Apply V_S , then V_G	[same as (a)]	Negative supply
(d)	Apply V_D	Low noise High gain High power Lower efficiency Gain easily adjusted by varying R_S	Unipolar, incorporating R_S automatic transient protection
(e)	Apply V_G	[same as (d)]	Negative unipolar, incorporating R_S

Figure 3.9.6 Five basic dc bias networks. (From G. D. Vendelin [3.5]; reproduced with permission of *Microwaves & RF*)

positive before the gate, the transistor will operate momentarily beyond its safe operating region. Therefore, the proper turn-on sequence is: first apply a negative bias to the gate (i.e., $V_G < 0$) and then apply the drain voltage ($V_D > 0$). One method to accomplish the previous turn-on procedure is to turn both sources at the same time and to include a long RC time constant network in the V_D supply and a short RC time constant network in the negative supply V_G .

The bias networks in Figs. 3.9.6d and 3.9.6e use a source resistor. The source resistor provides automatic transient protection. However, the source resistor will degrade the noise-figure performance, and the source bypass capacitor can cause low-frequency oscillations.

The decoupling capacitors shown in Fig. 3.9.6 are sometimes shunted with zener diodes. The zener diodes provide additional protection against transients, reverse biasing, and overvoltage.

The dc bias network of a GaAs FET must provide a stable quiescent point. It is not difficult to show that the negative feedback resistor R_s decreases the effect of variations of I_D with respect to temperature and I_{DSS} .

The selection of the dc quiescent point in a GaAs FET depends on the particular application. Figure 3.9.7 shows typical GaAs FET characteristics with four quiescent points located at A, B, C, and D.

For low-noise, low-power application, the quiescent point A is recommended. At A, the FET operates at a low value of current (i.e., $I_{DS} \approx 0.15I_{DSS}$).

For low noise and higher power gain, the recommended quiescent point is at B. The bias voltage remains the same as for point A, but the drain current is increased to $I_{DS} \approx 0.9I_{DSS}$.

The GaAs FET output power level can be increased by selecting the quiescent point at C with $I_{DS} \approx 0.5I_{DSS}$. The quiescent point at C maintains class A operation. For higher efficiency, or to operate the GaAs FET in class

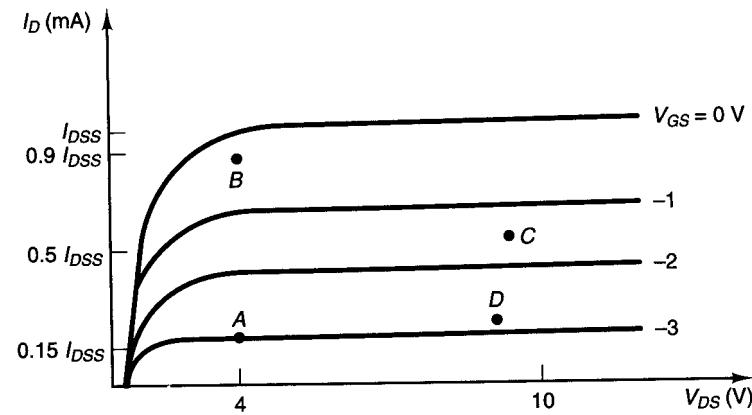


Figure 3.9.7 Typical GaAs FET characteristics and recommended quiescent points.

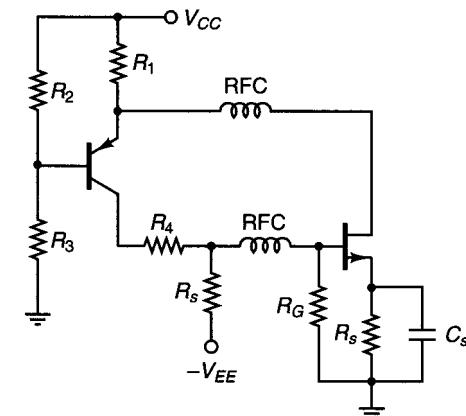


Figure 3.9.8 Active bias for a common-source GaAs FET.

AB or B, the drain-to-source current must be decreased and the quiescent point D is recommended.

An active bias network for a common-source GaAs FET is shown in Fig. 3.9.8.

PROBLEMS

- 3.1 (a)** Show that $G_T \leq G_A$ and $G_T \leq G_p$. When is the equality sign satisfied?
(b) Show that (3.2.3) can be obtained from (3.2.1) when $\Gamma_s = \Gamma_{IN}^*$, and (3.2.4) from (3.2.2) when $\Gamma_L = \Gamma_{OUT}^*$.
- 3.2 (a)** Show that the transducer power gain is given by $G_T = |S_{21}|^2$ when the source and load impedances to the transistor are equal to the reference impedance Z_0 (usually 50 Ω).
(b) Determine the expression for G_p and G_A when the source and load impedances are real and equal to Z_0 .
- 3.3 (a)** A microwave amplifier diagram is shown in Fig. 3.2.2. Determine G_T , G_A , and G_p if $\Gamma_s = 0.49 \angle -150^\circ$, $\Gamma_L = 0.56 \angle 90^\circ$, and the S parameters of the transistor are

$$S_{11} = 0.54 \angle 165^\circ \quad S_{12} = 0.09 \angle 20^\circ$$

$$S_{21} = 2 \angle 30^\circ \quad S_{22} = 0.5 \angle -80^\circ$$

- (b)** Calculate P_{AVS} , P_{IN} , P_{AVN} , and P_L if $E_1 = 10 \angle 30^\circ$, $Z_1 = 50 \Omega$, and $Z_2 = 50 \Omega$.

- 3.4** The S parameters of a transistor are

$$S_{11} = 0.7 \angle 30^\circ \quad S_{12} = 0$$

$$S_{21} = 4 \angle 90^\circ \quad S_{22} = 0.5$$

The transistor is used in the amplifier shown in Fig. P3.4, where the output matching network produces $\Gamma_L = 0.5 \angle 90^\circ$. Determine the values of G_T , G_p , and G_A .

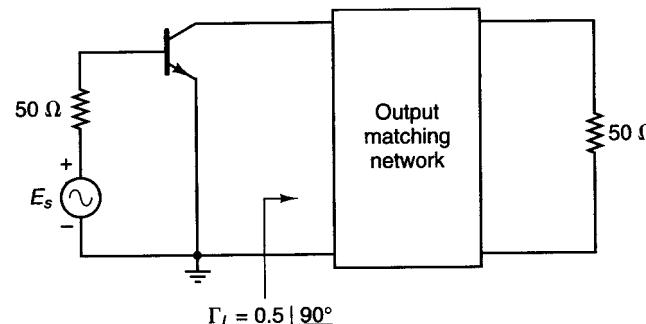


Figure P3.4

3.5 In each of the stability circle drawings shown in Fig. P3.5, indicate clearly the possible locations for a stable source reflection coefficient.

3.6 Two output stability circles are shown in Fig. P3.6. Determine the stable region for the load reflection coefficient.

3.7 The scattering parameters for three different transistors are given below. Determine the stability in each case and in a potentially unstable case, draw the input and output stability circles.

(a) $S_{11} = 0.674 | -152^\circ$
 $S_{12} = 0.075 | 6.2^\circ$
 $S_{21} = 1.74 | 36.4^\circ$
 $S_{22} = 0.6 | -92.6^\circ$

(b) $S_{11} = 0.385 | -55^\circ$
 $S_{12} = 0.045 | 90^\circ$
 $S_{21} = 2.7 | 78^\circ$
 $S_{22} = 0.89 | -26.5^\circ$

(c) $S_{11} = 0.7 | -50^\circ$
 $S_{12} = 0.27 | 75^\circ$
 $S_{21} = 5 | 120^\circ$
 $S_{22} = 0.6 | 80^\circ$

3.8 The S parameters of a GaAs FET at a given Q point are

f (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
4	0.9 -67°	0.076 43°	2.3 118°	0.68 -39°
6	0.84 -97°	0.112 24°	2.06 87°	0.6 -58°
8	0.73 -140°	0.135 -5°	2.04 53°	0.47 -85°
10	0.67 -178°	0.146 -27°	1.81 18°	0.42 -120°
14	0.63 115°	0.133 -66°	1.42 -38°	0.36 -172°

Draw the input stability circles (at each frequency) in a Smith chart and the output stability circles in another Smith chart. Indicate the unstable regions.

3.9 This problem analyzes some interesting (theoretical) cases of the stability circles. The S parameters of several two-port networks are

- (a) $S_{11} = 1/\sqrt{2}$, $S_{12} = 1/\sqrt{2} | -180^\circ$, $S_{21} = 1/\sqrt{2}$, and $S_{22} = 1/\sqrt{2}$.
- (b) $S_{11} = 1/\sqrt{2}$, $S_{12} = 1.707$, $S_{21} = 1.707$, and $S_{22} = 1/\sqrt{2}$.
- (c) $S_{11} = 1/\sqrt{2}$, $S_{12} = 0.292$, $S_{21} = 0.292$, and $S_{22} = 1/\sqrt{2}$.
- (d) $S_{11} = 1$, $S_{12} = \sqrt{2}$, $S_{21} = \sqrt{2}$, and $S_{22} = 1$.

Determine K and $|\Delta|$ and draw the input and output stability circles.

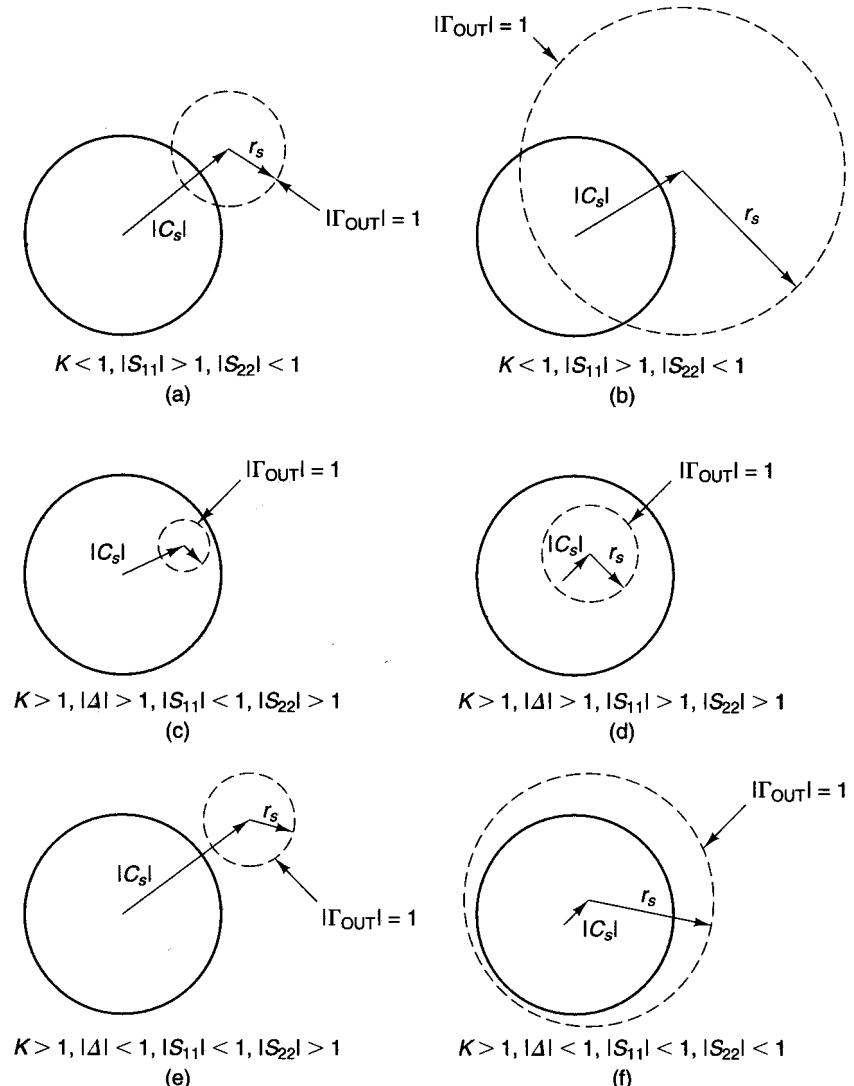


Figure P3.5

3.10 (a) Show that in the limit as S_{12} approaches zero, the center and radius of the stability circles are $C_S \approx 1/S_{22}$, $r_s \approx 0$, $C_L \approx 1/S_{11}$, and $r_L \approx 0$.

(b) The S parameters of a two-port network are

$$\begin{aligned} S_{11} &= 2 | 90^\circ & S_{12} &= 0 \\ S_{21} &= 2 & S_{22} &= 0.1 | 45^\circ \end{aligned}$$

Draw the stability circles and show the unstable regions.

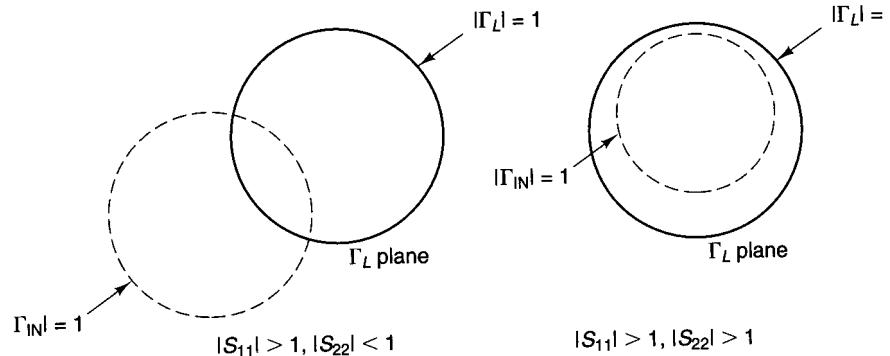


Figure P3.6

- 3.11** Show that the source stability circle does not enclose the center of the Smith chart when

$$|S_{22}| < 1 \text{ and } |\Delta| < S_{11}$$

or

$$|S_{22}| > 1 \text{ and } |\Delta| > S_{11}$$

Also show that the load stability circle does not enclose the center of Smith chart when

$$|S_{11}| < 1 \text{ and } |\Delta| < S_{22}$$

or

$$|S_{11}| > 1 \text{ and } |\Delta| > S_{22}$$

- 3.12** The conditions for unconditional stability were analyzed by considering the values in the Γ_s and Γ_L plane that result in $|\Gamma_{IN}| < 1$ and $|\Gamma_{OUT}| < 1$. An alternative approach is to consider the values in the Γ_{IN} and Γ_{OUT} plane, where $|\Gamma_{IN}| < 1$ and $|\Gamma_{OUT}| < 1$, that results in $|\Gamma_s| < 1$ and $|\Gamma_L| < 1$. Using this approach, show that the plot of the $|\Gamma_s| = 1$ circle in the Γ_{OUT} plane has radius and center given by

$$r_{OUT} = \frac{|S_{12}S_{21}|}{1 - |S_{11}|^2}$$

and

$$C_{OUT} = S_{22} + \frac{S_{12}S_{21}S_{11}^*}{1 - |S_{11}|^2} = \frac{S_{22} - \Delta S_{11}^*}{1 - |S_{11}|^2}$$

Also, the plot of the $|\Gamma_L| = 1$ circle in the Γ_{IN} plane has radius and center given by

$$r_{IN} = \frac{|S_{12}S_{21}|}{1 - |S_{22}|^2}$$

and

$$C_{IN} = S_{11} + \frac{S_{12}S_{21}S_{22}^*}{1 - |S_{22}|^2} = \frac{S_{11} - \Delta S_{22}^*}{1 - |S_{22}|^2}$$

- 3.13 (a)** Show that

$$|\Delta| \leq |S_{11}| |S_{22}| + |S_{12}S_{21}|$$

and

$$|S_{11}| |S_{22}| \leq |\Delta| + |S_{12}S_{21}|$$

Substitute these inequalities in (3.3.13) and verify that

$$(1 - |\Delta|)^2 > (|S_{11}|^2 - |S_{22}|^2)^2$$

Therefore, show that

$$B_1 B_2 > 0$$

and

$$B_1 + B_2 = 2(1 - |\Delta|^2)$$

where B_1 is given by (3.3.21) and

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2$$

- (b)** Use the previous results to show that the conditions $K > 1$ and $B_1 > 0$ are similar to $K > 1$ and $B_2 > 0$ (see Appendix C).

- (c)** Show that the condition $|\Delta| < 1$ is similar to $B_1 > 0$.

- 3.14** Show how resistive loading can stabilize a transistor whose S parameters at $f = 750$ MHz are

$$\begin{aligned} S_{11} &= 0.69 \angle -78^\circ & S_{12} &= 0.033 \angle 41.4^\circ \\ S_{21} &= 5.67 \angle 123^\circ & S_{22} &= 0.84 \angle -25^\circ \end{aligned}$$

Consider the four types of resistive loading shown in Fig. 3.3.9.

- 3.15** Prove that the maximum unilateral transducer power gain in (3.4.6) is obtained when $\Gamma_s = S_{11}^*$ and $\Gamma_L = S_{22}^*$.

- 3.16 (a)** Design a microwave transistor amplifier for $G_{TU,\max}$ using a BJT whose S parameters in a 50Ω system at $V_{CE} = 10$ V, $I_C = 20$ mA, and $f = 1$ GHz are

$$\begin{aligned} S_{11} &= 0.706 \angle -160^\circ \\ S_{12} &= 0 \\ S_{21} &= 5.01 \angle 85^\circ \\ S_{22} &= 0.508 \angle -20^\circ \end{aligned}$$

- (b)** Draw the constant-gain circles for $G_s = 2, 1, 0$, and -1 dB.

- 3.17** The scattering parameters of a GaAs FET in a 50Ω system are

$$\begin{aligned} S_{11} &= 2.3 \angle -135^\circ \\ S_{12} &= 0 \\ S_{21} &= 4 \angle 60^\circ \\ S_{22} &= 0.8 \angle -60^\circ \end{aligned}$$

- (a) Determine the unstable region in the Smith chart and construct the constant-gain circle for $G_s = 4$ dB.
 (b) Design the input matching network for $G_s = 4$ dB with the greatest degree of stability.
 (c) Draw the complete ac amplifier schematic.

3.18 A microwave amplifier is to be designed for $G_{TU,\max}$ using a transistor with

$$\begin{aligned} S_{11} &= 0.5 \angle 140^\circ & S_{12} &= 0 \\ S_{21} &= 5 \angle 45^\circ & S_{22} &= 0.6 \angle -95^\circ \end{aligned}$$

The S parameters were measured in a 50Ω system at $f = 900$ MHz, $V_{CE} = 15$ V, and $I_C = 15$ mA.

- (a) Determine $G_{TU,\max}$.
 (b) Design two different microstrip matching networks.
 (c) Draw the constant gain circle for $G_L = 1$ dB.
 (d) If the S parameters at 1 GHz are

$$\begin{aligned} S_{11} &= 0.48 \angle 137^\circ & S_{12} &= 0 \\ S_{21} &= 4.6 \angle 48^\circ & S_{22} &= 0.57 \angle -99^\circ \end{aligned}$$

calculate the gain G_T at 1 GHz for the designs in part (b).

3.19 (a) Verify the equations for a simultaneous conjugate match in (3.6.5) and (3.6.6).
 (b) Show that for small S_{12} , Γ_M and Γ_{ML} are close to S_{11}^* and S_{22}^* , respectively.

3.20 (a) Prove the identities

$$|C_1|^2 = |S_{11} - \Delta S_{22}^*|^2 = |S_{12}S_{21}|^2 + (1 - |S_{22}|^2)(|S_{11}|^2 - |\Delta|^2)$$

and

$$|C_2|^2 = |S_{22} - \Delta S_{11}^*|^2 = |S_{12}S_{21}|^2 + (1 - |S_{11}|^2)(|S_{22}|^2 - |\Delta|^2)$$

(b) Analyze the solutions to (3.6.5) and (3.6.6) when $|K| > 1$ with K negative.

3.21 Design a microwave transistor amplifier for $G_{T,\max}$ using a BJT whose S parameters in a 50Ω system at $V_{CE} = 10$ V, $I_C = 4$ mA, and $f = 750$ MHz are

$$\begin{aligned} S_{11} &= 0.277 \angle -59^\circ \\ S_{12} &= 0.078 \angle 93^\circ \\ S_{21} &= 1.92 \angle 64^\circ \\ S_{22} &= 0.848 \angle -31^\circ \end{aligned}$$

(This problem is based on a design given in Ref. [3.6].)

3.22 The output matching network shown in Fig. P3.22 was designed at 2 GHz for a simultaneous conjugate match.

- (a) Determine Γ_{ML} .
 (b) If the microstrip is alumina ($\epsilon_r = 10$) and $h = 30$ mils, determine the length of the 0.25λ line.

3.23 The matching network in Fig. P3.23 was designed for $\Gamma_{ML} = 0.718 \angle 103.9^\circ$. Determine Γ_x .

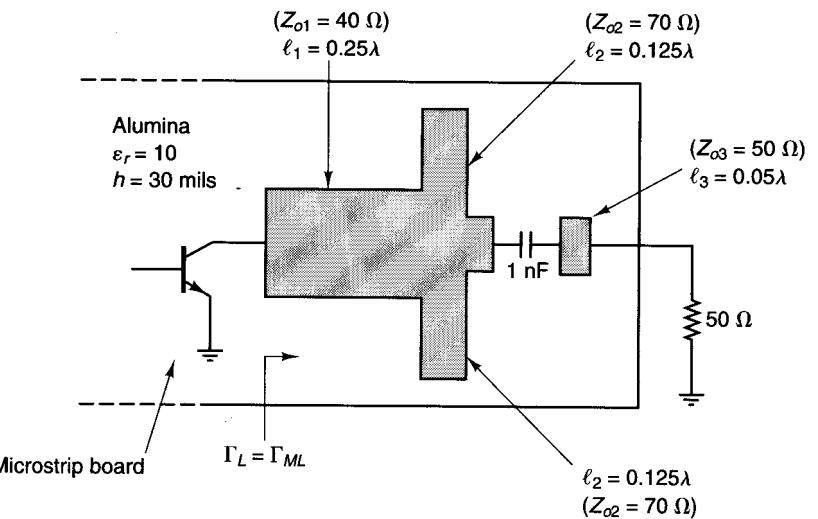


Figure P3.22

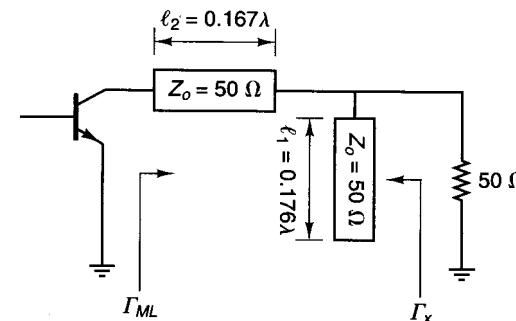


Figure P3.23

3.24 The S parameters from 3 GHz to 4 GHz of a BJT are

f (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
3	0.575 $\angle -173^\circ$	0.043 $\angle 25^\circ$	2.21 $\angle 48^\circ$	0.773 $\angle -58^\circ$
3.5	0.56 $\angle 180^\circ$	0.046 $\angle 25^\circ$	1.95 $\angle 37^\circ$	0.795 $\angle -64^\circ$
4	0.548 $\angle 173^\circ$	0.049 $\angle 24^\circ$	1.67 $\angle 29^\circ$	0.816 $\angle -71^\circ$

- (a) Can you design for a simultaneous conjugate match at 3.5 GHz?
 (b) Design an amplifier for $G_{T,\max}$ at 3.5 GHz.
 (b) Plot G_T in decibels versus frequency from 3 GHz to 4 GHz.

- 3.25** Design a microwave transistor amplifier at $f = 750$ MHz to have $G_p = 10$ dB using the BJT in Problem 3.21. For the value of Γ_L selected and with $\Gamma_s = \Gamma_{IN}^*$, determine $(VSWR)_{in}$ and $(VSWR)_{out}$.
(b) Determine the reflection coefficients for $G_{p,max}$ and show that they are identical to Γ_{Ms} and Γ_{ML} in Problem 3.21.

- 3.26** At 2 GHz, a GaAs FET has the following S parameters:

$$S_{11} = 0.7 \angle -65^\circ$$

$$S_{12} = 0.03 \angle 60^\circ$$

$$S_{21} = 3.2 \angle 110^\circ$$

$$S_{22} = 0.8 \angle -30^\circ$$

Determine the stability and design an amplifier with $G_p = 10$ dB. For the value of Γ_L selected and with $\Gamma_s = \Gamma_{IN}^*$, determine $(VSWR)_{in}$ and $(VSWR)_{out}$.

- 3.27** The S parameters of a transistor are

$$S_{11} = 0.78 \angle -102^\circ \quad S_{12} = 0.063 \angle 46^\circ$$

$$S_{21} = 2.43 \angle 84^\circ \quad S_{22} = 0.7 \angle -57^\circ$$

- (a)** Draw the $G_p = 10$ dB constant-gain circle.
(b) Select several values of Γ_L on the 10-dB gain circles. For each value of Γ_L selected, and with $\Gamma_s = \Gamma_{IN}^*$, determine $(VSWR)_{in}$ and $(VSWR)_{out}$.
(c) Draw the 15-dB, 20-dB, and 40-dB constant operating power-gain circle. Observe that as G_p becomes infinite, the power-gain circles approach the output stability circle.

- 3.28** Show that in a potentially unstable case when the operating power-gain circles and the output stability circle intersect the edge of the Smith chart, the points of intersection are identical (see Figs. 3.7.2 and 3.7.3).

- 3.29** The S parameters of a transistor are

$$S_{11} = 0.5 \angle 45^\circ \quad S_{12} = 0.4 \angle 145^\circ$$

$$S_{21} = 4 \angle 120^\circ \quad S_{22} = 0.4 \angle -40^\circ$$

- (a)** Draw the stability circles and show the stable regions.
(b) Calculate G_p if the transistor is used with $\Gamma_s = 0.2 \angle 145^\circ$ and $\Gamma_L = 0$.
(c) What is the value of the largest operating power gain that can be obtained?

- 3.30** Perform the analyses in Example 3.8.1 for the constant-gain circle $G_A = G_{A,max} - 2$ dB (i.e., $G_A = 7.66$ dB).

- 3.31** The parameters of a two-port network at 5 GHz are

$$S_{11} = 0.75 \angle -60^\circ \quad S_{12} = 0.3 \angle 70^\circ$$

$$S_{21} = 6 \angle 90^\circ \quad S_{22} = 0.5 \angle 60^\circ$$

- (a)** Verify that this two-port network is potentially unstable with $K > 1$ and $|A| > 1$.
(b) Draw the output stability circle.
(c) Calculate Γ_{ML} .
(d) Calculate $G_{p,min}$ and $G_{T,min}$.

- (e)** Draw the constant operating power-gain circles that are 2 dB, 5 dB, 10 dB, and 20 dB larger than $G_{p,min}$.
(f) Draw the input stability circle and calculate Γ_{Ms} .
(g) Determine $(VSWR)_{in}$ and $(VSWR)_{out}$ if the circuit is designed with $\Gamma_s = \Gamma_{Ms}$ and $\Gamma_L = \Gamma_{ML}$.

- 3.32** The S parameters of a GaAs FET at $f = 12$ GHz, $V_{DS} = 3.5$ V, and $I_D = 25$ mA are

$$S_{11} = 0.6 \angle 36^\circ \quad S_{12} = 0.14 \angle -85^\circ$$

$$S_{21} = 2.3 \angle -80^\circ \quad S_{22} = 0.15 \angle 45^\circ$$

- (a)** Determine $G_{p,max}$ and draw the constant G_p circle that is 1 dB less than $G_{p,max}$.
(b) Select several values of Γ_L on the $G_p = G_{p,max} - 1$ dB circle. For each Γ_L value, determine the values of Γ_s that lie on the constant $(VSWR)_{in} = 1.5$ circle, and draw the constant $(VSWR)_{in} = 1.5$ circle.
(c) Select several values of Γ_s on the $(VSWR)_{in} = 1.5$ circle. For each Γ_s value, calculate $(VSWR)_{out}$.

- 3.33 (a)** Map the $G_p = 10$ dB circle in Fig. 3.7.2 to the $\Gamma_s = \Gamma_{IN}^*$ plane.

- (b)** For $\Gamma_L = 0.1 \angle 97^\circ$, show the values of Γ_s that produce $(VSWR)_{in} = 2$.

- 3.34** Design a microwave amplifier using a BJT whose S parameters at 1 GHz are

$$S_{11} = 0.6 \angle -170^\circ \quad S_{12} = 0.03 \angle 50^\circ$$

$$S_{21} = 8 \angle 80^\circ \quad S_{22} = 0.45 \angle -30^\circ$$

Analyze the trade-offs between operating power gain, stability, and VSWRs.

- 3.35** Design a microwave amplifier using a GaAs FET whose S parameters at 6 GHz, $V_{DS} = 3$ V, and $I_{DS} = 20$ mA are

$$S_{11} = 0.6 \angle -170^\circ \quad S_{12} = 0.125 \angle -12^\circ$$

$$S_{21} = 3.1 \angle 25^\circ \quad S_{22} = 0.4 \angle -90^\circ$$

Analyze the trade-offs between available power gain, stability, and VSWRs.

- 3.36** The discrete amplifier in Fig. P3.36 was designed to operate at 500 MHz for $G_{T,max}$.

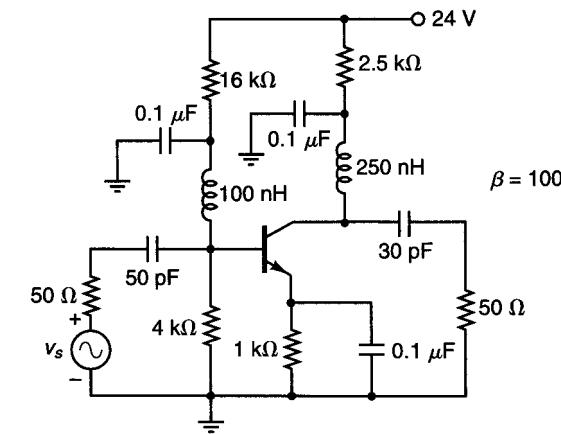


Figure P3.36

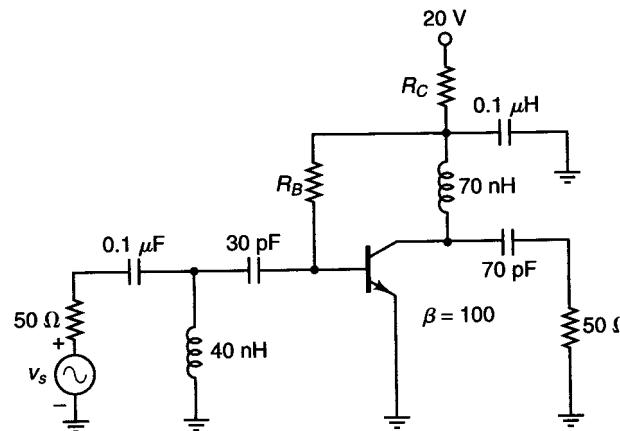


Figure P3.37

- (a) Draw the dc model.
 (b) Determine the approximate value of I_C and V_{CE} at the Q point.
 (c) Radio frequency coils could be connected in series with the $4\text{ k}\Omega$, $16\text{ k}\Omega$, and $2.5\text{ k}\Omega$ resistors. Are they necessary?
 (d) Draw the ac model.
 (e) Determine the value of Γ_s and Γ_L .
- 3.37** A 300-MHz low-noise amplifier is shown in Fig. P3.37. The dc bias circuit is recognized as the dc voltage feedback bias configuration in Fig. 3.9.2.
- (a) Draw the dc model.
 (b) Determine the values of R_L and R_B for a Q point at $10\text{ V}, 5\text{ mA}$.
 (c) Determine the values of Γ_s and Γ_L .
- 3.38** Design the bias circuit in Fig. 3.9.3 to have a Q point at $V_{CE} = 10\text{ V}$ and $I_C = 10\text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.75\text{ V}$.
- 3.39** Design the active bias network in Fig. 3.9.4 to set the quiescent point of transistor Q_2 at $V_{CE2} = 10\text{ V}$ and $I_{C2} = 10\text{ mA}$. Use $\beta = 100$ and $V_{BE} = 0.75\text{ V}$.
- 3.40** (a) Derive the stability factors for the dc bias networks in Fig. 3.9.2a.
 (b) Design the circuit for a quiescent point at $V_{CE} = 10\text{ V}$ and $I_C = 10\text{ mA}$. Use $h_{FE} = 50$ and $V_{CC} = 20\text{ V}$.
 (c) What happens to the quiescent point if h_{FE} changes from 50 to 100.
- 3.41** Design the dc bias network shown in Fig. 3.9.3 for $V_{CE} = 6\text{ V}$, $I_C = 1\text{ mA}$, and $S_i = 5$. Assume that $h_{FE} = 100$ and $I_{CBO} = 1\text{ }\mu\text{A}$ at 25°C . Calculate the resulting stability factors and find what happens to the operating point if the temperature increases to 75°C .
- 3.42** Design the active bias network in Fig. 3.9.8 to set the quiescent point of GaAs FET at $V_{DS} = 3\text{ V}$ and $I_{DS} = 10\text{ mA}$. The pinchoff voltage is -3 V , and the drain saturation current is 30 mA.

- [3.1] D. Woods, "Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in Terms of S Parameters," *IEEE Transactions on Circuits and Systems*, February 1976.
- [3.2] K. Kurokawa, "Power Waves and the Scattering Matrix," *IEEE Transactions on Microwave Theory and Techniques*, March 1965.
- [3.3] G. E. Bodway, "Two Port Power Flow Analysis Using Generalized Scattering Parameters," *Microwave Journal*, May 1967.
- [3.4] "Microwave Transistor Bias Considerations," Hewlett-Packard Application Note 944-1, April 1975.
- [3.5] G. D. Vendelin, "Five Basic Bias Designs for GaAs FET Amplifiers," *Microwaves & RF*, February 1978.
- [3.6] W. H. Froehner, "Quick Amplifier Design with Scattering Parameters," *Electronics*, October 1967.

REFERENCES

- [3.1] D. Woods, "Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in Terms of S Parameters," *IEEE Transactions on Circuits and Systems*, February 1976.

MATCHING NETWORKS AND SIGNAL FLOW GRAPHS

2.1 INTRODUCTION

The analysis of transmission-line problems and of matching circuits at microwave frequencies can be cumbersome in analytical form. The Smith chart provides a very useful graphical aid to the analysis of these problems. The Smith chart is basically a plot of all passive impedances in a reflection coefficient chart of unit radius. The reading accuracy from the Smith chart is sufficient for most practical microwave transistor amplifier design problems.

Matching circuits that provide optimum performance in a microwave amplifier can be easily and quickly designed using the normalized impedance and admittance Smith chart. The Smith chart is also used to present the frequency dependence of scattering parameters and other amplifier characteristics.

The characteristics of microstrip transmission lines are presented in this chapter. The mode of propagation in a microstrip line is assumed to be quasi-transverse electromagnetic. Although radiation losses in a microstrip line can be severe, the use of a thin material, having a high dielectric constant, between the top strip conductor and the ground plane of a microstrip line reduces the radiation losses to a minimum.

Microstrip lines find extensive use as passive circuit elements and as a medium in which the complete microwave amplifier can be built. The interconnection features of the microstrip line are unsurpassed. Transistors in chip or packaged form can be easily attached to the strip conductors of the microstrip line. Some practical circuit construction techniques using microstrips are presented.

In this chapter signal flow graphs are discussed. The description of two-port networks in terms of S parameters permits the use of signal flow graphs in the analysis of microwave amplifiers. Power-gain expressions, as well as other relations, are derived in Section 2.6 using signal flow graphs. In Section 2.7 the power-gain expressions are derived using direct manipulations of the traveling wave relations.

2.2 THE SMITH CHART

The Smith chart is the representation in the reflection coefficient plane, called the Γ plane, of the relation

$$\Gamma = \frac{Z - Z_o}{Z + Z_o} \quad (2.2.1)$$

for all values of Z , such that $\text{Re}[Z] \geq 0$. Z_o is the characteristic impedance of the transmission line or a reference impedance value. Defining the normalized impedance z as

$$z = \frac{Z}{Z_o} = \frac{R + jX}{Z_o} = r + jx$$

we can write (2.2.1) in the form

$$\Gamma = \frac{z - 1}{z + 1} \quad (2.2.2)$$

Figure 2.2.1a illustrates the properties of the transformation (2.2.2) for some values of z . For example, if $Z = 50 \Omega$ and $Z_o = 50 \Omega$, then $z = 1$ and $\Gamma = 0$. That is, the point $z = 1$ in the normalized z plane maps into the origin of the Γ plane. From (2.2.2) it also follows that the point $z = 0$ maps into the point $\Gamma = -1$ (i.e., $U = -1$ and $V = 0$).

Next we consider the mapping of normalized impedances having constant real and imaginary parts. For example, for $z = 1 + jx$ the corresponding values of Γ are

$$\Gamma = \frac{jx}{2 + jx} \quad \text{or} \quad \left\{ \begin{array}{ll} \Gamma = 0 & \text{for } x = 0 \\ \Gamma = 0.447 \angle 63.43^\circ & \text{for } x = \pm 1 \\ \Gamma = 0.707 \angle 45^\circ & \text{for } x = \pm 2 \\ \text{etc.} & \end{array} \right.$$

The mapping of the various points along $z = 1 + jx$ is shown in Fig. 2.2.1a. In fact, we will show that the mapping is a circle of radius $1/2$ centered at $U = 1/2$ and $V = 0$. Since this circle represents the mapping of all points with $r = 1$, the circle is known as the constant resistance circle for $r = 1$.

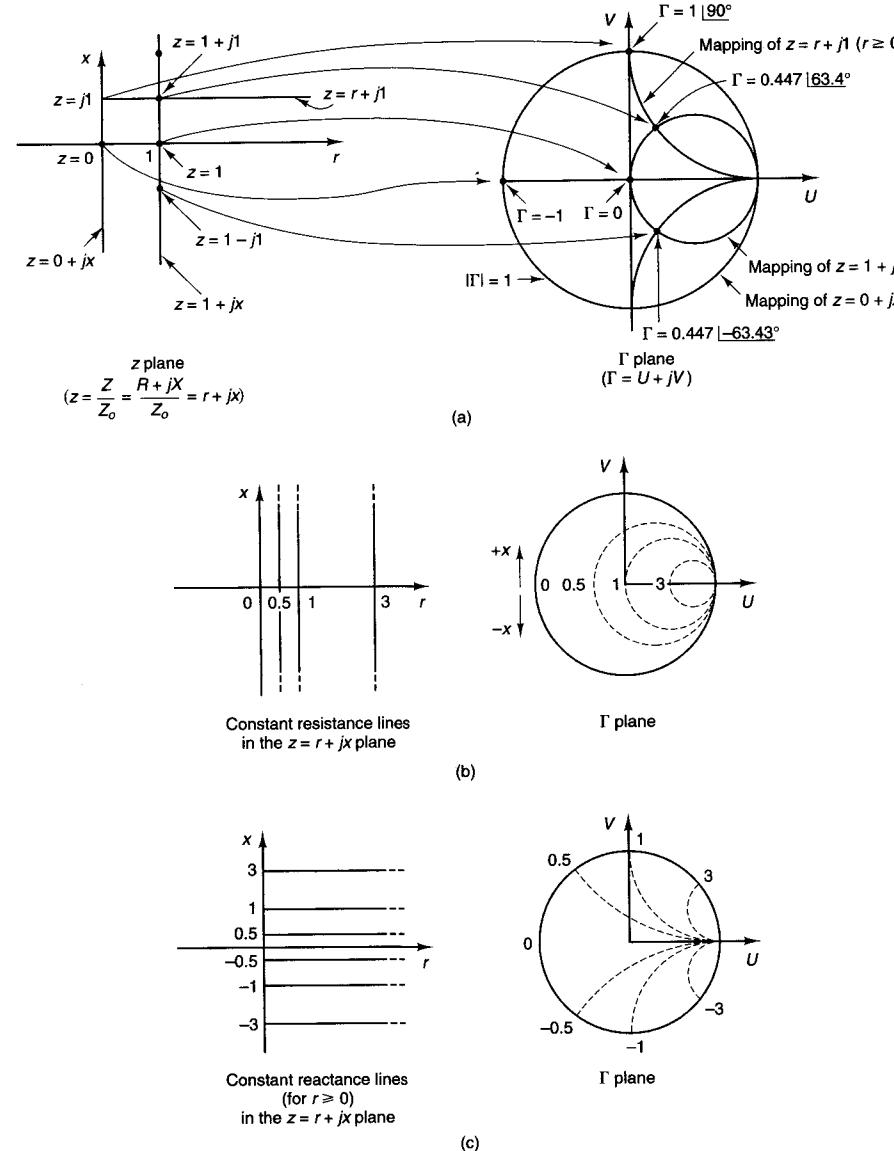


Figure 2.2.1 Development of the Smith chart.

From (2.2.2) it follows that the imaginary axis (i.e., $z = 0 + jx$) maps into the unit circle given by $|\Gamma| = 1$. Therefore, all passive impedances—that is, impedances having $r \geq 0$ —map inside the unit circle (i.e., $|\Gamma| \leq 1$) in the Γ plane.

Figure 2.2.1a also shows the mapping of $z = r + j1$ ($r \geq 0$) (i.e., a straight line having a constant imaginary value of one). For $z = r + j1$ ($r \geq 0$) it follows that the corresponding values of Γ lie along a portion of a circle inside $|\Gamma| \leq 1$

having radius 1, centered at $U = 1$ and $V = 1$. This circle is known as the constant reactance circle for $x = 1$. The portion of the circle outside the Smith chart corresponds to the mapping of $z = r + j1$ for $r < 0$ (i.e., for negative resistances).

The transformation (2.2.2) can be analyzed in general as follows. Let

$$\Gamma = U + jV = \frac{(r - 1) + jx}{(r + 1) + jx}$$

Then rationalize and separate the real and imaginary parts to obtain

$$U = \frac{r^2 - 1 + x^2}{(r + 1)^2 + x^2} \quad (2.2.3)$$

and

$$V = \frac{2x}{(r + 1)^2 + x^2} \quad (2.2.4)$$

Eliminating x from (2.2.3) and (2.2.4) results in

$$\left(U - \frac{r}{r + 1} \right)^2 + V^2 = \left(\frac{1}{r + 1} \right)^2$$

which is the equation of a family of circles centered at $U = r/(r + 1)$ and $V = 0$, with radii $1/(r + 1)$. The constant resistance circles for $r = 0, 0.5, 1$, and 3 are shown in Fig. 2.2.1b. Observe that in Fig. 2.2.1b the constant resistance circles in the Γ plane are labeled according to the constant resistance values (i.e., the value of r in the z plane).

Eliminating r from (2.2.3) and (2.2.4) results in

$$(U - 1)^2 + \left(V - \frac{1}{x} \right)^2 = \left(\frac{1}{x} \right)^2$$

which is the equation of a family of circles centered at $U = 1$ and $V = 1/x$, with radii $1/x$. The portion of constant-reactance circles for $x = -3, -1, -0.5, 0, 1, 0.5$, and 3 (with $r \geq 0$) is shown in Fig. 2.2.1c.

There is a one-to-one correspondence between points in the z plane and points in the Γ plane. The plot of the constant-resistance and constant-reactance circles for all values of z such that $\text{Re}[z] \geq 0$ in a graph is known as the *Smith chart*. The Smith chart is shown in Fig. 2.2.2. Observe that the upper half of the chart represents normalized impedances having a positive reactance (i.e., x is positive) and the lower half represents negative reactances (i.e., x is negative).

The Smith chart can also be used as an admittance chart. The appropriate transformation in this case is

$$\Gamma' = \frac{y - 1}{y + 1}$$

NAME	TITLE	DWG. NO.
SMITH CHART FORM 82-BSPR (9-66)	KAY ELECTRIC COMPANY, PINE BROOK, N.J. ©1966 PRINTED IN U.S.A.	DATE

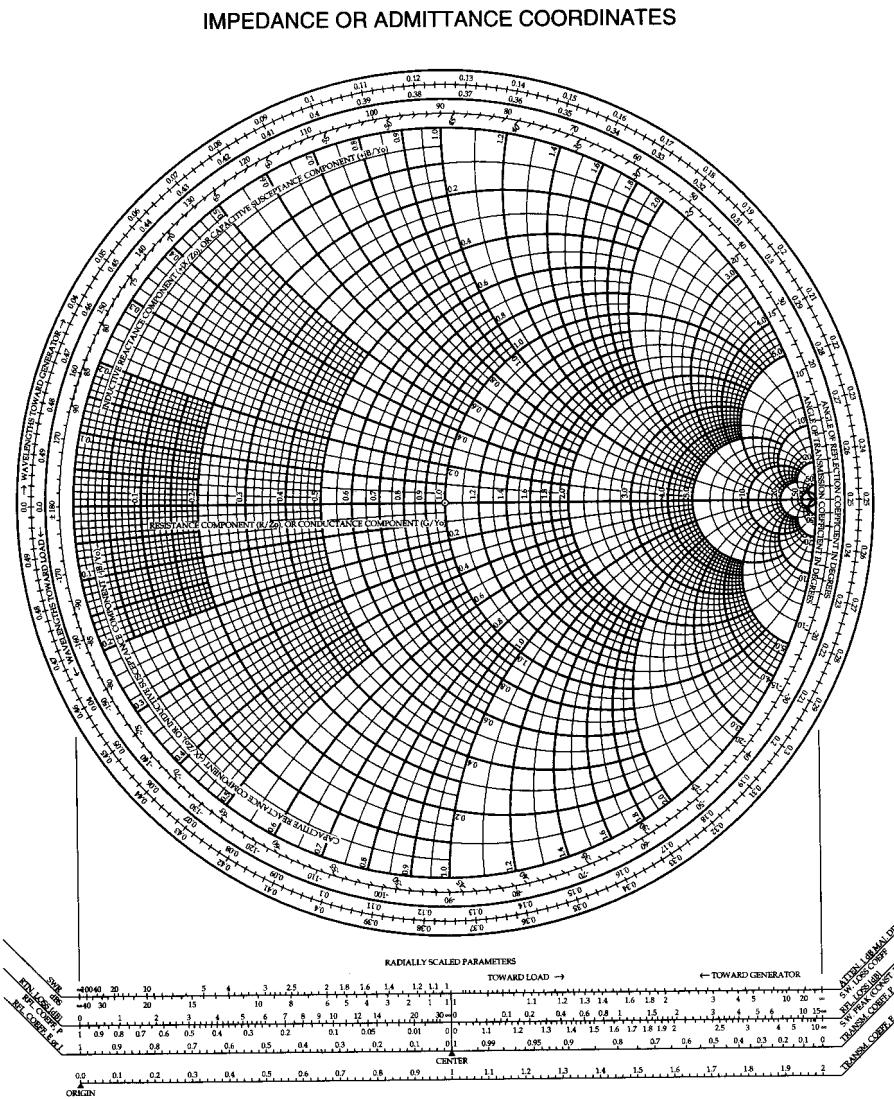


Figure 2.2.2 The Smith chart. (Reproduced with permission of Kay Electric Co., Pine Brook, N.J.)

Sec. 2.2 The Smith Chart

where the normalized admittance is $y = Y/Y_o$. Y_o is the characteristic admittance of the transmission line or a reference admittance value. Since $z = 1/y$, it follows that Γ and Γ' are related by $\Gamma' = -\Gamma$ or $\Gamma' = \Gamma e^{j\pi}$.

In the admittance chart, since

$$y = \frac{Y}{Y_e} = \frac{G + jB}{Y_e} = g + jb$$

the previous constant-resistance (r) circles become constant-conductance (g) circles and the constant-reactance (x) circles become constant-susceptance (b) circles. Observe that the upper half of the chart represents normalized admittances having a positive susceptance (i.e., b is positive) and the lower half represents negative susceptances (i.e., b is negative).

When needed for clarity, we will call a Smith chart used as an impedance chart a "Z Smith chart," and a Smith chart used as an admittance chart a "Y Smith chart."

Example 2.2.1

Locate in the Smith chart the following normalized impedances and admittances:

$$z_1 = 1 + j1, \quad z_2 = 0.4 + j0.5, \quad z_3 = 3 - j3, \quad z_4 = 0.2 - j0.6, \quad z_5 = 0$$

$$y_1 = 1 + j1, \quad y_2 = 0.4 + j0.5, \quad y_3 = 2 - j1.4, \quad y_4 = 0.5 - j0.2, \quad y_5 = \infty$$

Solution. The values of z 's and y 's are shown in Fig. 2.2.3. The Smith chart in Fig. 2.2.3a is obviously used as a Z Smith chart, and that in Fig. 2.2.3b as a Y Smith chart.

The conversion of a normalized impedance to a normalized admittance can be done easily in the Smith chart. Since from (2.2.2)

$$z = \frac{1 + \Gamma}{1 - \Gamma}$$

and

$$y = \frac{1}{z} = \frac{1 - \Pi}{1 + \Pi}$$

we observe that rotating Γ by $e^{i\pi}$ we obtain a new value of z (denoted by z') given by

$$z' = \frac{1 + \Gamma e^{j\pi}}{1 - \Gamma e^{j\pi}} = \frac{1 - \Gamma}{1 + \Gamma}$$

which is identical to the value of the admittance y . In other words, the numerical value of the impedance z' is identical to the value of the admittance $y = 1/z$.

Example 2.2.2

Find y for $z = 1 + j1$ using the Smith chart.

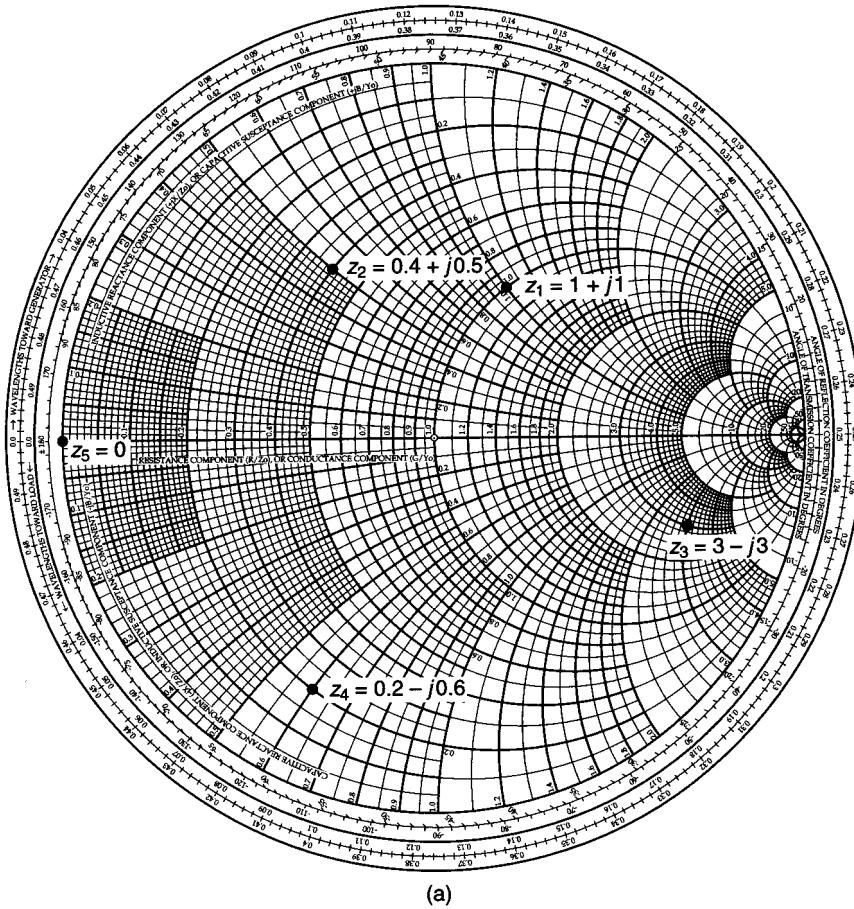


Figure 2.2.3 (a) Values of z in the Z Smith chart; (b) values of y in the Y Smith chart.

Solution. Locate the point $z = 1 + j1$ in Fig. 2.2.4. Associated with z there is a value of Γ . Rotating Γ by 180° (i.e., $e^{j\pi}$) results in the point z' , whose numerical value is that of y . The value of y is read as $0.5 - j0.5$, which of course agrees with

$$y = \frac{1}{z} = \frac{1}{1 + j1} = 0.5 - j0.5$$

Impedances having a negative real part will have a reflection coefficient whose magnitude is greater than 1. These impedances, therefore, map outside the Smith chart. Figure 2.2.5 shows a chart (known as the *compressed Smith chart*) that includes the Smith chart (i.e., $|\Gamma| \leq 1$) plus a portion of the negative impedance region.

An alternative way of handling negative resistances (i.e., $|z| > 1$) is to plot in the Smith chart $1/\Gamma^*$ and take the values of the resistance circles as being negative and the reactance circles as labeled.

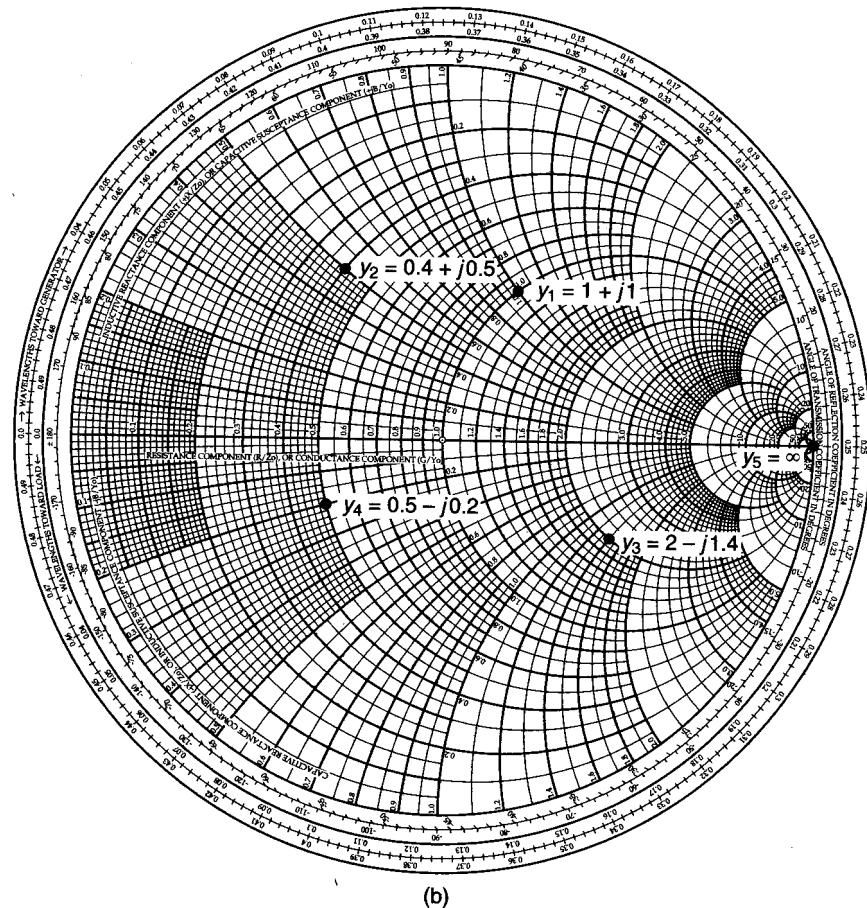


Figure 2.2.3 Continued

Example 2.2.3

Find the impedance whose reflection coefficient is $2.236|26.56^\circ$.

Solution. If we plot in the Smith chart shown in Fig. 2.2.6 the quantity

$$\frac{1}{\Gamma^*} = 0.447|26.56^\circ$$

the resulting z is $-2 + j1$. Of course, from (2.2.2),

$$\Gamma = \frac{-2 + j1 - 1}{-2 + j1 + 1} = 2.236|26.56^\circ$$

The use of the Smith chart in a transmission-line calculation follows from (1.3.35), (1.3.38), and (1.3.39). With $z = Z_L/Z_o$, we can conveniently write these equations in the form

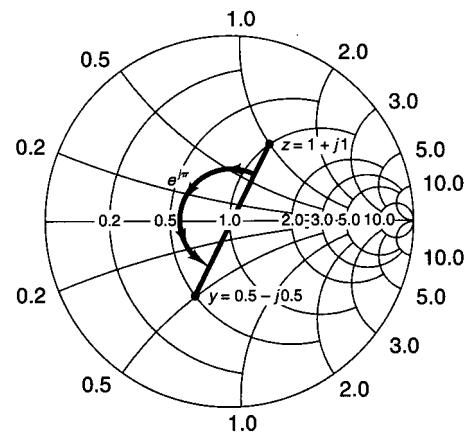


Figure 2.2.4 Conversion of z to y in the Smith chart.

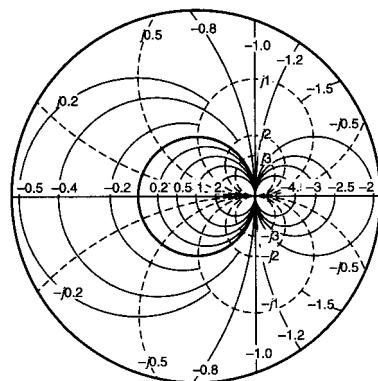


Figure 2.2.5 The compressed Smith chart. (From Ref. [1.1]; courtesy of Hewlett-Packard.)

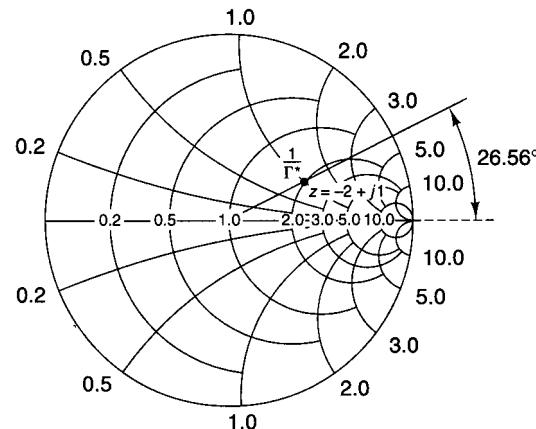


Figure 2.2.6 Negative resistances in the Smith chart.

$$\Gamma_0 = \frac{z - 1}{z + 1} \quad (2.2.5)$$

$$\Gamma_{IN}(d) = \Gamma_0 e^{-j2\beta d} \quad (2.2.6)$$

$$z_{IN}(d) = \frac{1 + \Gamma_{IN}(d)}{1 - \Gamma_{IN}(d)} \quad (2.2.7)$$

A typical transmission-line input impedance calculation involves the following steps:

1. Locate Γ_0 in the Z Smith chart for a given $z = Z_L/Z_o$ [i.e., (2.2.5)].
2. Rotate Γ_0 by $-2\beta d$ to obtain $\Gamma_{IN}(d)$ [i.e., (2.2.6)]. Observe that the rotation is along a vector of constant magnitude—namely, $|\Gamma_0| = |\Gamma_{IN}(d)|$.
3. Read the value of the normalized $z_{IN}(d)$ associated with $\Gamma_{IN}(d)$ [i.e., (2.2.7)].

Example 2.2.4

Find the input impedance, the load reflection coefficient, and the VSWR in a transmission line having an electrical length of 45° , characteristic impedance of 50Ω , and terminated in a load $Z_L = 50 + j50 \Omega$.

Solution. The transmission line is shown in Fig. 2.2.7a, where $z_L = Z_L/Z_o = 1 + j1$ and $\beta d = 2\pi d/\lambda = \pi/4$ or $d = \lambda/8 = 0.125\lambda$. In Fig. 2.2.7b, the point $z_L = 1 + j1$ is located and the vector representing Γ_0 drawn. To find Z_{IN} , we rotate along a constant $|\Gamma|$ radius a distance of -90° (i.e., $-2\beta d$)—that is, from 63.4° to -26.6° . The input impedance is read directly from the Smith chart as $z_{IN} = 2 - j1$ or $Z_{IN}(\lambda/8) = 100 - j50 \Omega$.

The previous calculations can also be made using the wavelength scales on the Smith chart. The input impedance is at a distance $d = 0.125\lambda$ from the load. From Fig. 2.2.7b, at z_L we read from the “wavelengths toward generator” scale a value of 0.162λ . This is an arbitrary value assigned to the load location. Next we add 0.125λ to obtain $0.162\lambda + 0.125\lambda = 0.287\lambda$. That is, we rotate toward the generator a distance $d = 0.125\lambda$ along a constant $|\Gamma|$ circle to reach the input of the line, which is found to be located at 0.287λ . Hence z_{IN} is read from the Smith chart to be $z = 2 - j1$ or $100 - j50 \Omega$.

The magnitude and phase of Γ_0 are read as indicated in Fig. 2.2.7b. Observe the linear scale for the magnitude of the reflection coefficient. The distance from the origin to z_L can be measured with a ruler or compass and superimposed on the linear scale. The reading of Γ_0 gives $\Gamma_0 = 0.447 |63.4^\circ|$.

Finally, the VSWR can be calculated from (1.3.44), or the distance from the origin to z_L can be measured and superimposed on the VSWR scale. The value obtained is 2.62. It can also be shown that the value of the maximum resistance in the line is numerically equal to the VSWR. This value is indicated in Fig. 2.2.7b as $VSWR = r_{max} = 2.62$.

Example 2.2.5

(a) Determine the length l of the 50Ω short-circuited transmission line shown in Fig. 2.2.8a so that the input impedance is $Z_{IN}(l) = j100\Omega$.

(b) Determine the length l of the 50Ω open-circuited transmission line shown in Fig. 2.2.8b so that the input impedance is $Z_{IN}(l) = j100\Omega$.

Solution. (a) In the short-circuited transmission line, $z_L = 0$. From Fig. 2.2.8a, the length l required to transform the load impedance $z_L = 0$ to the input impedance $z_{IN}(l) = j100/50 = j2\Omega$ is $l = 0.176\lambda$. Observe that in a short-circuited line the motion is along the edge of the chart (since $|\Gamma| = 1$ in a short-circuited line).

The length could have been calculated using (1.3.45). That is,

$$Z_{IN}(l) = j100 = j50 \tan \beta l$$

which gives $\tan \beta l = 2$ or $\beta l = 63.43^\circ = 0.352\pi$. Then

$$l = \frac{0.352\pi\lambda}{2\pi} = 0.176\lambda$$

(b) In the open-circuited transmission line, $z_L = \infty$. Therefore, from Fig. 2.2.8b the length l is 0.426λ [i.e., $(0.5\lambda - 0.25\lambda) + 0.176\lambda = 0.426\lambda$].

In many cases it is convenient to make transmission-line calculations using the Y Smith chart. The following two examples illustrate some transmission-line calculations using the Y Smith chart.

Example 2.2.6

(a) Determine the input admittance of a short-circuited transmission line having a length of $\lambda/8$ and $Y_o = 1/Z_o = 20\text{ mS}$.

(b) Determine the input admittance of an open-circuited transmission line having a length of $\lambda/8$ and $Y_o = 1/Z_o = 20\text{ mS}$.

Solution. (a) For the short-circuited line, the load admittance is $y_L = \infty$. Plotting y_L in the Y Smith chart shown in Fig. 2.2.9a and rotating along the constant gamma circle $|\Gamma| = 1$ a distance $l = \lambda/8$, we obtain $y_{IN}(l) = -j$ or

$$Y_{IN}(l) = y_{IN}(l)Y_o = -j(20 \times 10^{-3}) = -j20\text{ mS}$$

The input impedance is $Z_{IN}(l) = 1/Y_{IN}(l) = j50\Omega$.

(b) In the open-circuited line, the load admittance is $y_L = 0$. Therefore, as shown in Fig. 2.2.9b, at $l = \lambda/8$ we obtain $y_{IN}(l) = j$ or

$$Y_{IN}(l) = y_{IN}(l)Y_o = j(20 \times 10^{-3}) = j20\text{ mS}$$

The input impedance is $Z_{IN}(l) = 1/Y_{IN}(l) = -j50\Omega$.

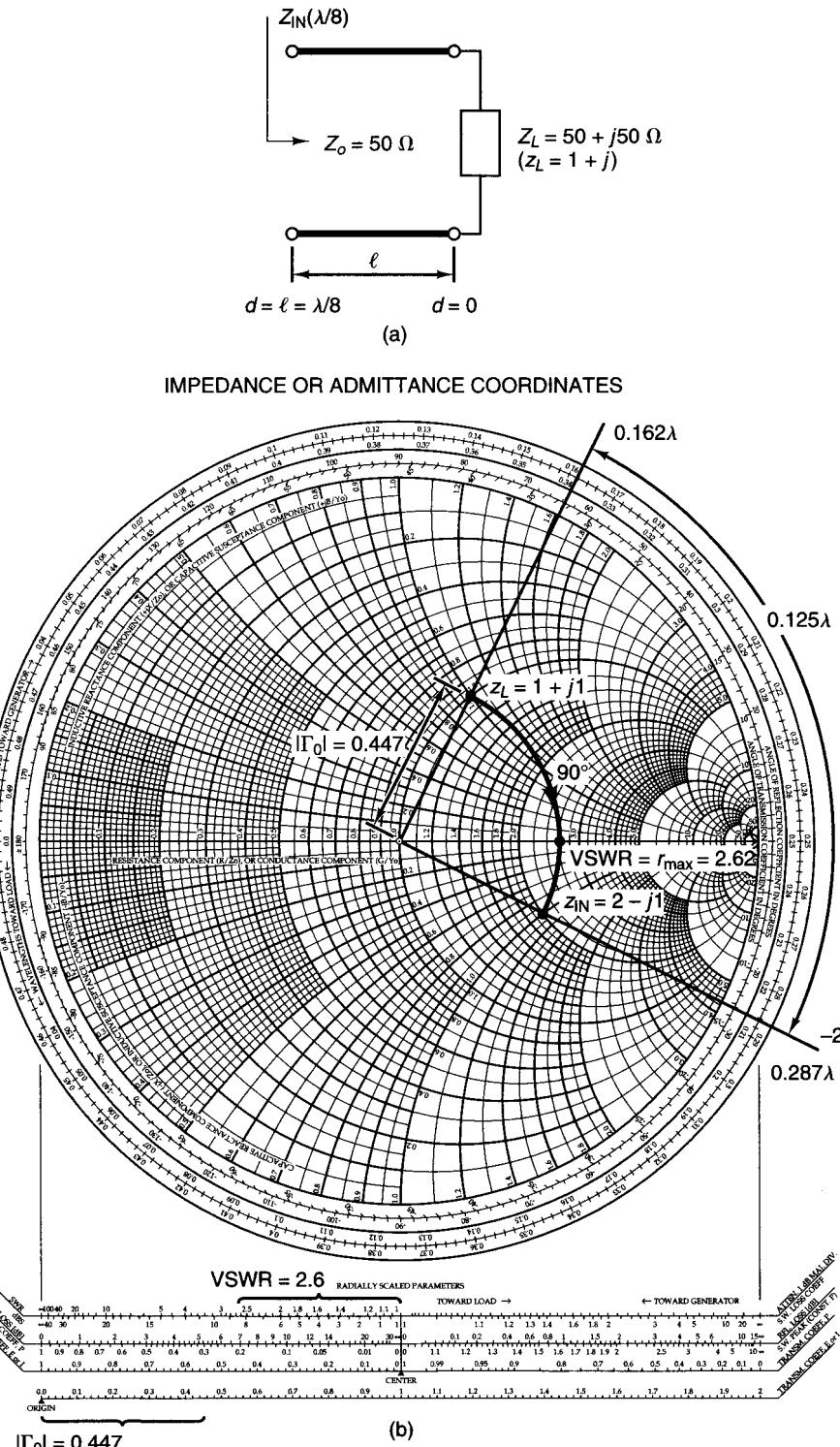


Figure 2.2.7 Typical transmission-line calculation using the Smith chart.

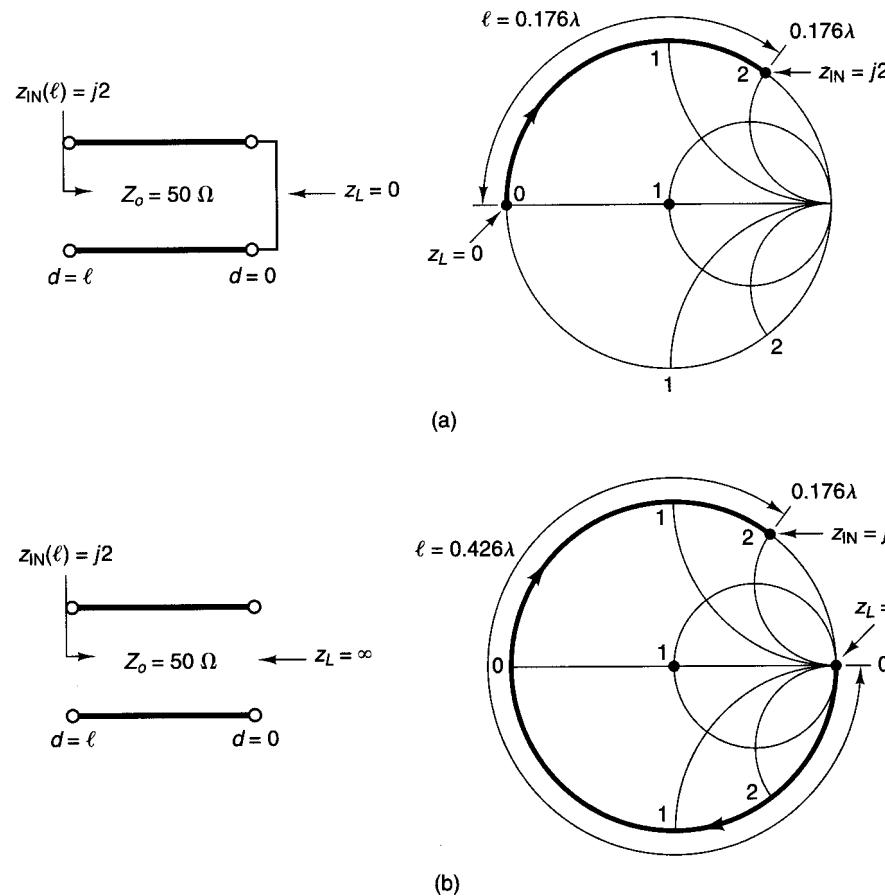


Figure 2.2.8 Circuit diagrams for Example 2.2.5.

Example 2.2.7

Solve Example 2.2.4 using a Y Smith chart.

Solution. Referring to Fig. 2.2.7a, the admittance associated with the load $z_L = Z_L/Z_o = 1 + j1$ can be obtained by rotating Γ_0 by 180° in Fig. 2.2.7b to obtain $y_L = 0.5 - j0.5$. The load y_L is shown in Fig. 2.2.10. At y_L , the “wavelengths toward generator” scale reads 0.412λ . Adding 0.125λ to 0.412λ results in the input being located at 0.537λ . Since the wavelength scale repeats every $\lambda/2$, it follows that 0.537λ is equivalent to 0.037λ . The motion from y_L to $y_{IN}(\lambda/8)$ along a constant $|\Gamma|$ circle is shown in Fig. 2.2.10. The value of $y_{IN}(l)$ is read from the Y Smith chart to be $0.4 + j0.2$, or $Y_{IN}(l) = (8 + j4) \text{ mS}$.

Comparing with Example 2.2.4, we observe that $y_{IN}(l) = 0.4 + j0.2$ is the admittance associated with $z_{IN}(l) = 1/y_{IN}(l) = 2 - j1$, and therefore $Z_{IN}(l) = 100 - j50 \Omega$, as expected.

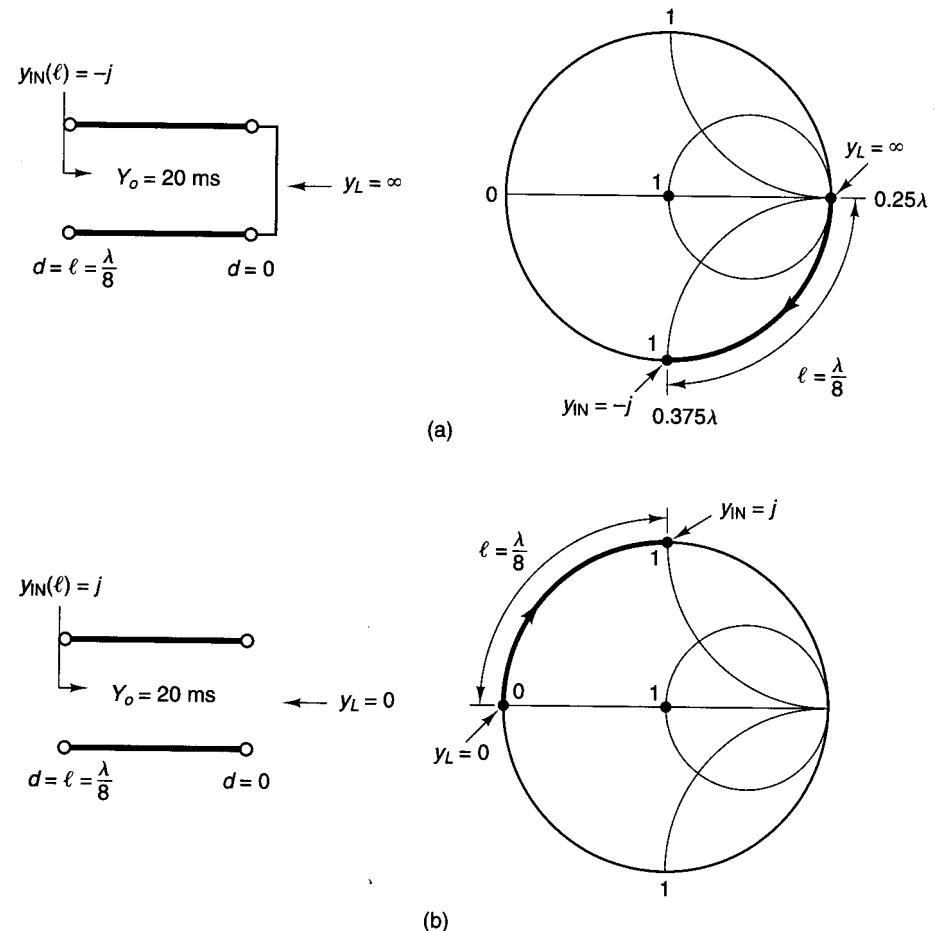


Figure 2.2.9 Circuit diagrams for Example 2.2.6.

2.3 THE NORMALIZED IMPEDANCE AND ADMITTANCE SMITH CHART

The impedance-to-admittance conversion can also be obtained by superimposing two Smith charts and rotating one of the charts by 180° . The rotated chart represents admittances and the other chart represents impedances. The superposition of the original and the rotated chart is known as the *normalized impedance and admittance coordinates Smith chart*. We will refer to this Smith chart as the *ZY Smith chart*. The *ZY Smith chart* is shown in Fig. 2.3.1, where the impedance values are shown in red and the admittance values in green. (See Fig. 2.3.1 in color on the inside cover of this book.)

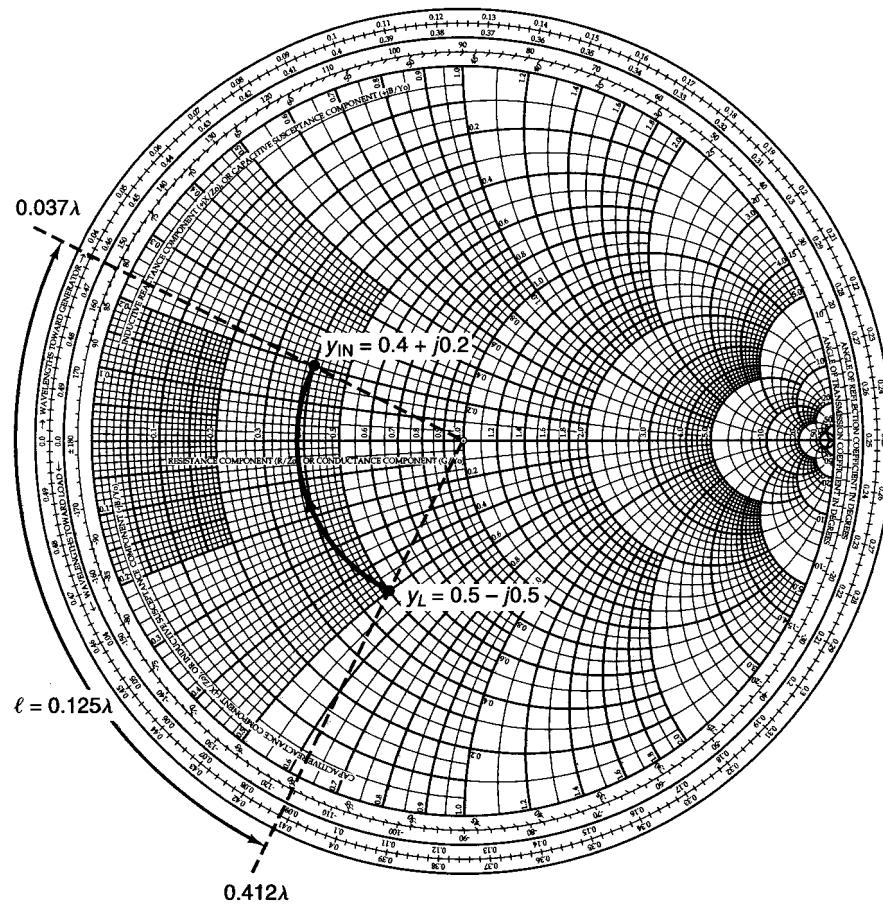


Figure 2.2.10 Solution in the Y Smith chart to the problem in Example 2.2.4.

On the left edge of the ZY Smith chart, one finds the symbols $+X_s$ and $-X_s$ to remind us that when using impedances values (red curves), the reactances are positive in the upper half of the chart and negative in the lower half. On the right edge of the ZY Smith chart, one finds $+B_p$ and $-B_p$ to remind us that when using the admittance values (i.e., green curves), the susceptances are negative in the upper half of the chart and positive in the lower half.

In the ZY Smith chart, for a given value of z the associated value of y is read directly from the admittance coordinates (shown in green), and vice versa.

Example 2.3.1

Find y for $z = 1 + j1$ using the ZY Smith chart.

NAME	TITLE	DWG. NO.
SMITH CHART FORM ZY-01-N	ANALOG INSTRUMENTS COMPANY, NEW PROVIDENCE, N.J. 07974	DATE

NORMALIZED IMPEDANCE AND ADMITTANCE COORDINATES

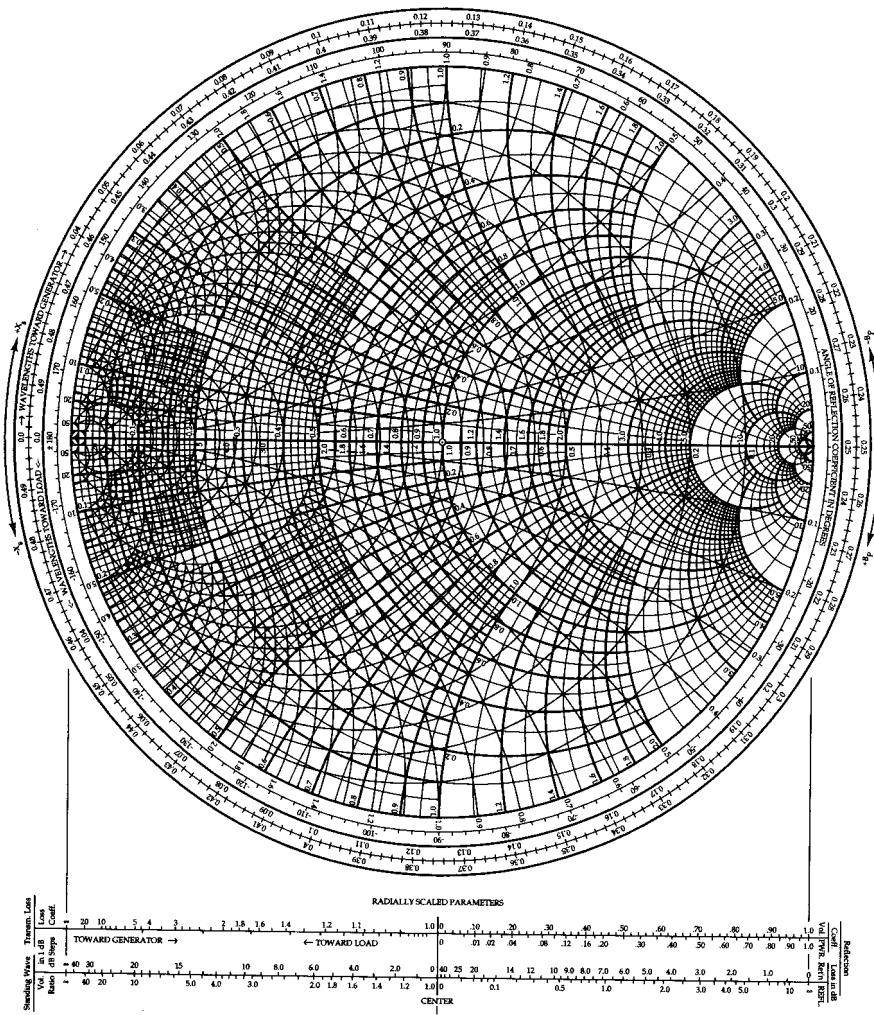


Figure 2.3.1 The normalized impedance and admittance coordinates Smith chart. (Reproduced with permission of Analog Instruments Co., New Providence, N.J.)

Solution. We can locate in the ZY Smith chart in Fig. 2.3.1 the point $z = 1 + j1$ (red curves) and read directly from the green curves the value $y = 0.5 - j0.5$.

The various Smith charts can be used to represent the frequency response of circuits, or from the frequency response of a circuit in the Smith chart an equivalent circuit model can be developed. The simplest frequency responses that can be represented in the Smith chart are those of a series RL circuit, a series RC circuit, a parallel RL circuit, and a parallel RC circuit.

In a series RL circuit, the normalized impedance is

$$z = \frac{Z}{Z_o} = \frac{R + j\omega L}{Z_o} = r + jx$$

As the frequency increases from f_a to f_b , the plot of the normalized impedance z follows a constant r circle with the reactance increasing linearly as a function of frequency, such as the typical plot for the series RL circuit shown in Fig. 2.3.2a, where $r = 0.2 \Omega$ and the reactance increases from $j0.24$ at f_a to $j0.5$ at f_b . In a series RL circuit, the reactance must change linearly with frequency.

The effect of adding a parallel capacitor to the series RL circuit is also shown in Fig. 2.3.2a. It is observed that the admittance of the capacitor (i.e., $y = jb = j\omega C/Y_o$) adds to the admittance of the series RL circuit. For the series RL circuit with a capacitor C in parallel in Fig. 2.3.2a, it is seen that at f_b the susceptance of the capacitor (i.e., $j0.9$) adds to the susceptance of the series RL circuit (i.e., $-j1.7$) to produce a series susceptance of $-j0.8$. Hence, the susceptance of the capacitor is $j0.9$ and at f_b the admittance of the circuit is $y = 0.7 - j0.8$. At f_a , the susceptance of the capacitor is such that the admittance of the circuit is $y = 2 - j1.4$.

In Fig. 2.3.2a, typical impedance plots for a series RC circuit and a series RC circuit with an inductor in parallel are also shown. In Fig. 2.3.2b, the admittance plots of parallel RL and parallel RC circuits are shown, as well as the effects of adding a series C to the parallel RL circuit and a series L to the parallel RC circuit.

Example 2.3.2

The frequency response of the normalized impedance (with $Z_o = 50 \Omega$) of a one-port network is shown in Fig. 2.3.3a as the frequency varies from 500 MHz to 1 GHz. Determine an equivalent circuit for the one-port network and the element values.

Solution. Since the frequency response follows a constant resistance circle of $r = 0.4$ and the reactance increases linearly with frequency, it follows that a series RL circuit simulates the behavior shown in Fig. 2.3.3a.

The value of R , with $Z_o = 50 \Omega$, is $R = rZ_o = 0.4(50) = 20 \Omega$. The value of L follows from

$$\frac{j\omega_b L}{Z_o} - \frac{j\omega_a L}{Z_o} = j0.4 - j0.2$$

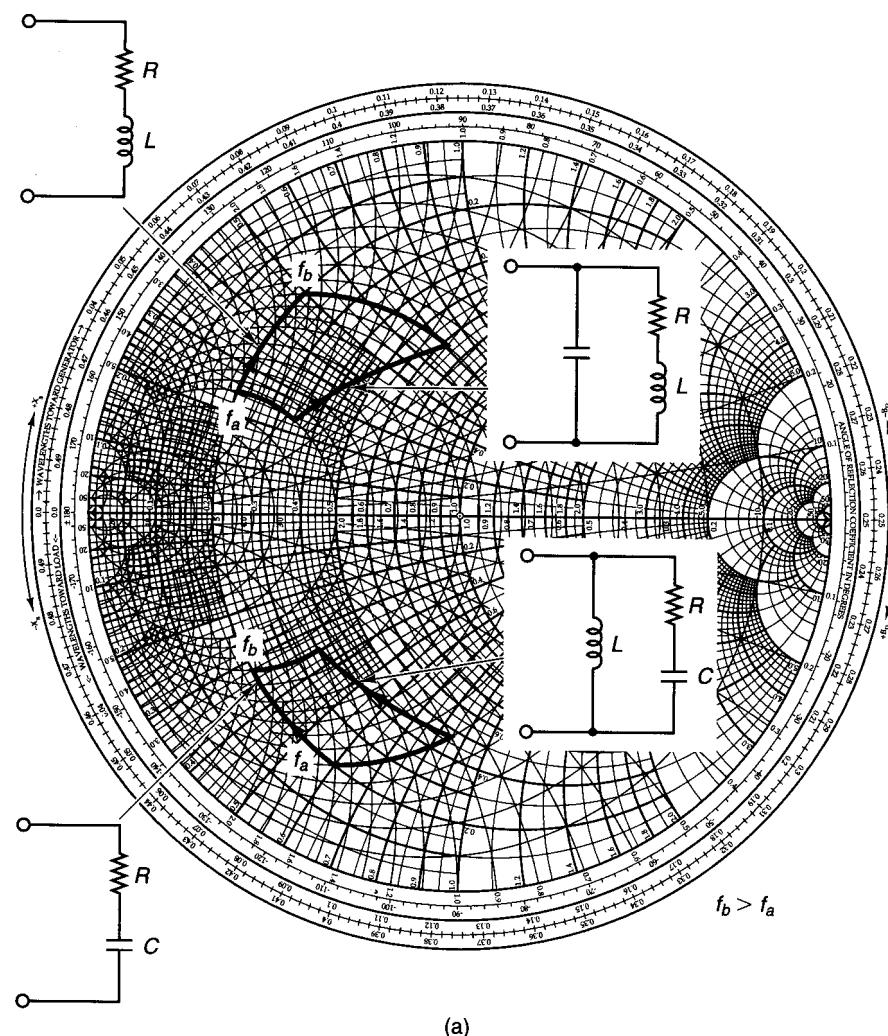


Figure 2.3.2 Characteristics of some networks in the ZY Smith chart.

or

$$L = 50 \frac{(0.4 - 0.2)}{(\omega_b - \omega_a)} = 50 \frac{0.2}{2\pi(1 \times 10^9 - 500 \times 10^6)} = 3.18 \text{ nH}$$

A typical plot of S_{11} for a transistor in the common-emitter configuration is shown in Fig. 1.10.2. It is observed that S_{11} for this transistor in chip form follows a constant-resistance circle, with a capacitive reactance at the lower frequencies and an inductive reactance at the higher frequencies. The equivalent

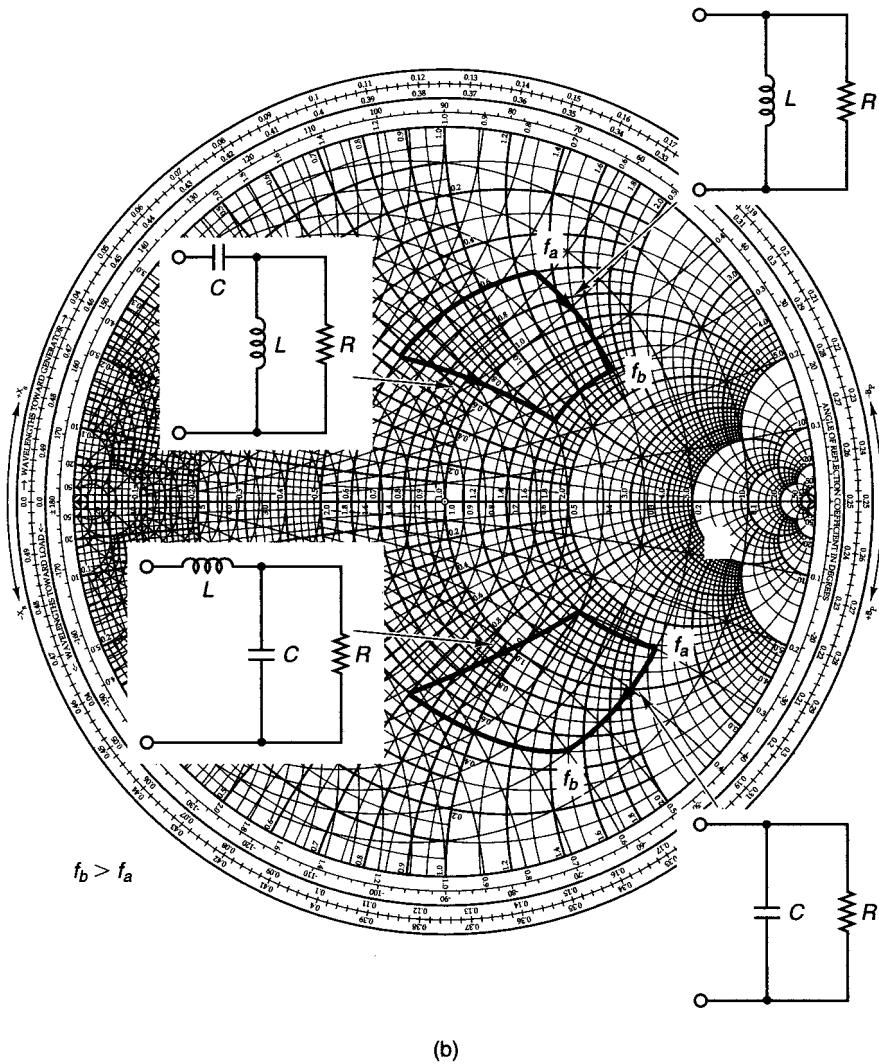


Figure 2.3.2 Continued

circuit for this transistor in chip form which exhibits the behavior of S_{11} is shown in Fig. 2.3.4a. The resistance R represents the base-to-emitter resistance plus any contact resistance. The capacitance C is due to the junction capacitance from base to emitter. The inductance L is due to the reflection properties of a transistor where the emitter resistance, when $h_{fe}(\omega)$ is complex, produces an inductive reactance across the base-to-emitter terminals.

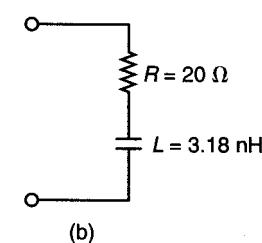
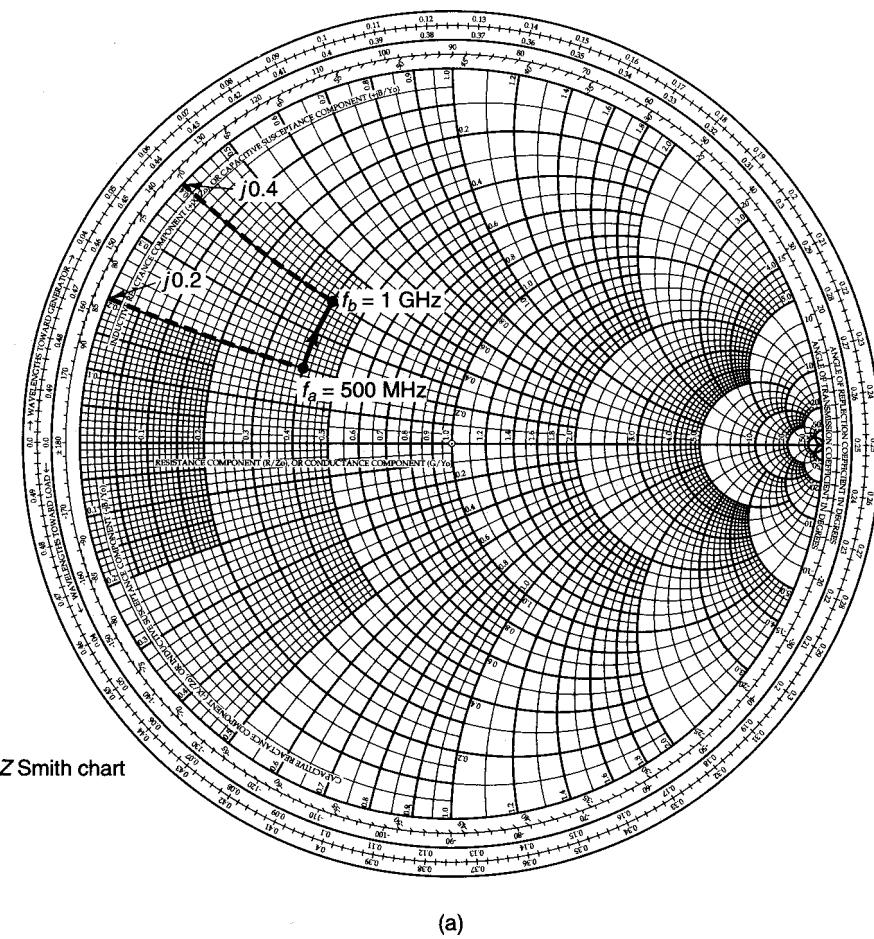


Figure 2.3.3 (a) Frequency response of a one-port network; (b) equivalent circuit of the one-port network.

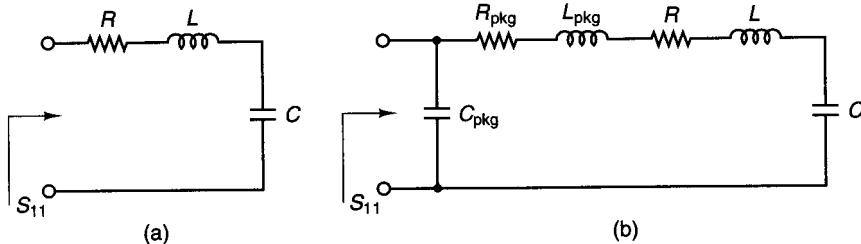


Figure 2.3.4 (a) Input equivalent circuit for a transistor in chip form; (b) input equivalent circuit for a packaged transistor.

Example 2.3.3

The frequency response of \$S_{11}\$ for a transistor chip is shown in Fig. 1.10.2, and its input equivalent network is shown in Fig. 2.3.4a. Determine the values of \$R\$, \$L\$, and \$C\$.

Solution. From Fig. 1.10.2, at \$f_a = 1\$ GHz the input impedance associated with \$S_{11}\$ is \$Z_{\text{IN}} = 50(0.2 - j0.2) = 10 - j10 \Omega\$, and at \$f_b = 10\$ GHz the input impedance is \$Z_{\text{IN}} = 50(0.2 + j0.15) = 10 + j7.5 \Omega\$. Hence, at \$\omega_a = 2\pi f_a\$ we obtain

$$10 - j10 = R + j\left(\omega_a L - \frac{1}{\omega_a C}\right) \quad (2.3.1)$$

and at \$\omega_b = 2\pi f_b\$ we obtain

$$10 + j7.5 = R + j\left(\omega_b L - \frac{1}{\omega_b C}\right) \quad (2.3.2)$$

From (2.3.1) and (2.3.2), it follows that \$R = 10 \Omega\$ and

$$-10 = 2\pi 10^9 L - \frac{1}{2\pi 10^9 C} \quad (2.3.3)$$

$$7.5 = 2\pi 10^{10} L - \frac{1}{2\pi 10^{10} C} \quad (2.3.4)$$

The simultaneous solution of (2.3.3) and (2.3.4) is \$L = 0.1024 \text{ nH}\$ and \$C = 14.95 \text{ pF}\$.

The equivalent circuit for the transistor in packaged form is a little more difficult to obtain. One equivalent circuit that will closely simulate the response of \$S_{11}\$ is shown in Fig. 2.3.4b. The resistor \$R_{\text{pkg}}\$ represents the resistance of the package, the package inductance is \$L_{\text{pkg}}\$, and the package capacitance is \$C_{\text{pkg}}\$.

2.4 IMPEDANCE MATCHING NETWORKS

The need for matching networks arises because amplifiers, in order to deliver maximum power to a load or to perform in a certain desired way, must be properly terminated at both the input and the output ports. Figure 2.4.1 illustrates a

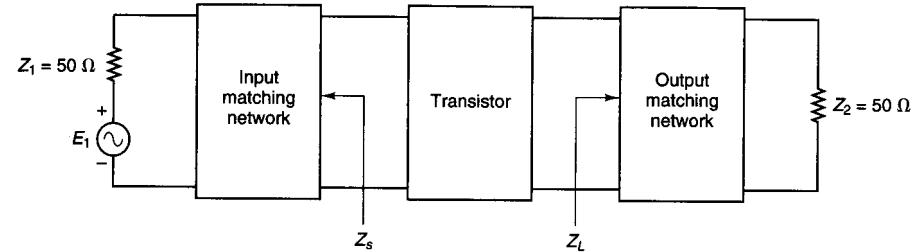


Figure 2.4.1 Block diagram of a microwave amplifier.

typical situation in which a transistor, in order to deliver maximum power to the \$50\Omega\$ load, must have the terminations \$Z_s\$ and \$Z_L\$. The input matching network is designed to transform the generator impedance (shown as \$50\Omega\$) to the source impedance \$Z_s\$, and the output matching network transforms the \$50\Omega\$ termination to the load impedance \$Z_L\$.

Although many different types of matching networks can be designed, the eight Ell sections (also denoted as \$L\$ sections) shown in Fig. 2.4.2 are not only simple to design but quite practical. The matching networks are lossless in order not to dissipate any of the signal power.

The ZY Smith chart can be used conveniently in the design of matching networks. The effect of adding a series reactance element to an impedance or a parallel susceptance element to an admittance, in the ZY Smith chart, is illustrated in the following example.

Example 2.4.1

(a) Illustrate the effect of adding a series inductor \$L\$ (\$z_L = j0.8\$) to an impedance \$z\$ (\$z = 0.3 - j0.3\$) in the ZY Smith chart.

Solution. Figure 2.4.3 shows that the effect of adding a series inductance with \$z_L = j0.8\$ is to move along a constant-resistance circle from a reactance value of \$-0.3\$ to a reactance of \$0.5\$. In other words, the motion is in a clockwise direction along a constant-resistance circle.

(b) Illustrate the effect of adding a series capacitor \$C\$ (\$z_C = -j0.8\$) to an impedance \$z\$ (\$z = 0.3 - j0.3\$) in the ZY Smith chart.

Solution. Figure 2.4.4 shows that the effect of adding a series capacitor with \$z_C = -j0.8\$ is to move along a constant-resistance circle from a reactance value of \$-0.3\$ to a reactance of \$-1.1\$. In other words, the motion is in a counterclockwise direction along a constant-resistance circle.

(c) Illustrate the effect of adding a shunt inductor \$L\$ (\$y_L = -j2.4\$) to an admittance \$y\$ (\$y = 1.6 + j1.6\$) in the ZY Smith chart.

Solution. Figure 2.4.5 shows that the effect of adding a shunt inductor with \$y_L = -j2.4\$ is to move along a constant-conductance circle from a susceptance of \$1.6\$ to a

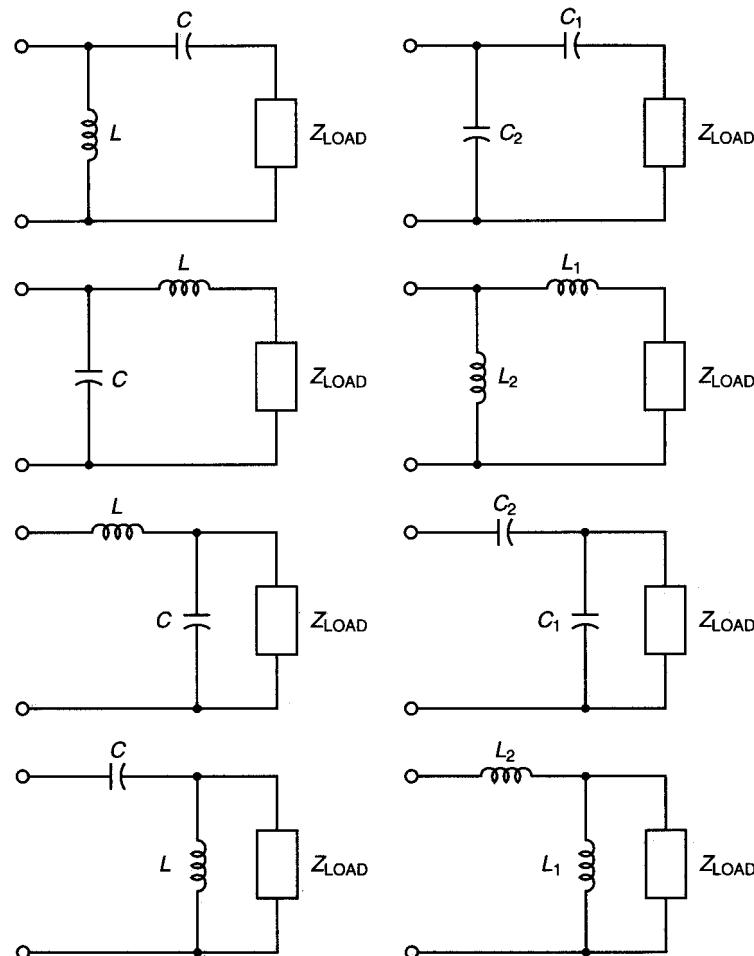


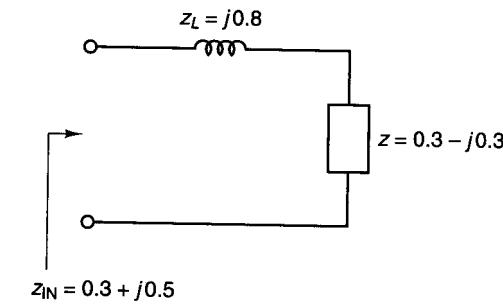
Figure 2.4.2 Matching networks.

susceptance of -0.8 . In other words, the motion is in a counterclockwise direction along a constant-conductance circle.

(d) Illustrate the effect of adding a shunt capacitor C ($y_C = j3.4$) to an admittance y ($y = 1.6 + j1.6$) in the ZY Smith chart.

Solution. Figure 2.4.6 shows that the effect of adding a shunt capacitor with $y_C = j3.4$ is to move along a constant-conductance circle from a susceptance of 1.6 to a susceptance of 5 . In other words, the motion is in a clockwise direction along a constant-conductance circle.

In conclusion, adding a series reactance produces a motion along a constant-resistance circle in the ZY Smith chart, and adding shunt susceptance



(a)

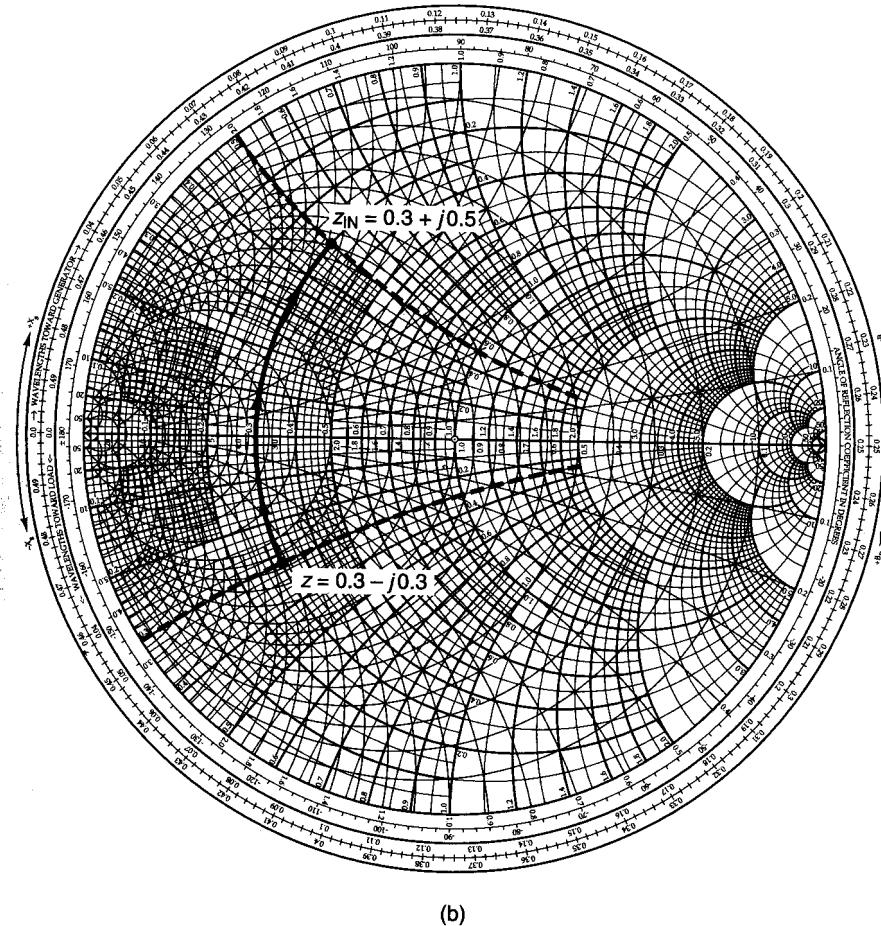
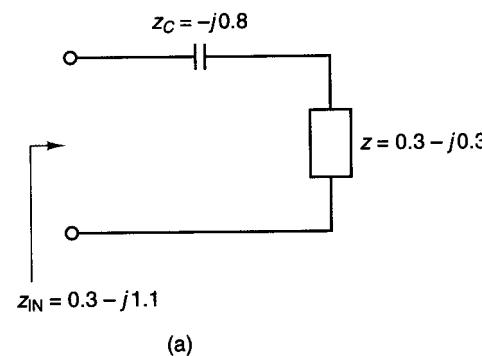
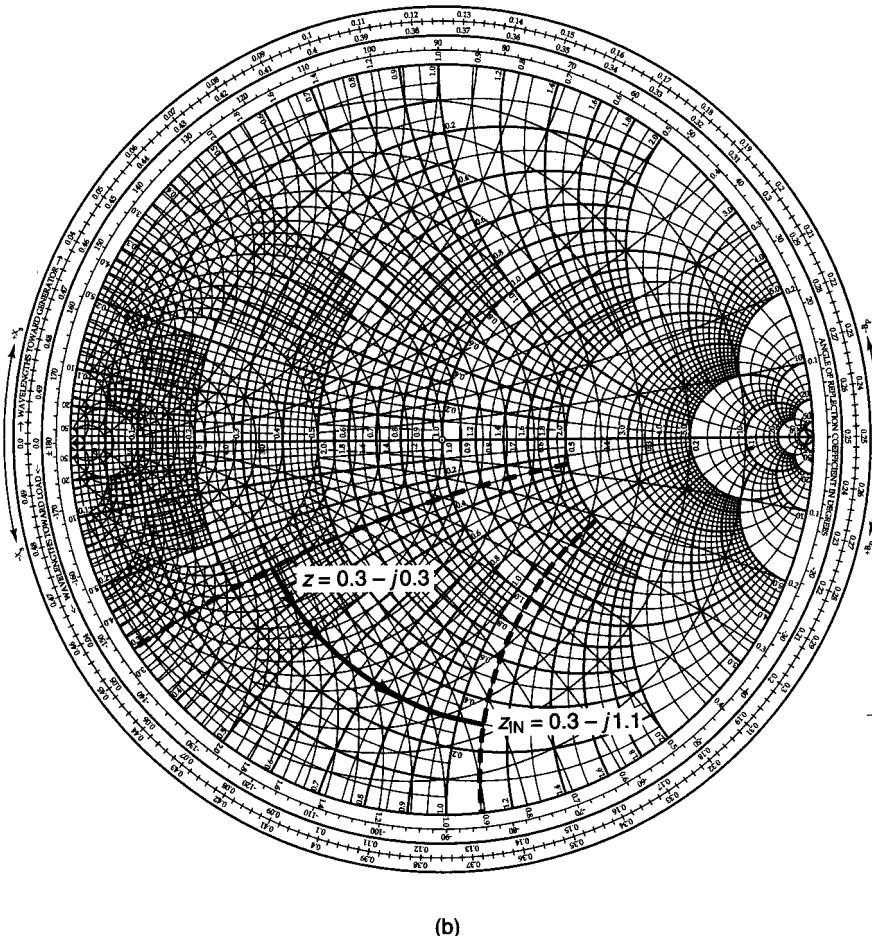


Figure 2.4.3 Effect of adding a series inductor to an impedance in the ZY Smith chart.

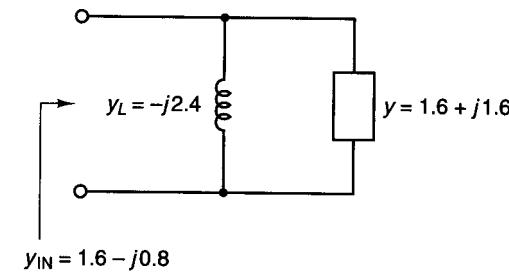


(a)

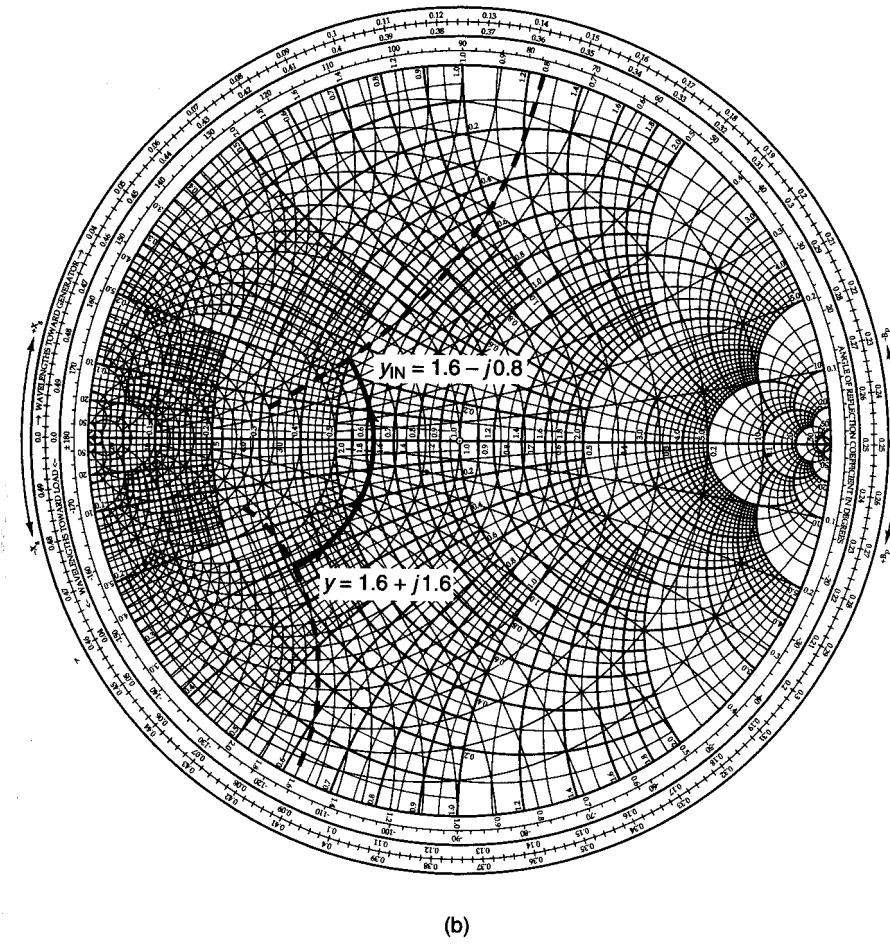


(b)

Figure 2.4.4 Effect of adding a series capacitor to an impedance in the ZY Smith chart.



(a)



(b)

Figure 2.4.5 Effect of adding a shunt inductor to an admittance in the ZY Smith chart.

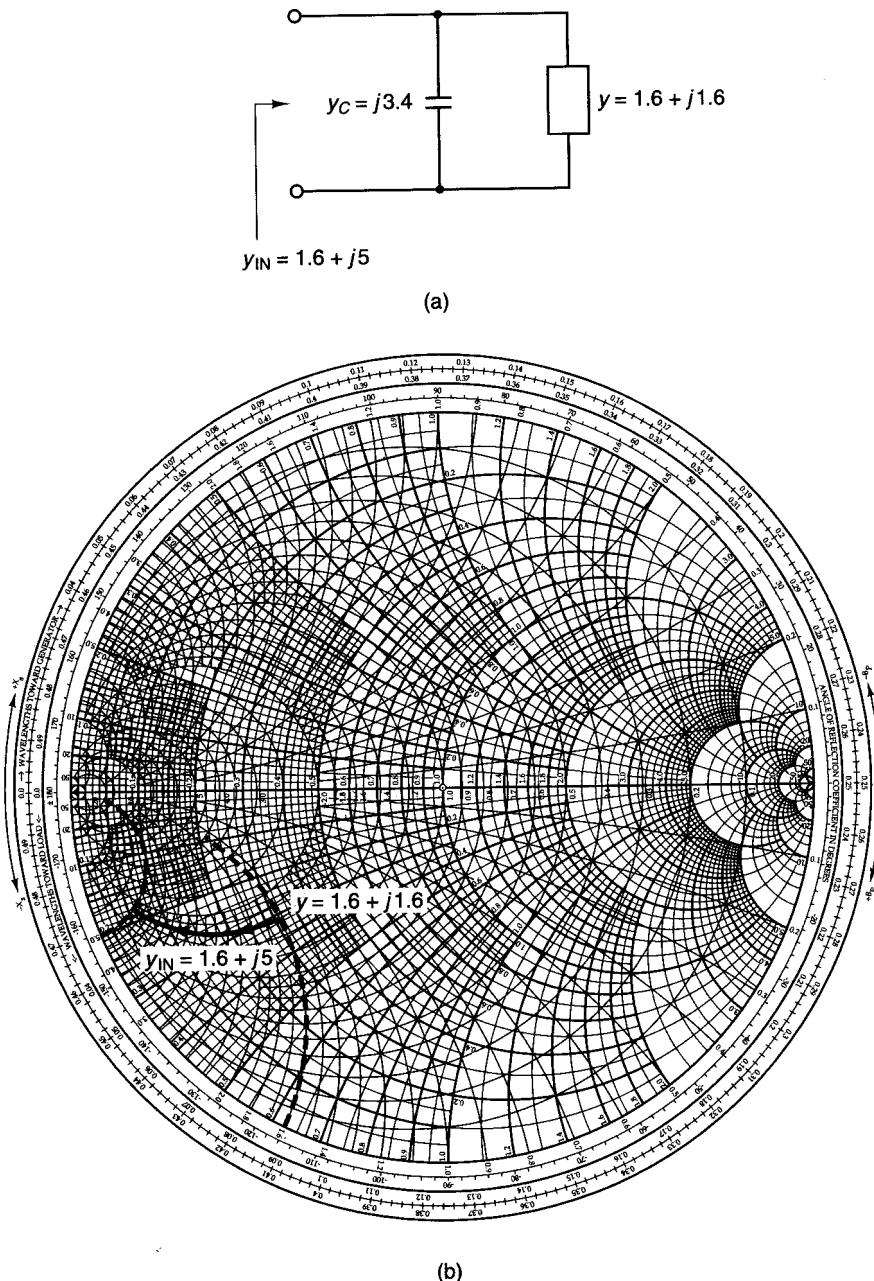


Figure 2.4.6 Effect of adding a shunt capacitor to an admittance in the ZY Smith chart.

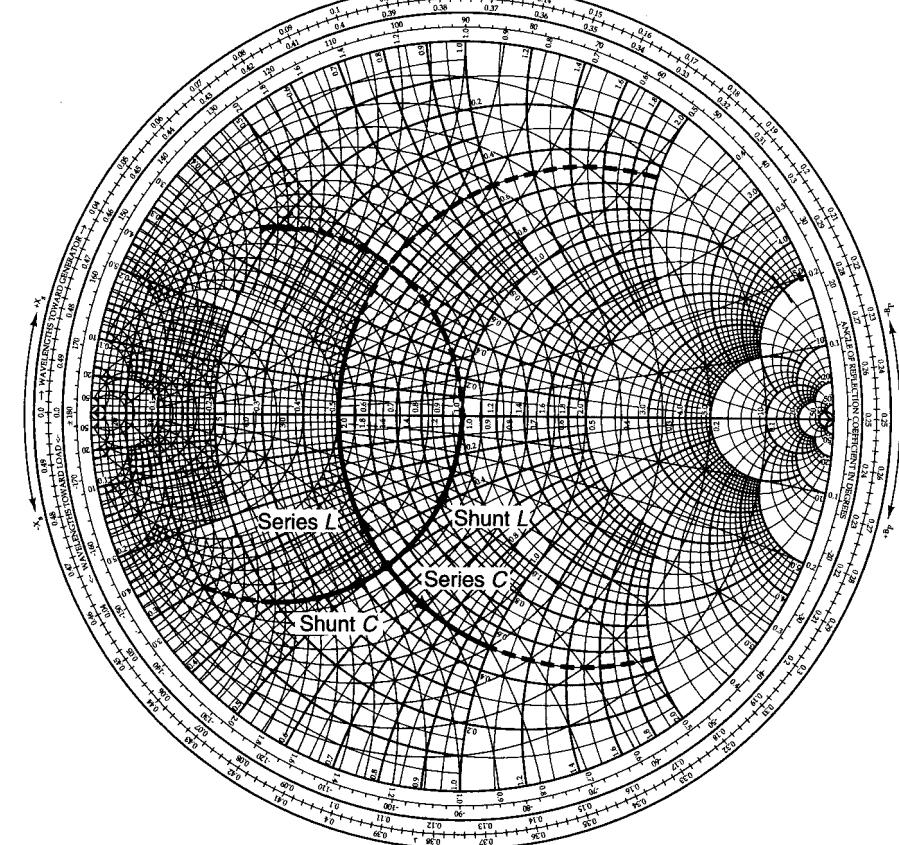


Figure 2.4.7 Effect of adding series and shunt elements in the ZY Smith chart.

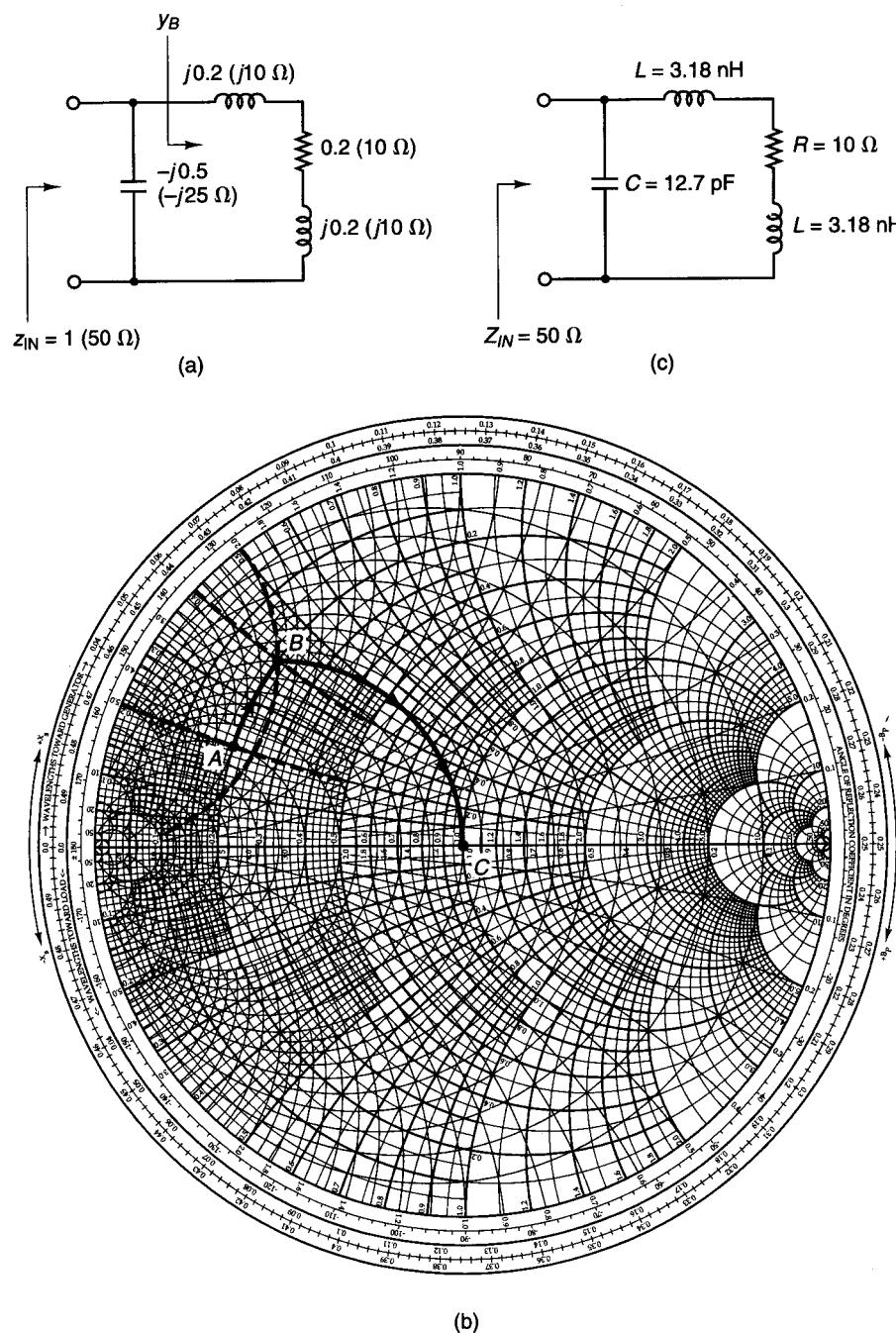
produces a motion along a constant-conductance circle in the ZY Smith chart. The four types of motions are illustrated in Fig. 2.4.7.

Designing a matching network in the ZY Smith chart consists of moving along a constant-resistance or constant-conductance circle from one value of impedance or admittance to another. Each motion along a constant-resistance or constant-conductance circle gives the value of an appropriate element. The following examples illustrate the use of the ZY Smith chart in the design of matching networks.

Example 2.4.2

A load $Z_{LOAD} = 10 + j10 \Omega$ is to be matched to a $50\text{-}\Omega$ line. Design two matching networks and specify the values of L and C at a frequency of 500 MHz.

Solution. Selecting the series L -shunt C network shown in Fig. 2.4.8a, the matching network is designed as shown in Fig. 2.4.8b. (See Figure 2.4.8b on the inside cover.)

Figure 2.4.8 Design of a series L -shunt C matching network.

The motion from point A [i.e., $z_{\text{LOAD}} = (10 + j10)/50 = 0.2 + j0.2$] to point B is along a constant-resistance circle, and we obtain for the inductor impedance $z_L = j0.4 - j0.2 = j0.2$. Observe that point B is along the unit constant-conductance circle. The admittance at point B is $y_B = 1 - j2$. The motion from point B to point C (i.e., the origin) is along a constant-conductance circle, and we obtain the capacitor admittance $y_C = 0 - (-j2) = j2$ (or $z_C = 1/j2 = -j0.5$). Therefore, at point C , $y_{\text{IN}} = z_{\text{IN}} = 1$ (or $Z_{\text{IN}} = 50 \Omega$) and the network is matched to a 50Ω line. At 500 MHz, the value of L is

$$L = \frac{10}{2\pi(500 \times 10^6)} = 3.18 \text{ nH}$$

and the value of C is

$$C = \frac{1}{25(2\pi)500 \times 10^6} = 12.74 \text{ pF}$$

The matching network at 500 MHz is shown in Fig. 2.4.8c.

The second matching network is shown in Fig. 2.4.9a and the ZY Smith chart design in Fig. 2.4.9b. (See Figure 2.4.9b on the inside cover.) The motion from A to B in Fig. 2.4.9b is along a constant-resistance circle; therefore, the impedance of the series capacitor is $z_C = -j0.4 - j0.2 = -j0.6$. The motion from B to C is along a constant-conductance circle; therefore, the admittance of the shunt inductor is $y_L = 0 - j2 = -j2$ (or $z_L = 1/-j2 = j0.5$). The design at 500 MHz is shown in Fig. 2.4.9c.

Example 2.4.3

Design the matching network shown in Fig. 2.4.10a to transform a 50Ω load at the input to an admittance $Y_{\text{OUT}} = (8 - j12) \times 10^{-3} \text{ S}$ at the output.

Solution. Figure 2.4.10b illustrates a motion in the ZY Smith chart from the origin (i.e., $z_{\text{LOAD}} = 50/50 = 1$) to $y_{\text{OUT}} = 50(8 - j12) \times 10^{-3} = 0.4 - j0.6$. (See Figure 2.4.10b on the inside cover.) The motion from A to B produces a series capacitor having an impedance of $z_C = -j1.21$. The motion from B to C produces a shunt inductor having an admittance of $y_L = -j0.6 - j0.49 = -j1.09$ (or $z_L = 1/-j1.09 = j0.917$). The matching network is shown in Fig. 2.4.10c.

Example 2.4.4

Design a matching network to transform the load $Z_{\text{LOAD}} = 100 + j100 \Omega$ to an input impedance of $Z_{\text{IN}} = 50 + j20 \Omega$.

Solution. Using a normalized value of $Z_o = 100 \Omega$, we have $z_{\text{LOAD}} = Z_{\text{LOAD}}/Z_o = 1 + j$ and $z_{\text{IN}} = Z_{\text{IN}}/Z_o = 0.5 + j0.2$. Figure 2.4.11a shows one possible solution for the matching network. The motion from A to B is produced by a shunt capacitor whose normalized admittance is $y_C = j0.86 - (-j0.5) = j1.36$ (or $z_C = 1/y_C = -j0.735$). Then, the impedance of the capacitor is

$$Z_C = 100z_C = 100(-j0.735) = -j73.5 \Omega$$

The motion from B to C requires an inductor having a normalized impedance value of $z_L = j0.2 - (-j0.87) = j1.07$. Then

$$Z_L = 100z_L = 100(j1.07) = j107 \Omega$$

The matching network is shown in Fig. 2.4.11b.

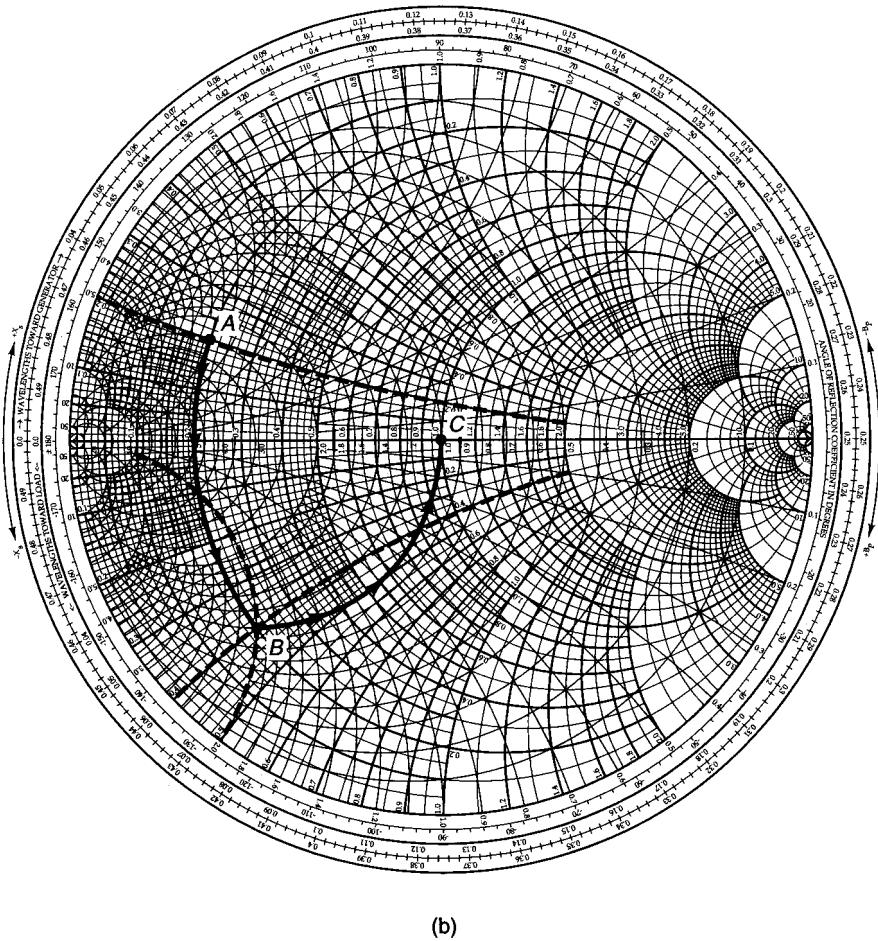
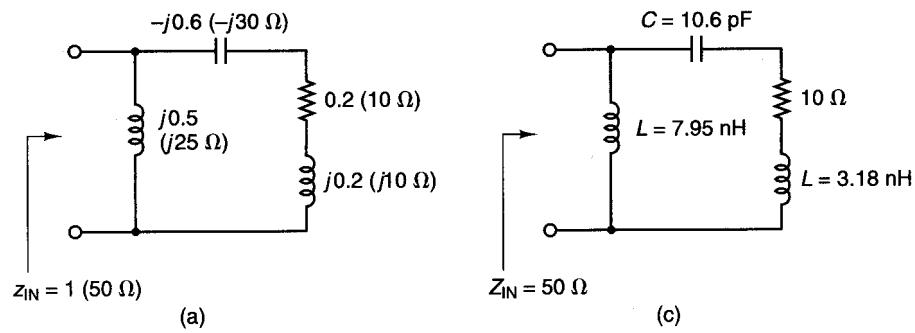


Figure 2.4.9 Design of a series C-shunt L matching network.

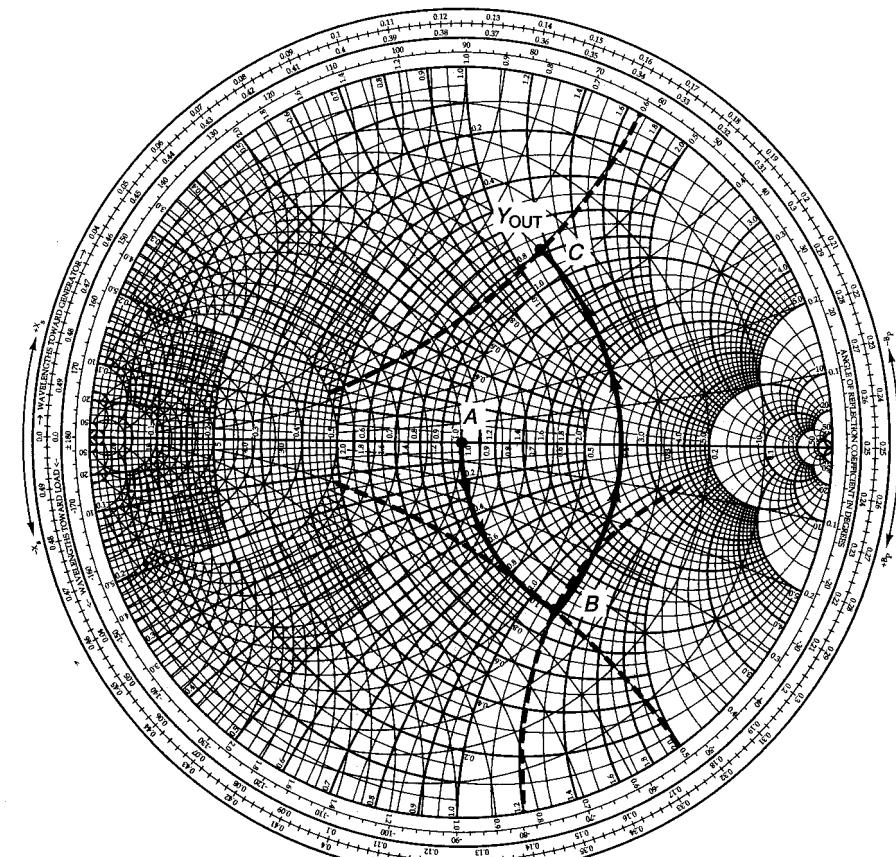
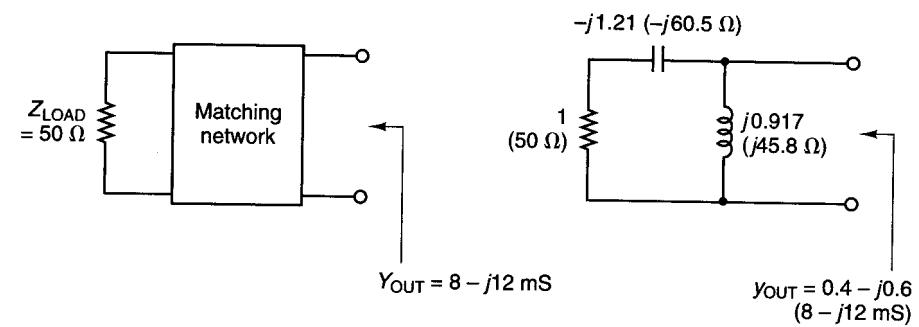


Figure 2.4.10 Matching a 50Ω load to a given Y_{OUT} using a series C-shunt L matching network.

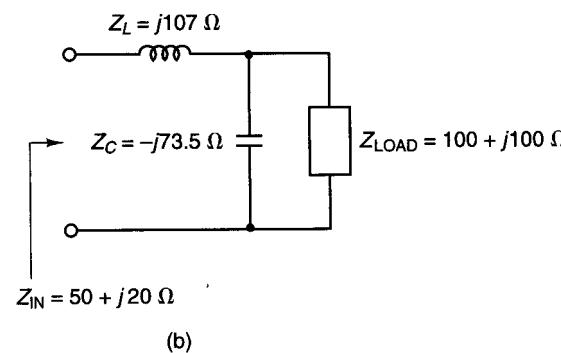
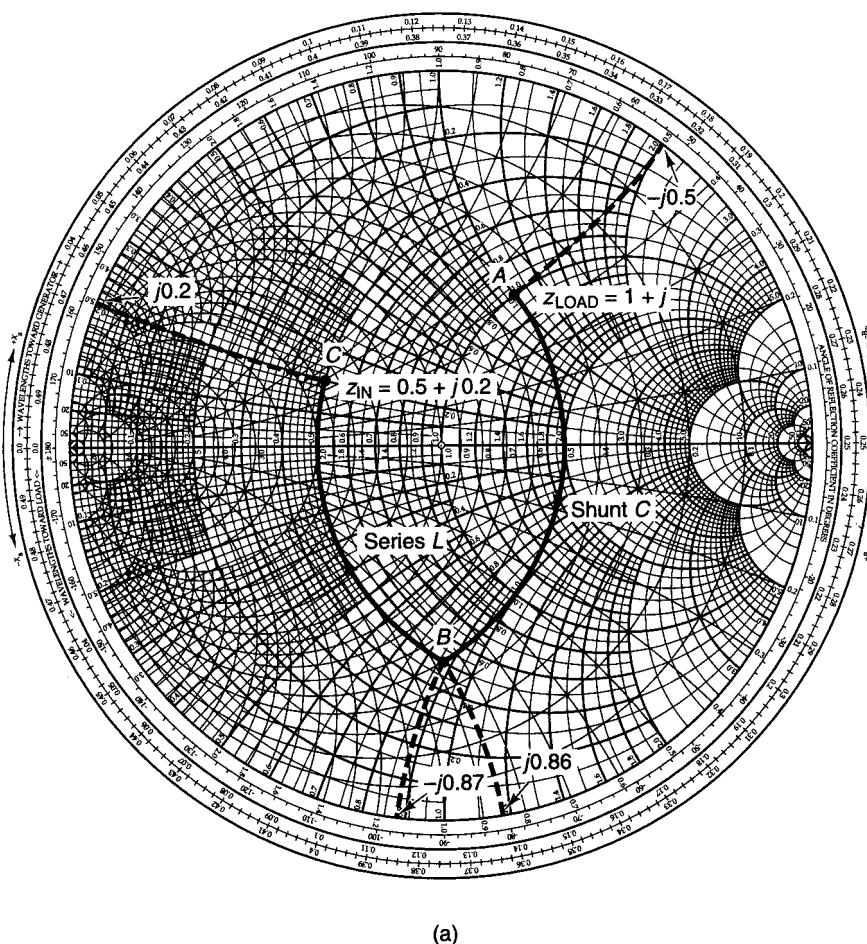


Figure 2.4.11 Matching a complex load to a given complex impedance Z_{IN} using a shunt C-series L network.

Sometimes a specific matching network cannot be used to accomplish a given match. For example, any load impedance falling in the marked region in Fig. 2.4.12a cannot be matched to 50Ω with the network in Fig. 2.4.12b because adding a series L produces motion, in a clockwise direction, away from any constant-conductance circle that passes through the origin.

Another observation regarding the Ell matching networks is that only those with an inductor and a capacitor can be used to provide a match between a resistive load and an input resistance.

In a resonant circuit, the ratio of its resonant frequency f_o to its bandwidth (BW) is known as the loaded Q of the circuit. That is,

$$Q_L = \frac{\omega_o}{BW}$$

If the bandwidth is expressed in hertz, we write

$$Q_L = \frac{f_o}{BW} \quad (2.4.1)$$

The Ell matching networks in Fig. 2.4.2 are used to provide a match at a certain frequency. The frequency response of an Ell network can be classified as either a (two-pole) low-pass filter or a high-pass filter. At each node of the Ell matching networks, there is an equivalent series input impedance, denoted by $R_s + jX_s$. Hence, a circuit node Q , denoted by Q_n , can be defined at each node as

$$Q_n = \frac{|X_s|}{R_s} \quad (2.4.2)$$

If the equivalent parallel input admittance at the node is $G_p + jB_p$, the circuit node Q can be expressed in the form

$$Q_n = \frac{|B_p|}{G_p} \quad (2.4.3)$$

For example, the normalized series input impedance of the circuit in Fig. 2.4.8a at point B (see Fig. 2.4.8b) is $0.2 + j0.4$; therefore, the circuit node Q , from (2.4.2), is $Q_n = 0.4/0.2 = 2$. Also, using the Smith chart, the equivalent parallel admittance is readily found to be $1 - j2$, producing a circuit node Q , according to (2.4.3), of $Q_n = 2/1 = 2$, which agrees with the result from (2.4.2).

The circuit in Fig. 2.4.8a is redrawn in Fig. 2.4.13a. The circuit is excited with a source voltage having a 50Ω source impedance (i.e., for a proper match at the input), and the output voltage is taken across the real part of the load impedance (i.e., the 10Ω resistor). For convenience, the two 3.18-nH inductors are combined, and the circuit can be viewed as performing a match between a 10Ω load and a 50Ω input resistance. The frequency response of the circuit

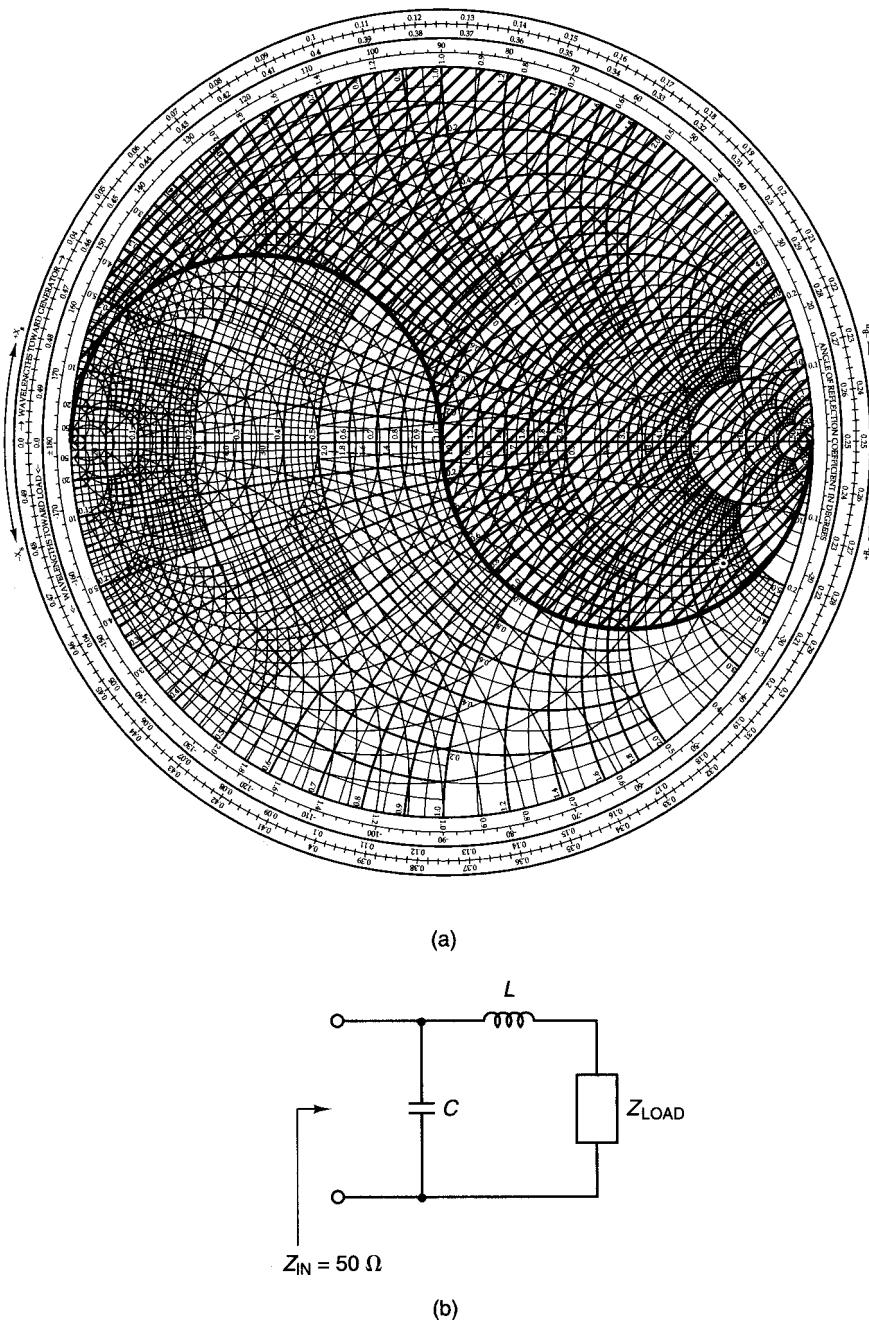


Figure 2.4.12 Forbidden region in the ZY Smith chart to match a given Z_{LOAD} to 50 Ω using a series L-shunt C matching network.

is illustrated in Fig. 2.4.13b. Figure 2.4.13b shows that a peak occurs at the frequency f_o where the match occurs, and the filter attenuates the frequencies above and below f_o .

For a narrowband range of frequencies around f_o , the filter can be viewed as a bandpass filter with a loaded Q , which can be calculated using (2.4.1). The equivalent bandpass filter is shown in Fig. 2.4.13c, as well as the frequency response. This circuit was obtained by changing $Z_B = 10 + j20 \Omega$ to an equivalent admittance (i.e., 50 Ω in parallel with $j25 \Omega$). The tuned circuit in Fig. 2.4.13c is loaded by the resistance $R_T = 50 \parallel 50 = 25 \Omega$. Hence, the loaded Q of the bandpass filter is given by

$$Q_L = \frac{\omega_o}{\text{BW}} = \omega_o R_T C = \frac{|B_C|}{G_T} = \frac{R_T}{|X_C|} = \frac{25}{25} = 1 \quad (2.4.4)$$

In Fig. 2.4.13c, since the gain at $f_o = 500$ MHz is -6 dB, it follows that the value of $|v_B|$ is $|v_s|/2$. Then, from Fig. 2.4.13a, we have

$$|v_o| = |v_B| \frac{10}{|10 + j20|} = \frac{|v_s|}{2} (0.447) = |v_s|(0.224)$$

or

$$20 \log \frac{|v_o|}{|v_s|} = -13 \text{ dB}$$

in agreement with the gain at $f_o = 500$ MHz in Fig. 2.4.13b.

The equivalent bandpass filter helps to explain the *bell-shape* response of the circuit in the neighborhood of f_o . Away from the neighborhood of f_o , the frequency responses of the circuit in Figs. 2.4.13a and 2.4.13c are quite different.

A question arises about the relation between the circuit node Q and the loaded Q associated with the response in the neighborhood of f_o . Referring to Fig. 2.4.13a, it is also observed that the circuit node Q looking toward the 50- Ω source is 2 (since the normalized admittance of the 50- Ω source and the $-j25 \Omega$ capacitor is $1 + j2$). Comparing Q_n in (2.4.3) with Q_L in (2.4.4), it follows that in the neighborhood of f_o the loaded Q of the Ell matching networks is given by

$$Q_L = \frac{Q_n}{2} \quad (2.4.5)$$

Using (2.4.5), the loaded Q of the circuit in Fig. 2.4.13a is

$$Q_L = \frac{Q_n}{2} = \frac{2}{2} = 1$$

Hence, the approximate bandwidth of the circuit in Fig. 2.4.13a is

$$\text{BW} \approx \frac{f_o}{Q_L} = \frac{500 \times 10^6}{1} = 500 \text{ MHz}$$

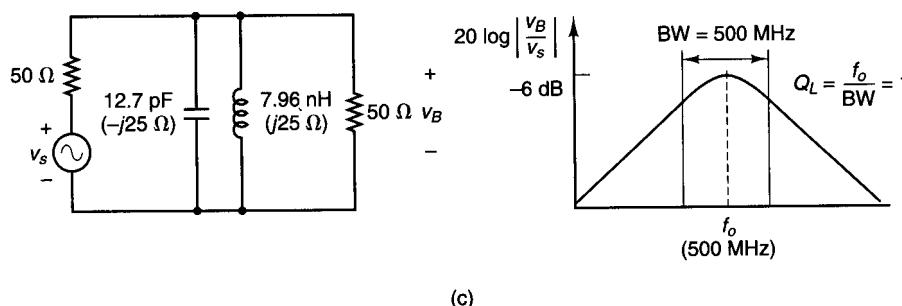
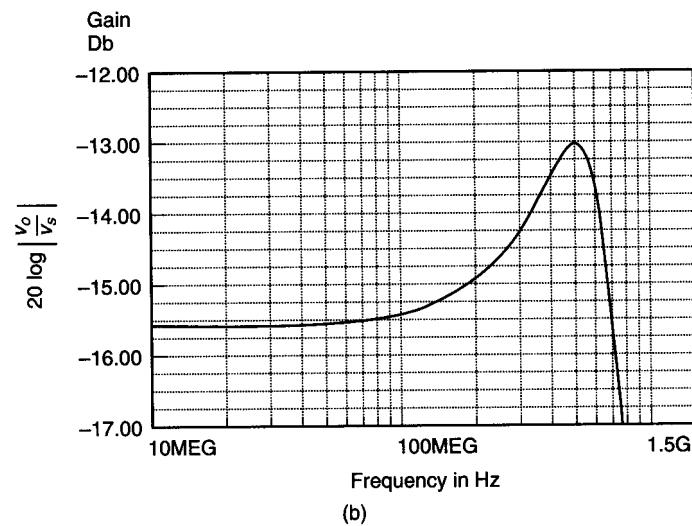
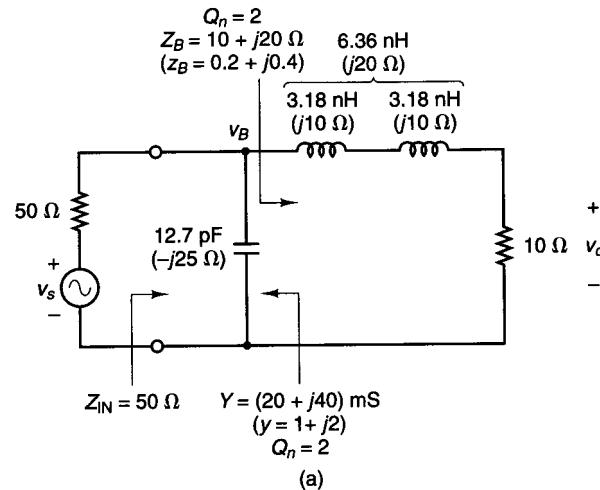


Figure 2.4.13 (a) The matching network from Fig. 2.4.8a; (b) its frequency response (the program *Microcap III* was used); (c) equivalent bandpass filter configuration in the neighborhood of f_o , and its frequency response.

The frequency response in Fig. 2.4.13b shows that the 3-dB point for $f > f_o$ occurs at $f = 720 \text{ MHz}$. For $f < f_o$, there is no 3-dB point. This is due to the low value of Q_n in Fig. 2.4.13a and, consequently, to the frequency response characteristics of the circuit for $f < f_o$. However, if we assume that the bell-shape response is symmetrical around f_o , it follows that the bandwidth is approximately

$$\text{BW} \approx 2(720 \times 10^6 - 500 \times 10^6) = 440 \text{ MHz}$$

in reasonable agreement with the predicted bandwidth.

The following example illustrates the bandpass characteristics of an Ell network when a high value of Q_n is used.

Example 2.4.5

The low-pass Ell network shown in Fig. 2.4.14a was designed to transform a $200\text{-}\Omega$ load to an input resistance of 20\Omega (see Fig. 2.4.14b) at 500 MHz . Determine the loaded Q of the circuit.

Solution. From Fig. 2.4.14b, the impedance at point B is $z_B = 0.1 - j0.3$. Hence, the circuit node Q is $Q_n = 0.3/0.1 = 3$. Also, observe that the normalized impedance looking toward the source is $z_B = 0.1 + j0.3$, giving $Q_n = 3$.

From (2.4.4), it follows that Q_L is

$$Q_L = \frac{Q_n}{2} = \frac{3}{2} = 1.5$$

and the expected bandwidth is

$$\text{BW} \approx \frac{f_o}{Q_L} = \frac{500 \times 10^6}{1.5} = 333.33 \text{ MHz}$$

The frequency response of the circuit in Fig. 2.4.14c shows that $f_o = 500 \text{ MHz}$ and $\text{BW} = 650 \times 10^6 - 275 \times 10^6 = 375 \text{ MHz}$. Hence,

$$Q_L = \frac{500}{375} = 1.33$$

These values agree with the expected bandwidth and Q_L .

In conclusion, in order to obtain a high value of Q_L , the circuit node Q must be high. For the low-pass matching circuit, the attenuation characteristics for frequencies above f_o are obviously better than for frequencies below f_o . For bandpass applications, the attenuation of the harmonics above f_o is important, and the Ell filters will provide an attenuation of -12 dB/octave .

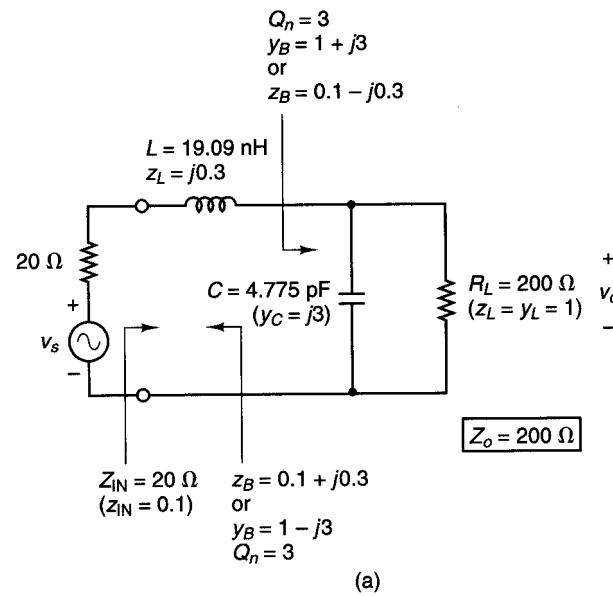


Figure 2.4.14 (a) The low-pass matching circuit for Example 2.4.5; (b) the design in the *ZY* Smith chart; (c) the frequency response (the program *Microcap-III* was used).

In many cases, the matching network provides a match to a complex impedance (or admittance). The frequency response of these circuits can be obtained using an excitation whose impedance (or admittance) is the complex conjugate of the impedance (or admittance) presented by the matching network. In some cases, it is convenient to represent the source excitation by its Norton's equivalent circuit. For example, in Fig. 2.4.10c the matching network was designed to present an output admittance of $Y_{\text{OUT}} = (8 - j12) \text{ mS}$. The frequency response of this circuit can be performed using a current source in parallel with a source admittance of $(8 + j12) \text{ mS}$.

When a high value of Q_L is a design consideration, the Ell matching networks in Fig. 2.4.2 might not be suitable since the Q_L cannot be controlled when matching with two elements. In fact, with the Ell matching networks we have to accept the resulting Q_n in the design. For example, in the design shown in Fig. 2.4.13a the value of Q_n is determined by the equivalent impedance at point *B*—namely, $z = 0.2 + j0.4$; therefore, in this circuit Q_n is fixed at the value of 2.

Higher values of Q_L than those obtained with the Ell circuits can be obtained using matching circuits with three elements. The addition of a third element to an Ell matching network results in either the lossless Tee network (also denoted as *T* network) or the lossless Pi network (also denoted as *Π* network) shown in Fig. 2.4.15.

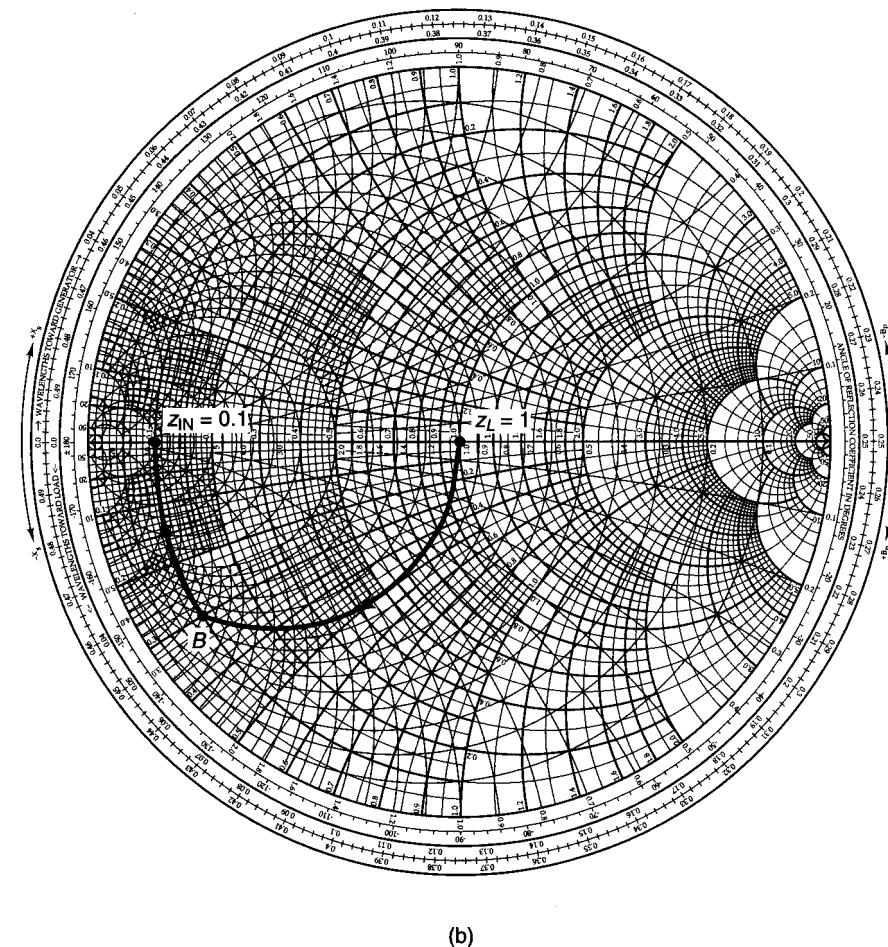


Figure 2.4.14 Continued

The addition of a third element introduces flexibility in the selection of the loaded Q , since the equivalent series impedance (or the equivalent parallel admittance) at the nodes in the circuit will determine various values of Q_n . Obviously, a high value of Q_n in the circuit will result in a high value of Q_L . However, it is not simple to exactly relate Q_n to Q_L in these circuits. The Q of a Tee or Pi network is normally taken as the highest value of Q_n in the circuit.

Several equivalent series input impedances can have the same Q_n . For example, the normalized impedances $z = 0.5 \pm j0.5$ have a Q_n of 1, as well as the impedances $z = 1 \pm j1$. Constant Q_n contours can be drawn on the Smith chart. This is illustrated in Fig. 2.4.16, where constant Q_n contours of 1, 5, and 10 are shown.

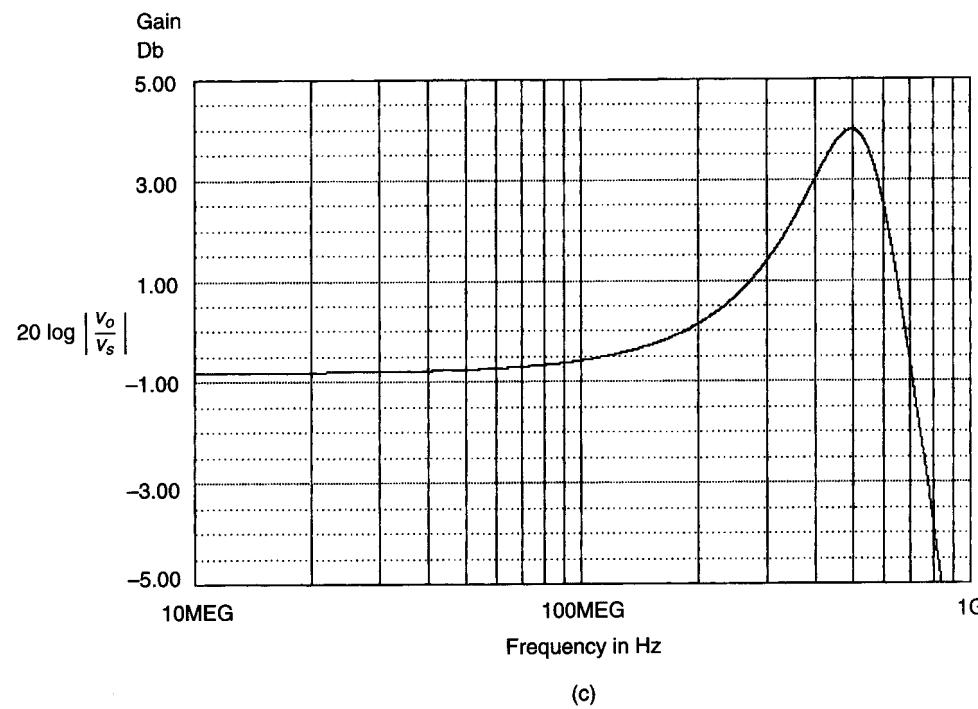


Figure 2.4.14 Continued

The upper and lower parts of the constant- Q contours can be shown to satisfy a circle equation as follows. Since

$$z = r + jx = \frac{1 + \Gamma}{1 - \Gamma} = \frac{1 - U^2 - V^2}{(1 - U)^2 + V^2} + j \frac{2U}{(1 - U)^2 + V^2}$$

then

$$Q_n = \frac{|x|}{r} = \frac{2U}{1 - U^2 - V^2}$$

which can be written as

$$U^2 + \left(V \pm \frac{1}{Q_n} \right)^2 = 1 + \frac{1}{Q_n^2} \quad (2.4.6)$$

The plus sign applies when x is positive, and the minus sign when x is negative.

Equation (2.4.6) is recognized as the equation of a circle. For $x > 0$, the center in the Γ plane is at $(0, -1/Q_n)$, and for $x < 0$ at $(0, 1/Q_n)$; the radius of the circle is

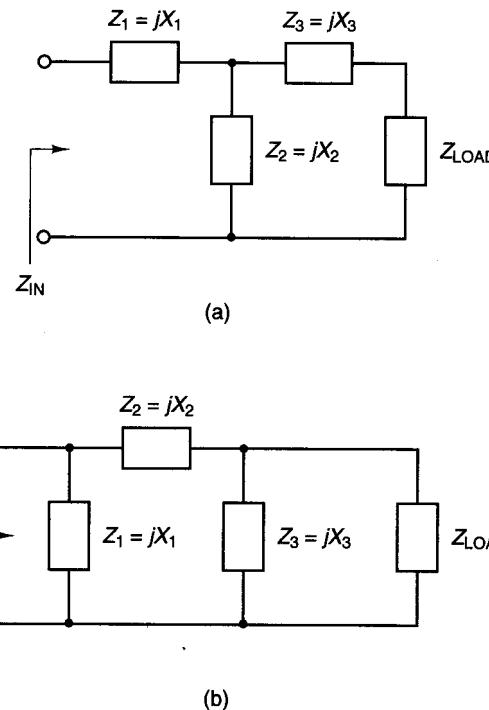


Figure 2.4.15 (a) Lossless Tee network; (b) lossless Pi network.

$$\sqrt{1 + \frac{1}{Q_n^2}}$$

For example, for the $Q_n = 5$ contour in Fig. 2.4.16, the upper and lower parts of the contour are simply one half of a circle centered at $(0, \mp 0.2)$, respectively, with radius of

$$\sqrt{1 + \frac{1}{25}} = 1.02$$

The following example illustrates the design of Tee and Pi networks for a given Q_n factor.

Example 2.4.6

Design two Tee networks to transform the load impedance $Z_{LOAD} = 50 \Omega$ to the input impedance $Z_{IN} = 10 - j15 \Omega$ with a Q_n of 5.

Solution. The design of a Tee matching network with a Q_n of 5 is illustrated in Fig. 2.4.17a, where a normalizing impedance of 50Ω was used. The motion from A to B produces a series inductor with impedance $z_{L_1} = j2$; the motion from B to C produces a

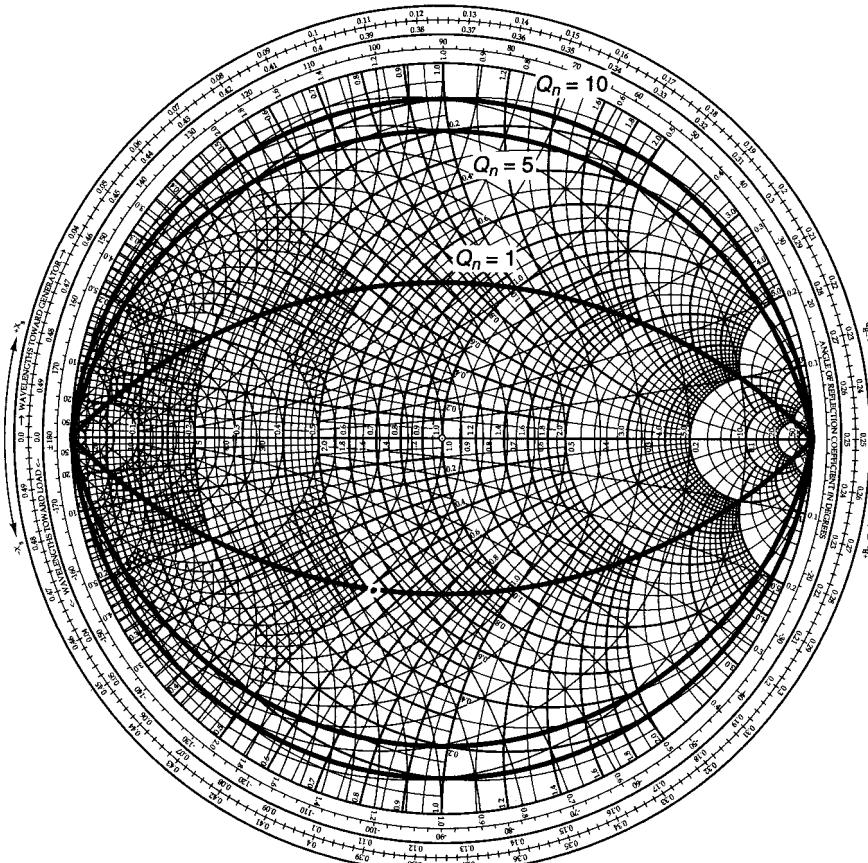


Figure 2.4.16 Constant Q_n contours for $Q_n = 1, 5$, and 10 .

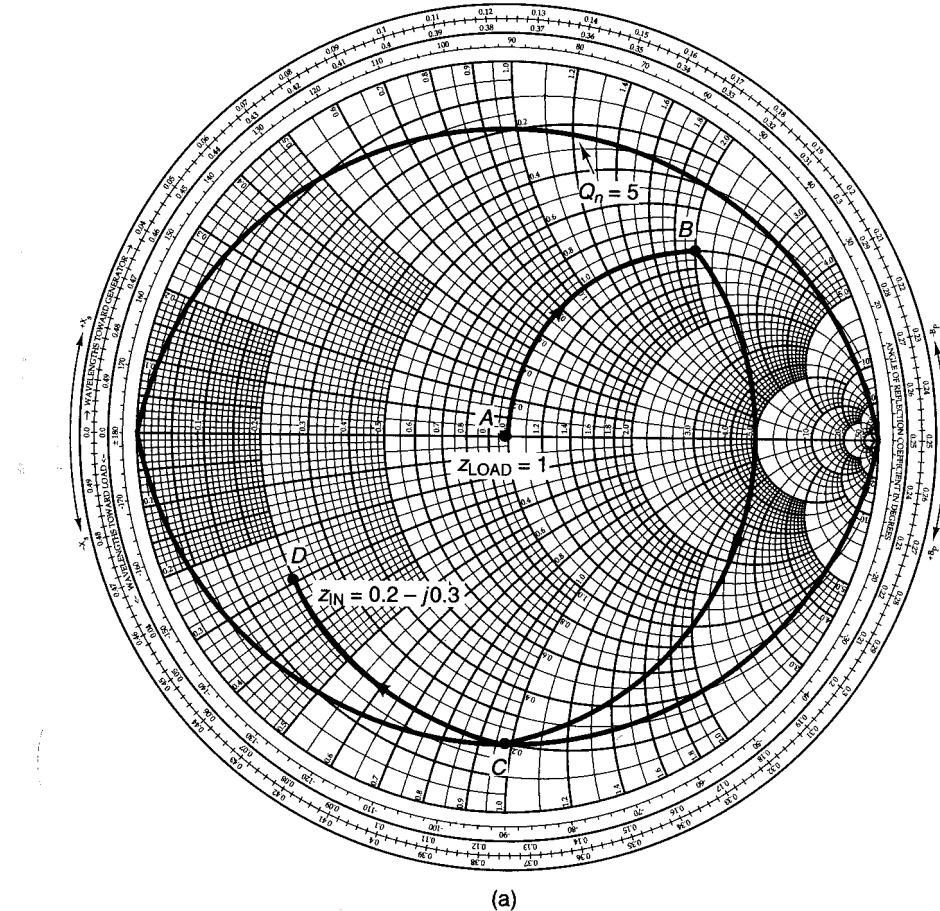
shunt capacitor with admittance $y_C = j0.96 - (-j0.4) = j1.36$ (or $z_C = 1/j1.36 = -j0.735$); and the motion from C to D produces a series inductor with impedance $z_{L_2} = -j0.3 - (-j1) = j0.7$. The impedance value at point C , which touches the $Q_n = 5$ contour, determines the Q of the network. The resulting Tee network is shown in Fig. 2.4.17b, and its frequency response at 500 MHz is shown in Fig. 2.4.17c. From Fig. 2.4.17c, the loaded Q is calculated to be

$$Q_L = \frac{f_o}{\text{BW}} = \frac{500 \times 10^6}{568 \times 10^6 - 382 \times 10^6} = 2.7$$

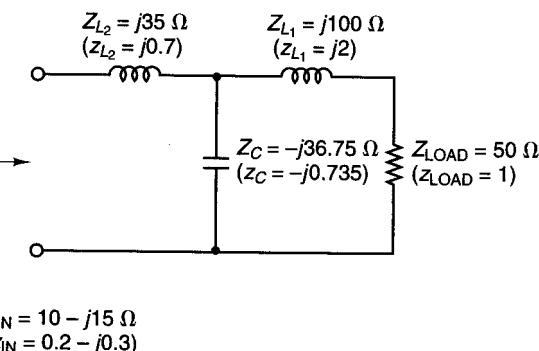
It is simple to verify that the other Tee network design, shown in Fig. 2.4.17d, transforms $Z_{\text{LOAD}} = 50 \Omega$ to $Z_{\text{IN}} = 10 - j15 \Omega$.

Example 2.4.7

Design two Pi networks to transform the load impedance $Z_{\text{LOAD}} = 50 \Omega$ to the input impedance $Z_{\text{IN}} = 150 \Omega$ with a Q_n of 5.



(a)



(b)

Figure 2.4.17 (a) Smith chart design of a Tee network for Example 2.4.6; (b) the Tee matching network; (c) the frequency response (the program *Microcap-III* was used); (d) another Tee network design.

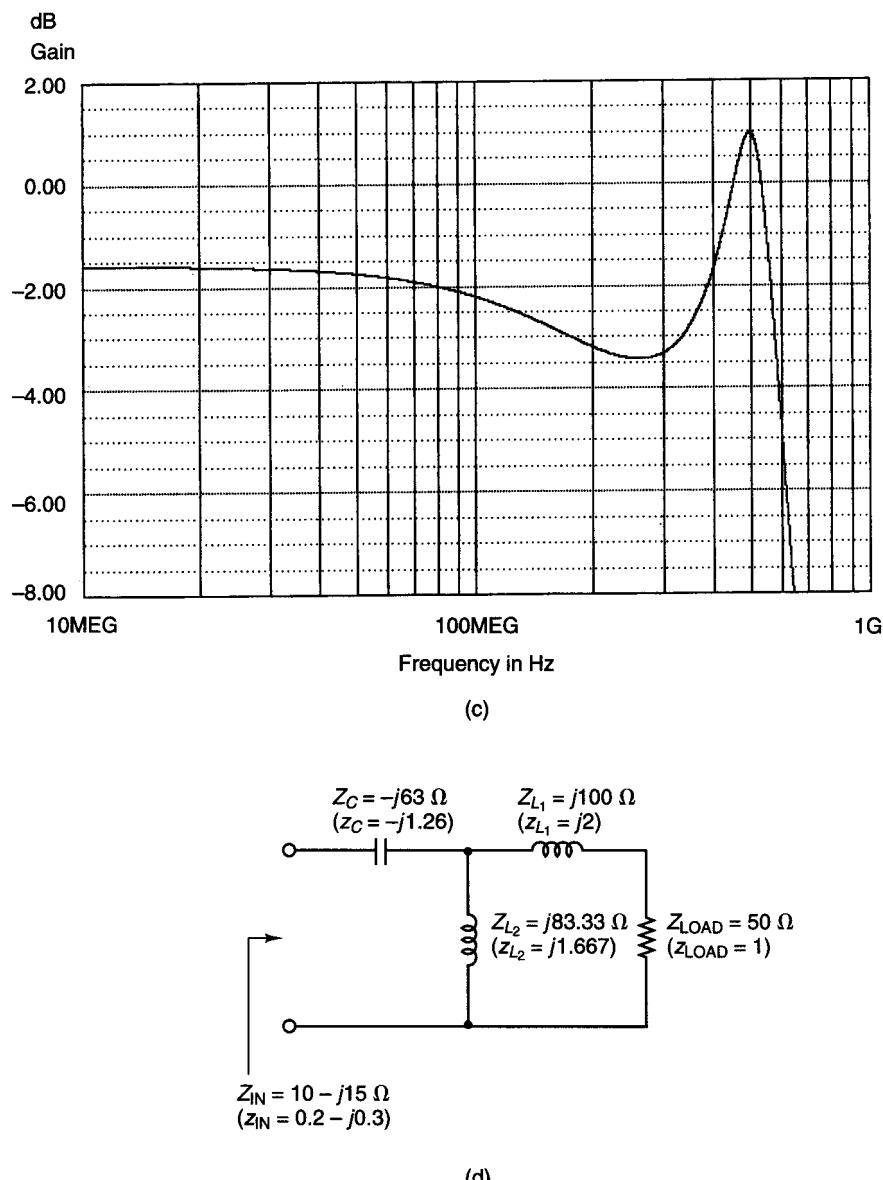


Figure 2.4.17 Continued

Solution. The design of a Pi network with a Q_n of 5 is illustrated in Fig. 2.4.18a. The motion from A to B requires a shunt inductor with admittance $y_{L_1} = -j2.9$ (or $z_{L_1} = j0.345$); the motion from B to C is produced by a series inductor with impedance $z_{L_2} = j0.55 - j0.31 = j0.24$. Point C is on the $Q_n = 5$ contour; therefore, the circuit Q is 5. Finally, the motion from C to D produces a shunt capacitor whose admittance is $y_C = 0 - (-j1.75) = j1.75$ (or $z_C = -j0.571$). The resulting Pi network is shown in Fig. 2.4.18b.

It is simple to verify that the other Pi network design, shown in Fig. 2.4.18c, transforms $Z_{LOAD} = 50 \Omega$ to $Z_{IN} = 150 \Omega$.

The design of lossless matching networks is accomplished by moving along constant-resistance and constant-conductance circles in the Smith chart. A question commonly asked is, "What happens if one moves along a constant-reactance or constant-susceptance circle in the Smith chart?". The answer is simple: A motion along a constant-reactance circle introduces a series resistance, and a motion along a constant-susceptance circle introduces a shunt resistance. The resistor is positive if the motion is along the constant resistance circle in the direction that r increases, or along the constant susceptance circle in the direction that g increases. It is important to emphasize that resistors are lossy elements and, in general, are avoided in the design of matching networks.

Example 2.4.8

A matching network is designed as shown in the ZY Smith chart in Fig. 2.4.19a to transform a 50Ω load to the input impedance $Z_{IN} = 50(1.6 - j1.2) \Omega$. Draw the matching circuit and determine the component values.

Solution. The motion from A to B results in a series inductor of value $z_L = j0.6$ (or $Z_L = 50(j0.6) = j30 \Omega$). The motion from B to C , along the constant-reactance circle $x_L = 0.6$, results in a series resistor. The impedance at B is $z_B = 1 + j0.6$, and the impedance at C is $z_C = 2.3 + j0.6$. Therefore, the value of the series resistor is $r = r_C - r_B = 2.3 - 1 = 1.3$, or $R = 50r = 50(1.3) = 65 \Omega$. Finally, the motion from C to D results in a capacitor of value $y_C = j0.3 - (-j0.1) = j0.4$ [or $z_C = -j2.5$, $Z_C = 50(-j2.5) = -j125 \Omega$]. The matching circuit is shown in Fig. 2.4.19b.

In a microwave amplifier (see Fig. 2.4.1), the input and output matching networks provide the appropriate ac impedances to the transistor. The transistor must also be biased at an appropriate quiescent point. A complete microwave amplifier contains both dc bias components and the ac matching networks. RFCs, bypass capacitors, and coupling capacitors need to be introduced so the dc bias components do not affect the ac performance of the amplifier. An example of a discrete microwave amplifier is shown in Fig. 2.4.20a. The capacitors denoted by C_B are bypass capacitors (they behave like short circuits at the frequency of operation). The RFCs behave like open circuits at the frequency of operation. The resistors R_1, R_2, R_C , and R_E set the quiescent point of the transistor. The input matching circuit consists of C_1 and L_1 , and the output matching circuit consists of C_2 and L_2 . In addition, the capacitors C_1 and C_2 act like coupling capacitors. That is, they isolate the dc bias circuit from the

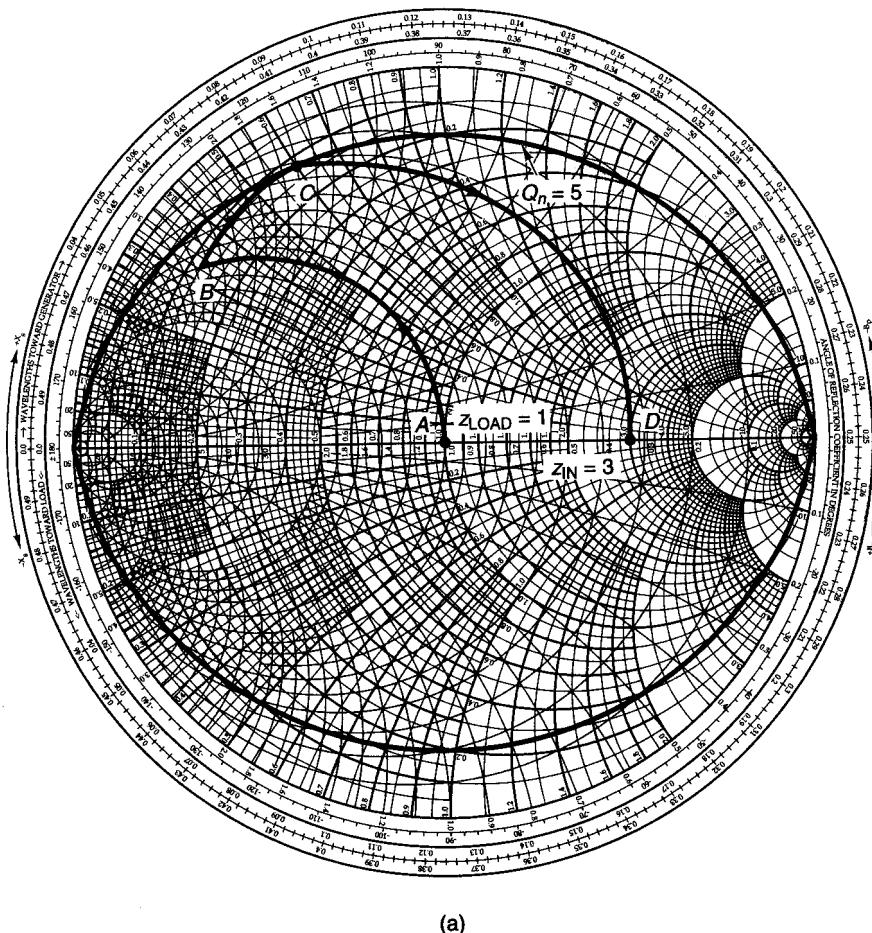


Figure 2.4.18 (a) Smith chart design of a Pi network for Example 2.4.7; (b) the Pi matching network; (c) another Pi network design.

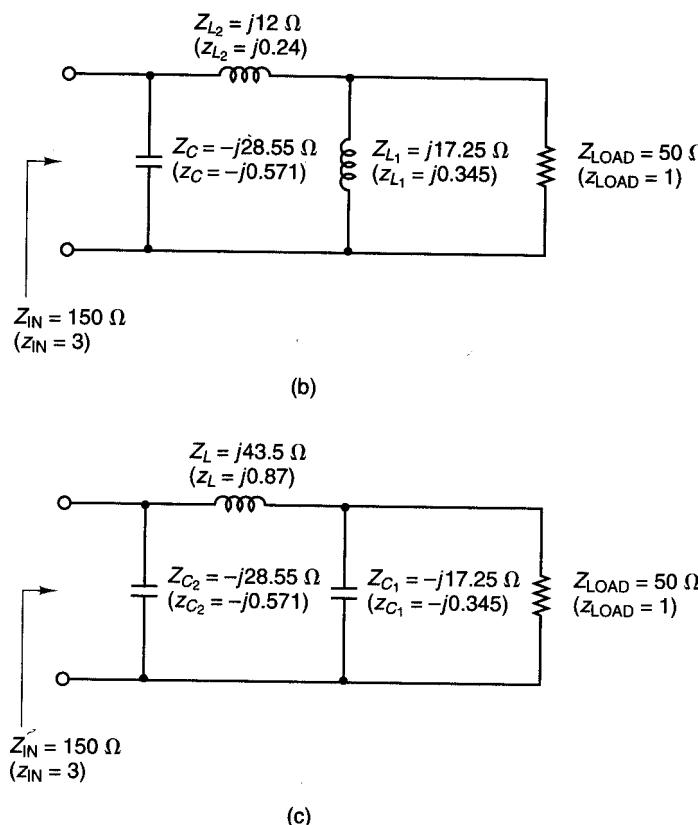
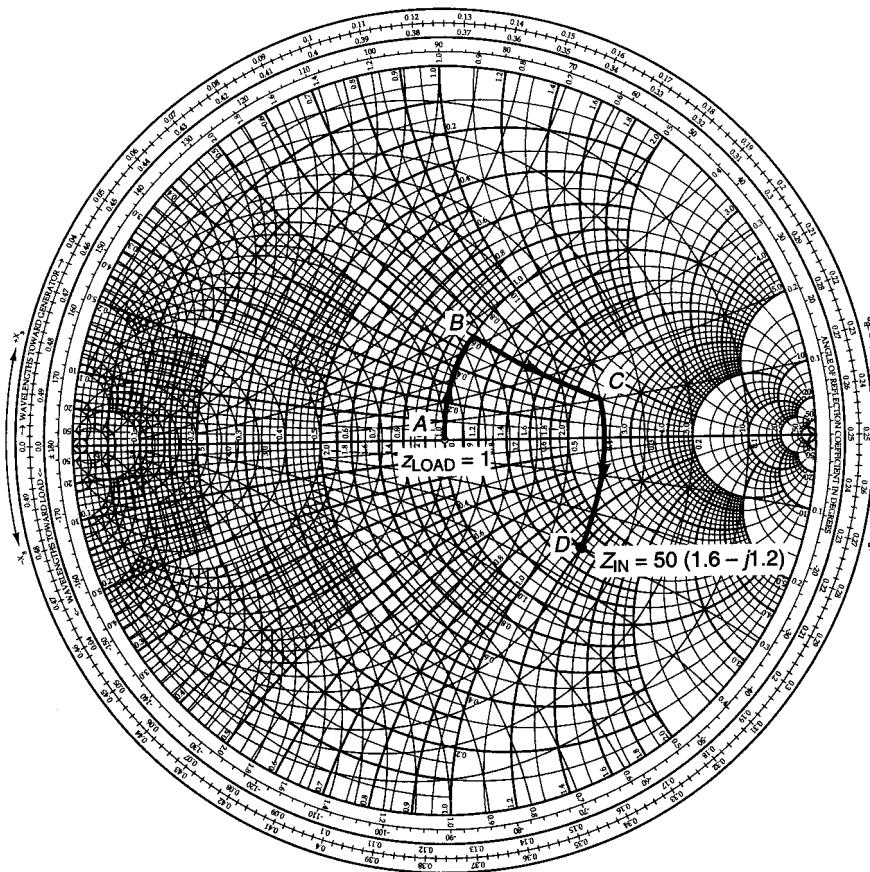
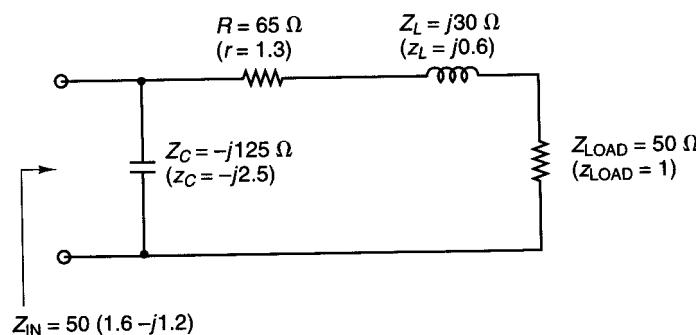


Figure 2.4.18 Continued



(a)



(b)

Figure 2.4.19 (a) Matching of $z_{LOAD} = 1$ to $z_{IN} = 1.6 - j1.2$; (b) the matching circuit.

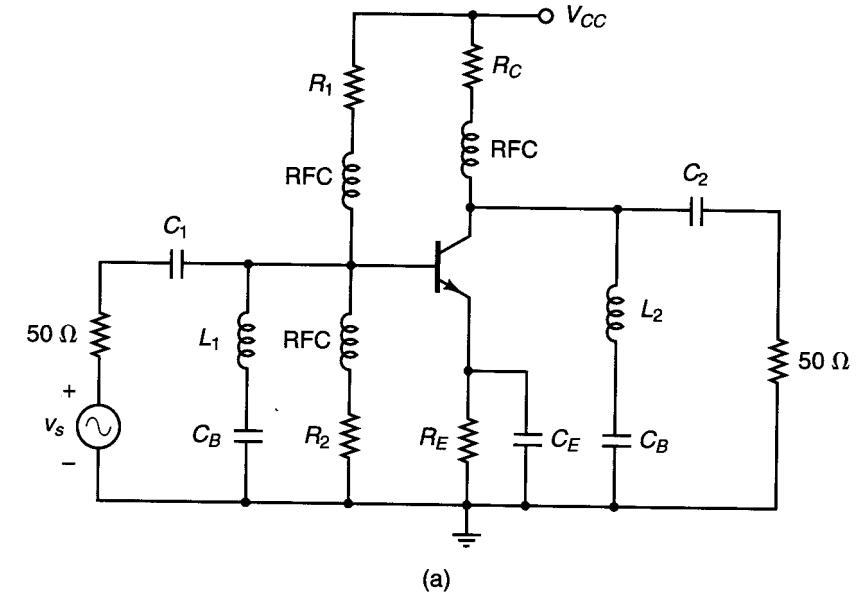


Figure 2.4.20 (a) A discrete microwave amplifier; (b) the dc model; (c) the ac model.

input source. The dc model of the amplifier is shown in Fig. 2.4.20b, and the ac model in Fig. 2.4.20c.

2.5 MICROSTRIP MATCHING NETWORKS

Microstrip lines

Microstrip lines are used extensively in building microwave transistor amplifiers because they are easily fabricated using printed-circuit techniques. Network interconnections and the placement of lumped and transistor devices

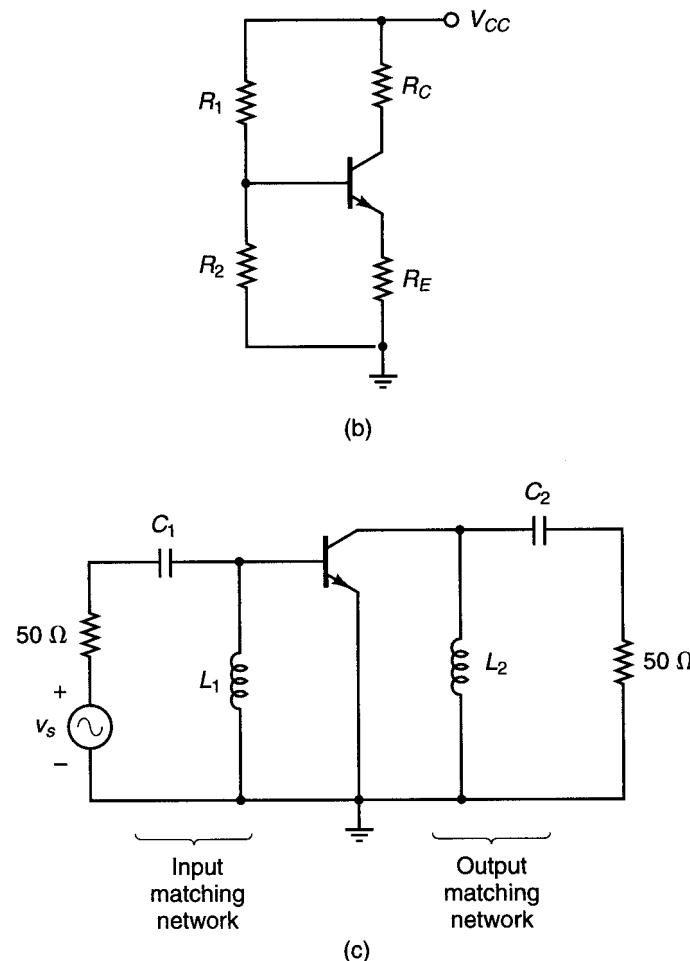


Figure 2.4.20 Continued

are easily made on its metal surface. The superior performance characteristics of the microstrip line make it one of the most important mediums of transmission in microwave transistor amplifiers and in microwave integrated-circuit technology.

A microstrip line is, by definition, a transmission line consisting of a strip conductor and a ground plane separated by a dielectric medium. Figure 2.5.1

illustrates the microstrip geometry. The dielectric material serves as a substrate and is sandwiched between the strip conductor and the ground plane. The relative dielectric constant of the substrate, ϵ_r , and ϵ are related by $\epsilon = \epsilon_r \epsilon_0$, where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m. Some typical dielectric substrates are RT/Duroid® (a trademark of Rogers Corporation, Chandler, Arizona), which is available with several values of ϵ_r (e.g., $\epsilon = 2.23\epsilon_0$, $\epsilon = 6\epsilon_0$, $\epsilon = 10.5\epsilon_0$, etc.); quartz ($\epsilon = 3.7\epsilon_0$), alumina ($\epsilon = 9\epsilon_0$); and Epsilam-10® ($\epsilon = 10\epsilon_0$).

The electromagnetic field lines in the microstrip are not contained entirely in the substrate. Therefore, the propagating mode in the microstrip is not a pure transverse electromagnetic mode (TEM mode) but a quasi-TEM. Assuming a quasi-TEM mode of propagation in the microstrip line, the phase velocity is given by

$$v_p = \frac{c}{\sqrt{\epsilon_{eff}}} \quad (2.5.1)$$

where c is the speed of light (i.e., 3×10^8 m/s) and ϵ_{eff} is the effective relative dielectric constant of the microstrip. The effective relative dielectric constant of the microstrip is related to the relative dielectric constant of the dielectric substrate and also takes into account the effect of the external electromagnetic fields (i.e., fringing effects must be considered).

Since $Z_o = \sqrt{L/C}$ and $v_p = 1/\sqrt{LC}$, the characteristic impedance of the microstrip line can be expressed in the form

$$Z_o = \frac{1}{v_p C} \quad (2.5.2)$$

where C is the capacitance per unit length of the microstrip. The wavelength in the microstrip line is given by

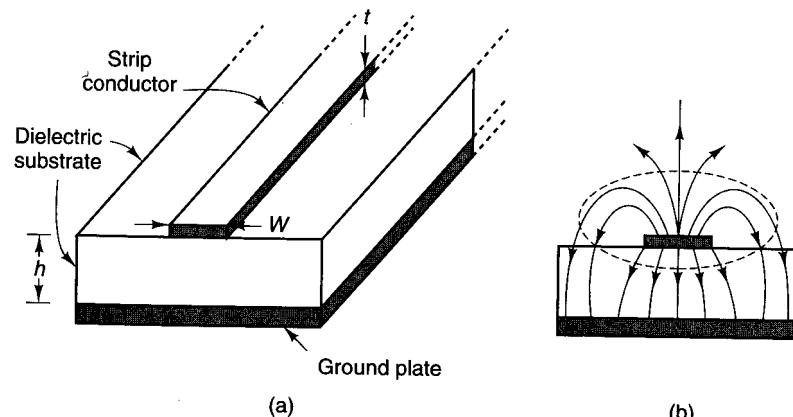


Figure 2.5.1 Microstrip geometry and field configuration. In (b), the solid lines represent electric field lines and the dashed line represents a magnetic field line.

$$\lambda = \frac{v_p}{f} = \frac{c}{f\sqrt{\epsilon_{eff}}} = \frac{\lambda_0}{\sqrt{\epsilon_{eff}}} \quad (2.5.3)$$

where λ_0 is the free-space wavelength.

As seen from (2.5.1), (2.5.2), and (2.5.3), the evaluation of v_p , Z_o , and λ in a microstrip line requires the evaluation of ϵ_{eff} and C . There are different methods for determining ϵ_{eff} and C and, of course, closed-form expressions are of great importance in microstrip-line design. The evaluation of ϵ_{eff} and C based on a quasi-TEM mode is accurate for design purposes at lower microwave frequencies. However, at higher microwave frequencies the longitudinal components of the electromagnetic fields are significant and the quasi-TEM assumption is no longer valid.

A useful set of relations for the characteristic impedance, assuming zero or negligible thickness of the strip conductor (i.e., $t/h < 0.005$), is as follows [2.1]:

For $W/h \leq 1$:

$$Z_o = \frac{60}{\sqrt{\epsilon_{eff}}} \ln\left(8 \frac{h}{W} + 0.25 \frac{W}{h}\right) \quad (2.5.4)$$

where

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[\left(1 + 12 \frac{h}{W}\right)^{-1/2} + 0.04 \left(1 - \frac{W}{h}\right)^2 \right] \quad (2.5.5)$$

For $W/h \geq 1$:

$$Z_o = \frac{120\pi/\sqrt{\epsilon_{eff}}}{W/h + 1.393 + 0.667 \ln(W/h + 1.444)} \quad (2.5.6)$$

where

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + 12 \frac{h}{W}\right)^{-1/2} \quad (2.5.7)$$

Plots of the characteristic impedance, as well as the normalized wavelength, as a function of W/h are shown in Figs. 2.5.2 and 2.5.3.

Based on the results in (2.5.3), (2.5.5), and (2.5.7) and/or in experimental data, the wavelength in the microstrip line, assuming zero or negligible thickness (i.e., $t/h \leq 0.005$) for the strip conductor, is given by the following relations [2.3]:

For $W/h \geq 0.6$:

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_r}} \left[\frac{\epsilon_r}{1 + 0.63(\epsilon_r - 1)(W/h)^{0.1255}} \right]^{1/2} \quad (2.5.8)$$

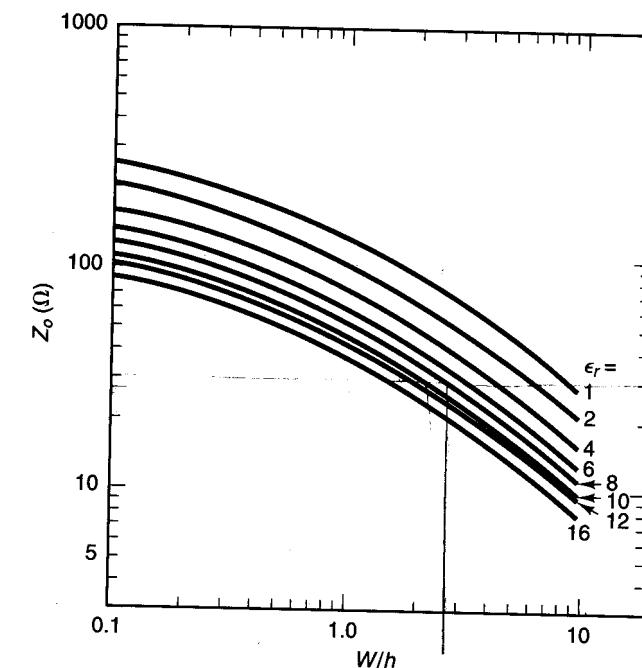


Figure 2.5.2 Characteristic impedance of the microstrip line versus W/h . (From H. Sobol [2.2]; copyright 1971, IEEE; reproduced with permission of IEEE.)

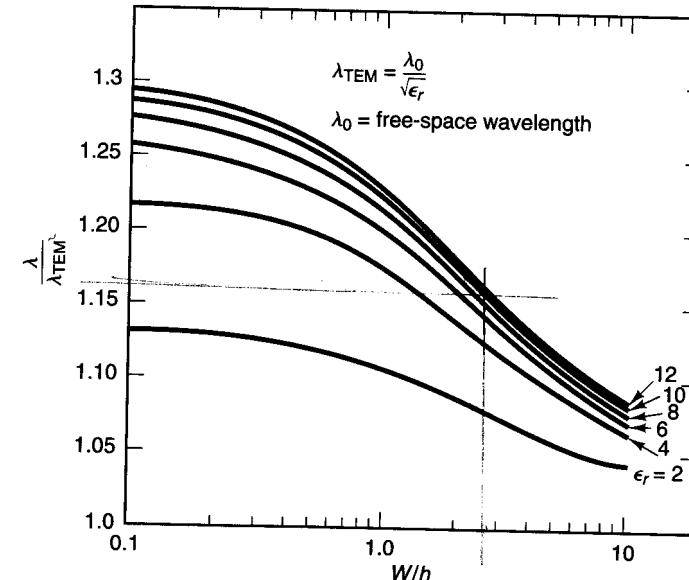


Figure 2.5.3 Normalized wavelength of the microstrip line versus W/h . (From H. Sobol [2.2]; copyright 1971, IEEE; reproduced with permission of IEEE.)

For $W/h < 0.6$:

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_r}} \left[\frac{\epsilon_r}{1 + 0.6(\epsilon_r - 1)(W/h)^{0.0297}} \right]^{1/2} \quad (2.5.9)$$

For design purposes, a set of equations relating Z_o and ϵ_r to the ratio W/h of the microstrip line is desirable. Assuming zero or negligible thickness of the strip conductor (i.e., $t/h \leq 0.005$), the expressions are as follows [2.2]:

For $W/h \leq 2$:

$$\frac{W}{h} = \frac{8e^A}{e^{2A} - 2} \quad (2.5.10)$$

For $W/h \geq 2$:

$$\frac{W}{h} = \frac{2}{\pi} \left\{ B - 1 - \ln(2B - 1) + \frac{\epsilon_r - 1}{2\epsilon_r} \left[\ln(B - 1) + 0.39 - \frac{0.61}{\epsilon_r} \right] \right\} \quad (2.5.11)$$

where

$$A = \frac{Z_o}{60} \sqrt{\frac{\epsilon_r + 1}{2}} + \frac{\epsilon_r - 1}{\epsilon_r + 1} \left(0.23 + \frac{0.11}{\epsilon_r} \right)$$

and

$$B = \frac{377\pi}{2Z_o \sqrt{\epsilon_r}}$$

The zero or negligible thickness formulas given in (2.5.4) to (2.5.11) can be modified to include the thickness of the strip conductor. The first-order effect of a strip conductor of finite thickness t is to increase the capacitance. Therefore, an approximate correction is made by replacing the strip width W by the effective width W_{eff} . The following relations for W_{eff}/h are useful when $t < h$ and $t < W/2$:

For $W/h \geq 1/2\pi$:

$$\frac{W_{\text{eff}}}{h} = \frac{W}{h} + \frac{t}{\pi h} \left(1 + \ln \frac{2h}{t} \right)$$

For $W/h \leq 1/2\pi$:

$$\frac{W_{\text{eff}}}{h} = \frac{W}{h} + \frac{t}{\pi h} \left(1 + \ln \frac{4\pi W}{t} \right)$$

The restrictions $t < h$ and $t < W/2$ are usually satisfied since for dielectric substrates a typical thickness is $t = 0.002$ in.

Various substrate materials are available for the construction of microstrip lines, with practical values of ϵ_r ranging from 2 to 10. The substrate material comes plated on both sides with copper, and an additional layer of gold

plating on top of the cooper is usually added after the circuit pattern is etched in order to prevent oxidation. Typical plating thickness of copper is from 1/2 mils to 2 mils (1 inch = 1000 mils).

The value of ϵ_r and the dielectric thickness (h) determine the width of the microstrip line for a given Z_o . These parameters also determine the speed of propagation in the line, and consequently its length. Typical dielectric thickness are 25, 30, 40, 50, and 100 mils.

Example 2.5.1

A microstrip material with $\epsilon_r = 10$ and $h = 1.016$ mm is used to build a transmission line. Determine the width for the microstrip transmission line to have a characteristic impedance of 50Ω . Also determine the wavelength and the effective relative dielectric constant of the microstrip line.

Solution. Figures 2.5.2 and 2.5.3 can be used to obtain the approximate values of W , λ , and ϵ_{ff} . From Fig. 2.5.2, with $Z_o = 50 \Omega$ and $\epsilon_r = 10$, the value of W/h is approximately 1. Hence,

$$W = h = 1.016 \text{ mm} = 0.1016 \text{ cm}$$

Sometimes W and h are expressed in mils. Since 1 inch = 1000 mils and 1 inch = 2.54 cm, it follows that the width and height in mils are

$$W = h = 0.1016 \left(\frac{1000}{2.54} \right) = 40 \text{ mils}$$

From Fig. 2.5.3, with $W/h = 1$ and $\epsilon_r = 10$, it follows that the value of $\lambda/\lambda_{\text{TEM}}$ is approximately 1.23, or

$$\lambda = 1.23\lambda_{\text{TEM}} = 1.23 \frac{\lambda_0}{\sqrt{\epsilon_r}} = \frac{1.23}{\sqrt{10}} \lambda_0 = 0.389\lambda_0$$

Hence [see (2.5.3)],

$$\epsilon_{\text{ff}} = \left(\frac{1}{0.389} \right)^2 = 6.61$$

More accurate values for W , λ , and ϵ_{ff} can be obtained using (2.5.8) to (2.5.11). From (2.5.10),

$$A = \frac{50}{60} \sqrt{\frac{10 + 1}{2}} + \frac{10 - 1}{10 + 1} \left(0.23 + \frac{0.11}{10} \right) = 2.1515$$

and

$$\frac{W}{h} = \frac{8e^{2.1515}}{e^{2(2.1515)} - 2} = 0.9563$$

Then

$$W = 0.9563(40) = 38.2 \text{ mils}$$

From (2.5.8),

$$\lambda = \frac{\lambda_0}{\sqrt{10}} \left[\frac{10}{1 + 0.63(10 - 1)(0.9563)^{0.1255}} \right]^{1/2} = 0.387\lambda_0$$

and

$$\epsilon_{ff} = \left(\frac{1}{0.387} \right)^2 = 6.68$$

The width and ϵ_{ff} of several 50-Ω microstrip lines for various values of ϵ_r and h were calculated and are tabulated in Fig. 2.5.4. For example, for a microstrip line having an alumina substrate with $h = 25$ mils and $\epsilon_r = 9.6$, it follows from Fig. 2.5.4 that $W = 24.7$ mils for $Z_o = 50 \Omega$ and $\epsilon_{ff} = 6.46$.

The formulas presented thus far are valid at frequencies where the quasi-TEM assumption can be made. When the quasi-TEM assumption is not valid, ϵ_{ff} and Z_o are functions of frequency and, therefore, the microstrip line becomes dispersive. The phase velocity of the microstrip line decreases with increasing frequency. Therefore, $\epsilon_{ff}(f)$ increases with frequency. Also, the characteristic impedance of the microstrip line increases with frequency, and it follows that the effective width $W_{eff}(f)$ decreases.

The frequency below which dispersion may be neglected is given by

$$f_o(\text{GHz}) = 0.3 \sqrt{\frac{Z_o}{h\sqrt{\epsilon_r - 1}}}$$

where h must be expressed in centimeters.

An analytical expression that shows the effect of dispersion in $\epsilon_{ff}(f)$ is [2.1]

$$\epsilon_{ff}(f) = \epsilon_r - \frac{\epsilon_r - \epsilon_{ff}}{1 + G(f/f_p)^2} \quad (\text{f in GHz})$$

	$\epsilon_r = 2.23$	4.54	6	9.6	10	30
$h = 25$ mils	$W = 76.4$ mils	46.7	37.5	24.7	23.8	6.01
	$\epsilon_{ff} = 1.91$	3.42	4.33	6.46	6.68	17.7
30	$W = 91.7$	56.1	45.0	29.7	28.5	7.21
	$\epsilon_{ff} = 1.91$	3.42	4.33	6.46	6.68	17.7
40	$W = 122.2$	74.7	60.0	39.6	38.0	9.6
	$\epsilon_{ff} = 1.91$	3.43	4.34	6.47	6.69	17.8
50	$W = 152.8$	93.4	75.0	49.4	47.6	12.0
	$\epsilon_{ff} = 1.91$	3.43	4.34	6.48	6.71	17.8
100	$W = 305.6$	186.8	150.1	98.9	95.1	24.1
	$\epsilon_{ff} = 1.91$	3.45	4.37	6.55	6.78	18.1

Figure 2.5.4 Width and ϵ_{ff} of microstrip lines for $Z_o = 50 \Omega$ for various ϵ_r and h values.

where

$$f_p = \frac{Z_o}{8\pi h} \quad (h \text{ in cm})$$

and

$$G = 0.6 + 0.009Z_o$$

Observe that when $f_p \gg f$, then $\epsilon_{ff}(f) \approx \epsilon_{ff}$. In other words, high-impedance lines on thin substrates are less dispersive.

The expression for the dispersion in Z_o is [2.1]

$$Z_o(f) = \frac{377h}{W_{eff}(f)\sqrt{\epsilon_{ff}}}$$

where

$$W_{eff}(f) = W + \frac{W_{eff}(0) - W}{1 + (f/f_p)^2}$$

and

$$W_{eff}(0) = \frac{377h}{Z_o(0)\sqrt{\epsilon_{ff}(0)}}$$

Another characteristic of the microstrip line is its attenuation. The attenuation constant is a function of the microstrip geometry, the electrical properties of the dielectric substrate and the conductors, and the frequency.

There are two types of losses in a microstrip line: a dielectric substrate loss and the ohmic skin loss in the conductors. The losses can be expressed as a loss per unit length along the microstrip line in terms of the attenuation factor α . Since the power carried by a wave traveling in the positive direction in a quasi-TEM mode is given by

$$P^+(z) = \frac{1}{2} \frac{|V^+|^2}{Z_o} e^{-2az} = P_0 e^{-2az} \quad (2.5.12)$$

where $P_0 = |V^+|^2/2Z_o$ is the power at $z = 0$. Then, from (2.5.12), we can write

$$\alpha = \frac{-dP(z)/dz}{2P(z)} = \alpha_d + \alpha_c$$

where α_d is the dielectric loss factor and α_c the conduction loss factor.

A useful set of expressions for calculating α_d is [2.1]

For a dielectric with low losses:

$$\alpha_d = 27.3 \frac{\epsilon_r}{\sqrt{\epsilon_{ff}}} \frac{\epsilon_{ff} - 1}{\epsilon_r - 1} \frac{\tan \delta}{\lambda_0} \quad \frac{\text{dB}}{\text{cm}} \quad (2.5.13)$$

where the loss tangent δ is given by

$$\tan \delta = \frac{\sigma}{\omega \epsilon}$$

For a dielectric with high losses:

$$\alpha_d = 4.34 \frac{\epsilon_{ff} - 1}{\sqrt{\epsilon_{ff}(\epsilon_r - 1)}} \left(\frac{\mu_0}{\epsilon_0} \right)^{1/2} \sigma \quad \frac{\text{dB}}{\text{cm}} \quad (2.5.14)$$

In (2.5.13) and (2.5.14), σ is the conductivity of the dielectric and $\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$.

A set of expressions for calculating α_c is [2.1]

For $W/h \rightarrow \infty$:

$$\alpha_c = \frac{8.68}{Z_o W} R_s$$

where

$$R_s = \sqrt{\frac{\pi f \mu_0}{\sigma}}$$

For $W/h \leq 1/2\pi$:

$$\alpha_c = \frac{8.68 R_s P}{2\pi Z_o h} \left[1 + \frac{h}{W_{\text{eff}}} + \frac{h}{\pi W_{\text{eff}}} \left(\ln \frac{4\pi W}{t} + \frac{t}{W} \right) \right]$$

For $1/2\pi < W/h \leq 2$:

$$\alpha_c = \frac{8.68 R_s}{2\pi Z_o h} PQ$$

For $W/h \geq 2$:

$$\alpha_c = \frac{8.68 R_s Q}{Z_o h} \left\{ \frac{W_{\text{eff}}}{h} + \frac{2}{\pi} \ln \left[2\pi e \left(\frac{W_{\text{eff}}}{2h} + 0.94 \right) \right] \right\}^{-2} \left[\frac{W_{\text{eff}}}{h} + \frac{W_{\text{eff}}/\pi h}{(W_{\text{eff}}/2h) + 0.94} \right]$$

where

$$P = 1 - \left(\frac{W_{\text{eff}}}{4h} \right)^2$$

and

$$Q = 1 + \frac{h}{W_{\text{eff}}} + \frac{h}{\pi W_{\text{eff}}} \left(\ln \frac{2h}{t} - \frac{t}{h} \right)$$

In dielectric substrates, the dielectric losses are normally smaller than conductor losses. However, dielectric losses in silicon substrates can be of the same order or larger than conductor losses.

The quality factor Q of a microstrip line is calculated from

$$Q = \frac{\beta}{2\alpha}$$

where

$$\beta = \frac{2\pi}{\lambda}$$

and α is the total loss. Therefore,

$$Q = \frac{\pi}{\lambda \alpha}$$

or in decibels we can write

$$\begin{aligned} Q &= \frac{8.686\pi}{\lambda \alpha} \quad \text{dB} \\ &= \frac{27.3}{\alpha} \quad \frac{\text{dB}}{\lambda} \end{aligned}$$

where we used the fact that 1 dB = 8.686 nepers.

A microstrip line also has radiation losses. The effect of radiation losses can be accounted for in terms of the radiation quality factor Q_r given by [2.1]

$$Q_r = \frac{Z_o}{480\pi(h/\lambda_0)F}$$

where

$$F = \frac{\epsilon_{ff}(f) + 1}{\epsilon_{ff}(f)} - \frac{(\epsilon_{ff}(f) - 1)^2}{2[\epsilon_{ff}(f)]^{2/3}} \ln \frac{\sqrt{\epsilon_{ff}(f)} + 1}{\sqrt{\epsilon_{ff}(f)} - 1}$$

is known as the *radiation factor*.

The total Q , called Q_T , of a microstrip resonator can be expressed as

$$\frac{1}{Q_T} = \frac{1}{Q_c} + \frac{1}{Q_d} + \frac{1}{Q_r}$$

where Q_d and Q_c are the quality factors of the dielectric (i.e., $Q_d = \pi/\lambda \alpha_d$) and conductor (i.e., $Q_c = \pi/\lambda \alpha_c$), respectively.

Microstrip lines can be etched to a typical accuracy of ± 1 mils for width greater than 5 mils. Circuit layouts using microstrips are easily done using a layout CAD program. For example, the Hewlett-Packard MDS program has a circuit layout capability that generates the circuit pattern from the circuit schematic. If a layout program is not available, a typical circuit layout first

involves making a scaled drawing of the circuit. A typical scale factor is 10. For example, a width of 24.7 mils in alumina microstrip with $h = 25$ mils produces a line with $Z_o = 50 \Omega$. A 1-to-1 scale drawing of a circuit using these lines would be almost impossible to obtain. Thus, it is common practice to draw the circuit in graph paper using a 10-to-1 scale. On a 10-to-1 scale, the previous $50-\Omega$ line would have a width of 247 mils, which is certainly a drawable quantity.

Electronic components come in various sizes. Some standard component size are as follows:

	Width (mils)	Length (mils)	Height (mils)
Transistors	100	120	20
Diodes	100	120	20
Chip capacitors	50	80	20
Chip coils (RFCs)	80	100	50
Chip resistors	50	80	15

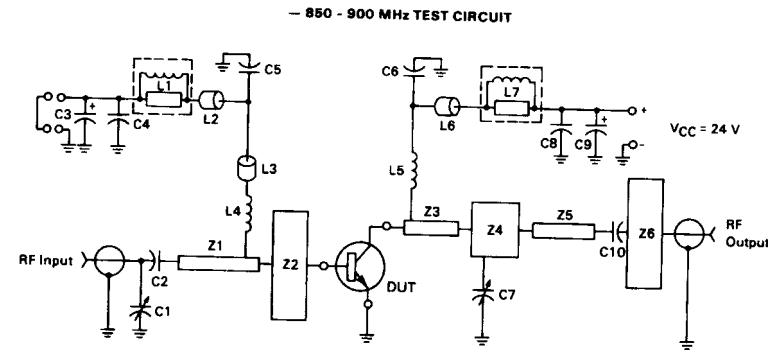
Lumped components and microstrip lines are drawn in graph paper, scaled by a factor of 10, to obtain the complete circuit layout. For example, a typical chip resistor in the scaled layout would have dimensions of $W = 500$ mils by $L = 800$ mils. Once the circuit drawing is done on a 10-to-1 scale, it is brought back to a 1-to-1 scale through photographic methods in which the reduction process is exact.

Manufacturers commonly provide test circuits for their transistors, as well as the photomaster of the test circuit (see Fig. 2.5.5).

Design of Matching Networks

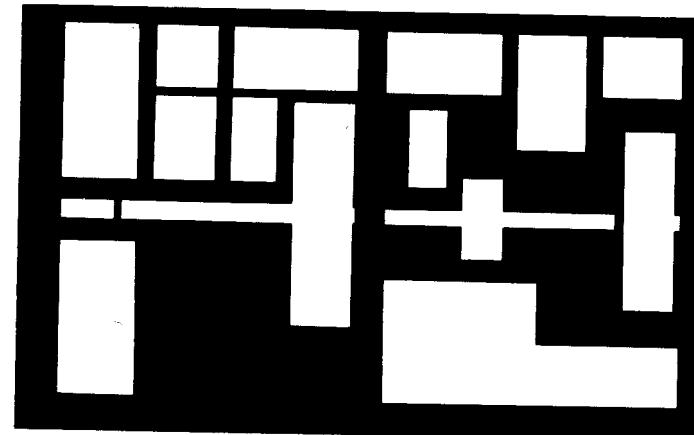
The impedance transforming properties of transmission lines can be used in the design of matching networks. A microstrip line can be used as a series transmission line, as an open-circuited stub, or as a short-circuited stub. In fact, a series microstrip line together with a short- or open-circuited shunt stub can transform a $50-\Omega$ resistor into any value of impedance.

The matching circuit configuration in which a short-circuited stub is connected in parallel with the load Z_L followed by a series microstrip transmission line is shown in Fig. 2.5.6a. The characteristic impedance of the microstrip lines is shown as 50Ω . Another way that is used to draw the schematic in Fig. 2.5.6a is shown in Fig. 2.5.6b. For comparison purposes, the same schematic drawn with $50-\Omega$ two-wire transmission lines is shown in



C1, C7 — Johanson 0.5 ~ 4.0 pF Giga-Trim
C2, C5, C6 — 91 pF Mini Underwood Micas
C3, C9 — 1.0 μ F Electrolytic
C4, C8 — 250 pF Unelco
C10 — 39 pF Mini Underwood
L1, L7 — 10 Turns Around 10 Ω 1/2 W Resistor
L2, L3, L6 — Ferrite Bead
L4, L5 — 4 Turns 26 AWG 0.1" ID
Z1, Z2, Z3, Z4, Z5, Z6 — Distributed Microstrip Elements (see photomask)
Board Material — Glass Teflon $\epsilon_r = 2.55$ $t = 0.031"$

— PHOTOMASTER FOR TEST FIXTURE



NOTE: The Printed Circuit Board shown is 75% of the original.

— 850-900 MHZ TEST CIRCUIT

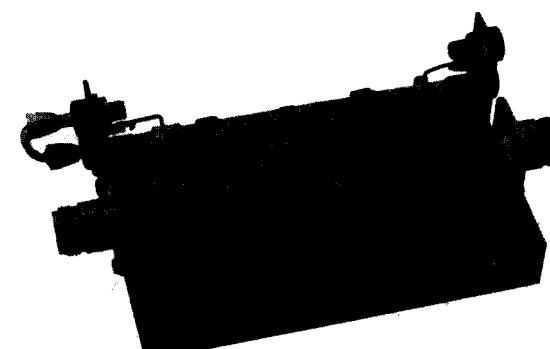


Figure 2.5.5 Test circuit, photomaster, and circuit construction for the MRF890 transistor. (From *Motorola RF Device Data*, Vol. 1, 6th edition; copyright of Motorola, used by permission.)

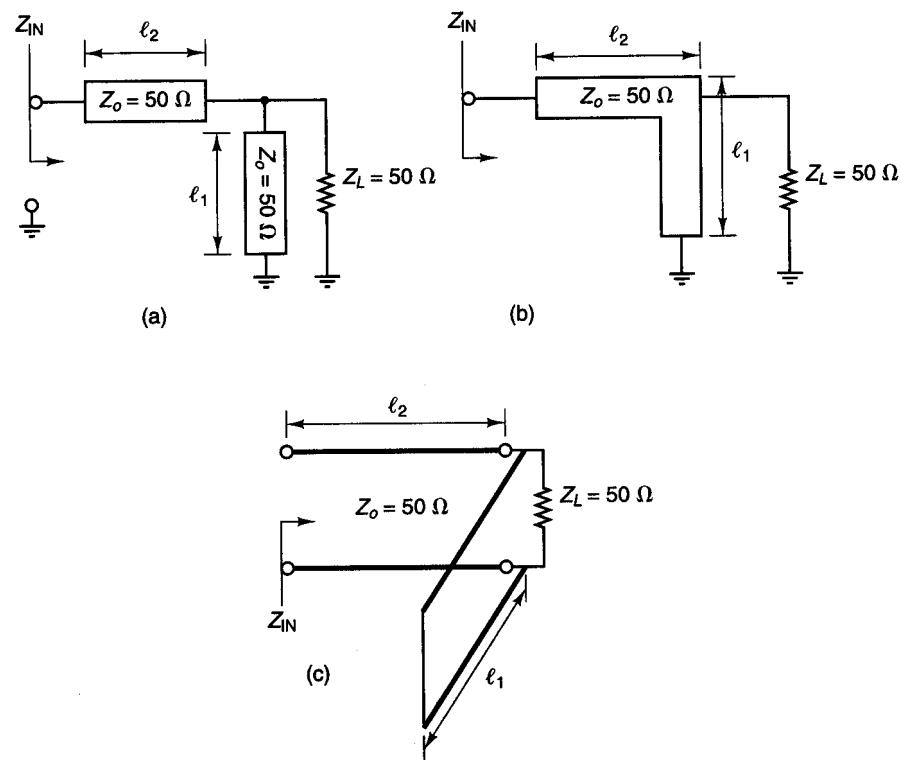


Figure 2.5.6 (a) A matching network using microstrip lines; (b) an alternative drawing; (c) schematic using two-wire transmission lines.

Fig. 2.5.6c. The shunt element in Fig. 2.5.6 could have been an open-circuited shunt stub instead of the short-circuited shunt stub shown in the figure. While the design procedure is discussed for a $50\text{-}\Omega$ load, it can be applied to an arbitrary load impedance.

The Smith chart design procedure for the matching circuit in Fig. 2.5.6a is now discussed. Since the admittance of the shunt stub adds to the load admittance, it is convenient to use the Y Smith chart to design the circuit. Figure 2.5.7a shows the matching circuit where $z_L = Z_L/Z_o = 50/50 = 1$ (or $y_L = 1/z_L = 1$). The normalized admittance of the shunt stub is written, for convenience, in the form $y_{sc} = jb_s$ ($b_s > 0$) for a capacitive susceptance, or $y_{sc} = -jb_s$ ($b_s > 0$) for an inductive susceptance. Using this notation, b_s is always positive and the length l_1 determines the value of $\pm jb_s$. The admittance y_x is given by

$$y_x = y_L + y_{sc} = 1 \pm jb_s$$

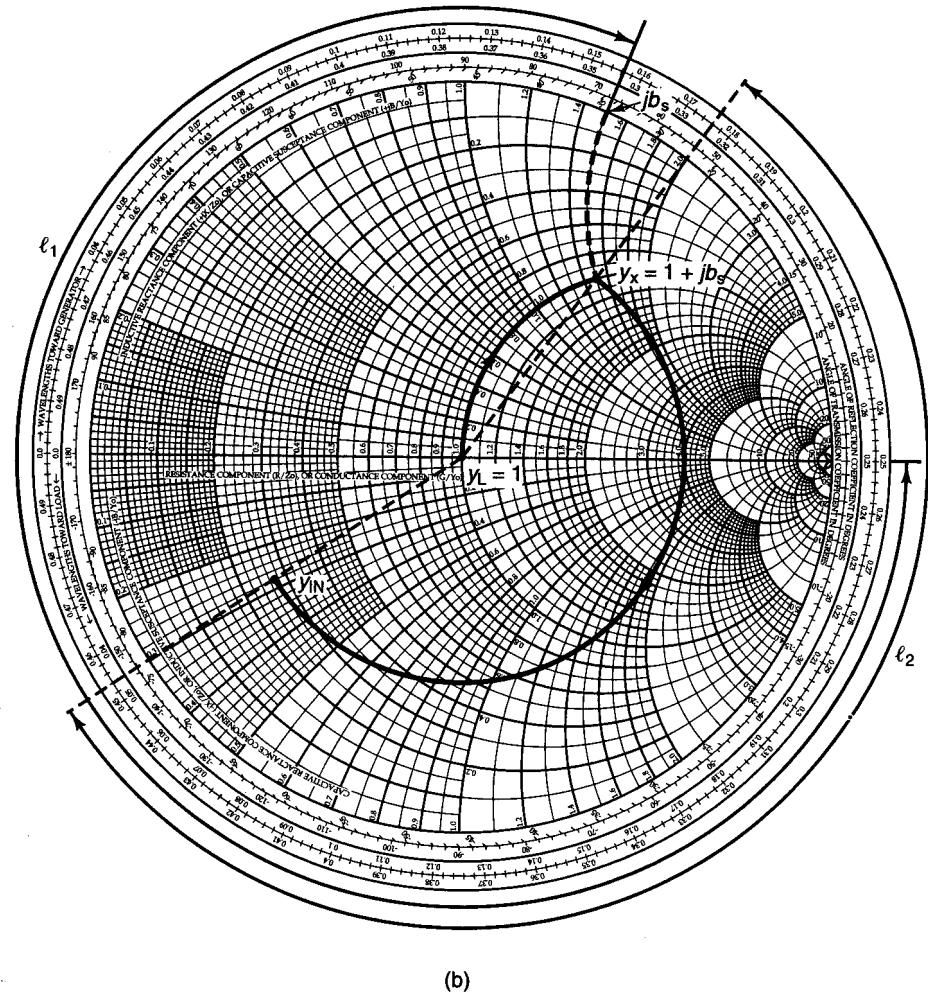
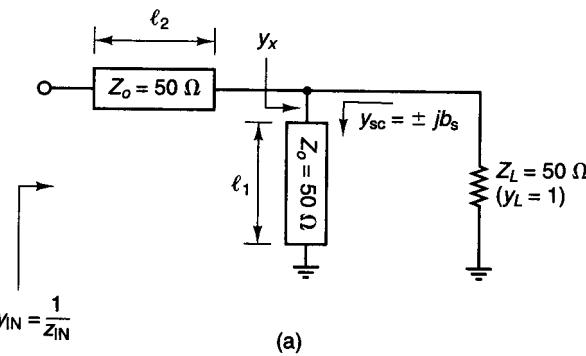


Figure 2.5.7 (a) A microstrip network to transform $y_L = 1$ to y_{IN} ; (b) the design in the Y Smith chart; with $y_{sc} = jb_s$ ($b_s > 0$); (c) the design in the Y Smith chart with $y_{sc} = -jb_s$ ($b_s > 0$).

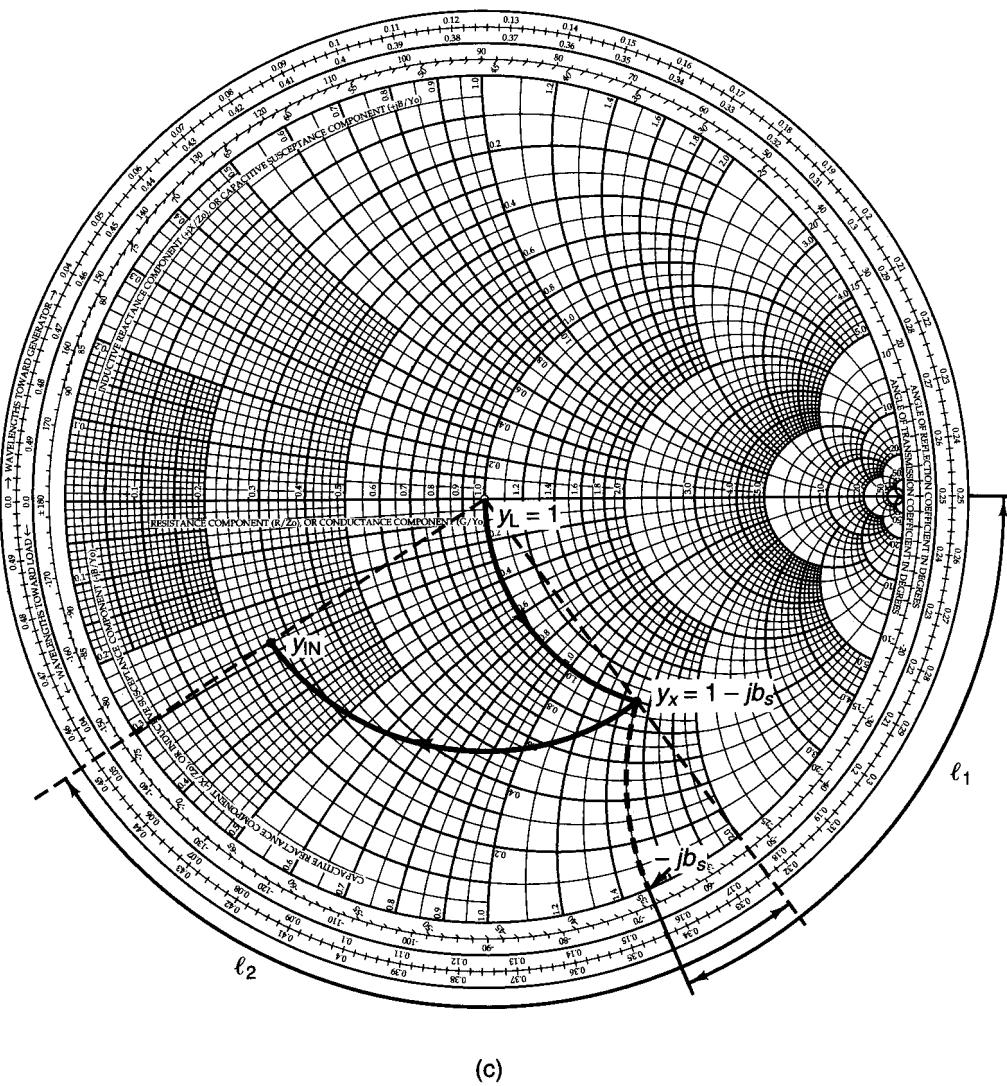


Figure 2.5.7 Continued

As shown in the Y Smith chart in Fig. 2.5.7b, the addition of $y_{sc} = jb_s$ to y_L produces a motion along the unity constant-conductance circle from $y_L = 1$ to $y_x = 1 + jb_s$. The design of the series transmission line of length l_2 is such that y_x is transformed to the admittance y_{IN} (a typical y_{IN} is shown in Fig. 2.5.7b). Consequently, the value of b_s must be selected so that y_x and y_{IN} are on a con-

stant $|\Gamma|$ circle. The readings of l_1 and l_2 in the Y Smith chart are also shown in Fig. 2.5.7b. Observe that l_1 is the length of the short-circuited stub that produces the admittance $y_{sc} = jb_s$. The value of l_1 is easily read from the edge of the Y Smith chart.

The matching design with $y_{sc} = -jb_s$ is illustrated in Fig. 2.5.7c. In this case, the shunt stub of length l_1 produces $y_{sc} = -jb_s$ and the series microstrip line of length l_2 is designed to change y_x to y_{IN} .

If an open-circuited shunt stub is used instead of the short-circuited shunt stub shown in Fig. 2.5.6, the design procedure is quite similar. In fact, the only difference is that the length l_1 in Figs. 2.5.7a and 2.5.7b is read starting from an open-circuited termination (i.e., starting from $y = 0$).

In Fig. 2.5.7a, the shunt stub simply behaves like either a shunt capacitor (see Fig. 2.5.7b with $y_{sc} = jb_s$) or a shunt inductor (see Fig. 2.5.7c with $y_{sc} = -jb_s$), changing $y_L = 1$ to the admittance $y_x = 1 \pm jb_s$. In fact, the same design procedure applies if the shunt stub is replaced by a lumped capacitor or inductor with admittance jb_s or $-jb_s$, respectively.

A practical design using the matching network topology in Fig. 2.5.6 is found in Example 2.5.2, Design 1.

The microstrip matching circuit shown in Fig. 2.5.8a can be designed to match an arbitrary load impedance Z_L to a 50Ω input impedance. This configuration resembles the one in Fig. 2.5.6. While in Fig. 2.5.6 the matching was from $Z_L = 50\Omega$ to an arbitrary Z_{IN} , in Fig. 2.5.8a the matching is from an arbitrary Z_L to $Z_{IN} = 50\Omega$. The design procedure in the Y Smith chart is shown in Fig. 2.5.8b, where the load admittance $y_L = 1/Z_L$ is transformed to $y_{IN} = 1/Z_{IN} = 1$. The length l_1 is selected so that y_L is transformed to the admittance $y_x = 1 + jb_s$ ($b_s > 0$). The motion from y_L to y_x is along a constant $|\Gamma|$ circle. Since $y_x = 1 + jb_s$, we design the short-circuited stub admittance to be $y_{sc} = -jb_s$ so that $y_{IN} = 1$. That is,

$$y_{IN} = y_x + y_{sc} = (1 + jb_s) - jb_s = 1$$

In other words, y_x must be located on the unit conductance circle at the location $y_x = 1 + jb_s$, and $y_{sc} = -jb_s$ cancels the imaginary part of y_x . The length l_2 (measured as shown in Fig. 2.5.8b) produces the short-circuited admittance $y_{sc} = -jb_s$.

The matching design with $y_{sc} = jb_s$ is illustrated in Fig. 2.5.8c. In this case, $y_x = 1 - jb_s$, so $y_{IN} = y_x + y_{sc} = (1 - jb_s) + jb_s = 1$.

Another practical microstrip matching circuit, shown in Fig. 2.5.9a, uses a series quarter-wave line with characteristic impedance Z_{o1} followed by a short-circuited shunt stub of length $\lambda/8$ or $3\lambda/8$ and characteristic impedance Z_{o2} to transform a 50Ω load to any value of input impedance. An open-circuited shunt stub can be used instead of the short-circuited shunt stub.

The design procedure for the matching circuit in Fig. 2.5.9a is as follows. Letting $Y_{IN} = G_{IN} + jB_{IN}$, where $B_{IN} > 0$, the quarter-wave transformer is used

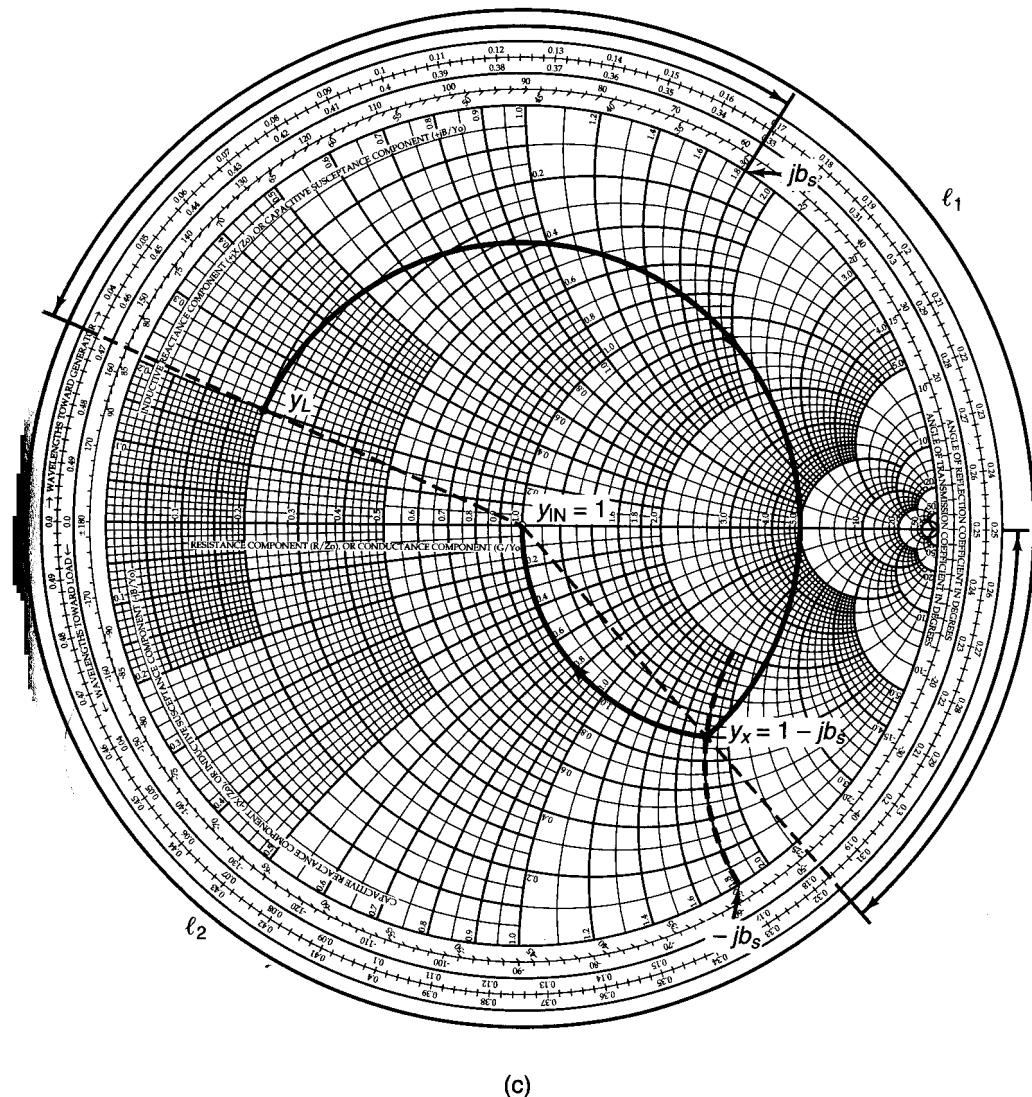
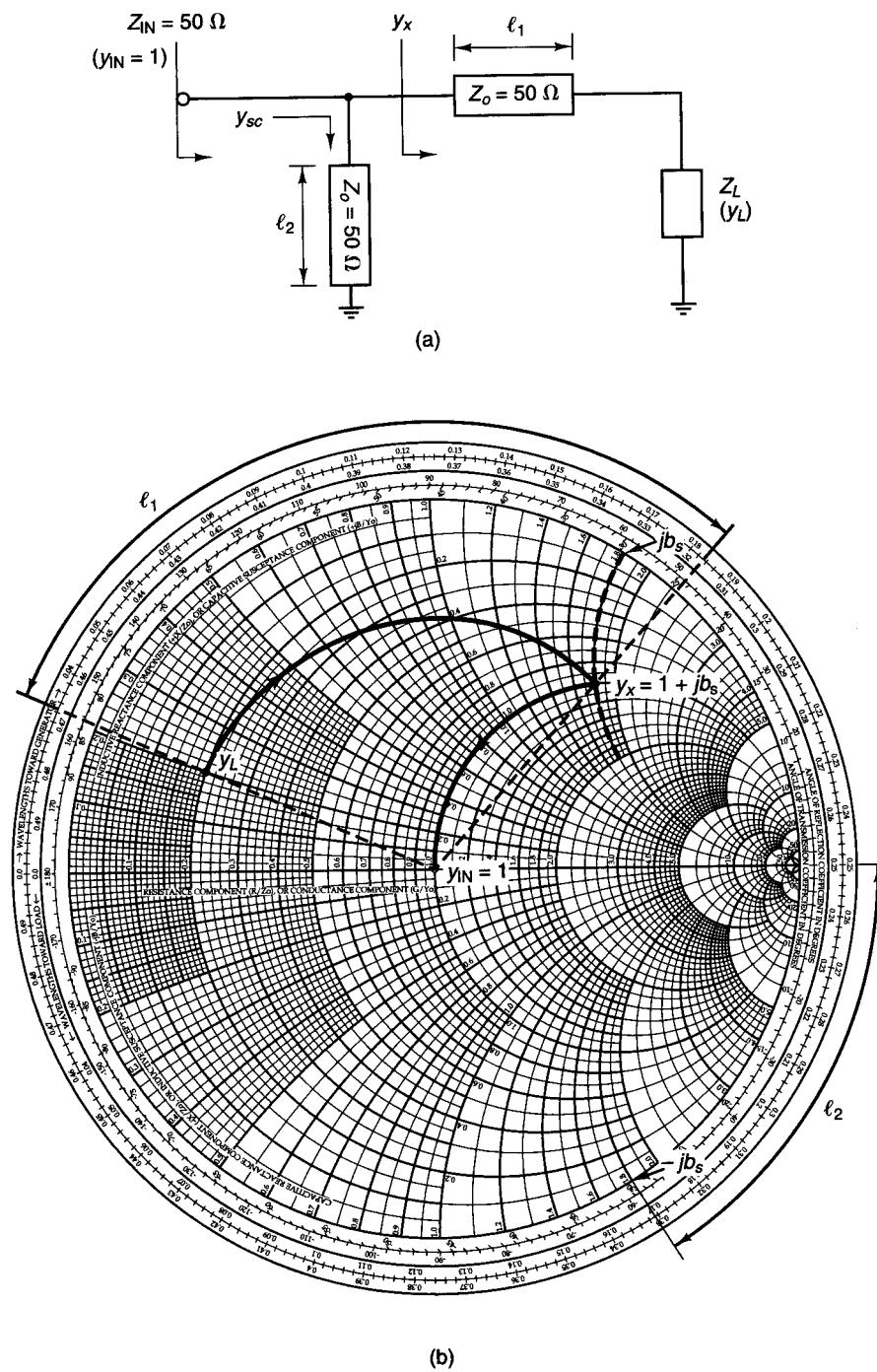


Figure 2.5.8 (a) A microstrip network to transform Z_{LOAD} to Z_{IN} ; (b) the design in the Y Smith chart with $y_{sc} = -jb_s$; (c) the design in the Y Smith chart with $y_{sc} = jb_s$.

to transform $Z_L = 50 \Omega$ to the input resistance $R_{IN} = 1/G_{IN}$. The short-circuited shunt stub is designed to produce the admittance jB_{IN} . Hence, the addition of G_{IN} and jB_{IN} produces the desired Y_{IN} .

The characteristic impedance of the quarter-wave line Z_{o1} is calculated using

$$Z_{o1} = \sqrt{Z_L R_{IN}} = \sqrt{50 R_{IN}} \quad (2.5.15)$$

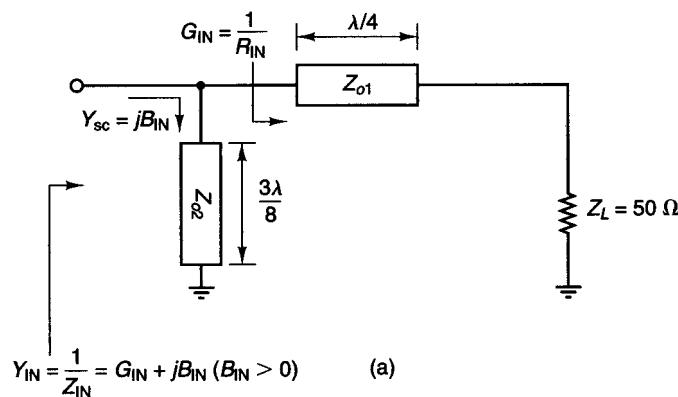


Figure 2.5.9 (a) A microstrip matching network; (b) admittance produced by a short-circuited stub of length $3\lambda/8$; (c) admittance produced by an open-circuited shunt stub of length $\lambda/8$.

Then, since a short-circuited shunt stub of length $3\lambda/8$ produces a shunt admittance of value $Y_{sc} = jY_{o2}$ (see Fig. 2.5.9b), we let jY_{o2} be equal to jB_{IN} . That is,

$$Y_{o2} = B_{IN}$$

or

$$Z_{o2} = \frac{1}{Y_{o2}} = \frac{1}{B_{IN}} \quad (2.5.16)$$

In conclusion, a characteristic impedance for the quarter-wave line Z_{o1} given by (2.5.15) produces the input conductance G_{IN} , and a characteristic impedance for the shunt stub Z_{o2} given by (2.5.16) produces the input susceptance B_{IN} . If an open-circuited shunt stub is used instead of the short-circuited shunt stub, then its length would have been $\lambda/8$ to produce $Y_{oc} = jY_{o2}$ (see Fig. 2.5.9c).

If the input admittance is given by $Y_{IN} = G_{IN} - jB_{IN}$, where $B_{IN} > 0$, as shown in Fig. 2.5.10a, Z_{o1} is calculated using (2.5.15), and a short-circuited shunt stub of length is $\lambda/8$ will produce $Y_{sc} = -jY_{o2}$ (see Fig. 2.5.10b). Therefore, letting $-jY_{o2}$ equal $-jB_{IN}$, we obtain

$$-jY_{o2} = -jB_{IN}$$

or

$$Z_{o2} = \frac{1}{Y_{o2}} = \frac{1}{B_{IN}}$$

If an open-circuited shunt stub is used instead of the short-circuited shunt stub, its length would have been $3\lambda/8$ to produce $Y_{oc} = -jY_{o2}$ (see Fig. 2.5.10c).

A practical design using the matching network topology in Fig. 2.5.9a is found in Example 2.5.2, Design 2.

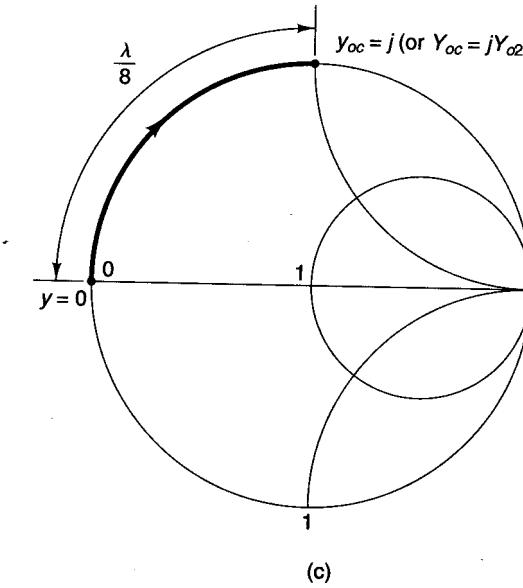
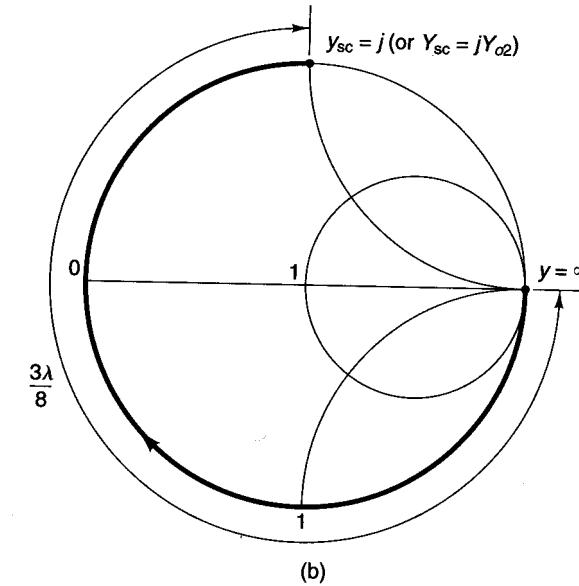


Figure 2.5.9 Continued

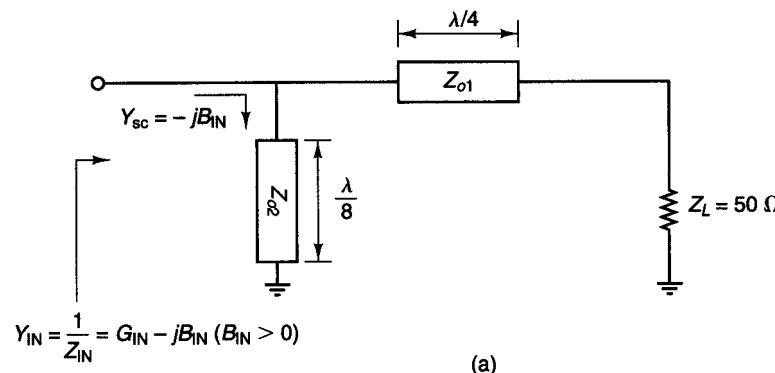


Figure 2.5.10 (a) A microstrip matching network; (b) admittance produced by a short-circuited stub of length $\lambda/8$; (c) admittance produced by an open-circuited shunt stub of length $3\lambda/8$.

The microstrip matching circuits shown in Fig. 2.5.9a and 2.5.10a can also be designed by using any practical value of Z_{o2} but setting the shunt stub length l_2 to produce the admittance of $Y_{sc} = \pm jB_{IN}$. The configuration is shown in Fig. 2.5.11.

The following examples illustrate the uses of microstrip lines in matching networks.

Example 2.5.2

Design two microstrip matching networks for the amplifier shown in Fig. 2.5.12. The normalized impedances and admittances associated with Γ_s and Γ_L can be read, to reasonable accuracy, from the ZY chart—namely,

$$y_s = \frac{1}{z_s} = \frac{1}{0.245 + j0.165} = 2.8 - j1.9$$

and

$$y_L = \frac{1}{z_L} = \frac{1}{0.325 + j0.83} = 0.4 - j1.05$$

In order to design the input matching network, we locate y_s in the Y Smith chart shown in Fig. 2.5.13a. The shortest length of microstrip line plus stub is obtained by using an open-circuited shunt stub of length 0.159λ to move from the origin (i.e., 50Ω) to point A on the Smith chart, and then using a transmission line length of 0.099λ to move from A to y_s .

Next, we locate y_L in Fig. 2.5.13b and follow a similar procedure. In this case, the shortest length of microstrip line plus stub is obtained by using a short-circuited shunt

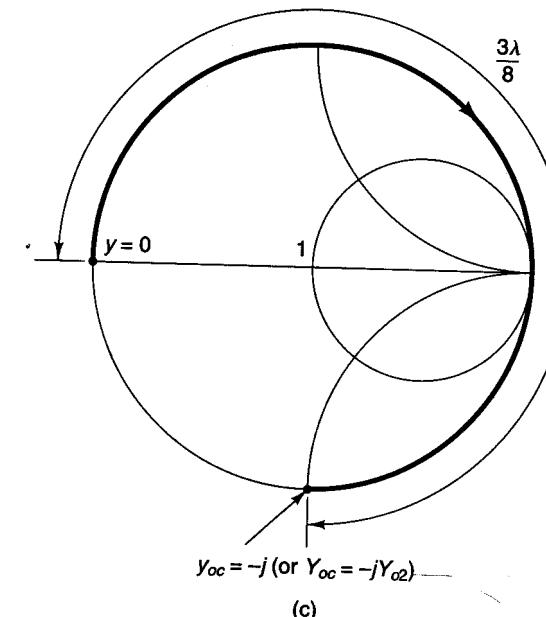
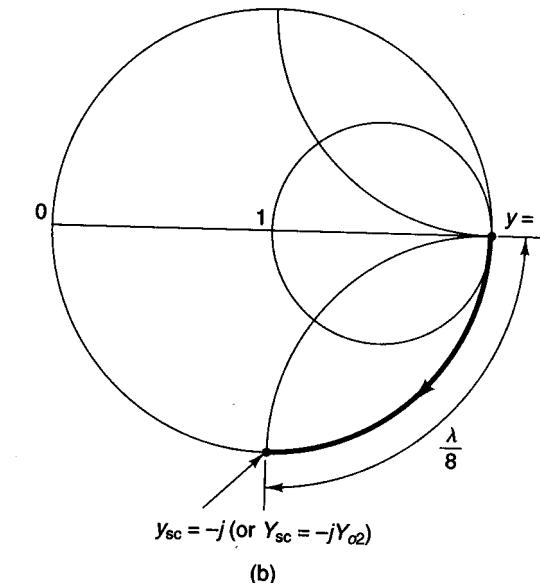


Figure 2.5.10 Continued

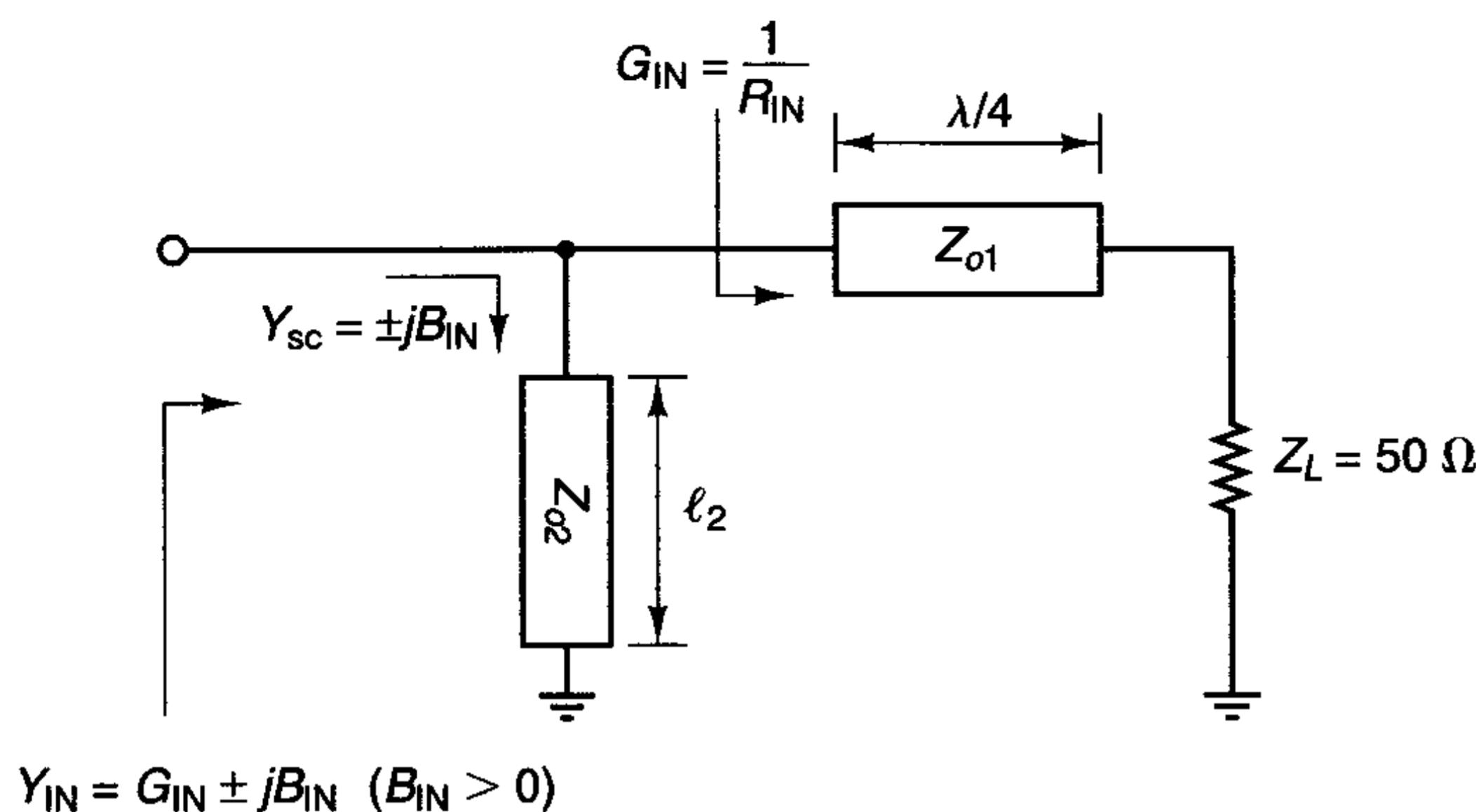


Figure 2.5.11 A microstrip matching circuit.

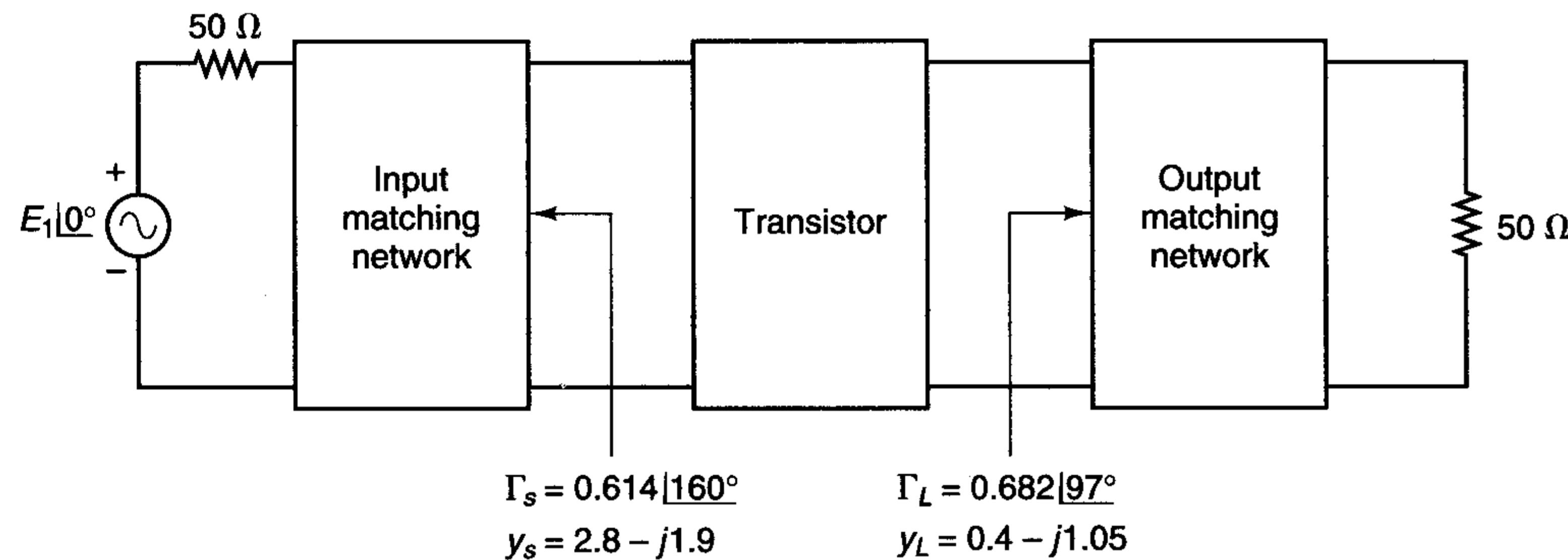


Figure 2.5.12 Amplifier block diagram.

stub of length 0.077λ to move from the origin to point *B*. Then a series transmission line of length 0.051λ is used to move from *B* to y_L .

The complete design, showing the transistor, the microstrip matching network, and the dc supply, is shown in Fig. 2.5.14. The characteristic impedance of all microstrip lines is 50Ω .

The capacitors C_A are coupling capacitors. Typical values for the chip capacitors C_A are 200 to 1000 pF, high- Q capacitor. The bypass capacitors C_B (i.e., chip capacitors, 50 to 500 pF) provide the ac short circuits for the 0.077λ and $\lambda/4$ short-circuited stubs. The $\lambda/4$ short-circuited stub, high-impedance line (denoted by $Z_o \gg$), provides the dc path for the base supply voltage. It also presents an open circuit to the ac signal at the base of the transistor. The narrowest practical line (i.e., large Z_o) should be used for the $\lambda/4$ short-circuited stub to avoid unwanted ac coupling. Typical dc bias circuits are shown in Figs. 3.9.2 and 3.9.4.

To minimize transition interaction between the shunt stubs and the series transmission lines, the shunt stubs are usually balanced along the series transmission line. A schematic of the amplifier using balanced shunt stubs is shown in Fig. 2.5.15. The

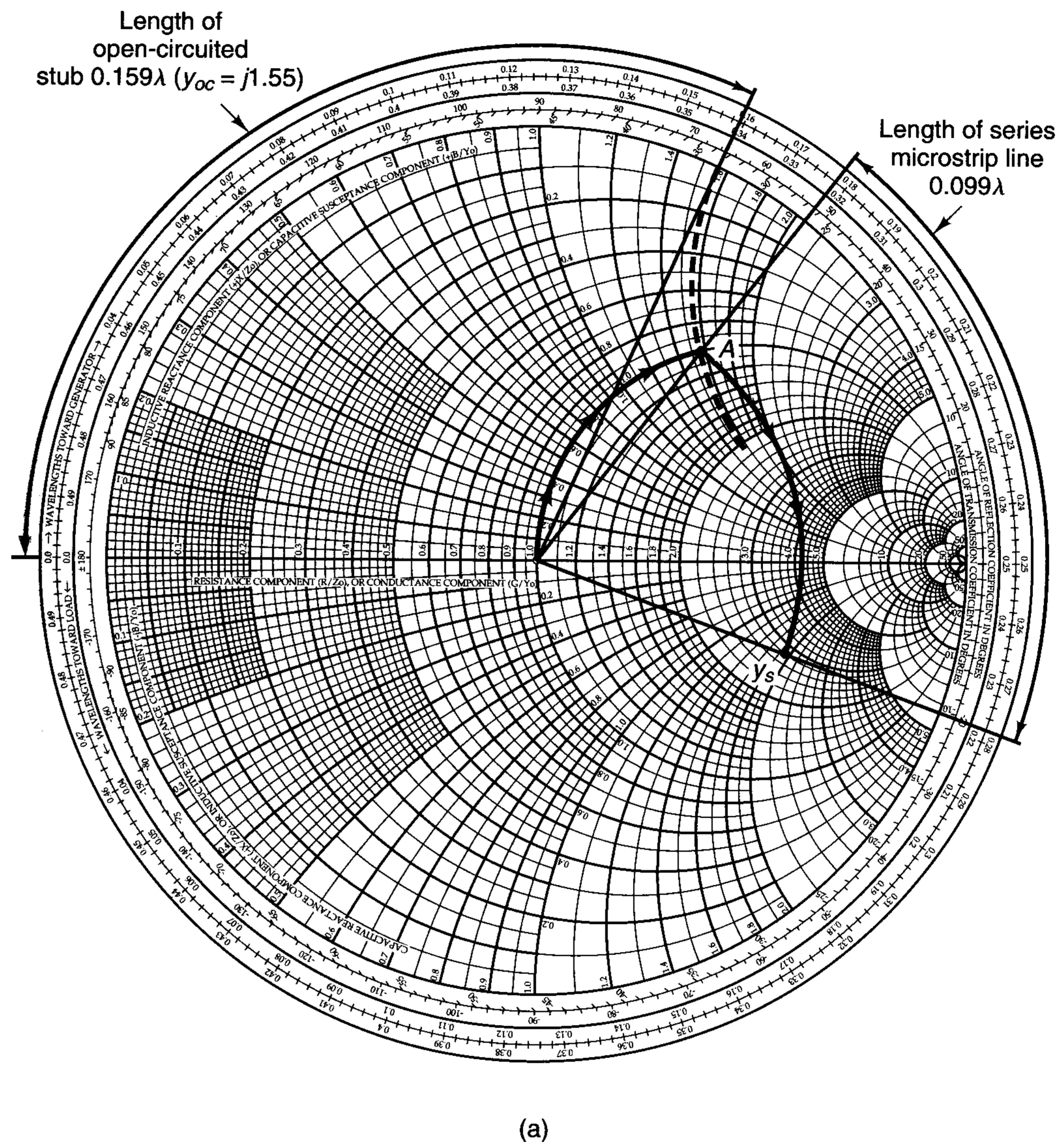


Figure 2.5.13 (a) Input matching network design; (b) output matching network design.

schematic also shows that $50\text{-}\Omega$ lines were added on both sides of C_A to provide a soldering area.

In Fig. 2.5.15, two parallel shunt stubs must provide the same admittance as the single stub in Fig. 2.5.14. Therefore, the admittance of each side of the balanced stub must be equal to half of the total admittance. For example, each side of the input balanced shunt stubs must have an admittance of $y = j1.55/2 = j0.775$. Using the Smith

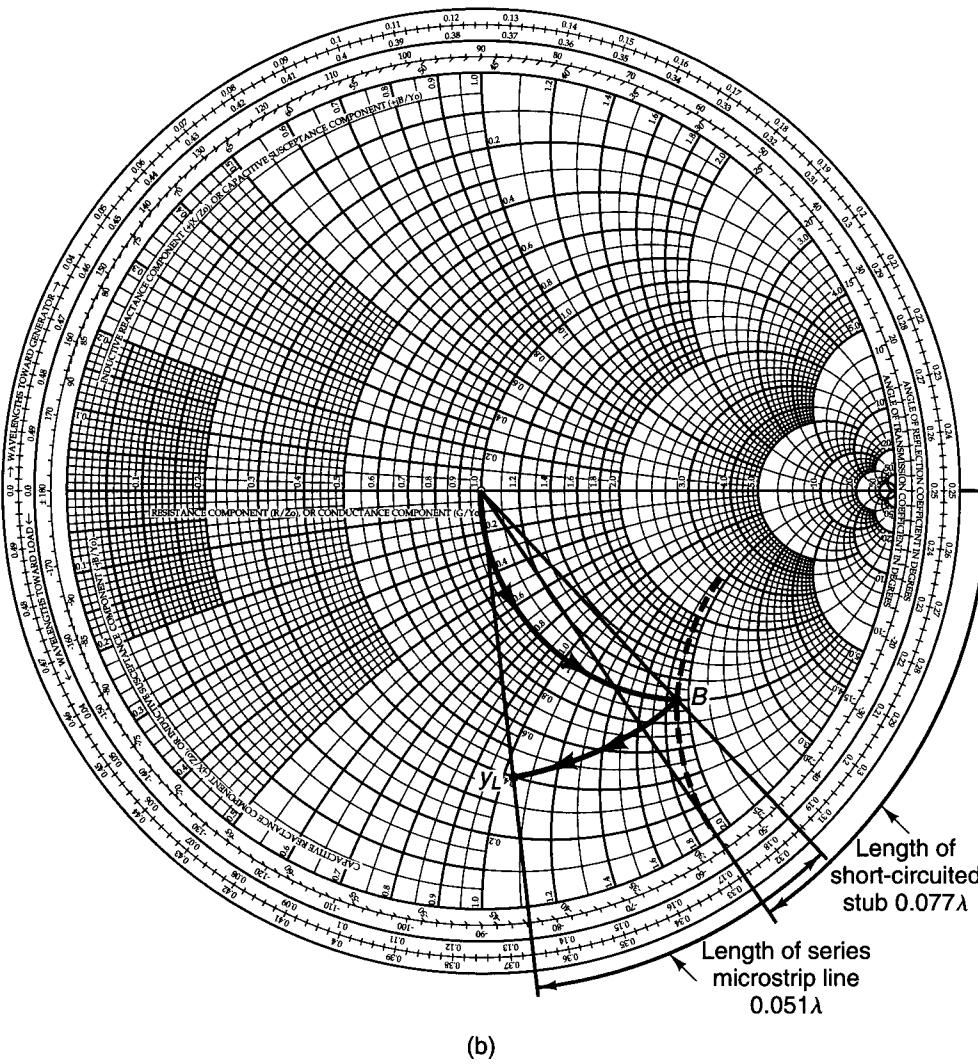
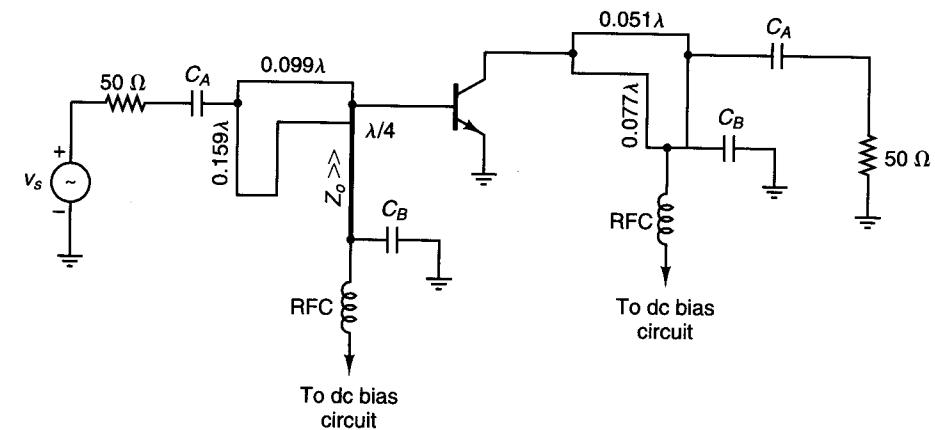
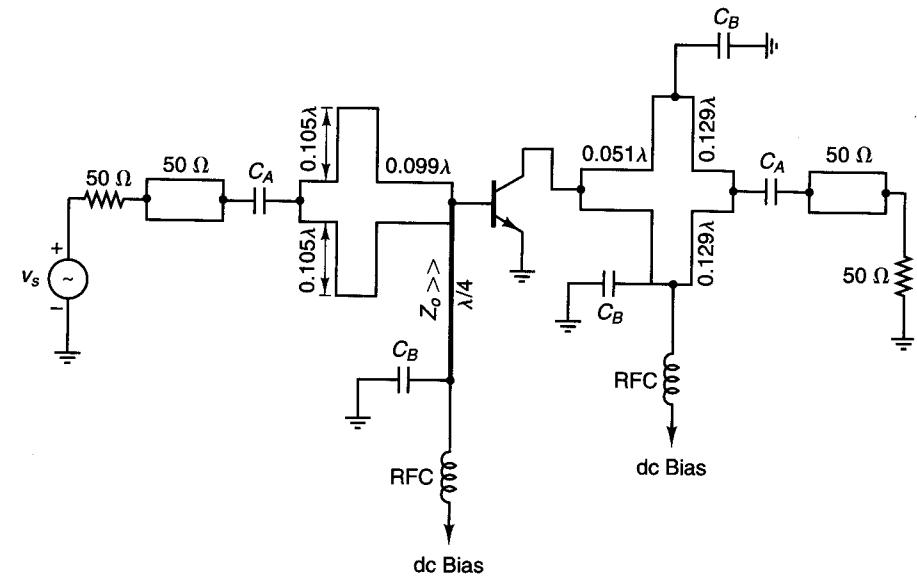


Figure 2.5.13 Continued

chart, we obtain that the length of each side must be 0.105λ . Observe that the length of the shunt stubs in Fig. 2.5.14 is not equal to the total length of the balance stubs in Fig. 2.5.15. Of course, a simple check will show that the admittance seen by the series transmission line is the same in both cases.

If we use RT/Duroid® with $\epsilon_r = 2.23$ and $h = 0.7874$ mm to build the amplifier, we find from (2.5.8) to (2.5.11) (or from Figs. 2.5.2 and 2.5.3) that a characteristic impedance of 50Ω is obtained with $W = 2.42$ mm and $\epsilon_{eff} = 1.91$. The microstrip wave-

Figure 2.5.14 Complete amplifier schematic. The characteristic impedance of the microstrip lines is 50Ω .Figure 2.5.15 Complete amplifier schematic using balanced shunt stubs. The characteristic impedance of the microstrip lines is 50Ω .

length in the 50Ω Duroid microstrip line is $\lambda = \lambda_0/\sqrt{1.91} = 0.7236\lambda_0$, where $\lambda_0 = 30\text{ cm}$ at $f = 1\text{ GHz}$. For a characteristic impedance of 100Ω in the $\lambda/4$ line, the width must be $W = 0.7\text{ mm}$. The line lengths in Fig. 2.5.15 are

$$0.105\lambda = 2.28\text{ cm}$$

$$0.099\lambda = 2.15\text{ cm}$$

$$0.051\lambda = 1.10\text{ cm}$$

$$0.129\lambda = 2.80\text{ cm}$$

$$\lambda/4 = 5.43\text{ cm}$$

Design 2: This method uses microstrip lines with different characteristic impedances, as shown in Fig. 2.5.10a. The design requires the transformation of 50Ω to $Y_s = (2.8 - j1.9)/50 = 0.056 - j0.038\text{ S}$. A quarter-wave transformer can be used to transform the source impedance of 50Ω to the resistance $1/0.056 = 17.86\Omega$. The characteristic impedance of the quarter-wave transformer is

$$Z_{o1} = \sqrt{50(17.86)} = 29.9\Omega$$

An open-circuited shunt stub can be used to obtain the admittance $-j0.038\text{ S}$. Therefore, as shown in Fig. 2.5.10c, an open-circuited shunt stub of length $3\lambda/8$ looks like a shunt inductor having the admittance $-jY_{o2}$. Equating $-jY_{o2}$ to $-j0.038\text{ S}$, we find the characteristic impedance Z_{o2} to be

$$Z_{o2} = \frac{1}{Y_{o2}} = \frac{1}{0.038} = 26.32\Omega$$

If the design is done using a short-circuited shunt stub (see Fig. 2.5.10a), its length would be $\lambda/8$ and $Z_{o2} = 26.3\Omega$.

Similarly, for the output matching network [$Y_L = (0.4 - j1.05)/50 = 0.008 - j0.021\text{ S}$], a quarter-wave line of characteristic impedance

$$Z_{o1} = \sqrt{50(125)} = 79.1\Omega$$

transforms the 50Ω load to a resistance of value $1/0.008 = 125\Omega$. An open-circuited shunt stub of length $3\lambda/8$ and characteristic impedance $Z_{o2} = 1/Y_{o2} = 1/0.021 = 47.6\Omega$ produces the required admittance of $-j0.021\text{ S}$.

The complete amplifier is shown in Fig. 2.5.16a. Figure 2.5.16b shows the amplifier using balanced shunt stubs of length $3\lambda/8$ to minimize the microstrip transition interaction. Observe that in the balance stubs the lengths were kept at $3\lambda/8$, but the characteristic impedance was doubled. For example, in Fig. 2.5.16b each half of the input balance stub must provide the admittance $-j0.038/2$ (since each half must contribute half of the total admittance). Therefore, the value of Z_{o2} for the balanced stubs at the input is $Z_{o2} = 2/0.038 = 52.6\Omega$.

Example 2.5.3

Design a microstrip matching network to transform the load $Z_L = 75 - j60\Omega$ to an input impedance of value $Z_{IN} = 15 + j30\Omega$.

Solution. In this design, let us select a Z_o different from 50Ω —for example, $Z_o = 75\Omega$. With $Z_o = 75\Omega$, the design consists of transforming a normalized load $z_L = Z_L/Z_o = 1 - j0.8$ (or $y_L = 0.61 + j0.49$) to the normalized input impedance $z_{IN} =$

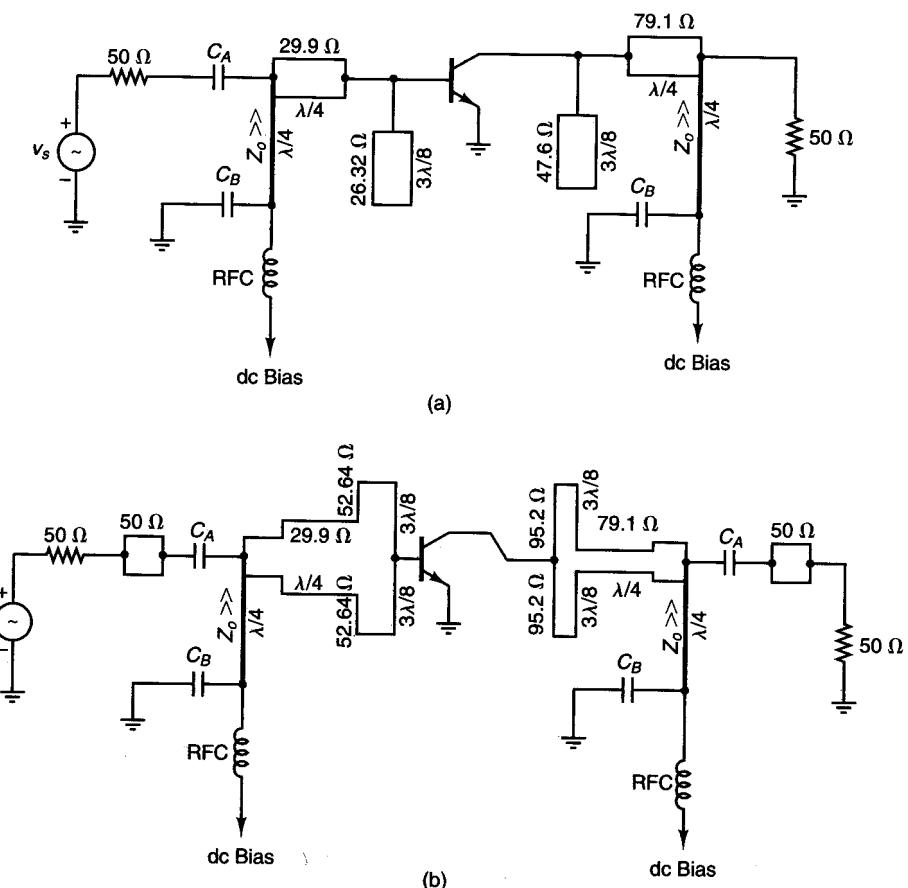


Figure 2.5.16 Matching network design using microstrip lines with different characteristic impedances.

$Z_{IN}/Z_o = 0.2 + j0.4\Omega$ (or $y_{IN} = 1 - j2$). The matching topology selected is shown in Fig. 2.5.17a, and the design in the Y Smith chart is shown in Fig. 2.5.17b. From Fig. 2.5.17b, the shunt admittance required to move from y_L to point A is $j1.5 - j0.49 = j1.01$. The admittance at point A is $y_A = 0.61 + j1.5$. An open-circuited shunt stub of length $l_1 = 0.126\lambda$ provides the admittance of $j1.01$. Then a series transmission line of length $l_2 = 0.313\lambda - 0.164\lambda = 0.149\lambda$ moves the admittance value, along a constant $|\Gamma|$ circle, from that at point A to y_{IN} .

A microstrip matching network which can be easily designed using a Z Smith chart is shown in Fig. 2.5.18a. This matching network uses a $\lambda/4$ line with characteristic impedance Z_{o1} to transform the 50Ω load ($z_L = 1$) to a resistance R_x ($r_x = R_x/50$) that lies on the constant $|\Gamma|$ circle that passes through $z_{IN} = Z_{IN}/50$. The value of Z_{o1} is given by

$$Z_{o1} = \sqrt{50R_x}$$

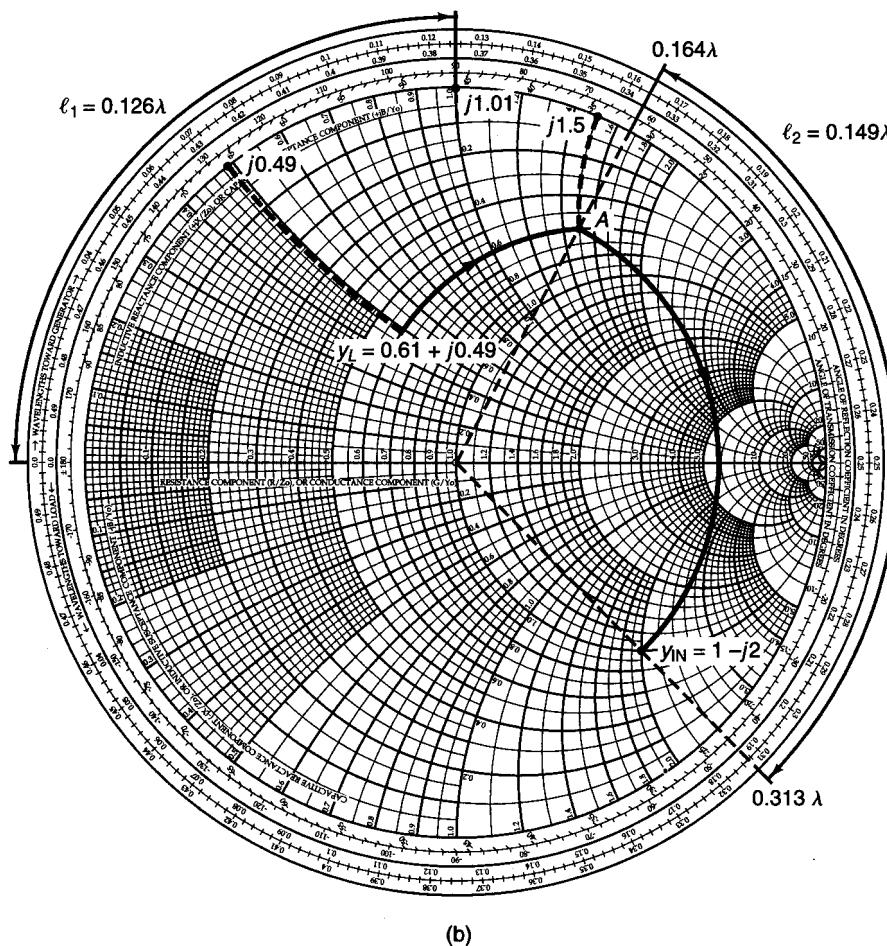
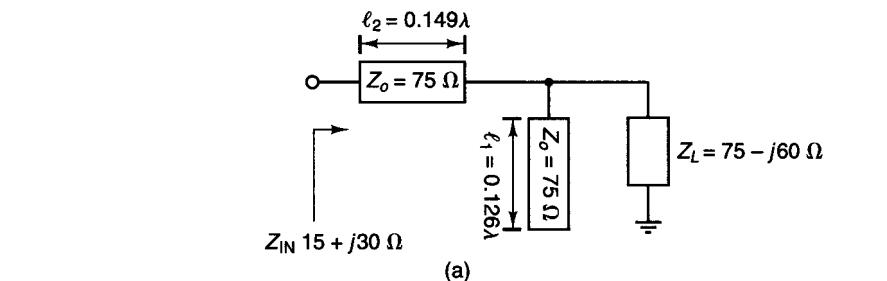


Figure 2.5.17 (a) Matching network for Example 2.5.3; (b) design in the Y Smith chart using $Z_o = 75 \Omega$.

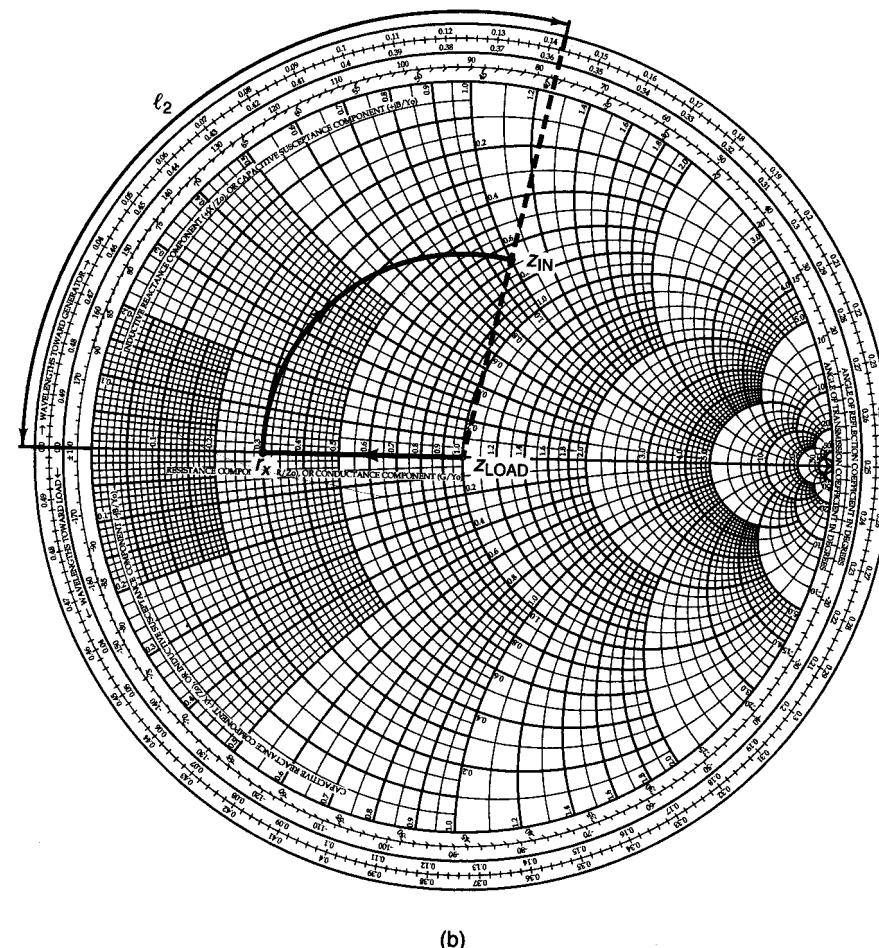
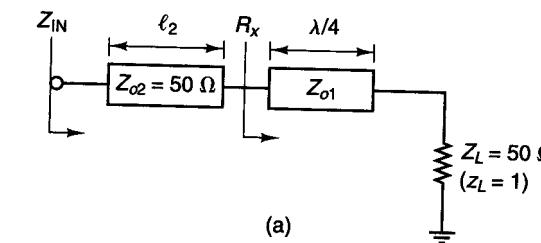


Figure 2.5.18 (a) A microstrip matching circuit; (b) design in the Z Smith chart that results in Z_{o1} smaller than 50Ω ; (c) design in the Z Smith chart that results in Z_{o1} greater than 50Ω .

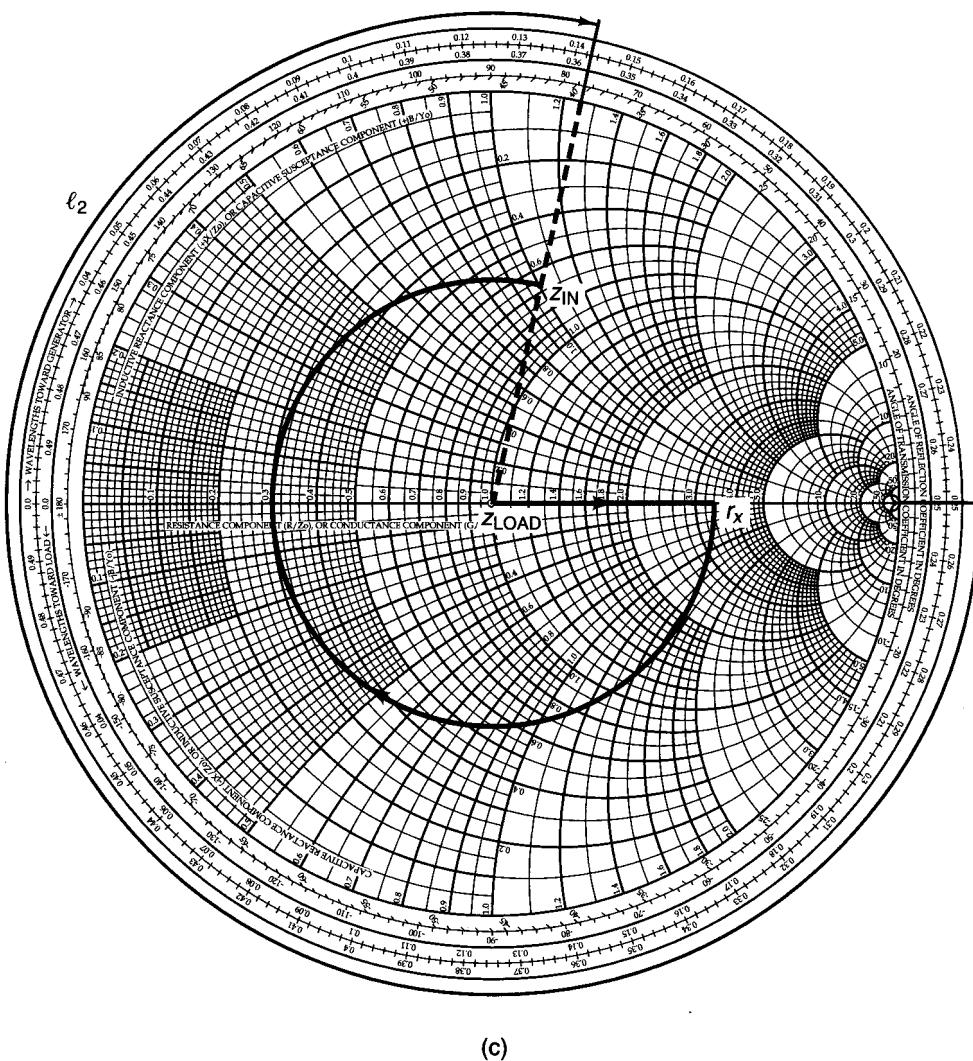


Figure 2.5.18 Continued

Then, the 50Ω line of length l_2 changes the normalized resistance r_x to the input impedance Z_{IN} . The design procedure in the Z Smith chart for an arbitrary value of Z_{IN} is shown in Fig. 2.5.18b. An alternate solution is shown in Fig. 2.5.18c. The solution described in Fig. 2.5.18b produces a Z_{o1} smaller than 50Ω , while the solution in Fig. 2.5.18c produces a Z_{o1} greater than 50Ω .

Example 2.5.4

Design the microstrip matching network in Fig. 2.5.18a to transform a 50Ω load to the input impedance $Z_{IN} = 33 + j50\Omega$.

Solution. The location of $Z_{IN} = Z_{IN}/50 = 0.66 + j1$ corresponds to the Z_{IN} shown in Fig. 2.5.18b. From the constant $|\Gamma|$ circle through Z_{IN} , we observe that $r_x = 0.3$, or $R_x = 50(0.3) = 15\Omega$. The $\lambda/4$ line is designed to transform $Z_L = 50\Omega$ to $R_x = 15\Omega$. Hence,

$$Z_{o1} = \sqrt{50(15)} = 27.4\Omega$$

Then, the 50Ω series transmission line of length $l_2 = 0.143\lambda$ produces an input impedance equal to $Z_{IN} = 0.66 + j1$, or $Z_{IN} = 50(Z_{IN}) = 33 + j50\Omega$.

The design of matching networks containing lumped components and microstrip transmission lines can also be done using the various Smith charts. The following example illustrates one such design.

Example 2.5.5 [cont'd from previous]

(a) An oscillator is designed at 2.5 GHz using the output matching topology shown in Fig. 2.5.19a. The length of the microstrips is shown for $\epsilon_{eff} = 1$ (i.e., for $v = c = 3 \times 10^{10}\text{ cm/s}$). The matching network uses a varactor diode as a voltage-variable capacitor for the control of the oscillator frequency. Determine the value of the load reflection coefficient.

(b) Specify the width, height, and length of the microstrip lines if they are constructed using an alumina substrate ($\epsilon_r = 9.6$).

Solution. (a) The wavelength in free space is

$$\lambda_0 = \frac{c}{f} = \frac{3 \times 10^{10}}{2.5 \times 10^9} = 12\text{ cm}$$

Hence, the shunt microstrip of length $l_1 = 43\text{ mm}$ (or $l_1 = 0.358\lambda_0$) has an open-circuited admittance of $y_{oc} = -j1.25$. This shunt microstrip acts like a shunt inductor. Hence, the admittance value of 50Ω in parallel with the shunt microstrip produces the admittance $y_A = 1 - j1.25$, shown as point A in the Y Smith chart in Fig. 2.5.19b.

The series microstrip of length $l_2 = 10\text{ mm}$ (or $l_2 = 0.083\lambda_0$) produces the matching from point A to point B (see Fig. 2.5.19b). This motion is along a constant $|\Gamma|$ circle. The admittance at B is $y_B = 0.41 - j0.53$. Then, the shunt capacitance of the varactor diode (3 pF or $y_C = j2.36$) produces the motion from point B to point C . At point C the admittance is $0.41 + j1.83$, which corresponds to $\Gamma_L = 0.83|-124.5^\circ|$.

In practice, the capacitance of the varactor diode is 3 pF when a specific dc voltage is applied to it. Let us assume that 3 pF occurs when the dc voltage is 4 V . A practical dc bias circuit for the varactor diode and for the transistor is shown in Fig. 2.5.19c.

(b) If the microstrip lines are built using alumina with $\epsilon_r = 9.6$, then from Fig. 2.5.4 a characteristic impedance of 50Ω can be obtained with $W = 24.7\text{ mils}$ and $h = 25\text{ mils}$.

It also follows that $\epsilon_{eff} = 6.46$ (or $\lambda = \lambda_0/\sqrt{\epsilon_{eff}} = 12/\sqrt{6.46} = 4.72\text{ cm}$) in the alumina. Hence, the length of the shunt stub, denoted by l'_1 , is

$$l'_1 = \frac{l_1}{\sqrt{\epsilon_{eff}}} = \frac{43}{\sqrt{6.46}} = 16.9\text{ mm}$$

and that of the series microstrip line, denoted by l'_2 is

$$l'_2 = \frac{l_2}{\sqrt{\epsilon_{eff}}} = \frac{10}{\sqrt{6.46}} = 3.93\text{ mm}$$

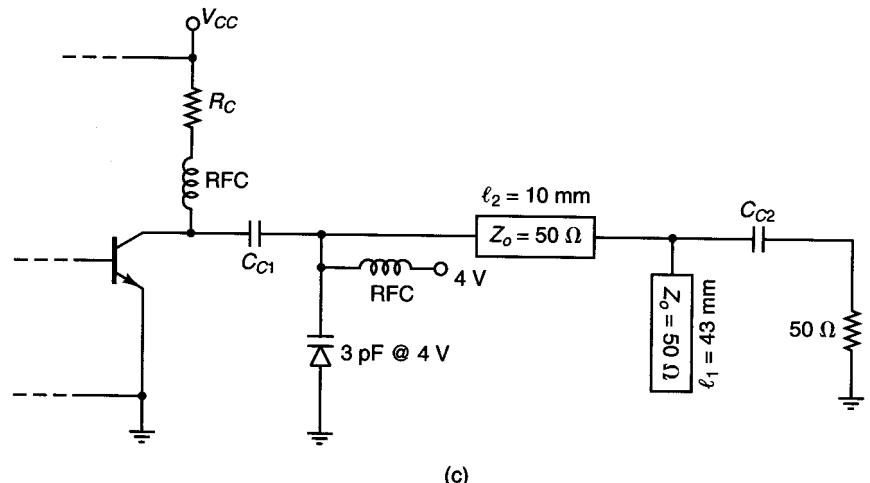
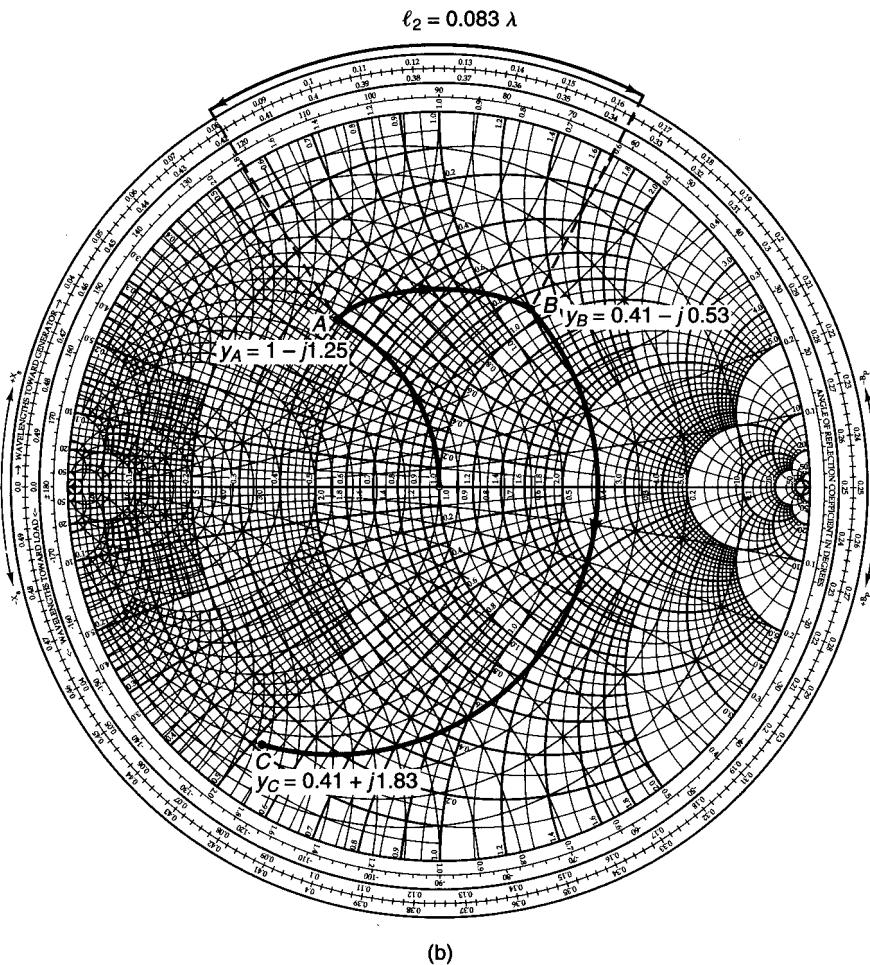
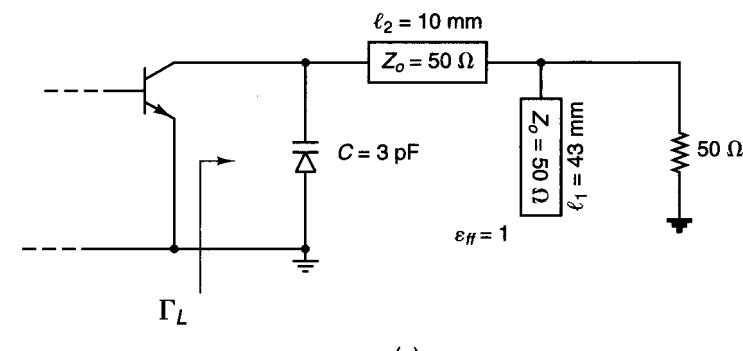


Figure 2.5.19 (a) Circuit schematic for Example 2.5.5; (b) calculation of Γ_L using the Y Smith chart; (c) an implementation of the dc bias circuit.

Example 2.5.6

Design a three-element microstrip matching network to transform a 50Ω termination to a load reflection coefficient given by $\Gamma_L = 0.48 \angle 72^\circ$.

Solution. We will solve this problem using the ZY Smith chart. The solution using this chart should be of interest to the reader (it is recommended that the reader also work this problem using the Y Smith chart). Figure 2.5.20a shows Γ_L in the ZY Smith chart, as well as the path selected for the matching network, using $Z_o = 50 \Omega$. The first element transforms the normalized admittance from point A to point B in Fig. 2.5.20a. The normalized admittance at B is $y_B = 1 - j0.82$. As shown in Fig. 2.5.20b, this can be implemented with a short-circuited shunt stub of length $l_1 = 0.141\lambda$. The second element produces the admittance $y_C = 0.5 - j0.3$ at point C. This element is implemented using a series microstrip line of length $l_2 = 0.099\lambda$. Finally, the third element changes the susceptance along a constant conductance circle of 0.5 from $-j0.3$ to $-j0.6$. This element can be implemented using a short-circuited shunt stub of length $l_3 = 0.203\lambda$ (i.e., having an admittance of $-j0.3$). The matching network is shown in Fig. 2.5.20b.

2.6 SIGNAL FLOW GRAPHS AND APPLICATIONS

A signal flow graph is a convenient technique to represent and analyze the transmission and reflection of waves in a microwave amplifier. Once the signal flow graph is developed, relations between the variables can be obtained using Mason's rule. The flow graph technique permits expressions, such as power gains and voltage gains of complex microwave amplifiers, to be derived easily. Certain rules are followed in constructing a signal flow graph:

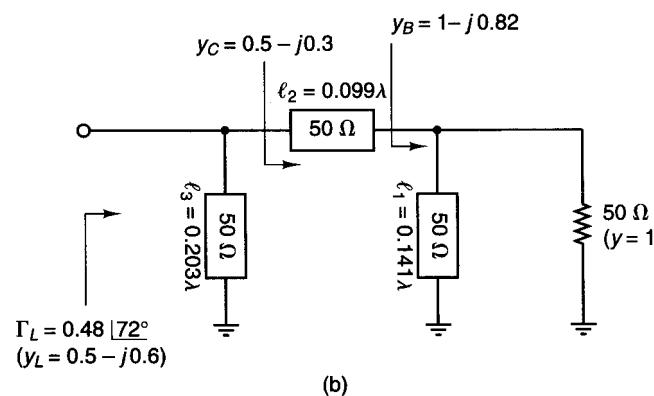
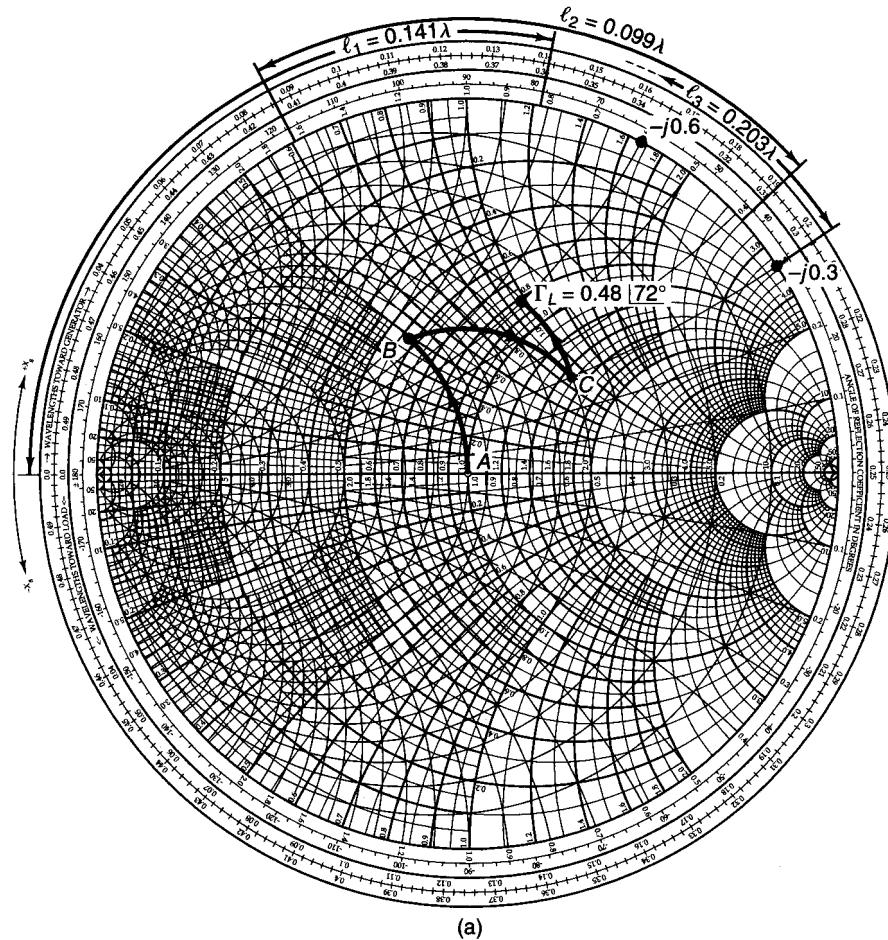


Figure 2.5.20 (a) Design in the ZY Smith chart for Example 2.5.6; (b) the matching network.

1. Each variable is designated as a node.
2. The S parameters and reflection coefficients are represented by branches.
3. Branches enter dependent variable nodes and emanate from independent variable nodes. The independent variable nodes are the incident waves, and the reflected waves are dependent variables nodes.
4. A node is equal to the sum of the branches entering it.

The signal flow graph of the S parameters of a two-port network is shown in Fig. 2.6.1. Observe that b_1 and b_2 are the dependent nodes and a_1 and a_2 the independent nodes. The complete signal flow graph of the two-port network is shown in Fig. 2.6.2.

The signal flow graph in Fig. 2.6.2 shows the relationship between the traveling waves. The incident wave a_1 at port 1 gets partly transmitted (i.e., $S_{21}a_1$) to become part of b_2 , and partly reflected (i.e., $S_{11}a_1$) to become part of b_1 . Similarly, the incident wave a_2 at port 2 gets partly transmitted (i.e., $S_{12}a_2$) to become part of b_1 and partly reflected (i.e., $S_{22}a_2$) to become part of b_2 .

In order to obtain the signal flow graph of a microwave amplifier, we need to obtain the signal flow graph of a signal generator with some internal impedance and the signal flow graph of a load impedance.

Figure 2.6.3a shows a voltage-source generator with impedance Z_s . At the terminals we can write

$$V_g = E_s + I_g Z_s \quad (2.6.1)$$

Using (1.4.1) and (1.4.2), we can express (2.6.1) in terms of traveling waves—namely,

$$V_g^+ + V_g^- = E_s + \left(\frac{V_g^+}{Z_o} - \frac{V_g^-}{Z_o} \right) Z_s$$

Solving for V_g^- , we obtain

$$b_g = b_s + \Gamma_s a_g \quad (2.6.2)$$

where

$$b_g = \frac{V_g^-}{\sqrt{Z_o}}$$

$$a_g = \frac{V_g^+}{\sqrt{Z_o}}$$

$$b_s = \frac{E_s \sqrt{Z_o}}{Z_s + Z_o}$$

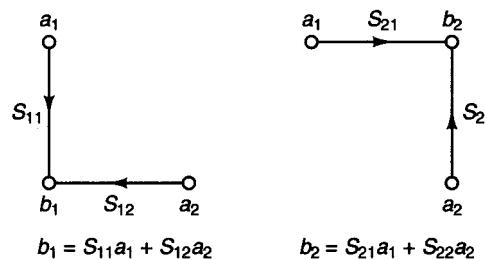


Figure 2.6.1 Signal flow graph for the scattering parameter equations.

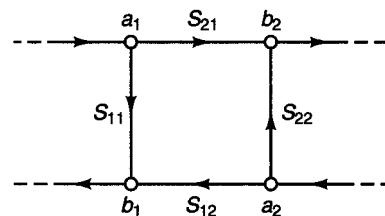


Figure 2.6.2 Signal flow graph of a two-port network.

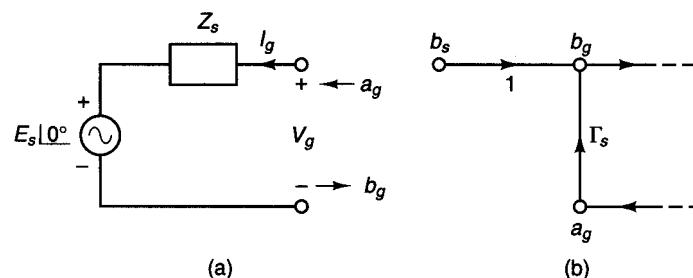


Figure 2.6.3 Signal flow graph of a voltage-source generator.

and

$$\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o}$$

From (2.6.2), the signal flow graph in Fig. 2.6.3b follows.

For the load impedance shown in Fig. 2.6.4a, we can write

$$V_L = Z_L I_L$$

In terms of traveling waves, we obtain

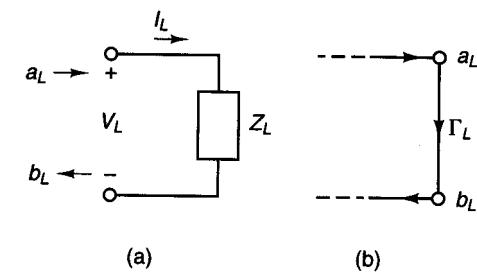


Figure 2.6.4 Signal flow graph of a load impedance.

$$V_L^+ + V_L^- = Z_L \left(\frac{V_L^+}{Z_o} - \frac{V_L^-}{Z_o} \right)$$

or

$$b_L = \Gamma_L a_L \quad (2.6.3)$$

where

$$b_L = \frac{V_L^-}{\sqrt{Z_o}}$$

$$a_L = \frac{V_L^+}{\sqrt{Z_o}}$$

and

$$\Gamma_L = \frac{Z_L - Z_o}{Z_L + Z_o}$$

The signal flow graph follows from (2.6.3) and is shown in Fig. 2.6.4b.

We can now combine the signal flow graph for the two-port network in Fig. 2.6.2 with the signal flow graphs of the signal generator (i.e., Fig. 2.6.3b) and the load (i.e., Fig. 2.6.4b). Observe that the nodes b_g , a_g , b_L , and a_L are identical to a_1 , b_1 , a_2 , and b_2 , respectively. The resulting signal flow graph of a microwave amplifier is shown in Fig. 2.6.5.

To determine the ratio or transfer function T of a dependent to an independent variable, we apply Mason's rule—namely,

$$T = \frac{P_1[1 - \sum L(1)^{(1)} + \sum L(2)^{(1)} - \dots] + P_2[1 - \sum L(1)^{(2)} + \dots] + \dots}{1 - \sum L(1) + \sum L(2) - \sum L(3) + \dots}$$

where the different terms are defined as follows.

The terms P_1 , P_2 , and so on are the different paths connecting the dependent and independent variables whose transfer function T is to be determined. A path is defined as a set of consecutive, codirectional branches along which no node is encountered more than once as we move in the graph from the independent to

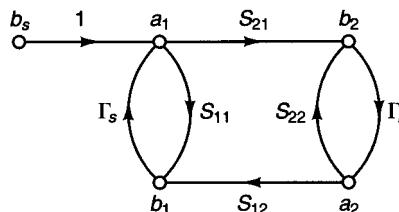


Figure 2.6.5 Signal flow graph of a microwave amplifier.

the dependent node. The value of the path is the product of all branch coefficients along the path. For example, in Fig. 2.6.5, b_s is the only independent variable. To determine the ratio b_1/b_s , we identify two paths, $P_1 = S_{11}$ and $P_2 = S_{21}\Gamma_L S_{12}$.

The term $\Sigma L(1)$ is the sum of all first-order loops. A first-order loop is defined as the product of the branches encountered in a round trip as we move from a node in the direction of the arrows back to that original node. In Fig. 2.6.5, $S_{11}\Gamma_s$, $S_{21}\Gamma_L S_{12}\Gamma_s$, and $S_{22}\Gamma_L$ are first-order loops.

The term $\Sigma L(2)$ is the sum of all second-order loops. A second-order loop is defined as the product of any two nontouching first-order loops. In Fig. 2.6.5, $S_{11}\Gamma_s$ and $S_{22}\Gamma_L$ do not touch; therefore, the product $S_{11}\Gamma_s S_{22}\Gamma_L$ is a second-order loop.

The term $\Sigma L(3)$ is the sum of all third-order loops. A third-order loop is defined as the product of three nontouching first-order loops. In Fig. 2.6.5, there are no third-order loops. Of course, the terms $\Sigma L(4)$, $\Sigma L(5)$, and so on represent fourth-, fifth-, and higher-order loops.

The terms $\Sigma L(1)^{(P)}$ is the sum of all first-order loops that do not touch the path P between the independent and dependent variables. In Fig. 2.6.5, for the path $P_1 = S_{11}$ we find that $\Sigma L(1)^{(1)} = \Gamma_L S_{22}$, and for the path $P_2 = S_{21}\Gamma_L S_{12}$ we find that $\Sigma L(1)^{(2)} = 0$.

The term $\Sigma L(2)^{(P)}$ is the sum of all second-order loops that do not touch the path P between the independent and dependent variables. In Fig. 2.6.5, we find that $\Sigma L(2)^{(P)} = 0$. Of course, $\Sigma L(3)^{(P)}$, $\Sigma L(4)^{(P)}$, and so on represent higher-order loops that do not touch the path P .

For the transfer function b_1/b_s in Fig. 2.6.5, we have found that $P_1 = S_{11}$, $P_2 = S_{21}\Gamma_L S_{12}$, $\Sigma L(1) = S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s$, $\Sigma L(2) = S_{11}\Gamma_s S_{22}\Gamma_L$, and $\Sigma L(1)^{(1)} = \Gamma_L S_{22}$. Therefore, using Mason's rule, we obtain

$$\frac{b_1}{b_s} = \frac{S_{11}(1 - \Gamma_L S_{22}) + S_{21}\Gamma_L S_{12}}{1 - (S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s) + S_{11}\Gamma_s S_{22}\Gamma_L}$$

Applications of Signal Flow Graphs

The first application of signal flow graph analysis is in the calculation of the input reflection coefficient, called Γ_{IN} , when a load is connected to the output of a two-port network. The signal flow graph is shown in Fig. 2.6.6.

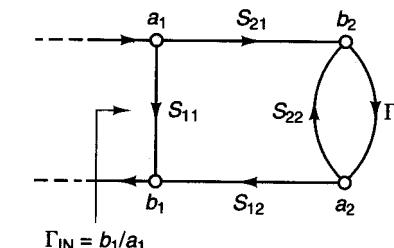


Figure 2.6.6 Signal flow graph for the input reflection coefficient Γ_{IN} .

The input reflection coefficient Γ_{IN} is defined as

$$\Gamma_{IN} = \frac{b_1}{a_1}$$

Observing that $P_1 = S_{11}$, $P_2 = S_{21}\Gamma_L S_{12}$, $\Sigma L(1) = S_{22}\Gamma_L$, and $\Sigma L(1)^{(1)} = S_{22}\Gamma_L$, we can use Mason's rule to obtain

$$\begin{aligned} \Gamma_{IN} &= \frac{S_{11}(1 - S_{22}\Gamma_L) + S_{21}\Gamma_L S_{12}}{1 - S_{22}\Gamma_L} \\ &= S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \end{aligned} \quad (2.6.4)$$

If $\Gamma_L = 0$, it follows from (2.6.4) that $\Gamma_{IN} = S_{11}$. Also, when there is no transmission from the output to the input (i.e., when $S_{12} = 0$), it follows that $\Gamma_{IN} = S_{11}$. When $S_{12} = 0$, we call the device represented by the two-port a unilateral device.

Similarly, we can calculate the output reflection coefficient $\Gamma_{OUT} = b_2/a_2$ with $b_s = 0$ from the signal flow graph shown in Fig. 2.6.7. The expression for Γ_{OUT} is

$$\Gamma_{OUT} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \quad (2.6.5)$$

Next, we use signal flow graphs to calculate power gain and voltage gain. The square of the magnitude of the incident and reflected waves represents

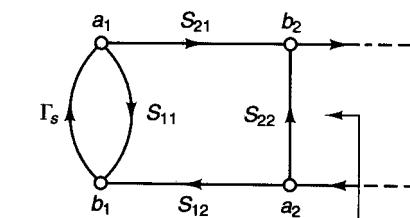


Figure 2.6.7 Signal flow graph for the output reflection coefficient Γ_{OUT} .

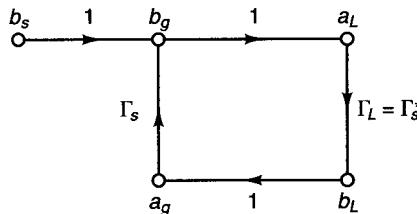


Figure 2.6.8 Signal flow graph of a voltage source connected to a conjugate matched load.

power. Therefore, the power delivered to the load in Fig. 2.6.5 is given by the difference between the incident and reflected power—namely,

$$P_L = \frac{1}{2} |b_2|^2 - \frac{1}{2} |a_2|^2 = \frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2) \quad (2.6.6)$$

The power available from a source is defined as the power delivered by the source to a conjugately matched load. Figure 2.6.8 shows the signal flow graph of a source connected to a conjugate match load (i.e., $\Gamma_L = \Gamma_s^*$). Therefore, the power available from the source, in Fig. 2.6.8, is given by

$$P_{AVS} = \frac{1}{2} |b_g|^2 - \frac{1}{2} |a_g|^2 \quad (2.6.7)$$

Observing that $b_g = b_s + b_s \Gamma_s \Gamma_s^*$ and $a_g = b_g \Gamma_s^*$, we obtain

$$b_g = \frac{b_s}{1 - |\Gamma_s|^2} \quad (2.6.8)$$

and

$$a_g = \frac{b_s \Gamma_s^*}{1 - |\Gamma_s|^2} \quad (2.6.9)$$

Substituting (2.6.8) and (2.6.9) into (2.6.7) gives

$$P_{AVS} = \frac{\frac{1}{2} |b_s|^2}{1 - |\Gamma_s|^2} \quad (2.6.10)$$

The previous results could have also been obtained as follows. Observe that the power delivered to the load Γ_L in Fig. 2.6.8 is given by

$$P_L = |a_L|^2 (1 - |\Gamma_L|^2) = \frac{\frac{1}{2} |b_s|^2 (1 - |\Gamma_L|^2)}{|1 - \Gamma_s \Gamma_L|^2}$$

Therefore, with $\Gamma_L = \Gamma_s^*$ the power delivered to the load is equal to the available power from the source, and (2.6.10) follows.

The transducer power gain, called G_T , is defined as the ratio of the power delivered to a load to the power available from the source. From (2.6.6) and (2.6.10), we obtain

$$G_T = \frac{P_L}{P_{AVS}} = \frac{|b_2|^2}{|b_s|^2} (1 - |\Gamma_L|^2) (1 - |\Gamma_s|^2) \quad (2.6.11)$$

The ratio b_2/b_s can be obtained using Mason's rule—namely,

$$\begin{aligned} \frac{b_2}{b_s} &= \frac{S_{21}}{1 - (S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s) + S_{11}\Gamma_s S_{22}\Gamma_L} \\ &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s} \end{aligned} \quad (2.6.12)$$

Substituting (2.6.12) into (2.6.11) results in

$$G_T = \frac{|S_{21}|^2 (1 - |\Gamma_s|^2) (1 - |\Gamma_L|^2)}{|(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s|^2} \quad (2.6.13)$$

The denominator of (2.6.13) can be further manipulated, and G_T can be expressed in the form

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_{IN}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.6.14)$$

or

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.6.15)$$

where Γ_{IN} and Γ_{OUT} are given by (2.6.4) and (2.6.5), respectively.

The power gain G_P is defined as the ratio of the power delivered to the load P_L to the input power to the network P_{IN} . P_L is given by (2.6.6), and the input power is given by

$$P_{IN} = \frac{1}{2} |a_1|^2 - \frac{1}{2} |b_1|^2 = \frac{1}{2} |a_1|^2 (1 - |\Gamma_{IN}|^2)$$

Hence, we can express the power gain in the form

$$G_P = \frac{P_L}{P_{IN}} = \frac{|b_2|^2 (1 - |\Gamma_L|^2)}{|a_1|^2 (1 - |\Gamma_{IN}|^2)}$$

Dividing the numerator and denominator by $|b_s|^2$ gives

$$G_P = \frac{P_L}{P_{IN}} = \frac{\frac{|b_2|^2}{|b_s|^2} (1 - |\Gamma_L|^2)}{\frac{|a_1|^2}{|b_s|^2} (1 - |\Gamma_{IN}|^2)} \quad (2.6.16)$$

The ratio b_2/b_s is given by (2.6.12), and using Mason's rule, the ratio a_1/b_s is

$$\frac{a_1}{b_s} = \frac{1 - S_{22}\Gamma_L}{1 - (S_{11}\Gamma_s + S_{22}\Gamma_L + S_{21}\Gamma_L S_{12}\Gamma_s) + S_{11}\Gamma_s S_{22}\Gamma_L} \quad (2.6.17)$$

Substituting (2.6.12) and (2.6.17) into (2.6.16), we can express G_p in the form

$$G_p = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (2.6.18)$$

The available power gain G_A is defined as the ratio of the power available from the network P_{AVN} to the power available from the source P_{AVS} . The power available from the network is the power delivered by the network to a conjugate matched load. That is,

$$\begin{aligned} P_{AVN} &= P_L \Big|_{\Gamma_L=\Gamma_{OUT}^*} = \left[\frac{1}{2} |b_2|^2 - \frac{1}{2} |a_2|^2 \right] \Big|_{\Gamma_L=\Gamma_{OUT}^*} \\ &= \left[\frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2) \right] \Big|_{\Gamma_L=\Gamma_{OUT}^*} = \frac{1}{2} |b_2|^2 (1 - |\Gamma_{OUT}|^2) \end{aligned} \quad (2.6.19)$$

Therefore, from (2.6.19) and (2.6.10), G_A is given by

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{|b_2|^2}{|b_s|^2} (1 - |\Gamma_{OUT}|^2) (1 - |\Gamma_s|^2) \quad (2.6.20)$$

From (2.6.12), the ratio b_2/b_s with $\Gamma_L = \Gamma_{OUT}^*$ can be expressed as

$$\begin{aligned} \frac{b_2}{b_s} &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s} \\ &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - \Gamma_{OUT}\Gamma_L)} \Big|_{\Gamma_L=\Gamma_{OUT}^*} \\ &= \frac{S_{21}}{(1 - S_{11}\Gamma_s)(1 - |\Gamma_{OUT}|^2)} \end{aligned} \quad (2.6.21)$$

Substituting (2.6.21) into (2.6.20) results in the expression

$$G_A = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{OUT}|^2} \quad (2.6.22)$$

The voltage gain of the amplifier is defined as the ratio of the output voltage to the input voltage. That is,

$$A_v = \frac{a_2 + b_2}{a_1 + b_1}$$

Dividing by b_s gives

$$A_v = \frac{a_2/b_s + b_2/b_s}{a_1/b_s + b_1/b_s}$$

Therefore, we need to calculate the ratios a_2/b_s , b_2/b_s , a_1/b_s , and b_1/b_s using Mason's rule. The expression for A_v can be shown to be

$$A_v = \frac{S_{21}(1 + \Gamma_L)}{(1 - S_{22}\Gamma_L) + S_{11}(1 - S_{22}\Gamma_L) + S_{21}\Gamma_L S_{12}} \quad (2.6.23)$$

2.7 POWER-GAIN EXPRESSIONS: ALTERNATE DERIVATIONS

In the previous section, the power-gain expressions were derived using signal flow graph theory. In this section, the power-gain expressions are derived using a direct manipulation of the S -parameters relations.

For the microwave amplifier shown in Fig. 2.7.1, the source and load reflection coefficients in a Z_o system are

$$\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o} \quad (2.7.1)$$

and

$$\Gamma_L = \frac{Z_L - Z_o}{Z_L + Z_o} \quad (2.7.2)$$

For the transistor, the input and output traveling waves measured in a Z_o system are related by

$$b_1 = S_{11}a_1 + S_{12}a_2 \quad (2.7.3)$$

and

$$b_2 = S_{21}a_1 + S_{22}a_2 \quad (2.7.4)$$

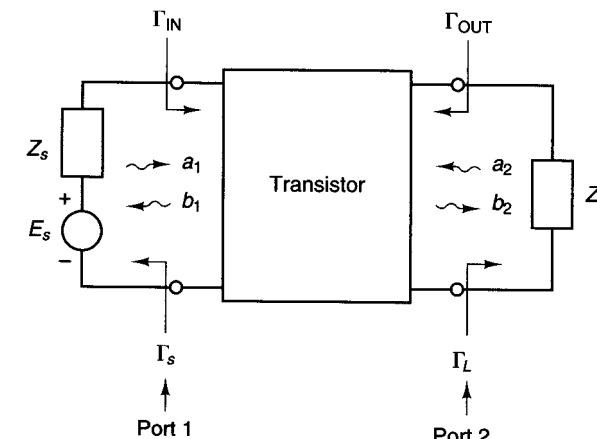


Figure 2.7.1 Block diagram of a microwave amplifier.

One should observe that the concepts of a reflection coefficient and traveling waves can be used even if there are no transmission lines at port 1 and port 2 in Fig. 2.7.1. The reflection coefficients Γ_s and Γ_L and the traveling waves a_1 , b_1 , a_2 , and b_2 are the reflection coefficients and traveling waves that would exist if transmission lines with characteristic impedances Z_o were inserted at ports 1 and 2. Alternatively, we could think of the reflection coefficients and the traveling waves as those in a transmission line of zero length and characteristic impedance Z_o connected at ports 1 and 2.

The input reflection coefficient Γ_{IN} is

$$\Gamma_{IN} = \frac{b_1}{a_1}$$

The ratio b_1/a_1 can be evaluated as follows. Since

$$a_2 = \Gamma_L b_2 \quad (2.7.5)$$

substituting (2.7.5) into (2.7.4) gives

$$b_2 = S_{21}a_1 + S_{22}\Gamma_L b_2$$

or

$$b_2 = \frac{S_{21}a_1}{1 - S_{22}\Gamma_L} \quad (2.7.6)$$

Then, substituting (2.7.5) and (2.7.6) into (2.7.3) gives

$$b_1 = S_{11}a_1 + S_{12}\Gamma_L b_2 = S_{11}a_1 + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} a_1$$

or

$$\Gamma_{IN} = \frac{b_1}{a_1} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}$$

The output reflection coefficient is defined as

$$\Gamma_{OUT} = \left. \frac{b_2}{a_2} \right|_{E_s=0}$$

With $E_s = 0$, it follows that

$$a_1 = \Gamma_s b_1 \quad (2.7.7)$$

Substituting (2.7.7) into (2.7.3) gives

$$b_1 = S_{11}\Gamma_s b_1 + S_{12}a_2$$

or

$$b_1 = \frac{S_{12}a_2}{1 - S_{11}\Gamma_s} \quad (2.7.8)$$

Then, substituting (2.7.7) and (2.7.8) into (2.7.4) gives

$$b_2 = S_{21}\Gamma_s b_1 + S_{22}a_2 = \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} a_2 + S_{22}a_2$$

Therefore, Γ_{OUT} is given by

$$\Gamma_{OUT} = \left. \frac{b_2}{a_2} \right|_{E_s=0} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s}$$

The power delivered to the input port of the transistor is

$$P_{IN} = \frac{1}{2} |a_1|^2 - \frac{1}{2} |b_1|^2 = \frac{1}{2} |a_1|^2(1 - |\Gamma_{IN}|^2) \quad (2.7.9)$$

At the input port (see Fig. 2.7.2),

$$V_1 = E_s + I_1 Z_s \quad (2.7.10)$$

Therefore, in terms of traveling waves, (2.7.10) is written as

$$a_1 = b_s + \Gamma_s b_1 \quad (2.7.11)$$

where

$$a_1 = \frac{V_1^-}{\sqrt{Z_o}}$$

$$b_1 = \frac{V_1^+}{\sqrt{Z_o}}$$

$$b_s = \frac{E_s \sqrt{Z_o}}{Z_s + Z_o}$$

and

$$\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o}$$

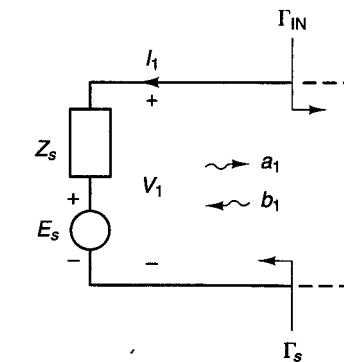


Fig. 2.7.2 The input port.

Except for a nonnomenclatural difference, (2.7.11) is identical to (2.6.2).

Since $b_1 = \Gamma_{IN}a_1$, (2.7.11) can be written as

$$a_1 = b_s + \Gamma_s \Gamma_{IN} a_1$$

or

$$a_1 = \frac{b_s}{1 - \Gamma_s \Gamma_{IN}} \quad (2.7.12)$$

Substituting (2.7.12) into (2.7.9) gives

$$P_{IN} = \frac{1}{2} |b_s|^2 \frac{1 - |\Gamma_{IN}|^2}{|1 - \Gamma_s \Gamma_{IN}|^2} \quad (2.7.13)$$

The power available from the source is equal to the input power when $\Gamma_{IN} = \Gamma_s^*$. With $\Gamma_{IN} = \Gamma_s^*$, (2.7.13) gives

$$P_{AVS} = P_{IN}|_{\Gamma_{IN}=\Gamma_s^*} = \frac{\frac{1}{2} |b_s|^2}{1 - |\Gamma_s|^2} \quad (2.7.14)$$

Observe that (2.7.14) agrees with (2.6.10).

Substituting (2.7.14) into (2.7.13), we can express P_{IN} in the form

$$P_{IN} = P_{AVS} \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_{IN}|^2)}{|1 - \Gamma_s \Gamma_{IN}|^2} \quad (2.7.15)$$

Equation (2.7.15) can be written as

$$P_{IN} = P_{AVS} M_s \quad (2.7.16)$$

where

$$M_s = \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_{IN}|^2)}{|1 - \Gamma_s \Gamma_{IN}|^2} \quad (2.7.17)$$

The factor M_s is known as the source mismatch factor (or the source mismatch loss). This factor is used to quantize what portion of P_{AVS} is delivered to the input of the transistor. Observe that if $\Gamma_{IN} = \Gamma_s^*$, (2.7.17) gives $M_s = 1$ and it follows that $P_{IN} = P_{AVS}$. This fact is expressed in the form

$$P_{IN} = P_{AVS}|_{\Gamma_{IN}=\Gamma_s^*}$$

A Thévenin's equivalent circuit at the output port of the transistor is shown in Fig. 2.7.3, where Z_{OUT} (i.e., the output impedance) is the Thévenin's impedance seen from the output terminals of the transistor. From Fig. 2.7.3, we have

$$V_L = E_{TH} - I_L Z_{OUT}$$

This expression is similar to (2.7.10) and can be expressed in terms of traveling waves as

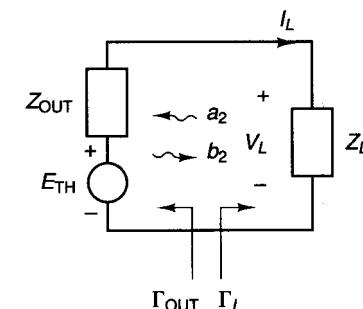


Figure 2.7.3 A Thévenin's equivalent circuit at the output port of the transistor.

$$b_2 = b_{TH} + \Gamma_{OUT} a_2 \quad (2.7.18)$$

where

$$b_2 = \frac{V_L^-}{\sqrt{Z_o}}$$

$$a_2 = \frac{V_L^+}{\sqrt{Z_o}}$$

$$b_{TH} = \frac{E_{TH} \sqrt{Z_o}}{Z_{OUT} + Z_o}$$

and

$$\Gamma_{OUT} = \frac{Z_{OUT} - Z_o}{Z_{OUT} + Z_o}$$

Since $a_2 = \Gamma_L b_2$, (2.7.18) gives

$$b_2 = b_{TH} + \Gamma_{OUT} \Gamma_L b_2$$

or

$$b_2 = \frac{b_{TH}}{1 - \Gamma_{OUT} \Gamma_L} \quad (2.7.19)$$

The power delivered to the load Z_L is

$$P_L = \frac{1}{2} |b_2|^2 - \frac{1}{2} |a_2|^2 = \frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2) \quad (2.7.20)$$

Substituting (2.7.19) into (2.7.20), we obtain

$$P_L = \frac{1}{2} |b_{TH}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT} \Gamma_L|^2} \quad (2.7.21)$$

The power available from the network P_{AVN} is equal to the power delivered to the load when $\Gamma_L = \Gamma_{OUT}^*$. With $\Gamma_L = \Gamma_{OUT}^*$, (2.7.21) gives

$$P_{AVN} = P_L |_{\Gamma_L = \Gamma_{OUT}^*} = \frac{\frac{1}{2} |b_{TH}|^2}{1 - |\Gamma_{OUT}|^2} \quad (2.7.22)$$

Substituting (2.7.22) into (2.7.21), we can express P_L in the form

$$P_L = P_{AVN} \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_{OUT}|^2)}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.7.23)$$

Equation (2.7.23) can be written in the form

$$P_L = P_{AVN} M_L \quad (2.7.24)$$

where

$$M_L = \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_{OUT}|^2)}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.7.25)$$

Observe the similarity between (2.7.24) and (2.7.16). The factor M_L is known as the load mismatch factor. This factor is used to quantize what portion of P_{AVN} is delivered to the load. For $\Gamma_L = \Gamma_{OUT}^*$, (2.7.25) gives $M_L = 1$ and it follows that $P_L = P_{AVN}$. This fact is expressed in the form

$$P_L = P_{AVN} |_{\Gamma_L = \Gamma_{OUT}^*}$$

The power gain is given by

$$G_p = \frac{P_L}{P_{IN}}$$

Therefore, using (2.7.9) and (2.7.20), we can write

$$G_p = \frac{P_L}{P_{IN}} = \frac{\frac{1}{2} |b_2|^2 (1 - |\Gamma_L|^2)}{\frac{1}{2} |a_1|^2 (1 - |\Gamma_{IN}|^2)} \quad (2.7.26)$$

Then, substituting (2.7.6) into (2.7.26), we obtain

$$G_p = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (2.7.27)$$

The transducer power gain is given by

$$G_T = \frac{P_L}{P_{AVS}} = \frac{P_L}{P_{IN} P_{AVS}} = G_p \frac{P_{IN}}{P_{AVS}} \quad (2.7.28)$$

From (2.7.16), the ratio P_{IN}/P_{AVS} is M_s ; therefore, (2.7.28) can be written as

$$G_T = G_p M_s \quad (2.7.29)$$

Substituting (2.7.17) and (2.7.27) into (2.7.29), we obtain

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_s \Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (2.7.30)$$

Manipulating the denominator, (2.7.30) can also be written in the form

$$G_T = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{OUT}\Gamma_L|^2} \quad (2.7.31)$$

The available power gain can be expressed in the form

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{P_L}{P_{AVS}} \frac{P_{AVN}}{P_L} = \frac{G_T}{M_L}$$

Then, using (2.7.25) and (2.7.31), we can write G_A in the form

$$G_A = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{OUT}|^2} \quad (2.7.32)$$

It is of interest to investigate further the significance of the mismatch factors M_s and M_L . From (2.7.16), the factor M_s relates P_{AVS} to P_{IN} . Consider the network in Fig. 2.7.4a. For this network,

$$P_{AVS} = \frac{1}{8} \frac{|E_s|^2}{R_s}$$

and the input power is given by

$$P_{IN} = \frac{1}{2} \left| \frac{E_s}{Z_s + Z_{IN}} \right|^2 R_{IN} = \frac{1}{8} \frac{|E_s|^2}{R_s} \left(\frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} \right) = P_{AVS} \left(\frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} \right) \quad (2.7.33)$$

Comparing (2.7.33) with (2.7.16), it follows that the mismatch factor M_s is the term in parentheses in (2.7.33)—namely,

$$M_s = \frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} \quad (2.7.34)$$

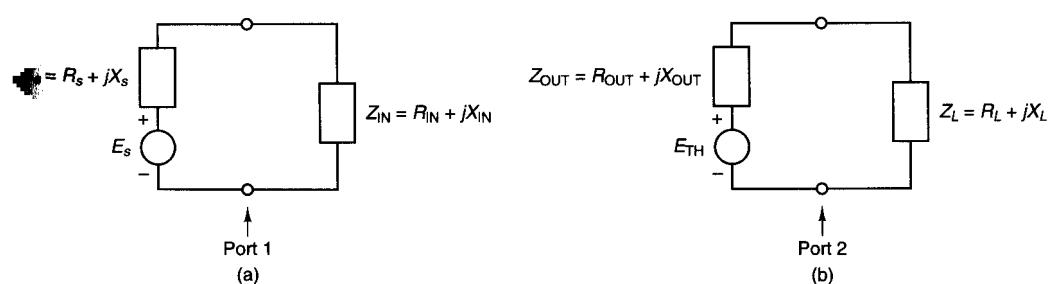


Figure 2.7.4 (a) The input port; (b) the output port.

Although it is not obvious at first sight, (2.7.34) and (2.7.17) are identical. To prove this, we use a normalizing impedance Z_o and write

$$Z_s = Z_o \frac{1 + \Gamma_s}{1 - \Gamma_s}$$

Then, the real part of Z_s is

$$R_s = \frac{1}{2} (Z_s + Z_s^*) = \frac{Z_o}{2} \left[\frac{1 + \Gamma_s}{1 - \Gamma_s} + \frac{1 + \Gamma_s^*}{1 - \Gamma_s^*} \right] = Z_o \frac{1 - |\Gamma_s|^2}{|1 - \Gamma_s|^2} \quad (2.7.35)$$

Similarly, we express the real part of Z_{IN} as

$$R_{IN} = Z_o \frac{1 - |\Gamma_{IN}|^2}{|1 - \Gamma_{IN}|^2} \quad (2.7.36)$$

Substituting (2.7.35) and (2.7.36) into (2.7.34), and simplifying a bit, it follows that

$$M_s = \frac{4R_s R_{IN}}{|Z_s + Z_{IN}|^2} = \frac{(1 - |\Gamma_{IN}|^2)(1 - |\Gamma_s|^2)}{|1 - \Gamma_s \Gamma_{IN}|^2}$$

At the output port in Fig. 2.7.4b, the factor M_L is given by

$$M_L = \frac{4R_{OUT} R_L}{|Z_{OUT} + Z_L|^2}$$

which is identical to (2.7.25).

Power-Gain Expressions in Terms of S_p Parameters

Consider the two-port network in Fig. 2.7.5, where the two-port network is characterized by its S_p parameters. In Section 1.7, the expression for G_T in terms of S_p parameters was derived [see (1.7.27)]. That is,

$$G_T = \frac{P_L}{P_{AVS}} = |S_{p21}|^2 \quad (2.7.37)$$

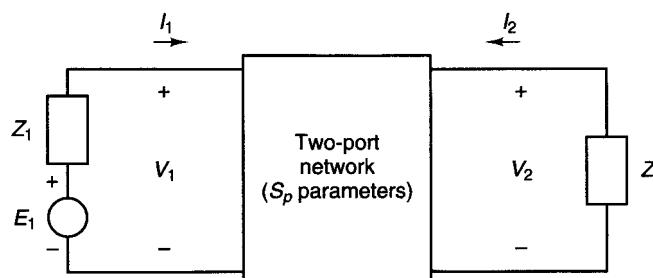


Figure 2.7.5 Two-port network characterized by its S_p parameters.

Using (2.7.28), the power gain G_p can be expressed as

$$G_p = G_T \frac{P_{AVS}}{P_{IN}} \quad (2.7.38)$$

In Fig. 2.7.5, the input power is given by

$$P_{IN} = P_{AVS}(1 - |S_{p11}|^2) \quad (2.7.39)$$

Substituting (2.7.37) and (2.7.39) into (2.7.38) gives

$$G_p = \frac{|S_{p21}|^2}{1 - |S_{p11}|^2} \quad (2.7.40)$$

Using (2.7.24), the power available from the network can be expressed as

$$G_A = \frac{G_T}{M_L} = \frac{G_T}{P_L} P_{AVN} \quad (2.7.41)$$

The power delivered to the load Z_L in Fig. 2.7.5 is given by

$$P_L = P_{AVN}(1 - |S_{p22}|^2) \quad (2.7.42)$$

Substituting (2.7.37) and (2.7.42) into (2.7.41) gives

$$G_A = \frac{|S_{p21}|^2}{1 - |S_{p22}|^2} \quad (2.7.43)$$

The power-gain expressions in terms of S_p parameters are given in (2.7.37), (2.7.40), and (2.7.43). Next, we will show that these relations are equivalent to the power-gain expressions in terms of S parameters given in (2.7.30), (2.7.27), and (2.7.32).

The derivation of the S_p parameters in terms of the S parameters is a difficult one. It consists of expressing the S parameters and S_p parameters in terms of z parameters and then eliminating the z parameters between the two relations to obtain the relation between the S_p parameters in terms of S parameters. The results are as follows:

$$S_{p11} = \frac{(1 - \Gamma_s)}{(1 - \Gamma_s^*)} \frac{(1 - \Gamma_L S_{22})(S_{11} - \Gamma_s^*) + S_{12} S_{21} \Gamma_L}{D} \quad (2.7.44)$$

$$S_{p22} = \frac{(1 - \Gamma_L)}{(1 - \Gamma_L^*)} \frac{(1 - \Gamma_s S_{11})(S_{22} - \Gamma_L^*) + S_{12} S_{21} \Gamma_s}{D} \quad (2.7.45)$$

$$S_{p21} = \frac{(1 - \Gamma_s)}{|1 - \Gamma_s|} \frac{|1 - \Gamma_L| S_{21} [(1 - |\Gamma_s|^2)(1 - |\Gamma_L|^2)]^{1/2}}{(1 - \Gamma_s^*)(1 - \Gamma_L^*) D} \quad (2.7.46)$$

and

$$S_{p12} = \frac{(1 - \Gamma_L)}{|1 - \Gamma_L|} \frac{|1 - \Gamma_s| S_{12} [(1 - |\Gamma_s|^2)(1 - |\Gamma_L|^2)]^{1/2}}{(1 - \Gamma_s^*)(1 - \Gamma_L^*) D} \quad (2.7.47)$$

where

$$D = (1 - \Gamma_s S_{11})(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_s$$

Substituting (2.7.46) into (2.7.37), we obtain the expression for G_T in terms of S parameters—namely,

$$G_T = |S_{p21}|^2 = \frac{1 - |\Gamma_s|^2}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{\text{OUT}}\Gamma_L|^2}$$

which is recognized as (2.7.30). Similarly, substituting (2.7.44) and (2.7.46) into (2.7.40), we obtain the expression for G_p in (2.7.27); and substituting (2.7.45) and (2.7.46) into (2.7.43) gives the expression for G_A in (2.7.32).

2.8 VSWR CALCULATIONS

The VSWR in a transmission line is given by (1.3.44). Since the power delivered to the load in a transmission line, excited by a source with impedance Z_o , is given by

$$P_L = P_{\text{AVS}}(1 - |\Gamma_0|^2)$$

it follows that the magnitude of the load reflection coefficient, $|\Gamma_0|$ (which is related to the VSWR), provides a measure of what portion of P_{AVS} is delivered to the load.

For example, for a VSWR = 1 it follows that $|\Gamma_0| = 0$ and $P_L = P_{\text{AVS}}$ (i.e., all the incident power is delivered to the load). For a VSWR = 1.5, it follows that $|\Gamma_0| = 0.2$ and the ratio of the incident to the reflected power from the load is $|\Gamma_0|^2 = 0.04$, or 4%. Hence, 4% of the incident power is reflected by the load (96% of the incident power is delivered to the load). For a VSWR = 2, then $|\Gamma_0| = 1/3$ and $|\Gamma_0|^2 = 0.11$, or 11%. In this case, 11% of the incident power is reflected by the load.

The values of the input and output VSWR are important to the microwave amplifier designer. For example, many microwave amplifiers require the input VSWR to be less than 1.5. In other designs, higher VSWR values must be tolerated in order to obtain other performances, such as a particular noise performance.

Figure 2.8.1 shows the input portion of a microwave amplifier. The reflection coefficient at the input of the lossless matching network, normalized to Z_o , is denoted by Γ_a . The input VSWR, denoted by $(\text{VSWR})_{\text{in}}$, is related to Γ_a by

$$(\text{VSWR})_{\text{in}} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} \quad (2.8.1)$$

where

$$\Gamma_a = \frac{Z_a - Z_o}{Z_a + Z_o}$$

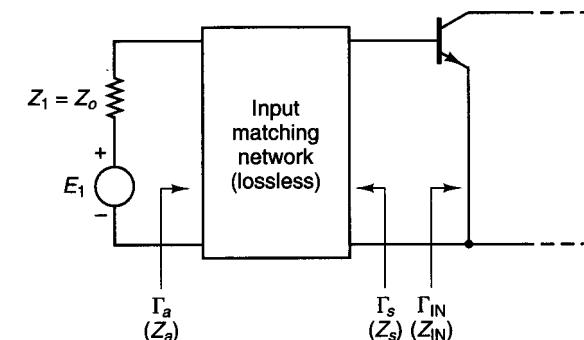


Figure 2.8.1 Input portion of a microwave amplifier.

Since the input power is given by

$$P_{\text{IN}} = P_{\text{AVS}}(1 - |\Gamma_a|^2)$$

and from (2.7.16) we have

$$P_{\text{IN}} = P_{\text{AVS}}M_s$$

it follows that

$$M_s = 1 - |\Gamma_a|^2$$

or

$$|\Gamma_a| = \sqrt{1 - M_s} \quad (2.8.2)$$

Equations (2.8.1) and (2.8.2) show that $(\text{VSWR})_{\text{in}}$ can be calculated from a knowledge of the mismatch factor M_s .

Substituting (2.7.17) into (2.8.2) and simplifying results in the following expression for $|\Gamma_a|$:

$$|\Gamma_a| = \sqrt{1 - \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_{\text{IN}}|^2)}{|1 - \Gamma_s\Gamma_{\text{IN}}|^2}} = \left| \frac{\Gamma_{\text{IN}} - \Gamma_s^*}{1 - \Gamma_{\text{IN}}\Gamma_s} \right| \quad (2.8.3)$$

This relation shows that $|\Gamma_a|$ can be calculated from a knowledge of Γ_{IN} and Γ_s .

The concept of a reflection coefficient and of a VSWR can be used in a microwave amplifier even if there are no transmission lines. For example, in Fig. 2.8.1 this can occur when the input matching network is implemented using lumped components. The reflection coefficients and $(\text{VSWR})_{\text{in}}$ in Fig. 2.8.1 can be thought of as those of a transmission line of zero length with characteristic impedance Z_o .

Similar relations can be defined for the output VSWR. The output portion of a microwave amplifier is shown in Fig. 2.8.2. For Fig. 2.8.2, we have

$$(\text{VSWR})_{\text{out}} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} \quad (2.8.4)$$

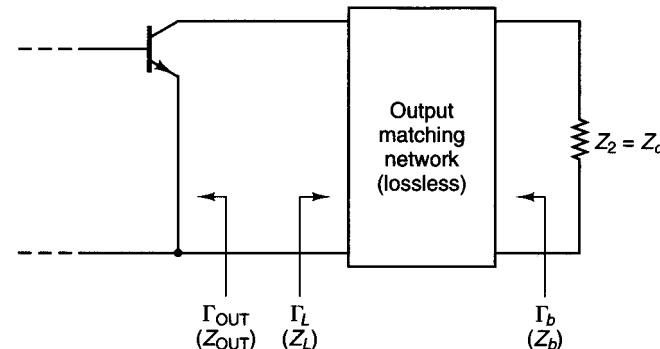


Figure 2.8.2 Output portion of a microwave amplifier.

$$\Gamma_b = \frac{Z_b - Z_o}{Z_b + Z_o}$$

$$|\Gamma_b| = \sqrt{1 - M_L} \quad (2.8.5)$$

and

$$|\Gamma_b| = \sqrt{1 - \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_{\text{OUT}}|^2)}{|1 - \Gamma_{\text{OUT}}\Gamma_L|^2}} = \left| \frac{\Gamma_{\text{OUT}} - \Gamma_L^*}{1 - \Gamma_{\text{OUT}}\Gamma_L} \right| \quad (2.8.6)$$

Example 2.8.1

(a) The input portion of the microwave amplifier in Fig. 2.5.14 is drawn in Fig. 2.8.3a. Assume that the S parameters of the transistor are such that the resulting Γ_{IN} is $0.614| -160^\circ$. Calculate $(\text{VSWR})_{\text{in}}$.

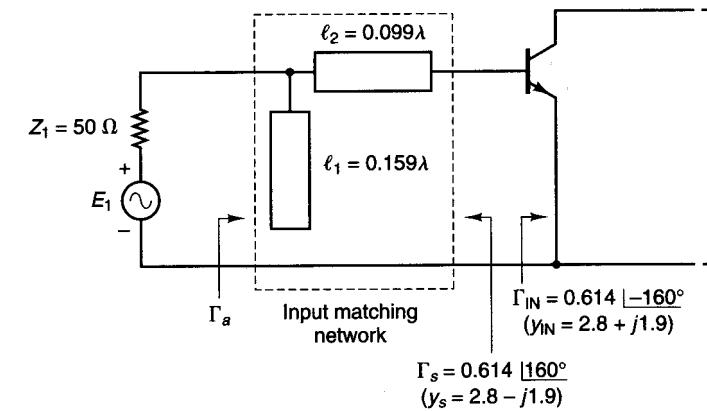
(b) Calculate $(\text{VSWR})_{\text{in}}$ in Fig. 2.8.3a if Γ_{IN} is $0.4| -145^\circ$.

Solution. (a) From (2.8.3), with $\Gamma_s = \Gamma_{\text{IN}}^* = 0.614|160^\circ$, it follows that $|\Gamma_a| = 0$. Therefore, using (2.8.1), we obtain $(\text{VSWR})_{\text{in}} = 1$.

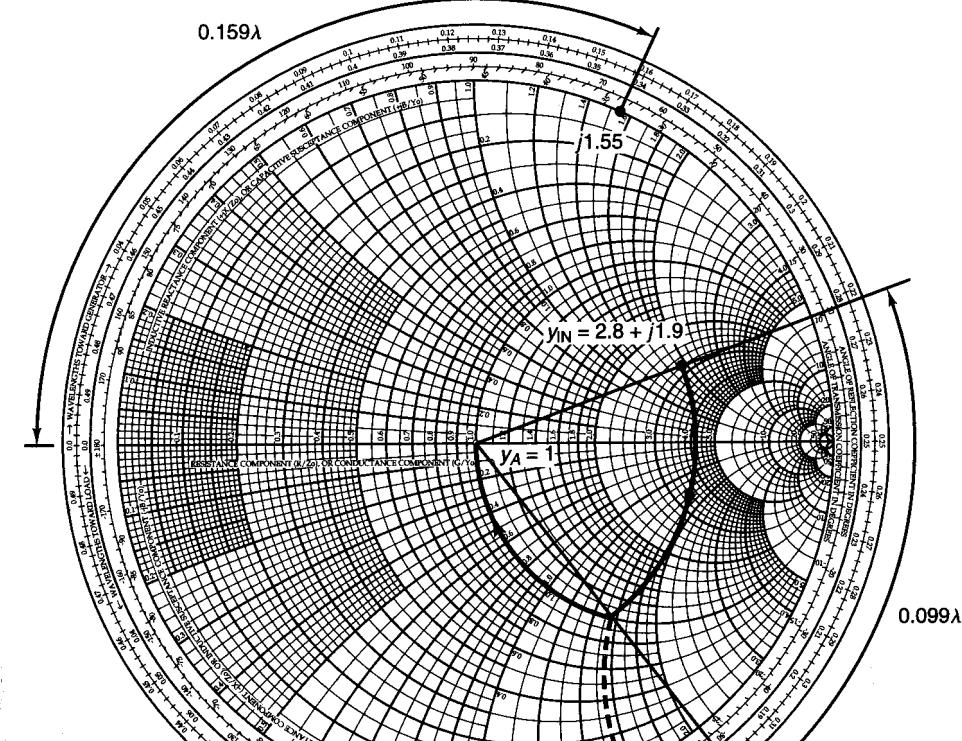
The previous calculation illustrates the fact that when $\Gamma_s = \Gamma_{\text{IN}}^*$, a conjugate matched condition exists at the input of the transistor and at the source. In other words, $Z_a = 50 \Omega$ (since $|\Gamma_a| = 0$) and the source sees a matched input impedance. This can also be verified by evaluating Z_a for the circuit in Fig. 2.8.3a. In Fig. 2.8.3b, we have plotted the normalized input admittance $y_{\text{IN}} = 1/z_{\text{IN}} = 2.8 + j1.9$ and calculated y_a to be 1, thus verifying that $Z_a = 50 \Omega$.

(b) In this case, Γ_{IN} is given as $0.4| -145^\circ$. Therefore, $\Gamma_s \neq \Gamma_{\text{IN}}^*$, so the input of the transistor is not conjugate matched for maximum power transfer. Using (2.8.3), we obtain

$$|\Gamma_a| = \left| \frac{0.4| -145^\circ - 0.614| -160^\circ}{1 - 0.4| -145^\circ(0.614|160^\circ)} \right| = 0.327$$



(a)



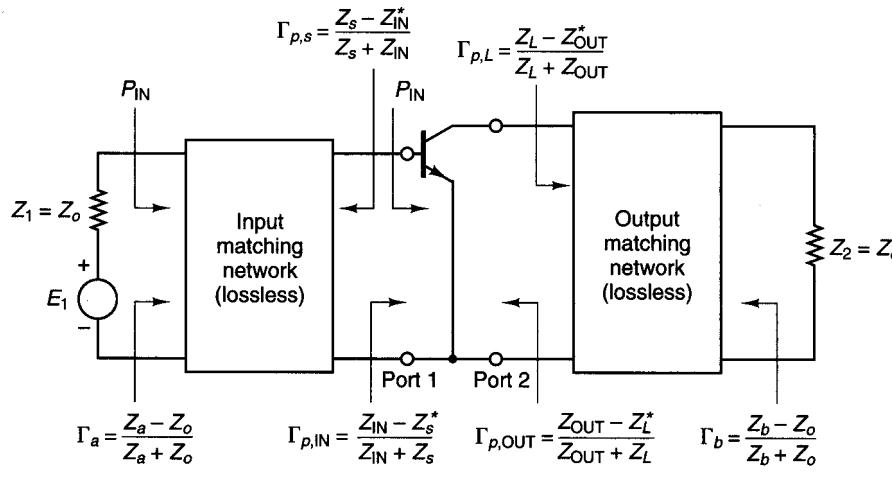
(b)

Figure 2.8.3 (a) Circuit for Example 2.8.1; (b) calculation of Z_a using the Smith chart.

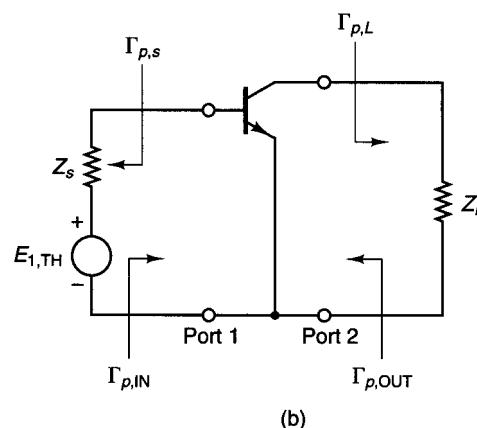
Therefore, the value of $(\text{VSWR})_{\text{in}}$ is

$$(\text{VSWR})_{\text{in}} = \frac{1 + 0.327}{1 - 0.327} = 1.97$$

Let us proceed a little further with the analysis of the VSWR. Consider Fig. 2.8.4a, where the matching networks are assumed to be lossless. The indicated coefficients Γ_a , Γ_b , $\Gamma_{p,\text{IN}}$, and $\Gamma_{p,\text{OUT}}$ are power reflection coefficients. Since the impedances Z_1 and Z_2 are real (i.e., $Z_1 = Z_2 = Z_o$), it follows that Γ_a and Γ_b are identical to the traveling-wave reflection coefficients. The equivalent circuits at the input and output ports of the transistor are shown in Fig. 2.8.4b.



(a)



(b)

Figure 2.8.4 (a) Power reflection coefficients in a microwave amplifier; (b) an equivalent circuit.

The input power at the generator input in Fig. 2.8.4a can be written as

$$P_{\text{IN}} = P_{\text{AVS}}(1 - |\Gamma_a|^2) = P_{\text{AVS}}M_s \quad (2.8.7)$$

Since the matching circuits are assumed to be lossless, at the input of the transistor we can write

$$P_{\text{IN}} = P_{\text{AVS}}(1 - |\Gamma_{p,\text{IN}}|^2) = P_{\text{AVS}}M'_s \quad (2.8.8)$$

where $\Gamma_{p,\text{IN}}$ is the power reflection coefficient at the input of the transistor—namely,

$$\Gamma_{p,\text{IN}} = \frac{Z_{\text{IN}} - Z_s^*}{Z_{\text{IN}} + Z_s} \quad (2.8.9)$$

and

$$M'_s = 1 - |\Gamma_{p,\text{IN}}|^2$$

In writing (2.8.8), we used the fact that P_{AVS} is the same at the input and output ports of the lossless matching circuit. The parameter $\Gamma_{p,\text{IN}}$ is identical to the value of S_{p11} associated with the two-port network in Fig. 2.8.4b.

Equating (2.8.7) to (2.8.8) gives

$$M_s = M'_s \quad (2.8.10)$$

or

$$|\Gamma_a| = |\Gamma_{p,\text{IN}}|$$

It also follows that $|\Gamma_{p,\text{IN}}| = |\Gamma_{p,s}|$. From the previous results, $(\text{VSWR})_{\text{in}}$ can also be expressed in terms of $\Gamma_{p,\text{IN}}$ as

$$(\text{VSWR})_{\text{in}} = \frac{1 + |\Gamma_{p,\text{IN}}|}{1 - |\Gamma_{p,\text{IN}}|} \quad (2.8.11)$$

Equation (2.8.10) shows that the mismatch factor is an invariant quantity in a lossless matching network. That is, the mismatch factor M_s at the input of the input matching network is equal to the mismatch factor M'_s at the output of the input matching network. Using (2.7.34), we can express (2.8.10) in terms of impedances as

$$\frac{4Z_oR_a}{|Z_o + Z_a|^2} = \frac{4R_sR_{\text{IN}}}{|Z_s + Z_{\text{IN}}|^2}$$

where $R_a = \text{Re}[Z_a]$, $R_s = \text{Re}[Z_s]$, and $R_{\text{IN}} = \text{Re}[Z_{\text{IN}}]$.

Similarly, at the output of the amplifier in Fig. 2.8.4a, we can write

$$M_L = M'_L$$

$$M_L = 1 - |\Gamma_b|^2 = \frac{4Z_oR_b}{|Z_o + Z_b|^2}$$

$$M'_L = 1 - |\Gamma_{p,\text{OUT}}|^2 = \frac{4R_L R_{\text{OUT}}}{|Z_L + Z_{\text{OUT}}|^2}$$

$$|\Gamma_b| = |\Gamma_{p,\text{OUT}}| = |\Gamma_{p,L}|$$

$$\Gamma_{p,\text{OUT}} = \frac{Z_{\text{OUT}} - Z_L^*}{Z_{\text{OUT}} + Z_L}$$

and

$$(\text{VSWR})_{\text{out}} = \frac{1 + |\Gamma_{p,\text{OUT}}|}{1 - |\Gamma_{p,\text{OUT}}|} \quad (2.8.12)$$

where $R_b = \text{Re}[Z_b]$, $R_L = \text{Re}[Z_L]$, and $R_{\text{OUT}} = \text{Re}[Z_{\text{OUT}}]$. The parameter $\Gamma_{p,\text{OUT}}$ is identical to the value of S_{p22} associated with the two-port network in Fig. 2.8.4b.

Example 2.8.2

- (a) Calculate the $(\text{VSWR})_{\text{in}}$ in Example 2.8.1, part (a), using (2.8.11).
- (b) Calculate the $(\text{VSWR})_{\text{in}}$ in Example 2.8.1, part (b), using (2.8.11). Also, evaluate M'_s .

Solution. (a) With $\Gamma_s = \Gamma_{\text{IN}}^*$, it follows that $Z_s = Z_{\text{IN}}^* = 12.308 + j8.297 \Omega$. Therefore, from (2.8.9) it follows that $\Gamma_{p,\text{IN}} = 0$ and from (2.8.11), $(\text{VSWR})_{\text{in}} = 1$.

(b) With $\Gamma_{\text{IN}} = 0.4|-145^\circ$ (or $Z_{\text{IN}} = 23.136 - j12.638 \Omega$), it follows from (2.8.9) that

$$\Gamma_{p,\text{IN}} = \frac{Z_{\text{IN}} - Z_s^*}{Z_{\text{IN}} + Z_s} = \frac{23.136 - j12.638 - (12.308 - j8.297)}{23.136 - j12.638 + (12.308 + j8.297)} = 0.327|-14.86^\circ$$

and

$$(\text{VSWR})_{\text{in}} = \frac{1 + 0.327}{1 - 0.327} = 1.97$$

As expected, $|\Gamma_a| = |\Gamma_{p,\text{IN}}| = |\Gamma_{p,s}| = 0.327$.

The mismatch factor M_s is

$$M_s = M'_s = 1 - |\Gamma_{p,\text{IN}}|^2 = 1 - (0.327)^2 = 0.893$$

PROBLEMS

- 2.1 (a)** Locate in the Z Smith chart the impedance $Z = 100 + j100 \Omega$ using a reference impedance of 50Ω .
- (b)** Find the value of the normalized admittance y using the Z Smith chart, and evaluate the value of Y (i.e., $Y = yY_o = y/Z_o$).
- (c)** Find y using the ZY Smith chart.
- (d)** Repeat parts (a) to (c) for the following impedances: $Z = 50 - j100 \Omega$, $Z = 25 - j25 \Omega$, $Z = j50 \Omega$, and $Z = j0 \Omega$.
- 2.2 (a)** Show that impedances having a negative real part (i.e., $z = -r + jx$) have a reflection coefficient whose magnitude is greater than 1.

Problems

- (b)** Prove that negative resistances can be handled in the Smith chart by plotting $1/\Gamma^*$ and interpreting the resistance circles as being negative and the reactance circles as marked.
- (c)** Locate in the Smith chart the impedances $Z_1 = -20 + j16 \Omega$ and $Z_2 = -200 + j25 \Omega$ and find the associated reflection coefficient. Normalize the impedances to 50Ω .
- (d)** Work the problem in part (c) in the compressed Smith chart.
- (e)** What is the value of $|\Gamma|$ on the boundary of the compressed Smith chart in Fig. 2.2.5?

- 2.3** Show that the impedance along a transmission line repeats itself at every $\lambda/2$ distance. That is,

$$Z(d) = Z\left(d + \frac{n\lambda}{2}\right), \quad n = 1, 2, 3, \dots$$

- 2.4** Show that the impedance along a transmission line can be expressed in the form

$$Z(d) = R(d) + jX(d) = |Z(d)|e^{j\theta_d}$$

where

$$R(d) = Z_o \frac{1 - |\Gamma|^2}{1 - 2|\Gamma| \cos \phi + |\Gamma|^2}$$

$$X(d) = Z_o \frac{2|\Gamma| \sin \phi}{1 - 2|\Gamma| \cos \phi + |\Gamma|^2}$$

$$|Z(d)| = Z_o \sqrt{\frac{1 + 2|\Gamma| \cos \phi + |\Gamma|^2}{1 - 2|\Gamma| \cos \phi + |\Gamma|^2}}$$

$$\theta_d = \tan^{-1} \frac{X(d)}{R(d)} = \tan^{-1} \left(\frac{2|\Gamma| \sin \phi}{1 - |\Gamma|^2} \right)$$

$$\Gamma = |\Gamma_0|e^{j\phi}, \quad \Gamma_0 = |\Gamma_0|e^{j\phi_l}, \quad \phi = \phi_l - 2\beta d$$

- 2.5** Find the input impedance, the load reflection coefficient, and the VSWR in a transmission line having an electrical length of 90° , $Z_o = 50 \Omega$, and terminated in the load $Z_L = 50 + j100 \Omega$. Work the problem in both the Z and Y Smith charts.
- 2.6 (a)** Prove that the maximum normalized resistance in a transmission line is numerically equal to the VSWR.
- (b)** Prove that the minimum normalized resistance in a transmission line is numerically equal to $1/\text{VSWR}$.
- 2.7 (a)** Determine the length l of the $50-\Omega$ short-circuited transmission line shown in Fig. 2.2.8a so that the input impedance is $Z_{\text{IN}}(l) = -j25 \Omega$.
- (b)** Determine the length l of the $50-\Omega$ open-circuited transmission line shown in Fig. 2.2.8b so that the normalized input admittance is $y_{\text{IN}}(l) = j2$.
- 2.8** The normalized admittance (with $Z_o = 50 \Omega$) of a one-port network is shown in Fig. P2.8 as the frequency varies from 500 MHz to 1 GHz. Determine an equivalent circuit for the one-port network and the element values.
- 2.9 (a)** In the series RC circuit shown in Fig. 2.3.2, the value of C is 50 pF. Determine the values of f_a and f_b .

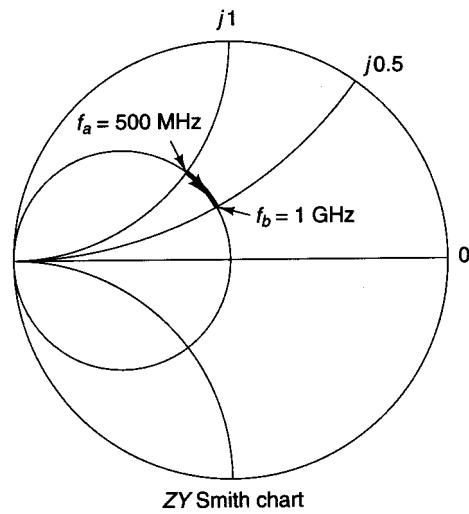


Figure P2.8

- (b) In Example 2.3.2, assume that at $f_b = 1 \text{ GHz}$ the normalized impedance is $j0.6$. At $f_a = 500 \text{ MHz}$, the normalized impedance remains at $j0.2$. Determine an equivalent circuit for the one-port network and the element values. Observe that in this problem the reactance does not increase linearly with frequency.
- 2.10 Design four different Ell matching networks to match the load $Z_{\text{LOAD}} = 10 + j40 \Omega$ to a 50Ω transmission line.
- 2.11 Design the matching network in Fig. P2.11 that provides $Y_L = (4 - j4) \times 10^{-3} \text{ S}$ to the transistor. Find the element values at 700 MHz.

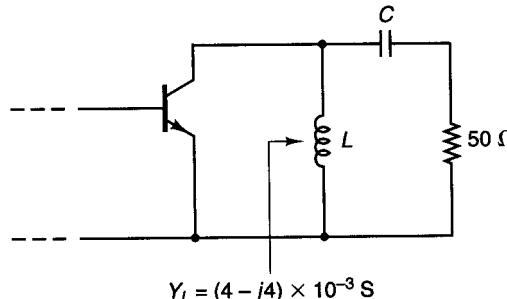


Figure P2.11

- 2.12 Two types of Ell matching networks are shown in Fig. P2.12. Select one that can match the load $Y_{\text{LOAD}} = (8 - j12) \times 10^{-3} \text{ S}$ to a 50Ω transmission line. Find the element values at $f = 1 \text{ GHz}$.

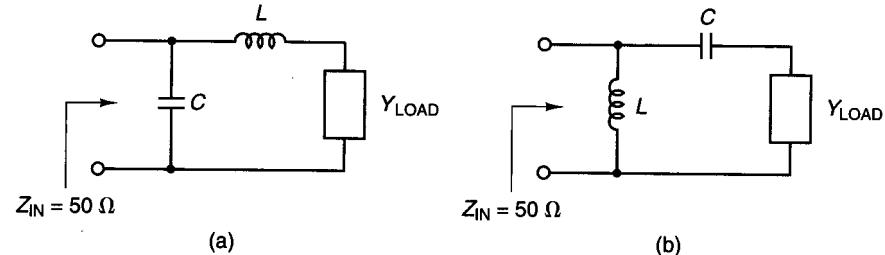


Figure P2.12

- 2.13 Show the impedance-admittance path in the ZY Smith chart for the circuit shown in Fig. P2.13 at $\omega = 10^9 \text{ rad/s}$, and evaluate Z_{IN} .

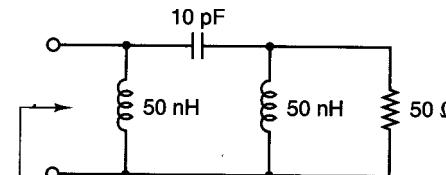


Figure P2.13

- 2.14 Design a two-element matching network to transform the load $Z_{\text{LOAD}} = 100 - j100 \Omega$ to an input impedance of $Z_{\text{IN}} = 25 + j25 \Omega$.
- 2.15 (a) Design a Tee matching network to transform $Z_{\text{LOAD}} = 50 \Omega$ to the input impedance $Z_{\text{IN}} = 20 + j20 \Omega$ with a Q of 5.
 (b) Design a Pi matching network to transform $Z_{\text{LOAD}} = 50 \Omega$ to the input impedance $Z_{\text{IN}} = 25 \Omega$ with a Q of 2.5.
- 2.16 (a) Use Figs. 2.5.2 and 2.5.3 to calculate W , λ , and ϵ_{eff} for a characteristic impedance of 50Ω using RT/Duroid® with $\epsilon_r = 2.23$ and $h = 0.7874 \text{ mm}$.
 (b) Use (2.5.8) through (2.5.11) to show that for RT/Duroid® with $\epsilon_r = 2.23$ and $h = 0.7874 \text{ mm}$, a 50Ω characteristic impedance is obtained with $W/h = 3.073$. Also, $\epsilon_{\text{eff}} = 1.91$ and $\lambda = 0.7236\lambda_0$.
- 2.17 The 10-nH inductor shown in Fig. P2.17 is to be implemented at 1 GHz using an open-circuited microstrip line with $Z_o = 50 \Omega$. The microstrip material has $\epsilon_r = 6$ and $h = 25 \text{ mils}$. Determine the width and length of the microstrip line.

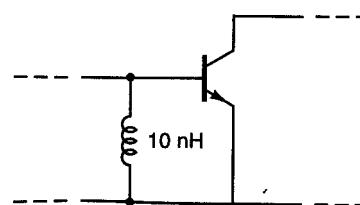


Figure P2.17

- 2.18** In the amplifier shown in Fig. 2.5.16b, calculate the width and the length of the lines at $f = 1$ GHz

- 2.19** (a) Design the matching circuit shown in Fig. P2.19 to transform the 50Ω load to the input impedance $Z_{IN} = 100 - j100 \Omega$.

- (b) What is the length l_1 if the short-circuited shunt stub is replaced by an open-circuited shunt stub?

- (c) Repeat parts (a) and (b) for $Z_{IN} = 100 + j100 \Omega$.

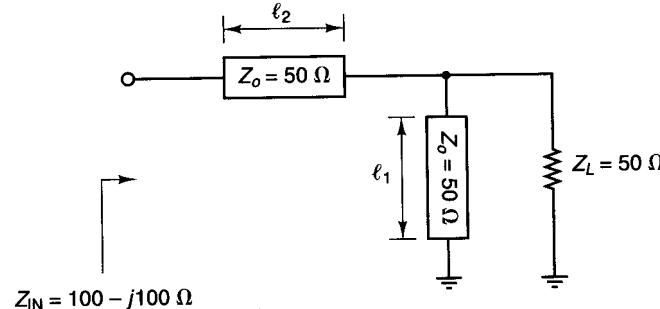


Figure P2.19

- 2.20** (a) Design a single-stub matching system (see Fig. P2.20) to match the load $Z_L = 15 + j25 \Omega$ to a 50Ω transmission line. The characteristic impedance of the short-circuited stub is 50Ω .

- (b) Design the single-stub matching system in Fig. P2.20 assuming that the characteristic impedance of the stub is 100Ω .

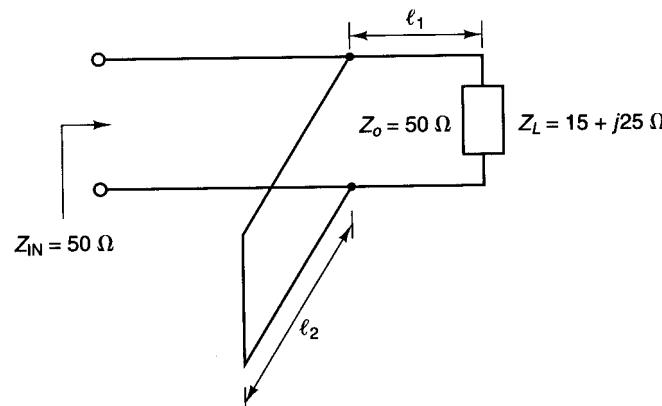


Figure P2.20

- 2.21** Design the matching circuits shown in Fig. P2.21 to transform the 50Ω load impedance to the input admittance given in the figures.

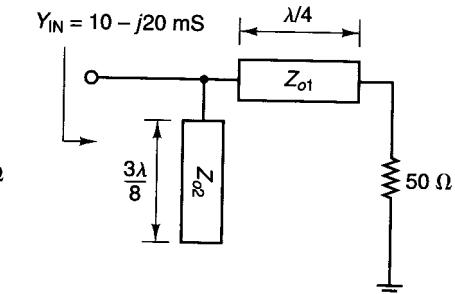
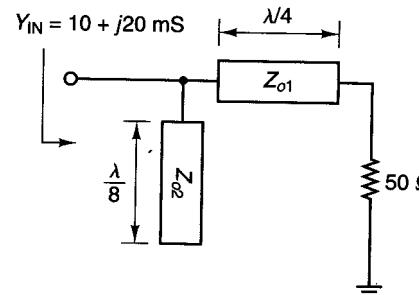
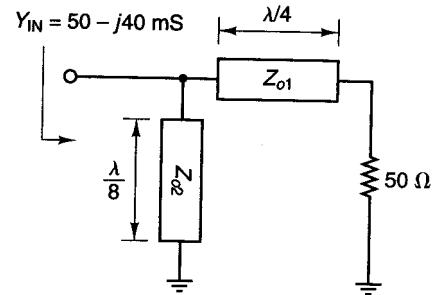
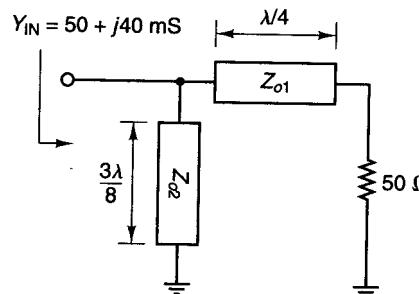


Figure P2.21

- 2.22** (a) Design the matching networks shown in Fig. P2.22 to produce the source reflection coefficient $\Gamma_s = 0.5|90^\circ|$. In Fig. P2.22b, the appropriate length for the short-circuited stub must be selected (i.e., $\lambda/8$ or $3\lambda/8$).

- (b) Design the balance form of the shunt stubs.

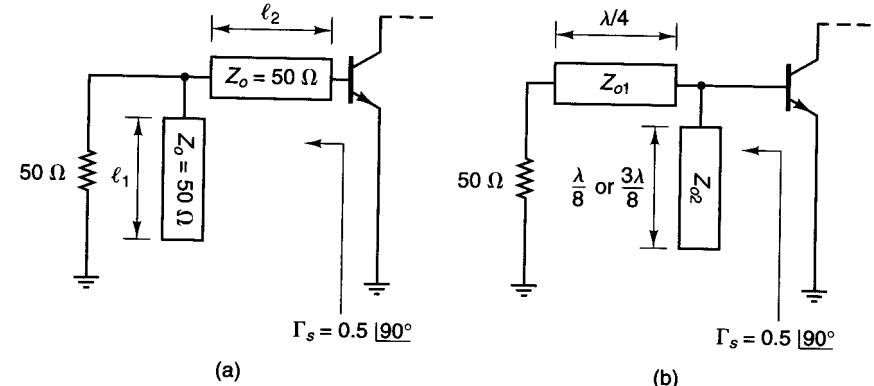


Figure P2.22

- 2.23** In the design shown in Fig. 2.5.16, the balance stubs of length $3\lambda/8$ were designed with characteristic impedances of 52.64Ω and 95.2Ω , respectively. Design the length of the balance stubs if the characteristic impedances are left at 26.32Ω and 47.6Ω , respectively.

2.24 (a) Design the matching network in Fig. P2.24 to produce a load reflection coefficient of $\Gamma_L = 0.4 \angle -120^\circ$ to the transistor. The appropriate length for the balance stubs must be selected (i.e., $\lambda/8$ or $3\lambda/8$).
(b) Design the length of the balance stubs if the impedance Z_{o2} is changed to $Z_{o2}/2$.

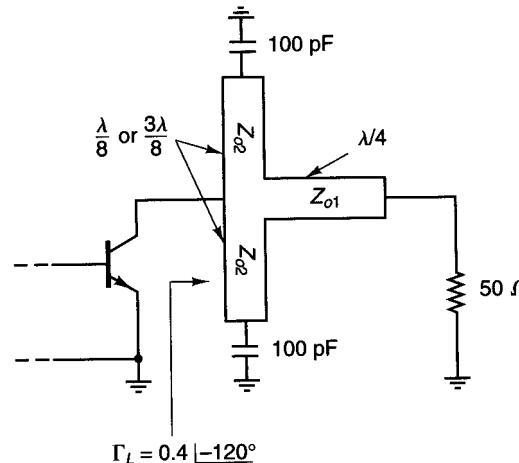


Figure P2.24

- 2.25** Design two microstrip matching networks for an amplifier whose reflection coefficients at $f = 800 \text{ MHz}$, in a $50\text{-}\Omega$ system, are $\Gamma_s = 0.8|160^\circ$ and $\Gamma_L = 0.7|20^\circ$. Show the diagram for the complete amplifier using balanced shunt stubs.

2.26 (a) Determine the value of Γ_s in Fig. P2.26a.
 (b) Determine the value of Γ_s in Fig. P2.26b. Observe that the 100-pF capacitor is a short circuit to the ac signal.

2.27 (a) Determine the value of Γ_L in Fig. P2.27. The lengths shown are for $\epsilon_{ff} = 1$ and $f = 6 \text{ GHz}$.
 (b) Show the balance form of the shunt stubs.

2.28 The input matching network shown in Fig. P2.28 was designed to obtain a certain gain. If the input reflection coefficient is $\Gamma_{IN} = 0.5|100^\circ$, determine the value of the impedance Z_A seen by the source.

2.29 Design a microstrip matching network to transform the load impedance $Z_L = 50 - j50 \Omega$ to the input impedance $Z_{IN} = 25 + j25 \Omega$ in Fig. P2.29.

2.30 Determine the value of Γ_L in Fig. P2.30.

2.31 Design a two-element matching network, as shown in the Smith chart in Fig. P2.31, that produces $\Gamma_s = 0.57|116^\circ$.

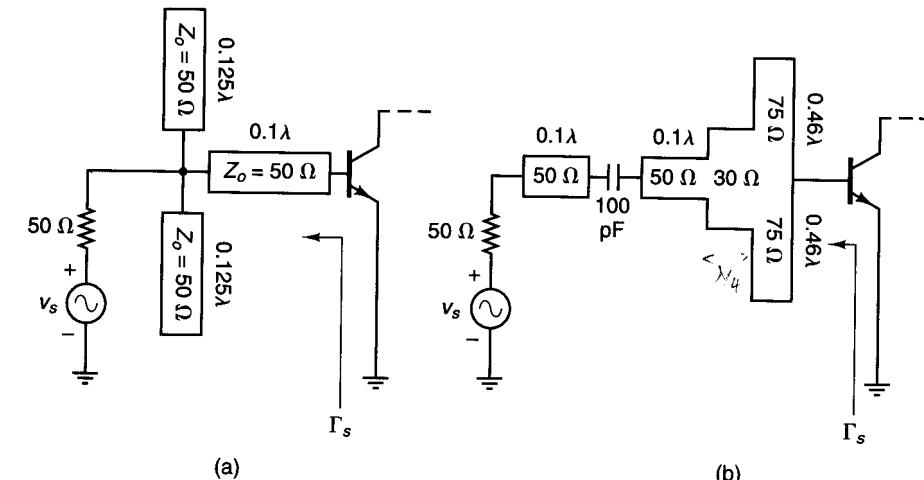


Figure P2.2

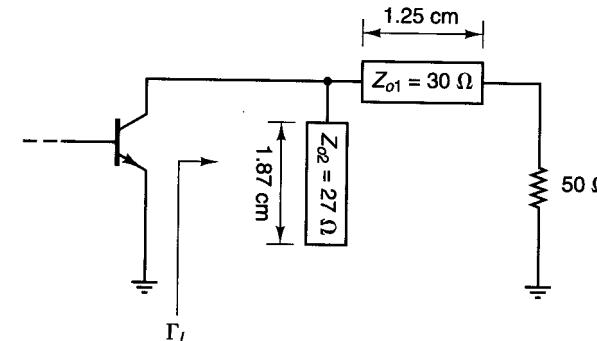


Figure P2.2

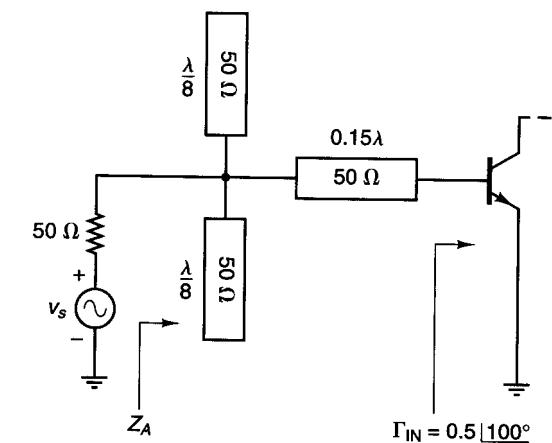


Figure P2.2

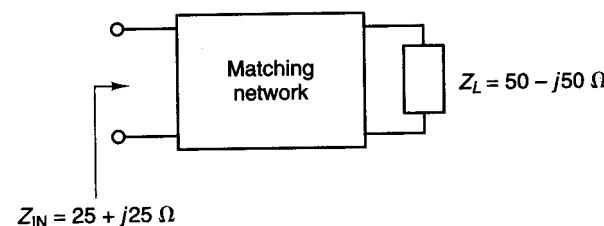


Figure P2.29

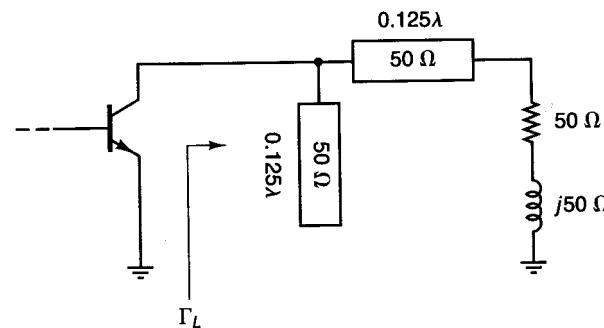


Figure P2.30

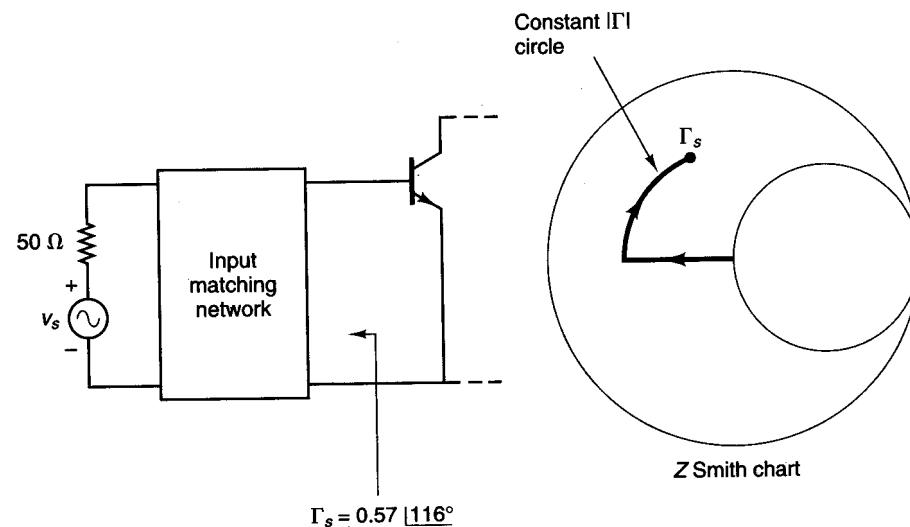


Figure P2.31

- 2.32** Design the matching networks in Fig. P2.32 to match the load $Z_L = 100 + j100 \Omega$ to a $50\text{-}\Omega$ transmission line.

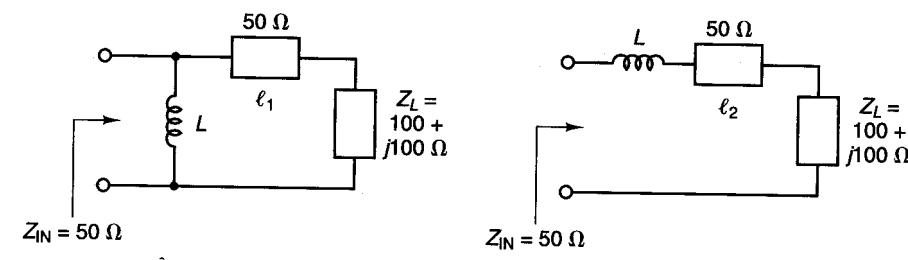


Figure P2.32

- 2.33** Design the matching networks in Fig. P2.33 to match a $50\text{-}\Omega$ load to the impedance Z_{IN} .

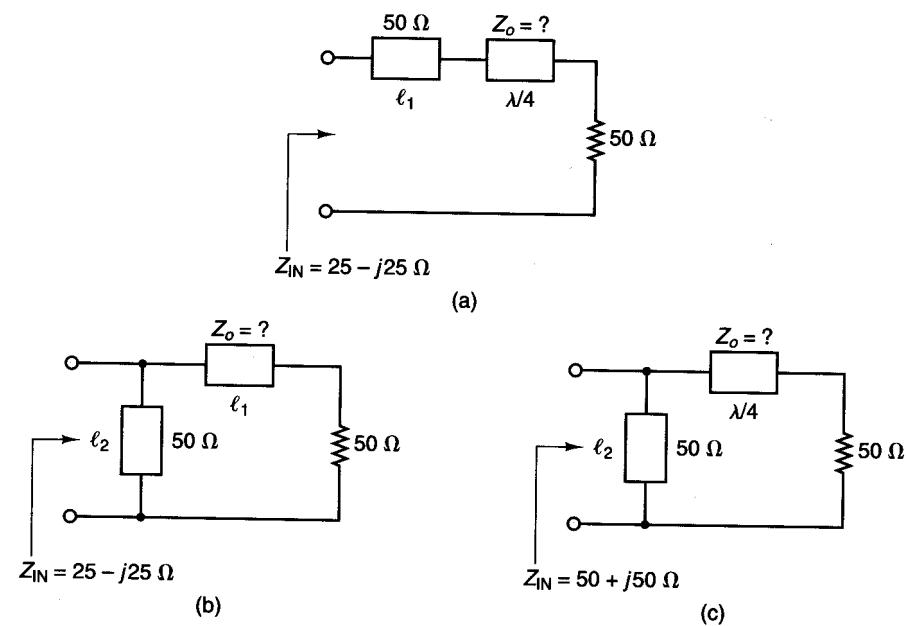


Figure P2.33

- 2.34** Verify the expressions for G_T in (2.6.14) and (2.6.15).
2.34 Verify the expression for A_v in (2.6.23).

- 2.36** (a) The output portion of a microwave amplifier is shown in Fig. P2.36. Calculate $(VSWR)_{out}$ if $\Gamma_{out} = 0.682 \angle -97^\circ$.
 (b) Verify that Z_b is 50Ω when $\Gamma_L = \Gamma_{out}^*$.
 (c) Calculate $(VSWR)_{out}$ if $\Gamma_{out} = 0.5 \angle -60^\circ$.

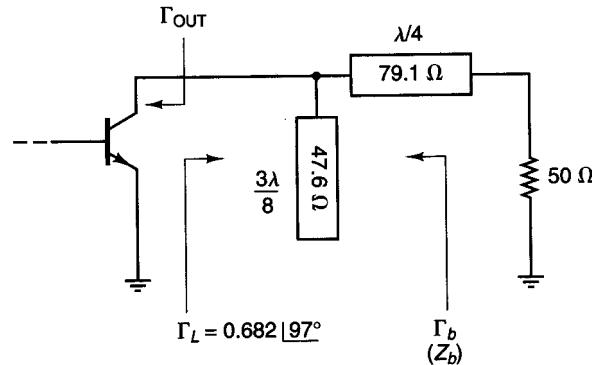


Figure P2.36

REFERENCES

- [2.1] I. J. Bahl and D. K. Trivedi, "A Designer's Guide to Microstrip Line," *Microwaves*, May 1977.
- [2.2] H. Sobol, "Applications of Integrated Circuit Technology to Microwave Frequencies," *Proceedings of the IEEE*, August 1971.
- [2.3] H. Sobol, "Extending IC Technology to Microwave Equipment," *Electronics*, March 1967.

- 2.37** (a) The input portion of a microwave amplifier is shown in Fig. P2.37. Calculate $(VSWR)_{in}$ if $\Gamma_{in} = 0.545 \angle -77.7^\circ$.
 (b) Verify that Z_a is 50Ω when $\Gamma_s = \Gamma_{in}^*$.
 (c) Calculate $(VSWR)_{in}$ if $\Gamma_{in} = 0.4 \angle 45^\circ$.

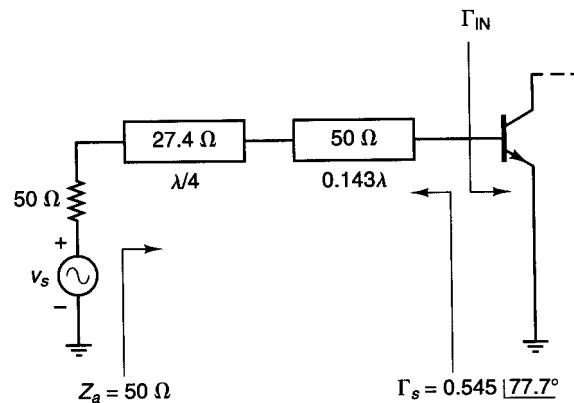


Figure P2.37



Scan page
using app

Understanding the Relevance of Harmonic Impedance Matching in Amplifier Design

Steve Dudkiewicz, Marc Schulze Tenberge and Giampiero Esposito
Maury Microwave Corp., Ontario, Calif.

Travis Barbieri
Freescale Semiconductor, Phoenix, Ariz.

Today's modern commercial and military communication systems are demanding better performance with regards to power, efficiency, linearity and operating bandwidth. As such, extra considerations must be placed on designing the internal components of the systems, including the low-noise and power amplifiers, to maximize performance.

In order to reach higher efficiencies, significant research has been performed on designing high efficiency amplifiers by matching one or more harmonic impedances. An equally large effort has gone into designing commercial test systems which aid in the systematic identification of ideal matching impedances at the fundamental and harmonic frequencies, referred to as harmonic load-pull, in order to maximize performance.

Before venturing into a design project, it is important to ask several questions: Does the application require an amplifier with harmonic matching? If so, which test system is best suited to reach the design goals? This paper explores various types of amplifiers in order to identify which can or cannot take advantage of harmonic matching, and to compare and contrast various harmonic load-pull methodologies as they relate to amplifier design.

HARMONIC IMPEDANCE MATCHING AND ITS RELEVANCE IN AMPLIFIER DESIGN

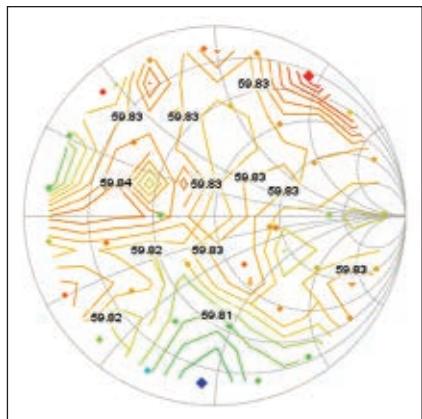
Amplifiers are designed for various applications ranging from highly linear LNAs and PAs operating in Class A condition, to highly nonlinear PAs operating in advanced classes (E, F, G, J and their inverses). Some are designed using unmatched transistors while others are designed using partially (pre-matched

or harmonically-terminated) - or fully-matched components. The design objective and the type of transistor used will dictate the need for harmonic impedance matching.

Amplifiers operating under linear (small-signal) conditions do not produce power at harmonic frequencies, and the output power of a device under test is linearly proportional to its input power. Because no power exists at harmonic frequencies, terminating the harmonic impedances should have no effect on the performance of a stable device under test (DUT). Several mathematical methods using S-parameters, including unilateral design and conjugate match, exist for determining the ideal input and output matches for maximum power and gain.^{1,2} Alternatively, fundamental-frequency load-pull can be used to identify ideal matching impedances for a given figure of merit.

Modern commercial and military systems may be required to operate over several octaves or over a decade in order to meet the frequency spectrum requirements of the application. As such, wideband amplifiers play a critical role in the overall performance of the radio or radar system, often dictating output power or gain flatness over the bandwidth. When designing an amplifier, the match at the fundamental frequency strongly influences the power and gain performance parameters. When dealing with wideband amplifiers, the ideal impedance match must be determined for subsets of the overall frequency range, and the matching network synthesized to achieve the desired wideband response. In this case, it is entirely possible that the harmonic frequencies ($2f_0$, $3f_0$...) of the lower frequency band overlap a fundamental frequency in the middle or upper por-

Technical Feature

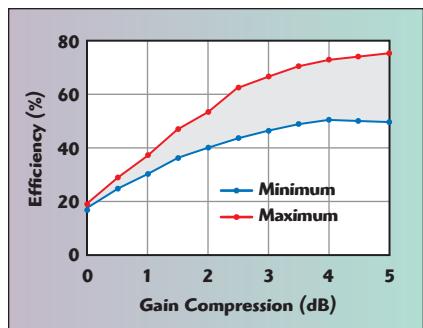


▲ Fig. 1 Drain efficiency load-pull contours at f_0 on a harmonically-terminated transistor.

tion of the frequency band. When this happens, the ability to independently match the harmonic impedances at each fundamental frequency is drastically reduced or even eliminated.^{3,4} Consider a design example of a wideband amplifier operating between 3.1 and 10.6 GHz. For a fundamental frequency of 3.1 GHz, a theoretical Class-F amplifier would require a short at the second harmonic of 6.2 GHz and an open at the third harmonic of 9.3

GHz. However, 6.2 and 9.3 GHz are required fundamental frequencies of the wideband amplifier, and the short/open terminations could yield low performance at those frequencies.

Commercial packaged transistors are available from multiple vendors with varying degrees of integrated matching, varying from completely unmatched to partially matched and fully matched. Completely unmatched transistors can be tuned for maximum performance at a given frequency, power and bias by determining ideal loading conditions at the fundamental and harmonic frequencies, if harmonic power exists. Partially-matched and fully-matched transistors offer less flexibility as the internal matching structure within the packaged component limits the ability to significantly alter the match presented to the internal transistor. Partially- and fully-matched packaged transistors are commonly offered with optimum harmonic terminations already implemented for specific applications such as the design of commercial wireless base stations and handsets; therefore the advantages of



▲ Fig. 2 Drain efficiency as a function of compression, with harmonic termination.

presenting additional harmonic terminations outside of the DUT package are practically eliminated.

Figure 1 shows the results of harmonic load-pull performed on a harmonically-terminated Freescale LDMOS Class-F transistor with integrated harmonic matching operating at 960 MHz with $V_{dd} = 28$ V, $I_{dq} = 300$ mA and 35 W output power at 1 dB gain compression. No improvement or trend can be seen when varying the harmonic terminations across the Smith Chart.

Harmonic impedance matching becomes critical in designing highly efficient amplifiers operating under compression or saturation for specific bands of operation. Under these conditions transistors will exhibit deep nonlinearities and put out power at one or more harmonic frequencies. It is under these nonlinear operating conditions that advanced classes of operation (E, F, G, J and their inverses) are achievable by terminating the harmonic impedances to ideal values.

In general, when power exists at the harmonic frequencies due to compression, power-added efficiency (PAE) can be improved by reflecting the energy back towards the device under test. This generally occurs as the magnitude of reflection $|\Gamma_L|$ approaches 1 at a specific phase angle (dependent on the reference plane of the measurement), with lower PAEs as the magnitude of reflection decreases.⁵⁻⁸

Figure 2 shows the change in drain efficiency at 2.5 GHz for a 10 W GaN transistor with varying levels of gain compression, from 0 dB (nearly small signal linear operating condition) to 5 dB (highly compressed approaching saturation) for the harmonic terminations which result in minimum (blue) and maximum (red) efficiency. The level at which termi-

High Power Couplers

Features

- High Performance Design
- 600 W Average/10kW Peak Power Handling
- 0.75 to 12 GHz in Bands
- Low Loss Air Dielectric
- Operation to +85 C Without Degradation
- Multi-Octave Bandwidths

Applications

- High Power Measurement Systems
- Power Amplifier Monitoring
- EW Systems Testing

NEW!



Visit us at
IMS Booth
#1826

Visit our new website with interactive catalog and online RFQ!

www.WeinschelAssociates.com

19212 Orbit Drive
Gaithersburg, MD 20879
Voice: 301.963.4630
Fax: 301.963.8640
RF@WeinschelAssociates.com

WA WEINSCHEL
ASSOCIATES
BROADBAND RF
& MICROWAVE
SOLUTIONS

Technical Feature

nating the harmonic impedances has an effect increases with the amount of harmonic power output by the device under test.

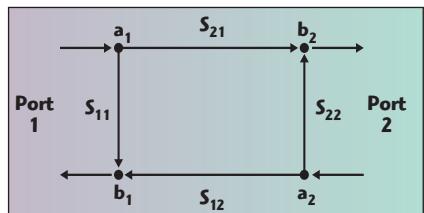
MODERN HARMONIC LOAD-PULL TECHNIQUES

Harmonic load-pull techniques have existed for decades, since the invention of the earliest closed-loop and open-loop active load-pull techniques between 1979 and 1990.^{9,10} While state-of-the-art at their time, these systems had inherent stability and processing issues that limited their commercial applications. Throughout the '90s and '00s, passive mechanical tuners were configured for harmonic load-pull using multiplexers to combine tuners in parallel¹¹ and advanced mathematics to internally/externally cascade tuners in-line.¹² As with all passive systems, achievable magnitude of reflection at the DUT was limited by the tuning range of the tuning network and the losses of the components used to connect to the device.

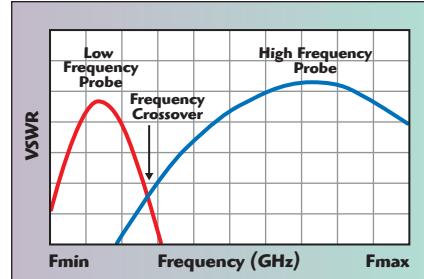
Modern open-loop active load-pull systems were introduced in the

'00s and '10s in order to overcome the weaknesses of the earliest active systems as well as the limitations of purely passive systems.¹³ Each system, passive harmonic, open-loop active harmonic, hybrid active harmonic and mixed-signal active harmonic has its own strengths and weaknesses, each of which should be clearly understood. The key topics that differentiate the load-pull techniques are measurement method, available magnitude of reflection at the DUT reference, methodology of harmonic control, tuning accuracy and speed, and system cost.

Load-pull systems can be based on scalar or vector measurements of power waves. Scalar-based systems use power meters or spectrum analyzers to measure scalar values at a specified marker or the entire signal, which are then de-embedded to the device reference plane. Vector-based systems use a vector analyzer, calibrated at the DUT reference plane, to directly measure vector a- and b-waves (more accurately, without de-embedding) from which performance parameters



▲ Fig. 3 Two-port network defined by S-parameters and a- and b-waves.



▲ Fig. 4 Typical wideband probe response of a passive impedance tuner.

are calculated. A two-port network defined by S-parameters and a- and b-waves is shown in **Figure 3**. While passive load-pull methodologies can be either scalar- or vector-based, active load-pull methodologies require a vector-receiver to measure the a- and b-waves and determine the terminations presented to the DUT.

Passive impedance tuners are wideband in nature, which means the tuning element (probe/slug) inside the tuner creates a continuum of reflection vectors over a large bandwidth, possibly affecting multiple harmonics, as shown in **Figure 4**. With a single element, it is possible to control the impedance at one frequency of interest; however, the tuner will present uncontrolled wideband impedances at higher frequencies, including the harmonics. With n tuning elements internally cascaded in a single box or externally cascaded using multiple single-element tuners, it is possible to control the impedance presented at n frequencies. Therefore, a two-element tuner configuration can present controlled impedances at two frequencies, and so on.¹²

In a traditional scalar harmonic load-pull system comprised of cascaded tuning elements, the maximum magnitude of reflection achievable at any frequency is the summation of the reflections of the elements, minus the losses of the interconnection between the tuners and DUT such that $R_L(DUT) = R_L(\text{tuner}) + R_L(\text{interconnection})$. In a typical setup using

When it comes to THIN FILM We get Small

IMS Booth # 3312

Thin film circuits with < 25 micron traces and spaces

METRIGRAPHICS®
Micron-Scale Circuit Development & Manufacturing
FLEX • THIN FILM • ELECTROFORMING
www.metrigraphicsllc.com

Fill your brain with our online thin film guidelines.

The Industry's Largest Selection of Off-the-Shelf RF/Microwave Amplifiers

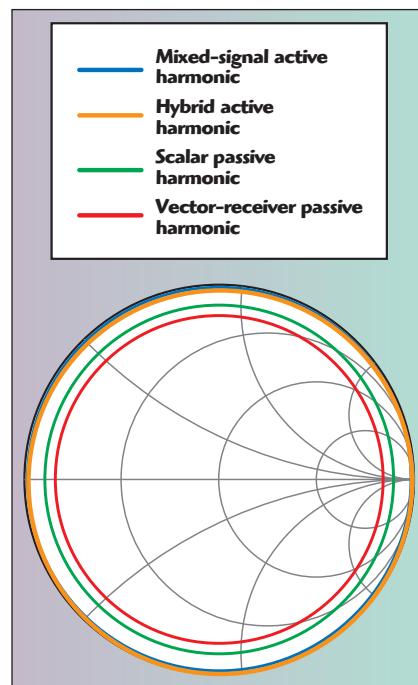


**All in-stock
and available
to ship the
same day**

- Frequencies from DC to 40 GHz
- Gain ranging from 10 to 60 dB
- P1dB from 2 mW to 100 Watts
- Noise figures as low as 0.8 dB
- Gain variation down to ± 0.3 dB

pasterнак.com
866.727.8376

Technical Feature



▲ Fig. 5 Maximum reflection of various load-pull methods at the second harmonic ($2f_0$).

a 50 ohm test fixture at 2.5 GHz, a realistic $|\Gamma_L|=0.93$ is achievable at the fundamental frequency while $|\Gamma_L|=0.9$ is achievable at the second harmonic.

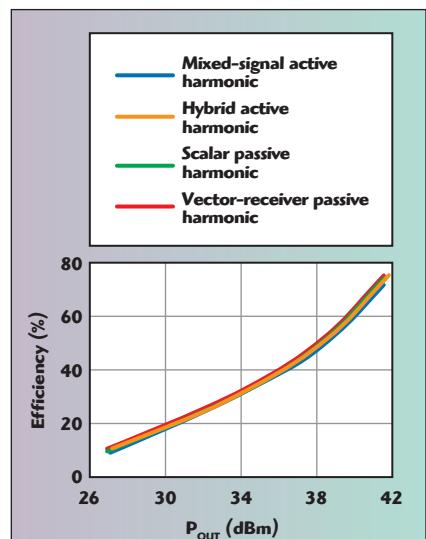
In a modern vector-receiver passive harmonic load-pull system, low-loss couplers are added between the tuner and DUT, thereby increasing the losses (R_L interconnection term) and decreasing achievable magnitude of reflection to 0.91 at the fundamental frequency and 0.85 at the second harmonic.

Open-loop active load-pull replaces the passive mechanical tuner with an active tuning chain consisting of a magnitude and phase controllable signal source. Instead of using a passive tuner to reflect energy back to the DUT, the signal source creates a new signal which is injected into the output of the DUT, satisfying $\Gamma_L = a^2/b^2$.

In a hybrid active load-pull system using a passive tuner for fundamental impedance control and active tuning chains at the harmonics, realistic $|\Gamma_L|=0.86$ to 0.91 is achievable at the fundamental frequency while $|\Gamma_L|>1$ (limited only by injection power) is achievable at the second harmonic.

A purely active load-pull system would have no limits on achievable magnitude of reflection at any frequency so long as the active tuning chain can produce sufficient power to satisfy $\Gamma_L = a^2/b^2$.

It is important to note that as



▲ Fig. 6 Drain efficiency vs. output power using harmonic load-pull methods with identical loads.

frequency increases, the insertion losses of the interconnections generally increase and the tuning range of passive tuning systems decrease.

Figure 5 compares the typical magnitudes of reflection achievable at the harmonic frequencies for the various harmonic load-pull methodologies at $f_0=2.5$ GHz.

Active load-pull systems often make use of commercial VNAs which act as both vector-receiver and active tuning chains, depending on the quantity of available signal sources within the instrument. The VNA measures the a- and b-waves presented by the DUT, software calculates the resulting injection signal required to achieve $\Gamma_L = a^2/b^2$ at the DUT reference plane and commands the source to create that signal, and the VNA measures the resulting wave at the DUT reference plane for accuracy. An iterative software algorithm adjusts the active injection signal magnitude and phase until the desired reflection is achieved within a predefined convergence limit. The process is repeated for each desired impedance and power.

A unique form of active load-pull is referred to as mixed-signal active load-pull (MSALP) and utilizes wideband arbitrary waveform generation with up-conversion, and wideband data analysis with down-conversion, instead of classic active load-pull single frequency generation and analysis methodology. Because of its wideband nature, MSALP uses a time-slotted approach to present many impedanc-



World's Smallest YIG Oscillator

INTRODUCING THE
DOM-624
6~24 GHz YIG Oscillator



Visit us at
IMS 2015
Booth # 613

VIDA's 6 ~ 24 GHz patented Differential Oscillator Module features excellent phase noise and spurious response. The small size fits all magnetic packages. Use your existing magnetic tuning package or design your own with VIDA application notes. Standard Oscillator package is a sealed 1mm thick by 4mm diameter pill package.

- 6 to 24GHz operating range
- 1mm x 4mm package
- Bias =± 5V<200mA
- Θn <-145dBc@1 MHz Offset
- Harmonics <-20 dBc
- Spurious <-70 dBc

VIDA Products Inc

3551 Westwind Blvd.,
Santa Rosa, CA 95403
Phone: 707-541-7000
info@vidaproducts.com

www.vidaproducts.com

Technical Feature

es, near-simultaneously, to the DUT, resulting in a much faster tuning and measurement time.¹³

Regardless of the method used, the same impedances presented at the same frequencies should yield the same measurement results. **Figure 6** demonstrates a comparison of measurement results between scalar passive, vector-receiver passive, hybrid active and mixed-signal active harmonic load-pull systems for the same f_0 , $2f_0$ and $3f_0$ load impedances.

An investigation into the performance of drain efficiency shows that maximum efficiency improvement normally occurs at maximum magnitude of reflection of the harmonic load impedance. In order to yield the highest possible efficiency, a $|\Gamma_L| = 1$ should be presented at the harmonic frequencies.

Figure 7 shows the change in drain efficiency at 2.5 GHz for a 10 W GaN transistor with varying magnitudes of reflection at a fixed phase of reflection with fixed gain compression and bias. The impact of harmonic tuning increases with its magnitude: presenting a second harmonic termination at $|\Gamma_L| = 0.85$ (typical of a vector-receiver passive load-pull system) results in a drain efficiency of ~75 percent whereas $|\Gamma_L| = 0.99$ (possible with active load-pull) results in a drain efficiency of ~80 percent.

While each technique has different reflection limitations, each technique also has its own characterization, calibration and measurement advantages and disadvantages. In general, using a passive mechanical tuner requires a tuner characterization process, in

which S-parameters of a tuning element are mapped into a database for various motor positions. Using n elements results in a time increase by a factor of n. Modern tuners equipped with an LXI™ interface used in conjunction with a modern VNA can be characterized in ~11 minutes per element for 700 to 1000 states, while two single-element tuners or a single two-element tuner requires ~22 minutes for characterization.

System calibration varies between scalar and vector systems. Scalar calibration involves a "power calibration" in which the losses through the input chain are calculated based on deembedding power meter readings, a process which takes several minutes. Vector-based load-pull calibration is two-fold or three-fold, and involves a vector calibration, an absolute power calibration, and an optional phase calibration for nonlinear VNA measurements. The two-step process takes ~5 minutes while the three-step process takes ~7 minutes.

Tuning time varies between passive and active load-pull methodologies, and the measurement instrument used. Scalar-based, passive load-pull relies on the movement of multiple tuning elements and a slow acquisition with averaging from a power meter. Vector-receiver passive load-pull relies on the same movement of multiple tuning elements; however, the measurement is faster as it uses a vector-receiver in place of the power meter. Traditional open-loop active load-pull measurements can be

TABLE 1
COMPARISON OF LOAD-PULL MEASUREMENT TIME (MINUTES UNLESS STATED)

Setup	Tuner Cal	System Cal	Step 1	Step 2	Step 3
			<i>f_0 Load-Pull, Fixed $2f_0$ at 50 ohms (35 Loads, 16 Powers)</i>	<i>$2f_0$ Load-Pull, fixed f_0 at Optimized Value (20 Loads, 16 Powers)</i>	<i>f_0 Load-Pull, fixed $2f_0$ at Optimized Value (35 Loads, 16 Powers)</i>
Scalar Harmonic (2 tuning elements)	22	3	11.1	6.4	11.1
Vector-Receiver Harmonic (2 tuning elements)	22	5	5.3	3.1	5.3
Hybrid-Active Harmonic (1 tuning element)	11	5	4.2	7.3	7.5
Mixed-Signal Active (0 tuning elements)	No Tuner	5	15 seconds	35 seconds	50 seconds

Hi-Q Capacitors

Reliable • Fast Turnarounds
• Competitive Pricing

RF/Microwave & HF/UHF

Low ESR/ESL



Case Size: 0505, 1111 & EIA sizes

NEW Broadband

01005BB: 16kHz - 67GHz

- Insertion Loss: < 1db
- Value: 100nF
- 4 WVDC

0201BB: 16kHz - 65GHz

- Insertion Loss: < 1db
- Value: 100nF
- 16 WVDC

0402BB: 16kHz - 35GHz

- Insertion Loss: < 1db
- Value: 100nF
- 25 WVDC



Power Assemblies



Series / Parallel Combinations

Aerospace • Aviation • Military Commercial • Medical Telecommunications

- Unmatched customer service
- Online store for immediate availability
- Design kits in stock
- Inventory programs

Call us today

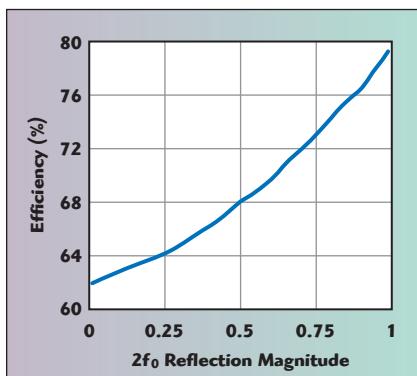
631-425-0938

sales@passiveplus.com

www.PassivePlus.com

See us at IMS Booth 1831

Technical Feature



▲ Fig. 7 Drain efficiency as a function of $2f_0$ reflection magnitude at a constant phase.

faster or slower than a passive tuner, depending on the accuracy of the desired tuning and the resulting number of iterations required to converge on the desired impedance (~ 30 dB tuning accuracy results in faster measurements while ~ 50 dB tuning accuracy may result in slower measurements). Because mixed-signal active load-pull uses a time-slotted approach with wideband signal generation and analysis, it can present many impedances, near-simultaneously, to the DUT, resulting in a much faster tuning and measurement time.

Table 1 shows the time associated with tuner characterization (if applicable), system calibration, fundamental impedance load-pull with power sweeps for a fixed second harmonic impedance, and second harmonic impedance load-pull with power sweeps for a fixed fundamental impedance. For this comparison, commercial load-pull systems from Maury Microwave were used. The tuners were MT982ML01 LXI™-certified harmonic tuners; the vector-receiver used was N5242A PNA-X; the active tuning chain consisted of N5242A PNA-X second internal source with an external amplifier; the software used was the MT930-series IVCAD measurement and modeling device characterization suite; the mixed-signal active load-pull system was Maury Microwave's MT2000.¹⁴

CONCLUSION

It is essential to plan ahead and understand the design objectives and limitations before launching into an extensive and time-consuming design process. With regards to amplifier design, it is important to determine

whether the objective is a small-signal linear amplifier, a narrowband highly efficient amplifier or a wideband amplifier, and whether it will use unmatched, partially-matched or fully-matched packaged transistors. Only when the objectives have been defined will it be possible to determine whether the design can or cannot make use of harmonic impedance terminations. If it has been determined that harmonic tuning is required, additional thought must be given to the ideal methodology and technique, evaluating the desired magnitudes of reflection, the equipment that will be dedicated to the load-pull station, and the time allocated to the test and measurement process. ■

References

1. D. M. Pozar, *Microwave Engineering*, Wiley & Sons, Inc., New York, 2005.
2. High Frequency Techniques: An Introduction to RF and Microwave Engineering, Joseph F. White, John Wiley & Sons Inc., 2004.
3. P. Colantonio, F. Giannini, R. Giofre and L. Piazzon, "High-Efficiency Ultra-Wideband Power Amplifier in GaN technology," *Electron Letters*, Vol. 44, No. 2, pp. 130–131, 2008.
4. C. M. Andersson, J. Moon, C. Fager, B. Kim, and N. Rorsman, "Decade Bandwidth High Efficiency GaN HEMT Power Amplifier Designed with Resistive Harmonic Loading," *IEEE MTT-S International Microwave Symposium Digest*, June 2012.
5. T. Maier, V. Carrubba, R. Quay, F. van Raay, and O. Ambacher, "Active Harmonic Source-/Load-Pull Measurements of AlGaN/GaN HEMTs at X-Band Frequencies," 83rd ARFTG Microwave Measurement Conference, 2014, pp. 1–4.
6. Travis A. Barbieri, Basim Noori "Improvements in High Power LDMOS Amplifier Efficiency Realized Through the Application of Mixed-Signal Active Loadpull," 82nd ARFTG Microwave Measurement Conference, 2013.
7. T. Thirivikraman and J. Hoffman, "Design of an Ultra-High Efficiency GaN High-Power Amplifier for SAR Remote Sensing," IEEE Aerospace Conference, pp. 1–6, 2013.
8. J. Moon, J. Lee, R. Pengelly, R. Baker and B. Kim, "Highly Efficient Saturated Power Amplifier," *IEEE Microwave Magazine*, Vol. 13, No. 1, pp. 125–131, February 2012.
9. B. Stancil and D. P. Poulin, "Harmonic Load-Pull," *MTT-S International Microwave Symposium Digest*, pp. 185–187, 1979.
10. R. Larose, F. Ghannouchi and R. Bosisio, "A New Multi-harmonic Load-pull Method for Nonlinear Device Characterization and Modeling," *IEEE International Microwave Symposium*, pp. 443–446, 1990.
11. Maury Microwave, Application Note 5C-053, "A Comparison of Harmonic Tuning Methods for Load Pull Systems."
12. Maury Microwave, Application Note 5C-081, "Cascading Tuners for High VSWR and Harmonic Load Pull."
13. Squillante, M.; Marchetti, M.; Spirito, M.; and de Vreede, L.C.N., "A Mixed-signal Approach for High-Speed Fully Controlled Multidimensional Load-Pull Parameters Sweep," 73rd ARFTG Microwave Measurement Conference, 2009.
14. Maury Microwave website www.maurymw.com/MW_RF/.

LNA-Basics



WHITE PAPER

Designing Ultra Low Noise Amplifiers for Infrastructure Receiver Applications

By Fikret Altunkilic, Alan Miller, Vivian Tzanakos and Michele Lewis

Introduction

Amplifiers used in wireless infrastructure receiver applications have the key requirements of low noise, high linearity, and unconditional stability. In order to meet these needs, Skyworks has developed a new family of low noise amplifiers (LNAs) implemented in 0.5 μm enhancement mode (E-Mode) pHEMT. The SKY67101-396LF, covering 0.7–1.0 GHz, and the SKY67100-396LF, covering 1.7–2.0 GHz, are appropriate for application in GSM, WCDMA, TDSCDMA, and LTE infrastructure receiver chains. For low cost and efficient use of PCB space, the LNAs are housed in a 2 x 2 mm QFN package and share a common layout among the frequency bands.

Specifications

Noise figure (NF) performance is a key parameter in receiver systems as it describes the capability for reception of low-level signals. The lower the noise figure, the better the receiver sensitivity. Linearity performance, represented as third order intercept (IP3), is also important as it indicates the ability to amplify signals with low intermodulation distortion from signals close in frequency. Unconditional stability is the ability for the amplifier to provide oscillation free performance under any input or output load conditions. Other specifications including current consumption, return loss, and human body model electro static discharge (HBM ESD) performance are also important and need to be taken into consideration in an LNA design. To obtain optimized performance with minimum trade-offs, special design techniques have been incorporated. Target specifications for low- and high-band LNAs are shown in Table 1.

Parameter	Low-Band 0.9 GHz SKY67101	High-Band 1.9 GHz SKY67100
Noise figure (NF, dB)	0.55	0.65
Gain (G_A , dB)	18	17.5
Input Return Loss ($ S_{11} $, dB)	>20	>20
Output Return Loss ($ S_{22} $, dB)	>20	>12
Isolation ($ S_{12} $, dB)	>30	>30
Output 1dB Compression (O_{P1dB} , dBm)	>18	>18
3 rd Order Intercept Point (OIP3, dBm)	>34	>34
Supply Voltage (V, V)	4	4
Supply Current (I, mA)	55	55
HBM Electrostatic Discharge Rating (ESD, V)	>250	>250

Table 1. LNA Specifications at $V_{DD} = 5$ V, Temp. = 25 °C

Technology and Topology Selection

A 0.5 μm enhancement mode pHEMT technology has been selected for excellent RF performance, superior low noise and high linearity. It also simplifies MMIC design by only requiring a positive voltage at the gate of FETs. This enables direct source connection to ground without requiring extra components for a self-bias structure. Very accurate models are available for circuit simulations.

A cascode LNA topology is used in the design because of its wide bandwidth, high gain and high-reverse isolation.

Design Steps

The design of the SKY67101-396LF 900 MHz LNA is detailed in this paper. The SKY67100-396LF 1900 MHz design was designed by a frequency scaling technique using the same methodology. Measured and simulated results of low and high bands are shown in the “Comparison of Measured and Simulated Results” section.

Biassing Circuit

An active biasing circuit is shown in Figure 1 that stabilizes the current consumption of the LNA around approximately 55 mA over temperature, process, and supply voltage. R1 is used to set the voltage at pin 4 to set the total bias current. Any variation in supply bias is stabilized by the active bias circuitry. Stabilized gate voltage is fed to pin 2 through the L1 inductor. These components are also used for input impedance and noise figure source impedance matching.

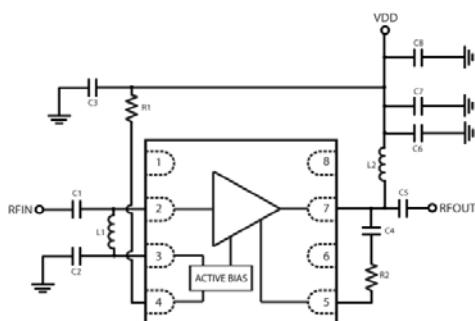


Figure 1. Bias, Matching Components and Package

Measured vs. simulated supply current variation over temperature (-40 °C to +80 °C) is about 3 mA as shown in Figure 2.

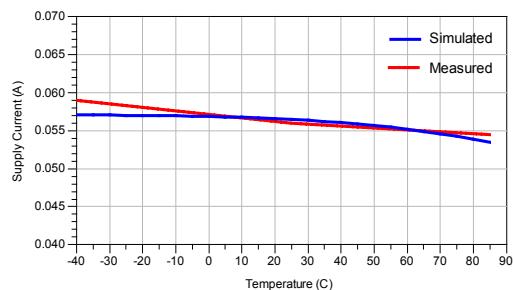


Figure 2. Compensated Results of Measured vs. Simulated Current over Temperature

Noise Figure and Input Match

Noise figure and input return loss are the main factors in an LNA design. The first stage of the cascode design was designed for optimum noise figure, output impedance match, and P_{1dB} at the target drain source current (I_{ds}). The buffer stage was designed for optimal IP3 performance, output match, and P_{1dB} while not degrading other performance specifications. The topology is made stable over almost all impedances by using source degeneration (it is made fully, unconditionally stable after adding an inter-stage network, output network, transmission line loss, and SMT component parasitic impedance. See “Linearity” and “Stability” sections). Figure 3 shows gain and NF_{min} (minimum noise figure) trade-off over frequency of the topology.

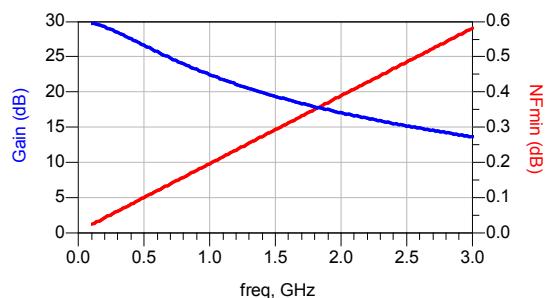


Figure 3. Gain vs. NF_{min} Trade-Off over Frequency

Figure 4 shows constant NF circles, source stability circle and available gain circles of the cascode topology in the source stable region of the Smith chart at 900 MHz.

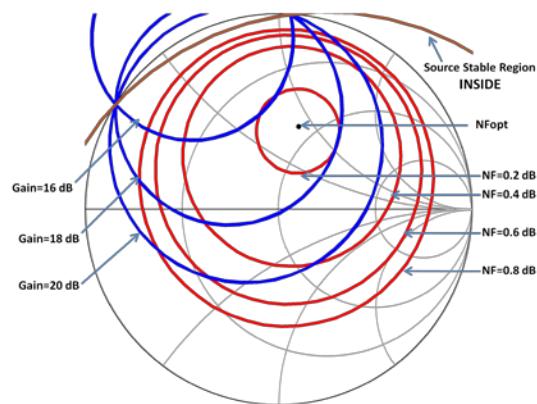


Figure 4. Available Gain and Noise Circles

By considering the parasitic effects of SMT components and transmission line losses, a source impedance point of $Z_s = 64 + j44$ ohms is selected in the 0.4 dB noise circle and 18 dB gain circle for the trades-off between noise, gain, and input return loss matching. The input matching network is realized by C1, C2, and L1. C1 and L1 are selected as high-Q components to achieve best NF. C1 is also used for DC blocking. Simulated gain, input return loss, and noise figure can be seen in the “Comparison of Simulated and Measured Results” section.

Linearity (OIP3) and $P_{1\text{dB}}$

Input and output terminations in band and out of band directly affect the linearity performance of the amplifier. The input and output loads of the amplifier can be swept directly through source and load pull technique. In this section, after the source is matched, the load-pull measurement is performed.

After the source is matched to an impedance of $Z_s = 64 + j44$ ohms for desired NF, input return loss, and gain at the bias current, $P_{1\text{dB}}$ and OIP3 are determined by the output matching and feedback network. Simulation models are used to estimate OIP3 at 0.9 GHz, for two tones separated by 5 MHz, with input power $P_{\text{IN}} = -20$ dBm for each tone. Figure 5 shows load-pull impedances on the Smith chart and the circle shows the best OIP3 region at 0.9 GHz. Figure 6 shows OIP3 and delivered power contours at 0.9 GHz.

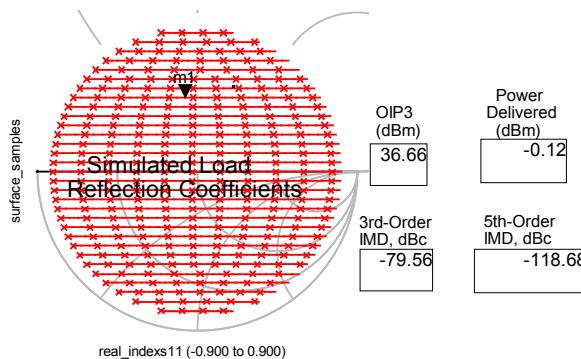


Figure 5. Simulated Load Pull for OIP3, Delivered Power and IMD3

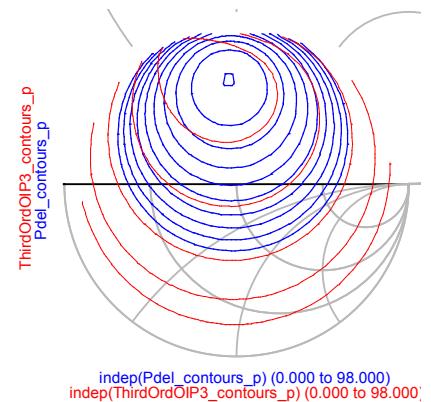


Figure 6. OIP3 and Delivered Power Contours

Final load pull simulation and matching should be done after connecting input and output matching circuits as shown in Figure 1. After matching the source and load, simulated OIP3 and $P_{1\text{dB}}$ results are shown in Figure 7 and Figure 8, respectively

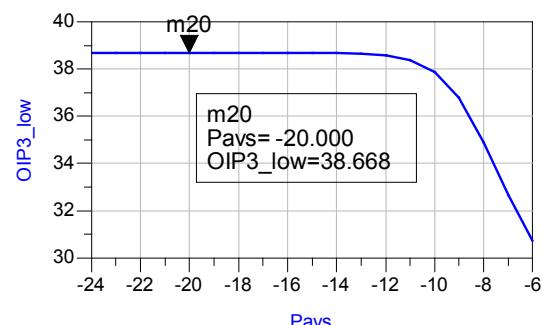


Figure 7. Simulated OIP3 After Source and Load Are Matched

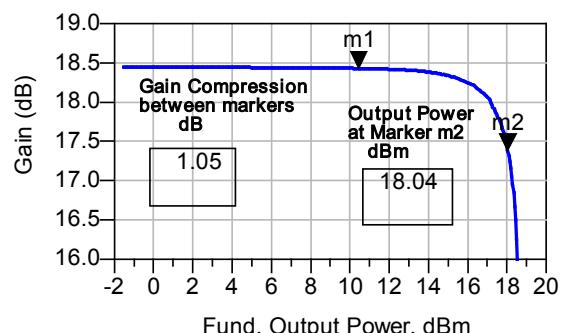


Figure 8. Simulated $P_{1\text{dB}}$ After Source and Load Are Matched

Stability

Stability is one of the most important requirements to consider for an LNA. Typical specifications dictate unconditionally stable operation up to 18 GHz. Each stage must be designed for unconditionally stable operation, including all the external components and biasing over all conditions. A very low-noise device with high gain usually becomes highly unstable in most of this frequency range. In order to stabilize the device and meet these requirements, multiple stability design techniques have been employed.

In order to solve stability problems at low to operating frequencies, usually some value of a source inductor is used. A source degeneration inductor used for input and noise-figure match can be used for stability purpose as well. One technique that is commonly used is to have a series-parallel LR network. This network behaves like low impedance at low frequencies and high impedance at higher frequencies.

Another technique that is commonly used is a shunt-series CR network connected from drain to ground. This network behaves like a shunt resistor at high frequencies and high impedance at low frequencies. A shunt resistor connected to ground helps to stabilize the device.

Another method that is used to improve stability is the shunt feedback between the output and the input of the device. However, this method degrades the noise figure. Therefore, it is not used in the first stage design but it is used in the second stage (buffer) design. This feedback also helps the IP3, return loss (RL) and gain adjustment.

Stabilizing circuits are integrated into the cascode LNAs. Final simulated stability and measured stability results of SKY67100 and SKY67101 are presented in Figures 16 and 17.

ESD Consideration

Electrostatic discharge (ESD) is the transfer of electrostatic charge between bodies or surfaces at different electrostatic potential and can be particularly destructive for semiconductors. ESD must be considered in the early phase of product development. Power clamps, diode, and stacked diodes ESD protection circuits are used in the design to achieve **Class 1A (>250 V)** HBM rating between all pin combinations. ESD protection circuits were used in different parts of the design, with careful attention paid to ensure that small-signal, large-signal, and noise figure performance were not degraded.

Layout

The SKY67100/SKY67101 application test board layout was designed to achieve the lowest possible noise figure and best stability performance. The test board was realized using 10-mils-thick Rogers 4350B substrate on 50-mils-thick FR4 supporting substrate. The Rogers 4350B material was selected for the RF circuit for its low dielectric constant (ϵ_r) and low ϵ_r variation over temperature for the best possible noise performance. The lower cost FR4 material is used to support remaining layers and provide added mechanical rigidity and thickness. Microstrip line widths and spacing were designed to accept commonly-used 0402 sized surface mount components while maintaining a uniform 50 ohm system. Copper thickness is 1.4 mils to reduce circuit losses and their additive effects on noise figure. The evaluation board is biased with a single 4.0 V power supply.

Component Selection

The test board schematic is shown in Figure 9. The input components C1, C2 and L1 determine the input match and noise figure of the device. For optimum noise figure to be achieved, it is recommended that high-Q components are used.

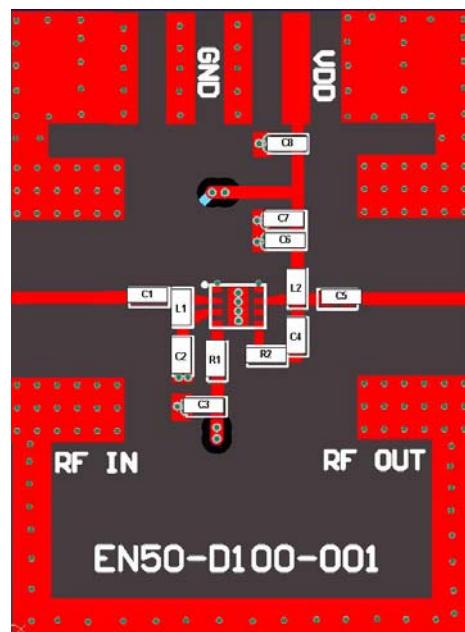


Figure 9. SKY67101-396LF Evaluation Board Layout

Components R2 and C4 comprise the feedback circuitry of the device if gain adjustment is needed.

The output match is realized through components L2 and C5. L2 is also used for decoupling at the bias line with components C6, C7 and C8. Linearity can be optimized by tuning the output match.

Comparison of Simulated and Measured Results

Figures 10 and 11 illustrate the measured and simulated gain profiles of SKY67100-396LF (1.9 GHz) and SKY67101-396LF (0.9 GHz) over a broadband frequency range. The gain for the SKY67101 is 18.2 dB at 0.9 GHz and the gain for SKY67100 is 17.67 dB at 1.9 GHz.

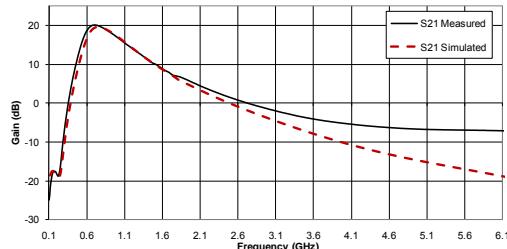


Figure 10. SKY67101 Simulated and Measured Gain

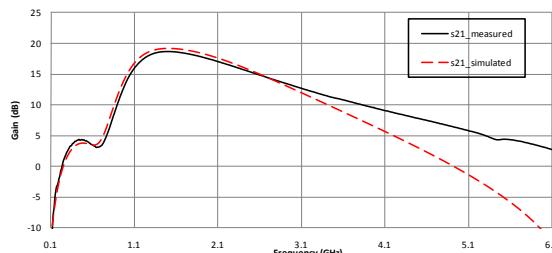


Figure 11. SKY67100 Simulated and Measured Gain

In Figure 12, the input and output return losses of the SKY67101 device are depicted. Both input and output return losses measure greater than 20 dB at 0.9 GHz.

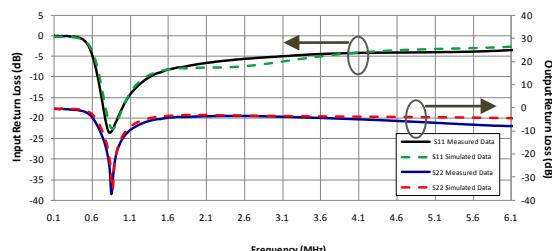


Figure 12. SKY67101 Simulated and Measured Return Losses

Input and output return losses of the SKY67100 device are depicted in Figure 13.

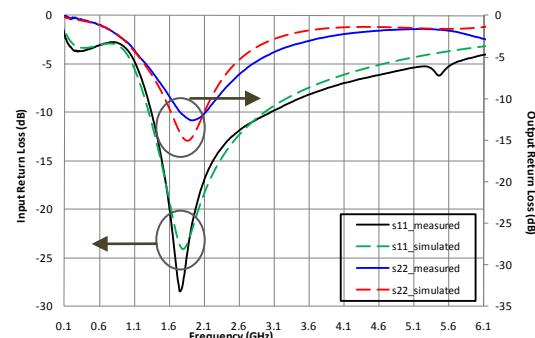


Figure 13. SKY67100 Simulated and Measured Input and Output Return Loss

In-band OIP3 and $P_{1\text{dB}}$ performance are shown in Figures 14 and 15. OIP3 measurements were taken at 900 ± 5 MHz for the SKY67101 and at 1950 ± 5 MHz for the SKY67100.

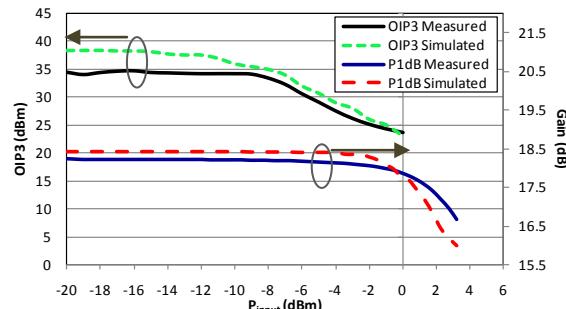


Figure 14. SKY67101 Simulated and Measured Large Signal Data

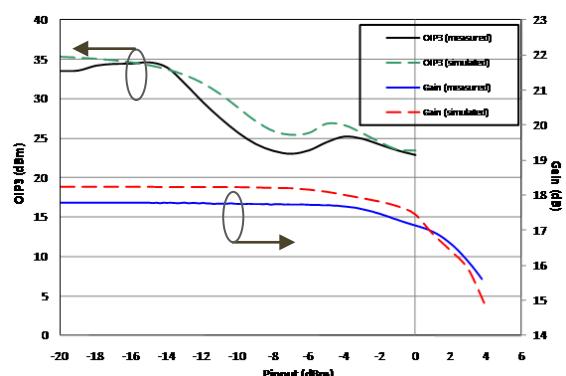


Figure 15. SKY67100 Simulated and Measured Large Signal Data

Figures 16 and 17 exhibit the measured versus simulated NF performance for the SKY67101 and SKY67100 respectively. A correction factor of 0.05 dB was applied to account for losses due to the input connector and evaluation board transmission line to the first matching component for the SKY67101 and 0.1 dB was applied for the SKY67100.

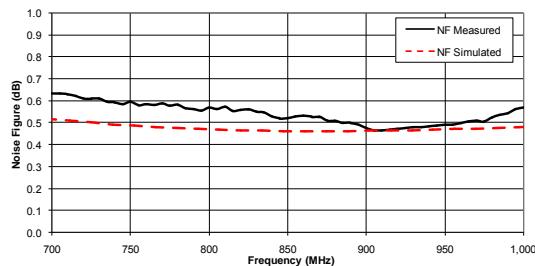


Figure 16. SKY67101 Simulated and Measured Noise Figure (NF)

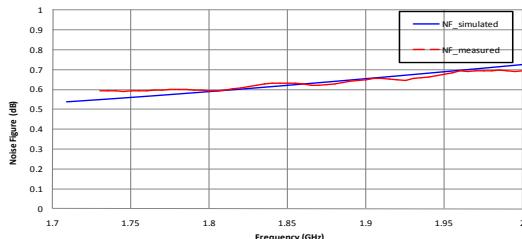


Figure 17. SKY67100 Simulated and Measured Noise Figure (NF)

Figures 18 and 19 illustrate the measured and simulated stability performance of the SKY67101 and SKY67100 respectively. Both devices exhibit unconditional stability over the broadband range as $B > 0$ and Rollet's stability factor $K > 1$.

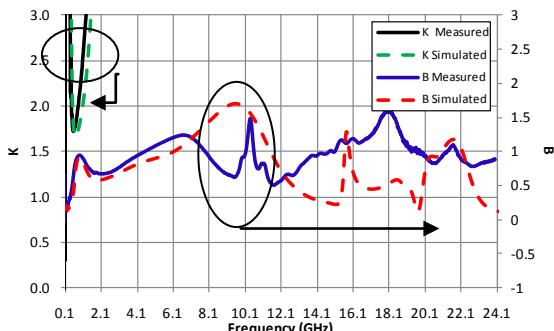


Figure 18. SKY67101 Simulated and Measured Stability

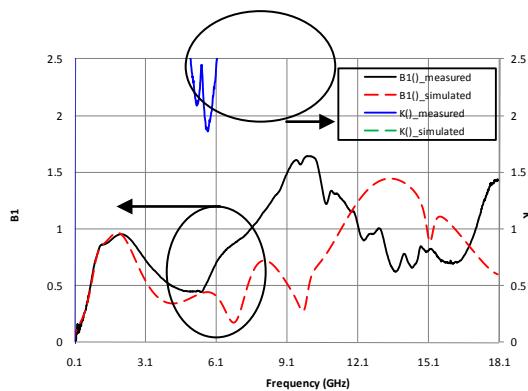


Figure 19. SKY67100 Simulated and Measured Stability

Conclusion

This paper has presented the design of a pair of low noise, high linearity amplifiers, SKY67100 and SKY67101. These LNAs are implemented using enhancement mode pHEMT devices in a cascode topology. The LNAs are appropriate for receiver applications in various wireless infrastructure products. They both utilize a common pinout and layout in a small, low cost 2 x 2 mm QFN package.

Copyright © 2010 Skyworks Solutions, Inc. All Rights Reserved.

Information in this document is provided in connection with Skyworks Solutions, Inc. ("Skyworks") products or services. These materials, including the information contained herein, are provided by Skyworks as a service to its customers and may be used for informational purposes only by the customer. Skyworks assumes no responsibility for errors or omissions in these materials or the information contained herein. Skyworks may change its documentation, products, services, specifications or product descriptions at any time, without notice. Skyworks makes no commitment to update the materials or information and shall have no responsibility whatsoever for conflicts, incompatibilities, or other difficulties arising from any future changes.

No license, whether express, implied, by estoppel or otherwise, is granted to any intellectual property rights by this document. Skyworks assumes no liability for any materials, products or information provided hereunder, including the sale, distribution, reproduction or use of Skyworks products, information or materials, except as may be provided in Skyworks Terms and Conditions of Sale.

THE MATERIALS, PRODUCTS AND INFORMATION ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, WHETHER EXPRESS, IMPLIED, STATUTORY, OR OTHERWISE, INCLUDING FITNESS FOR A PARTICULAR PURPOSE OR USE, MERCHANTABILITY, PERFORMANCE, QUALITY OR NON-INFRINGEMENT OF ANY INTELLECTUAL PROPERTY RIGHT; ALL SUCH WARRANTIES ARE HEREBY EXPRESSLY DISCLAIMED. SKYWORKS DOES NOT WARRANT THE ACCURACY OR COMPLETENESS OF THE INFORMATION, TEXT, GRAPHICS OR OTHER ITEMS CONTAINED WITHIN THESE MATERIALS. SKYWORKS SHALL NOT BE LIABLE FOR ANY DAMAGES, INCLUDING BUT NOT LIMITED TO ANY SPECIAL, INDIRECT, INCIDENTAL, STATUTORY, OR CONSEQUENTIAL DAMAGES, INCLUDING WITHOUT LIMITATION, LOST REVENUES OR LOST PROFITS THAT MAY RESULT FROM THE USE OF THE MATERIALS OR INFORMATION, WHETHER OR NOT THE RECIPIENT OF MATERIALS HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Skyworks products are not intended for use in medical, lifesaving or life-sustaining applications, or other equipment in which the failure of the Skyworks products could lead to personal injury, death, physical or environmental damage. Skyworks customers using or selling Skyworks products for use in such applications do so at their own risk and agree to fully indemnify Skyworks for any damages resulting from such improper use or sale.

Customers are responsible for their products and applications using Skyworks products, which may deviate from published specifications as a result of design defects, errors, or operation of products outside of published parameters or design specifications. Customers should include design and operating safeguards to minimize these and other risks. Skyworks assumes no liability for applications assistance, customer product design, or damage to any equipment resulting from the use of Skyworks products outside of stated published specifications or parameters.

Skyworks, the Skyworks symbol, and "Breakthrough Simplicity" are trademarks or registered trademarks of Skyworks Solutions, Inc., in the United States and other countries. Third-party brands and names are for identification purposes only, and are the property of their respective owners. Additional information, including relevant terms and conditions, posted at www.skyworksinc.com, are incorporated by reference.

Tutorial-1

Low Noise Amplifier (LNA) Design

Complied by Rashad M. Ramzan

Objective:

Low noise amplifiers are one of the basic building blocks of any communication system. The purpose of the LNA is to amplify the received signal to acceptable levels with minimum self-generated additional noise. Gain, NF, non-linearity and impedance matching are four most important parameters in LNA design.

The objective of this tutorial is to outline the basic tradeoffs between different amplifying topologies w.r.t gain, NF and impedance matching. After this comparison it is concluded that inductor degenerated common source topology gives the best performance to meet the gain, NF, and impedance matching goals with minimum power consumption in case of narrow band designs.

Goals:

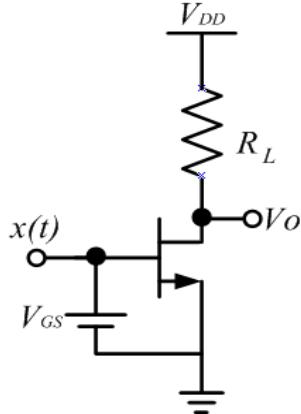
After this tutorial, students should be able to

- Calculate the gain, input impedance and NF of common gate, common source, and shunt feedback amplifiers.
- Understand the basic equations and tradeoff between different LNA topologies.
- Perform the calculation for inductor degenerated common source topology and understand the tradeoff between the gain, NF, and impedance matching.

A supplement tutorial LNA lab is also part of this course which guides through different analyses to design a practical LNA.

Problem-1.1(Tutorial)

NMOS transistor is racing horse in LNA design arena due to its higher mobility compared to PMOS transistors. Calculate the IP3 of NMOS CS amplifier shown below. Assume that NMOS transistor is in saturation.



- a) Consider simplified square law model. (HW)

$$I_D = \frac{K_n}{2} (V_{GS} - V_T)^2$$

- b) Consider the short channel effects as:

$$I_D = \frac{K_n}{2} \left[\frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)} \right]$$

θ = Velocity Saturation, Mobility Degradation

$$V_{GS} - V_T = 0.2V \quad \text{and} \quad \theta = 0.1V^{-1}$$

Observe that this transistor is not a very “short channel” device as $\theta \ll 1$.

- c) What conclusion can be drawn from part b) about the bias current and transconductance of the transistor for higher IP3?

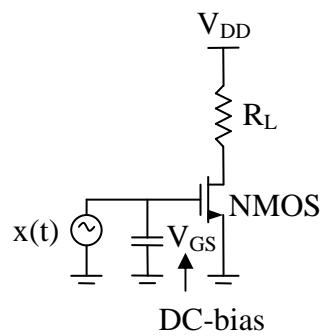
Solution:

- a). Homework answer: $IP3 = \infty$

b).

$$y(x) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \quad \dots \dots \dots (1)$$

$$x(t) = A \cos \omega_1 t + A \cos \omega_2 t \Rightarrow A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|}$$



$$I_D = \frac{K_n}{2} \frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)}$$

Here we assume a small signal $x(t)$ around the bias $(V_{GS} - V_T)$, so

$$I_D = \frac{K_n}{2} \frac{[(V_{GS} - V_T) + x(t)]^2}{1 + \theta(V_{GS} - V_T + x(t))}$$

we define $V_{GS} - V_T = \Delta V$ ----- Bias voltage

$$I_D = \frac{K_n}{2} \frac{[x(t) + \Delta V]^2}{\theta(x(t) + \Delta V) + 1} \quad \text{-----(2)}$$

$$V_o = I_D R_L \Rightarrow V_o = \frac{K_n R_L}{2} \frac{(x(t) + \Delta V)^2}{1 + \theta(x(t) + \Delta V)} \text{ and we put } \frac{K_n R_L}{2} = K$$

$$\theta \ll 1 \text{ so } (x(t) + \Delta V) \text{ is also small} \Rightarrow \frac{1}{1 + \rho} = 1 - \frac{\rho}{2}$$

$$\frac{1}{1 + \theta(x(t) + \Delta V)} \approx 1 - \frac{\theta(x(t) + \Delta V)}{2}$$

$$V_o = K(x(t) + \Delta V)^2 \left(1 - \frac{\theta(x(t) + \Delta V)}{2} \right)$$

$$V_o = K(x(t) + \Delta V)^2 - (x(t) + \Delta V)^3 \frac{K\theta}{2}$$

$$V_o = K\Delta V^2 - \frac{K\theta}{2}\Delta V^3 + \left(2K\Delta V - \frac{3K\theta}{2}\Delta V^2 \right)x(t)$$

$$+ \left(K - \frac{3K\theta}{2}\Delta V \right)x^2(t) - \frac{K\theta}{2}x^3(t) \quad \text{-----(3)}$$

Comparing (1) & (3)

$$\alpha_1 = 2K\Delta V - \frac{3K\theta}{2}\Delta V^2, \quad \alpha_2 = K - \frac{3K\theta}{2}\Delta V, \quad \alpha_3 = -\frac{K\theta}{2}$$

$$A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|} = \sqrt{\frac{4}{3} \times \frac{2K\Delta V - \frac{3}{2}K\theta\Delta V^2}{\frac{K\theta}{2}}} = \sqrt{\frac{8}{3} \left(\frac{2\Delta V}{\theta} - 3\Delta V^2 \right)}$$

$$A_{IP3} = \sqrt{\frac{8}{3} \frac{2\Delta V}{\theta}} = \sqrt{\frac{16}{3} \frac{\Delta V}{\theta}} \quad \text{As } \theta \ll 1 \quad 3\Delta V^2 \text{ can be ignored.}$$

$$A_{IP3} = \sqrt{\frac{16}{3} \frac{\Delta V}{\theta}} = \sqrt{\frac{16}{3} \frac{(V_{GS} - V_T)}{\theta}} \quad \dots \dots \dots \quad (4)$$

Please, note that this formula only holds for small value of θ .

A large gate bias voltage ($V_{GS} - V_T$) improves IP3.

Put $\Delta V = 0.2 \text{V}$, $\theta = 0.1 \text{V}^{-1}$

$$A_{IP3} = \sqrt{\frac{16}{3} \frac{0.2}{0.1}} = 3.27 \text{Volts}$$

$$IIP3(\text{dBm}) = 10 \log \left[\left(\frac{3.27}{\sqrt{2}} \right)^2 \cdot \frac{1}{50} \right] \text{dBm} \cong 20 \text{dBm}$$

c). From $I_D = \frac{K_n}{2} \frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)}$ the NMOS transconductance can be found as

$$g_m = \frac{K_n}{2} \frac{(V_{GS} - V_T)(2 + \theta(V_{GS} - V_T))}{(1 + \theta(V_{GS} - V_T))^2}.$$

By comparison of those two formulas we find

$$\frac{I_D}{g_m} = (V_{GS} - V_T) \times \frac{1 + \theta(V_{GS} - V_T)}{2 + \theta(V_{GS} - V_T)} \cong \frac{V_{GS} - V_T}{2}$$

and hence, (4) can be rewritten as

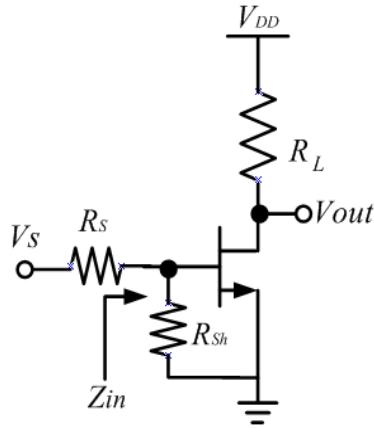
$$A_{IP3} \cong \sqrt{\frac{32}{3\theta} \frac{I_D}{g_m}}$$

As shown, IIP3 is decided by the ratio I_D/g_m which is constant for a given gate bias voltage. Using e.g. a wider transistor does not change this ratio and only the power consumption is increased.

Problem-1.2 (Tutorial)

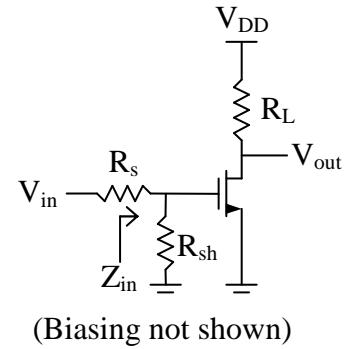
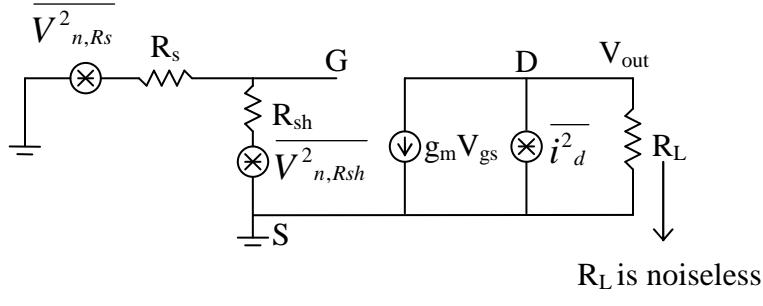
It is preferred in current RF designs that the input of LNA be matched to 50Ω . The easiest way is to shunt the gate with a resistor of 50Ω .

- a) Calculate the gain, input impedance and NF in absence of gate noise. Assume that $R_{sh}=R_L$ for NF derivation.
- b) What are the disadvantages of shunt resistor with reference to gain and NF?



Solution:

a). (Please read assumption in the problem statement carefully)



$$F = \frac{\text{Total output noise power}}{\text{Output noise due to input source}}$$

$$\overline{V^2}_{m,Rs} = 4kTR_s\Delta f$$

$$\text{Gain} \Big|_{\text{Gate}} = -g_m R_L$$

$$\overline{V^2}_{m,Rsh} = 4kTR_{sh}\Delta f$$

$$A = g_m R_L \left(\frac{R_{sh}}{R_s + R_{sh}} \right) \text{ for } R_{sh} = R_s$$

$$\overline{i^2}_d = 4kTg_m\Delta f$$

$$A = -g_m \frac{R_L}{2}$$

Using superposition, considering one at a time and shorting / opening other sources.

$$\overline{V^2}_{on,Rs} = \overline{V^2}_{n,Rs} \times g_m^2 R_L^2 \times \left(\frac{R_{sh}}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2}_{on,Rsh} = \overline{V^2}_{n,Rsh} \times g_m^2 R_L^2 \times \left(\frac{R_s}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2}_{no,d} = \overline{i^2}_d \times R^2_L$$

$$F = \frac{\overline{V^2}_{on,Rs} + \overline{V^2}_{on,Rsh} + \overline{V^2}_{no,d}}{\overline{V^2}_{on,Rs}} = 1 + \frac{\overline{V^2}_{on,Rsh} + \overline{V^2}_{o,d}}{\overline{V^2}_{on,Rs}}$$

$$F = 1 + \frac{\frac{4kTR_{sh}\Delta f \times g_m^2 R^2_L \times R^2_{sh}}{(R_s + R_{sh})^2}}{\frac{4kTR_s\Delta f \times g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}} + \frac{\frac{4kT\gamma g_m \Delta f \times R^2_L}{4kTR_s\Delta f \times g_m^2 R^2_L \times R^2_s}}{\frac{4kTR_s\Delta f \times g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}}$$

In case of impedance match $R_s = R_{sh}$

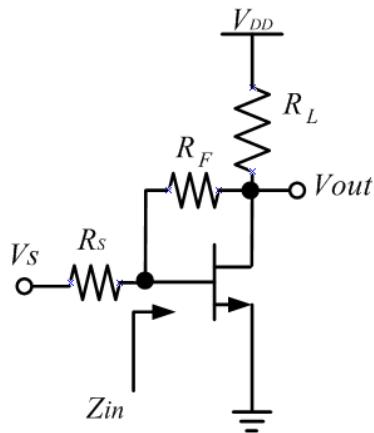
$$F = 1 + 1 + \frac{\frac{R_L^2 \gamma g_m}{R_s \times \frac{g_m^2 R_s^2 \times R_L^2}{4R_s^2}}} = 2 + \frac{\frac{R_L^2 \gamma g_m}{R_s \times \frac{g_m^2 \times R_L^2}{4}}} = 2 + \frac{4\gamma}{g_m R_s}$$

b).

- Poor Noise Figure
- Input signal attenuated by voltage divider
- R_{sh} adds extra noise.
- At high frequency, shunt L is needed to tune out C_{gs}
- Reduced gain.

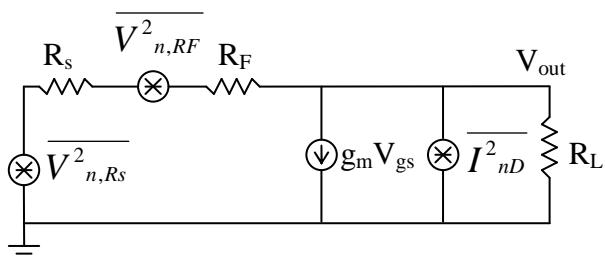
Problem-1.3 (Tutorial)

Another approach to get 50Ω input impedance match is shunt feedback amplifier shown below.

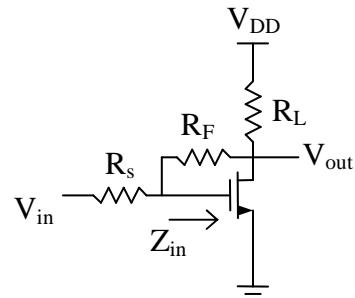


- a) Calculate the gain, input impedance and NF neglecting the gate noise. The gate-drain, gate-bulk, and gate-source capacitance can be neglected as well.
- b) What are the disadvantages of shunt feedback amplifier with reference to gain and NF?

Solution:



(Equivalent noise model ignoring gate noise), R_L is noiseless



(Biasing not shown)

$$\overline{I^2_{nD}} = 4kT\gamma g_m \Delta f, \overline{V^2_{n,RS}} = 4kTR_s \Delta f$$

$$F = \frac{\overline{V^2_{n,out}}}{A_{v,tot}^2 \overline{V^2_{RS}}} = \frac{\text{Total input noise power}}{\text{Output noise power due to input source}}$$

Here $A_{v,tot}$ = Gain from V_{in} to V_{out}

Again using superposition theorem

$$F = \frac{\overline{V^2_{n,out}}}{A_{v,tot}^2 \overline{V^2_{RS}}} = \frac{\overline{V^2_{n,RS,out}} + \overline{V^2_{n,RF,out}} + \overline{V^2_{n,D,out}}}{A_{v,tot}^2 \overline{V^2_{RS}}}$$

Gain Calculation

$$V_{in} = i_{in}(R_s + R_F) + V_{out}$$

$$V_{out} = (i_{in} - g_m V_{gs}) R_L$$

$$V_{gs} = i_{in} R_F + V_o$$

$$A_{v,tot} = \frac{V_{out}}{V_{in}} = \frac{R_L(1 - g_m R_L)}{R_s + R_F + R_L + g_m R_s R_L}$$

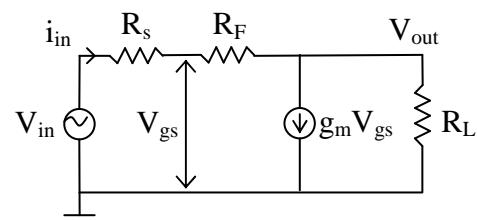
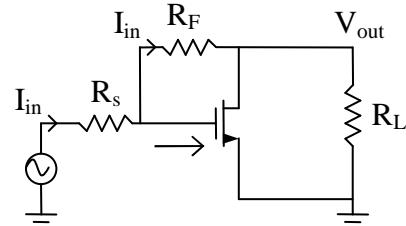
If $R_F \gg R_s$ & $g_m R_F \gg 1$

$$A_{v,tot} = \frac{-g_m R_L}{\frac{R_s}{R_F} + 1 + R_L + \frac{1 + g_m R_s}{R_F}} \approx -g_m R_L$$

$$A_{v,tot} \approx -g_m R_L$$

$$\text{Also } Z_{in} = \frac{R_F + R_L}{1 + g_m R_L}$$

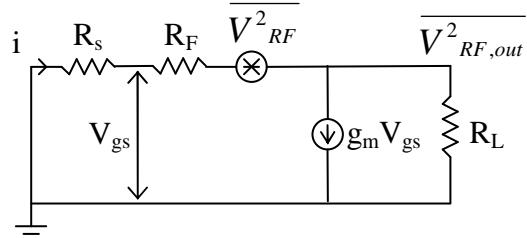
By ignoring C_{gs} , we have considered real part only.



For source resistance

$$\overline{V_{nRS,out}^2} = A_{v,tot}^2 \overline{V_{nRS}^2} \quad \dots \dots \dots (1)$$

For feedback resistance



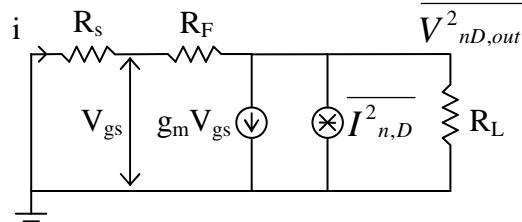
$$V_{gs} = -iR_s = iR_F - V_{RF} + V_{RF,out}$$

$$V_{RF,out} = R_L(i - g_m V_{gs})$$

$$V_{RF,out} = V_{RF} \frac{1}{1 + \frac{R_s + R_F}{R_L(1 + g_m R_s)}} = V_{RF} \frac{R_L}{R_F} (1 + g_m R_s)$$

$$\overline{V_{n,RF,out}^2} = \overline{V_{n,RF}^2} \left[\frac{R_L}{R_F} (1 + g_m R_s) \right]^2 \quad \dots \dots \dots (2)$$

Similarly



$$\frac{V_{nD,out}}{R_L} + I_{nD} + g_m V_{gs} + \frac{V_{nD,out}}{R_S + R_F} = 0$$

$$V_{gs} = R_S \frac{V_{nD,out}}{R_S + R_F}$$

$$V_{nD,out} = \frac{I_{nD}}{\frac{1}{R_L} + \frac{1}{R_S + R_F} + \frac{g_m R_S}{R_S + R_F}} \approx I_{nD} R_L$$

So,

$$\overline{V_{nD,out}^2} = I_{nD}^2 R_L^2 \quad \dots \dots \dots (3)$$

Combaining (1) (2) & (3)

$$F = 1 + \frac{\overline{V^2}_{n,RF} \left[\frac{R_L}{R_F} (1 + g_m R_s) \right]^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}} + \frac{\overline{I^2}_{nD} R_L^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}}$$

$$A_{v,tot} = -g_m R_L, \quad \overline{V^2}_{n,RS} = 4kTR_S \Delta f, \quad \overline{V^2}_{M,RF} = 4kTR_F \quad \& \quad \overline{I^2}_{nD} = 4kT\gamma g_m$$

$$F = 1 + \frac{R_s}{R_F} \left(1 + \frac{1}{g_m R_s} \right)^2 + \frac{\gamma}{g_m R_s}$$

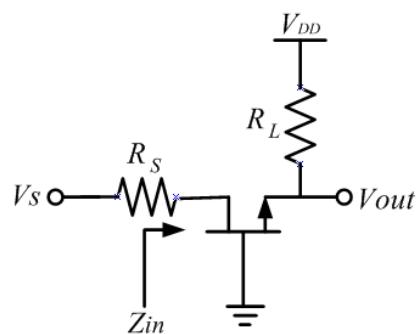
b).

NF \downarrow $g_m R_s \uparrow$ & $R_F \uparrow$ usually $R_s = 50\Omega$

- Better performance than CS amplifier
- R_F induces noise
- At higher $f \uparrow$ a shunt inductor needed to tune out C_{gs}
- Broadband Amp @ Lower frequency
- To make NF \downarrow $R_F > R_s$ and $g_m R_s \gg 1$

Problem-1.4 (HW)

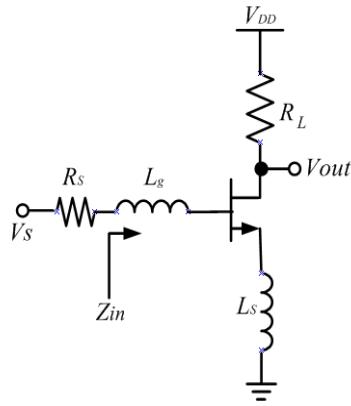
Common gate amplifier also offers 50Ω input impedance match and solves the input matching problem.



- c) Calculate the gain, input impedance and NF in absence of gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- a) What are the disadvantages of common gate amplifier with reference to gain and NF?

Problem-1.5 (Tutorial)

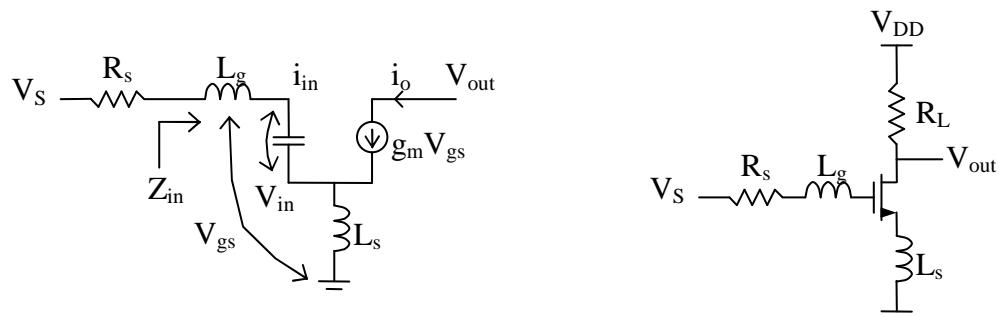
The disadvantages of the amplifiers discussed in Problem-2, 3 & 4 can be circumvented by using the source degenerated LNA shown below.



- Calculate the input impedance. This inductor source degenerated amplifier presents a noiseless resistance for 50Ω for input power match. How we can cancel the imaginary part of complex input impedance so that the LNA presents 50Ω real input resistance at input port.
- Calculate the NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- C_{gd} bridges the input and output ports. The reverse isolation of this LNA is very poor. Why reverse isolation is important? Suggest the modification to improve reverse isolation.

Solution:

a).



(Biasing not shown)

From model above we can write

$$V_{in} = i_{in} \left(j\omega L_g + j\omega L_s \right) + i_{in} \left(\frac{1}{j\omega C_{gs}} \right) + i_o j\omega L_s \quad \dots \dots \dots (1)$$

$$i_o = g_m V_{gs} = g_m i_{in} \times \frac{1}{j\omega C_{gs}} \quad \dots \dots \dots (2)$$

Substituting (2) in (1)

$$V_{in} = i_{in} \left[j\omega (L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}} \right]$$

$$Z_{in} = \frac{V_{in}}{I_{in}} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

For matching $L_g + L_s$ are canceled out by C_{gs} . So at frequency of interest

$$\omega_o(L_g + L_s) = \frac{1}{\omega_o C_{gs}} \Rightarrow \omega_o^2 = \frac{1}{(L_g + L_s)C_{gs}}$$

$$\text{And } R_S = 50\Omega = \frac{g_m}{C_{gs}} L_s$$

Notes:

a). L_s is typically small and may be realized by the bond wire for source.

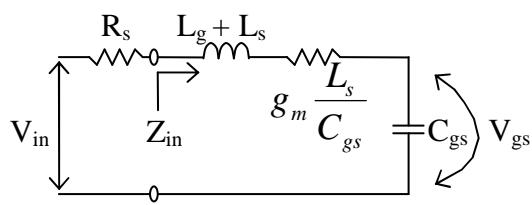
b). L_g can be implemented by spiral/external inductor.

b).

From part a)

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

We can draw this circuit as



Here

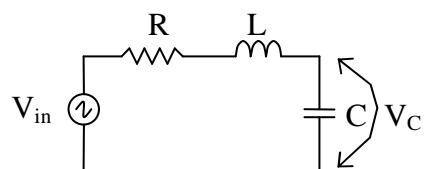
$$Q_{in} = \frac{\omega_o(L_g + L_s)}{R_s + \frac{g_m L_s}{C_{gs}}} = \frac{\omega_o(L_g + L_s)}{R_s + \omega_T L_s}$$

$$\because \omega_T \equiv \frac{g_m}{C_{gs}} \quad \text{frequency of current gain equal 1}$$

$$Q_{in} = \frac{1}{\omega_o \left(R_s + \frac{g_m L_s}{C_{gs}} \right) C_{gs}} \quad \text{for match load } R_S = \frac{g_m L_s}{C_{gs}}$$

Reference:

For series RLC Circuit



$$Q_s = \frac{1}{R} \sqrt{\frac{L}{C}} = \frac{\omega_o L}{R} = \frac{1}{\omega_o R C}$$

$$\text{and } V_C = Q_s V_{in}$$

$$Q_{in} = \frac{1}{2\omega_o R_s C_{gs}}$$

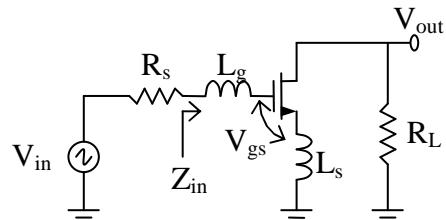
Gain

$$V_{gs} = Q_{in} V_{in}$$

$$g_m = \frac{I_{out}}{V_{gs}}$$

$$G_m = \frac{I_{out}}{V_{in}} = \frac{V_{gs} g_m}{V_{in}} = Q_{in} g_m$$

$$G_m = Q_{in} g_m$$



so, $\frac{V_{out}}{V_{in}} = -G_m R_L$ where $G_m = Q_{in} g_m$

Noise Figure:

$$F = \frac{\text{Total noise power at output}}{\text{noise power at output due to input source}}$$

For this calculation we ignore channel noise.

$$F = \frac{\overline{V^2}_{nRS,OUT} + \overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}} = 1 + \frac{\overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}}$$

$$\overline{V^2}_{nD,OUT} = \overline{i^2}_{n,D} R_L^2 \quad \overline{i^2}_{n,D} = 4kT \gamma g_m \Delta f$$

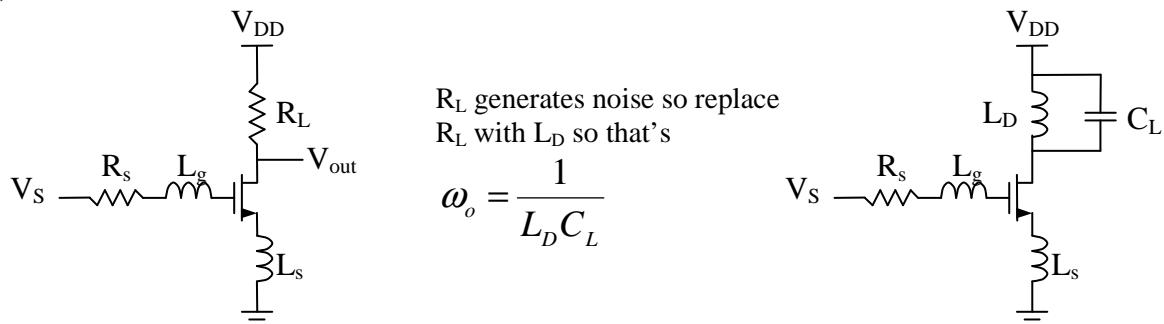
$$\overline{V^2}_{nRS,OUT} = \overline{V^2}_{n,RS} G_m^2 R_L^2 \quad \overline{V^2}_{n,RS} = 4kT R_S \Delta f \quad \& \quad G_m = Q_{in} g_m$$

$$F = 1 + \frac{\overline{i^2}_{n,D} R_L^2}{\overline{V^2}_{n,RS} Q_{in}^2 g_m^2 R_L^2} \quad \overline{i^2}_{n,D} = 4kT \gamma g_m, \quad \overline{V^2}_{n,RS} = 4kT R_S$$

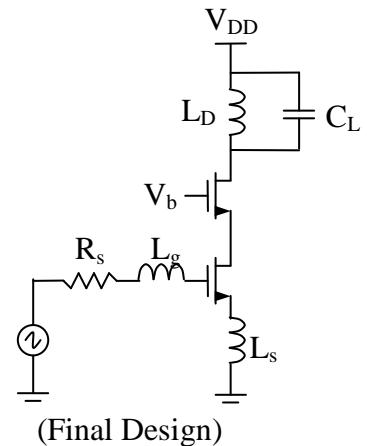
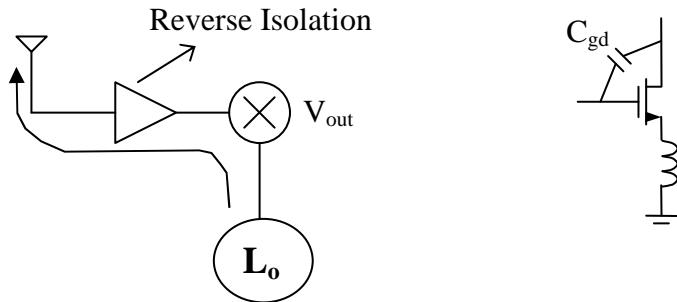
$$F = 1 + \frac{\gamma}{g_m R_S Q_{in}^2}$$

Notes:

- Very good NF value
- Narrow band matching
- NF \downarrow with Q^2
- The Q is dependent upon $L_g + L_s$, L_s usually small so Q depends mainly upon L_g

C). Drawbacks**i).**

The C_L can be considered the input capacitance of the following mixer or filter.

ii).

Reverse isolation depends upon capacitance between output and input.
To make it less the cascode architecture can be used.

Problem-1.6 (HW)

Fill-in the Table below, use the data from Problem-1.2, 1.4, 1.3 and 1.5

Type of LNA	Z_{in}	Noise Factor	Gain	NF (dB)
Shunt Resistor	R_{sh}	$2 + \frac{4\gamma}{g_m R_s}$	$\frac{-g_m R_L}{2}$	
Common Gate				
Shunt Feedback				
Source Degenerated				

- Calculate the NF for all above amplifiers. Assume $\gamma=2$, $g_m = 20\text{mS}$, $R_s = 50\Omega$, $R_f = 500\Omega$, and $Q_{in} = 2$.
- Which is the best topology for Narrow Band LNA design at high frequency?

Problem-1.7 (Tutorial)

Real Design: We will design the inductor-source-degenerated LNA shown in Fig below to meet the specification outlined for IEEE802.11b standard. The first cut approximate values are calculated as a starting point for simulation.

LNA Specification:

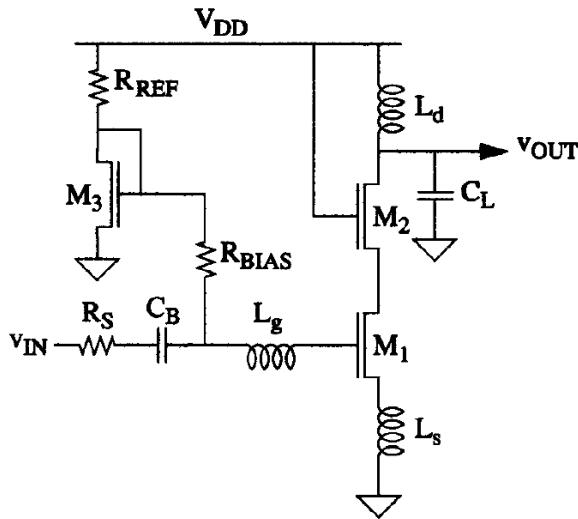
$NF < 2.5 \text{ dB}$, Gain $> 15 \text{ dB}$, $IIP3 > -5 \text{ dBm}$, Centre Frequency = 2.4 GHz

Load Capacitance = 1pF

Technology Parameters for 0.35um CMOS:

$$L_{eff} = 0.35\mu m, \quad \mu_n C_{ox} = 170 \mu A/V^2, \quad C_{ox} = 4.6 mF/m^2, \quad \mu_p C_{ox} = 58 \mu A/V^2, \quad \gamma = 2$$

$$\delta = 4, \quad |C| = 0.395, \quad \alpha = 0.85$$



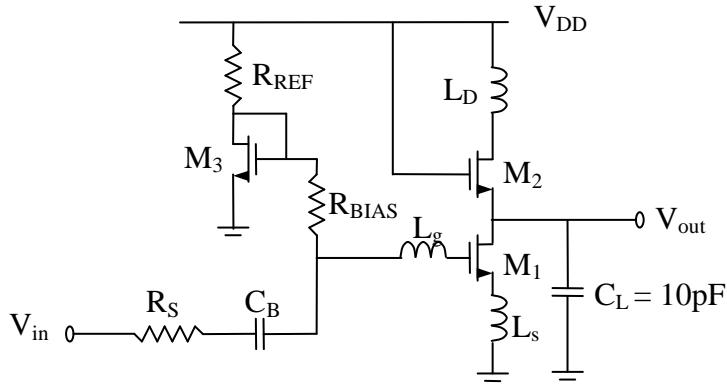
Solution:

$$\text{Technology } 0.35\mu m \text{ CMOS: } \left\{ \begin{array}{l} \mu_o C_{ox} = 170 \mu A/V^2, \mu_p C_{ox} = 58 \mu A/V^2, \\ C_{ox} = 4.6 mF/m^2, \gamma = 2, L_{eff} = 0.35\mu m \end{array} \right\}$$

$$\delta = 4, \quad |C| = 0.395, \quad \alpha = 0.85$$

Design Parameters

$NF < 2.5 \text{ dB}$, Gain $> 15 \text{ dB}$, $IIP3 > -5 \text{ dBm}$, $f_0 = 2.4 \text{ GHz}$



Component Description

L_s – Matches input impedance

L_g – Sets the Resonant Frequency $f_O = 2.4$ GHz

M_3 – Biasing transistor which forms current mirror with M_1

L_d – Tuned output increases the gain and also work as band pass filter with C_L

M_2 – Isolates tuned input from output to increase reverse isolation, also reduces the effect of Miller capacitance C_{gd}

C_B – BC blocking capacitor chosen to have negligible reactance at $f_O = 2.4$ GHz

R_{BIAS} – Large enough so that its equivalent current noise is small enough to be ignored. (Don't consider it as voltage noise source. Why??)

Design Procedure

Size of M_1 :

From the noisy two-port theory (see the course book or lecture notes) the optimal input matching and minimum noise figure is given by:

$$G_{opt} = \alpha \omega C_{gs} \sqrt{\frac{\delta}{5\gamma} (1 - |C|^2)} = \frac{1}{50\Omega} \quad \text{-----(A)}$$

$$F_{min} = 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_T} \sqrt{\gamma C (1 - |C|^2)} = 1 + 2.3 \frac{\omega}{\omega_T} \quad \text{-----(B)}$$

From (A)

$$C_{gs} \approx 2.7 \text{ pF} \Rightarrow W_{M1} \approx 3C_{gs} / 2C_{OX} L_{eff} \approx 2.5 \text{ mm} \quad (\text{not feasible - huge size, huge power !})$$

Conclusion: We will not go for the global minimum noise figure. Instead, we will look into the constraint power design approach.

Solution:

LNA NF will be optimized for given power which is higher than the global minimum NF.

In this case the optimum transistor width is given by:

$$W_{opt} = \frac{1}{3\omega_o L_{eff} C_{ox} R_s}$$

while the minimum power-constraint NF :

$$F_{min,p} = 1 + 2.4 \frac{\gamma}{\alpha} \frac{\omega}{\omega_T} \Rightarrow F_{min,p} = 1 + 5.6 \frac{\omega}{\omega_T} \quad \text{-----(C)}$$

(B) is the global minimum noise figure.

(C) is the minimum NF for a given power consumption.

In practice the difference is usually 0.5dB to 1dB (no big deal for Lower Power)

Step - 1:

$$I_1 = I_2 = 5mA \text{ (Limited Power consumption)}$$

Step - 2:

$$W_{M1} = \frac{1}{3\omega_0 L_{eff} C_{ox} R_s}$$

$$W_{M1} = \frac{1}{3 \times 0.35\mu \times 4.6m \times 50 \times \omega_o} \quad \left. \begin{cases} R_s = 50\Omega, C_{ox} = 4.6mF/m^2, \\ \mu_n C_{ox} = 170\mu A/V, L_{eff} = 0.35\mu m, \\ \omega_o = 2\pi f_o, f_o = 2.4GHz \end{cases} \right\}$$

$$W_{M1} = 3.9 \times 10^{-4}$$

$$W_{M1} = 3.9 \times 10^{-4} = 390\mu m$$

Step - 3:

$$C_{gs1} = \frac{2}{3} W_{M1} L_{eff} C_{ox}$$

$$C_{gs1} = \frac{2}{3} \times 390\mu \times 0.35\mu \times 4.6m = 0.41pF$$

$$g_{m1} = \sqrt{2\mu_n C_{ox} \left(\frac{W}{L} \right)_{M1} I_{DM1}} \quad \text{or} \quad g_{m1} = \frac{2I_{DM1}}{V_{GS} - V_T} \quad (\text{for short channel model})$$

$$g_{m1} = \sqrt{2 \times 170\mu \times \left(\frac{390}{0.35} \right) \times 5m} = 43mA/V$$

$$\omega_T \approx \frac{g_m}{C_{gs1}} = \frac{43 \text{ mA/V}}{0.41 \text{ pF}} = 104G \text{ rad/Sec}$$

Assuming $\gamma = 2$

$$\text{Now } F_{\min} = 1 + 5.6 \frac{\omega_o}{\omega_T}$$

$$F_{\min} = 1 + 5.6 \frac{2\pi 2.4 G}{104 G} \approx 2.55 \text{ dB}$$

$$NF \approx 2.55 \text{ dB}$$

This NF is very close to the specified value. If we increase I_D then ω_T should increase slightly as well and hence, a lower NF value can be achieved at expense of more power.

Step - 4:

Source and gate inductance such that they cancel C_{gs} and set 50Ω input impedance

$$\omega_o = 2\pi f_o = 2\pi 2.4 = 15G \text{ rad/Sec}$$

From previous problem

$$R_s = R_{transformed} = \frac{g_m L_s}{C_{gs}} \cong \omega_T L_s$$

$$L_s = \frac{R_s}{\omega_T} = \frac{50}{100G} \cong 0.5nH$$

$L_s = 0.5nH$ can be implemented using the bond wire.

$$\text{Now } L_g + L_s = \frac{1}{\omega_0^2 C_{gs1}}$$

$$L_g + L_s = \frac{1}{(15G)^2 \times 0.41 \text{ pF}} = 10.81nH$$

$$L_g \approx 10nH$$

Step - 5:

$$L_d = \frac{1}{\omega_o^2 C_L} \quad \because C_L = 1 \text{ pF}$$

$$L_d = \frac{1}{(15G)^2 \times 1 \text{ pF}} \cong 4.4nH$$

$$L_d = 4.4nH$$

Step - 6:

Size of M3 is chosen to minimize power consumption

$$W_{M3} = 70\mu m, \quad R_{REF} = 2k\Omega \Rightarrow I_3 = 0.6mA$$

$R_{BIAS} = 2k\Omega$ (Large enough so that it's equivalent current noise can be neglected)

$$C_B = 10pF \quad (X_C \approx 6.6\Omega \text{ so good value @ 2.4G} \quad X_B = \frac{1}{2\pi f_o C_B} = 6.6\Omega)$$

Step - 7:

Size M2 = M3

So that they can have shared Drain Area..

(Note: You will simulate same LNA circuit in LAB # 2)

Tutorial-2

Low Noise Amplifier (LNA) Design

Written By: Rashad.M.Ramzan
rashad.ramzan@nu.edu.pk

Objective:

Low noise amplifiers are one of the basic building blocks of any communication system. The purpose of the LNA is to amplify the received signal to acceptable levels with minimum self generated additional noise. Gain, NF, non-linearity and impedance matching are four most important parameters in LNA design.

The objective of this tutorial is to outline the basic tradeoffs between different amplifying topologies w.r.t gain, NF and impedance matching. After this comparison it is concluded that inductor degenerated common source topology gives the best performance to meet the gain, NF, and impedance matching goals with minimum power consumption in case of narrow band designs.

Goals:

After this tutorial, students should be able to

- Calculate the gain, input impedance and NF of common gate, common source, and shunt feedback amplifiers.
- Understand the basic equations and tradeoff between different LNA topologies.
- Perform the calculation for inductor degenerated common source topology and understand the tradeoff between the gain, NF, and impedance matching.

A supplement tutorial LNA lab is also part of this course which takes the circuit from Problem-2.8 and guides through different analysis to design and practical LNA.

Problem-2.1(Tutorial)

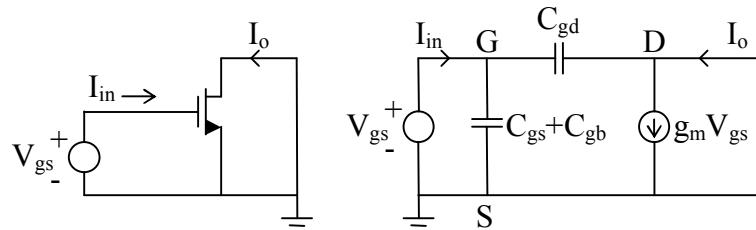
ω_T is single figure of merit for high frequency transistors. This is defined as frequency at which current gain is extrapolated to fall to unity.

Although the dc gate current of an MOS transistor is essentially zero, the high-frequency behavior of the transistor is controlled by the capacitive elements in the small-signal model, which cause the gate current to increase as frequency increases.

- Derive the expression for ω_T .
- For RF design we always use minimum length transistors. Why?

Solution:

a).



$$V_{sb} = V_{ds} = 0$$

So g_{mb} , r_o , C_{sb} , C_{db} have no effect on calculations. (This is drawback of ω_T definition)

$$i_i = j\omega(C_{gs} + C_{gb} + C_{gd})V_{gs}$$

$$i_i \approx g_m V_{gs}$$

$$\frac{i_o}{i_{in}} = \frac{g_m}{j\omega(C_{gs} + C_{gb} + C_{gd})}$$

According to definition $\frac{i_o}{i_{in}} = 1$ at ω_T

$$\omega = \omega_T = \frac{g_m}{C_{gs} + C_{gb} + C_{gd}}$$

C_{gb} and C_{gd} are small compared to C_{gs}

$$\text{So, } \omega_T \approx \frac{g_m}{C_{gs} + C_{gd}} \approx \frac{g_m}{C_{gs}} \quad \dots \dots \dots \quad (1)$$

$$\text{b). } I_D = \frac{\mu_o C_{ox}}{2} \frac{W}{L} (V_{gs} - V_T)^2$$

$$g_m = \frac{\partial I_D}{\partial V_{gs}} = \mu_o C_{ox} \frac{W}{L} (V_{gs} - V_T) \quad \dots \dots \dots \quad (2)$$

And $C_{gs} = C_o \times WL$ ----- (3)

Put (2) & (3) in (1)

$$\omega_T = \frac{g_m}{C_{gs}} = \frac{\mu_o C_{ox} W (V_{gs} - V_t)}{L \cdot C_{ox} \cdot W \cdot L} = \frac{\mu_o (V_{gs} - V_t)}{L^2}$$

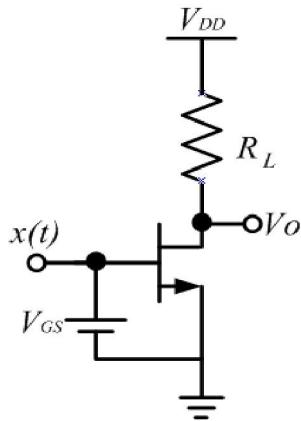
$$\omega_T = \frac{\mu_o (V_{gs} - V_t)}{L^2}$$

This means that $\omega_T \propto \frac{1}{L^2}$ so that's why minimum L is preferred. But this approximation holds

for long channel devices for short channel $\omega_T \propto \frac{1}{L}$ instead of $\frac{1}{L^2}$.

Problem-2.2(Tutorial)

NMOS transistor is racing horse in LNA design arena due to its higher mobility compared to PMOS transistors. Calculate the IP3 of NMOS CS amplifier shown below. Assume that NMOS transistor is in saturation.



- a) Consider simplified square law model. (**HW**)

$$I_D = \frac{K_n}{2} (V_{GS} - V_T)^2$$

- b) Consider the short channel effects as (**Tutorial**)

$$I_D = \frac{K_n}{2} \left[\frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)} \right]$$

$\theta = \text{Velocity Saturation, Mobility Degradation}$

$$V_{GS} - V_T = 0.2V \quad \text{and} \quad \theta = 0.1V^{-1}$$

- c) What conclusion can be drawn from part b) about the bias current and transconductance of transistor for higher IP3?

Solution:

a). Home work Ans: $IP3 = \alpha$

b). From Razavi

$$y(x) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \quad \dots \dots \dots (1)$$

$$x(t) = A \cos \omega_l t + A \cos \omega_2 t \Rightarrow IP3 = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|}$$

$$I_D = \frac{K_n}{2} \frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)}$$

Here we assume that small signal $x(t)$ over-rides $(V_{GS} - V_T)$.

So,

$$I_D = \frac{K_n}{2} \frac{[(V_{GS} - V_T) + x(t)]^2}{1 + \theta(V_{GS} - V_T + x(t))}$$

$$\& V_{GS} - V_T = \Delta V \quad \dots \dots \dots \text{Large signal}$$

$$X(t) \quad \dots \dots \dots \text{Small signal}$$

$$I_D = \frac{K_n}{2} \frac{[x(t) + \Delta V]^2}{\theta(x(t) + \Delta V) + 1}$$

$$V_o = I_D R_L \Rightarrow V_o = \frac{K_n R_L}{2} \frac{(x(t) + \Delta V)^2}{1 + \theta(x(t) + \Delta V)} \text{ put } \frac{K_n R_L}{2} = K$$

$$\theta \ll 1 \text{ So } (x(t) + \Delta V) \text{ is also small} \Rightarrow \frac{1}{1 + x} = 1 - \frac{x}{2}$$

$$\frac{1}{1 + \theta(x(t) + \Delta V)} \approx 1 - \frac{\theta(x(t) + \Delta V)}{2}$$

$$V_o = K(x(t) + \Delta V)^2 \left(1 - \frac{\theta(x(t) + \Delta V)}{2} \right)$$

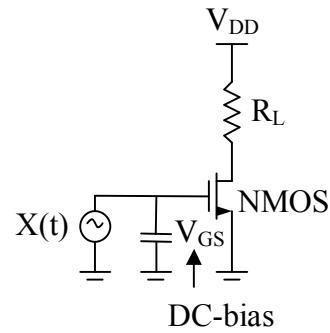
$$V_o = K(x(t) + \Delta V)^2 - (x(t) + \Delta V)^3 \frac{K\theta}{2}$$

$$V_o = K\Delta V^2 - \frac{K\theta}{2} \Delta V^3 + \left(2K\Delta V - \frac{3K\theta}{2} \Delta V^2 \right) x(t) + \left(K - \frac{3K\theta}{2} \Delta V \right) x^2(t) - \frac{K\theta}{2} x^3(t)$$

----- (2)

Small signal components

$$V_o = \left(2K\Delta V - \frac{3K\theta}{2} \Delta V^2 \right) x(t) + \left(K - \frac{3K\theta}{2} \Delta V \right) x^2(t) - \frac{K\theta}{2} x^3(t) \quad \dots \dots \dots (3)$$



Comparing (1) & (3)

$$\alpha_1 = 2K\Delta V - \frac{3K\theta}{2}\Delta V^2, \quad \alpha_2 = K - \frac{3K\theta}{2}\Delta V, \quad \alpha_3 = -\frac{K\theta}{2}$$

$$IP3 = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|} = \sqrt{\frac{4}{3} \times \frac{\frac{2K\Delta V - \frac{3}{2}K\theta\Delta V^2}{K\theta}}{\frac{2}{2}}} = \sqrt{\frac{8}{3} \left(\frac{2\Delta V}{\theta} - 3\Delta V^2 \right)}$$

$$IP3 = \sqrt{\frac{8}{3} \frac{2\Delta V}{\theta}} = \sqrt{\frac{16}{3} \frac{\Delta V}{\theta}} \quad \text{As } \theta \ll 1 \quad 3\Delta V^2 \text{ can be ignored.}$$

$$IP3 = \sqrt{\frac{16}{3} \frac{\Delta V}{\theta}} = \sqrt{\frac{16}{3} \frac{(V_{GS} - V_T)}{\theta}} \quad \dots \dots \dots \quad (4) \quad \therefore I_D = \frac{\mu_o C_{ox}}{2} \frac{W}{L} (V_{GS} - V_T)^2$$

Put $\Delta V = 0.2V$, $\theta = 0.1$

$$IP3 = \sqrt{\frac{16}{3} \frac{0.2}{0.1}} = 3.27 \text{ Volts}$$

$$IP3(\text{dBm}) = 10 \log \left[\left[\left(\frac{3.27}{\sqrt{2}} \right)^2 \cdot \frac{1}{50} \right] / 1 \text{ mW} \right] \cong 20 \text{ dBm}$$

$$\begin{aligned} g_m &= \frac{\partial I_D}{V_{GS}} = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T) \\ \frac{g_m}{I_D} &= \frac{2}{V_{GS} - V_T} = \frac{2}{\Delta V} \end{aligned} \quad \text{This is just an approximation with } I_D \text{ & } g_m$$

$$\frac{g_m}{I_D} = \frac{2}{V_{GS} - V_T} = \frac{2}{\Delta V}$$

$$\Rightarrow IP3 \cong \sqrt{\frac{32}{3\theta} \sqrt{\frac{I_D}{g_m}}} \quad \therefore g_m = \sqrt{\mu_n C_{ox} \frac{W}{L} I_D}$$

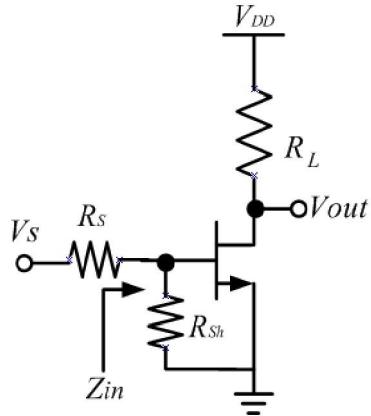
c).

- To increase IIP3 $I_D \uparrow$ (high power) or $g_m \downarrow$ (high noise)
- g_m also depends upon I_D , so when $I_D \uparrow$ $g_m \uparrow$ but at that rate $\propto \sqrt{I_D}$
- When W increases $g_m \uparrow$ for same I_D (Power consumption) so this decreases IP3
- The above observations are for long-channel. But for short channel $(4) \Rightarrow (V_{GS} - V_T) \uparrow$ then $I_D \uparrow$ any how.

Problem-2.3 (Tutorial)

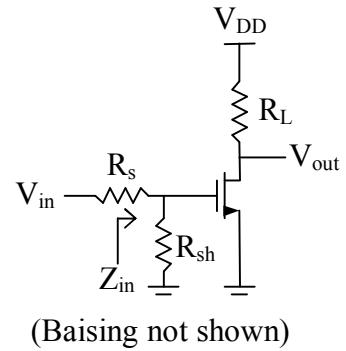
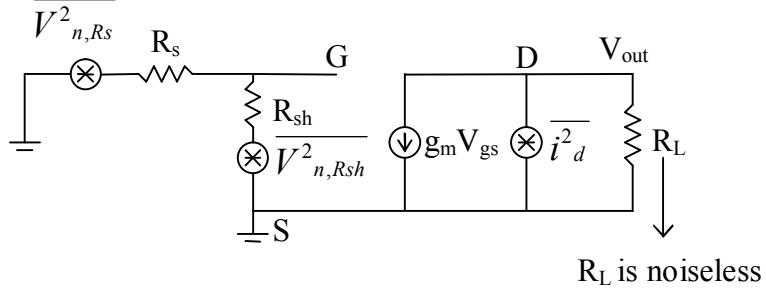
It is preferred in current RF designs that the input of LNA be matched to 50Ω (Razavi, Pg168). The easiest way is to shunt the gate with a resistor of 50Ω .

- Calculate the gain, input impedance and NF in absence of gate noise. Assume that $R_{sh}=R_L$ for NF derivation.
- What are the disadvantage of shunt resistor with reference to gain and NF?



Solution:

a). (Please read assumption in problem statement carefully)



$$F = \frac{\text{Total output noise power}}{\text{Output noise due to input source}}$$

$$\overline{V^2_{m, Rs}} = 4KTR_s \Delta f$$

$$\text{Gain} \Big|_{\text{Gate}} = -g_m R_L$$

$$\overline{V^2_{m, Rsh}} = 4KTR_{sh} \Delta f$$

$$A = g_m R_L \left(\frac{R_{sh}}{R_s + R_{sh}} \right) \text{ for } R_{sh} = R_s$$

$$\overline{i^2_d} = 4KT \gamma g_m \Delta f$$

$$A = -g_m \frac{R_L}{2}$$

Using superposition, considering one at a time and shorting / opening other sources.

$$\overline{V^2_{on, Rs}} = \overline{V^2_{n, Rs}} \times g_m^2 R_L^2 \times \left(\frac{R_{sh}}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2_{on, Rsh}} = \overline{V^2_{n, Rsh}} \times g_m^2 R_L^2 \times \left(\frac{R_s}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2_{no,d}} = \overline{i^2_d} \times R^2_L$$

$$F = \frac{\overline{V^2_{on,Rs}} + \overline{V^2_{on,Rsh}} + \overline{V^2_{no,d}}}{\overline{V^2_{on,Rs}}} = 1 + \frac{\overline{V^2_{on,Rsh}} + \overline{V^2_{o,d}}}{\overline{V^2_{on,Rs}}}$$

$$F = 1 + \frac{4KTR_{sh}\Delta f \times \frac{g_m^2 R^2_L \times R^2_{sh}}{(R_s + R_{sh})^2}}{4KTR_s\Delta f \times \frac{g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}} + \frac{4KT\gamma g_m \Delta f \times R^2_L}{4KTR_s\Delta f \times \frac{g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}}$$

In case of impedance match $R_s = R_{sh}$

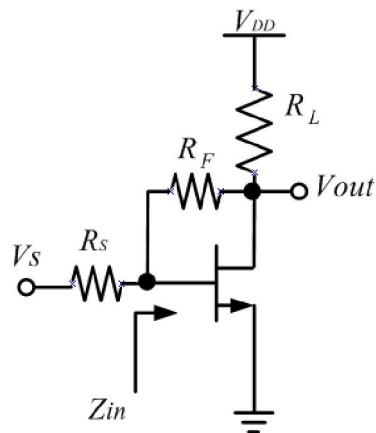
$$F = 1 + 1 + \frac{R^2_L \gamma g_m}{R_s \times \frac{g_m^2 R^2_s \times R^2_L}{4R^2_s}} = 2 + \frac{R^2_L \gamma g_m}{R_s \times \frac{g_m^2 \times R^2_L}{4}} = 2 + \frac{4\gamma}{g_m R_s}$$

b).

- Poor Noise Figure
- Input signal attenuated by voltage divider
- R_{sh} adds extra noise.
- At high frequency, shunt L is needed to tune out C_{gs}
- Reduced gain

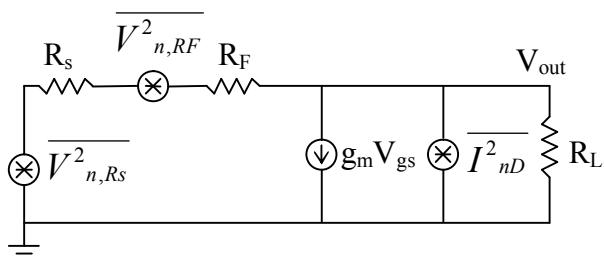
Problem-2.4 (Tutorial)

Another approach to get 50Ω input impedance match is shunt feedback amplifier shown below.

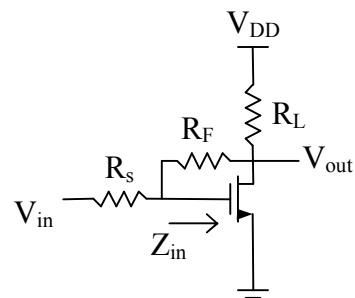


- a) Calculate the gain, input impedance and NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- b) What are the disadvantage of shunt feedback amplifier with reference to gain and NF?

Solution:



(Equivalent noise model ignoring gate noise), RL is noiseless



(Basing not shown)

$$\overline{I^2_{nD}} = 4KT\gamma g_m \Delta f, \overline{V^2_{n,RS}} = 4KTR_s \Delta f$$

$$NF = \frac{\overline{V^2_{n,out}}}{A^2_{v,tot} \overline{V^2_{RS}}} = \frac{\text{Total input noise power}}{\text{Output noise power due to input source}}$$

Here $A_{v,tot}$ = Gain from V_{in} to V_{out}

Again using superposition theorem

$$NF = \frac{\overline{V^2_{n,out}}}{A^2_{v,tot} \overline{V^2_{RS}}} = \frac{\overline{V^2_{n,RS,out}} + \overline{V^2_{n,RF,out}} + \overline{V^2_{n,D,out}}}{A^2_{v,tot} \overline{V^2_{RS}}}$$

Gain Calculation

$$V_{in} = i_{in}(R_s + R_F) + V_{out}$$

$$V_{out} = (i_{in} - g_m V_{gs}) R_L$$

$$V_{gs} = i_{in} R_F + V_o$$

$$A_{v,tot} = \frac{V_{out}}{V_{in}} = \frac{R_L(1 - g_m R_L)}{R_s + R_F + R_L + g_m R_S R_L}$$

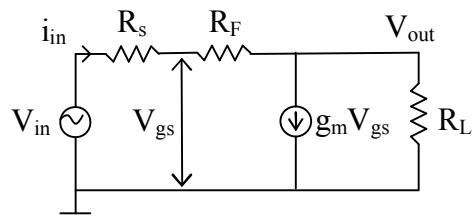
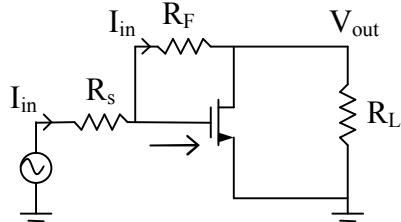
If $R_F \gg R_s$ & $g_m R_F \gg 1$

$$A_{v,tot} = \frac{-g_m R_L}{\frac{R_s}{R_F} + 1 + R_L + \frac{1 + g_m R_s}{R_F}} \approx -g_m R_L$$

$$A_{v,tot} \approx -g_m R_L$$

$$\text{Also } Z_{in} = \frac{R_F + R_L}{1 + g_m R_L}$$

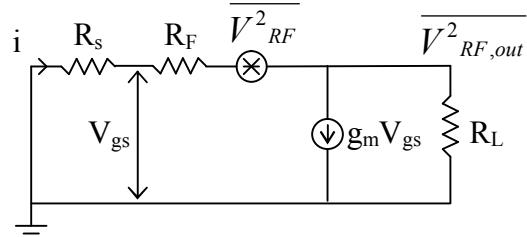
By ignoring C_{gs} , we have considered real part only.



For source resistance

$$\overline{V^2_{nRS,out}} = A_{v,tot}^2 \overline{V^2_{nRS}} \quad \text{-----(1)}$$

For feedback resistance



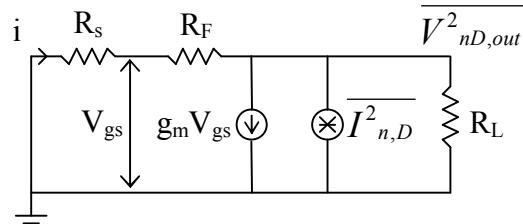
$$V_{gs} = -iR_s = iR_F - V_{RF} + V_{RF,out}$$

$$V_{RF,out} = R_L(i - g_m V_{gs})$$

$$V_{RF,out} = V_{RF} \frac{1}{1 + \frac{R_S + R_F}{R_L(1 + g_m R_s)}} = V_{RF} \frac{R_L}{R_F} (1 + g_m R_s)$$

$$\overline{V^2_{nRF,out}} = \overline{V^2_{nRF}} \left[\frac{R_L}{R_F} (1 + g_m R_s) \right]^2 \quad \text{-----(2)}$$

Similarly



$$\frac{V_{nD,out}}{R_L} + \overline{I_{nD}} + g_m V_{gs} + \frac{V_{nD,out}}{R_S + R_F} = 0$$

$$V_{gs} = R_S \frac{V_{nD,out}}{R_S + R_F}$$

$$V_{nD,out} = \frac{I_{nD}}{\frac{1}{R_L} + \frac{1}{R_S + R_F} + \frac{g_m R_S}{R_S + R_F}} \approx I_{nD} R_L$$

So,

$$\overline{V^2_{nD,out}} = \overline{I^2_{nD}} R_L^2 \quad \text{-----(3)}$$

Combaining (1) (2) & (3)

$$NF = 1 + \frac{\overline{V^2}_{n,RF} \left[\frac{R_L}{R_F} (1 + g_m R_S) \right]^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}} + \frac{\overline{I^2}_{nD} R_L^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}}$$

$$A_{v,tot} = -g_m R_L, \quad \overline{V^2}_{n,RS} = 4KTR_S \Delta f, \quad \overline{V^2}_{M,RF} = 4KTR_F \quad \& \quad \overline{I^2}_{nD} = 4KT\gamma g_m$$

$$NF = 1 + \frac{R_S}{R_F} \left(1 + \frac{1}{g_m R_S} \right)^2 + \frac{\gamma}{g_m R_S}$$

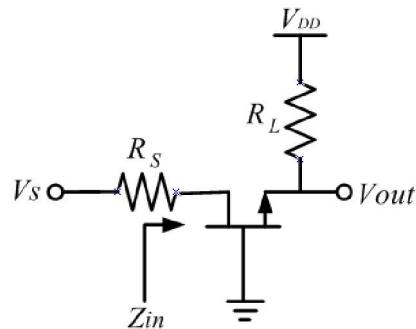
b).

$NF \downarrow$ $g_m R_S \uparrow$ & $R_F \uparrow$ usually $R_S = 50\Omega$

- Better performance than CS amplifier
- R_F induces noise
- At $f \uparrow$ need shunt inductor to tune out C_{gs}
- Broadband Amp @ Lower frequency
- To make $NF \downarrow$ $R_F > R_S$ $g_m R_S \gg 1$

Problem-2.5 (HW)

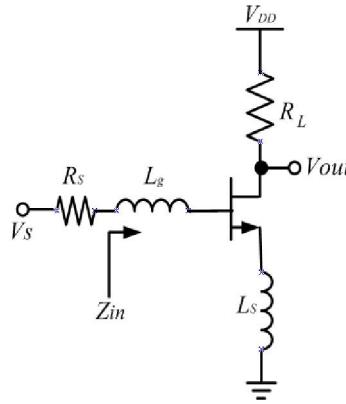
Common gate amplifier also offers 50Ω input impedance match and solves the input matching problem.



- c) Calculate the gain, input impedance and NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- a) What are the disadvantage of common gate amplifier with reference to gain and NF?

Problem-2.6 (Tutorial)

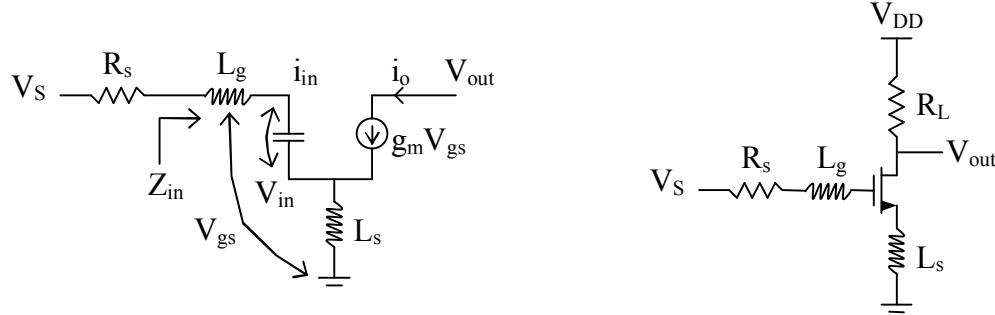
The disadvantages of three types of amplifiers in Problem-3, 4 & 5 can be circumvented by using source degenerated LNA shown below.



- Calculate the input impedance. This inductor source degenerated amplifier presents a noiseless resistance for 50Ω for input power match. How we can cancel the imaginary part of complex input impedance so that the LNA presents 50Ω real input resistance at input port.
- Calculate the NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- C_{gd} bridges the input and output ports. The reverse isolation of this LNA is very poor. Why reverse isolation is important? Suggest the modification to improve reverse isolation.

Solution:

a).



(Biasing not shown)

From model above we can write

$$V_{in} = i_{in} \left(j\omega L_g + j\omega L_s \right) + i_{in} \left(\frac{1}{j\omega C_{gs}} \right) + i_o j\omega L_s \quad \text{-----(1)}$$

$$i_o = g_m V_{gs} = g_m i_{in} \times \frac{1}{j\omega C_{gs}} \quad \text{-----(2)}$$

Substituting (2) in (1)

$$V_{in} = i_{in} \left[j\omega (L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}} \right]$$

$$Z_{in} = \frac{V_{in}}{I_{in}} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

For matching $L_g + L_s$ are canceled out by C_{gs} . So at frequency of interest

$$\omega_o(L_g + L_s) = \frac{1}{\omega_o C_{gs}} \Rightarrow \omega_o^2 = \frac{1}{(L_g + L_s)C_{gs}}$$

$$\text{And } R_S = 50\Omega = \frac{g_m}{C_{gs}} L_s$$

Notes:

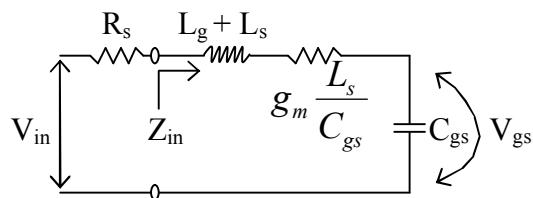
- a). L_s is typically small and may be realized by bond wire.
- b). L_g can be implemented by spiral/external inductor.

b).

From part a)

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

We can draw this circuit as



Here

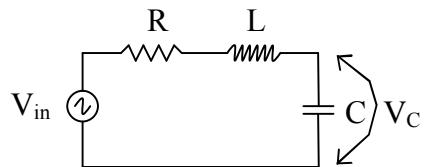
$$Q_{in} = \frac{\omega_o(L_g + L_s)}{R_S + \frac{g_m L_s}{C_{gs}}} = \frac{\omega_o(L_g + L_s)}{R_S + W_T L_s}$$

$$\therefore \omega_T \cong \frac{g_m}{C_{gs}}$$

$$Q_{in} = \frac{1}{\omega_o \left(R_S + \frac{g_m L_s}{C_{gs}} \right) C_{gs}} \quad \text{For match load } R_S = \frac{g_m L_s}{C_{gs}}$$

Reference:

For series RLC Circuit



$$Q_s = \frac{1}{R} \sqrt{\frac{L}{C}} = \frac{\omega_o L}{R} = \frac{1}{\omega_o R C}$$

$$\text{And } V_C = Q_s V_{in}$$

For problem (1)

$$\omega_T \cong \frac{g_m}{C_{gs}} \quad \text{Unity gain frequency}$$

for current

$$Q_{in} = \frac{1}{2\omega_o R_s C_{gs}}$$

Gain

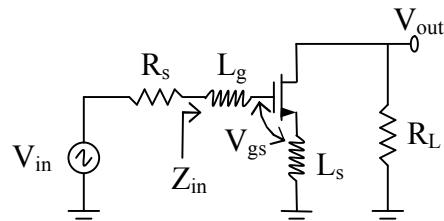
$$V_{gs} = Q_{in} V_{in}$$

$$g_m = \frac{V_{out}}{V_{gs}}$$

$$G_m = \frac{V_{out}}{V_{in}} = \frac{V_{gs} g_m}{V_{in}} = Q_{in} g_m$$

$$G_m = Q g_m$$

So, Gain = $-G_m R_L$ or $-Q_{in} g_m R_L$ & $G_m = Q_{in} g_m$

**Noise Figure:**

$$NF = \frac{\text{Total noise power at output}}{\text{noise power at output due to input source}}$$

For this calculation we ignore channel noise.

$$F = \frac{\overline{V^2}_{nRS,OUT} + \overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}} = 1 + \frac{\overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}}$$

$$\overline{V^2}_{nD,OUT} = \overline{i^2}_{n,D} R^2_L \quad \overline{i^2}_{n,D} = 4KT\gamma g_m \Delta f$$

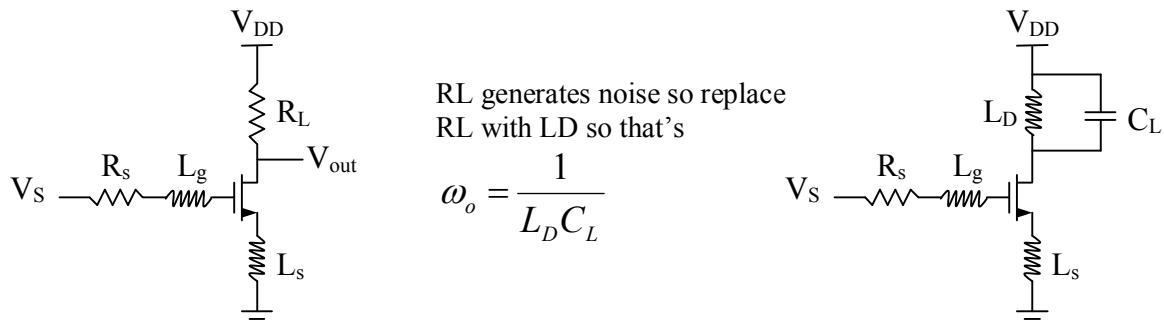
$$\overline{V^2}_{nRS,OUT} = \overline{V^2}_{n,RS} G^2 m R^2_L \quad \overline{V^2}_{n,RS} = 4KTR_S \Delta f \quad \& \quad G_m = Q_{in} g_m$$

$$F = 1 + \frac{\overline{i^2}_{n,D} R^2_L}{\overline{V^2}_{n,RS} Q^2_{in} g^2_m R^2_L} \quad \overline{i^2}_{n,D} = 4KT\gamma g_m, \quad \overline{V^2}_{n,RS} = 4KTR_S$$

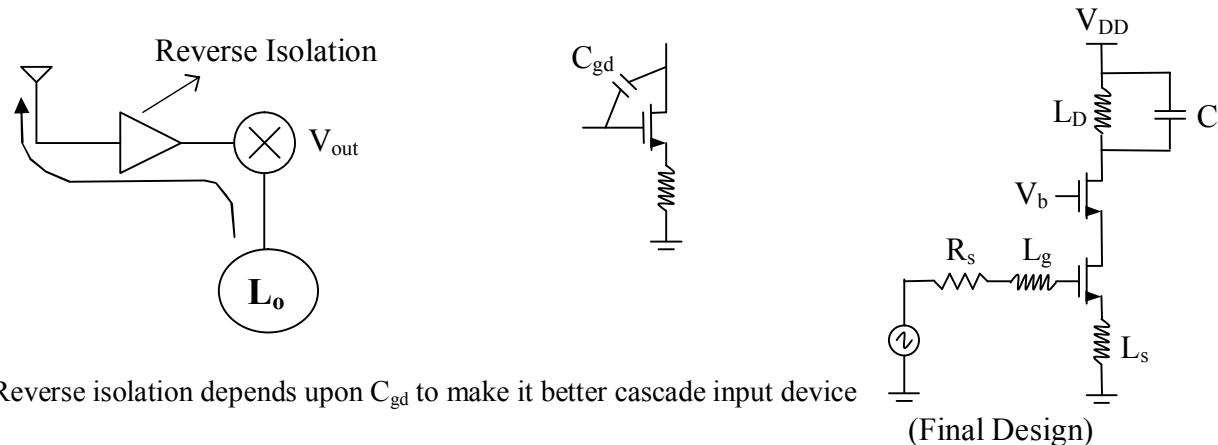
$$F = 1 + \frac{\gamma}{g_m R_s Q^2_{in}}$$

Notes:

- Very good NF value
- Narrow band matching
- NF \downarrow with Q^2
- The Q is dependent upon $L_g + L_s$, L_s is small so Q depend upon L_g

C). Draw Backs**i).**

The C_L can be the input capacitance of mixer or filter.

ii).**Problem-2.7 (HW)**

Fill-in the Table below, use the data from Problem-2.4, 2.5, 2.6 and 2.7

Type of LNA	Z_{in}	Noise Factor	Gain	NF (dB)
Shunt Resistor	R_{sh}	$2 + \frac{4\gamma}{g_m R_s}$	$\frac{-g_m R_L}{2}$	
Common Gate				
Shunt Feedback				
Source Degenerated				

- Calculate the NF for all above amplifiers. Assume $\gamma=2$, $g_m = 20mS$, $R_s = 50\Omega$, $R_f = 500\Omega$, and $Q_{in} = 2$.
- Which is best topology for Narrow Band LNA design at high frequency?

Problem-2.8 (Tutorial)

Real Design: We will design the inductor source degenerated LNA shown in Fig below to meet the specification outlined for IEEE802.11 (b) standard. The first cut approximate values are calculated as a starting point for simulation. In **LAB3: Design of LNA** you will take the same design and modify these component values to meet the specification.

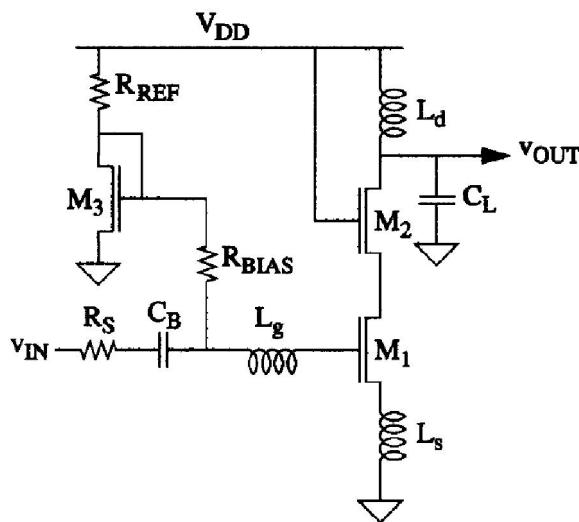
LNA Specification:

NF < 2.5 db, Gain > 15dB, IP3 > -5dBm, Centre Frequency = 2.4 GHz

S11 < -20dB, S22 < -10dB, Load Capacitance = 1pF

Technology Parameters for 0.35um CMOS:

$$L_{\text{eff}} = 0.35 \mu m, \quad \mu_n C_{\text{ox}} = 170 \mu A/V^2, \quad C_{\text{ox}} = 4.6 mF/m^2, \quad \mu_p C_{\text{ox}} = 58 \mu A/V^2, \quad \gamma = 2$$

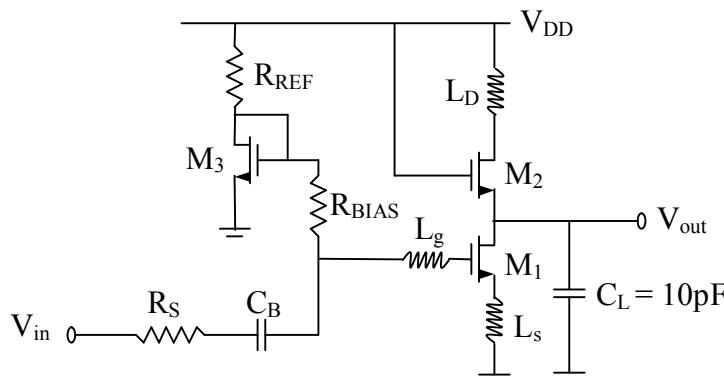


Solution:

$$\text{Technology: } \left\{ \begin{array}{l} \mu_o C_{\text{ox}} = 170 \mu A/V^2, \mu_p C_{\text{ox}} = 58 \mu A/V^2, 0.35 \mu m \\ \text{CMOS}, C_{\text{ox}} = 4.6 mF/m^2, \gamma = 2, f_o = 2.4 GHz \\ (\text{IEE802.11(b) Standard}), \mu_o C_{\text{ox}} = 170 \mu A/V^2 \end{array} \right\}$$

Design Parameters

$$R_S = 50 \Omega, V_{DD} = 3.3V, C_L = 10 pF, NF < 2.5 dB$$



Component Description

L_s – Matches input impedance

L_g – Set the Resonant Frequency $f_0 = 2.4$ GHz

M_3 – Biasing transistor which forms current mirror with M_1

L_d – Tuned output increases the gain and also work as band pass filter with C_L

M_2 – Isolate tuned input and tuned output increases reverse isolation, also reduces the effect of Miller capacitance C_{gd}

C_B – BC Blocking capacitor chosen to have negligible reactance at $f_0 = 2.4$ GHz

R_{BIAS} – Large enough so that its equivalent current noise is small enough to be ignored. (Don't consider it as voltage noise source. Why??)

Design Procedure

Size of M_1 :

We will not go for global minimum noise figure as given by two-point noise theory (See lecture on LNA Slide # 10)

$$G_{opt} = \alpha \omega C_{gs} \sqrt{\frac{\delta}{5\gamma} (1 - |C|^2)} = \frac{1}{50\Omega} \quad \text{-----(1)}$$

$$C_{gs} \approx 4 \text{ pF} \Rightarrow W_{M1} \approx 4 \text{ mm}!! \text{ (not possible)}$$

Solution:

A & B are from Thomas. H. Lee book (LNA Chapter)

LNA NF will be optimized for given Power

* It will not be best NF globally.

$$W_{opt} = \frac{1}{3\omega_o L_{eff} C_{ox} R_s}$$

$$F_{min,p} = 1 + 2.4 \frac{\gamma}{\alpha} \frac{\omega}{\omega_T} \Rightarrow F_{min,p} = 1 + 5.6 \frac{\omega}{\omega_T} \quad \text{-----(A)}$$

From (1) we can derive

$$F_{min,p} = 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_T} \sqrt{\gamma C (1 - |C|^2)} = 1 + 2.3 \frac{\omega}{\omega_T}$$

$$F_{min,p} = 1 + 2.3 \frac{\omega}{\omega_T} \quad \text{-----(B)}$$

(a) is minimum NF for a given power consumption.

(b) is global minimum noise figure.

The difference is usually 0.5dB to 1dB (no big deal for Lower Power)

Step - 1:

$$I_1 = I_2 = 5mA \text{ (Low Power consumption)}$$

Step - 2:

$$W_{M1} = \frac{1}{3WL_{eff}C_{ox}R_S}$$

$$W_{M1} = \frac{1}{3 \times 0.35\mu \times 4.6m \times 50 \times \omega_o} \quad \left. \begin{array}{l} R_S = 50\Omega, C_{ox} = 4.6mF/m^2, \\ \mu_n C_{ox} = 170\mu A/V, L_{eff} = 0.35\mu m, \\ \omega_o = 2\pi f_o, f_o = 2.4GHz \end{array} \right\}$$

$$W_{M1} = 3.9 \times 10^{-4}$$

$$W_{M1} = 3.9 \times 10^{-4} = 390\mu m$$

Step - 3:

$$C_{gs1} = \frac{2}{3} W_{M1} L_{eff} C_{ox}$$

$$C_{gs1} = \frac{2}{3} \times 390\mu \times 0.35\mu \times 4.6m = 0.41pF$$

$$g_{m1} = \sqrt{2\mu_n C_{ox} \left(\frac{W}{L} \right)_{M1} I_{DM1}}$$

$$g_{m1} = \sqrt{2 \times 170\mu \times \left(\frac{390}{0.35} \right) \times 5m} = 43mA/V$$

$$\omega_T \approx \frac{g_{m1}}{C_{gs1}} = \frac{43mA/V}{0.41pF} = 104Grad/Sec$$

Assuming $\gamma = 2$

$$\text{Now } F_{\min} = 1 + 5.6 \frac{\omega_o}{\omega_T}$$

$$F_{\min} = 1 + 5.6 \frac{2\pi 2.4G}{104G} \approx 2.55dB$$

$$NF \approx 2.55dB$$

It's very close to what we derive, if the value is higher we can increase I_D to increase ω_T and hence low NF on expense of power.

Step - 4:

Source and gate inductance such that they cancel C_{gs} and set 50Ω input impedance

$$\omega_o = 2\pi f_o = 2\pi 2.4 = 15 \text{Grad}/\text{Sec}$$

From previous problem

$$R_s = R_{Transformed} = g_m \frac{L_s}{C_{gs}} = L_s \omega_T$$

$$L_s = \frac{R_s}{\omega_T} = \frac{50}{100G} \cong 0.5nH$$

$L_s = 0.5nH$ can be implemented using Band wire.

$$\text{Now } L_g + L_s = \frac{1}{(\omega_o^2 C_{gs})}$$

$$L_g + L_s = \frac{1}{(15G)^2 \times 0.41pF} = 10.81nH$$

$$L_g \approx 10nH$$

Step - 5:

$$L_d = \frac{1}{\omega_o^2 C_L} \quad \because C_L = 1pF$$

$$L_d = \frac{1}{(15G)^2 \times 1pF} \cong 4.4nH$$

$$L_d = 4.4nH$$

Step - 6:

Size of M3 is chosen to minimize power consumption

$$W_{M3} = 70\mu m, \quad R_{REF} = 2K\Omega \Rightarrow I_3 = 0.6mA$$

$R_{BIAS} = 2K\Omega$ (Large enough so that it's equivalent current noise can be neglected)

$$C_B = 10pF \quad (X_C \approx 6.6\Omega \text{ so good value @ 2.4G} \quad X_B = \frac{1}{2\pi f_o C_B} = 6.6\Omega)$$

Step - 7:

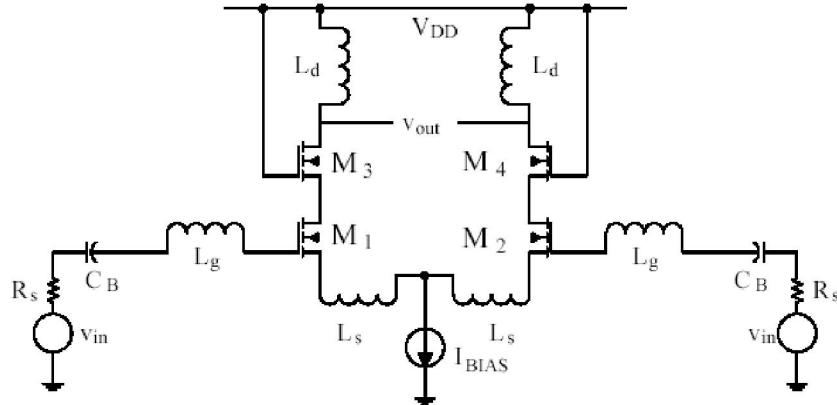
Size M2 = M3

So that they can have shared Drain Area..

(Note: We will simulate same design in LAB # 2)

Problem-2.9 (Point to Ponder):

Connecting two Inductor source degenerated LNA as shown in Figure make the differential LNA. Differential LNA has many advantages: higher common mode rejection ratio, less sensitivity to the ground inductance variation L_s compared to single ended counterpart..



- Compare intuitively the NF of single ended and differential if both have same power consumption.
- If low power is not parameter of interest, which LNA has lower NF?

Instructions:

For hand calculation of NF you can ignore the gate noise of the device and noise generated by the load resistance R_L .

Acknowledgement: The major part of this tutorial was developed, while author was employed by Linkoping University, Sweden.

LNA Design

Iulian Rosu, YO3DAC / VA3IUL
<http://www.qsl.net/va3iul/>

An LNA combines a low noise figure, reasonable gain, and stability without oscillation over entire useful frequency range.

The Low Noise Amplifier (LNA) always operates in Class A, typically at 15-20% of its maximum useful current. Class A is characterized by a bias point more or less at the center of maximum current and voltage capability of the device used, and by RF current and voltages that are sufficiently small relative to the bias point that the bias point does not shift.

The smallest signal that can be received by a receiver defines the receiver sensitivity.

The largest signal can be received by a receiver establishes the upper power level limit of what can be handled by the system while preserving voice or data quality. The dynamic range of the receiver, the difference between the largest possible received signal and the smallest possible received signal, defines the quality of the receiver chain. The LNA function, play an important role in the receiver designs. Its main function is to amplify extremely low signals without adding noise, thus preserving the required Signal-to-Noise Ratio (SNR) of the system at extremely low power levels. Additionally, for large signal levels, the LNA amplifies the received signal without introducing any distortions, which eliminates channel interference.

- An LNA design presents a considerable challenge because of its simultaneous requirement for high gain, low noise figure, good input and output matching and unconditional stability at the lowest possible current draw from the amplifier.
- Although Gain, Noise Figure, Stability, Linearity and input and output match are all equally important, they are interdependent and do not always work in each other's favor.
- Carefully selecting a transistor and understanding parameter trade-offs can meet most of these conditions.
- Low noise figure and good input match is really simultaneously obtained without using feedback arrangements.
- Unconditional stability will always require a certain gain reduction because of either shunt or series resistive loading of the collector. High IP3 requires higher current draw, although the lowest possible noise figure is usually achieved at lower current levels.
- Envelope termination technique can be used to improve IP3 performance while operating LNA at low current levels.
- Additional improvement of IP3 can also be achieved by proper power output matching (1dB compression point match or P1dB match). The P1dB match, being different from conjugate match, reduces the gain although improving IP3 performance.
- Transistor selection is the first and most important step in an LNA design. The designer should carefully review the transistor selection, keeping the most important LNA design trade-offs in mind.

The transistor should exhibit high gain, have a low noise figure, and offer high IP3 performance at the lowest possible current consumption, while preserving relatively easy matching at frequency of operation.

Examination of a data sheet is a good starting point in a transistor evaluation for LNA design.

The transistor's S-parameters should be published at different collector/emitter voltages and different current levels for frequencies ranging from low to high values. The data sheet should also contain noise parameters, which are essential for low noise design. Spice models for the transistor and its package are also useful for IP3 and P1dB simulations.

The designer should first look at the main design parameters as: Noise, Gain, and IP3, and decide what Vce and Ic levels will produce optimal performance.

The forward transducer power gain represents the gain from transistor itself with its input and output presented with $50\ \Omega$ impedance.

The manufacturer of the transistor at multiple frequencies and different Vce and current levels provides the S21 values.

Additional gain can be obtained from source and load matching circuits.

Maximum Stable Gain and Maximum Power Gain (G_{max}) are good indicators of additional obtainable gain from the LNA circuit.

LNA linearity is another important parameter. A figure of merit for linearity is IP3. A two-tone test is used for derivation of IP3.

As a rule of thumb for bipolar junction transistors (BJT), the Output-IP3 can be estimated from the following formula:

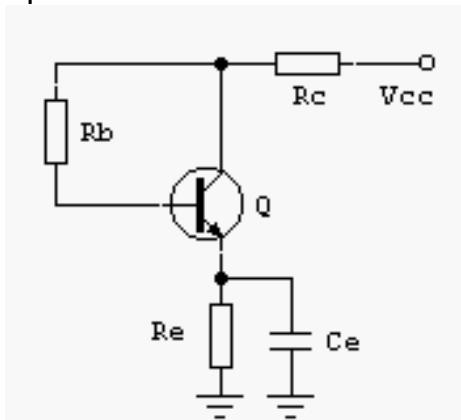
$$OIP3_{[\text{dBm}]} = 10 \log (V_{ce} [\text{volt}] * I_c [\text{mA}] * 5)$$

- RF performance of the LNA depends by many variables as:
 - Frequency
 - DC Biasing and Power Dissipation
 - Stability
 - Input and Output Matching
 - Layout and Grounding
 - EM Shielding
 - Supply decoupling
 - Temperature

1. DC Biasing (BJT)

represent the first step in LNA design.

- The chosen DC bias circuit should exhibit stable thermal performance and reduce the influence of h_{FE} spread.
- The resistive feedback arrangement is the simplest form of DC biasing that fulfills all the major requirements.
- Two bias feedback arrangements are possible: one with a combination of R_c and R_b and a second one with simple R_e and C_e combination.



The operation of the R_c and R_b is simple: R_c and R_b will establish a biasing point. If the device current increases, the voltage drop across R_c increases, reducing the voltage seen by the base, thereby providing feedback. Because the operation class of the LNA is going to be Class-A (constant current draw for dynamic range of power levels), a stable biasing point over different temperatures is required. For different lot of transistors small variation in h_{FE} can be expected. For R_b to have little influence on source matching, which is crucial for noise performance, the feedback network should be decoupled with an inductor (making biasing invisible at RF band of operation).

Another possible bias feedback can be realized with emitter resistor and capacitor. C_e should be selected carefully, because R_e will also have a direct effect on RF gain of LNA. C_e should present a short at frequency of operation to limit its influence on gain and noise performance of the circuit.

Other biasing methods are suitable for Class-A networks. These are usually closed feedback arrangements with dynamic bias control provided by active components. Although suitable for LNA application, these active feedback bias networks increase complexity of the LNA network, introduce additional components and increase the real-estate area of the solution.

2. Stability Design

should be the next step in LNA design.

- Unconditional stability of the circuit is the goal of the LNA designer.
- Unconditional stability means that with any load present to the input or output of the device, the circuit will not become unstable – will not oscillate. Instabilities are primarily caused by three phenomena: internal feedback of the transistor, external feedback around the transistor caused by external circuit, or excess gain at frequencies outside of the band of operation.
- S-parameters provided by manufacturer of the transistor will aid in stability analysis: numerical and graphical.
- Numerical analysis consists of calculating a term called Rollett Stability Factor (K-factor).
- When K-factor is greater than unity, the circuit will be unconditionally stable for any combinations of source and load impedance.
- When K-factor is less than unity, the circuit is potentially unstable and oscillation may occur with a certain combination of source and /or load impedance present to the transistor.

The K-factor represents a quick check for stability at given biasing condition. A sweep of the K-factor over frequency for a given biasing point should be performed to ensure unconditional stability outside of the band of operation.

The designer's goal is to design an LNA circuit that is unconditionally stable for the complete range of frequencies where the device has a substantial gain.

An LNA designer can use at least five methods for circuit stabilization.

- The first one consists of resistive loading of the input. This method, although capable of improving the stability of the circuit, also degrades the noise of the LNA and is almost never used.
- Output resistive loading is preferred method of circuit stabilization. This method should be carefully used because it effects are lower gain and lower P1dB point (thus IP3 point).
- The third method uses collector to base resistor-inductor-capacitor (RLC) feedback to lower the gain at the lower frequencies and hence improve the stability of the circuit.

- The fourth method consists of filter matching, usually used at the output of the transistor, to decrease the gain at a specific narrow bandwidth frequency. This method is frequently used for eliminating gain at high frequencies, much above the band of operation. Short circuit quarter wave lines designed for problematic frequencies, or simple capacitors with the same resonant frequency as the frequency of oscillation (or excessive gain) can be used to stabilize the circuit.
- The final stabilization method can be realized with a simple emitter feedback inductor. A small inductor can make the circuit more stable at higher frequencies. But if the source inductance is increased, the K-factor at higher frequencies eventually falls below 1. This effect limits the amount of source inductance that can safely be used.
- To get the best LNA stability performances have to accommodate the full range of expected variations in operating parameters as:
 - Component package parasitics
 - Component values
 - Temperature
 - Supply voltage
- Most common causes for LNA instability are:
 - Insufficient RF decoupling between supply lines of the amplifier bias.
 - Parasitic inductance in GND connections.
 - Excess in-band and/or out-of-band Gain.
 - Electro-Magnetic coupling and Feedback.
- Always check stability of your LNA well beyond band-of-interest checking for both, small-signal stability and for large-signal stability.
- Use stability circles on Smith Chart (for both, source and load) to verify legitimacy of chosen Z_{in} and Z_{out} impedances.

3. Noise Matching and Input Return Loss (IRL)

The next step in LNA design consists of Noise Match and Input Return Loss (IRL).

- IRL defines how well the circuit is matched to 50Ω matching of the source.
- A typical approach in LNA design is to develop an input matching circuit that terminates the transistor with conjugate of Gamma optimum (Γ_{opt}), which represents the terminating impedance of the transistor for the best noise match.

In many cases, this means that the input return loss of the LNA will be sacrificed.

The optimal IRL can be achieved only when the input-matching network terminates the device with a conjugate of S_{11} , which in many cases is different from the conjugate of Γ_{opt} .

- In 1928 [H. Nyquist](#) showed that the noise from any impedance is determined by its resistive component. Consequently, if an ideal lossless element is used to provide feedback, than the minimum noise measure is unaffected.
- An emitter (or source) inductor feedback can rotate S_{11} closer to Γ_{opt} , which can help obtaining close to minimum Noise Figure and respectable IRL simultaneously.

The additional series inductance provides lossless negative series feedback and also reduce the overall available gain of the network and can be used in balancing trade-offs between the gain, IIP3 and stability in LNA design. Have to mention that this inductive degeneration does not seriously impact Noise Figure performance, as resistive degeneration does.

At high frequencies this inductance will be achieved with small strip lines (stubs) connected directly to the emitters of the transistor.

The inductive reactance of the stubs is usually no greater than $10\ \Omega$ and the line lengths are typically $\sim 2\text{mm}$ or less with characteristic impedances $50\ \Omega$ or greater.

- To design an LNA for minimum Noise Figure, determine (experimentally or from the data sheet) the source resistance and bias point that produce the minimum Noise Figure for that device.
- Then force the actual source impedance to “look like” that optimum value with all stability considerations still applying. If the Rollet stability factor (K) is calculated to be less than 1 (K is defined as a figure of merit for LNA stability), then you must be careful in choosing the source and load-reflection coefficients.
- A typical method used in designing input matching network is to display noise circles and gain/loss circles of the input network on the same Smith chart. This provides a visual tool in establishing an input matching network for the best Input Return Loss and noise trade off.

Using Noise Figure from Datasheets

Generally for microwave transistors following a datasheet the minimum Noise Figure (F_{\min}) at higher frequencies is based on measurements, while the F_{\min} s at lower frequencies are extrapolated.

- F_{\min} represents the true minimum Noise Figure of the device when the device is presented with an impedance matching network that transforms the source impedance, typically 50Ω , to an impedance represented by the reflection coefficient Γ_{opt} .
- The designer must develop a matching network that will present Γ_{opt} to the device with minimal associated circuit losses. To accomplish this have to minimize the number of components needed on the LNA input.
- The Noise Figure of the completed amplifier is equal to the Noise Figure of the device plus the losses of the matching network preceding the device.
- The Noise Figure of the device is equal to F_{\min} only when the device is presented with Γ_{opt} .
- If the reflection coefficient of the matching network is other than Γ_{opt} , then the Noise Figure of the device will be greater than F_{\min} .
- The losses of the matching networks are non-zero and they will also add to the noise figure of the device creating a higher amplifier noise figure. The losses of the matching networks are related to the Q of the components and associated printed circuit board loss.
- Γ_{opt} is typically fairly low at higher frequencies and increases as frequency is lowered.
- For FET devices larger gate width devices will typically have a lower Γ_{opt} as compared to narrower gate width devices.
- Typically for FETs, the higher Γ_{opt} usually infers that an impedance much higher than 50Ω is required for the device to produce F_{\min} . At VHF frequencies and even lower L Band frequencies, the required impedance can be in the vicinity of several thousand ohms. Matching to such high impedance requires very hi-Q components in order to minimize circuit losses.

4. Output Matching

The last step in LNA design involves output matching of the transistor.

An additional resistor, either in series or parallel, has been placed on the collector of the transistor for circuit stabilization.

Conjugate matching has been exclusively used for narrow band LNA design to maximize the gain out of the circuit.

With additional IP3 requirement forced on the LNA, the trade-off between IP3 and gain must be considered.

Linearity matching is widely known by high-power amplifier designers. The so-called load pulling is used to establish IP3 and gain impedance contours. The load pulling can be realized by using the non-linear Spice model of the transistor with simulation software.

Harmonic balance can be used for establishing two-tone environment.

The load pulling method sweeps impedance of the whole Smith chart and plots contours of the constant gain and IP3 numbers. The optimal gain impedance does not match the optimal IP3 point, which means that the design will have to be realized by means of a trade off.

Typically, the designer should design the LNA circuit at the point where the gain does not degrade as much, and the IP3 is still respectable. If one were to draw a line between the optimal gain and IP3 impedance points, every point on that straight line will represent a good area of trade-off, with the ends representing the two optimal points.

The rule of thumb for 1dB gain compression point (P_{1dB}) and IP3 is:

$$IP3 = P_{1dB} + 10 \text{ [dBm]}$$

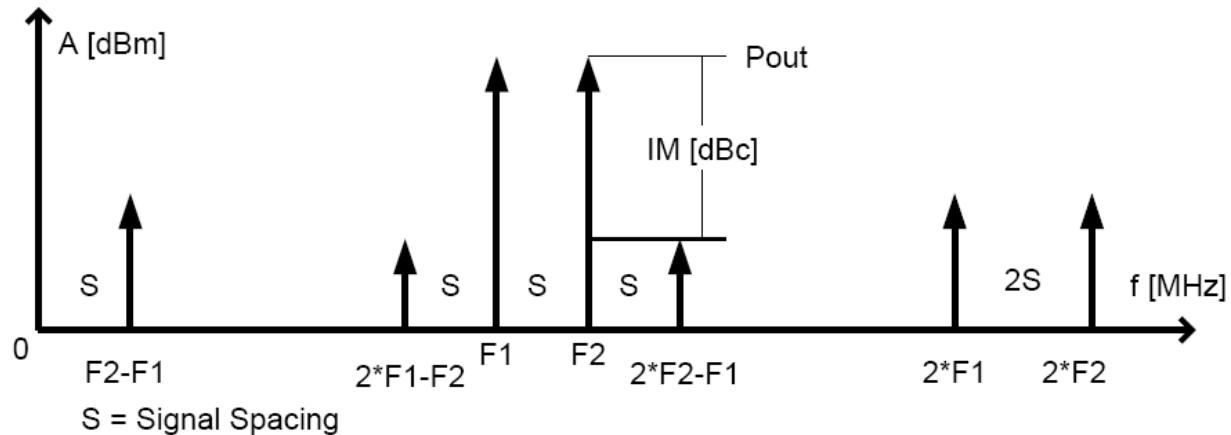
That means that by knowing the gain compression point (P_{1dB}), can estimates the IP3 levels.

- The 10dB rule can further be improved with appropriate bypassing of the base and the collector.

As previously indicated, the IIP3 is established by injecting two equal-in-magnitude signals with small frequency offset (S) into an active circuit.

As the active circuit approaches non-linear region, close to P_{1dB} , the two carriers will generate distortion products, both in and out of band.

In example below we have two signals, with output levels P_{out} and frequencies F_1 and F_2 .

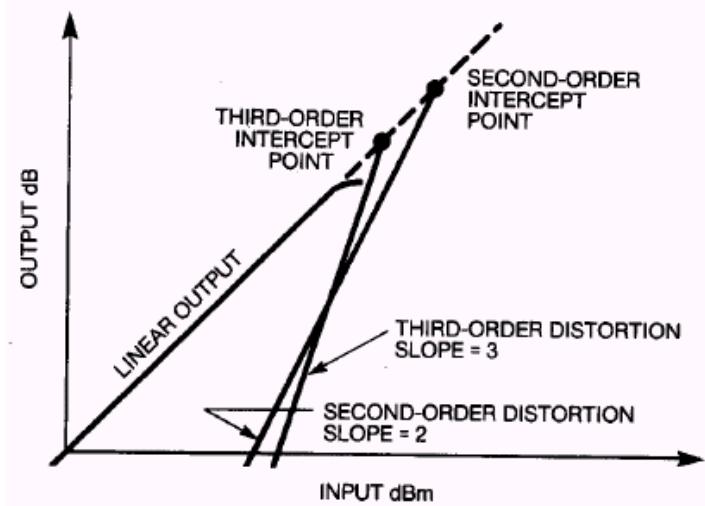


$$OIP3[\text{dBm}] = P_{out}[\text{dBm}] + IM[\text{dBc}] / 2$$

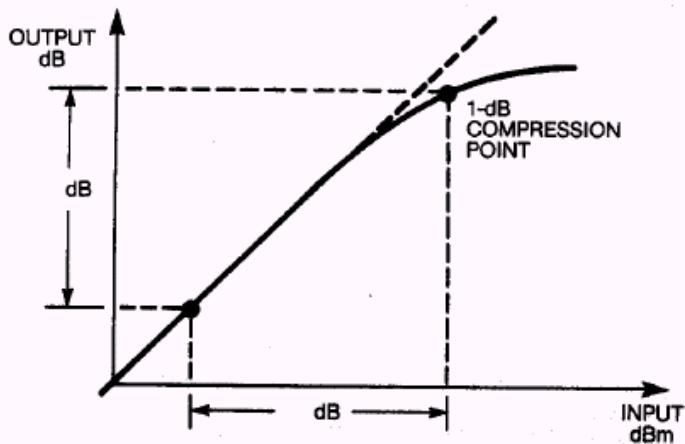
Where IM [dBc] is the difference in amplitude between one of the two equal amplitude test tones, present at the amplifier output, and the level of the highest 3rd-order distortion product.

- For every dB increase in input power, the third order products (IM3) will increase 3dB.
- For every dB increase in input power, the second order products (IM2) will increase 2dB.

Plotting third order products versus input power predicts a 3:1 response which intersects the 1:1 response at the third order intercept point.



Second and Third-Order Distortion Slopes



1dB Gain Compression Point (P1dB)

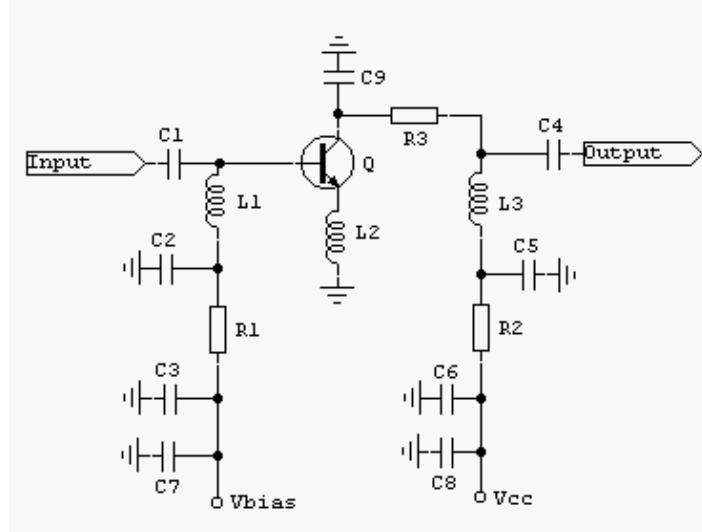
The relation between Input-IP3 (IIP3) and Output-IP3 (OIP3) is defined as:

$$\text{IIP3[dBm]} = \text{OIP3[dBm]} - \text{Gain[dB]}$$

The low frequency IM2 products (F2-F1), can modulate the base-emitter and collector-emitter LNA supply voltages. To improve the linearity the fluctuation of the base and the collector shall be stabilized with low impedance at so called video frequencies or baseband frequencies (between DC and usually up to 40MHz).

- The designer should exhibit caution during bypassing design. A poor selection of the by-pass capacitors could also degrade IP3 performance.
- As a rule of thumb, the impedance of bypassing circuit should be lower than 25% of the input impedance of the transistor at particular frequency spacing.
- Although preserving the gain performance of the LNA, the bypassing method (also known as an envelope termination technique) can improve LNA's IIP3 performance without increasing current consumption.

5. LNA components and the effect on IP3



- Any mismatch due to noise matching C1/L1 improves Input-IP3.
- Increasing L2 reduces gain and improves Input-IP3, but watch for microwave oscillation with excessive inductance.
- C9 can be used to improve IP3- provides gain roll-off at 2^*F1 or 2^*F2 .
- Printed circuit board losses R3 provide Q1 stability while reducing IP3. R3 less than 27 Ω for about a dB reduction in IP3.
- C3 and C6 provide a HF/VHF termination for Q1. Depending on spacing of signals used to test IP3, values may not be large enough – may necessitate additional low frequency bypassing in the form of C7 and C8. Typical values are 0.01 to 0.1 uF.
- The combination of C2/C5, C3/C7 and C6/C8 must provide low impedance at F2 – F1. May have to add resistance between caps to decrease Q.
- C7 and C8 also used to minimize power supply noise from modulating the DC.
- Capacitor C2, C3 and C7 performs the low-frequency bypass function and an improvement in IP3 of approximately 5 to 10 dB can be expected by using this method. Using extra charge storage on the drain may see the same effect, but the results are not nearly as dramatic.
- The closer together the two input test tones F1 and F2 are in frequency, the lower frequency the product or beat tone ($F_2 - F_1$) is. Therefore, as input test tones F1 and F2 come closer together, more capacitance is needed to achieve best possible bypassing of the low frequency product ($F_2 - F_1$).

For a test tone separation of 1MHz, 0.1 uF was found to be more adequate for this application. For best results, the transistor should see a low impedance path at low frequencies between this additional bypass caps and its terminals. For this reason, a coil rather than a high value resistor is used to bring the gate bias voltage and isolate the RF from the DC bias network. For example a value of 15nH for L1, has negligible impedance up to tens of MHz, but provides enough impedance at 2 GHz to nearly isolate the gate of the transistor from the bias network within LNA's normal operating frequency range.

It is important to note that bypassing the $F_2 - F_1$ product as described here does not affect the compression point of the amplifier, but only the IP3 (3rd-order intercept point).

As a result, if this bypassing is used, the general rule of thumb stating that there is approximately a 10 dB difference between IP3 and 1 dB gain compression point (P_{1dB}) is no longer valid.

6. Real issues in LNA design

- An LNA is a design that minimizes the Noise Figure of the system by matching the device to its noise matching impedance, or Gamma optimum (Γ_{opt}).
- Gamma optimum (Γ_{opt}) occurs at impedance where the noise of the device is terminated.
- All devices exhibit noise energy. To minimize this noise as seen from the output port, one must match the input load to the conjugate noise impedance of the device. Otherwise the noise will be reflected back from the load to the device and amplified. While this gives a minimum noise figure, it often results in slightly reduced gain as well as possibility increasing the potential instabilities.

Noise match often comes close to S_{11} conjugate (S_{11}^*) under non-feedback conditions. As a result, the input impedance to the amplifier will not be matched to 50 ohms. Γ_{opt} , as presented in data sheets, is the actual measured load at which the minimum noise figure is found.

- Noise Figure for BJT LNA increases more rapidly than FET LNA, as Collector/Drain current is increased.
- FETs allows for better trade-off between high-linearity and low Noise Figure than BJTs.

A further complication on LNA design is that the input load of the amplifier is usually less than ideal. It is either connected to an antenna, which can change its impedance with changing the environment, or to a filter, which by very physics of a reflective network will have very bad match out of band. These mismatches could cause the device to become unstable out of band and some cases in band. As the gain of the device increases, the difficulties in yielding a stable design become increasingly more challenging.

To avoid overloading the LNA, an input filter is commonly used. Since the device is not matched to S_{11}^* , the input of the LNA will not be 50 ohms. This can cause distortions in the pass band of the filter when connected to the input of the LNA, as filter are intended to operate in their characteristic impedance, typically 50 ohms.

Printed inductors or transmission lines are free as compared to SMT inductors, which typically cost 10 to 25 times as much as resistors or capacitors in volume. Printing an inductor is easy and results in highly repeatable results. Printed inductors usually exhibit poor Q due to the lossy dielectric, and, if a ground plane exists, they are no more than a high impedance transmission line.

As shown a transmission line can replace an inductor to some degree, but inductors and high impedance transmission lines have a different trajectory on the Smith Chart. High impedance transmission line can be made to look more like printed inductors in cases where the backside of the PCB is suspended away from a grounded chassis. This is accomplished by removing the backside ground plane of the PCB directly under the printed inductor.

In this case beware of digital noise coupling into the input of the LNA from circuitry on the opposite side.

- The next concern is what load impedance to match. Remember matching to the conjugate of S_{22}^* is only valid if the input is conjugate matched. Since S_{12} is non-zero, whatever load is present to the input will cause the output load change.
- Another issue is stability, especially if a filter is going to be used at the input. The output port can potentially give difficulties since the input is very restricted by its match.

The designer must replace the ideal sources in the bias circuit and ideal values in the matching circuit with equivalent real components. This often presents the designer with a new

set of problems. First, the bias network must be robust enough to function properly over a range of power-supply voltages and temperatures. This introduces additional complexity into the bias network. The real components in the bias network the resistors and large capacitors operate at DC voltages, so frequency effects are not a problem. The matching network, however, contains real capacitors and inductors that operate at RF frequencies.

Real components differ from ideal ones in several respects. First, real components have a price associated with them. There is a trade-off between price and performance of these parts. The competitiveness of today's markets often forces designers to use inexpensive components in their designs.

- Real discrete components have a finite resistance called Equivalent Series Resistance (ESR). The ESR introduces losses that result in lower gain and noise figure. Although typically only a few tenths of an ohm in value, ESR will affect the matching networks.
- Discrete components also have a Q value, measured at a particular frequency that can contribute to unwanted resonance.
- High-Q networks are sensitive to variations in process, voltage, temperature, and component value.
- A component's Series Resonant Frequency (SRF) is the frequency where it will behave erratically. For example, if an inductor is operated at or above its SRF, it might behave as a capacitor. To avoid this, select components where the SRF is much higher than the operating frequency.
- Also, lead-through-hole parts have leads that add series inductance to a design, and surface-mount parts have pads that add shunt capacitance to a circuit.
- Another issue is that of packaging a completed design. If the circuit is to be integrated and sold as an Integrated Circuit (IC), it must be packaged. The package introduces several negative effects. In an IC, the bond wires add unwanted inductance (L) and the bond pads add unwanted capacitance (C).
- Isolation between pins in the package is also important. Lack of pin-to-pin isolation in a feedback circuit can lead to major reliability problems and stability concerns. The additional inductance in the emitter of the collector-emitter section can severely degrade the noise figure of the circuit.
- Additionally, several grounds are usually needed to improve the performance of RF circuits, but the package has a limited number of pins. After using the input, output, and power-supply pins, there may not be enough ground pins to accommodate an adequate design.

All of these factors can degrade the circuit's performance from the ideal, and the designer must carefully take them into account.

7. CMOS LNA Design

- A few comparison characteristics between CMOS and BJT LNAs:
 - The DC currents of CMOS and BJT LNAs are close; therefore the transconductance (g_m) of CMOS transistor is lower than the BJT one's.
 - The g_m/I ratio of CMOS is lower than that of BJT.
 - In CMOS technologies, a high f_T is achieved through a smaller C_{gs} , while in BJT technologies the same f_T is obtained through a higher g_m .
 - Smaller C_{gs} means CMOS tuned circuits tend to have higher Q, a disadvantage in withstanding component or process variation.

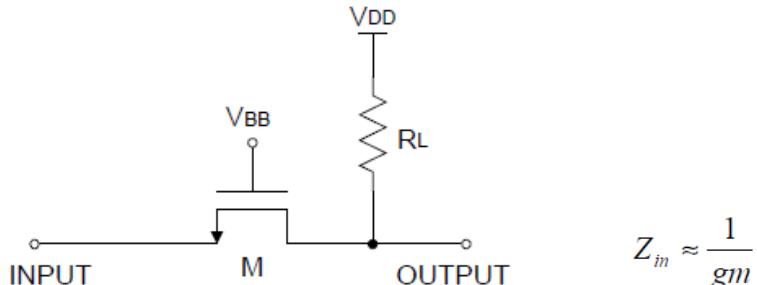
The CMOS LNA input quality factor (independent of L_s) is defined as follows:

$$Q_{gs} = \frac{1}{\omega_o C_{gs} Z_0}$$

There are two types of methods commonly used to design an LNA in CMOS circuits:

- Common-Gate
- Cascode amplifier

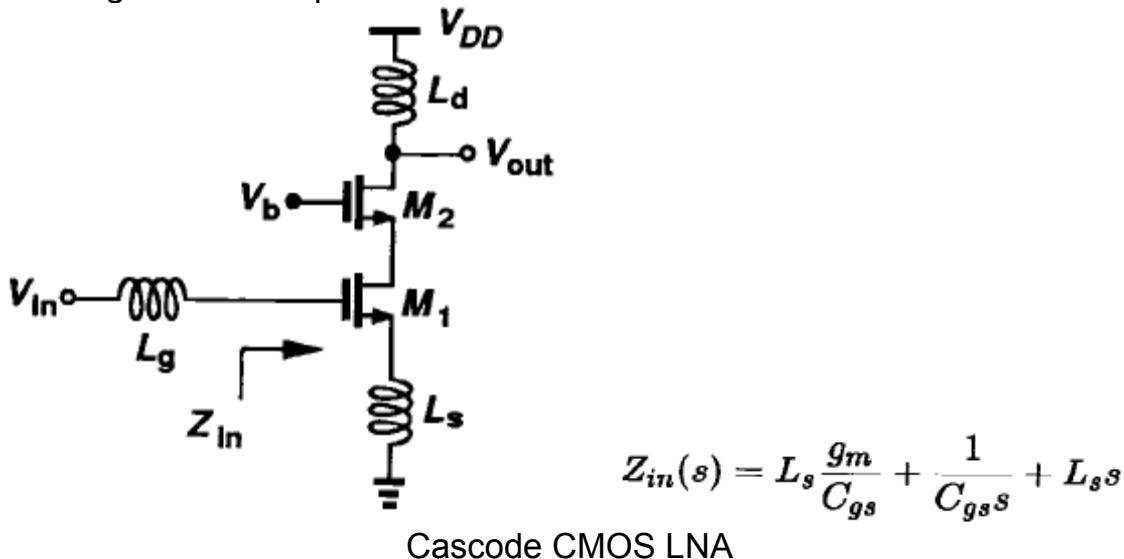
While the Common-Gate stage provides a wide-band input matching and is less sensitive to parasitics, it has an inherently high noise figure.



Common-Gate CMOS LNA

- With the increasing of the operating frequency, the parasitic transistor capacitance C_{gs} starts playing roles, which degrades the amplifier performance in the high frequency. In the narrow band application, a shunt inductor is added in the input to resonate with C_{gs} to have a good impedance matching in the designed frequency.
- Due to the lower quality factor of the resonant network, Common-Gate it is more robust against the process and electrical variation.
- Due to the missing of the C_{gd} path from the input to the output, the Common-Gate LNA shows better reverse isolation and stability versus Common-Source LNA.

Therefore, in most CMOS applications where the noise figure is critical issue, a cascode LNA with inductive degeneration is preferable.



Cascode CMOS LNA

- For a given unit-gain frequency as we lower the bias current, the noise figure decreases.
- For a given Q, higher g_m improves the noise figure.

- For a given source resistance of 50 ohms, as we reduce L_s , unit-gain frequency increases but the minimum value of L_s is limited by parasitic and sensitivity issues.
- By proper choice of g_m , L_s , and C_{gs} , the input resistance can be equal to 50 ohms source resistance and the input reactance (imaginary part of impedance) can be resonated out by a series inductor (L_s).
- Inductor degeneration (L_s) also improves the linearity by forming a negative series feedback.
- As we lower the bias current, while keeping unit-gain frequency constant, C_{gs} decreases, leading to higher Q.
- A high Q matching networks has several drawbacks:
 - Circuit becomes very sensitive to component variations and parasitics.
 - The input matching circuit which in this case contains a series inductor, inserts a large amount of loss at the input (even for a high-Q off-chip inductor)
- Another source of noise in the cascode topology is the noise introduced by the cascode device, M2, added to improve stability of the amplifier.
 - At high frequencies the capacitance at the drain of M1 reduces the impedance of this node, increasing the output noise, so to minimize the noise is very important to minimize this capacitance.
 - To improve the noise performance of the cascode design, the parasitic capacitance at the drain of M1 is resonated out by adding an inductor to the source of cascode.

This inductor should be sized carefully in order to resonate the unwanted capacitances at the desired frequency of operation.

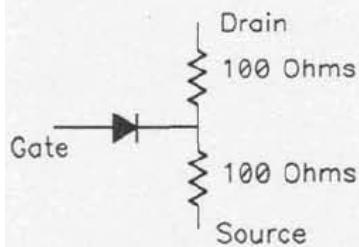
8. GaAs FET LNA Design

In case of GaAs FETs generally the input matching network transforms 50 ohms to the input impedance of the transistor, which typically at VHF frequencies is about 3000 ohms. A high-quality circuit usually is used to transform the impedance up and in the same time to filter out unwanted signals.

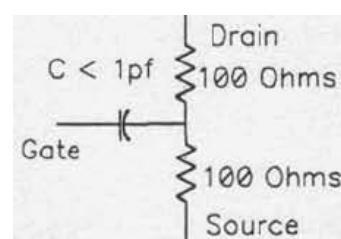
As frequency increases to microwave region (above 10GHz), the GaAs FET input impedance drops to few tens of ohms. HEMT's type transistors have even higher impedance at VHF frequencies (about 5000 ohms) dropping to 50 ohms above 10GHz.

Biasing the GaAs FET LNA

The input of a GaAs FET is a small Schottky diode, when the input of a HEMT is basically a small Tunnel diode. In normal operation these diodes are negatively biased, effectively making the gate a low value capacitor.



DC Equivalent Model of a GaAs FET

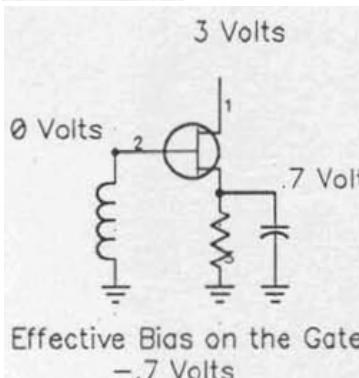


RF Equivalent of a biased GaAs FET

There are two common ways to supply the required negative voltage to the gate of a GaAs FET.

- This method requires power supply to be more complex than the LNA, but usually is the preferred method in most of the designs.
- Have the advantage to ground directly the source, and have excellent control of all DC parameters of the GaAs device.
- One of biggest issue of this biasing method is, the negative supply must turn ON first, and turn OFF last, making the power supply to be complex.

1.



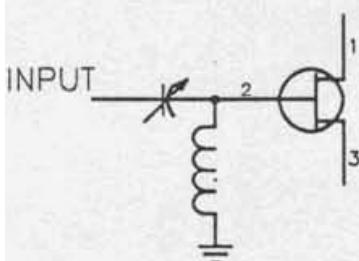
2.

- The self-bias method makes the voltage drop across the source resistor raises the substrate of the device above DC ground, giving to the gate a reverse bias.
- The method has the advantage using only one positive supply and also the thermal run-away is almost eliminated.
- The device operates in a current limiting mode; more current gives a greater negative bias to the gate, turning OFF the device.
- RF bypassing of the source becomes difficult at microwave frequencies, and need to use capacitors with low parasitic inductance.

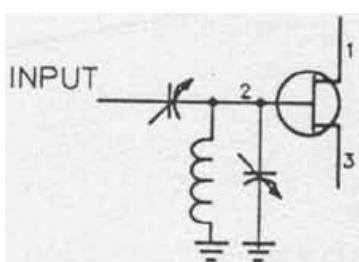
GaAs FET LNA - Input Matching

There are few topologies for matching of GaAs FETs, each of them having pros and cons.

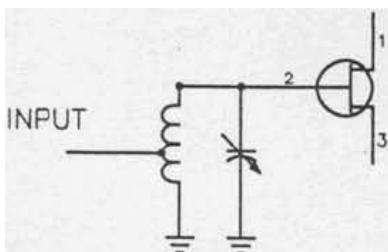
Remember that, not always want to tune for best Noise Figure or highest gain.



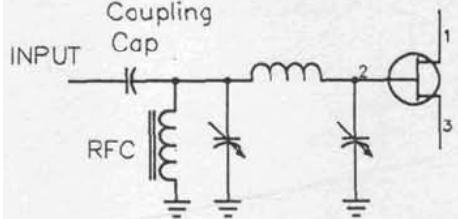
- This low insertion loss and simple input match circuit, works well up to UHF frequencies. There are not too many choices of tuning and a match from 50 ohms to Gamma Optimum (best Noise Figure) depends on the FET's internal stray capacitance from gate to the ground.



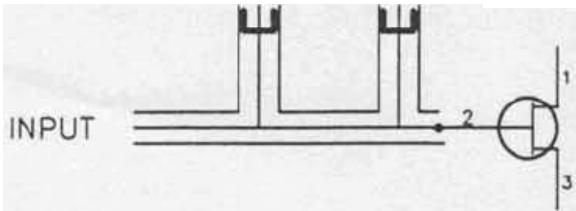
- This input match gives a high performance below 500MHz.
- Can get the best Noise Figure, Gamma Optimum that can be reached.
- Because the input impedance of the GaAs FET below 500MHz is high, the Q or Bandwidth of the input circuit can be varied with little impact on the Noise Figure.



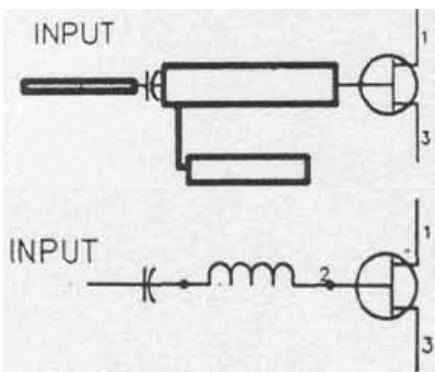
- The tapped L input is fine for VHF range high performance LNAs.
- It has very low loss, input is grounded, but the circuit has very wide bandwidth.



- The Pi input matching network works from low frequencies up to few GHz, and virtually can match any impedance.
- In the Pi matching network the second capacitor is tuned at a very low value, due to the FET input stray capacitance.
- This method requires high quality capacitors.



- The stub circuit is versatile and capable of matching almost of any input impedance, but the assembly is very large below 1GHz and relative difficult to supply the bias voltage.



- The microstrip stub circuit is limited to microwave frequencies, is narrowband, but has great repeatability.
- Just a small inductance in the gate lead can match the impedance in the 1GHz to 2GHz region. The circuit is simple and has low loss, broad bandwidth and excellent Noise Figure.

GaAs FET LNA - Construction and Operating Tips

- Always leave the LNA powered up. When operating, the gate of the GaAs FET is negatively biased at 0.5V-0.7V. The FET is damaged when the gate has positive voltage and conducts current. You need enough RF at the input to overcome the negative bias before damage the transistor.
Typically the circuit needs 4 times more RF leakage power to damage a turned ON GaAs than a turned OFF GaAs FET device.
- Use very short leads when the circuit works at microwave frequencies.
- Use small coils and avoid mutual coupling between them placing the inductors 90° to each other.
- Use Ferrite beads which behave as both, RF absorber and a low-Q RF choke.
- If use trimmer caps, connect their rotors to the ground or to the connectors.
When the microwave LNA is placed inside of a shielded box this become a cavity at some frequency, and the LNA might become an oscillator. Absorbers placed inside of the box help preventing this happen.

Typical characteristics of different configurations of High-Frequency GaAs FET LNAs

Characteristic	Common-Source (CS)	Common-Gate (CG)	Cascode
NFmin	Best	Better	Good
Gain	Moderate	Low	High
Bandwidth	Narrow	Very Wide	Wide
Stability	Compensation often required	RF Decoupling of the Gate is important	Good
Reverse Isolation	Lower	Moderate	Best

References

1. *RF Design Magazine* (1994-2002)
2. *Applied Microwave & Wireless Magazine* (1998-2002)
3. *Microwaves & RF Magazine* (1998-2002)
4. *U.L. Rohde, D.P. Newkirk - RF/Microwave Circuit Design – John Wiley & Sons, Inc-2000*
5. *G. Gonzales – Microwave Transistor Amplifiers – Prentice-Hall – 1984*
6. *Design of Analog CMOS Integrated Circuits – B. Razavi*
7. *GaAs FET Pre Amp Cookbook – K. Britain WA5VJB*
8. *Agilent Technologies – IP3 Measurements Data Sheets*
9. *Avago Technologies – Application Datasheets*
10. *LNA Design Trade-Offs in the Working World - Freescale*

Tutorial-1

Low Noise Amplifier (LNA) Design

Complied by Rashad M. Ramzan

Objective:

Low noise amplifiers are one of the basic building blocks of any communication system. The purpose of the LNA is to amplify the received signal to acceptable levels with minimum self-generated additional noise. Gain, NF, non-linearity and impedance matching are four most important parameters in LNA design.

The objective of this tutorial is to outline the basic tradeoffs between different amplifying topologies w.r.t gain, NF and impedance matching. After this comparison it is concluded that inductor degenerated common source topology gives the best performance to meet the gain, NF, and impedance matching goals with minimum power consumption in case of narrow band designs.

Goals:

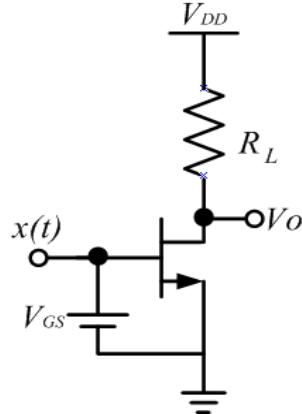
After this tutorial, students should be able to

- Calculate the gain, input impedance and NF of common gate, common source, and shunt feedback amplifiers.
- Understand the basic equations and tradeoff between different LNA topologies.
- Perform the calculation for inductor degenerated common source topology and understand the tradeoff between the gain, NF, and impedance matching.

A supplement tutorial LNA lab is also part of this course which guides through different analyses to design a practical LNA.

Problem-1.1(Tutorial)

NMOS transistor is racing horse in LNA design arena due to its higher mobility compared to PMOS transistors. Calculate the IP3 of NMOS CS amplifier shown below. Assume that NMOS transistor is in saturation.



- a) Consider simplified square law model. (HW)

$$I_D = \frac{K_n}{2} (V_{GS} - V_T)^2$$

- b) Consider the short channel effects as:

$$I_D = \frac{K_n}{2} \left[\frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)} \right]$$

θ = Velocity Saturation, Mobility Degradation

$$V_{GS} - V_T = 0.2V \quad \text{and} \quad \theta = 0.1V^{-1}$$

Observe that this transistor is not a very “short channel” device as $\theta \ll 1$.

- c) What conclusion can be drawn from part b) about the bias current and transconductance of the transistor for higher IP3?

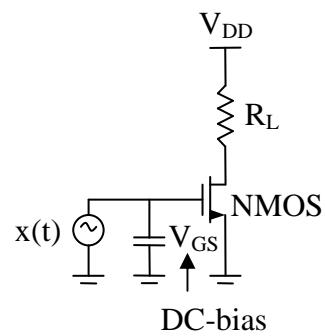
Solution:

- a). Homework answer: $IP3 = \infty$

b).

$$y(x) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \quad \dots \dots \dots (1)$$

$$x(t) = A \cos \omega_1 t + A \cos \omega_2 t \Rightarrow A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|}$$



$$I_D = \frac{K_n}{2} \frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)}$$

Here we assume a small signal $x(t)$ around the bias $(V_{GS} - V_T)$, so

$$I_D = \frac{K_n}{2} \frac{[(V_{GS} - V_T) + x(t)]^2}{1 + \theta(V_{GS} - V_T + x(t))}$$

we define $V_{GS} - V_T = \Delta V$ ----- Bias voltage

$$I_D = \frac{K_n}{2} \frac{[x(t) + \Delta V]^2}{\theta(x(t) + \Delta V) + 1} \quad \text{-----(2)}$$

$$V_o = I_D R_L \Rightarrow V_o = \frac{K_n R_L}{2} \frac{(x(t) + \Delta V)^2}{1 + \theta(x(t) + \Delta V)} \text{ and we put } \frac{K_n R_L}{2} = K$$

$$\theta \ll 1 \text{ so } (x(t) + \Delta V) \text{ is also small} \Rightarrow \frac{1}{1 + \rho} = 1 - \frac{\rho}{2}$$

$$\frac{1}{1 + \theta(x(t) + \Delta V)} \approx 1 - \frac{\theta(x(t) + \Delta V)}{2}$$

$$V_o = K(x(t) + \Delta V)^2 \left(1 - \frac{\theta(x(t) + \Delta V)}{2} \right)$$

$$V_o = K(x(t) + \Delta V)^2 - (x(t) + \Delta V)^3 \frac{K\theta}{2}$$

$$V_o = K\Delta V^2 - \frac{K\theta}{2}\Delta V^3 + \left(2K\Delta V - \frac{3K\theta}{2}\Delta V^2 \right)x(t)$$

$$+ \left(K - \frac{3K\theta}{2}\Delta V \right)x^2(t) - \frac{K\theta}{2}x^3(t) \quad \text{-----(3)}$$

Comparing (1) & (3)

$$\alpha_1 = 2K\Delta V - \frac{3K\theta}{2}\Delta V^2, \quad \alpha_2 = K - \frac{3K\theta}{2}\Delta V, \quad \alpha_3 = -\frac{K\theta}{2}$$

$$A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|} = \sqrt{\frac{4}{3} \times \frac{2K\Delta V - \frac{3}{2}K\theta\Delta V^2}{\frac{K\theta}{2}}} = \sqrt{\frac{8}{3} \left(\frac{2\Delta V}{\theta} - 3\Delta V^2 \right)}$$

$$A_{IP3} = \sqrt{\frac{8}{3} \frac{2\Delta V}{\theta}} = \sqrt{\frac{16}{3} \frac{\Delta V}{\theta}} \quad \text{As } \theta \ll 1 \quad 3\Delta V^2 \text{ can be ignored.}$$

Please, note that this formula only holds for small value of θ .

A large gate bias voltage ($V_{GS} - V_T$) improves IP3.

Put $\Delta V = 0.2V$, $\theta = 0.1 V^{-1}$

$$A_{IP3} = \sqrt{\frac{16}{3} \frac{0.2}{0.1}} = 3.27 Volts$$

$$IIP3(dBm) = 10 \log \left[\left(\frac{3.27}{\sqrt{2}} \right)^2 \cdot \frac{1}{50} \right] \text{lmW} \cong 20 dBm$$

c). From $I_D = \frac{K_n}{2} \frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)}$ the NMOS transconductance can be found as

$$g_m = \frac{K_n}{2} \frac{(V_{GS} - V_T)(2 + \theta(V_{GS} - V_T))}{(1 + \theta(V_{GS} - V_T))^2}.$$

By comparison of those two formulas we find

$$\frac{I_D}{g_m} = (V_{GS} - V_T) \times \frac{1 + \theta(V_{GS} - V_T)}{2 + \theta(V_{GS} - V_T)} \cong \frac{V_{GS} - V_T}{2}$$

and hence, (4) can be rewritten as

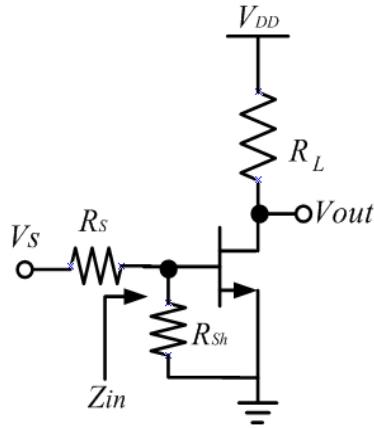
$$A_{IP3} \equiv \sqrt{\frac{32}{3\theta} \frac{I_D}{g_m}}$$

As shown, IIP3 is decided by the ratio I_D/g_m which is constant for a given gate bias voltage. Using e.g. a wider transistor does not change this ratio and only the power consumption is increased.

Problem-1.2 (Tutorial)

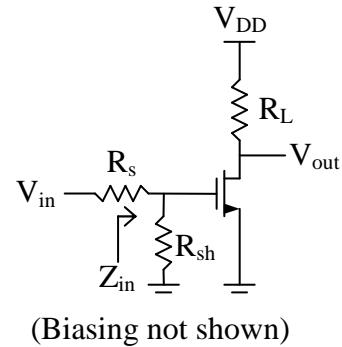
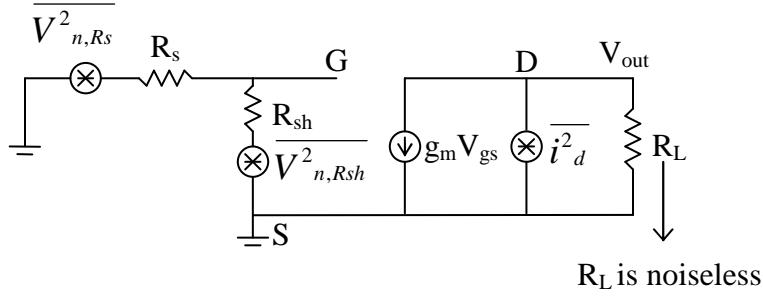
It is preferred in current RF designs that the input of LNA be matched to $50\ \Omega$. The easiest way is to shunt the gate with a resistor of $50\ \Omega$.

- a) Calculate the gain, input impedance and NF in absence of gate noise. Assume that $R_{sh}=R_L$ for NF derivation.
 - b) What are the disadvantages of shunt resistor with reference to gain and NF?



Solution:

a). (Please read assumption in the problem statement carefully)



$$F = \frac{\text{Total output noise power}}{\text{Output noise due to input source}}$$

$$\overline{V^2_{m,Rs}} = 4kTR_s\Delta f$$

$$\text{Gain} \Big|_{\text{Gate}} = -g_m R_L$$

$$\overline{V^2_{m,Rsh}} = 4kTR_{sh}\Delta f$$

$$A = g_m R_L \left(\frac{R_{sh}}{R_s + R_{sh}} \right) \text{ for } R_{sh} = R_s$$

$$\overline{i^2_d} = 4kTg_m\Delta f$$

$$A = -g_m \frac{R_L}{2}$$

Using superposition, considering one at a time and shorting / opening other sources.

$$\overline{V^2_{on,Rs}} = \overline{V^2_{n,Rs}} \times g_m^2 R_L^2 \times \left(\frac{R_{sh}}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2_{on,Rsh}} = \overline{V^2_{n,Rsh}} \times g_m^2 R_L^2 \times \left(\frac{R_s}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2}_{no,d} = \overline{i^2}_d \times R^2_L$$

$$F = \frac{\overline{V^2}_{on,Rs} + \overline{V^2}_{on,Rsh} + \overline{V^2}_{no,d}}{\overline{V^2}_{on,Rs}} = 1 + \frac{\overline{V^2}_{on,Rsh} + \overline{V^2}_{o,d}}{\overline{V^2}_{on,Rs}}$$

$$F = 1 + \frac{\frac{4kTR_{sh}\Delta f \times g_m^2 R^2_L \times R^2_{sh}}{(R_s + R_{sh})^2}}{\frac{4kTR_s\Delta f \times g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}} + \frac{\frac{4kT\gamma g_m \Delta f \times R^2_L}{4kTR_s\Delta f \times g_m^2 R^2_L \times R^2_s}}{\frac{4kTR_s\Delta f \times g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}}$$

In case of impedance match $R_s = R_{sh}$

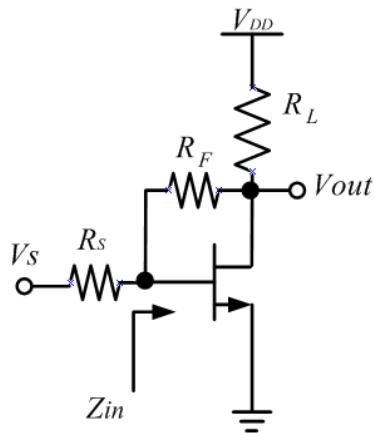
$$F = 1 + 1 + \frac{\frac{R_L^2 \gamma g_m}{R_s \times \frac{g_m^2 R_s^2 \times R_L^2}{4R_s^2}}} = 2 + \frac{\frac{R_L^2 \gamma g_m}{R_s \times \frac{g_m^2 \times R_L^2}{4}}} = 2 + \frac{4\gamma}{g_m R_s}$$

b).

- Poor Noise Figure
- Input signal attenuated by voltage divider
- R_{sh} adds extra noise.
- At high frequency, shunt L is needed to tune out C_{gs}
- Reduced gain.

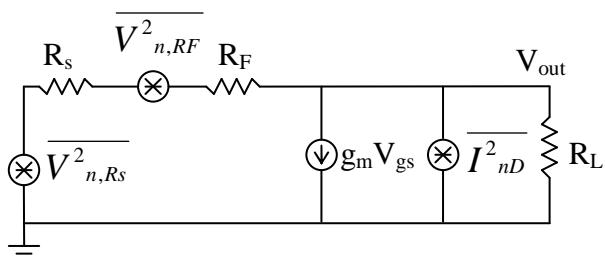
Problem-1.3 (Tutorial)

Another approach to get 50Ω input impedance match is shunt feedback amplifier shown below.

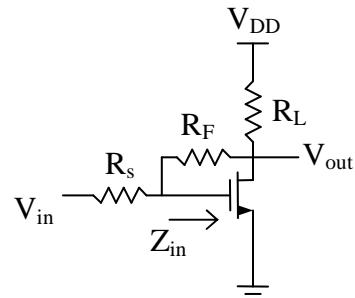


- a) Calculate the gain, input impedance and NF neglecting the gate noise. The gate-drain, gate-bulk, and gate-source capacitance can be neglected as well.
- b) What are the disadvantages of shunt feedback amplifier with reference to gain and NF?

Solution:



(Equivalent noise model ignoring gate noise), R_L is noiseless



(Biasing not shown)

$$\overline{I^2_{nD}} = 4kT\gamma g_m \Delta f, \overline{V^2_{n,RS}} = 4kTR_s \Delta f$$

$$F = \frac{\overline{V^2_{n,out}}}{A_{v,tot}^2 \overline{V^2_{RS}}} = \frac{\text{Total input noise power}}{\text{Output noise power due to input source}}$$

Here $A_{v,tot}$ = Gain from V_{in} to V_{out}

Again using superposition theorem

$$F = \frac{\overline{V^2_{n,out}}}{A_{v,tot}^2 \overline{V^2_{RS}}} = \frac{\overline{V^2_{n,RS,out}} + \overline{V^2_{n,RF,out}} + \overline{V^2_{n,D,out}}}{A_{v,tot}^2 \overline{V^2_{RS}}}$$

Gain Calculation

$$V_{in} = i_{in}(R_s + R_F) + V_{out}$$

$$V_{out} = (i_{in} - g_m V_{gs}) R_L$$

$$V_{gs} = i_{in} R_F + V_o$$

$$A_{v,tot} = \frac{V_{out}}{V_{in}} = \frac{R_L(1 - g_m R_L)}{R_s + R_F + R_L + g_m R_s R_L}$$

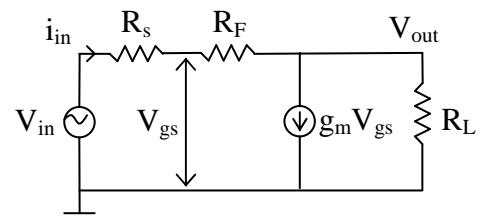
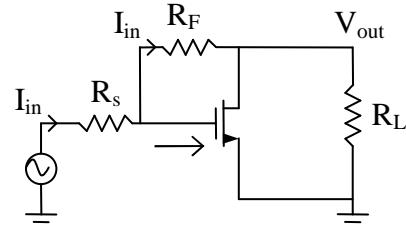
If $R_F \gg R_s$ & $g_m R_F \gg 1$

$$A_{v,tot} = \frac{-g_m R_L}{\frac{R_s}{R_F} + 1 + R_L + \frac{1 + g_m R_s}{R_F}} \approx -g_m R_L$$

$$A_{v,tot} \approx -g_m R_L$$

$$\text{Also } Z_{in} = \frac{R_F + R_L}{1 + g_m R_L}$$

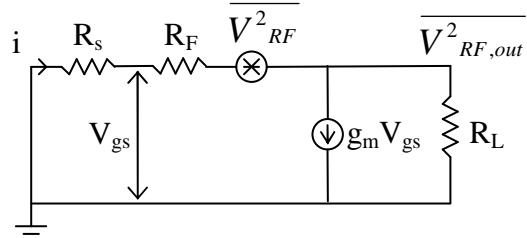
By ignoring C_{gs} , we have considered real part only.



For source resistance

$$\overline{V_{nRS,out}^2} = A_{v,tot}^2 \overline{V_{nRS}^2} \quad \text{-----(1)}$$

For feedback resistance



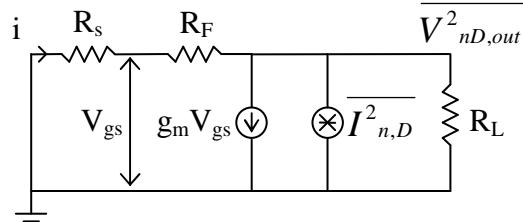
$$V_{gs} = -iR_s = iR_F - V_{RF} + V_{RF,out}$$

$$V_{RF,out} = R_L(i - g_m V_{gs})$$

$$V_{RF,out} = V_{RF} \frac{1}{1 + \frac{R_s + R_F}{R_L(1 + g_m R_s)}} = V_{RF} \frac{R_L}{R_F} (1 + g_m R_s)$$

$$\overline{V_{n,RF,out}^2} = \overline{V_{n,RF}^2} \left[\frac{R_L}{R_F} (1 + g_m R_s) \right]^2 \quad \text{-----(2)}$$

Similarly



$$\frac{V_{nD,out}}{R_L} + I_{nD} + g_m V_{gs} + \frac{V_{nD,out}}{R_S + R_F} = 0$$

$$V_{gs} = R_S \frac{V_{nD,out}}{R_S + R_F}$$

$$V_{nD,out} = \frac{I_{nD}}{\frac{1}{R_L} + \frac{1}{R_S + R_F} + \frac{g_m R_S}{R_S + R_F}} \approx I_{nD} R_L$$

So,

$$\overline{V_{nD,out}^2} = I_{nD}^2 R_L^2 \quad \text{-----(3)}$$

Combaining (1) (2) & (3)

$$F = 1 + \frac{\overline{V^2}_{n,RF} \left[\frac{R_L}{R_F} (1 + g_m R_s) \right]^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}} + \frac{\overline{I^2}_{nD} R_L^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}}$$

$$A_{v,tot} = -g_m R_L, \quad \overline{V^2}_{n,RS} = 4kTR_S \Delta f, \quad \overline{V^2}_{M,RF} = 4kTR_F \quad \& \quad \overline{I^2}_{nD} = 4kT\gamma g_m$$

$$F = 1 + \frac{R_s}{R_F} \left(1 + \frac{1}{g_m R_s} \right)^2 + \frac{\gamma}{g_m R_s}$$

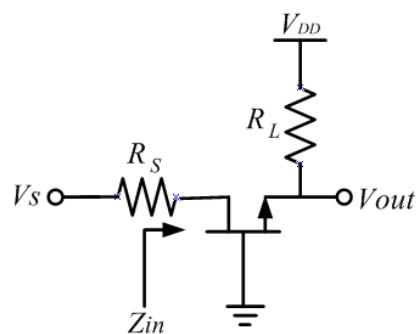
b).

NF \downarrow $g_m R_s \uparrow$ & $R_F \uparrow$ usually $R_s = 50\Omega$

- Better performance than CS amplifier
- R_F induces noise
- At higher $f \uparrow$ a shunt inductor needed to tune out C_{gs}
- Broadband Amp @ Lower frequency
- To make NF \downarrow $R_F > R_s$ and $g_m R_s \gg 1$

Problem-1.4 (HW)

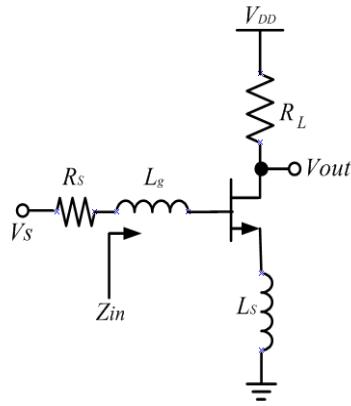
Common gate amplifier also offers 50Ω input impedance match and solves the input matching problem.



- c) Calculate the gain, input impedance and NF in absence of gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- a) What are the disadvantages of common gate amplifier with reference to gain and NF?

Problem-1.5 (Tutorial)

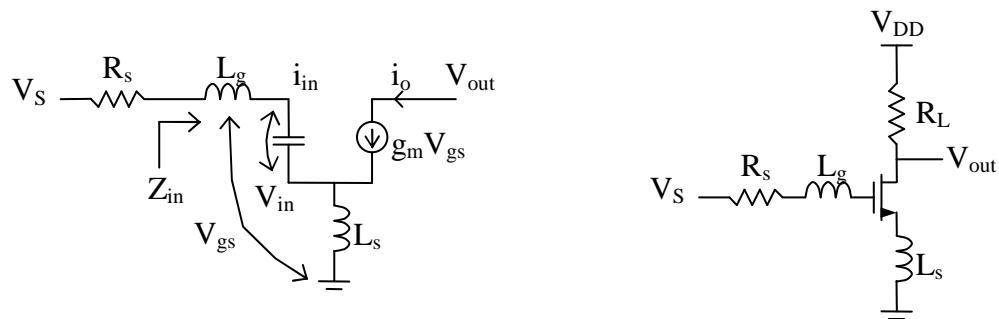
The disadvantages of the amplifiers discussed in Problem-2, 3 & 4 can be circumvented by using the source degenerated LNA shown below.



- Calculate the input impedance. This inductor source degenerated amplifier presents a noiseless resistance for 50Ω for input power match. How we can cancel the imaginary part of complex input impedance so that the LNA presents 50Ω real input resistance at input port.
- Calculate the NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- C_{gd} bridges the input and output ports. The reverse isolation of this LNA is very poor. Why reverse isolation is important? Suggest the modification to improve reverse isolation.

Solution:

a).



(Biasing not shown)

From model above we can write

$$V_{in} = i_{in} \left(j\omega L_g + j\omega L_s \right) + i_{in} \left(\frac{1}{j\omega C_{gs}} \right) + i_o j\omega L_s \quad \dots \dots \dots (1)$$

$$i_o = g_m V_{gs} = g_m i_{in} \times \frac{1}{j\omega C_{gs}} \quad \dots \dots \dots (2)$$

Substituting (2) in (1)

$$V_{in} = i_{in} \left[j\omega (L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}} \right]$$

$$Z_{in} = \frac{V_{in}}{I_{in}} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

For matching $L_g + L_s$ are canceled out by C_{gs} . So at frequency of interest

$$\omega_o(L_g + L_s) = \frac{1}{\omega_o C_{gs}} \Rightarrow \omega_o^2 = \frac{1}{(L_g + L_s)C_{gs}}$$

$$\text{And } R_S = 50\Omega = \frac{g_m}{C_{gs}} L_s$$

Notes:

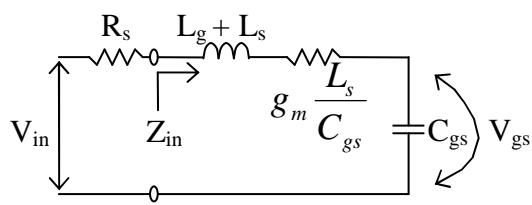
- a). L_s is typically small and may be realized by the bond wire for source.
- b). L_g can be implemented by spiral/external inductor.

b).

From part a)

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

We can draw this circuit as



Here

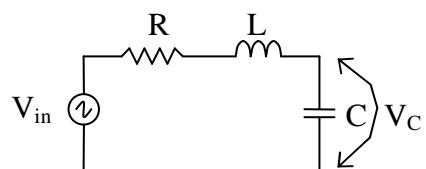
$$Q_{in} = \frac{\omega_o(L_g + L_s)}{R_s + \frac{g_m L_s}{C_{gs}}} = \frac{\omega_o(L_g + L_s)}{R_s + \omega_T L_s}$$

$$\because \omega_T \cong \frac{g_m}{C_{gs}} \quad \text{frequency of current gain equal 1}$$

$$Q_{in} = \frac{1}{\omega_o \left(R_s + \frac{g_m L_s}{C_{gs}} \right) C_{gs}} \quad \text{for match load } R_S = \frac{g_m L_s}{C_{gs}}$$

Reference:

For series RLC Circuit



$$Q_s = \frac{1}{R} \sqrt{\frac{L}{C}} = \frac{\omega_o L}{R} = \frac{1}{\omega_o R C}$$

$$\text{and } V_C = Q_s V_{in}$$

$$Q_{in} = \frac{1}{2\omega_o R_s C_{gs}}$$

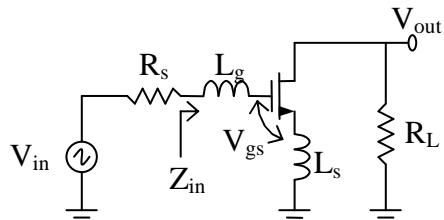
Gain

$$V_{gs} = Q_{in} V_{in}$$

$$g_m = \frac{I_{out}}{V_{gs}}$$

$$G_m = \frac{I_{out}}{V_{in}} = \frac{V_{gs} g_m}{V_{in}} = Q_{in} g_m$$

$$G_m = Q_{in} g_m$$



$$\text{so, } \frac{V_{out}}{V_{in}} = -G_m R_L \quad \text{where} \quad G_m = Q_{in} g_m$$

Noise Figure:

$$F = \frac{\text{Total noise power at output}}{\text{noise power at output due to input source}}$$

For this calculation we ignore channel noise.

$$F = \frac{\overline{V^2}_{nRS,OUT} + \overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}} = 1 + \frac{\overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}}$$

$$\overline{V^2}_{nD,OUT} = \overline{i^2}_{n,D} R_L^2 \quad \overline{i^2}_{n,D} = 4kT \gamma g_m \Delta f$$

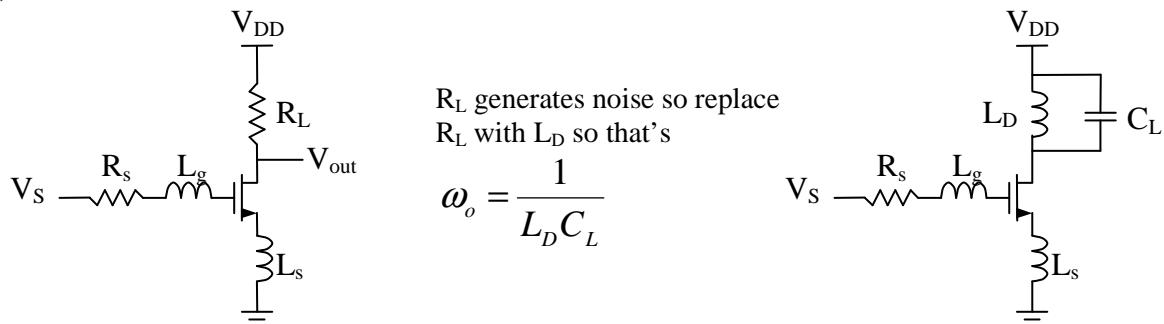
$$\overline{V^2}_{nRS,OUT} = \overline{V^2}_{n,RS} G_m^2 R_L^2 \quad \overline{V^2}_{n,RS} = 4kT R_S \Delta f \quad \& \quad G_m = Q_{in} g_m$$

$$F = 1 + \frac{\overline{i^2}_{n,D} R_L^2}{\overline{V^2}_{n,RS} Q_{in}^2 g_m^2 R_L^2} \quad \overline{i^2}_{n,D} = 4kT \gamma g_m, \quad \overline{V^2}_{n,RS} = 4kT R_S$$

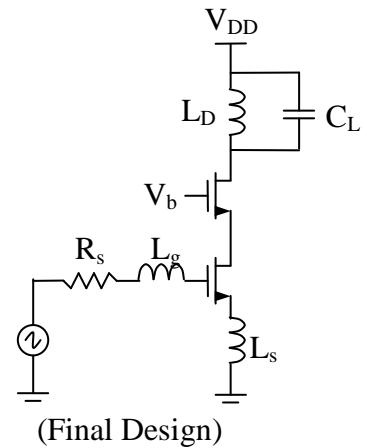
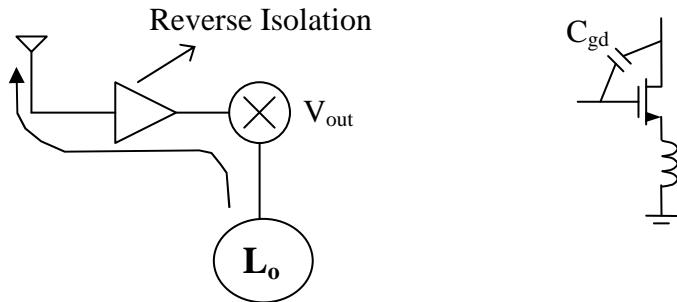
$$F = 1 + \frac{\gamma}{g_m R_S Q_{in}^2}$$

Notes:

- Very good NF value
- Narrow band matching
- NF \downarrow with Q^2
- The Q is dependent upon $L_g + L_s$, L_s usually small so Q depends mainly upon L_g

C). Drawbacks**i).**

The C_L can be considered the input capacitance of the following mixer or filter.

ii).

Reverse isolation depends upon capacitance between output and input.
To make it less the cascode architecture can be used.

Problem-1.6 (HW)

Fill-in the Table below, use the data from Problem-1.2, 1.4, 1.3 and 1.5

Type of LNA	Z_{in}	Noise Factor	Gain	NF (dB)
Shunt Resistor	R_{sh}	$2 + \frac{4\gamma}{g_m R_s}$	$\frac{-g_m R_L}{2}$	
Common Gate				
Shunt Feedback				
Source Degenerated				

- Calculate the NF for all above amplifiers. Assume $\gamma=2$, $g_m = 20\text{mS}$, $R_s = 50\Omega$, $R_f = 500\Omega$, and $Q_{in} = 2$.
- Which is the best topology for Narrow Band LNA design at high frequency?

Problem-1.7 (Tutorial)

Real Design: We will design the inductor-source-degenerated LNA shown in Fig below to meet the specification outlined for IEEE802.11b standard. The first cut approximate values are calculated as a starting point for simulation.

LNA Specification:

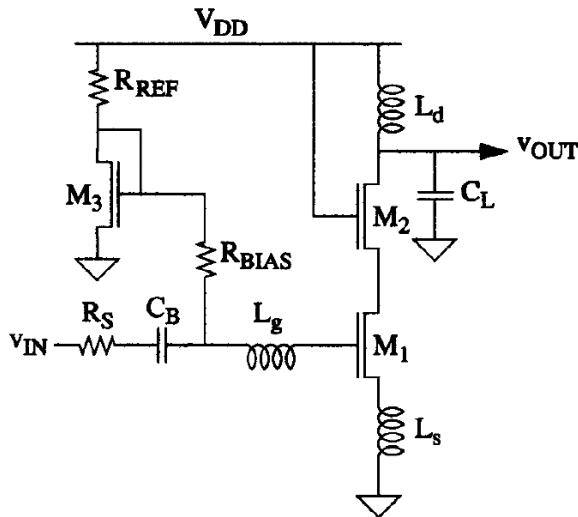
$NF < 2.5 \text{ dB}$, Gain $> 15 \text{ dB}$, $IIP3 > -5 \text{ dBm}$, Centre Frequency = 2.4 GHz

Load Capacitance = 1pF

Technology Parameters for 0.35um CMOS:

$$L_{eff} = 0.35\mu m, \quad \mu_n C_{ox} = 170 \mu A/V^2, \quad C_{ox} = 4.6 mF/m^2, \quad \mu_p C_{ox} = 58 \mu A/V^2, \quad \gamma = 2$$

$$\delta = 4, \quad |C| = 0.395, \quad \alpha = 0.85$$



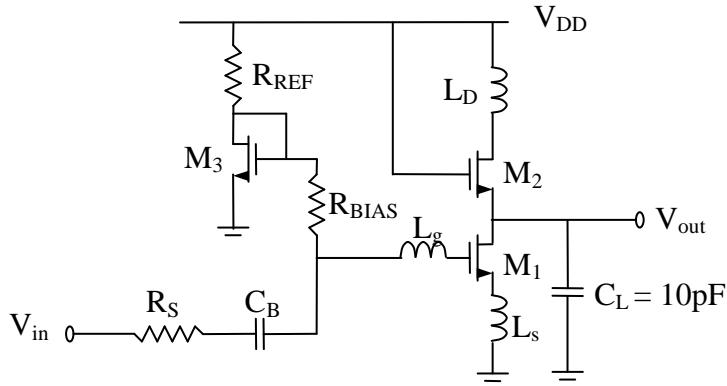
Solution:

$$\text{Technology } 0.35\mu m \text{ CMOS: } \left\{ \begin{array}{l} \mu_o C_{ox} = 170 \mu A/V^2, \mu_p C_{ox} = 58 \mu A/V^2, \\ C_{ox} = 4.6 mF/m^2, \gamma = 2, L_{eff} = 0.35\mu m \end{array} \right\}$$

$$\delta = 4, \quad |C| = 0.395, \quad \alpha = 0.85$$

Design Parameters

$NF < 2.5 \text{ dB}$, Gain $> 15 \text{ dB}$, $IIP3 > -5 \text{ dBm}$, $f_0 = 2.4 \text{ GHz}$



Component Description

L_s – Matches input impedance

L_g – Sets the Resonant Frequency $f_O = 2.4$ GHz

M_3 – Biasing transistor which forms current mirror with M_1

L_d – Tuned output increases the gain and also work as band pass filter with C_L

M_2 – Isolates tuned input from output to increase reverse isolation, also reduces the effect of Miller capacitance C_{gd}

C_B – BC blocking capacitor chosen to have negligible reactance at $f_O = 2.4$ GHz

R_{BIAS} – Large enough so that its equivalent current noise is small enough to be ignored. (Don't consider it as voltage noise source. Why??)

Design Procedure

Size of M_1 :

From the noisy two-port theory (see the course book or lecture notes) the optimal input matching and minimum noise figure is given by:

$$G_{opt} = \alpha \omega C_{gs} \sqrt{\frac{\delta}{5\gamma} (1 - |C|^2)} = \frac{1}{50\Omega} \quad \text{-----(A)}$$

$$F_{min} = 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_T} \sqrt{\gamma C (1 - |C|^2)} = 1 + 2.3 \frac{\omega}{\omega_T} \quad \text{-----(B)}$$

From (A)

$$C_{gs} \approx 2.7 \text{ pF} \Rightarrow W_{M1} \approx 3C_{gs} / 2C_{OX} L_{eff} \approx 2.5 \text{ mm} \quad (\text{not feasible - huge size, huge power !})$$

Conclusion: We will not go for the global minimum noise figure. Instead, we will look into the constraint power design approach.

Solution:

LNA NF will be optimized for given power which is higher than the global minimum NF.

In this case the optimum transistor width is given by:

$$W_{opt} = \frac{1}{3\omega_o L_{eff} C_{ox} R_s}$$

while the minimum power-constraint NF :

$$F_{min,p} = 1 + 2.4 \frac{\gamma}{\alpha} \frac{\omega}{\omega_T} \Rightarrow F_{min,p} = 1 + 5.6 \frac{\omega}{\omega_T} \quad \text{-----(C)}$$

(B) is the global minimum noise figure.

(C) is the minimum NF for a given power consumption.

In practice the difference is usually 0.5dB to 1dB (no big deal for Lower Power)

Step - 1:

$$I_1 = I_2 = 5mA \text{ (Limited Power consumption)}$$

Step - 2:

$$W_{M1} = \frac{1}{3\omega_0 L_{eff} C_{ox} R_s}$$

$$W_{M1} = \frac{1}{3 \times 0.35\mu \times 4.6m \times 50 \times \omega_o} \quad \left. \begin{cases} R_s = 50\Omega, C_{ox} = 4.6mF/m^2, \\ \mu_n C_{ox} = 170\mu A/V, L_{eff} = 0.35\mu m, \\ \omega_o = 2\pi f_o, f_o = 2.4GHz \end{cases} \right\}$$

$$W_{M1} = 3.9 \times 10^{-4}$$

$$W_{M1} = 3.9 \times 10^{-4} = 390\mu m$$

Step - 3:

$$C_{gs1} = \frac{2}{3} W_{M1} L_{eff} C_{ox}$$

$$C_{gs1} = \frac{2}{3} \times 390\mu \times 0.35\mu \times 4.6m = 0.41pF$$

$$g_{m1} = \sqrt{2\mu_n C_{ox} \left(\frac{W}{L} \right)_{M1} I_{DM1}} \quad \text{or} \quad g_{m1} = \frac{2I_{DM1}}{V_{GS} - V_T} \quad (\text{for short channel model})$$

$$g_{m1} = \sqrt{2 \times 170\mu \times \left(\frac{390}{0.35} \right) \times 5m} = 43mA/V$$

$$\omega_T \approx \frac{g_m}{C_{gs1}} = \frac{43 \text{ mA/V}}{0.41 \text{ pF}} = 104G \text{ rad/Sec}$$

Assuming $\gamma = 2$

$$\text{Now } F_{\min} = 1 + 5.6 \frac{\omega_o}{\omega_T}$$

$$F_{\min} = 1 + 5.6 \frac{2\pi 2.4 G}{104G} \approx 2.55 \text{ dB}$$

$$NF \approx 2.55 \text{ dB}$$

This NF is very close to the specified value. If we increase I_D then ω_T should increase slightly as well and hence, a lower NF value can be achieved at expense of more power.

Step - 4:

Source and gate inductance such that they cancel C_{gs} and set 50Ω input impedance

$$\omega_o = 2\pi f_o = 2\pi 2.4 = 15G \text{ rad/Sec}$$

From previous problem

$$R_s = R_{transformed} = \frac{g_m L_s}{C_{gs}} \cong \omega_T L_s$$

$$L_s = \frac{R_s}{\omega_T} = \frac{50}{100G} \cong 0.5nH$$

$L_s = 0.5nH$ can be implemented using the bond wire.

$$\text{Now } L_g + L_s = \frac{1}{\omega_0^2 C_{gs1}}$$

$$L_g + L_s = \frac{1}{(15G)^2 \times 0.41 \text{ pF}} = 10.81nH$$

$$L_g \approx 10nH$$

Step - 5:

$$L_d = \frac{1}{\omega_o^2 C_L} \quad \because C_L = 1 \text{ pF}$$

$$L_d = \frac{1}{(15G)^2 \times 1 \text{ pF}} \cong 4.4nH$$

$$L_d = 4.4nH$$

Step - 6:

Size of M3 is chosen to minimize power consumption

$$W_{M3} = 70\mu m, \quad R_{REF} = 2k\Omega \Rightarrow I_3 = 0.6mA$$

$R_{BIAS} = 2k\Omega$ (Large enough so that it's equivalent current noise can be neglected)

$$C_B = 10pF \quad (X_C \approx 6.6\Omega \text{ so good value @ 2.4G} \quad X_B = \frac{1}{2\pi f_o C_B} = 6.6\Omega)$$

Step - 7:

Size M2 = M3

So that they can have shared Drain Area..

(Note: You will simulate same LNA circuit in LAB # 2)

Tutorial-2

Low Noise Amplifier (LNA) Design

Written By: Rashad.M.Ramzan
rashad.ramzan@nu.edu.pk

Objective:

Low noise amplifiers are one of the basic building blocks of any communication system. The purpose of the LNA is to amplify the received signal to acceptable levels with minimum self generated additional noise. Gain, NF, non-linearity and impedance matching are four most important parameters in LNA design.

The objective of this tutorial is to outline the basic tradeoffs between different amplifying topologies w.r.t gain, NF and impedance matching. After this comparison it is concluded that inductor degenerated common source topology gives the best performance to meet the gain, NF, and impedance matching goals with minimum power consumption in case of narrow band designs.

Goals:

After this tutorial, students should be able to

- Calculate the gain, input impedance and NF of common gate, common source, and shunt feedback amplifiers.
- Understand the basic equations and tradeoff between different LNA topologies.
- Perform the calculation for inductor degenerated common source topology and understand the tradeoff between the gain, NF, and impedance matching.

A supplement tutorial LNA lab is also part of this course which takes the circuit from Problem-2.8 and guides through different analysis to design and practical LNA.

Problem-2.1(Tutorial)

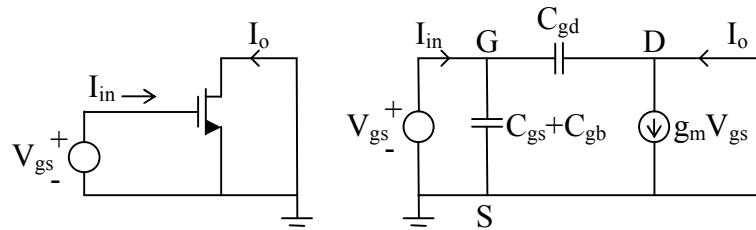
ω_T is single figure of merit for high frequency transistors. This is defined as frequency at which current gain is extrapolated to fall to unity.

Although the dc gate current of an MOS transistor is essentially zero, the high-frequency behavior of the transistor is controlled by the capacitive elements in the small-signal model, which cause the gate current to increase as frequency increases.

- Derive the expression for ω_T .
- For RF design we always use minimum length transistors. Why?

Solution:

a).



$$V_{sb} = V_{ds} = 0$$

So g_{mb} , r_o , C_{sb} , C_{db} have no effect on calculations. (This is drawback of ω_T definition)

$$i_i = j\omega(C_{gs} + C_{gb} + C_{gd})V_{gs}$$

$$i_i \approx g_m V_{gs}$$

$$\frac{i_o}{i_{in}} = \frac{g_m}{j\omega(C_{gs} + C_{gb} + C_{gd})}$$

According to definition $\frac{i_o}{i_{in}} = 1$ at ω_T

$$\omega = \omega_T = \frac{g_m}{C_{gs} + C_{gb} + C_{gd}}$$

C_{gb} and C_{gd} are small compared to C_{gs}

$$\text{So, } \omega_T \approx \frac{g_m}{C_{gs} + C_{gd}} \approx \frac{g_m}{C_{gs}} \quad \dots \dots \dots \quad (1)$$

$$\text{b). } I_D = \frac{\mu_o C_{ox}}{2} \frac{W}{L} (V_{gs} - V_T)^2$$

$$g_m = \frac{\partial I_D}{\partial V_{gs}} = \mu_o C_{ox} \frac{W}{L} (V_{gs} - V_T) \quad \dots \dots \dots \quad (2)$$

And $C_{gs} = C_o \times WL$ ----- (3)

Put (2) & (3) in (1)

$$\omega_T = \frac{g_m}{C_{gs}} = \frac{\mu_o C_{ox} W (V_{gs} - V_t)}{L \cdot C_{ox} \cdot W \cdot L} = \frac{\mu_o (V_{gs} - V_t)}{L^2}$$

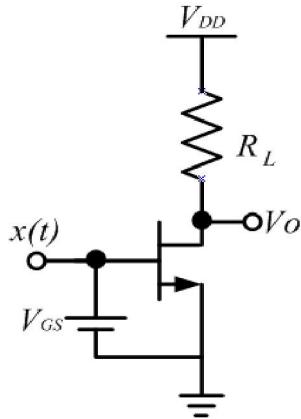
$$\omega_T = \frac{\mu_o (V_{gs} - V_t)}{L^2}$$

This means that $\omega_T \propto \frac{1}{L^2}$ so that's why minimum L is preferred. But this approximation holds

for long channel devices for short channel $\omega_T \propto \frac{1}{L}$ instead of $\frac{1}{L^2}$.

Problem-2.2(Tutorial)

NMOS transistor is racing horse in LNA design arena due to its higher mobility compared to PMOS transistors. Calculate the IP3 of NMOS CS amplifier shown below. Assume that NMOS transistor is in saturation.



- a) Consider simplified square law model. (**HW**)

$$I_D = \frac{K_n}{2} (V_{GS} - V_T)^2$$

- b) Consider the short channel effects as (**Tutorial**)

$$I_D = \frac{K_n}{2} \left[\frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)} \right]$$

$\theta = \text{Velocity Saturation, Mobility Degradation}$

$$V_{GS} - V_T = 0.2V \quad \text{and} \quad \theta = 0.1V^{-1}$$

- c) What conclusion can be drawn from part b) about the bias current and transconductance of transistor for higher IP3?

Solution:

a). Home work Ans: $IP3 = \alpha$

b). From Razavi

$$y(x) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \quad \dots \dots \dots (1)$$

$$x(t) = A \cos \omega_l t + A \cos \omega_2 t \Rightarrow IP3 = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|}$$

$$I_D = \frac{K_n}{2} \frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)}$$

Here we assume that small signal $x(t)$ over-rides $(V_{GS} - V_T)$.

So,

$$I_D = \frac{K_n}{2} \frac{[(V_{GS} - V_T) + x(t)]^2}{1 + \theta(V_{GS} - V_T + x(t))}$$

$$\& V_{GS} - V_T = \Delta V \quad \dots \dots \dots \text{Large signal}$$

$$X(t) \quad \dots \dots \dots \text{Small signal}$$

$$I_D = \frac{K_n}{2} \frac{[x(t) + \Delta V]^2}{\theta(x(t) + \Delta V) + 1}$$

$$V_o = I_D R_L \Rightarrow V_o = \frac{K_n R_L}{2} \frac{(x(t) + \Delta V)^2}{1 + \theta(x(t) + \Delta V)} \text{ put } \frac{K_n R_L}{2} = K$$

$$\theta \ll 1 \text{ So } (x(t) + \Delta V) \text{ is also small} \Rightarrow \frac{1}{1 + x} = 1 - \frac{x}{2}$$

$$\frac{1}{1 + \theta(x(t) + \Delta V)} \approx 1 - \frac{\theta(x(t) + \Delta V)}{2}$$

$$V_o = K(x(t) + \Delta V)^2 \left(1 - \frac{\theta(x(t) + \Delta V)}{2} \right)$$

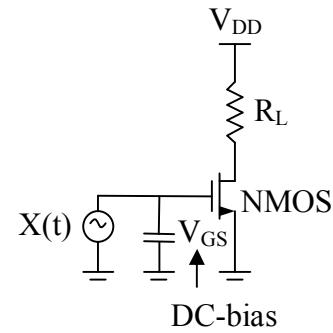
$$V_o = K(x(t) + \Delta V)^2 - (x(t) + \Delta V)^3 \frac{K\theta}{2}$$

$$V_o = K\Delta V^2 - \frac{K\theta}{2} \Delta V^3 + \left(2K\Delta V - \frac{3K\theta}{2} \Delta V^2 \right) x(t) + \left(K - \frac{3K\theta}{2} \Delta V \right) x^2(t) - \frac{K\theta}{2} x^3(t)$$

----- (2)

Small signal components

$$V_o = \left(2K\Delta V - \frac{3K\theta}{2} \Delta V^2 \right) x(t) + \left(K - \frac{3K\theta}{2} \Delta V \right) x^2(t) - \frac{K\theta}{2} x^3(t) \quad \dots \dots \dots (3)$$



Comparing (1) & (3)

$$\alpha_1 = 2K\Delta V - \frac{3K\theta}{2}\Delta V^2, \quad \alpha_2 = K - \frac{3K\theta}{2}\Delta V, \quad \alpha_3 = -\frac{K\theta}{2}$$

$$IP3 = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|} = \sqrt{\frac{4}{3} \times \frac{\frac{2K\Delta V - \frac{3}{2}K\theta\Delta V^2}{K\theta}}{\frac{2}{2}}} = \sqrt{\frac{8}{3} \left(\frac{2\Delta V}{\theta} - 3\Delta V^2 \right)}$$

$$IP3 = \sqrt{\frac{8}{3} \frac{2\Delta V}{\theta}} = \sqrt{\frac{16}{3} \frac{\Delta V}{\theta}} \quad \text{As } \theta \ll 1 \quad 3\Delta V^2 \text{ can be ignored.}$$

$$IP3 = \sqrt{\frac{16}{3} \frac{\Delta V}{\theta}} = \sqrt{\frac{16}{3} \frac{(V_{GS} - V_T)}{\theta}} \quad \dots \dots \dots \quad (4) \quad \therefore I_D = \frac{\mu_o C_{ox}}{2} \frac{W}{L} (V_{GS} - V_T)^2$$

Put $\Delta V = 0.2V$, $\theta = 0.1$

$$IP3 = \sqrt{\frac{16}{3} \frac{0.2}{0.1}} = 3.27 \text{ Volts}$$

$$IP3(\text{dBm}) = 10 \log \left[\left[\left(\frac{3.27}{\sqrt{2}} \right)^2 \cdot \frac{1}{50} \right] / \text{1mW} \right] \cong 20 \text{ dBm}$$

$$g_m = \frac{\partial I_D}{V_{GS}} = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T) \\ \frac{g_m}{I_D} = \frac{2}{V_{GS} - V_T} = \frac{2}{\Delta V} \quad \text{This is just an approximation with } I_D \& g_m$$

$$\frac{g_m}{I_D} = \frac{2}{V_{GS} - V_T} = \frac{2}{\Delta V}$$

$$\Rightarrow IP3 \cong \sqrt{\frac{32}{3\theta} \sqrt{\frac{I_D}{g_m}}} \quad \therefore g_m = \sqrt{\mu_n C_{ox} \frac{W}{L} I_D}$$

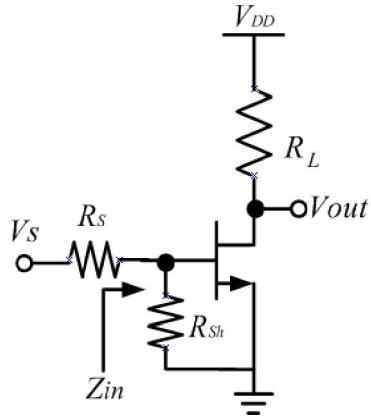
c).

- To increase IIP3 $I_D \uparrow$ (high power) or $g_m \downarrow$ (high noise)
- g_m also depends upon I_D , so when $I_D \uparrow$ $g_m \uparrow$ but at that rate $\propto \sqrt{I_D}$
- When W increases $g_m \uparrow$ for same I_D (Power consumption) so this decreases IP3
- The above observations are for long-channel. But for short channel $(4) \Rightarrow (V_{GS} - V_T) \uparrow$ then $I_D \uparrow$ any how.

Problem-2.3 (Tutorial)

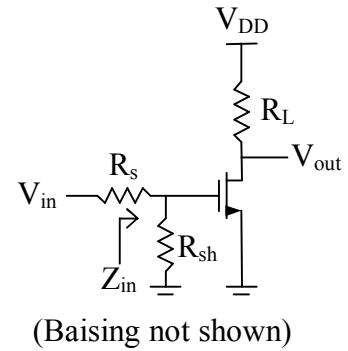
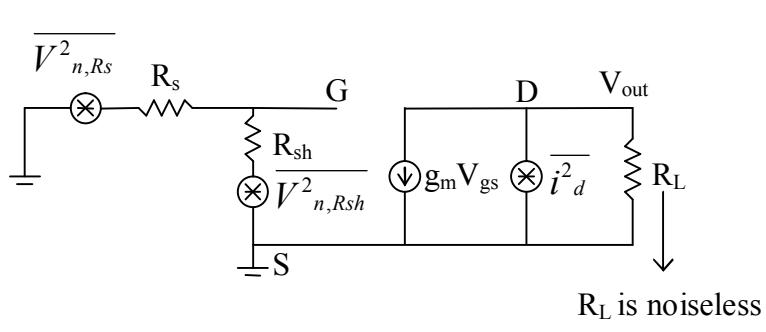
It is preferred in current RF designs that the input of LNA be matched to 50Ω (Razavi, Pg168). The easiest way is to shunt the gate with a resistor of 50Ω .

- Calculate the gain, input impedance and NF in absence of gate noise. Assume that $R_{sh}=R_L$ for NF derivation.
- What are the disadvantage of shunt resistor with reference to gain and NF?



Solution:

a). (Please read assumption in problem statement carefully)



$$F = \frac{\text{Total output noise power}}{\text{Output noise due to input source}}$$

$$\overline{V^2_{m,Rs}} = 4KTR_s \Delta f$$

$$\text{Gain} \Big|_{\text{Gate}} = -g_m R_L$$

$$\overline{V^2_{m,Rsh}} = 4KTR_{sh} \Delta f$$

$$A = g_m R_L \left(\frac{R_{sh}}{R_s + R_{sh}} \right) \text{ for } R_{sh} = R_s$$

$$\overline{i^2_d} = 4KT \gamma g_m \Delta f$$

$$A = -g_m \frac{R_L}{2}$$

Using superposition, considering one at a time and shorting / opening other sources.

$$\overline{V^2_{on,Rs}} = \overline{V^2_{n,Rs}} \times g_m^2 R_L^2 \times \left(\frac{R_{sh}}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2_{on,Rsh}} = \overline{V^2_{n,Rsh}} \times g_m^2 R_L^2 \times \left(\frac{R_s}{R_s + R_{sh}} \right)^2$$

$$\overline{V^2_{no,d}} = \overline{i^2_d} \times R^2_L$$

$$F = \frac{\overline{V^2_{on,Rs}} + \overline{V^2_{on,Rsh}} + \overline{V^2_{no,d}}}{\overline{V^2_{on,Rs}}} = 1 + \frac{\overline{V^2_{on,Rsh}} + \overline{V^2_{o,d}}}{\overline{V^2_{on,Rs}}}$$

$$F = 1 + \frac{4KTR_{sh}\Delta f \times \frac{g_m^2 R^2_L \times R^2_{sh}}{(R_s + R_{sh})^2}}{4KTR_s\Delta f \times \frac{g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}} + \frac{4KT\gamma g_m \Delta f \times R^2_L}{4KTR_s\Delta f \times \frac{g_m^2 R^2_L \times R^2_s}{(R_s + R_{sh})^2}}$$

In case of impedance match $R_s = R_{sh}$

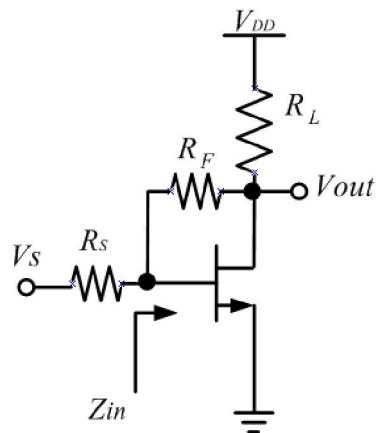
$$F = 1 + 1 + \frac{R^2_L \gamma g_m}{R_s \times \frac{g_m^2 R^2_s \times R^2_L}{4R^2_s}} = 2 + \frac{R^2_L \gamma g_m}{R_s \times \frac{g_m^2 \times R^2_L}{4}} = 2 + \frac{4\gamma}{g_m R_s}$$

b).

- Poor Noise Figure
- Input signal attenuated by voltage divider
- R_{sh} adds extra noise.
- At high frequency, shunt L is needed to tune out C_{gs}
- Reduced gain

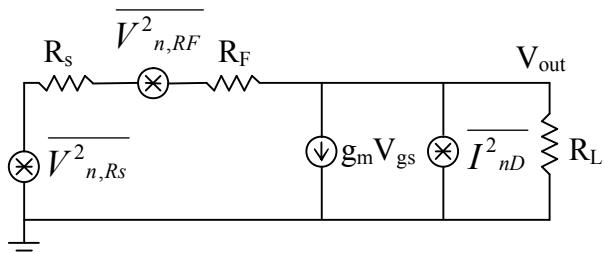
Problem-2.4 (Tutorial)

Another approach to get 50Ω input impedance match is shunt feedback amplifier shown below.

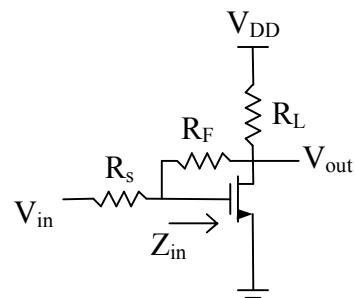


- a) Calculate the gain, input impedance and NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- b) What are the disadvantage of shunt feedback amplifier with reference to gain and NF?

Solution:



(Equivalent noise model ignoring gate noise), RL is noiseless



(Basing not shown)

$$\overline{I^2_{nD}} = 4KT\gamma g_m \Delta f, \overline{V^2_{n,RS}} = 4KTR_s \Delta f$$

$$NF = \frac{\overline{V^2_{n,out}}}{A^2_{v,tot} \overline{V^2_{RS}}} = \frac{\text{Total input noise power}}{\text{Output noise power due to input source}}$$

Here $A_{v,tot}$ = Gain from V_{in} to V_{out}

Again using superposition theorem

$$NF = \frac{\overline{V^2_{n,out}}}{A^2_{v,tot} \overline{V^2_{RS}}} = \frac{\overline{V^2_{n,RS,out}} + \overline{V^2_{n,RF,out}} + \overline{V^2_{n,D,out}}}{A^2_{v,tot} \overline{V^2_{RS}}}$$

Gain Calculation

$$V_{in} = i_{in}(R_s + R_F) + V_{out}$$

$$V_{out} = (i_{in} - g_m V_{gs}) R_L$$

$$V_{gs} = i_{in} R_F + V_o$$

$$A_{v,tot} = \frac{V_{out}}{V_{in}} = \frac{R_L(1 - g_m R_L)}{R_s + R_F + R_L + g_m R_s R_L}$$

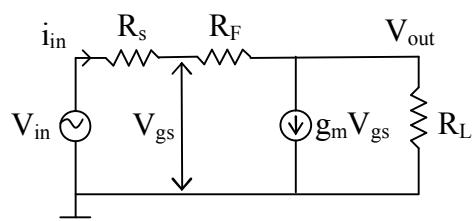
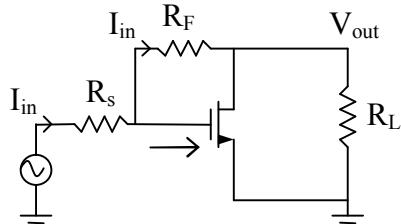
If $R_F \gg R_s$ & $g_m R_F \gg 1$

$$A_{v,tot} = \frac{-g_m R_L}{\frac{R_s}{R_F} + 1 + R_L + \frac{1 + g_m R_s}{R_F}} \approx -g_m R_L$$

$$A_{v,tot} \approx -g_m R_L$$

$$\text{Also } Z_{in} = \frac{R_F + R_L}{1 + g_m R_L}$$

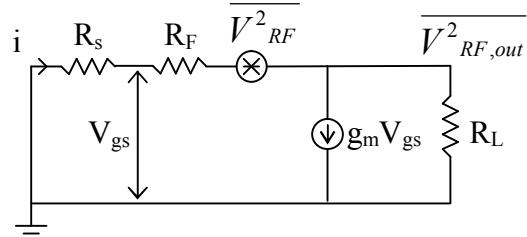
By ignoring C_{gs} , we have considered real part only.



For source resistance

$$\overline{V^2_{nRS,out}} = A_{v,tot}^2 \overline{V^2_{nRS}} \quad \text{-----(1)}$$

For feedback resistance



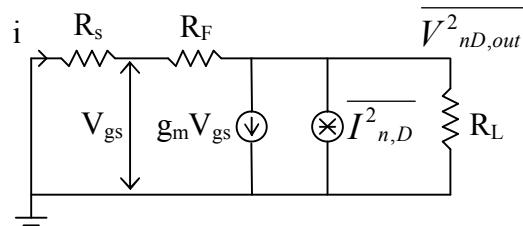
$$V_{gs} = -iR_s = iR_F - V_{RF} + V_{RF,out}$$

$$V_{RF,out} = R_L(i - g_m V_{gs})$$

$$V_{RF,out} = V_{RF} \frac{1}{1 + \frac{R_S + R_F}{R_L(1 + g_m R_s)}} = V_{RF} \frac{R_L}{R_F} (1 + g_m R_s)$$

$$\overline{V^2_{n,RF,out}} = \overline{V^2_{n,RF}} \left[\frac{R_L}{R_F} (1 + g_m R_s) \right]^2 \quad \text{-----(2)}$$

Similarly



$$\frac{V_{nD,out}}{R_L} + \overline{I_{nD}} + g_m V_{gs} + \frac{V_{nD,out}}{R_S + R_F} = 0$$

$$V_{gs} = R_S \frac{V_{nD,out}}{R_S + R_F}$$

$$V_{nD,out} = \frac{I_{nD}}{\frac{1}{R_L} + \frac{1}{R_S + R_F} + \frac{g_m R_S}{R_S + R_F}} \approx I_{nD} R_L$$

So,

$$\overline{V^2_{nD,out}} = \overline{I^2_{nD}} R_L^2 \quad \text{-----(3)}$$

Combaining (1) (2) & (3)

$$NF = 1 + \frac{\overline{V^2}_{n,RF} \left[\frac{R_L}{R_F} (1 + g_m R_S) \right]^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}} + \frac{\overline{I^2}_{nD} R_L^2}{A_{v,tot}^2 \overline{V^2}_{n,RS}}$$

$$A_{v,tot} = -g_m R_L, \quad \overline{V^2}_{n,RS} = 4KTR_S \Delta f, \quad \overline{V^2}_{M,RF} = 4KTR_F \quad \& \quad \overline{I^2}_{nD} = 4KT\gamma g_m$$

$$NF = 1 + \frac{R_S}{R_F} \left(1 + \frac{1}{g_m R_S} \right)^2 + \frac{\gamma}{g_m R_S}$$

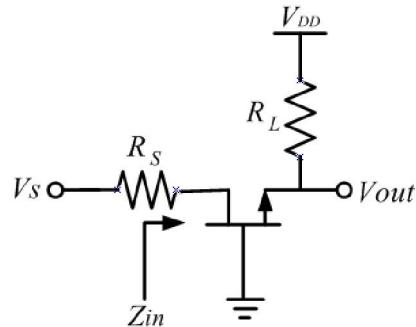
b).

$NF \downarrow$ $g_m R_S \uparrow$ & $R_F \uparrow$ usually $R_S = 50\Omega$

- Better performance than CS amplifier
- R_F induces noise
- At $f \uparrow$ need shunt inductor to tune out C_{gs}
- Broadband Amp @ Lower frequency
- To make $NF \downarrow$ $R_F > R_S$ $g_m R_S \gg 1$

Problem-2.5 (HW)

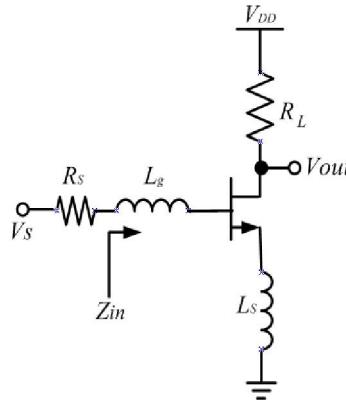
Common gate amplifier also offers 50Ω input impedance match and solves the input matching problem.



- c) Calculate the gain, input impedance and NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- a) What are the disadvantage of common gate amplifier with reference to gain and NF?

Problem-2.6 (Tutorial)

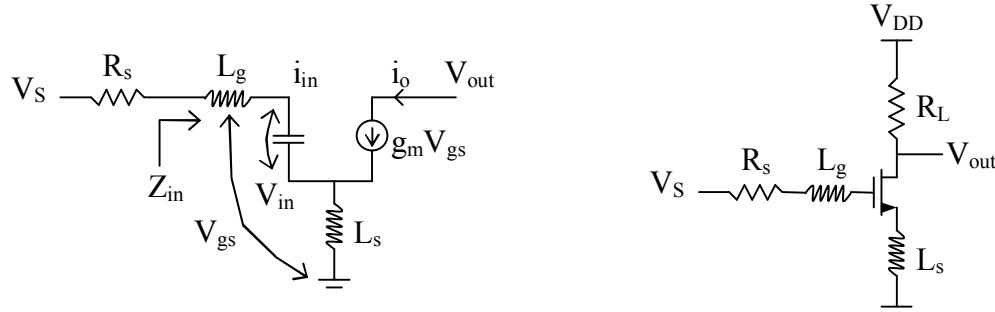
The disadvantages of three types of amplifiers in Problem-3, 4 & 5 can be circumvented by using source degenerated LNA shown below.



- Calculate the input impedance. This inductor source degenerated amplifier presents a noiseless resistance for 50Ω for input power match. How we can cancel the imaginary part of complex input impedance so that the LNA presents 50Ω real input resistance at input port.
- Calculate the NF in absence on gate noise. Neglect gate drain and gate to bulk and gate to source capacitance.
- C_{gd} bridges the input and output ports. The reverse isolation of this LNA is very poor. Why reverse isolation is important? Suggest the modification to improve reverse isolation.

Solution:

a).



(Biasing not shown)

From model above we can write

$$V_{in} = i_{in} \left(j\omega L_g + j\omega L_s \right) + i_{in} \left(\frac{1}{j\omega C_{gs}} \right) + i_o j\omega L_s \quad \text{-----(1)}$$

$$i_o = g_m V_{gs} = g_m i_{in} \times \frac{1}{j\omega C_{gs}} \quad \text{-----(2)}$$

Substituting (2) in (1)

$$V_{in} = i_{in} \left[j\omega (L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}} \right]$$

$$Z_{in} = \frac{V_{in}}{I_{in}} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

For matching $L_g + L_s$ are canceled out by C_{gs} . So at frequency of interest

$$\omega_o(L_g + L_s) = \frac{1}{\omega_o C_{gs}} \Rightarrow \omega_o^2 = \frac{1}{(L_g + L_s)C_{gs}}$$

$$\text{And } R_S = 50\Omega = \frac{g_m}{C_{gs}} L_s$$

Notes:

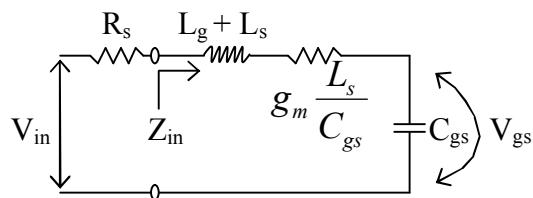
- a). L_s is typically small and may be realized by bond wire.
- b). L_g can be implemented by spiral/external inductor.

b).

From part a)

$$Z_{in} = j\omega(L_g + L_s) + \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}$$

We can draw this circuit as



Here

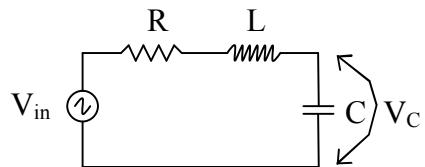
$$Q_{in} = \frac{\omega_o(L_g + L_s)}{R_S + \frac{g_m L_s}{C_{gs}}} = \frac{\omega_o(L_g + L_s)}{R_S + W_T L_s}$$

$$\therefore \omega_T \cong \frac{g_m}{C_{gs}}$$

$$Q_{in} = \frac{1}{\omega_o \left(R_S + \frac{g_m L_s}{C_{gs}} \right) C_{gs}} \quad \text{For match load } R_S = \frac{g_m L_s}{C_{gs}}$$

Reference:

For series RLC Circuit



$$Q_s = \frac{1}{R} \sqrt{\frac{L}{C}} = \frac{\omega_o L}{R} = \frac{1}{\omega_o R C}$$

$$\text{And } V_C = Q_s V_{in}$$

For problem (1)

$$\omega_T \cong \frac{g_m}{C_{gs}} \quad \text{Unity gain frequency}$$

for current

$$Q_{in} = \frac{1}{2\omega_o R_s C_{gs}}$$

Gain

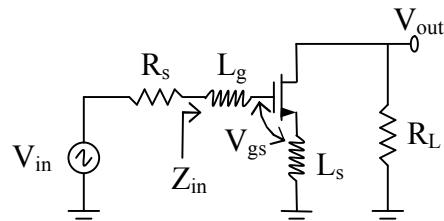
$$V_{gs} = Q_{in} V_{in}$$

$$g_m = \frac{V_{out}}{V_{gs}}$$

$$G_m = \frac{V_{out}}{V_{in}} = \frac{V_{gs} g_m}{V_{in}} = Q_{in} g_m$$

$$G_m = Q g_m$$

So, Gain = $-G_m R_L$ or $-Q_{in} g_m R_L$ & $G_m = Q_{in} g_m$

**Noise Figure:**

$$NF = \frac{\text{Total noise power at output}}{\text{noise power at output due to input source}}$$

For this calculation we ignore channel noise.

$$F = \frac{\overline{V^2}_{nRS,OUT} + \overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}} = 1 + \frac{\overline{V^2}_{nD,OUT}}{\overline{V^2}_{nRS,OUT}}$$

$$\overline{V^2}_{nD,OUT} = \overline{i^2}_{n,D} R_L^2 \quad \overline{i^2}_{n,D} = 4KT\gamma g_m \Delta f$$

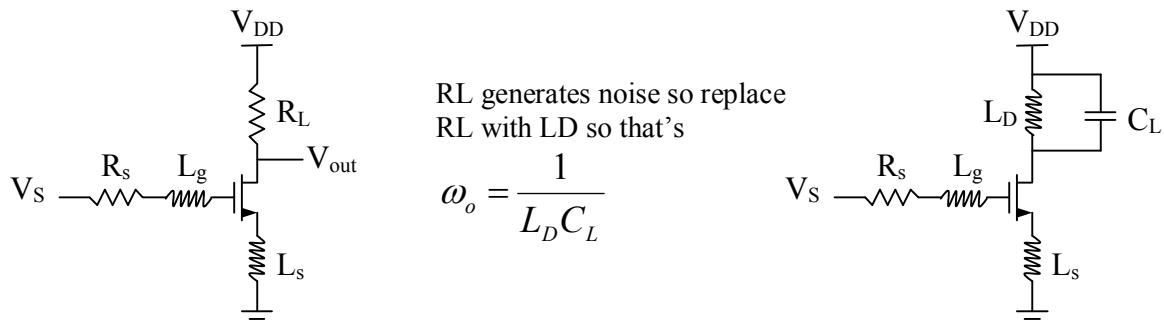
$$\overline{V^2}_{nRS,OUT} = \overline{V^2}_{n,RS} G_m^2 R_L^2 \quad \overline{V^2}_{n,RS} = 4KTR_S \Delta f \quad \& \quad G_m = Q_{in} g_m$$

$$F = 1 + \frac{\overline{i^2}_{n,D} R_L^2}{\overline{V^2}_{n,RS} Q_{in}^2 g_m^2 R_L^2} \quad \overline{i^2}_{n,D} = 4KT\gamma g_m, \quad \overline{V^2}_{n,RS} = 4KTR_S$$

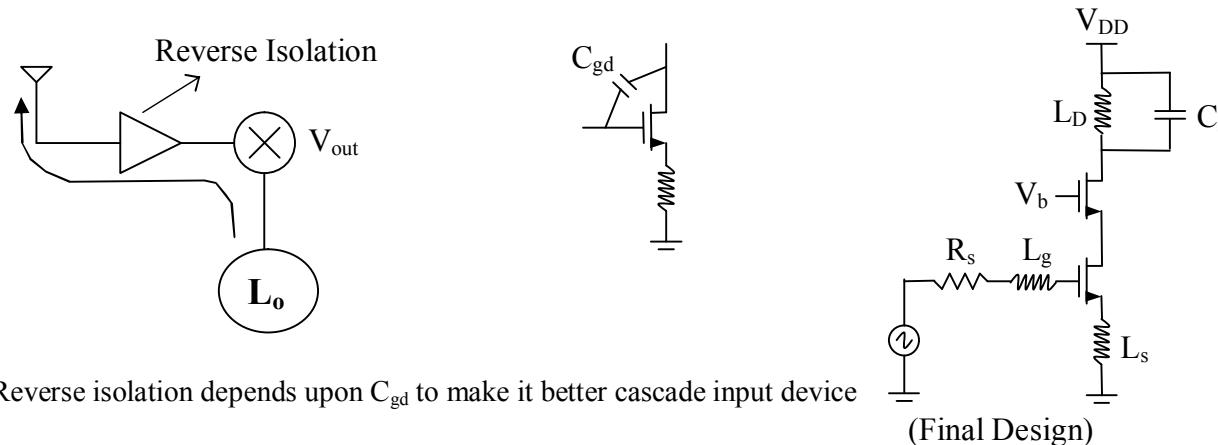
$$F = 1 + \frac{\gamma}{g_m R_s Q_{in}^2}$$

Notes:

- Very good NF value
- Narrow band matching
- NF \downarrow with Q^2
- The Q is dependent upon $L_g + L_s$, L_s is small so Q depend upon L_g

C). Draw Backs**i).**

The C_L can be the input capacitance of mixer or filter.

ii).**Problem-2.7 (HW)**

Fill-in the Table below, use the data from Problem-2.4, 2.5, 2.6 and 2.7

Type of LNA	Z_{in}	Noise Factor	Gain	NF (dB)
Shunt Resistor	R_{sh}	$2 + \frac{4\gamma}{g_m R_s}$	$\frac{-g_m R_L}{2}$	
Common Gate				
Shunt Feedback				
Source Degenerated				

- Calculate the NF for all above amplifiers. Assume $\gamma=2$, $g_m = 20mS$, $R_s = 50\Omega$, $R_f = 500\Omega$, and $Q_{in} = 2$.
- Which is best topology for Narrow Band LNA design at high frequency?

Problem-2.8 (Tutorial)

Real Design: We will design the inductor source degenerated LNA shown in Fig below to meet the specification outlined for IEEE802.11 (b) standard. The first cut approximate values are calculated as a starting point for simulation. In **LAB3: Design of LNA** you will take the same design and modify these component values to meet the specification.

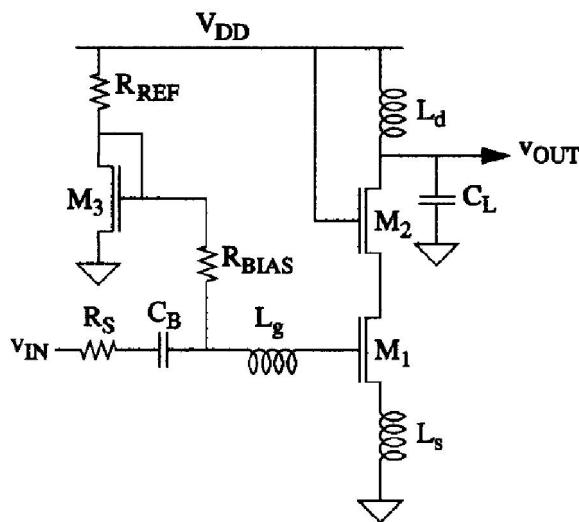
LNA Specification:

NF < 2.5 db, Gain > 15dB, IP3 > -5dBm, Centre Frequency = 2.4 GHz

S11 < -20dB, S22 < -10dB, Load Capacitance = 1pF

Technology Parameters for 0.35um CMOS:

$$L_{\text{eff}} = 0.35 \mu m, \quad \mu_n C_{\text{ox}} = 170 \mu A/V^2, \quad C_{\text{ox}} = 4.6 mF/m^2, \quad \mu_p C_{\text{ox}} = 58 \mu A/V^2, \quad \gamma = 2$$

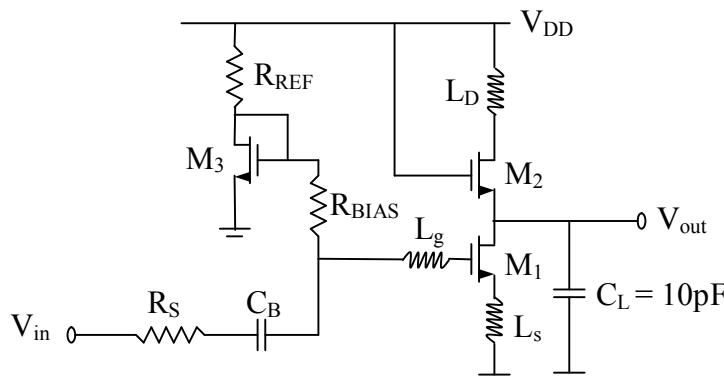


Solution:

$$\text{Technology: } \left\{ \begin{array}{l} \mu_o C_{\text{ox}} = 170 \mu A/V^2, \mu_p C_{\text{ox}} = 58 \mu A/V^2, 0.35 \mu m \\ \text{CMOS}, C_{\text{ox}} = 4.6 mF/m^2, \gamma = 2, f_o = 2.4 GHz \\ (\text{IEE802.11(b) Standard}), \mu_o C_{\text{ox}} = 170 \mu A/V^2 \end{array} \right\}$$

Design Parameters

$$R_S = 50 \Omega, V_{DD} = 3.3V, C_L = 10 pF, NF < 2.5 dB$$



Component Description

L_s – Matches input impedance

L_g – Set the Resonant Frequency $f_0 = 2.4$ GHz

M_3 – Biasing transistor which forms current mirror with M_1

L_d – Tuned output increases the gain and also work as band pass filter with C_L

M_2 – Isolate tuned input and tuned output increases reverse isolation, also reduces the effect of Miller capacitance C_{gd}

C_B – BC Blocking capacitor chosen to have negligible reactance at $f_0 = 2.4$ GHz

R_{BIAS} – Large enough so that its equivalent current noise is small enough to be ignored. (Don't consider it as voltage noise source. Why??)

Design Procedure

Size of M_1 :

We will not go for global minimum noise figure as given by two-point noise theory (See lecture on LNA Slide # 10)

$$G_{opt} = \alpha \omega C_{gs} \sqrt{\frac{\delta}{5\gamma} (1 - |C|^2)} = \frac{1}{50\Omega} \quad \text{-----(1)}$$

$$C_{gs} \approx 4 \text{ pF} \Rightarrow W_{M1} \approx 4 \text{ mm}!! \text{ (not possible)}$$

Solution:

A & B are from Thomas. H. Lee book (LNA Chapter)

LNA NF will be optimized for given Power

* It will not be best NF globally.

$$W_{opt} = \frac{1}{3\omega_o L_{eff} C_{ox} R_s}$$

$$F_{min,p} = 1 + 2.4 \frac{\gamma}{\alpha} \frac{\omega}{\omega_T} \Rightarrow F_{min,p} = 1 + 5.6 \frac{\omega}{\omega_T} \quad \text{-----(A)}$$

From (1) we can derive

$$F_{min,p} = 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_T} \sqrt{\gamma C (1 - |C|^2)} = 1 + 2.3 \frac{\omega}{\omega_T}$$

$$F_{min,p} = 1 + 2.3 \frac{\omega}{\omega_T} \quad \text{-----(B)}$$

(a) is minimum NF for a given power consumption.

(b) is global minimum noise figure.

The difference is usually 0.5dB to 1dB (no big deal for Lower Power)

Step - 1:

$$I_1 = I_2 = 5mA \text{ (Low Power consumption)}$$

Step - 2:

$$W_{M1} = \frac{1}{3WL_{eff}C_{ox}R_S}$$

$$W_{M1} = \frac{1}{3 \times 0.35\mu \times 4.6m \times 50 \times \omega_o} \quad \left. \begin{array}{l} R_S = 50\Omega, C_{ox} = 4.6mF/m^2, \\ \mu_n C_{ox} = 170\mu A/V, L_{eff} = 0.35\mu m, \\ \omega_o = 2\pi f_o, f_o = 2.4GHz \end{array} \right\}$$

$$W_{M1} = 3.9 \times 10^{-4}$$

$$W_{M1} = 3.9 \times 10^{-4} = 390\mu m$$

Step - 3:

$$C_{gs1} = \frac{2}{3} W_{M1} L_{eff} C_{ox}$$

$$C_{gs1} = \frac{2}{3} \times 390\mu \times 0.35\mu \times 4.6m = 0.41pF$$

$$g_{m1} = \sqrt{2\mu_n C_{ox} \left(\frac{W}{L} \right)_{M1} I_{DM1}}$$

$$g_{m1} = \sqrt{2 \times 170\mu \times \left(\frac{390}{0.35} \right) \times 5m} = 43mA/V$$

$$\omega_T \approx \frac{g_{m1}}{C_{gs1}} = \frac{43mA/V}{0.41pF} = 104Grad/Sec$$

Assuming $\gamma = 2$

$$\text{Now } F_{\min} = 1 + 5.6 \frac{\omega_o}{\omega_T}$$

$$F_{\min} = 1 + 5.6 \frac{2\pi 2.4G}{104G} \approx 2.55dB$$

$$NF \approx 2.55dB$$

It's very close to what we derive, if the value is higher we can increase I_D to increase ω_T and hence low NF on expense of power.

Step - 4:

Source and gate inductance such that they cancel C_{gs} and set 50Ω input impedance

$$\omega_o = 2\pi f_o = 2\pi 2.4 = 15 \text{Grad}/\text{Sec}$$

From previous problem

$$R_s = R_{Transformed} = g_m \frac{L_s}{C_{gs}} = L_s \omega_T$$

$$L_s = \frac{R_s}{\omega_T} = \frac{50}{100G} \cong 0.5nH$$

$L_s = 0.5nH$ can be implemented using Band wire.

$$\text{Now } L_g + L_s = \frac{1}{(\omega_o^2 C_{gs})}$$

$$L_g + L_s = \frac{1}{(15G)^2 \times 0.41pF} = 10.81nH$$

$$L_g \approx 10nH$$

Step - 5:

$$L_d = \frac{1}{\omega_o^2 C_L} \quad \because C_L = 1pF$$

$$L_d = \frac{1}{(15G)^2 \times 1pF} \cong 4.4nH$$

$$L_d = 4.4nH$$

Step - 6:

Size of M3 is chosen to minimize power consumption

$$W_{M3} = 70\mu m, \quad R_{REF} = 2K\Omega \Rightarrow I_3 = 0.6mA$$

$R_{BIAS} = 2K\Omega$ (Large enough so that it's equivalent current noise can be neglected)

$$C_B = 10pF \quad (X_C \approx 6.6\Omega \text{ so good value @ 2.4G} \quad X_B = \frac{1}{2\pi f_o C_B} = 6.6\Omega)$$

Step - 7:

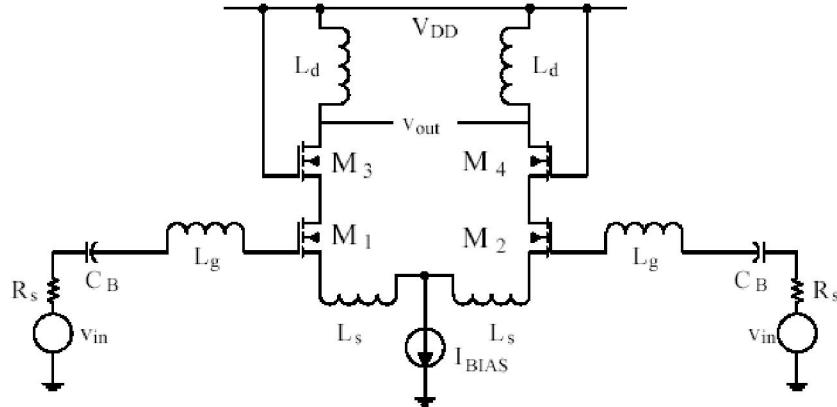
Size M2 = M3

So that they can have shared Drain Area..

(Note: We will simulate same design in LAB # 2)

Problem-2.9 (Point to Ponder):

Connecting two Inductor source degenerated LNA as shown in Figure make the differential LNA. Differential LNA has many advantages: higher common mode rejection ratio, less sensitivity to the ground inductance variation L_s compared to single ended counterpart..



- a) Compare intuitively the NF of single ended and differential if both have same power consumption.
- b) If low power is not parameter of interest, which LNA has lower NF?

Instructions:

For hand calculation of NF you can ignore the gate noise of the device and noise generated by the load resistance R_L .

Acknowledgement: The major part of this tutorial was developed, while author was employed by Linkoping University, Sweden.

PA Design

CHAPTER 2

POWER AMPLIFIER

2.0 Introduction

The main characteristics of an amplifier are Linearity, efficiency, output power, and signal gain. In general, there is a trade off between these characteristics. For example, improving amplifier's linearity will degrade its efficiency. Therefore knowing the importance degree of each one of these characteristics is an essential step in designing an Amplifier. This can be jugged based on the application. As an example high output power Amplifier is used in the transmitter side of a transceiver, whereas high linear amplifier used in the receiver side.

An amplifier is said to be linear if it preserves the details of the signal waveform, that is to say,

$$V_o(t) = A \cdot V_i(t) \quad (2.1)$$

where, V_i and V_o are the input and output signals respectively, and A is a constant gain representing the amplifier gain. But if the relationship between V_i and V_o contains the higher power of V_i , then the amplifier produces nonlinear distortion.

The amplifier's efficiency is a measure of its ability to convert the dc power of the supply into the signal power delivered to the load. The definition of the efficiency can be represented in an equation form as

$$\eta = \frac{\text{Signal power delivered to load}}{\text{DC power Supplied to output circuit}} . \quad (2.2)$$

For an ideal amplifier, the efficiency is one. Thus, the power delivered to the load is equal to the power taken from the DC supply. In this case, no power would be consumed in the amplifier. In reality, this is not possible, especially in high frequency realm of RF circuits. In many high frequency systems, the output stage and driver stage of an amplifier consumed power in the amplification process.

The gain of the amplifier (G) is equal to the magnitude of the output signal (X_o) over the magnitude of the input signal (X_i) as shown in the equation.

$$G = \frac{X_o}{X_i} . \quad (2.3)$$

G can be voltage, current, or power gain depending on the application.

The output power level plays an important role in evaluating the power amplifier. The power output capability factor, P_{MAX} , is the power output that would be produced with stresses of 1 Volt and 1 Amp on the drain of the field effect transistor (FET). Multiplication of P_{MAX} by the drain voltage and current ratings of a real device produces the maximum output power available from that device.

The power output capability factor is

$$P_{MAX} = \frac{\text{The Maximum Output Power}}{\text{The Peak Drain Voltage} \times \text{The Peak Drain Current}} . \quad (2.4)$$

2.1 Amplifier Classification

Amplifiers are classified according to their circuit configurations and methods of operation into different classes such as A, B, C, and F. These classes range from entirely linear with low efficiency to entirely non-linear with high efficiency. The analysis presented in this chapter assumes piecewise-linear operation of the active device. The majority of this

information is available in Solid State Radio Engineering by Krauss, Bostain, and Raab [1980].

The active device used in this research is the field effect transistor. The reason for choosing this type of transistor is its superior performance in the microwave range. The characteristics of the FET can be described by:

$$i_D = 0 \quad \text{cut-off region,}$$

$$i_D = g_m \cdot (V_{GS} - V_T) \quad \text{active region, (2.5)}$$

$$i_D = \frac{V_D}{R_{on}} \quad \text{saturation region.}$$

The regions of operation are defined by:

$$\text{cut-off region: } V_{GS} < V_T ,$$

$$\text{active region : } V_{GS} \geq V_T \text{ and } i_D < V_D/R_{on} ,$$

$$\text{saturation region: } V_{GS} \geq V_T \text{ and } i_D = V_D/R_{on} .$$

The term “saturation” is used here to denote the region where further increase in gate voltage produces no increase in drain current, that is to say, i_D is independent of V_{GS} .

2.2.1 Class A

The class-A amplifier has the highest linearity over the other classes. It operates in a linear portion of its characteristic; it is equivalent to a current source. As shown in figures.2.1 and 2.2, the configurations of class-A, B, and C amplifiers can be either a push-pull or a single ended tuned version. Figure.2.3 shows the load-line and current waveform for the class-A amplifier. To achieve high linearity and gain, the amplifier’s base and drain dc voltage should be chosen properly so that the amplifier operates in the linear region. The device, since it is on

(conducting) at all times, is constantly carrying current, which represents a continuous loss of power in the device.

As shown in Fig.2.3, the maximum ac output voltage V_{om} is slightly less than V_{DD} and the maximum ac output current I_{om} is equal to I_{dq} . In the inductor-less system, the output voltage V_{om} will not be able to rise above the supply voltage, therefore, the swing will be constrained to $V_{DD}/2$ and not V_{DD} . The drain voltage must have a dc component equal to that of the supply voltage and a fundamental-frequency component equal to that of the output voltage; hence

$$V_D(\mathbf{q}) = V_{DD} + V_{om} \cdot \sin \mathbf{q} \quad . \quad (2.6)$$

The dc power is

$$P_{dc} = V_{DD} \cdot I_{dq} \quad , \quad (2.7)$$

the maximum output power is

$$P_o = \frac{1}{2} \cdot V_{om} \cdot I_{om} \approx \frac{1}{2} \cdot V_{DD} \cdot I_{dq} \quad , \quad (2.8)$$

and the efficiency is

$$\eta = \frac{P_o}{P_{dc}} \cdot 100 = \frac{1}{2} \cdot \frac{V_{om}}{V_{DD}} \cdot 100 \leq 50\% \quad . \quad (2.9)$$

The difference between the dc power and output power is called power dissipation:

$$P_d = P_{dc} - P_o \quad . \quad (2.10)$$

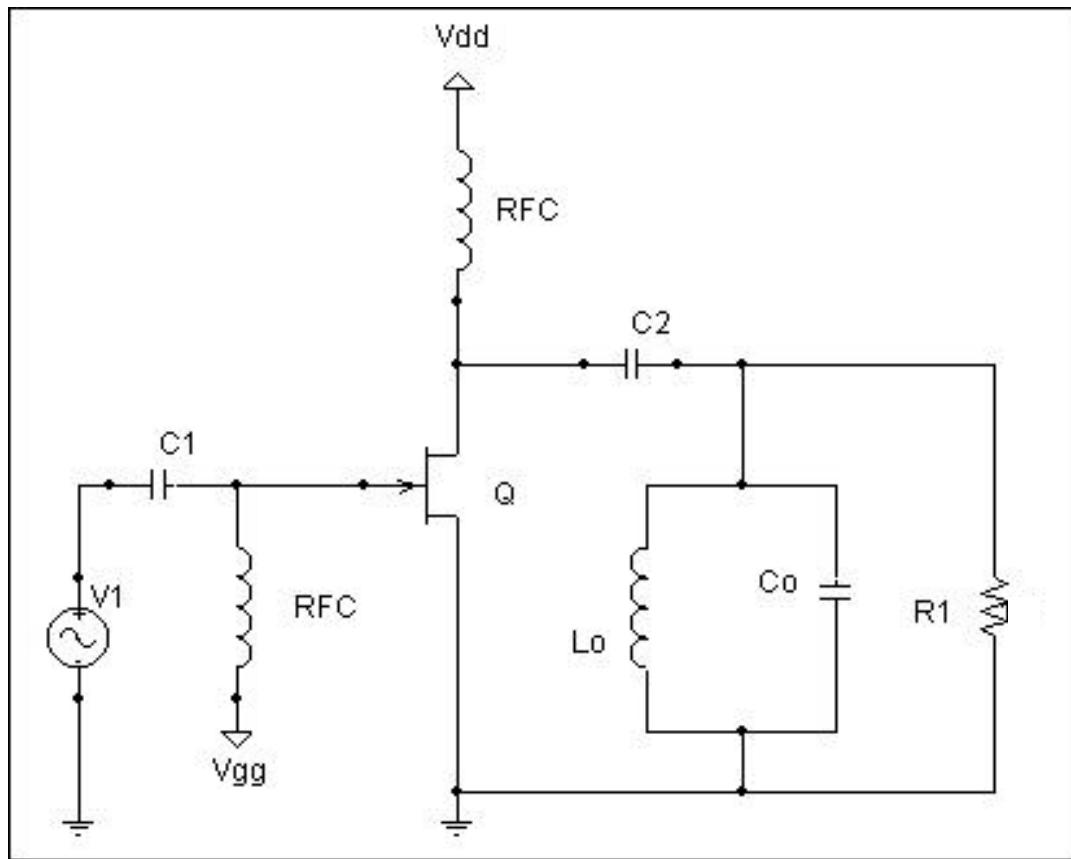
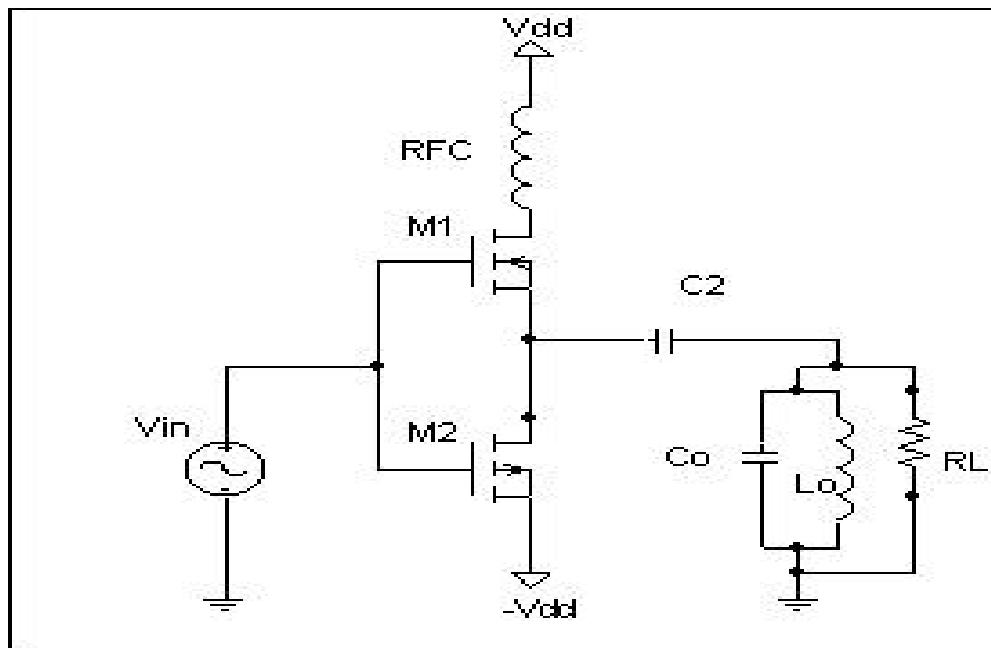
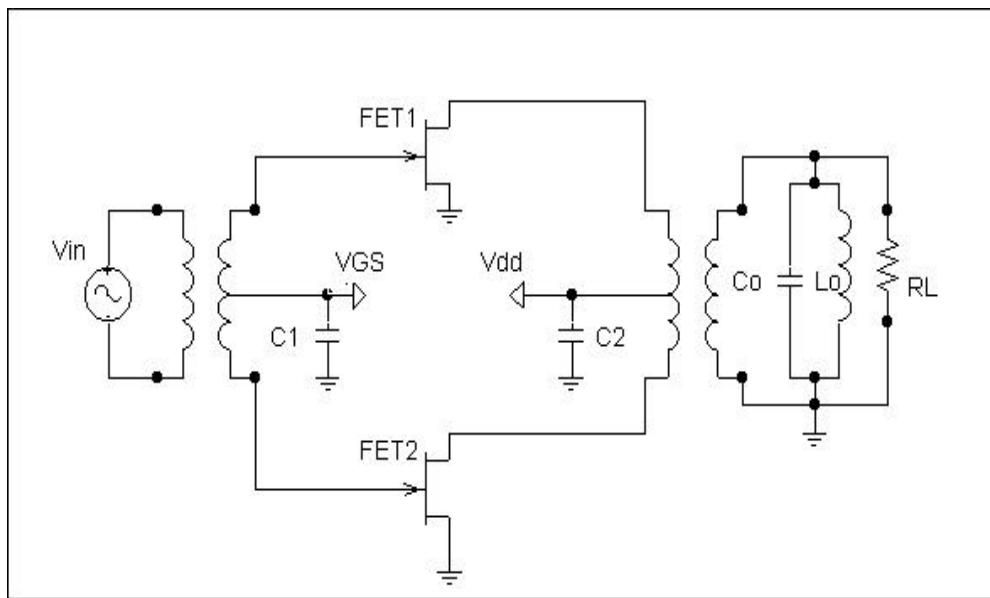


Figure 2.1. Single-ended Power Amplifier (Class A, B, or C)



(a)



(b)

Figure 2.2.

- a. Complementary Push-pull Power Amplifier (Class A, B, or C)*
- b. Transformer-coupled Push-pull Power Amplifier (Class A, B, or C)*

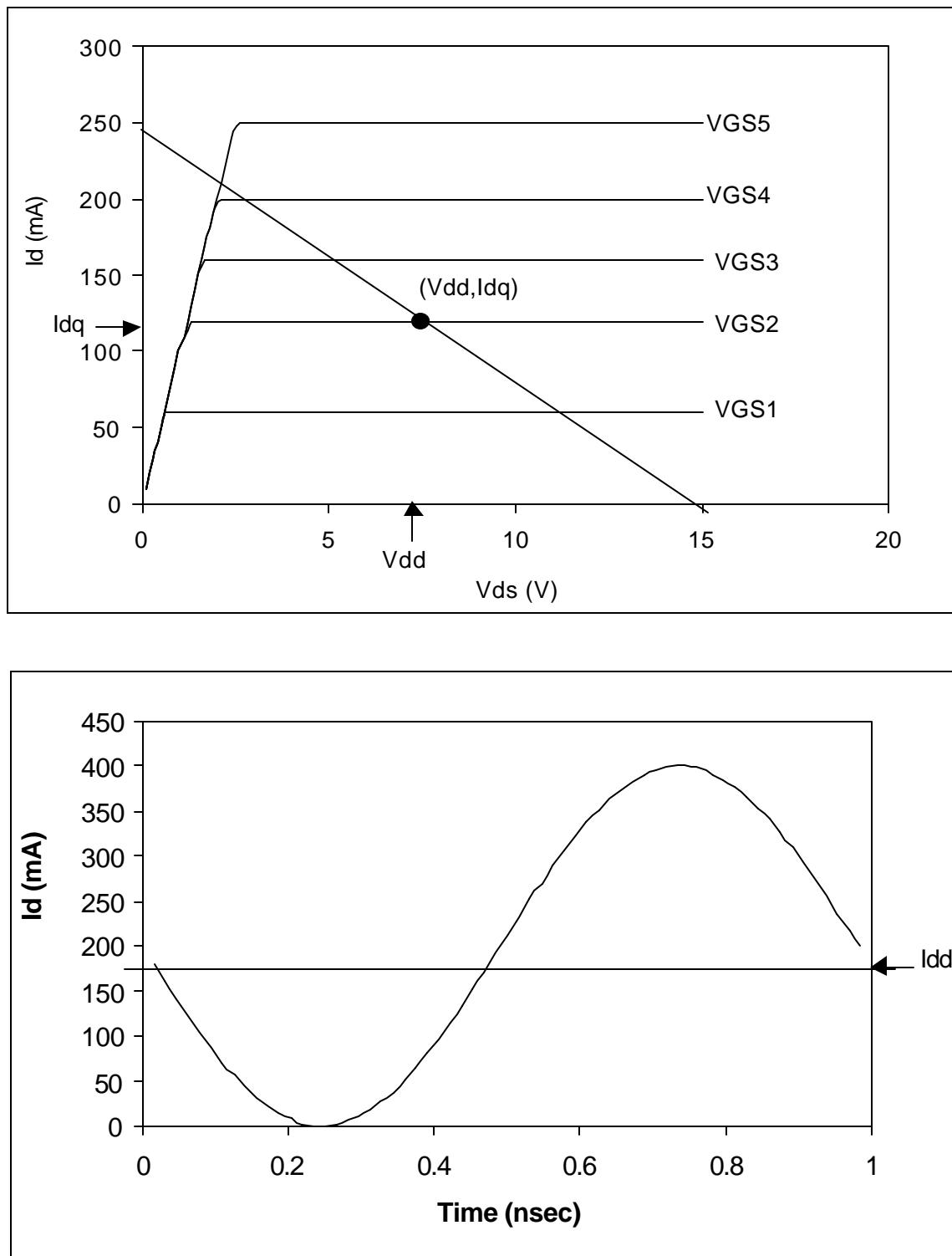


Figure 2.3. Load line and current waveform for the class-A power amplifier

2.2.2 Class B

The class-B amplifier operates ideally at zero quiescent current, so that the dc power is small. Therefore, its efficiency is higher than that of the class-A amplifier. The price paid for the enhancement in the efficiency is in the linearity of the device.

Figure 2.4 shows how the class-B amplifier operates. The output power for the single-ended class-B amplifier is

$$P_o = \frac{1}{2} \cdot I_{om} \cdot V_o . \quad (2.11)$$

the dc drain current is

$$I_{dc} = 2 \frac{I_{om}}{p} , \quad (2.12)$$

the dc power is

$$P_{dc} = 2 \frac{I_{om} \cdot V_{DD}}{p} , \quad (2.13)$$

and the maximum efficiency when $V_{om} = V_{DD}$ is

$$h = \frac{P_o}{P_{dc}} \cdot 100 = \frac{p}{4} \cdot \frac{V_{om}}{V_{DD}} \cdot 100 \leq 78.53% . \quad (2.14)$$

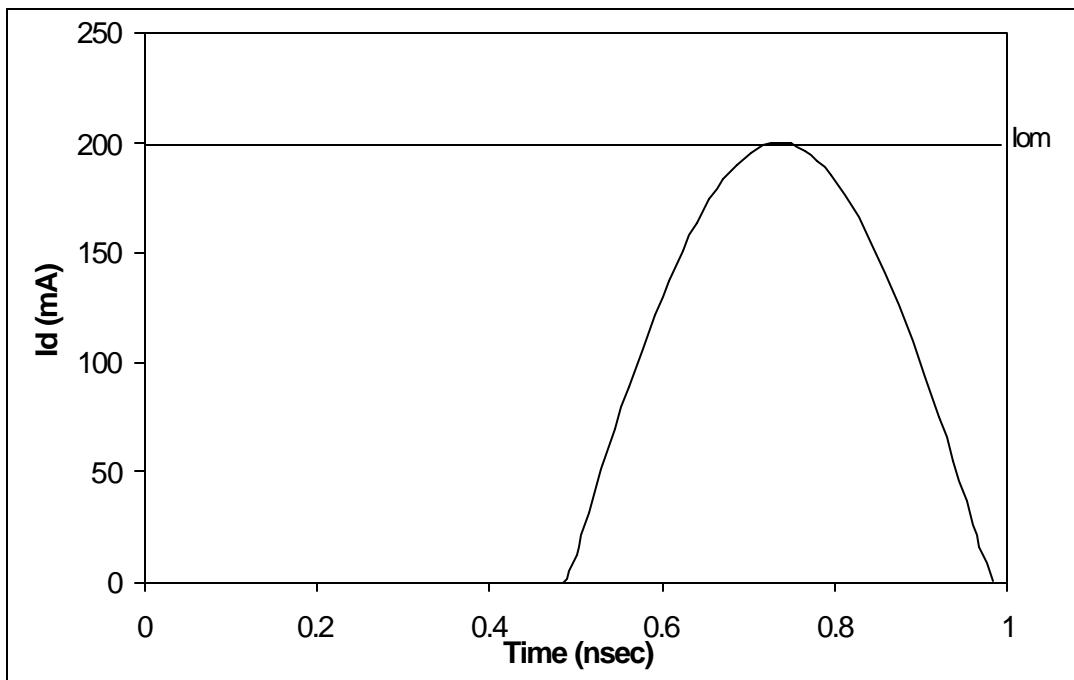
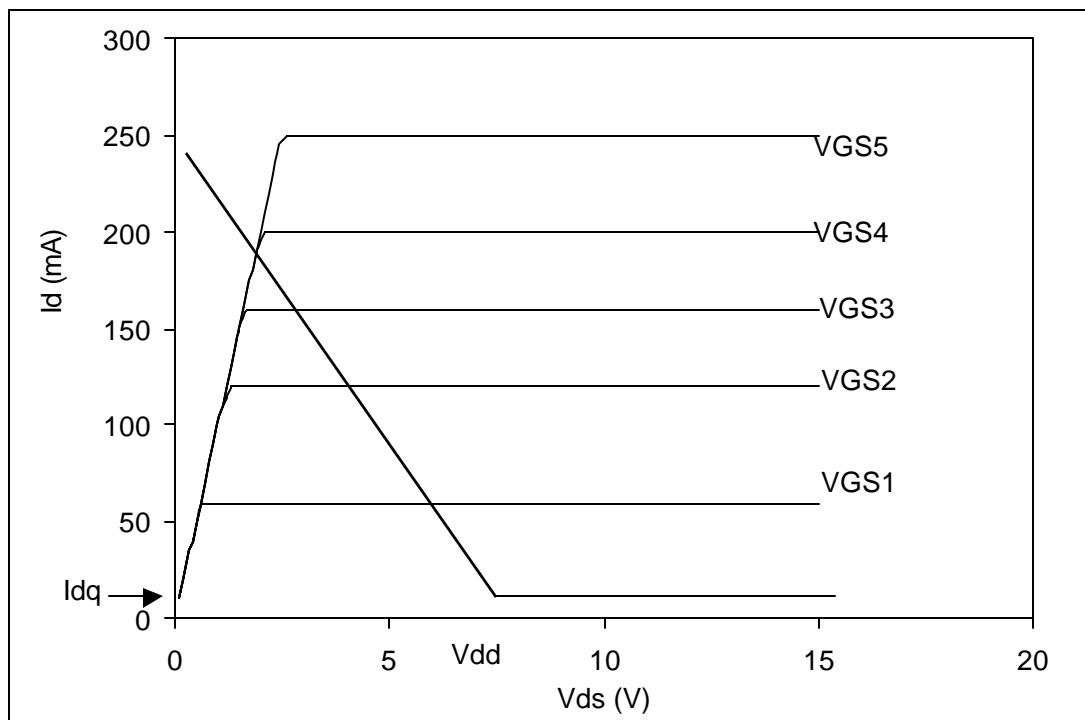


Figure 2.4. Load line and current waveform for the class-B power amplifier

2.2.3 Class AB

The class-AB amplifier is a compromise between class A and class B in terms of efficiency and linearity. The transistor is biased as close to pinch-off as possible, typically at 10 to 15 percent of I_{dss} . In this case, the transistor will be on for more than half a cycle, but less than a full cycle of the input signal.

2.2.4 Class C

The previous classes, A, B, and AB are considered linear amplifier, where the output signal's amplitude and phase are linearly related to the input signal's amplitude and phase. In the application where linearity is not an issue, and efficiency is critical, non-linear amplifier classes (C, D, E, or F) are used.

Class-C amplifier is the one biased so that the output current is zero for more than one half of an input sinusoidal signal cycle. Figure 2.5 illustrates the operation of the class-C amplifier. A tuned circuit or filter is a necessary part of the class-C amplifier.

Classes-A, AB, B, and C amplifiers can be defined in terms of the conduction angle Y as follows:

$$\text{Class of operation} = \begin{cases} \text{A , } y = p \\ \text{B , } y = \frac{p}{2} \\ \text{AB , } \frac{p}{2} < y < p \\ \text{C , } y < \frac{p}{2} \end{cases} \quad (2.15)$$

The conduction angle is

$$Y = \arccos\left(-\frac{I_{dq}}{I_{dd}}\right) . \quad (2.16)$$

The dc current is

$$\begin{aligned} I_{dc} &= \frac{1}{2p} \cdot \int_0^{2p} i_D(\mathbf{q}) d\mathbf{q} = \frac{1}{p} \cdot (I_{dq} \cdot y - I_{dd} \cdot \sin(y)) \\ &= \frac{I_{dd}}{p} \cdot (\sin(y) - y \cos(y)) . \end{aligned} \quad (2.17)$$

Also, the output voltage (V_o) can be obtained in term of Y as

$$\begin{aligned} V_o &= \frac{1}{2p} \cdot \int_0^{2p} i_D(\mathbf{q}) \cdot R \cdot d\mathbf{q} = \frac{R}{2p} \cdot [4I_{dq} \cdot \sin(y) + 2I_{dd} \cdot y + I_{dd} \cdot \sin(2y)] \\ &= \frac{I_{dd} \cdot R}{2p} \cdot [2y - \sin(2y)] . \end{aligned} \quad (2.18)$$

The output power is

$$P_o = \frac{V_o^2}{R} , \quad (2.19)$$

the dc power is

$$P_{dc} = V_{cc} \cdot I_{dd} , \quad (2.20)$$

and the maximum output voltage V_o is

$$V_{OMAX} = V_{DD} . \quad (2.21)$$

From the above equations the maximum efficiency is

$$h_{max} = \frac{P_{OMAX}}{P_i} = \frac{2y - \sin(2y)}{4 \cdot [\sin(y) - y \cdot \cos(y)]} . \quad (2.22)$$

Since the peak drain voltage and drain current are

$$V_{D_{MAX}} = 2V_{DD} , \quad (2.23)$$

and

$$I_{D_{MAX}} = I_{dq} + I_{dd} \quad (2.24)$$

respectively, the power output capability factor is

$$h_{\max} = \frac{P_{O_{MAX}}}{V_{D_{MAX}} \cdot I_{D_{MAX}}} = \frac{2y - \sin(2y)}{8p \cdot [1 - \cos(y)]} . \quad (2.25)$$

Figure.2.6 shows the maximum efficiency versus the conduction angle. Although it is shown that 100% efficiency is possible, it is impractical because the output power is zero, as shown in Fig.2.7.

Although the preceding analysis was for the single-ended amplifier configuration, a similar analysis can be done for the push-pull amplifier configuration. During the positive half of the signal swing, one device will push the current to the load, and during the negative half signal swing, the other device will pull the current from the load. For example, in a class-B push-pull power amplifier, every device is on for one half of the input cycle, which means that the conduction angle is equal to 180 degrees for each device. This is similar to two class-B single-ended power amplifiers connected in a parallel line. From this observation, it is possible to conclude that the efficiency of the push-pull power amplifier is the same as that of the single-ended power amplifier with the same conduction angle, and the output power capability of the push-pull power amplifier is twice that of the single-ended power amplifier. And this result is due to using two FETs.

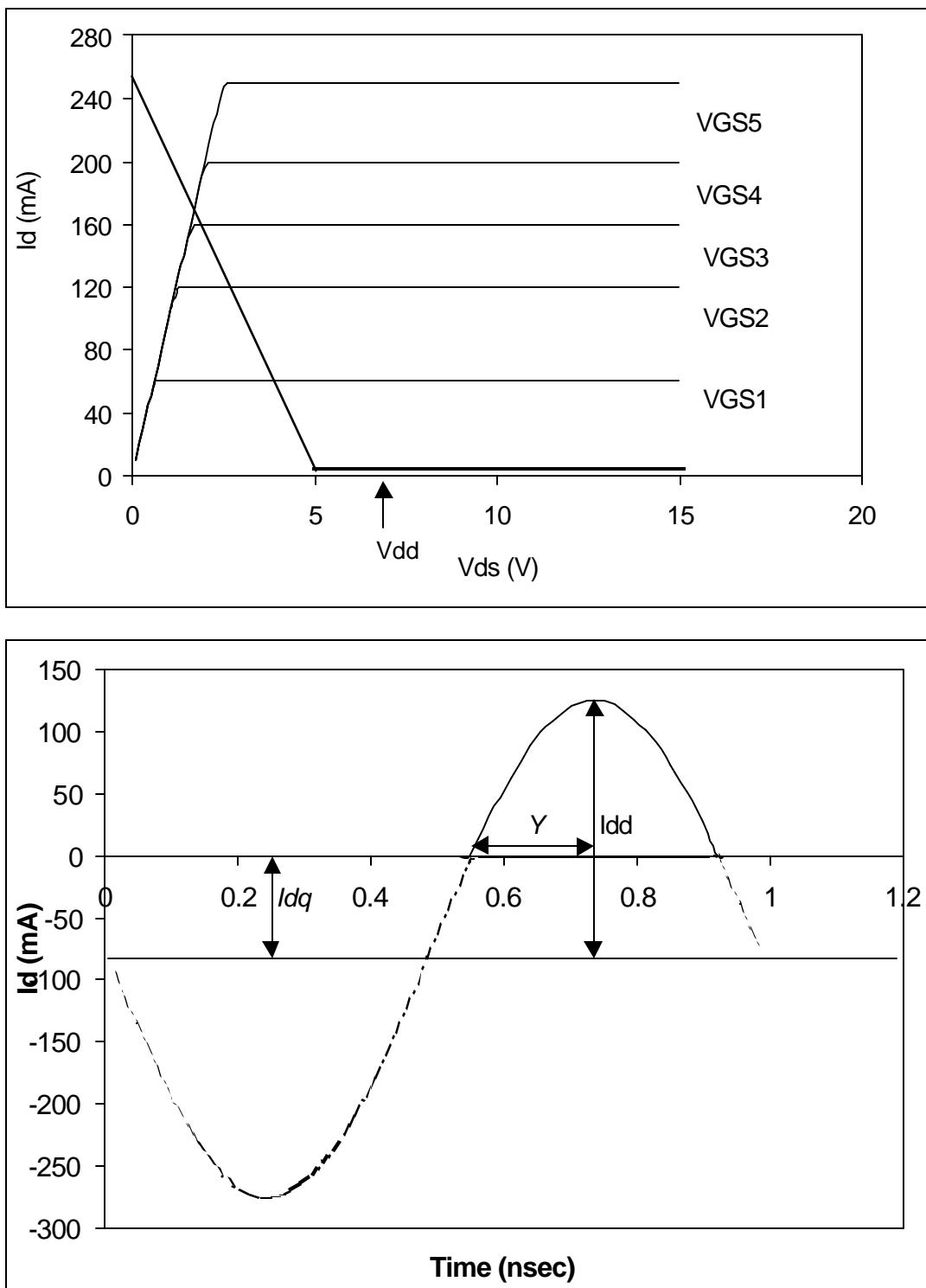


Figure 2.5. Load line and current waveform for the class-C power

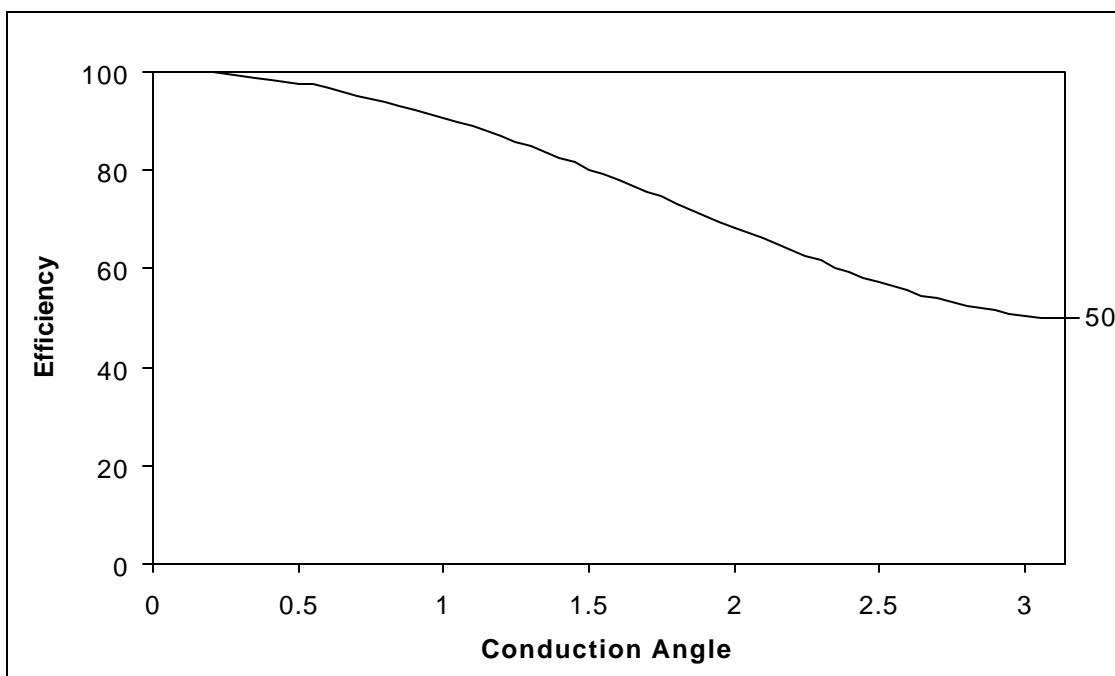


Figure 2.6. Efficiency vs. conduction angle

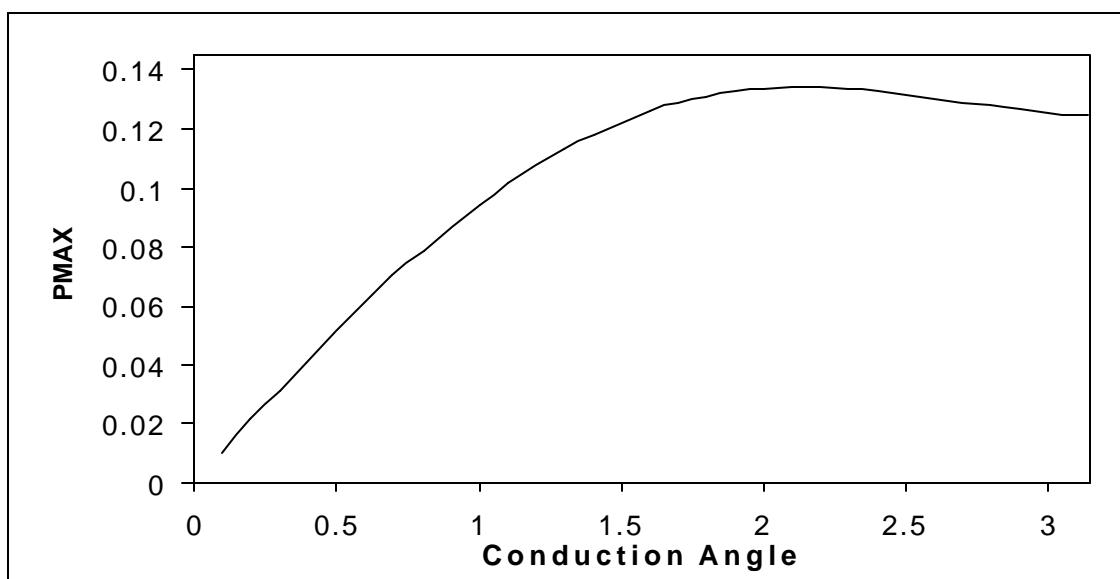


Figure 2.7. P_{MAX} vs. conduction angle

2.2.4 Class F

The class-F amplifier is one of the highest efficiency amplifiers. It uses harmonic resonators to achieve high efficiency, which resulted from a low dc voltage current product. In other words, the drain voltage and current are shaped to minimize their overlap region. Figure 2.8 shows a class-F amplifier. The inductor L_3 and capacitor C_3 are used to implement a third harmonic resonator that makes it possible to have a third harmonic component in the collector voltage. The output resonator is used to filter out the harmonic, keeping only the fundamental frequency at the output. The magnitude and the phase of the third harmonic control the flatness of the collector voltage and the power of amplifier.

The drain voltage is

$$V_d(\mathbf{q}) = V_{DD} + V_{om} \cdot \sin \mathbf{q} + V_{om3} \cdot \sin(3\mathbf{q}) \quad (2.26)$$

The setting $V_{om3} = \frac{V_{om}}{9}$ produces maximum flatness for the drain voltage. And, the maximum output occurs when the minimum point of $V_d(\mathbf{q})$ is zero. Hence,

$$V_{om} = \frac{9}{8} \cdot V_{DD} \quad (2.27)$$

The dc current is

$$I_{dc} = \frac{I_{dm}}{p} \quad , \quad (2.28)$$

the dc power is

$$P_{dc} = V_{DD} \cdot \frac{I_{dm}}{p} \quad , \quad (2.29)$$

the fundamental current is

$$I_{om} = \frac{I_{dm}}{2} \cdot \sin q \quad , \quad (2.30)$$

the maximum fundamental output power is

$$P_{o\max} = \frac{I_{dm}}{4} \cdot V_{om} \quad , \quad (2.31)$$

and, the maximum efficiency is

$$\eta_{\max} = \frac{P_{o\max}}{P_{dc}} \cdot 100 = \frac{\frac{I_{dm}}{4} \cdot \frac{9}{8} V_{DD}}{\frac{I_{dm}}{p}} \cdot 100 = 88.36\% \quad . \quad (2.32)$$

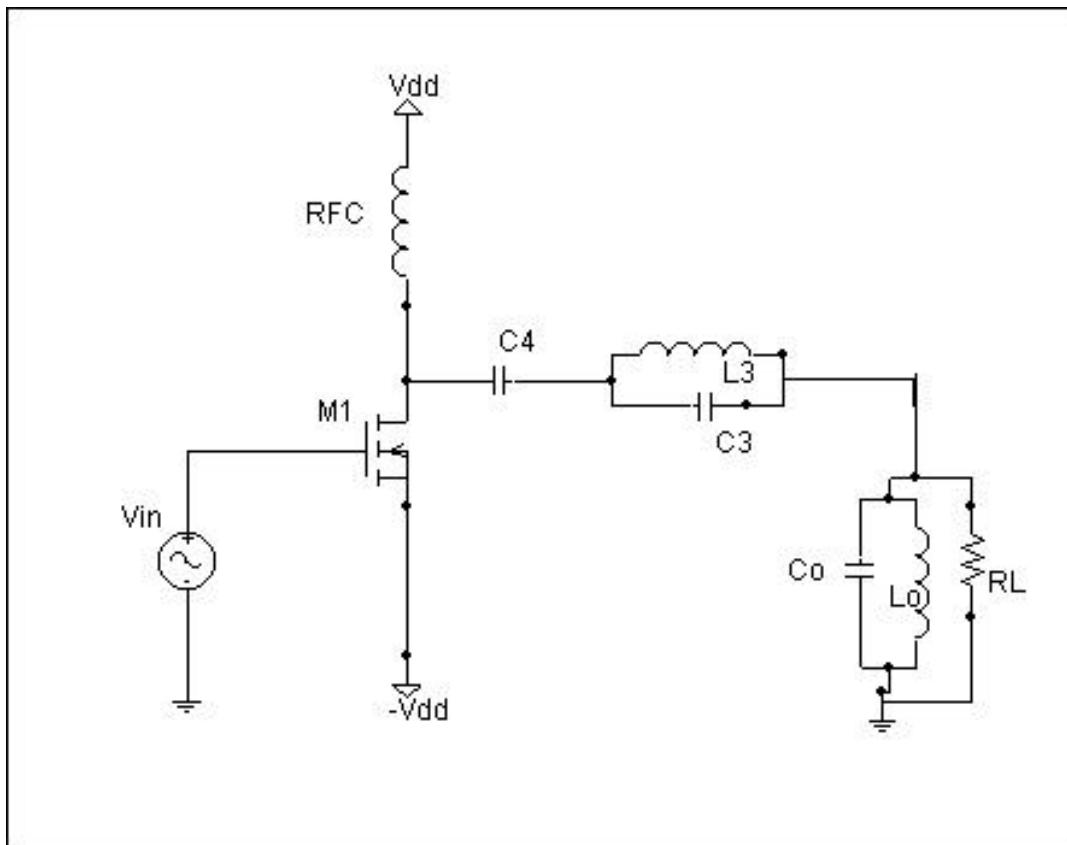


Figure 2.8. Single-ended power amplifier (class-F)

2.2.5 Other High-Efficiency Classes

There are other high-efficiency amplifiers such as D, E, G, H, and S. These classes use different techniques to reduce the average collector or drain power, which, in sequence, increase the efficiency. Classes D, E, and S use a switching technique, while classes G and H use resonators and multiple power-supply voltage to reduce the collector current-voltage product. A detailed analysis of class-E amplifier will be presented in Chapter 5.

Designers select the class type to be used based on the application requirements. Classes-A, AB, and B amplifiers have been used for linear applications such as amplitude modulation (AM), single-sideband modulation (SSB), and quadrature amplitude modulation (QAM). Also it can be used in linear and wide band applications such as the multi-carrier power amplifier. Classes C, D, E, F, G, and H have satisfied the need for narrowband tuned amplifiers of higher efficiency. Such applications include amplification of FM signals.

2.3 Main Physical Limitations

The descriptions of amplifiers in the previous sections have dealt with ideal devices. In reality, transistor amplifiers suffer from a number of limitations that influence amplifier operation and ultimately reduce their efficiency and output power.

In practical FET, there are four fundamental effects that force the operation of FET to deviate from the ideal case: the drain source resistance, the maximum channel current I_f , the open channel avalanche breakdown voltage, and the drain-source break down voltage [Robert, 1988]. Figure 2.9 shows I_{DS} - V_{DS} characteristics of a typical MESFET (ATF-46100).

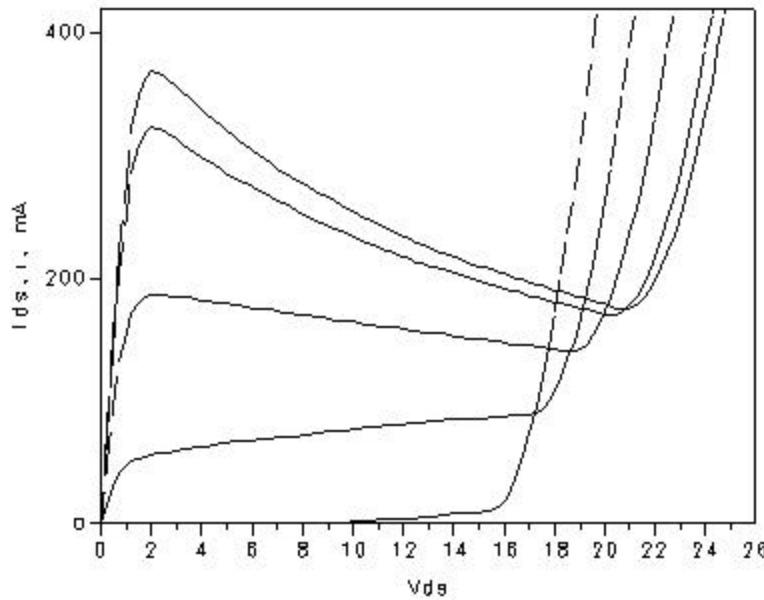


Figure 2.9. I_{DS} - V_{DS} characteristics of a typical MESFET.

2.4 Nonlinear MESFET Model

The development of a large-signal model for the transistor is an important step in the design of a nonlinear amplifier. The transistor model consists of linear and nonlinear circuit elements, where the latter are described by a set of nonlinear equations. Figure 2.10 shows a lumped-element model of the MESFET that can be used either in a small-signal or a large-signal analysis. R_g is the ohmic resistance of the gate, and R_s and R_d are the source and drain ohmic resistances, respectively. C_{ds} , and C_g and C_d are the drain-source capacitance and gate-channel capacitances respectively. Several authors have proposed number of nonlinear MESFET models such as the Curtice-Ettenberg model, the Staz model, and the Tom model.

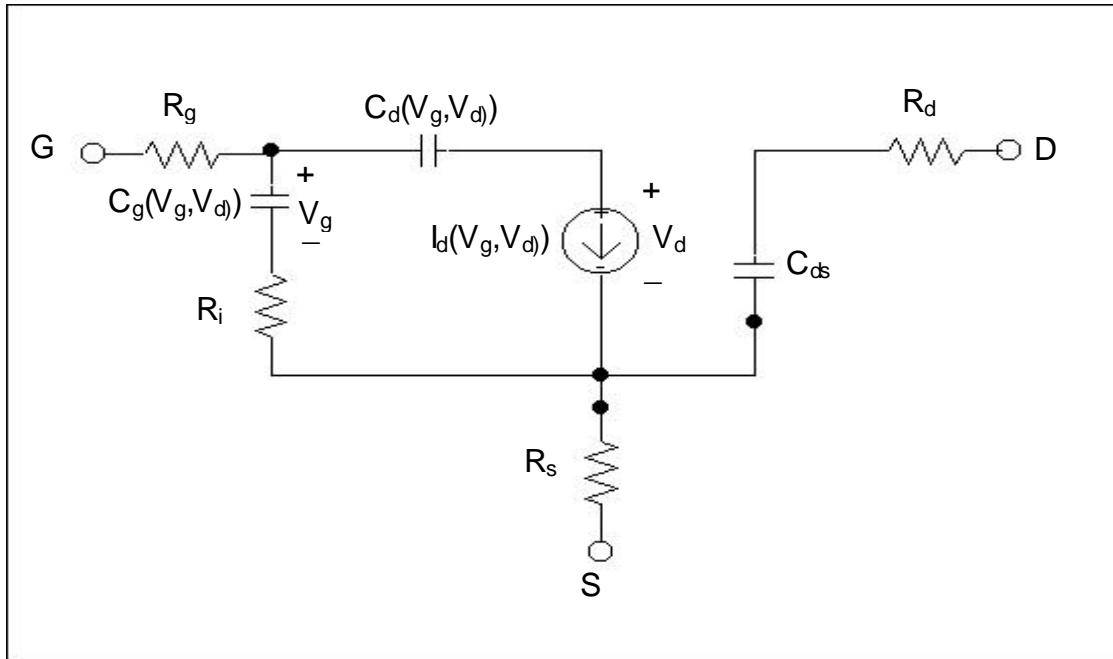


Figure 2.10 GaAs MESFET nonlinear equivalent circuit.

2.5 Nonlinear analysis

Having obtained a suitable model for the large signal behavior of the FET, the next step is to select the analysis method. There are two methods: frequency-domain techniques, and time domain techniques. Harmonic-balance analysis and Volterra-series analysis are the most important frequency-domain techniques. In the harmonic balance technique, the nonlinear circuit is partitioned into linear and nonlinear subcircuits. The linear subcircuit can be described by its Y, S, or any other parameters. The nonlinear elements are modeled by their *I/V* characteristics. The voltages at the interconnections between the two subcircuits are variables that, when determined, define all the voltages and currents in the network. In the Volterra-series analysis, the nonlinear elements are characterized by power series. Then the nonlinear transfer function can be obtained using the convolution.

In the time-domain techniques, conventional circuit theory is used to obtain time-domain differential equations that describe a nonlinear circuit. These differential equations are solved numerically. A major disadvantage of time-domain analysis is that a steady state solution, which is the only one of interest in amplifier design, often requires several cycles and, consequently, a long computation time.

More information about the nonlinear analysis methods for the amplifier circuits is available in *Nonlinear Microwave Circuits* by Stephen Maas [1987].

basics of Design

Optimizing Power Amplifier Design

Power amplifiers (PAs) play a critical role in any communication system, providing the necessary signal for all our wireless applications, from mobile networks to radar and guidance electronics. As people and things become more interconnected—and, increasingly, through wireless means—PAs are needed more than ever.

The complex world of PA design is becoming more intuitive thanks to software and hardware tools that let designers characterize, test, and refine their amplifier ideas quickly and effortlessly. For instance, several electronic design automation (EDA) tools have integrated schematic capture. The technique uses libraries of commercially available components for printed

circuit board-based design or process design kits (PDKs) for monolithic microwave integrated circuits (MMICs) used to build PAs. These EDA tools also have layout editing and electromagnetic analysis capabilities.

After conceptualizing circuit architectures from idealized circuit elements, designers can methodically incorporate more realistic models of their actual physical design as that information becomes available and modify amplifier details to achieve the desired performance based on simulation results of common tests.

PA designs are complex and can take on many possible topologies. While gain, the ratio of power at the output to that at the input, and output power (often measured at

the 1db compression point) are important to the application, mobile communication systems require that PA designs also rely on two main figures of merit: power-added efficiency (PAE) and linearity.

Power-added efficiency

PAE is defined as the difference between the output and input signals divided by the system bias. That is:

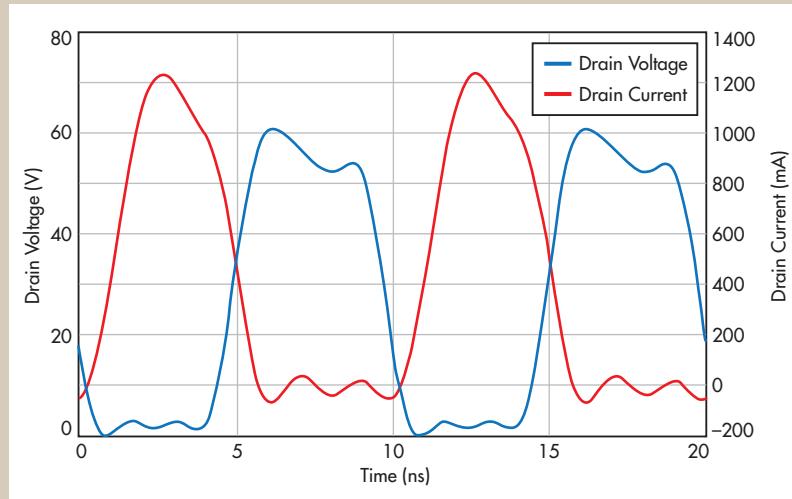
$$\text{PAE} = \frac{P_{\text{out, RF}} - P_{\text{in, RF}}}{P_{\text{DC, TOTAL}}}$$

where $P_{\text{out, RF}}$ is the power of the resulting RF signal, $P_{\text{in, RF}}$ is the power of the input RF signal, and $P_{\text{DC, TOTAL}}$ is the DC power supplied to the amplifier, commonly referred to as bias. PAE is usually represented as a percentage.

While PAE is a figure of merit on its own, it is also an indication of the PA's thermal performance since energy that does not make it into the output signal is dissipated as heat. Thermal performance is a specific concern in compact electronics where excess heat can degrade function and shorten device life.

Linearity

In addition to maximizing PAE, designers want their amplifiers to exhibit the best linearity possible. In a theoretical 100-percent linear PA, the output waveform would exactly match the shape of the input waveform with a larger amplitude. A graph of the output voltage $V_{\text{out}}(t)$ versus the input voltage $V_{\text{in}}(t)$ would show (correcting for fixed propagation delay through the am-



Real-world PAs have voltage and current waveforms that overlap in the time domain, detracting from efficiency and linearity. The area of overlap is proportional to the energy lost.

plifier) an instantaneous constant proportion.

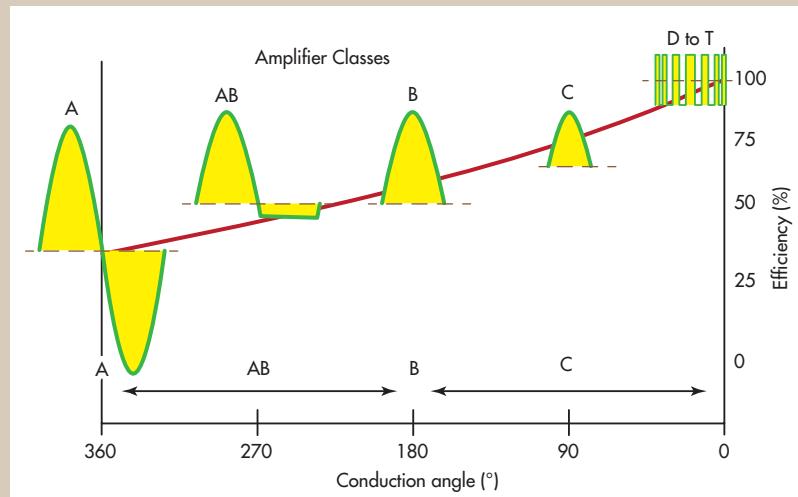
In reality, no PA is perfectly linear. AC circuits have complex impedance Z that includes both resistance R and complex reactance X . In addition, device nonlinearity will cause the time-domain output waveform to flatten, or clip, at the top and bottom. Just as sound is distorted through speakers at the upper end of their volumes, nonlinearity in amplified power signals increases as the output signal's peak-to-peak amplitude increases. The resulting distorted time-domain waveform generates spurious harmonic tones in the frequency domain.

Out-of-phase voltage and current waveforms create overlap in the time domain, with attendant energy lost proportional to the overlapping area as shown in the graph. Since nonlinearities and harmonics cannot be completely eliminated, designers manipulate them and their related impedances to boost specific PA performance metrics. Designers can engineer the resulting waveforms by terminating the device in specific impedances at the harmonic frequencies. Successful waveform engineering can significantly reduce the voltage-current overlap, thereby improving the PAE.

Designers use linearity and PAE concepts to choose the type of PA to construct. PAs are divided into classes based on their design and operation and summarized in the accompanying figure. Classes A, B, AB, and C are bias-type or linear PAs that operate with simple DC bias voltages and current. Classes D, E, and F are called switching-type or nonlinear amplifiers that treat the amplifying device as a switch. Although other classes exist, they are not typically used in RF designs.

Bias-type PAs

Simple Class A amplifiers have the highest gain and linearity but the lowest efficiencies. Their quiescent bias is set so output cur-



Bias-type PAs (A through C) are characterized by a conduction angle and clipping of the waveform. Switching-type PAs (D and higher) are more efficient but more complex.

rent flows constantly, resulting in a transistor conduction angle of 360 degrees. This middle bias avoids clipping and results in high linearity even when approaching the maximum operating frequency of the transistor. Ideal efficiency at peak envelope power (PEP) is 50 percent.

Class B sets its bias point to zero, resulting in a conduction angle of 180 degrees, only half the time. Cutting off half the waveform creates significant harmonics, so Class B PAs usually require output filters. On the other hand, ideal Class B PAs are 78.5 percent efficient at PEP.

As the name suggests, Class AB PAs strike a balance between the two previous classes. The quiescent point bias is set at 10 to 15 percent of $I_{C\max}$, the maximum current through the device. The result is a partially clipped waveform and a conduction angle between 180 and 360 degrees. Class AB PAs are 50 to 78.5 percent efficient and are sometimes used for their wider dynamic range.

Class C PAs have a quiescent bias point below zero, so their conduction angle is less than 180 degrees. Because the output waveform is severely clipped, Class C amplifiers exhibit poor linearity and can suffer transistor breakdowns. On

the other hand, their efficiency can approach 85 percent.

Switching-type PAs

Class D PAs use switching circuits with two or more transistors. The circuits generate a square-wave drain voltage and half-sinusoidal waveform for drain current that are completely out of phase for theoretical 100 percent efficiency. Device parasitics degrade this efficiency in practice, but although this perfection is unattainable in reality due to device parasitics, Class D PAs offer fast switching in their common applications range of low RF and audio frequencies and efficiency independent of load reactance.

The switching circuit in Class E amplifiers consists of a single transistor and parallel capacitor. The drain voltage waveform is the sum of the DC and RF currents. Ideally, the drain voltage has an inflection point at zero as the transistor turns on, creating a 100 percent efficiency. The output circuit filters out harmonics to output only the fundamental frequency.

Class F power amplifiers make use of harmonic frequencies by tuning the impedances of their input and output networks, often with transmission lines. The result is an approximation of the waveforms

used by Class D for high efficiency. Indeed, Class F PAs can reach 88.4 percent efficiency when tuning for three harmonics while still supporting high power signals. Class F

PAs are getting a lot of attention for new designs, but their design is far from simple. Designers use load-pull analysis and impedance matching to tune input networks, output

networks, and transmission lines for the best performance.

Load-pull analysis

Traditional load-pull systems connect the device under test (DUT) to a signal source, input and output tuners, and a power meter. This simplicity makes traditional load pull easy and potentially inexpensive to implement, but prevents measurement of amplitude or phase of the input signal. Additionally, the inability to characterize the input matching network and its associated losses impairs measurement accuracy.

Replacing the power meter with a vector receiver permits vector-receiver load pull, also known as real-time load pull. This technique characterizes the input and output traveling waves at the reference plane of the DUT. It can determine the input impedance and improve accuracy, but is only for non-modulated signals and narrow-band inputs.

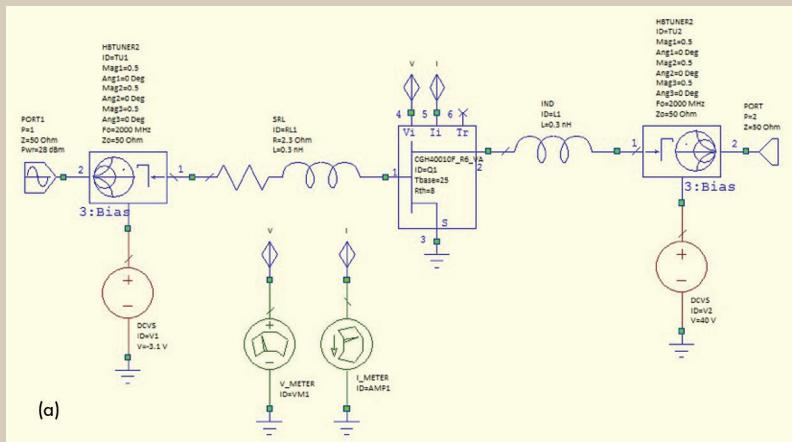
A variation on vector-receiver load pull, called open-loop active load pull, replaces the mechanical impedance tuner with a signal source with variable magnitude and phase. The system measures the output traveling waves and iteratively adjusts the input waves until the target impedance is reached. Testers can also perform harmonic load pull by merging multiple tuning chains at different frequencies. Open-loop load pull is expedited by the removal of the mechanical tuner, but the systems are expensive, and they can't accommodate wideband signals.

Virtual load pull

An alternative to manual load-pull testing is automated simulation within software design environments. One example comes from NI AWR Design Environment™, specifically within Microwave Office Load-Pull script.

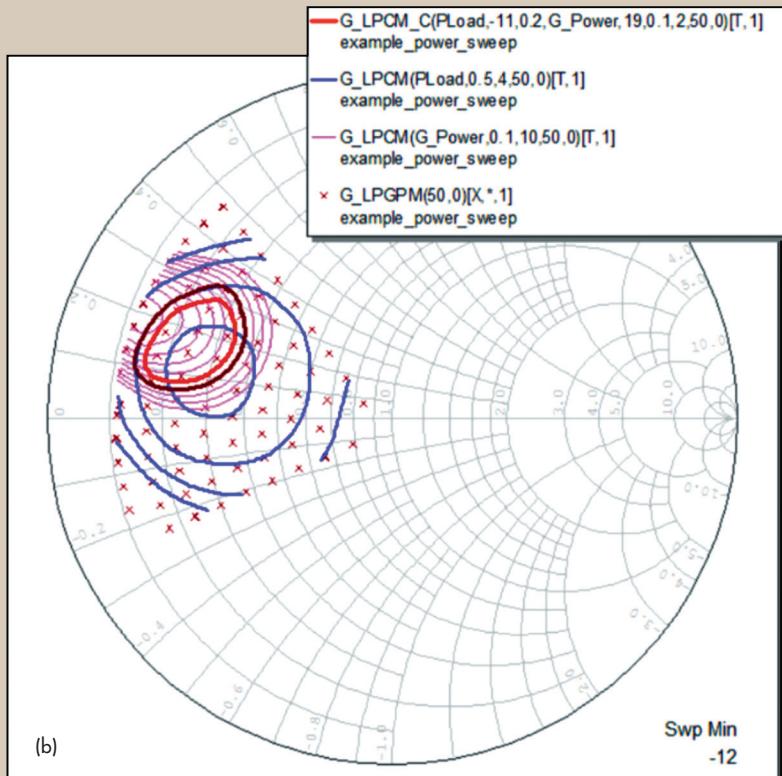
The wizard's basic setup features a source-pull tuner at the input and a load-pull tuner at the output, each integrated with bias Ts. A device

SAMPLE LOAD-PULL SIMULATION SETUP



Simulated load-pull can significantly speed the design and iteration process. This setup shows input on the left and output on the right, each with a bias network. The DUT is shown in the central rectangle.

SMITH CHART OF LOAD-PULL RESULTS



Plotting the results of a load-pull simulation on a Smith chart lets the designer quickly visualize the performance of the system over a swept range of frequencies. It also aids in the optimization of impedance-matching networks.

model is placed in the center and supplied with any additional parasitics such as those from wire bonds or test fixtures.

Using this setup, one popular approach is to first perform source-pull simulation to determine the input impedance which gives the best power gain at the fundamental frequency by simulating the DUT as directly connected to a standard 50 ohm load on the output. The source impedance, to be used in the subsequent load-pull simulations, is chosen based on this data.

Next, the designer can move on to the actual load pull simulations. The load-pull can be performed with only the fundamental impedance swept, but NI AWR Design Environment allows for a “nested” load pull, wherein the fundamental and harmonic impedances can be swept as part of a single simulation (for that matter, the source and load-pull simulations can also be nested).

In this way, designers can determine the fundamental load termination—as well as the second and third harmonic load terminations—for the best device performance. The load-pull tuners can also be set to a fixed impedance for any of the harmonics while the other harmonic terminations are swept.

In the case of Class F PAs, the goal is usually to have an open circuit presented to the device at the odd harmonic frequencies and a short circuit at the even harmonic frequencies. It is important to meet these conditions at the device’s current generator plane rather than at its package plane. NI AWR Design Environment allows for de-embedded load-pull results, so that any package effects can be accounted for. The load-pull results and time-domain waveforms can also be plotted at the actual device’s current generator plane.

Smith charts

The results of load-pull analysis are often represented on Smith charts. They display a PA’s frequency response in a vector form,

instead of the scalar form that can be plotted on simple x and y axes.

These polar plots show the reflection coefficient, Γ , a measure of impedance mismatch. That is:

$$\Gamma = (Z_L - Z_0) / (Z_L + Z_0)$$

where Z_L is the load impedance and Z_0 is the characteristic impedance, usually set to 50 ohms.

The boundary of the Smith chart represents a Γ of 1, or complete reflection of power. The center of the chart is $\Gamma = 0$, indicating perfect transmission of the power. The left edge of the chart represents $\Gamma = -1$ at an angle of 180 degrees for a short. The right edge corresponds with $\Gamma = +1$ at 0 degrees for an open circuit.

Also plotted on the Smith chart are impedances shown as constant resistance circles tangential to partial circles representing constant imaginary portions of the complex impedance or a constant real portion.

The results of load-pull analysis can be plotted on the Smith chart as shown in the figure. Impedances for matching networks can also be plotted for each frequency of interest, letting designers tune or optimize matching networks directly by comparing network impedances to the maximum performance points from the load-pull data including harmonic frequencies.

Impedance matching

To optimize performance, a designer creates input and output networks that match the source and load impedance to the values determined through the load/source pull measurements or simulation results.

It’s important to remember that impedance is a complex number. In many applications, like antennas and transmitters, the complex components of two matched impedances need to cancel each other.

Connections between PA components must be considered along with the components themselves. If

these lines have a significant electrical length—if they are a considerable percentage of the fundamental wavelength, λ —they need to be considered transmission lines with their own impedances.

Transmission line impedance is calculated by:

$$Z_0 = \sqrt{L/C}$$

where Z_0 is the characteristic impedance of the transmission line, L is its inductance in henry per meter (H/m), and C is its capacitance in farads per meter (F/m).

If transmission line impedance isn’t matched to the source and load impedances, standing waves will develop that prevent the load from absorbing all the power sent its way. Reflected power also has the potential to damage the source.

As discussed above, the amount of power reflected is depicted by the reflection coefficient Γ . Another representation of power lost is the voltage standing wave ratio (VSWR). That is:

$$\text{VSWR} = |Z_L| / |Z_0| \text{ for } |Z_L| > |Z_0| \text{ and } \text{VSWR} = |Z_0| / |Z_L| \text{ for } |Z_L| < |Z_0|$$

where Z_L is the load impedance. Thus, perfectly matched impedances provide a VSWR of 1. Reflection coefficient magnitude can also be computed from VSWR:

$$|\Gamma| = (\text{VSWR} - 1) / (\text{VSWR} + 1)$$

Several techniques let designers match load and source impedances. Where both impedances are wholly real, designers use a specific length of transmission line called a Q-section, or quarter-wave transformer, to match the impedances. The precise length of $\lambda/4$ matches the load impedance to the source impedance at a single frequency.

The impedance of the Q-section is calculated as:

$$Z_Q = \sqrt{(Z_0 * Z_L)}$$

where Z_Q is the Q-section’s characteristic impedance, Z_0 is the characteristic impedance of the input

transmission line from the source, and Z_L is the load impedance.

Another technique is the development of an L-network, a circuit containing a shunt inductor and a series capacitor. The capacitor blocks DC power from flowing back toward the source while the inductor blocks radio frequency or microwave inputs from flowing toward the power supply. The L-network's reactive resistance increases with input frequency instead of being limited to a single wavelength, however circuit Q is entirely dictated by input and output impedances.

L-networks can be designed for high-pass or low-pass operation.

Low-pass L-networks in RF designs can reduce the effect of unwanted harmonics and filter out noise and other undesirable signals. In these networks, the capacitance is placed on the input side of the L-network if source impedance is lower than load impedance and on the output side if source impedance is higher.

If this Q does not meet application requirements or is too low to act as a bandwidth filter, another option is T- or pi-networks. A pi-network is two capacitors connected by an inductor in the shape of the Greek letter π . Its design is basically two L-networks placed next to each other and separated by a virtual impedance. A T-network connects a capacitor between two inductors or more commonly, between an inductor and a second capacitor. These can also be viewed for design purposes as two cascaded L-networks.

Computerized matching

As with load-pull analysis, computer tools have evolved to help designers create the impedance matching networks they need. One example is the iMatch module for Microwave Office software.

The synthesis tool offer a new bandpass option to create an impedance-matching network. The

wizard lets users edit the terminations within the networks to match output impedance to load impedance over the desired frequency range. Designers can create terminations in the wizard or load them from a schematic or data file they have already created.

The software displays the results along with a Smith chart to let designers determine if the matching network is sufficient. When the design is finalized, the wizard synthesizes it into the overall design.

Today's power amplifiers target unprecedented levels of linearity and efficiency thanks to a greater understanding of waveform engineering, novel circuit architectures, switch-mode techniques, and improvements in design automation software. Along with load-pull simulation, these tools help designers iterate more quickly to achieve the performance required by tomorrow's communication systems and other wireless applications. ■

NI AWR Design Environment

Now Playing on a Screen Near You



Microwave Office | Visual System Simulator | Analog Office | AXIEM | Analyst

V12 NI AWR Design Environment redefines the term “user productivity” for designers of MMICs, RF PCBs, modules, and more. With the addition of new amplifier, radar, and antenna specific features, expanded third-party flows for EM, stability analysis, and DRC/LVS, as well as

additional speed and ease-of-use enhancements, it’s never been easier to streamline your design process, improve your end-product performance, and accelerate your time to market. Display NI AWR Design Environment on your desktop today. Get started at awrcorp.com/v12.

>> Learn more at ni.com/awr

RF Power Amplifiers

Iulian Rosu, YO3DAC / VA3IUL, <http://www.qsl.net/va3iul>

RF Power Amplifiers are used in a wide variety of applications including Wireless Communication, TV transmissions, Radar, and RF heating.

The basic techniques for RF power amplification can use classes as A, B, C, D, E, and F, for frequencies ranging from VLF (Very Low Frequency) through Microwave Frequencies.

RF Output Power can range from a few mW to MW, depend by application. The introduction of solid-state RF power devices brought the use of lower voltages, higher currents, and relatively low load resistances.

- Most important parameters that defines an RF Power Amplifier are:
 1. Output Power
 2. Gain
 3. Linearity
 4. Stability
 5. DC supply voltage
 6. Efficiency
 7. Ruggedness

Choosing the bias points of an RF Power Amplifier can determine the level of performance ultimately possible with that PA. By comparing PA bias approaches, can evaluate the trade-offs for: Output Power, Efficiency, Linearity, or other parameters for different applications.

- The Power Class of the amplification determines the type of bias applied to an RF power transistor.
- The Power Amplifier's Efficiency is a measure of its ability to convert the DC power of the supply into the signal power delivered to the load.

The definition of the efficiency can be represented in an equation form as:

$$\eta = \frac{\text{Signal power delivered to load}}{\text{DC power Supplied to output circuit}}$$
 or Power Added Efficiency:
$$PAE = \frac{P_o - P_{in}}{P_{DC}}$$

- Power that is not converted to useful signal is dissipated as heat. Power Amplifiers that has low efficiency have high levels of heat dissipation, which could be a limiting factor in particular design.
- In addition to the class of operation, the overall efficiency of a Power Amplifier is affected by factors such as dielectric and conductor losses. First quantify any loss in the circuit, then attempt to minimize it, and finally ensure that the mechanical and thermal design is adequate under all conditions.

Power Classes

There are many different classes of amplification available:

Class-A

Is defined, as an amplifier that is biased so that the output current flows at all the time, and the input signal drive level is kept small enough to avoid driving the transistor in cut-off. Another way of stating this is to say that the conduction angle of the transistor is 360° , meaning that the transistor conducts for the full cycle of the input signal. That makes Class-A the most linear of all amplifier types, where linearity means simply how closely the output signal of the amplifier resembles the input signal.

- Always have to remember this:

No transistor is perfectly linear, however the output signal of an amplifier is never an exact replica of the input signal.

- Linear amplification is required when the signal contains AM – Amplitude Modulation or a combination of both, Amplitude and Phase Modulation (SSB, TV video carriers, QPSK, QAM, OFDM).

Signals such as CW, FM or PM have constant envelopes (amplitudes) and therefore do not require linear amplification.

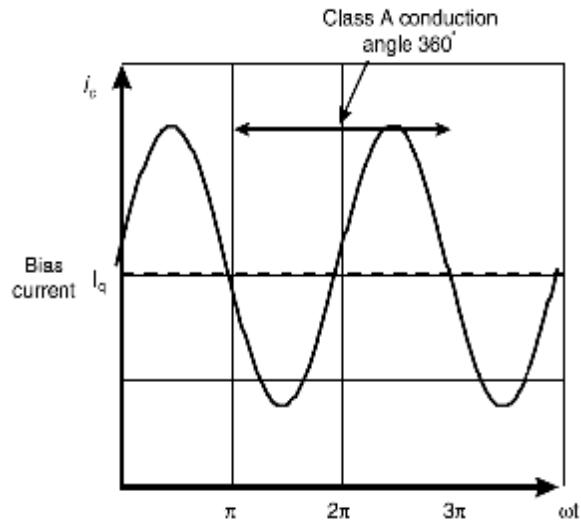
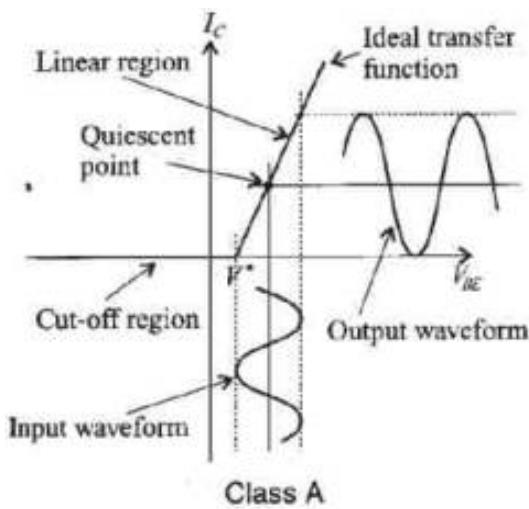
- The DC-power input is constant and the efficiency of an ideal Class-A PA is 50 % at PEP.
- The DC power consumption of a Class-A amplifier is independent of the output signal amplitude.

$$P_{DC} = V_{CC}^2 / R = V_{CC} \times I_{CQ} \text{ and } I_{CQ} \sim I_{MAX} / 2$$

- The amplification process in Class-A is inherently linear, hence increasing the quiescent current or decreasing the input signal level monotonically decreases IMD and harmonic levels.
- Since both positive and negative excursions of the drive affect the drain current, it has the highest gain of any PA.
- The absence of harmonics in the amplification process, allows Class-A to be used at frequencies close to the maximum capability (f_{max}) of the transistor. However, the efficiency is low. Class-A PAs are therefore typically used in applications requiring low power, high linearity, high gain, broadband operation, or high-frequency operation.
- The efficiency of real Class-A Pas is degraded by the on-state resistance or saturation voltage of the transistor. It is also degraded by the presence of load reactance, which in essence requires the PA to generate more output voltage or current to deliver the same power to the load.

$$\eta \text{ (Efficiency_Class-A)} = \text{Max_Load_Voltage} / (2 * V_{CC}^2)$$

- One important thing to mentioned is that: small signal S-parameters can be used in simulations if the large-signal amplifier is operating in Class-A.



Class-B

This is an amplifier in which the conduction angle for the transistor is approximately 180°.

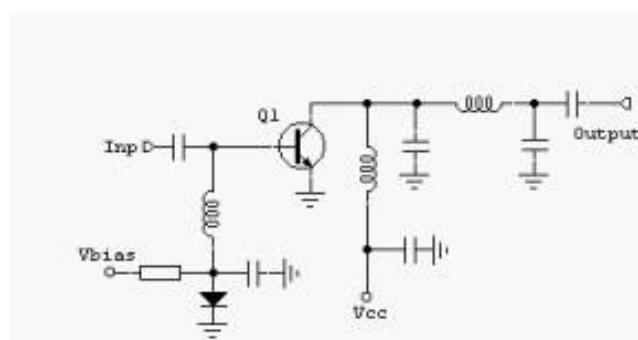
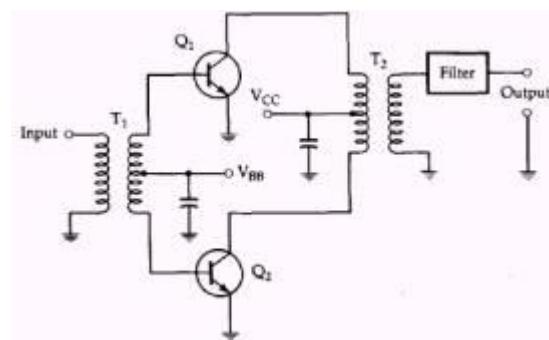
- Thus the transistor conducts only half of the time, either on positive or negative half cycle of the input signal.
- The same as in Class-A, the DC bias applied to the transistor determines the Class-B operation.
- Class-B amplifiers are more efficient than Class-A amplifiers. The instantaneous efficiency of a Class-B PA varies with the output voltage and for an ideal PA reaches $\pi/4$ (78.5 %) at PEP. However they are much less linear. Therefore a typical Class-B amplifier will produce quite a bit harmonic distortion that must be filtered from the amplified signal.

$$P_{DC} = (2 * V_{CC} * V) / (\Pi * R);$$

$$P_{LOAD} = V^2 / (2 * R);$$

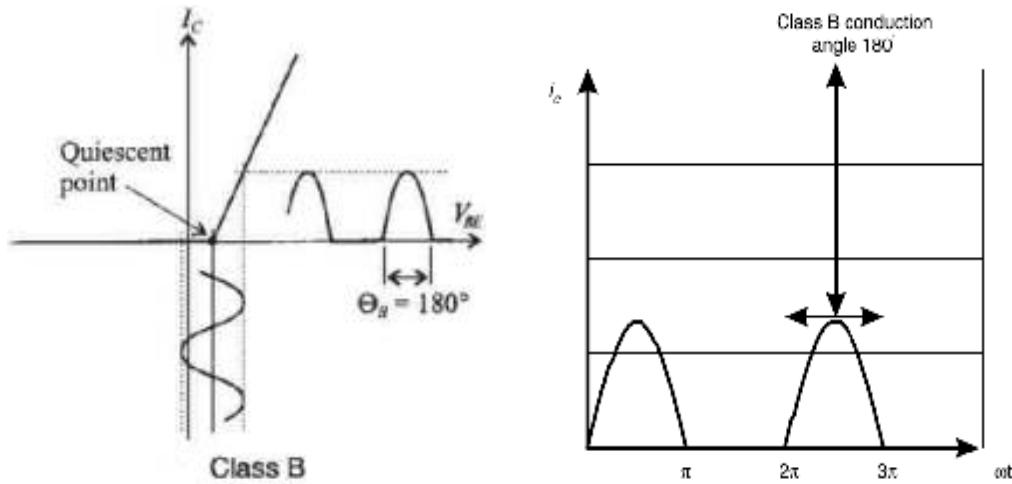
$$\eta \text{ (Efficiency_Class-B)} = (\Pi * V) / (4 * V_{CC})$$

Common configuration of Class-B amplifier is push-pull amplifier. In this configuration one transistor conducts during positive half cycles of the input signal and the second transistor conducts during the negative half cycle. In this way the entire input signal is reproduced at the output.



A single transistor may be used in a Class-B configuration. The only requirement in this case is that a resonant circuit must be placed in the output network of the transistor in order to “reproduce” the other half of the input signal.

- In practice, the quiescent current is on the order of 10 % of the peak collector current and adjusted to minimize crossover distortion caused by transistor nonlinearities at low outputs.
- In theory 6dB or more drive power is needed to achieve Class-B compared with Class-A. In practice this 6dB reduction in power gain is lower; for BJT amplifiers is lower than FETs, approximately 2dB.
- The efficiency of the push-pull power amplifier is the same as that of the single ended power amplifier with the same conduction angle, and the output power capability of the push-pull power amplifier is twice that of the single-ended power amplifier (3dB higher).



Class-AB

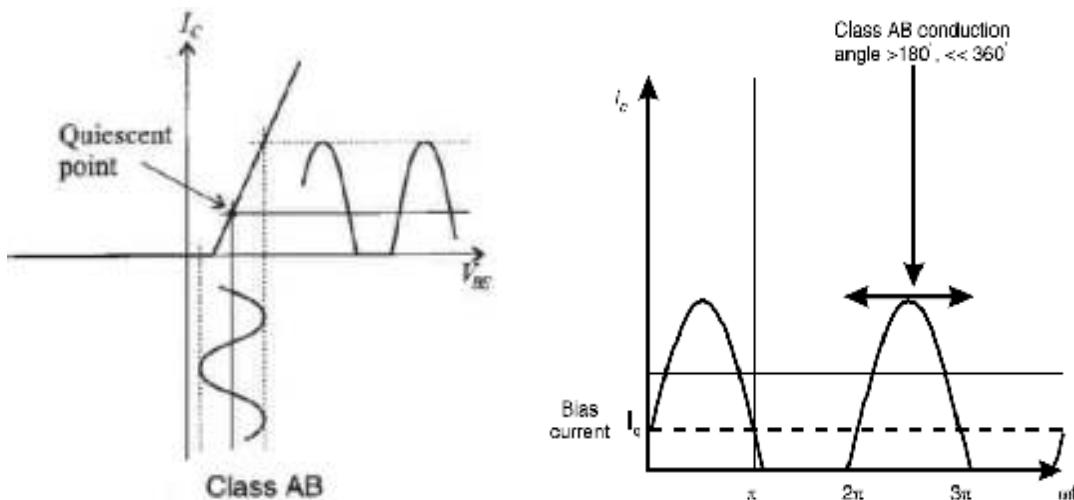
This amplifier is a compromise between Class-A and Class-B in terms of efficiency and linearity.

The transistor is biased typically to a quiescent point, which is somewhere in the region between the cutoff point and the Class A bias point, at 10 to 15 percent of $I_{C\max}$.

In this case, the transistor will be ON for more than half a cycle, but less than a full cycle of the input signal.

- Conduction angle in Class-AB is between 180° and 360° and Efficiency is between 50 % and 78.5 %
- Class-AB has higher efficiency than Class-A at price of linearity.
- Class-AB is not a linear amplifier; a signal with an amplitude-modulated envelope will be distorted significantly at this peak power level. The reason is in fact that in Class-AB operation the conduction angle is a function of drive level.

Experimentally was found that Class-AB often offers a wider dynamic range than either Class-A or Class-B operation. This is because gain compression in Class-AB comes from a different, and additional, source than Class-A. Saturation effects are primarily caused by the clipping of the RF voltage on the supply rails.



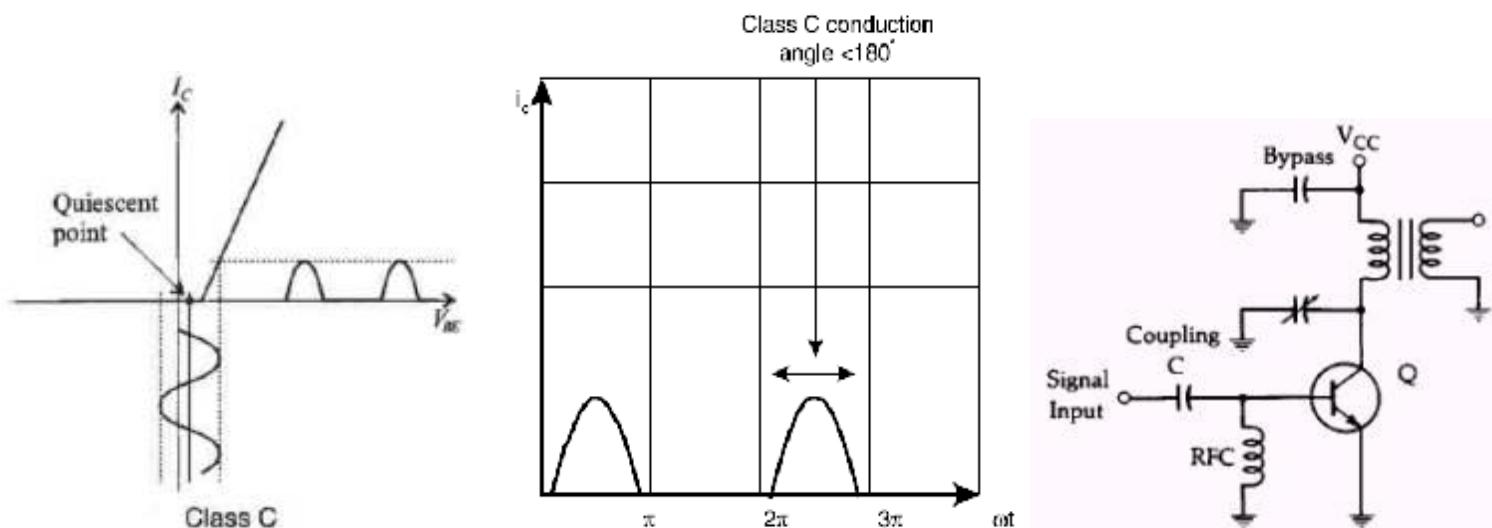
- Linearizing the response of a BJT PA in Class-AB includes the use of specific, and very low, impedance for the base bias supply voltage. This is a very different bias design issue in comparison to the simple current bias used in small signal BJT amplifiers, or the simple high impedance voltage bias used in FET PA's.
- Running the PA in a mid-AB condition the power gain may be 3dB higher than Class-B.
- Conventional Class-AB operation incurs odd degree nonlinearities in the process of improving efficiency. Theoretically to increases efficiency all the way up to 78.5 %, the device shall generate only even order nonlinearities. Such a device will not generate undesirable close-to-carrier intermodulation distortion.

Class-C

Is an amplifier where the conduction angle for the transistor is significantly less than 180°.

- The transistor is biased such that under steady-state conditions no collector current flows.
- The transistor idles at cut-off.

Class C Amplifier



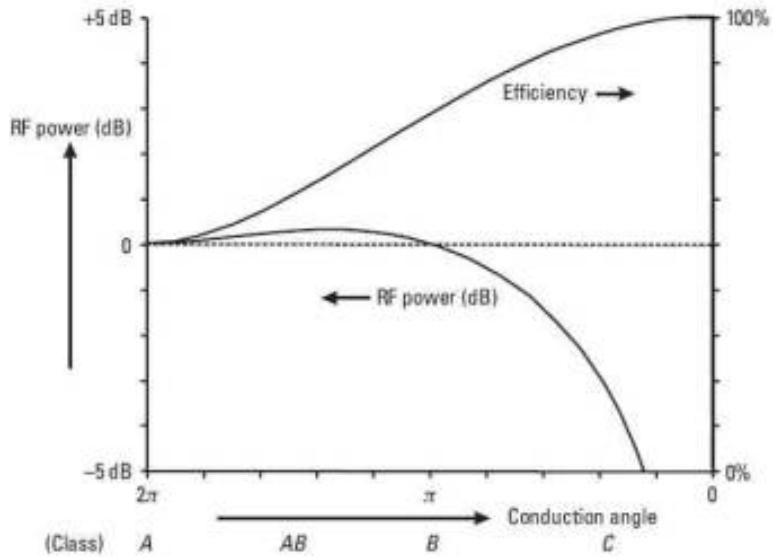
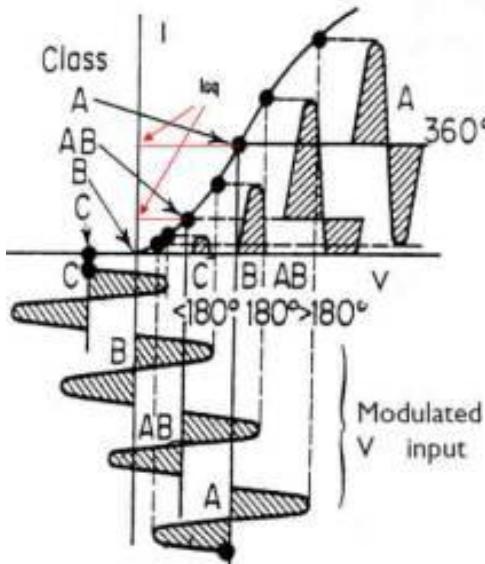
- Linearity of the Class-C amplifier is the poorest of the classes of amplifiers.
- The Efficiency of Class-C can approach 85 %, which is much better than either the Class-B or the Class-A amplifier.

In order to bias a transistor for Class-C operation, it is necessary to reverse bias of base-emitter junction. External biasing is usually not needed, because it is possible to force the transistor to provide its own bias, using an RF choke from base to ground.

One of the major problems with utilizing Class-C in solid-state applications is the large negative swing of the input voltage, which coincides with the collector/drain output voltage peaks. This is the worst condition for reverse breakdown in any kind of transistor, and even small amounts of leakage current flowing at this point of the cycle have an important effect on the efficiency. For this reason true Class-C operation is not often used in solid-state at higher RF and Microwave frequencies.

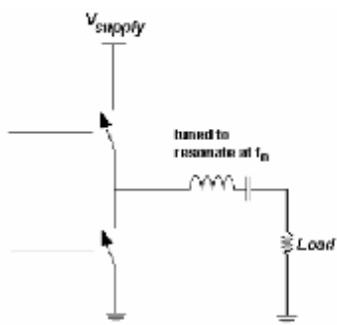
In order to survive Class-C operation, the transistor should have a collector voltage breakdown that is at least three times the active device's own DC voltage supply. The reason: Class-C amplifiers have low average output power (since the transistor conducts only for short, pulse-like periods), but demand very high input drive levels. Thus, the transistor's main Class-C failure mode is the low value of the active device's own reverse breakdown voltage, which is unfortunately exacerbated by the RF input signal voltage going negative just as the transistor's collector voltage reaches its positive peak. This is especially problematic and dangerous if the load changes from design expectations, such as occurs if the system sustains a damaged or missing antenna or feedline during operation.

Output waveforms and Efficiency vs Conduction Angle

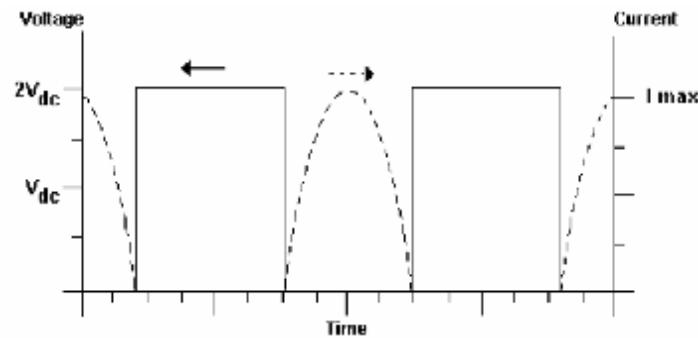


Class-D

The voltage mode Class D amplifier is defined as a switching circuit that results in the generation of a half-sinusoidal current waveform and a square voltage waveform. Class-D PAs use two or more transistors as switches to generate a square drain-voltage waveform. A series-tuned output filter passes only the fundamental-frequency component to the load,



Class-D amplifier



Class-D Voltage and Current waveforms

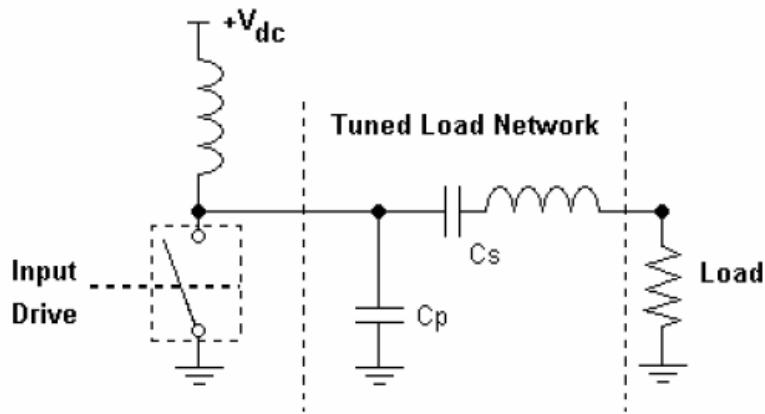
Class-D amplifiers suffer from a number of problems that make them difficult to realize, especially at high frequencies. First, the availability of suitable devices for the upper switch is limited. Secondly, device parasitics such as drain-source capacitance and lead inductance result in losses in each cycle. If realized, (they are common at low RF and audio frequencies) Class-D amplifiers theoretically can reach 100% efficiency, as there is no period during a cycle where the voltage and current waveforms overlap (current is drawn only through the transistor that is on).

- No real amplifier can be a true Class-D, as non-zero switch resistances and capacitive as well as inductive parasitics restrict the shape of the drain voltage waveform.
- A unique aspect of Class-D (with infinitely fast switching) is that efficiency is not degraded by the presence of reactance in the load.

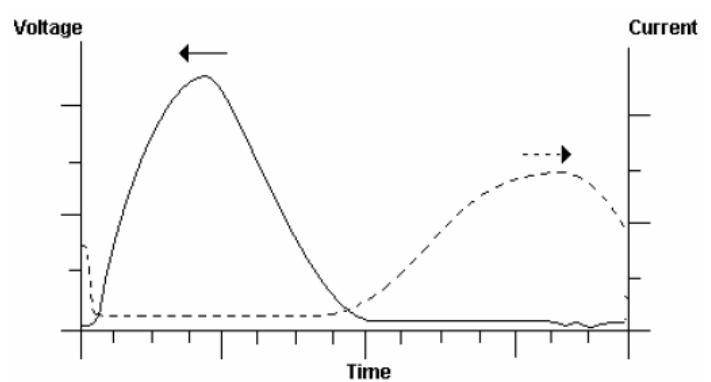
Class-E

Class-E employs a single transistor operated as a switch. The collector/drain voltage waveform is the result of the sum of the DC and RF currents charging the drain-shunt capacitance. In optimum class E, the drain voltage drops to zero and has zero slope just as the transistor turns on.

The result is an ideal efficiency of 100 %, elimination of the losses associated with charging the drain capacitance in class D, reduction of switching losses, and good tolerance of component variation.



Class-E amplifier

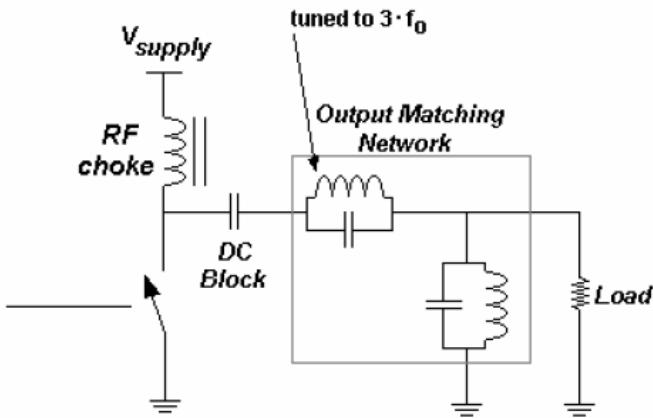


Class-E Voltage and Current waveforms

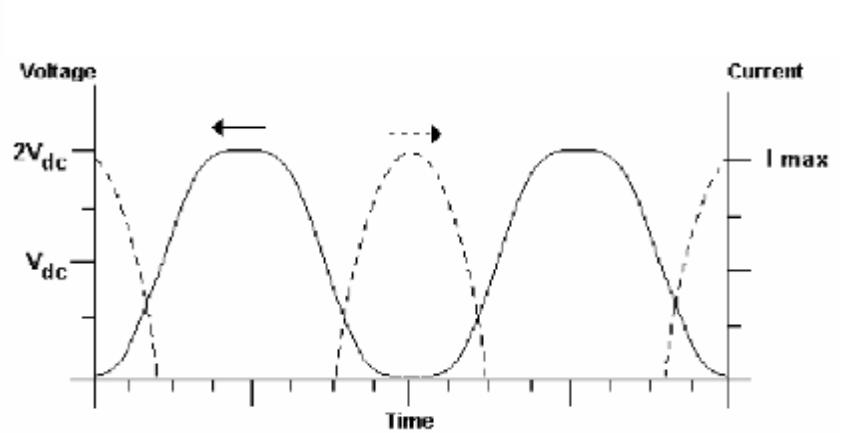
- A Class-E amplifier will exhibit an upper limit on its frequency of operation based on the output capacitance required for the output matching circuit that produces the waveforms described and shown above. Specifically, a Class-E amplifier (see figure II-6) for optimum efficiency requires an upper limit on capacitance C_s .
- If a given transistor has an intrinsic capacitance greater than C_{p_max} , it is not useable at the desired frequency. This C_s requirement implies that for high power at high frequencies, higher current densities are required, as the cross-sectional area of the switch corresponds directly to the device's intrinsic capacitance.

Class-F

Class-F boosts both efficiency and output by using harmonic resonators in the output network to shape the drain waveforms. The voltage waveform includes one or more odd harmonics and approximates a square wave, while the current includes even harmonics and approximates a half sine wave. Alternately (“inverse class F”), the voltage can approximate a half sine wave and the current a square wave.



Class-F amplifier



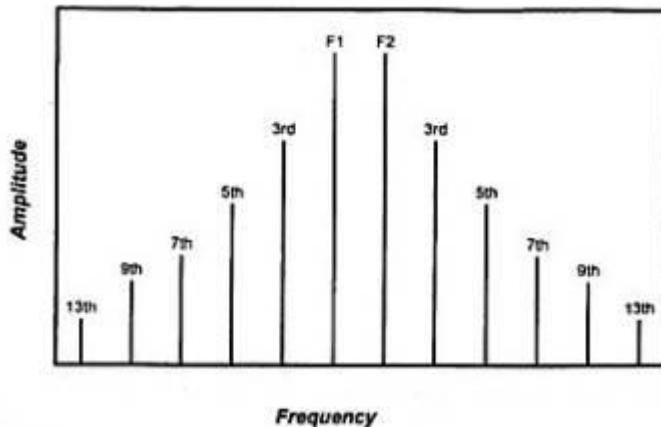
Class-F Voltage and Current waveforms

- The required harmonics can in principle be produced by current source operation of the transistor. However, in practice the transistor is driven into saturation during part of the RF cycle and the harmonics are produced by a self-regulating mechanism similar to that of saturating Class-C. Use of a harmonic voltage requires creating a high impedance (3 to 10 times the load impedance) at the collector/drain, while use of a harmonic current requires a low impedance (1/3 to 1/10 of the load impedance). While Class-F requires a more complex output filter than other PAs, the impedances must be correct at only a few specific frequencies. Lumped-element traps are used at lower frequencies and transmission lines are used at microwave frequencies. Typically, a shorting stub is placed a quarter or half-wavelength away from the collector/drain.
- Class-F amplifier designs intentionally squaring the voltage waveform through controlling the harmonic content of the output waveform. This is accomplished by implementing an output matching network which provides a high impedance ‘open circuit’ to the odd harmonics and low impedance ‘shorts’ to even harmonics. This results in a squared off (though for Class-F, truly squared) voltage waveform. The third harmonic only is peaked.

- Class-F amplifiers are capable of high efficiency (88.4% for traditionally defined Class-F, or 100% if infinite harmonic tuning is used).
- Class-F amplifier design is difficult mainly due to the complex design of the output matching network.

Power Amplifier Linearity

- When two or more signals are input to an amplifier simultaneously, the second, third, and higher-order intermodulation components (IM) are caused by the sum and difference products of each of the fundamental input signals and their associated harmonics.
- The rated PEP of a Power Amplifier is the maximum envelope power of a two-tone signal for which the amplifier intermodulation level is -30dBc.
- When two signals at frequencies f_1 and f_2 are input to any nonlinear amplifier, the following output components will result:
 - Fundamental: f_1, f_2
 - Second order: $2f_1, 2f_2, f_1 + f_2, f_1 - f_2$
 - Third order: $3f_1, 3f_2, 2f_1 \pm f_2, 2f_2 \pm f_1$,
 - Fourth order: $4f_1, 4f_2, 2f_2 \pm 2f_1$,
 - Fifth order: $5f_1, 5f_2, 3f_1 \pm 2f_2, 3f_2 \pm 2f_1$, + Higher order terms
- The odd order intermodulation products ($2f_1-f_2, 2f_2-f_1, 3f_1-2f_2, 3f_2-2f_1$, etc) are close to the two fundamental tone frequencies f_1 and f_2 .



- The nonlinearity of a Power Amplifier can be measured on the basis of generated spectra than on variations of the fundamental signal. The estimation of the amplitude change (in dB), of the intermodulation components (IM) versus fundamental level change, is equal to the order of nonlinearity.

For example for 1dB increase of fundamental level (f_1 and f_2), the level of IM2 will go up with 2dB, the level of IM3 will go up with 3dB, and so on.

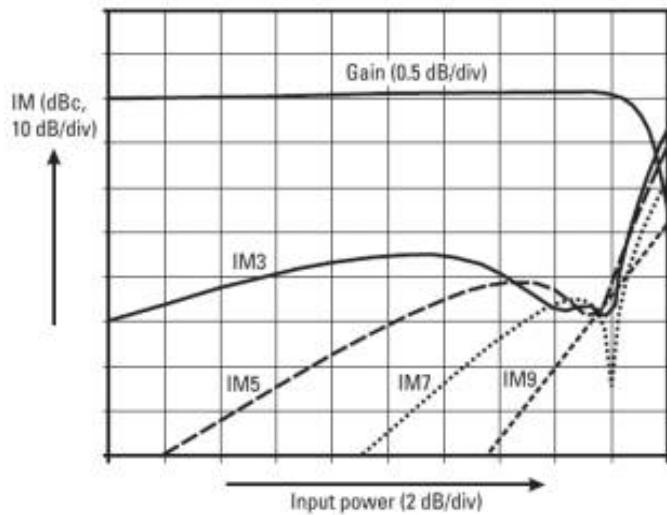
This is valid only for an amplifier that is not in compression.

- As a relation between the degree of the nonlinearity (third, fifth, etc) and the frequency of the side tone (such as IM3, IM5, etc), can be mentioned that the IM5 tones are not affected by third-degree nonlinearities, but IM3 tones are functions of both third- and fifth-degree (and higher) nonlinearities. That means at low signal amplitudes, where

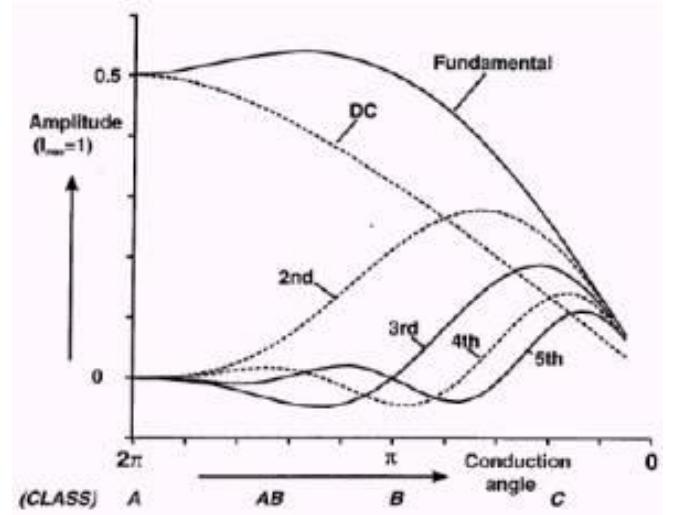
the fifth-order distortion products can be neglected, the amplitudes of the IM3 tones are proportional to the third power of the input amplitude.

With a fairly large signal amplitude, fifth-order products (which are dependent on a power of five) will start to affect the IM3 responses. As a result, the 3:1 amplitude estimate will no longer hold,

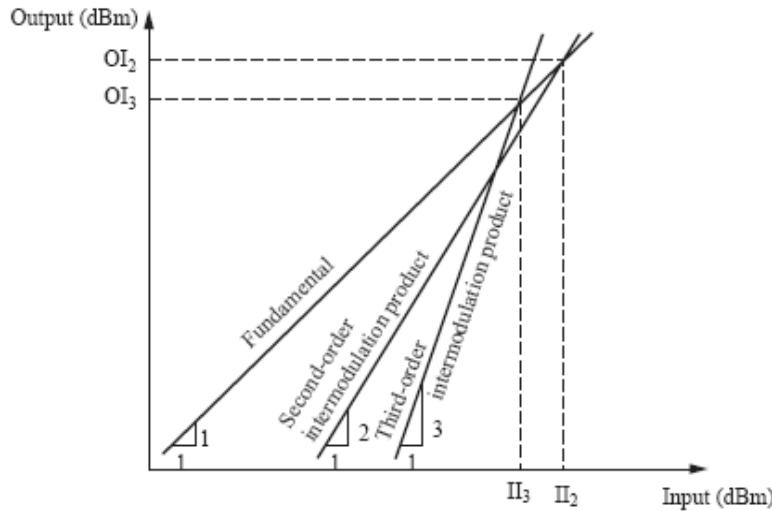
- If the phases of the third- and fifth-degree coefficients are equal, the fifth-degree nonlinearity will expand the IM3 responses. However, if the phases are the opposite, the IM3 distortion will be locally reduced. This explains why notches (sweet-spots) in the IM3 (and high-order) sidebands have been reported at certain amplitudes of output power.



IM(n) products vs Input



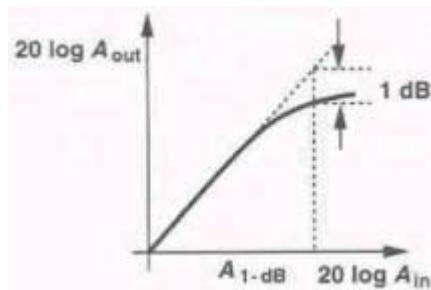
Re-growth of harmonic content vs Conduction Angle



Second- and third-order input and output intercept points Power

- Since the amount of device nonlinearities cannot be changed much, distortion is most effectively minimized by optimizing the impedances seen by the distortion current sources.
- In all the Power Amplifiers, the output level is a “compressive” or “saturating” function of the input level. The gain of the Power Amplifier approaches zero for sufficiently high input levels. In RF circuits this effect is quantified by the “1dB

compression point”, defined as the input signal level that causes the small-signal gain to drop by 1dB. This can be plotted in a log-log scale vs input level.



- Sometimes this Output Power vs Input Power characteristic is referred as AM-AM distortion.
- The asymmetry of side-band intermodulation products (IM) in a two-tone test is often dependent on the carrier spacing, but not in a monotonic fashion. The effect can be explained as an interaction between AM-AM and AM-PM distortion processes. On the other hand, the mere presence of both processes does not guarantee that asymmetry will occur.
- If there is a time lag, or phase shift as *measured in the envelope time domain*, between the AM-AM and AM-PM responses, or their individual frequency components, IM asymmetry will occur.

In the same time, reduction of AM-PM in the PA design would alleviate this problem.

- For RF power transistors, a primary cause of AM-PM effects appears to be the dynamic mistuning of the input match. Some deliberate mistuning on the high Q factor input match of RF power transistors might pay off in terms of improved AM-PM performance for the loss of a decibel or two of gain.

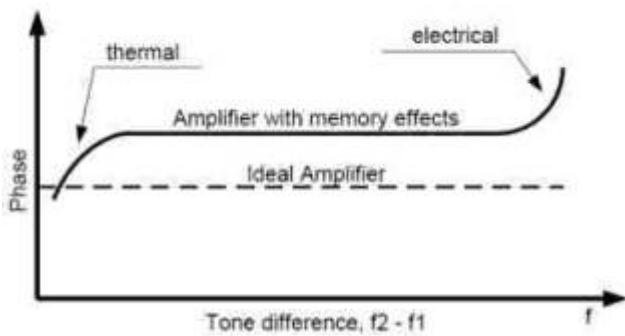
Memory Effect

- In a two or multi tone IMD test if the amplitude and/or phase of the IM signals is affected by the tone difference, the amplifier exhibits *memory effects*.
- Memory is caused by the storage of energy that has to be charged or discharged.
- Memory Effect could be explained as a time lag between AM-PM and AM-PM response of the amplifier. The Electrical Memory Effect is introduced by poor gate/base and drain/collector decoupling at low frequencies causing a distortion of the envelope currents which results in IMD asymmetry.

Low frequencies mean baseband/video frequencies or the frequency spacing between two tones.

The most significant Memory Effect appears in Class AB amplifiers, with reduced conduction angle where drain/collector varies with output power. In Class A amplifiers the Memory Effect reduces.

- Smooth memory effects are not usually harmful to the linearity of the PA itself. A phase rotation of 10° to 20° or an amplitude change of less than 0.5 dB, as a function of modulation frequency, has no dramatic effect on the linearity of the device,
- There are two memory effects: electrical and thermal.



- Electrical memory effects are produced by nonconstant node impedances within frequency bands as DC, Fundamental, and Harmonics. Most of these effects are generated by frequency dependent envelope impedance, and those within the DC band are the most harmful, because bias impedances are strongly frequency dependent.
- Thermal memory effects are generated by the junction temperature, which is modulated by the applied signal.
- Thermal effects will be much more prominent in a slow sweep, or a stepped CW test.
- The analysis and simulation show that the drain envelope impedance is the most important factor for reducing the memory effect and nonlinearity. A new matching topology is proposed for minimizing the drain and gate envelope impedances. The matching topology consists of a series LC circuit for shorting the device at a low frequency while maintaining a matchable impedance at the operating frequency. The circuits are connected to the gate and drain terminals, rather than to the bias lines, since the circuit can produce a very low impedance, not limited by the quarter-wavelength bias line. The amplifier, with the reduced envelope impedances, provides drastically reduced memory effects and very linear amplification performance for wideband signals.
- Simultaneous amplitude and delayed phase modulation does generate asymmetric sidebands.

Input/Output Matching and Load Line

- The input matching configuration, including the bias circuit, has an important impact on the operation of the RF Power Amplifiers.
- The input match will show different optima for maximum gain, best linearity, and highest efficiency. Optimization of the efficiency may involve substantial reduction in power gain.
- Correct handling of harmonics is a necessary feature on the input, as well as the output, match. Device used well below its cutoff frequency may require specific harmonic terminating circuit elements on the input.
- The performance of the output matching circuit is critical for a Power Amplifier.
- In a PA, impedances control how much power is delivered to the output and how much gain and noise are produced in the process, therefore matching network is critical for maximum performance.
- One aspect that's sometimes overlooked is the power dissipation in the output matching circuit. This power is lost in the capacitors, inductors, and other lossy

elements that are part of the matching network. This "dissipation loss" degrades the PA's efficiency and output power capability.

- Different implementations of the output match result in different losses and there are still significant design tradeoffs to be made between bandwidth and dissipation loss. For a PA, the loss of the output match is always a concern because of the large power levels involved.
- A capacitor's quality factor is reversal proportional to its capacitance. To minimize the dissipation loss of the output match, it's therefore necessary to design the output match with the lowest possible value of C. The tradeoff is between bandwidth and dissipation loss.
- Different capacitor technologies give different losses when used in output matching circuits.
- One way of understanding the loss mechanisms of an output match and to don't mix up mismatch loss and dissipation loss, is to simulate the match with loss-less components, then introduce loss into one component at a time.

$$\text{Mismatch Loss [dB]} = 10 \cdot \text{LOG} (1 - \Gamma^2)$$

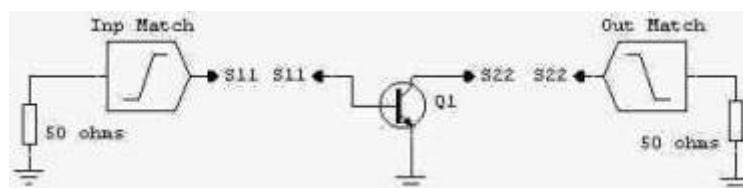
where reflection coefficient $\Gamma = (\text{VSWR}-1) / (\text{VSWR}+1)$

- Because the dissipation loss doesn't depend on the source impedance it's possible to use S21 to find the correct dissipation loss in a circuit simulation. The procedure involves using the complex conjugate of the simulated load line as the source impedance.
- Running at a low efficiency not only reduces talk time in a portable device, but it also creates significant problems with heating and reliability.
- The load line is set based on the needed Power Amplifier output power and available supply voltage. For example low voltage PA's (~3.5V for mobile devices) have a load line ranging from 1 to 5 Ω .

$$R_L = V_{\max} / I_{\max}$$

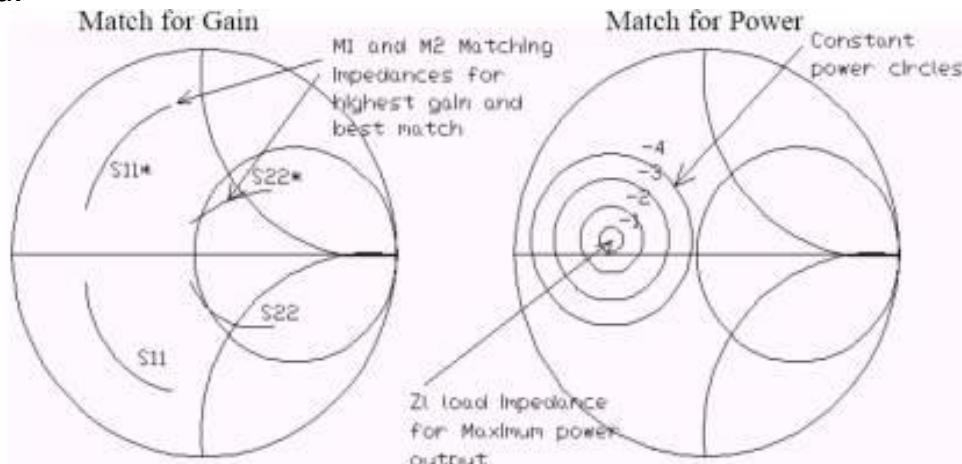
- Matching for maximum Gain occurs when the amplifier is unconditional stable and load impedance is equal to the complex conjugate of the same source impedance (conjugate matching). Complex conjugate simply refers to complex impedance having the same real part with an opposite reactance.

for example: if the source impedance is $Z_s=R+jX$, then its complex conjugate would be $Z_s^*=R-jX$

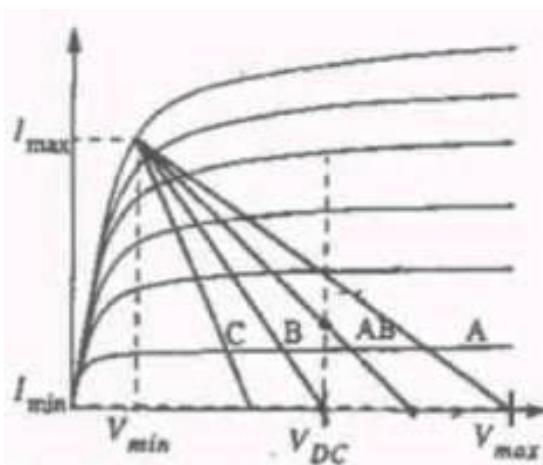


- Matching for maximum Output Power occurs when Optimum Load impedance (R_L) is equal to Source impedance (R_{gen}).

- The transistor's input and output impedances will also decrease with an increase in frequency, which further complicates the design of a PA's matching networks, especially since these impedances can be as low as 0.5Ω . Thus, when matching a discrete driver stage to its PA with maximum efficiency, we would normally want to implement a direct match from the true output impedance of the driver to the true input impedance of the PA, instead of first forming a $50\text{-}\Omega$ match at the output of the driver, and then another $50\text{-}\Omega$ match for the input of the power amplifier, as this would needlessly transform the impedances from low to high, and then back from high to low.
- By selecting a transistor with a high collector voltage requirement, we can increase its output impedance over a transistor that operates at lower values of collector voltage.
- Output Matching network for most high-powered amplifiers should normally consist of the T type, rather than the PI type. PI matching networks for high powered amplifiers sometimes result in unrealistic component values when matching for the higher operating frequencies encountered into a 50Ω load. Indeed, T networks are capable of much higher-frequency operation before this becomes a major problem. Both T and PI networks can be used, however, if the output impedance of the transistor is higher than its load.



Smith Chart representation for maximum gain and power matching



Load Line for different classes

Optimum Load Resistance

In the absence of collector output resistance information on the datasheet, it becomes necessary to make a simple calculation to determine the optimum load resistance for the transistor.

The value of load resistance is dependent upon power level required and is given by:

$$R_L = \frac{(V_{CC} - V_{SAT})^2}{2P} \quad \text{or} \quad R_L = \frac{(V_{CC})^2}{2P} \quad (\text{with less accuracy})$$

where,

V_{CC} = the supply voltage

V_{SAT} = the saturation voltage of the transistor

P = the output power level required in Watts

Note that this equation provides only the *load resistance*, when usually in the datasheets the manufacturer provides values of shunt output capacitance vs frequency for the RF power transistor.

Power Amplifier Bias Design

- There are many different ways to bias an amplifier, depending on the required temperature stability, efficiency, cost, device, power output, linearity, and so on.
- Power BJT transistor biasing:
 - Must force the DC (average) value of V_{CE} and I_C to desired values and keep them constant using feedback techniques.
 - Whether employing diode or transistor bias, it is essential to thermally connect these components to the RF transistor itself. This allows the semiconductor bias components to track the power amplifier's temperature variations.
- Power FET transistor biasing:

The gate biasing circuit has several functions:

 - To maintain a constant gate-to-source voltage, V_{GS} .
 - To be able to supply a negative and positive gate current, I_{GS} .
 - To protect the gate by limiting I_{GS} when the device goes into breakdown (drain-to-gate or gate-to-source) or when the gate-to-source junction is biased with a positive voltage. These abnormal operating conditions for the devices can be due to an operator error, an overdrive, a system problem or ESDs.
 - To stabilize the device in case a negative resistance appears in the gate at any frequency where the device has a positive gain.
 - To filter the signal, the products and the harmonics generated by the device input from low to high frequencies without affecting the device input matching circuit.
 - To isolate the gate from any signal coming from the drain through the bias circuits.

- Power LDMOS transistor biasing

The main consideration of the power LDMOS biasing is to achieve the linearity.

- This is done DC biasing of the LDMOS transistor for optimal drain current for a given power output.
- This bias needs to be held constant over temperature and time. Typically the target accuracy for bias current over temperature is $\pm 5\%$ but $\pm 3\%$ is much more desirable for a high performance design.
- The DC Bias on LDMOS amplifiers is set by applying a DC voltage to the gate (V_{gs}) and monitoring the Drain current (I_{dd}).
- Ideally, this I_{dd} will be constant over temperature, but since the V_{gs} of LDMOS amplifier devices varies with temperature, some type of temperature compensation is required. For optimal temperature compensation, in-circuit adjustments need to be made for both the temperature compensation as well as the V_{gs} bias itself.

Power Amplifier Design Issues

- Reflected power caused by a high VSWR condition between a PA and its load does not, in and of itself, cause a transistor's destruction or damage. Rather, a PA can be damaged or destroyed in a high VSWR environment simply because it is now looking at a completely *different* load impedance than it was designed for.
- High device power dissipation can then produce elevated heating of the transistor and/or excessively high voltages.
- Some of the DC bias voltage would be wasted if the resistance in the chokes were too high. Also the collector chokes must supply a very high impedance to the RF. If this impedance is not high enough, then some of the valuable RF output power generated by the PA will be wasted.
- Use low-ESR electrolytic capacitors at the PA's power supply, as this type of capacitor can immediately supply the needed current to the amplifier stage, and without pulling down the entire voltage supply during this critical transient turn-on time.
- When in saturation, a nonlinear PA's gain, PAE, and linearity are most affected by the reflections of its own harmonics back into its output port. These reflections are caused by the next stage, which will normally be a bandpass or lowpass filter, as well as an antenna.
- Instability in RF amplifiers can take the form of oscillations at almost any frequency, and may even damage or destroy the transistor. These spurious oscillations will arise at specific, or very wide ranging, frequency or frequencies, and over a particular bias, drive level, temperature, or output load impedance.
- Decreasing the low-frequency gain of a PA stage, which is naturally at an increased level, will assist in amplifier stability.
- The unavailability of a sufficient groundplane, or a groundplane that is excessively segmented, can create uncontrollable instability in a PA.

RF Power Amplifiers for Wideband Modulations

RF Power Amplifiers for wideband modulations as CDMA or WCDMA, which operate in the linear region, are not very efficient. Only a portion of the D.C. current is used to generate the RF power; a much larger portion turns into heat.

LDMOS and GaN (Gallium Nitride) devices are best suited for the output and driver stages because of higher gain, improved linearity, and very low on-resistance. High gain reduces the number of stages needed in the amplifier to attain the same output power, compared to the old generation systems built with bipolar transistors.

In a multi-stage linear power amplifier there are various factors that need to be considered for choosing the right transistor for each of the stages of the amplifier.

- The pre-driver is biased Class-A for attaining consistent performance for minimal effect on the linearity of the device due to minor changes in bias supply. Drain efficiency is not as much of a concern for the pre-driver as it is for the latter stages in the amplifier. The driver and the output stage for such a system are typically biased Class AB, for achieving best tradeoff between linearity and efficiency of the amplifier.
- The most common method used to determine the linearity of a transistor is to characterize the Intermodulation Distortion (IMD) measured with two tones spaced. Typically, a tone spacing up to 20 MHz should be used while tuning amplifiers for wide-band modulation applications.
- When transistors are used significantly backed-off from their peak power levels, it is all the more necessary that the IMD characteristics of the transistor at lower output power levels be taken into account. The profile of an IMD vs. Pout (drive-up) curve for a good transistor should have a large positive slope, even while attaining similar peak power capability, to get maximum Adjacent Channel Power Ratio (ACPR).
- The transistor used in the driver stage has similar linearity requirements as the output stage. In terms of ACPR, it needs be operated at an output power that gives a margin of at least 4 dB from the maximum allowed value for the output stage. In addition, it needs to have an input bandwidth about 2 to 2.5 times greater than the bandwidth of the modulating signal in order to maintain constant group delay and flat gain.
- One of the major factors determining the performance of the high power transistor in the wide modulation environment is the gain flatness. The transistor needs to have a flat gain across the band for its use in multiple channel amplifiers. Very flat gain response greatly simplifies the design of linearization schemes systems.
- Fast roll-off of gain at the edges of the band causes deterioration in the ACPR performance.
- To attain the intrinsic device linearity, the 3 dB bandwidth of bias networks needs to be at least two times greater than the modulation bandwidth.
- For attaining best ACPR response, it is necessary to have an excellent decoupling network at the Drain of the transistor, down to very low modulation frequencies. This can be achieved by using a high quality shunt capacitor. This technique helps in achieving maximal gain flatness, which is very critical for wideband applications.
- It is advisable to avoid ferrite components in the biasing network and to use a series resistor on the Gate bias network to prevent instability. To achieve flat gain response across the band, the traditional inductive feed should be avoided on both the Gate and Drain. Instead, a quarter wave line at the frequency of interest, properly decoupled with a chip-capacitor, has been shown to provide very flat gain across the entire bandwidth.

References:

1. *RF Circuit Design* – C. Bowick
2. *RF Power Amplifiers for Wireless Communications* – S. Cripps
3. *Advanced Techniques in RF Power Amplifier Design* – S. Cripps
4. *Distortion in RF Power Amplifiers* – J. Vuolevi
5. *Circuit Design for RF Transceivers* – D. Leenaerts, J. Tang, C. Vaucher
6. *Radio Frequency Transistors* - N. Dye, H. Granberg
7. *High Frequency Current Mode Class-D Amplifiers With High Output Power and Efficiency* - A. L. Long
8. *Complete Wireless Design* - C. Sayre
9. *Feedforward linear power amplifiers* – N. Pothecary
10. *Microwave Journal Magazine*; 1996 – 2005
11. *Portable Design Magazine*; 2002 - 2005
12. *High Frequency Electronics Magazine*; 2002 - 2007

[Home](http://www.qsl.net/va3iul) <http://www.qsl.net/va3iul>

RF and Microwave Power Amplifier and Transmitter Technologies — Part 1

By Frederick H. Raab, Peter Asbeck, Steve Cripps, Peter B. Kenington, Zoya B. Popovic, Nick Pothecary, John F. Sevic and Nathan O. Sokal

With this issue, we begin a four-part series of articles that offer a comprehensive overview of power amplifier technologies. Part 1 covers the key topics of amplifier linearity, efficiency and available RF power devices

background for the subject, and begins the technical discussion with material on signals, linearity, efficiency, and RF-power devices. At the end, there is a convenient summary of the acronyms used—this will be provided with all four installments. Author affiliations and contact information are also provided at the end of each part.

1. INTRODUCTION

The generation of significant power at RF and microwave frequencies is required not only in wireless communications, but also in applications such as jamming, imaging, RF heating, and miniature DC/DC converters. Each application has its own unique requirements for frequency, bandwidth, load, power, efficiency, linearity, and cost. RF power can be generated by a wide variety of techniques using a wide variety of devices. The basic techniques for RF power amplification via classes A, B, C, D, E, and F are reviewed and illustrated by examples from HF through Ka band. Power amplifiers can be combined into transmitters in a similarly wide variety of architectures, including linear, Kahn, envelope

tracking, outphasing, and Doherty. Linearity can be improved through techniques such as feedback, feedforward, and predistortion. Also discussed are some recent developments that may find use in the near future.

A power amplifier (PA) is a circuit for converting DC input power into a significant amount of RF/microwave output power. In most cases, a PA is not just a small-signal amplifier driven into saturation. There exists a great variety of different power amplifiers, and most employ techniques beyond simple linear amplification.

A transmitter contains one or more power amplifiers, as well as ancillary circuits such as signal generators, frequency converters, modulators, signal processors, linearizers, and power supplies. The classic architecture employs progressively larger PAs to boost a low-level signal to the desired output power. However, a wide variety of different architectures in essence disassemble and then reassemble the signal to permit amplification with higher efficiency and linearity.

Modern applications are highly varied. Frequencies from VLF through millimeter wave are used for communication, navigation, and broadcasting. Output powers vary from 10 mW in short-range unlicensed wireless systems to 1 MW in long-range broadcast transmitters. Almost every conceivable type of modulation is being used in one system or another. PAs and transmitters also find use in systems such as radar, RF heating, plasmas, laser drivers, magnetic-resonance imaging, and miniature DC/DC converters.

This series of articles is an expanded version of the paper, "Power Amplifiers and Transmitters for RF and Microwave" by the same authors, which appeared in the the 50th anniversary issue of the *IEEE Transactions on Microwave Theory and Techniques*, March 2002. © 2002 IEEE. Reprinted with permission.

No single technique for power amplification nor any single transmitter architecture is best for all applications. Many of the basic techniques that are now coming into use were devised decades ago, but have only recently been made practical because of advances in RF-power devices and supporting circuitry such as digital signal processing (DSP).

2. HISTORICAL DEVELOPMENT

The development of RF power amplifiers and transmitters can be divided into four eras:

Spark, Arc, and Alternator

In the early days of wireless communication (from 1895 to the mid 1920s), RF power was generated by spark, arc, and alternator techniques. The original RF-power device, the spark gap, charges a capacitor to a high voltage, usually from the AC mains. A discharge (spark) through the gap then rings the capacitor, tuning inductor, and antenna, causing radiation of a damped sinusoid. Spark-gap transmitters were relatively inexpensive and capable of generating 500 W to 5 kW from LF to MF [1].

The arc transmitter, largely attributed to Poulsen, was a contemporary of the spark transmitter. The arc exhibits a negative-resistance characteristic which allows it to operate as a CW oscillator (with some fuzziness). The arc is actually extinguished and reignited once per RF cycle, aided by a magnetic field and hydrogen ions from alcohol dripped into the arc chamber. Arc transmitters were capable of generating as much as 1 MW at LF [2].

The alternator is basically an AC generator with a large number of poles. Early RF alternators by Tesla and Fessenden were capable of operation at LF, and a technique developed by Alexanderson extended the operation to LF [3]. The frequency was controlled by adjusting the rotation speed and up to 200 kW could be

generated by a single alternator. One such transmitter (SAQ) remains operable at Grimeton, Sweden.

Vacuum Tubes

With the advent of the DeForest audion in 1907, the thermionic vacuum tube offered a means of electronically generating and controlling RF signals. Tubes such as the RCA UV-204 (1920) allowed the transmission of pure CW signals and facilitated the transition to higher frequencies of operation.

Younger readers may find it convenient to think of a vacuum tube as a glass-encapsulated high-voltage FET with heater. Many of the concepts for modern electronics, including class-A, -B, and -C power amplifiers, originated early in the vacuum-tube era. PAs of this era were characterized by operation from high voltages into high-impedance loads and by tuned output networks. The basic circuits remained relatively unchanged throughout most of the era.

Vacuum tube transmitters were dominant from the late 1920s through the mid 1970s. They remain in use today in some high power applications, where they offer a relatively inexpensive and rugged means of generating 10 kW or more of RF power.

Discrete Transistors

Discrete solid state RF-power devices began to appear at the end of the 1960s with the introduction of silicon bipolar transistors such as the 2N6093 (75 W HF SSB) by RCA. Power MOSFETs for HF and VHF appeared in 1974 with the VMP-4 by Siliconix. GaAs MESFETs introduced in the late 1970s offered solid state power at the lower microwave frequencies.

The introduction of solid-state RF-power devices brought the use of lower voltages, higher currents, and relatively low load resistances. Ferrite-loaded transmission line transformers enabled HF and VHF

PAs to operate over two decades of bandwidth without tuning. Because solid-state devices are temperature-sensitive, bias stabilization circuits were developed for linear PAs. It also became possible to implement a variety of feedback and control techniques through the variety of opamps and ICs.

Solid-state RF-power devices were offered in packaged or chip form. A single package might include a number of small devices. Power outputs as high as 600 W were available from a single packaged push-pull device (MRF157). The designer basically selected the packaged device that best fit the requirements. How the transistors were made was regarded as a bit of sorcery that occurred in the semiconductor houses and was not a great concern to the ordinary circuit designer.

Custom/Integrated Transistors

The late 1980s and 1990s saw a proliferation variety of new solid-state devices including HEMT, pHEMT, HFET, and HBT, using a variety of new materials such as InP, SiC, and GaN, and offering amplification at frequencies to 100 GHz or more. Many such devices can be operated only from relatively low voltages. However, many current applications need only relatively low power. The combination of digital signal processing and microprocessor control allows widespread use of complicated feedback and predistortion techniques to improve efficiency and linearity.

Many of the newer RF-power devices are available only on a made-to-order basis. Basically, the designer selects a semiconductor process and then specifies the size (e.g., gate periphery). This facilitates tailoring the device to a specific power level, as well as incorporating it into an RFIC or MMIC.

3. LINEARITY

The need for linearity is one of the principal drivers in the design of

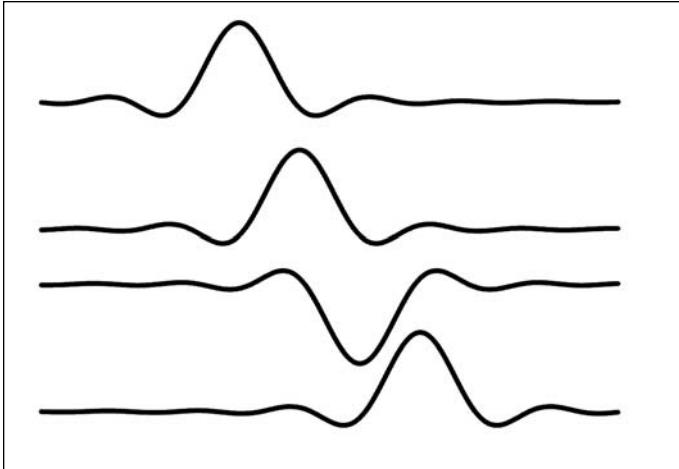


Figure 1 · SRRC data pulses.

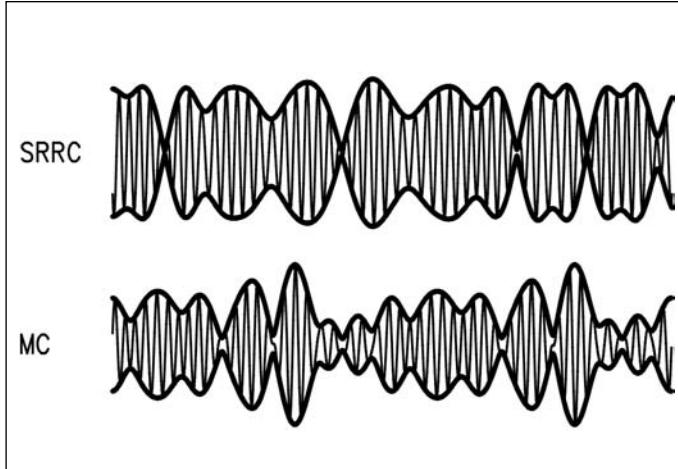


Figure 2 · RF waveforms for SRRC and multicarrier signals.

modern power amplifiers. Linear amplification is required when the signal contains both amplitude and phase modulation. It can be accomplished either by a chain of linear PAs or a combination of nonlinear PAs. Nonlinearities distort the signal being amplified, resulting in splatter into adjacent channels and errors in detection.

Signals such as CW, FM, classical FSK, and GMSK (used in GSM) have constant envelopes (amplitudes) and therefore do not require linear amplification. Full-carrier amplitude modulation is best produced by high level amplitude modulation of the final RF PA. Classic signals that require linear amplification include single sideband (SSB) and vestigial-sideband (NTSC) television. Modern signals that require linear amplification include shaped-pulse data modulation and multiple carriers.

Shaped Data Pulses

Classic FSK and PSK use abrupt frequency or phase transitions, or equivalently rectangular data pulses. The resultant RF signals have constant amplitude and can therefore be amplified by nonlinear PAs with good efficiency. However, the resultant sinc-function spectrum spreads signal energy over a fairly wide bandwidth. This was satisfactory for rela-

tively low data rates and a relatively uncrowded spectrum.

Modern digital signals such as QPSK or QAM are typically generated by modulating both I and Q subcarriers. The requirements for both high data rates and efficient utilization of the increasingly crowded spectrum necessitates the use of shaped data pulses. The most widely used method is based upon a raised-cosine channel spectrum, which has zero intersymbol interference during detection and can be made arbitrarily close to rectangular [4]. A raised-cosine channel spectrum is achieved by using a square-root raised-cosine (SRRC) filter in both the transmitter and receiver. The resultant SRRC data pulses (Figure 1) are shaped somewhat like sinc functions which are truncated after several cycles. At any given time, several different data pulses contribute to the I and Q modulation waveforms. The resultant modulated carrier (Figure 2) has simultaneous amplitude and phase modulation with a peak-to-average ratio of 3 to 6 dB.

Multiple Carriers and OFDM

Applications such as cellular base stations, satellite repeaters, and multi-beam “active-phased-array” transmitters require the simultaneous amplification of multiple signals.

Depending on the application, the signals can have different amplitudes, different modulations, and irregular frequency spacing.

In a number of applications including HF modems, digital audio broadcasting, and high-definition television, it is more convenient to use a large number of carriers with low data rates than a single carrier with a high data rate. The motivations include simplification of the modulation/demodulation hardware, equalization, and dealing with multipath propagation. Such Orthogonal Frequency Division Multiplex (OFDM) techniques [5] employ carriers with the same amplitude and modulation, separated in frequency so that modulation products from one carrier are zero at the frequencies of the other carriers.

Regardless of the characteristics of the individual carriers, the resultant composite signal (Figure 2) has simultaneous amplitude and phase modulation. The peak-to-average ratio is typically in the range of 8 to 13 dB.

Nonlinearity

Nonlinearities cause imperfect reproduction of the amplified signal, resulting in distortion and splatter. Amplitude nonlinearity causes the instantaneous output amplitude or

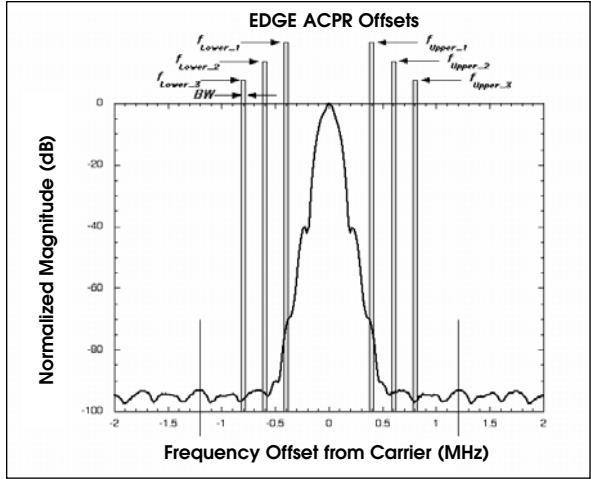


Figure 3 · ACPR offsets and bandwidths.

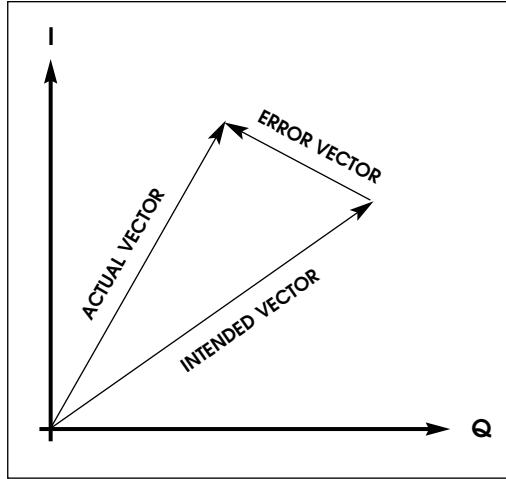


Figure 4 · Error vector.

envelope to differ in shape from the corresponding input. Such nonlinearities are the variable gain or saturation in a transistor or amplifier. Amplitude-to-phase conversion is a phase shift associated with the signal amplitude and causes the introduction of unwanted phase modulation into the output signal. Amplitude-to-phase conversion is often associated with voltage-dependent capacitances in the transistors. While imperfect frequency response also distorts a signal, it is a linear process and therefore does not generate out-of-band signals.

Amplitude nonlinearity and amplitude-to-phase conversion are described by transfer functions that

act upon the instantaneous signal voltage or envelope. However, memory effects can also occur in high-power PAs because of thermal effects and charge storage. Thermal effects are somewhat more noticeable in III-V semiconductors because of lower thermal conductivity, while charge-storage effects are more prevalent in overdriven BJT PAs.

Measurement of Linearity

Linearity is characterized, measured, and specified by various techniques depending upon the specific signal and application. The linearity of RF PAs is typically characterized by C/I, NPR, ACPR, and EVM (defined below).

The traditional measure of linearity is the carrier-to-intermodulation (C/I) ratio. The PA is driven with two or more carriers (tones) of equal amplitudes. Nonlinearities cause the production of intermodulation products at frequencies corresponding to sums and differences of multiples of the carrier frequencies

[6]. The amplitude of the third-order or maximum intermodulation distortion (IMD) product is compared to that of the carriers to obtain the C/I. A typical linear PA has a C/I of 30 dB or better.

Noise-Power Ratio (NPR) is a traditional method of measuring the linearity of PAs for broadband and noise-like signals. The PA is driven with Gaussian noise with a notch in one segment of its spectrum. Nonlinearities cause power to appear in the notch. NPR is the ratio of the notch power to the total signal power.

Adjacent Channel Power Ratio (ACPR) characterizes how nonlinearity affects adjacent channels and is widely used with modern shaped-pulse digital signals such as NADC and CDMA. Basically, ACPR is the ratio of the power in a specified band outside the signal bandwidth to the rms power in the signal (Figure 3). In some cases, the actual power spectrum $S(f)$ is weighted by the frequency response $H(f)$ of the pulse-shaping filter; i.e. (eq. 1)

$$ACPR_{lower} = \frac{\int_{f_c-f_o-BW/2}^{f_c-f_o+BW/2} |H(f)|^2 S(f) df}{\int_{f_L}^{f_U} |H(f)|^2 S(f) df}$$

STANDARD	Offset 1	Offset 2	BW (kHz)	Integration Filter	EVM (peak/rms)
NADC [13]	± 30 kHz -26 dBc	± 60 kHz -45 dBc	32.8 kHz	RRC $\alpha=0.35$	25%/12%
PHS [14]	± 600 kHz -50 dBc	± 900 kHz -55 dBc	37.5 kHz	RRC $\alpha=0.50$	25%/12%
EDGE [15]	± 400 kHz -58 dBc	± 600 kHz -66 dBc	30 kHz	None	22%/7.0%
TETRA [16]	25 kHz -60 dBc	50 kHz -70 dBc	25 kHz	RRC $\alpha=0.35$	30%/10%
IS-95 CDMA [17]	885 kHz -45 dBc	1980 kHz -55 dBc	30 kHz	None	N/A
W-CDMA (3G-PP) [18]	5.00 MHz -33 dB	10.0 MHz -43 dB	4.68 MHz	RRC $\alpha=0.22$	25%/N/A

Table 1 · ACPR and EVM requirements of various systems.

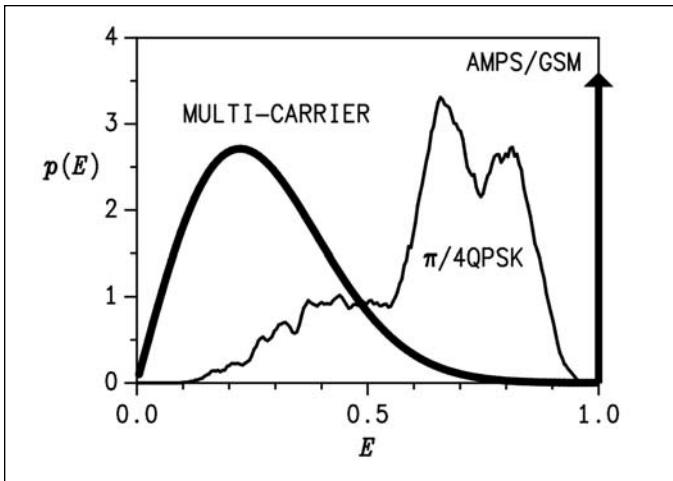


Figure 5 · Envelope PDFs.

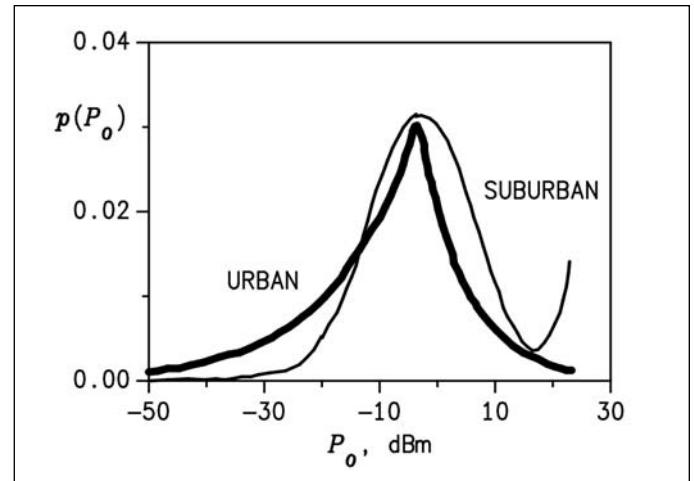


Figure 6 · Power-output PDFs.

where f_c is the center frequency, B is the bandwidth, f_o is the offset, and f_L and f_U are the band edges. The weighting, frequency offsets, and required ACPRs vary with application as shown in Table 1. ACPR can be specified for either upper or lower sideband. In many cases, two different ACPRs for two different frequency offsets are specified. ACPR2, based upon the outer band, is sometimes called "Alternate Channel Power Ratio."

Error Vector Magnitude (EVM) is a convenient measure of how nonlinearity interferes with the detection process. EVM is defined (Figure 4) as the distance between the desired and actual signal vectors, normalized to a fraction of the signal amplitude. Often, both peak and rms errors are specified (Table 1).

4. EFFICIENCY

Efficiency, like linearity, is a critical factor in PA design. Three definitions of efficiency are commonly used. Drain efficiency is defined as the ratio of RF output power to DC input power:

$$\eta = P_{out}/P_{in} \quad (2)$$

Power-added efficiency (PAE) incorporates the RF drive power by subtracting it from the output power;

i.e. $(P_{out} - P_{DR})/P_{in}$. PAE gives a reasonable indication of PA performance when gain is high; however, it can become negative for low gains. An overall efficiency such as $P_{out}/(P_{in} + P_{DR})$ is useable in all situations. This definition can be varied to include driver DC input power, the power consumed by supporting circuits, and anything else of interest.

Average Efficiency

The instantaneous efficiency is the efficiency at one specific output level. For most PAs, the instantaneous efficiency is highest at the peak output power (PEP) and decreases as output decreases. Signals with time-varying amplitudes (amplitude modulation) therefore produce time-varying efficiencies. A useful measure of performance is then the average efficiency, which is defined [7] as the ratio of the average output power to the average DC-input power:

$$\eta_{AVG} = P_{outAVG}/P_{inAVG} \quad (3)$$

This concept can be used with any of the three definitions of efficiency.

The probability-density function (PDF) of the envelope gives the relative amount of time an envelope spends at various amplitudes (Figure 5). Also used is the cumulative distri-

bution function (CDF), which gives the probability that the envelope does not exceed a specified amplitude. CW, FM, and GSM signals have constant envelopes and are therefore always at peak output. SRRC data modulation produces PDFs that are concentrated primarily in the upper half of the voltage range and have peak-to-average ratios on the order of 3 to 6 dB. Multiple carriers [8] produce random-phasor sums much like random noise and therefore have Rayleigh-distributed envelopes; i.e.,

$$p(E) = 2E\xi \exp(-V_2\xi) \quad (4)$$

Peak-to-average ratio ξ is typically between 6 and 13 dB.

The average input and output powers are found by integrating the product of their variation with amplitude and the PDF of the envelope. Two cases are of special interest. When the DC input current is constant (class-A bias), the DC input power is also constant. The average efficiency is then η_{PEP}/ξ . If the DC input current (hence power) is proportional to the envelope (as in class-B), the average efficiency is $(4/\pi)^{1/2}\eta_{PEP}$, for a Rayleigh-distributed signal. Thus for a multicarrier signal with a 10 dB peak-to-average ratio, ideal class-A and B PAs with PEP efficiencies of 50 and 78.5 percent,

respectively, have average efficiencies of only 5 and 28 percent, respectively.

Back-Off

The need to conserve battery power and to avoid interference to other users operating on the same frequency necessitates the transmission of signals whose peak amplitudes well below the peak output power of the transmitter. Since peak power is needed only in the worst-case links, the “back-off” is typically in the range of 10 to 20 dB.

For a single-carrier mobile transmitter, back-off rather than envelope PDF is dominant in determining the average power consumption and average efficiency. The PDF of the transmitting power (Figure 4) depends not only upon the distance, but also upon factors such as attenuation by buildings, multipath, and orientation of the mobile antenna [8], [9], [10]. To facilitate prediction of the power consumption, the envelope and back-off PDFs can be combined [11].

5. RF POWER TRANSISTORS

A wide variety of active devices is currently available for use in RF-power amplifiers, and RF-power transistors are available in packaged, die, and grown-to-order forms. Packaged devices are used at frequencies up to X band, and are dominant for high power and at VHF and lower frequencies. A given package can contain one or more die connected in parallel and can also include internal matching for a specific frequency of operation. Dice (chips) can be wire-bonded directly into a circuit to minimize the effects of the package and are used up to 20 GHz. In MMICs, the RF-power device is grown to order, allowing its size and other characteristics to be optimized for the particular application. This form of construction is essential for upper-microwave and millimeter-wave frequencies to minimize the effects of strays and interconnects. Virtually all RF power transistors

are npn or n-channel types because the greater mobility of electrons (vs. holes) results in better operation at higher frequencies.

Bipolar Junction Transistor (BJT)

The Si BJT is the original solid-state RF power device, originating in the 1960s. Since the BJT is a vertical device, obtaining a high collector-breakdown voltage is relatively simple and the power density is very high. Si BJTs typically operate from 28 V supplies and remain in use at frequencies up to 5 GHz, especially in high-power (1 kW) pulsed applications such as radar. While Si RF power devices have higher gain at high frequencies, their fundamental properties are basically those of ordinary bipolar transistors. The positive temperature coefficient of BJTs tends to allow current hogging, hot-spotting, and thermal runaway, necessitating carefully regulated base bias. Since RF power BJTs are generally composed of multiple, small BJTs (emitter sites), emitter ballasting (resistance) is generally employed to force even division of the current within a given package.

Metal-Oxide-Silicon Field-Effect Transistor (MOSFET)

MOSFETs are constructed with insulated gates. Topologies with both vertical and later current flow are used in RF applications, and most are produced by a double-diffusion process. Because the insulated gate conducts no DC current, MOSFETs are very easily biased.

The negative temperature coefficient of a MOSFET causes its drain current to decrease with temperature. This prevents thermal runaway and allows multiple MOSFETs to be connected in parallel without ballasting. The absence of base-charge storage time allows fast switching and also eliminates a mechanism for sub-harmonic oscillation. An overdriven (saturated) MOSFET can conduct drain current in either direction,

which is very useful in switching-mode operation with reactive loads.

Vertical RF power MOSFETs are useable through VHF and UHF. Gemini-packaged devices can deliver up to 1 kW at HF and 100s of watts at VHF. VMOS devices typically operate from 12, 28, or 50-V supplies, although some devices are capable of operation from 100 V or more.

Laterally Diffused MOS (LDMOS)

LDMOS is especially useful at UHF and lower microwave frequencies because direct grounding of its source eliminates bond-wire inductance that produces negative feedback and reduces gain at high frequencies. This also eliminates the need for the BeO insulating layer commonly used in other RF-power MOSFETs.

LDMOS devices typically operate from 28-V supplies and are currently available with power outputs of 120 W at 2 GHz. They are relatively low in cost compared to other devices for this frequency range and are currently the device of choice for use in high-power transmitters at 900 MHz and 2 GHz.

Junction FET (JFET)

JFETs for power applications are often called Static Induction Transistors (SITs). Impressive power and efficiency have been obtained from RF JFETs based upon Si, SiGe, and SiC at frequencies through UHF. However, the JFET has never become as popular as other RF-power FETs.

GaAs MEtal Semiconductor FET (GaAs MESFET)

GaAs MESFETs are JFETs based upon GaAs and a Schottky gate junction. They have higher mobility than do Si devices and are therefore capable of operating efficiently at higher frequencies. GaAs MESFETs are widely used for the production of microwave power, with capabilities of up 200 W at 2 GHz and 40 W at 20 GHz in packaged devices. These

devices have relatively low breakdown voltages compared to MOSFETs or JFETs and are typically operated from supply voltages (drain biases) of 5 to 10 V. Most MESFETs are depletion-mode devices and require a negative gate bias, although some enhance-mode devices that operate with a positive bias have been developed. Linearity is often poor due to input capacitance variation with voltage; the output capacitance is also often strongly bias- and frequency-dependent.

Heterojunction FET (HFET) / High-Electron-Mobility Transistor (HEMT)

HFETs and HEMTs improve upon the MESFET geometry by separating the Schottky and channel functions. Added to the basic MESFET structure is a heterojunction consisting of an n-doped AlGaAs Schottky layer, an undoped AlGaAs spacer, and an undoped GaAs channel. The discontinuity in the band gaps of AlGaAs and GaAs causes a thin layer of electrons ("two-dimensional electron gas or 2-DEG") to form below the gate at the interface of the AlGaAs and GaAs layers. Separation of the donors from the mobile electrons reduces collisions in the channel, improving the mobility, and hence high-frequency response, by a factor of about two.

AlGaAs has crystal-lattice properties similar to those of GaAs, and this makes it possible to produce a potential difference without lattice stress. The GaAs buffer contributes to a relatively high breakdown voltage. Their fabrication employs advanced epitaxial technologies (Molecular Beam Epitaxy or Metal Organic Chemical Vapor Deposition) which tends to increase their cost.

The GaAs HEMT is known in the literature by a wide variety of different names, including MODFET (Modulation-Doped FET), TEGFET (Two-dimensional Electron-Gas FET), and SDFET (Selectively Doped FET). It is also commonly called an HFET (Heterostructure FET),

although technically an "HFET" has a doped channel that provides the carriers (instead of the heterojunction). The acronyms "HFET" and "HJFET" (HeteroJunction FET) appear to be used interchangeably.

GaAs HEMTs/HFETs with f_T as high as 158 GHz are reported. PAs based upon these HEMTs exhibit 15-W outputs at 12 GHz with a power-added efficiency (PAE) of 50 percent. Outputs of 100 W are available at S band from packaged devices.

Pseudomorphic HEMT

The pseudomorphic HEMT (pHEMT) further improves upon the basic HEMT by employing an InGaAs channel. The increased mobility of In with respect to GaAs increases the bandgap discontinuity and therefore the number of carriers in the two-dimensional electron gas. The lattice mismatch between the GaInAs channel and GaAs substrate is also increased, however, and this limits the In content to about 22 percent.

The efficiency of PAs using pHEMTs does not begin to drop until about 45 GHz and pHEMTs are useable to frequencies as high as 80 GHz. Power outputs vary from 40 W at L band to 100 mW at V band. While pHEMTs are normally grown to order, a packaged device pHEMT has recently become available.

InP HEMT

The InP HEMT places an AlInAs/GaInAs heterojunction on an InP substrate. The lattices are more closely matched, which allows an In content of up to about 53 percent. This results in increased mobility, which in turn results in increased electron velocity, increased conduction-band discontinuity, increased two-dimensional electron gas, and higher transconductance. The thermal resistance is 40 percent lower than that of a comparable device built on a GaAs substrate.

The InP HEMT has higher gain

and efficiency than the GaAs pHEMT, with the PA efficiency beginning to drop at 60 GHz. However, it has a lower breakdown voltage (typically 7 V) and must therefore be operated from a relatively low drain-voltage supply (e.g., 2 V). This results in lower output per device and possibly loss in the combiners required to achieve a specified output power. Nonetheless, the InP HEMT generally has a factor-of-two efficiency advantage over the pHEMT and GaAs HEMT.

InP HEMTs have been fabricated with f_{max} as high as 600 GHz (0.1 μ m gate length), and amplification has been demonstrated at frequencies as high as 190 GHz. The efficiency does not begin to drop until about 60 GHz. Power levels range from 100 to 500 mW per chip.

Metamorphic HEMT (mHEMT)

The mHEMT allows channels with high-In content to be built on GaAs substrates. The higher electron mobility and higher peak saturation velocity result in higher gain than is possible in a pHEMT. mHEMTs are generally limited to low-power applications by their relatively low breakdown voltage (<3 V). However, an mHEMT capable of 6-V operation and a power output of 0.5 W has been recently reported.

Heterojunction Bipolar Transistor (HBT)

HBTs are typically based upon the compound-semiconductor material AlGaAs/GaAs. The AlGaAs emitter is made as narrow as possible to minimize base resistance. The base is a thin layer of p GaAs. The barrier is created by heterojunction (AlGaAs/GaAs) rather than the doping. The base can therefore be doped heavily to minimize its resistance. Base sheet resistance is typically two orders of magnitude lower than that of an ordinary BJT, and the frequency of operation is accordingly higher. The current flow is (in contrast to a MES-

FET) vertical so that surface imperfections have less effect upon performance. The use of a semi-insulating substrate and the higher electron mobility result in reduced parasitics. The DC curves are somewhat similar to those of a conventional BJT, but often contain a saturation resistance as well as saturation voltage. Currently available AlGaAs/GaAs HBTs are capable of producing several watts and are widely used in wireless handsets. GaAs HBTs are also widely used in MMIC circuits at frequencies up to X band and can operate in PAs at frequencies as high as 20 GHz.

SiGe HBT

The use of SiGe rather than Si in the base of the HBT both increases the maximum operating frequency and decreases the base resistance. However, they are generally less efficient than GaAs HBTs and can have lower breakdown voltages. One experimental SiGe HBT is capable of delivering over 200 W at L band.

InP HBT

The use of InP in an HBT further boosts mobility and therefore the high frequency response. In addition, InP HBTs have lower turn-on and knee voltages, resulting in higher gain and efficiency. InP HBTs for RF-power applications incorporate two heterojunctions (AlInAs/GaInAs and GaInAs/InP). The InP in the collector increases the breakdown voltage, allowing higher output power. To date, outputs of about 0.5 W at 20 GHz have been demonstrated, but it is anticipated that operation to 50 or 60 GHz will be possible.

SiC MESFET

The wide band gap of SiC results in both high mobility and high breakdown voltage. An SiC MESFET can therefore have a frequency response comparable to that of a GaAs MESFET, but breakdown voltages double that of Si LDMOS. This results in

power densities of 10 W/mm, which is ten times that of a GaAs MESFET. The high thermal conductivity of the SiC substrate is particularly useful in high-power applications. The higher operating voltage and associated higher load impedance greatly simplify output networks and power combining. SiC MESFETs typically operate from a 48-V supply. Devices with outputs of 10 W are currently available, and outputs of 60 W or more have been demonstrated experimentally. The cost of SiC devices is at presently about ten times that of Si LDMOS.

GaN HEMT

GaN offers the same high breakdown voltage of SiC, but even higher mobility. Its thermal conductivity is, however, lower, hence GaN devices must be built substrate such as SiC or diamond. While the GaN HEMT offers the promise of a high-power, high-voltage device operating at frequencies of 10 GHz or more, it is still in an experimental state. Power outputs of 8 W at 10 GHz with 30 percent efficiency have been demonstrated.

Monolithic Microwave Integrated Circuit (MMIC)

MMICs integrate RF power devices and matching/decoupling elements such as on-chip inductors, capacitors, resistors, and transmission lines. The proximity of these elements to the RF-power devices is essential for input, output, and inter-stage matching at microwave and millimeter-wave frequencies.

References

1. W. J. Bryon, "Arcs and sparks, Part 1," *Communications Quarterly*, vol. 4, no. 2, pp. 27-43, Spring 1994.
2. W. J. Bryon, "The arc method of producing continuous waves," *Communications Quarterly*, vol. 8, no. 3, pp. 47-65, Summer 1998.
3. K. M. MacIlvain and W. H. Freedman, *Radio Library*, Vol. III: *Radio Transmitters and Carrier Currents*, Scranton PA: International Textbook Company, 1928.
4. J. B. Groe and L. E. Larson, *CDMA Mobile Radio Design*, Norwood, MA: Artech House, 2000.
5. R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, Norwood, MA: Artech House, 2000.
6. H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: Wiley, 1980.
7. F. H. Raab, "Average efficiency of power amplifiers," *Proc. RF Technology Expo '86*, Anaheim, CA, pp. 474-486, Jan. 30-Feb. 1, 1986.
8. N. Potheccary, "Feedforward linear power amplifiers," in *Workshop Notes WFB*, Int'l. Microwave Symp., Boston, MA, June 16, 2001.
9. J. F. Sevic, "Statistical characterization of RF power amplifier efficiency for wireless communication systems," *Proc. Wireless Commun. Conf.*, Boulder, CO, pp. 1-4, Aug. 1997.
10. G. Hanington, P.-F. Chen, P. M. Asbeck, and L. E. Larson, "High-efficiency power amplifier using dynamic power-supply voltage for CDMA applications," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 8, pp. 1471-1476, Aug. 1999.
11. I. Kipnis, "Refining CDMA mobile-phone power control," *Microwaves & RF*, vol. 39, no. 6, pp. 71-76, June 2000.
12. J. Staudinger, "Applying switched gain stage concepts to improve efficiency and linearity for mobile CDMA power amplification," *Microwave Journal*, vol. 43, no. 9, pp. 152-162, Sept. 2000.
13. "Mobile station – base station interoperability standard for dual-mode cellular system," ANSI-136 Standard, Telecommun. Industries Assoc., 2000.
14. "Digital cellular communication systems," RCR STD-27, Research and Development Center for Radio Systems (RCR), April 1991.

Table 1 References

Acronyms Used in Part 1

AC	Alternating Current	MESFET	Metal Semiconductor FET
ACPR	Adjacent-Channel Power Ratio	mHEMT	Metamorphic HEMT
BJT	Bipolar-Junction Transistor	MMIC	Microwave Monolithic Integrated Circuit
C/I	Carrier-to-Intermodulation	MOSFET	Metal-Oxide-Silicon Field-Effect Transistor
CDF	Cumulative Distribution Function	NADC	North American Digital Cellular
CDMA	Code-Division Multiple Access	NPR	Noise-Power Ratio
CW	Continuous Wave	NTSC	National Television Standards Committee
DC	Direct Current	OFDM	Orthogonal Frequency-Division Multiplex
DSP	Digital Signal Processing	PA	Power Amplifier
EVM	Error-Vector Magnitude	PAE	Power-Added Efficiency
FET	Field-Effect Transistor	PDF	Probability-Density Function
FSK	Frequency-Shift Keying	PEP	Peak-Envelope Power
GMSK	Gaussian Minimum Shift Keying	pHEMT	Pseudomorphic HEMT
GSM	Global System for Mobile communication	PSK	Phase-Shift Keying
HBT	Heterojunction bipolar transistor	QAM	Quadrature Amplitude Modulation
HEMT	High Electron-Mobility Transistor	QPSK	Quadrature Phase Shift Keying
HFET	Heterojunction FET (also HJFET)	RF	Radio Frequency
IC	Integrated Circuit	SRRC	Square-Root Raised Cosine
JFET	Junction Field-Effect Transistor	SSB	Single SideBand
LDMOS	Laterally Diffused MOS (FET)		

15. "Digital cellular telecommunications system (phase 2+), radio transmission and reception," GSM 5.05 Standard, v. 8.4.1, European Telecommun. Standards Inst., 1999.

16. "Terrestrial trunked radio (TETRA) voice+data air interface," TETRA Draft Standard, European Telecommun. Standards Inst., 1999.

17. "Mobile station – base station interoperability standard for dual-mode wideband spread-spectrum cellular system," TIA/EIA IS-95 Interim Standard, Telecommun. Industries Assoc., July 1993.

18. "UE radio transmission and reception (FDD)," TS 25.101, v. 3.4.1, Third Generation Partnership Project, Technical Specification Group, 1999.

Series Notes

1. The remaining three parts of this series will appear in successive issues of *High Frequency Electronics* (July, September and November 2003 issues).

2. To maintain continuity, all figures, tables, equations and references will be numbered sequentially throughout the entire series.

3. Like all articles in *High Frequency Electronics*, this series will be archived and available for downloading (for personal use by individuals only) online at the magazine website: www.highfrequencyelectronics.com

Author Information

The authors of this series of articles are: Frederick H. Raab (lead author), Green Mountain Radio Research, e-mail: f.raab@ieee.org; Peter Asbeck, University of California at San Diego; Steve Cripps, Hywave Associates; Peter B. Kenington, Andrew Corporation; Zoya B. Popovic, University of Colorado; Nick Pothecary, Consultant; John F. Sevic, California Eastern Laboratories; and Nathan O. Sokal, Design Automation. Readers desiring more information should contact the lead author.

RF and Microwave Power Amplifier and Transmitter Technologies — Part 2

By Frederick H. Raab, Peter Asbeck, Steve Cripps, Peter B. Kenington, Zoya B. Popovich, Nick Pothecary, John F. Sevic and Nathan O. Sokal

Our multi-part series on power amplifier technologies and applications continues with a review of amplifier configurations, classes of operation, device characterization and example applications

methods for power amplifiers operating at HF through microwave frequencies.

6a. BASIC TECHNIQUES FOR RF POWER AMPLIFICATION

RF power amplifiers are commonly designated as classes A, B, C, D, E, and F [19]. All but class A employ various nonlinear, switching, and wave-shaping techniques. Classes of operation differ not in only the method of operation and efficiency, but also in their power-output capability. The power-output capability ("transistor utilization factor") is defined as output power per transistor normalized for peak drain voltage and current of 1 V and 1 A, respectively. The basic topologies (Figures 7, 8 and 9) are single-ended, transformer-coupled, and complementary. The drain voltage and current waveforms of selected ideal PAs are shown in Figure 10.

Class A

In class A, the quiescent current is large enough that the transistor remains at all times in the active region and acts as a current source, controlled by the drive.

This series of articles is an expanded version of the paper, "Power Amplifiers and Transmitters for RF and Microwave" by the same authors, which appeared in the the 50th anniversary issue of the *IEEE Transactions on Microwave Theory and Techniques*, March 2002. © 2002 IEEE. Reprinted with permission.

Part 1 of this series introduced basic concepts, discussed the characteristics of signals to be amplified, and gave background information on RF power devices. Part 2 reviews the basic techniques, ratings, and implementation

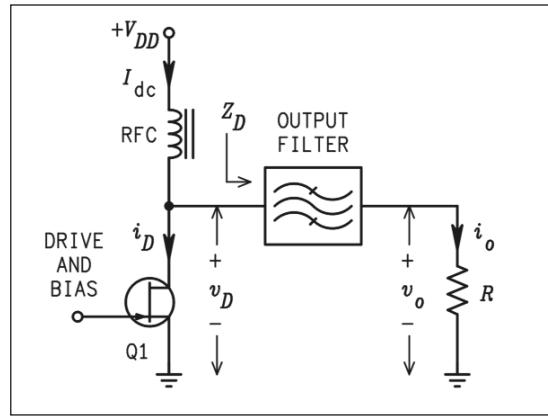


Figure 7 · A single-ended power amplifier.

Consequently, the drain voltage and current waveforms are (ideally) both sinusoidal. The power output of an ideal class-A PA is

$$P_o = V_{om}^2 / 2R \quad (5)$$

where output voltage V_{om} on load R cannot exceed supply voltage V_{DD} . The DC-power input is constant and the efficiency of an ideal PA is 50 percent at PEP. Consequently, the instantaneous efficiency is proportional to the power output and the average efficiency is inversely proportional to the peak-to-average ratio (e.g., 5 percent for $x = 10$ dB). The utilization factor is 1/8.

For amplification of amplitude-modulated signals, the quiescent current can be varied in proportion to the instantaneous signal envelope. While the efficiency at PEP is unchanged, the efficiency for lower ampli-

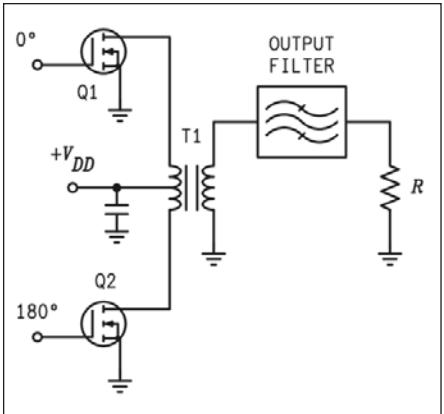


Figure 8 · Transformer-coupled push-pull PA.

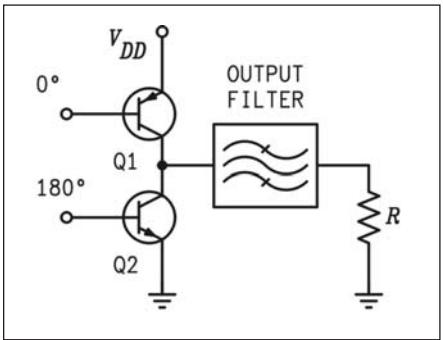


Figure 9 · Complementary PA.

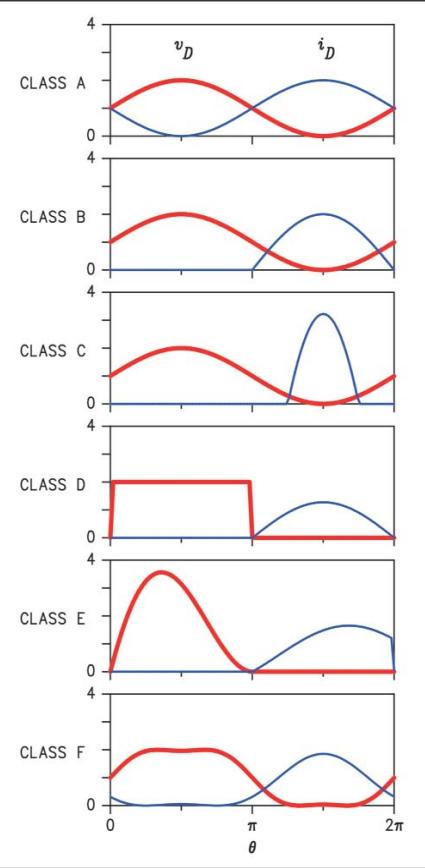


Figure 10 · Waveforms for ideal PAs.

tudes is considerably improved. In an FET PA, the implementation requires little more than variation of the gate-bias voltage.

The amplification process in class A is inherently linear, hence increasing the quiescent current or decreasing the signal level monotonically decreases IMD and harmonic levels. Since both positive and negative excursions of the drive affect the drain current, it has the highest gain of any PA. The absence of harmonics in the amplification process allows class A to be used at frequencies close to the maximum capability (f_{max}) of the transistor. However, the efficiency is low. Class-A PAs are therefore typically used in applications requiring low power, high linearity, high gain, broadband operation, or high-frequency operation.

The efficiency of real class-A PAs is degraded by the on-state resistance

or saturation voltage of the transistor. It is also degraded by the presence of load reactance, which in essence requires the PA to generate more output voltage or current to deliver the same power to the load.

Class B

The gate bias in a class-B PA is set at the threshold of conduction so that (ideally) the quiescent drain current is zero. As a result, the transistor is active half of the time and the drain current is a half sinusoid. Since the amplitude of the drain current is proportional to drive amplitude and the shape of the drain-current waveform is fixed, class-B provides linear amplification.

The power output of a class-B PA is controlled by the drive level and varies as given by eq. (5). The DC input current is, however, proportional to the drain current which is in

turn proportional to the RF-output current. Consequently, the instantaneous efficiency of a class-B PA varies with the output voltage and for an ideal PA reaches $\pi/4$ (78.5 percent) at PEP. For low-level signals, class B is significantly more efficient than class A, and its average efficiency can be several times that of class A at high peak-to-average ratios (e.g., 28 vs. 5 percent for $\xi = 10$ dB). The utilization factor is the same 0.125 of class A.

In practice, the quiescent current is on the order of 10 percent of the peak drain current and adjusted to minimize crossover distortion caused by transistor nonlinearities at low outputs. Class B is generally used in a push-pull configuration so that the two drain-currents add together to produce a sine-wave output. At HF and VHF, the transformer-coupled push-pull topology (Figure 8) is generally used to allow broadband operation with minimum filtering. The use of the complementary topology (Figure 9) has generally been limited to audio, LF, and MF applications by the lack of suitable p-channel transistors. However, this topology is attractive for IC implementation and has recently been investigated for low-power applications at frequencies to 1 GHz [20].

Class C

In the classical (true) class-C PA, the gate is biased below threshold so that the transistor is active for less than half of the RF cycle (Figure 10). Linearity is lost, but efficiency is increased. The efficiency can be increased arbitrarily toward 100 percent by decreasing the conduction angle toward zero. Unfortunately, this causes the output power (utilization factor) to decrease toward zero and the drive power to increase toward infinity. A typical compromise is a conduction angle of 150° and an ideal efficiency of 85 percent.

The output filter of a true class-C PA is a parallel-tuned type that

bypasses the harmonic components of the drain current to ground without generating harmonic voltages. When driven into saturation, efficiency is stabilized and the output voltage locked to supply voltage, allowing linear high-level amplitude modulation.

Classical class C is widely used in high-power vacuum-tube transmitters. It is, however, little used in solid-state PAs because it requires low drain resistances, making implementation of parallel-tuned output filters difficult. With BJTs, it is also difficult to set up bias and drive to produce a true class-C collector-current waveform. The use of a series-tuned output filter results in a mixed-mode class-C operation that is more like mistuned class E than true class C.

Class D

Class-D PAs use two or more transistors as switches to generate a square drain-voltage waveform. A series-tuned output filter passes only the fundamental-frequency component to the load, resulting in power outputs of $(8/\pi^2)V_{DD}^2/R$ and $(2/\pi^2)V_{DD}^2/R$ for the transformer-coupled and complementary configurations, respectively. Current is drawn only through the transistor that is on, resulting in a 100-percent efficiency for an ideal PA. The utilization factor ($1/2\pi = 0.159$) is the highest of any PA (27 percent higher than that of class A or B). A unique aspect of class D (with infinitely fast switching) is that efficiency is not degraded by the presence of reactance in the load.

Practical class-D PAs suffer from losses due to saturation, switching speed, and drain capacitance. Finite switching speed causes the transistors to be in their active regions while conducting current. Drain capacitances must be charged and discharged once per RF cycle. The associated power loss is proportional to $V_{DD}^3/2$ [21] and increases directly

with frequency.

Class-D PAs with power outputs of 100 W to 1 kW are readily implemented at HF, but are seldom used above lower VHF because of losses associated with the drain capacitance. Recently, however, experimental class-D PAs have been tested with frequencies of operation as high as 1 GHz [22].

Class E

Class E employs a single transistor operated as a switch. The drain-voltage waveform is the result of the sum of the DC and RF currents charging the drain-shunt capacitance. In optimum class E, the drain voltage drops to zero and has zero slope just as the transistor turns on. The result is an ideal efficiency of 100 percent, elimination of the losses associated with charging the drain capacitance in class D, reduction of switching losses, and good tolerance of component variation.

Optimum class-E operation requires a drain shunt susceptance $0.1836/R$ and a drain series reactance $1.15R$ and delivers a power output of $0.577V_{DD}^2/R$ for an ideal PA [23]. The utilization factor is 0.098. Variations in load impedance and shunt susceptance cause the PA to deviate from optimum operation [24, 25], but the degradations in performance are generally no worse than those for class A and B.

The capability for efficient operation in the presence of significant drain capacitance makes class E useful in a number of applications. One example is high-efficiency HF PAs with power levels to 1 kW based upon low-cost MOSFETs intended for switching rather than RF use [26]. Another example is the switching-mode operation at frequencies as high as K band [27]. The class-DE PA [28] similarly uses dead-space between the times when its two transistors are on to allow the load network to charge/discharge the drain capacitances.

Class F

Class F boosts both efficiency and output by using harmonic resonators in the output network to shape the drain waveforms. The voltage waveform includes one or more odd harmonics and approximates a square wave, while the current includes even harmonics and approximates a half sine wave. Alternately ("inverse class F"), the voltage can approximate a half sine wave and the current a square wave. As the number of harmonics increases, the efficiency of an ideal PA increases from the 50 percent (class A) toward unity (class D) and the utilization factor increases from $1/8$ (class A) toward $1/2\pi$ (class D) [29].

The required harmonics can in principle be produced by current-source operation of the transistor. However, in practice the transistor is driven into saturation during part of the RF cycle and the harmonics are produced by a self-regulating mechanism similar to that of saturating class C. Use of a harmonic voltage requires creating a high impedance (3 to 10 times the load impedance) at the drain, while use of a harmonic current requires a low impedance (1/3 to 1/10 of the load impedance). While class F requires a more complex output filter than other PAs, the impedances must be correct at only a few specific frequencies. Lumped-element traps are used at lower frequencies and transmission lines are used at microwave frequencies. Typically, a shorting stub is placed a quarter or half-wavelength away from the drain. Since the stubs for different harmonics interact and the open or short must be created at a "virtual drain" ahead of the drain capacitance and bond-wire inductance, implementation of suitable networks is a bit of an art. Nonetheless, class-F PAs are successfully implemented from MF through Ka band.

A variety of modes of operation in-between class C, E, and F are possi-

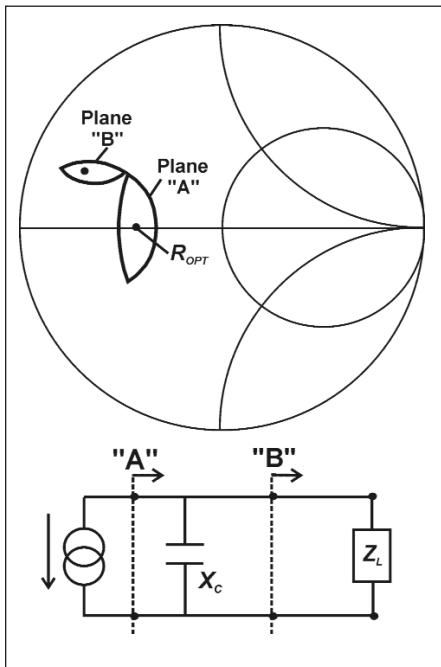


Figure 11 · Constant power contours and transformation.

ble. The maximum achievable efficiency [30] depends upon the number of harmonics, (0.5, 0.707, 0.8165, 0.8656, 0.9045 for 1 through 5 harmonics, respectively). The utilization factor depends upon the harmonic impedances and is highest for ideal class-F operation.

6b. LOAD-PULL CHARACTERIZATION

RF-power transistors are characterized by breakdown voltages and saturated drain currents. The combination of the resultant maximum drain voltage and maximum drain current dictates a range of load impedances into which useful power can be delivered, as well as an impedance for delivery of the maximum power. The load impedance for maximum power results in drain voltage and current excursions from near zero to nearly the maximum rated values.

The load impedances corresponding to delivery of a given amount of RF power with a specified maximum drain voltage lie along parallel-resis-

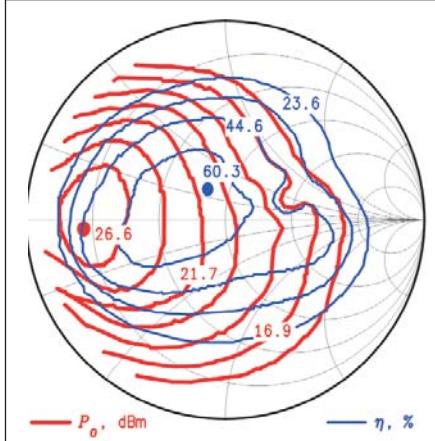


Figure 12 · Example load-pull contours for a 0.5-W, 836 MHz PA. (Courtesy Focus Microwaves and dBm Engineering)

tance lines on the Smith chart. The impedances for a specified maximum current analogously follow a series-resistance line. For an ideal PA, the resultant constant-power contour is football-shaped as shown in Figure 11.

In a real PA, the ideal drain is embedded behind the drain capacitance and bond-wire/package inductance. Transformation of the ideal drain impedance through these elements causes the constant-power contours to become rotated and distorted [31]. With the addition of second-order effects, the contours become elliptical. A set of power contours for a given PA somewhat resembles a set of contours for a conjugate match. However, a true conjugate match produces circular contours. With a power amplifier, the process is more correctly viewed as loading to produce a desired power output. As shown in the example of Figure 12, the power and efficiency contours are not necessarily aligned, nor do maximum power and maximum efficiency necessarily occur for the same load impedance. Sets of such "load-pull" contours are widely used to facilitate design trade-offs.

Load-pull analyses are generally iterative in nature, as changing one

parameter may produce a new set of contours. A variety of different parameters can be plotted during a load-pull analysis, including not only power and efficiency, but also distortion and stability. Harmonic impedances as well as drive impedances are also sometimes varied.

A load-pull system consists essentially of a test fixture, provided with biasing capabilities, and a pair of low-loss, accurately resettable tuners, usually of precision mechanical construction. A load-pull characterization procedure consists essentially of measuring the power of a device, to a given specification (e.g., the 1-dB compression point) as a function of impedance. Data are measured at a large number of impedances and plotted on a Smith chart. Such plots are, of course, critically dependent on the accurate calibration of the tuners, both in terms of impedance and losses. Such calibration is, in turn, highly dependent on the repeatability of the tuners.

Precision mechanical tuners, with micrometer-style adjusters, were the traditional apparatus for load-pull analysis. More recently, a new generation of electronic tuners has emerged that tune through the use varactors or transmission lines switched by pin diodes. Such electronic tuners [32] have the advantage of almost perfect repeatability and high tuning speed, but have much higher losses and require highly complex calibration routines. Mechanical tuners are more difficult to control using a computer, and move very slowly from one impedance setting to another.

In an active load-pull system, a second power source, synchronized in frequency and phase with the device input excitation, is coupled into the output of the device. By controlling the amplitude and phase of the injected signal, a wide range of impedances can be simulated at the output of the test device [33]. Such a

system eliminates the expensive tuners, but creates a substantial calibration challenge of its own. The wide availability of turn-key load-pull systems has generally reduced the application of active load-pull to situations where mechanical or electronic tuning becomes impractical (e.g., millimeter-wave frequencies).

6c. STABILITY

The stability of a small-signal RF amplifier is ensured by deriving a set of S-parameters from using measured data or a linear model, and then establishing the value of the k-factor stability parameter. If the k-factor is greater than unity, at the frequency and bias level in question, then expressions for matching impedances at input and output can be evaluated to give a perfect conjugate match for the device. Amplifier design in this context is mainly a matter of designing matching networks which present the prescribed impedances over the necessary specified bandwidth. If the k factor is less than unity, negative feedback or lossy matching must be employed in order to maintain an unconditionally stable design.

A third case is relevant to PA design at higher microwave frequencies. There are cases where a device has a very high k-factor value, but very low gain in conjugate matched condition. The physical cause of this can be traced to a device which has gain roll-off due to carrier-mobility effects, rather than parasitics. In such cases, introduction of some positive feedback reduces the k-factor and increases the gain in conjugately matched conditions, while maintaining unconditional stability. This technique was much used in the early era of vacuum-tube electronics, especially in IF amplifiers.

6d. MICROWAVE IMPLEMENTATION

At microwave frequencies, lumped elements (capacitors, inductors) become unsuitable as tuning compo-

nents and are used primarily as chokes and by-passes. Matching, tuning, and filtering at microwave frequencies are therefore accomplished with distributed (transmission-line) networks. Proper operation of power amplifiers at microwave frequencies is achieved by providing the required drain-load impedance at the fundamental and a number of harmonic frequencies.

Class F

Class-F operation is specified in terms of harmonic impedances, so it is relatively easy to see how transmission-line networks are used. Methods for using transmission lines in conjunction with lumped-element tuned circuits appear in the original paper by Tyler [34]. In modern microwave implementation, however, it is generally necessary to use transmission lines exclusively. In addition, the required impedances must be produced at a virtual ideal drain that is separated from the output network by drain capacitance, bond-wire/lead inductance.

Typically, a transmission line between the drain and the load provides the fundamental-frequency drain impedance of the desired value. A stub that is a quarter wavelength at the harmonic of interest and open at one end provides a short circuit at the opposite end. The stub is placed along the main transmission line at either a quarter or a half wavelength from the drain to create either an open or a short circuit at the drain [35]. The supply voltage is fed to the drain through a half-wavelength line bypassed on the power-supply end or alternately by a lumped-element choke. When multiple stubs are used, the stub for the highest controlled harmonic is placed nearest the drain. Stubs for lower harmonics are placed progressively further away and their lengths and impedances are adjusted to allow for interactions. Typically, “open” means three to ten times the fundamental-frequency impedance,

and “shorted” means no more 1/10 to 1/3 of the fundamental-frequency impedance [FR17].

A wide variety of class-F PAs have been implemented at UHF and microwave frequencies [36-41]. Generally, only one or two harmonic impedances are controlled. In the X-band PA from [42], for example, the output circuit provides a match at the fundamental and a short circuit at the second harmonic. The third-harmonic impedance is high, but not explicitly adjusted to be open. The 3-dB bandwidth of such an output network is about 20 percent, and the efficiency remains within 10 percent of its maximum value over a bandwidth of approximately 10 to 15 percent.

Dielectric resonators can be used in lieu of lumped-element traps in class-F PAs. Power outputs of 40 W have been obtained at 11 GHz with efficiencies of 77 percent [43].

Class E

The drain-shunt capacitance and series inductive reactance required for optimum class-E operation result in a drain impedance of $R + j0.725R$ at the fundamental frequency, $-j1.7846R$ at the second harmonic, and proportionately smaller capacitive reactances at higher harmonics. At microwave frequencies, class-E operation is approximated by providing the drain with the fundamental-frequency impedance and preferably one or more of the harmonic impedances [44].

An example of a microwave approximation of class E that provides the correct fundamental and second-harmonic impedances [44] is shown in Figure 13. Line l2 is a quarter-wavelength long at the second harmonic so that the open circuit at its end is transformed to a short at plane AA'. Line l1 in combination with L and C is designed to be also a quarter wavelength to translate the short at AA' to an open at the transistor drain. The lines l1 to l4 provide the desired impedance at the funda-

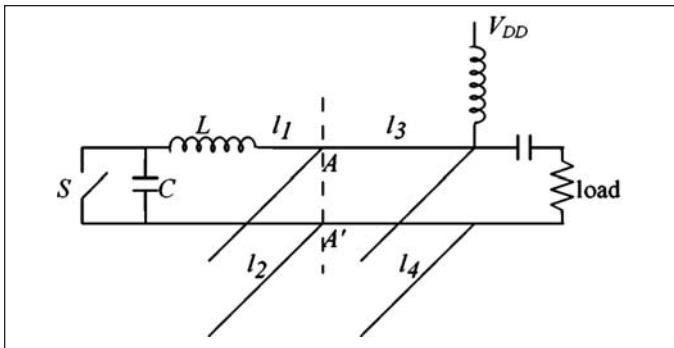


Figure 13 · Idealized microwave class-E PA circuit.



Figure 14 · Example X-band class-E PA.

mental. The implementation using an FLK052 MESFET is shown in Figure 14 produces 0.68 W at X band with a drain efficiency of 72 percent and PAE of 60 percent [42].

Methods exist for providing the proper impedances through the fourth harmonic [45]. However, the harmonic impedances are not critical [30], and many variations are therefore possible. Since the transistor often has little or no gain at the higher harmonic frequencies, those impedances often have little or no effect upon performance. A single-stub match is often sufficient to provide the desired impedance at the fundamental while simultaneously providing an adequately high impedance at the second harmonic, thus eliminating the need for an extra stub and reducing a portion of the losses associated with it. Most microwave class-E amplifiers operate in a suboptimum mode [46]. Demonstrated capabilities range from 16 W with 80-percent efficiency at UHF (LDMOS) to 100 mW with 60-percent efficiency at 10 GHz [47], [48], [44], [49], [50], [51]. Optical sampling of the waveforms [52] has verified that these PAs do indeed operate in class E.

Comparison

PAs configured for classes A (AB), E, and F are compared experimentally in [50] with the following conclusions. Classes AB and F have essentially the same saturated output

power, but class F has about 15 percent higher efficiency. Class E has the highest efficiency. Gain compression occurs at a lower power level for class E than for class F. For a given efficiency, class F produces more power. For the same maximum output power, the third order intermodulation products are about 10 dB lower for class F than for class E. Lower-power PAs implemented with smaller RF power devices tend to be more efficient than PAs implemented with larger devices [42].

Millimeter-Wave PAs

Solid-state PAs for millimeter-wave (mm-W) frequencies (30 to 100 GHz) are predominantly monolithic. Most Ka-band PAs are based upon pHEMT devices, while most W-band PAs are based upon InP HEMTs. Some use is also made of HBTs at the lower mm-W frequencies. Class A is used for maximum gain. Typical performance characteristics include 4 W with 30-percent PAE at Ka band [53], 250 mW with 25-percent PAE at Q band [54], and 200 mW with 10-percent PAE at W band [55]. Devices for operation at mm-W are inherently small, so large power outputs are obtained by combining the outputs of multiple low-power amplifiers in corporate or spatial power combiners.

6e. EXAMPLE APPLICATIONS

The following examples illustrate the wide variety of power amplifiers in use today:

HF/VHF Single Sideband

One of the first applications of RF-power transistors was linear amplification of HF single-sideband signals. Many PAs developed by Helge Granberg have been widely adapted for this purpose [56, 57]. The 300-W PA for 2 to 30 MHz uses a pair of Motorola MRF422 Si NPN transistors in a push-pull configuration. The PA operates in class AB push-pull from a 28-V supply and achieves a collector efficiency of about 45 percent (CW) and a two-tone IMD ratio of about -30 dBc. The 1-kW amplifier is based upon a push-pull pair of MRF154 MOSFETs and operates from a 50-V supply. Over the frequency range of 2 to 50 MHz it achieves a drain efficiency of about 58 percent (CW) with an IMD rating of -30 dBc.

13.56-MHz ISM Power Sources

High-power signals at 13.56 MHz are needed for a wide variety of Industrial, Scientific, and Medical (ISM) applications such as plasma generation, RF heating, and semiconductor processing. A 400-W class-E PA uses an International Rectifier IRFP450LC MOSFET (normally used for low-frequency switching-mode DC power supplies) operates from a 120-V supply and achieves a drain efficiency of 86 percent [58, 26]. Industrial 13.56-MHz RF power generators using class-E output stages have been manufactured since 1992 by Dressler Hochfrequenztechnik (Stolberg, Germany) and Advanced

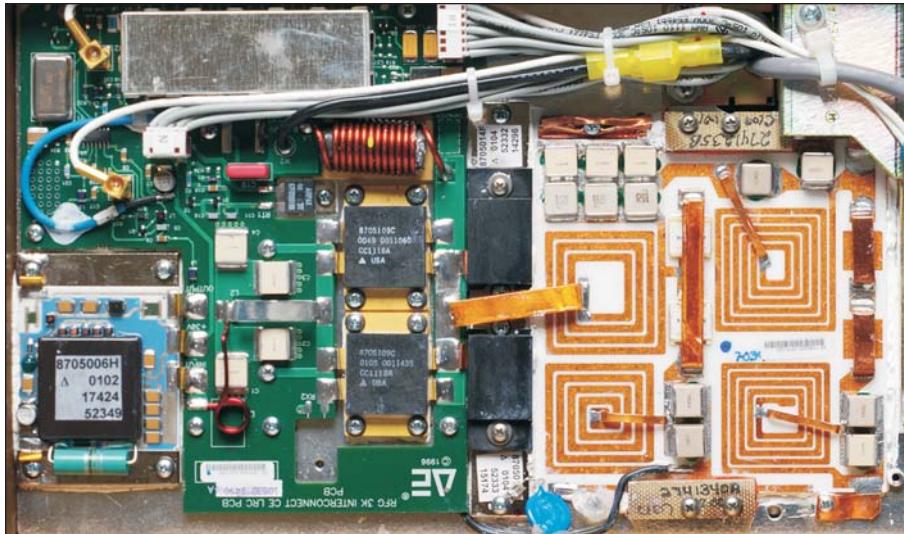


Figure 15 . 3-kW high efficiency PA for 13.56 ISM-band operation.
(Courtesy Advanced Energy)



Figure 16 . Output section of a 50-kW AM broadcast transmitter.
(Courtesy Harris)

Energy Industries (Ft. Collins, CO). They typically use RF-power MOSFETs with 500- to 900-V breakdown voltages made by Directed Energy or Advanced Power Technology and produce output powers of 500 W to with 3 kW with drain efficiencies of about 90 percent. The Advanced Energy Industries amplifier (Figure 15) uses thick-film-hybrid circuits to reduce size. This allows placement inside the clean-room facilities of semiconductor-manufacturing plants, eliminating the need for long runs of coaxial cable from an RF-power generator installed outside the clean-room.

VHF FM Broadcast Transmitter

FM-broadcast transmitters (88 to 108 MHz) with power outputs from 50 W to 10 kW are manufactured by Broadcast Electronics (Quincy, Illinois). These transmitters use up to 32 power-combined PAs based upon Motorola MRF151G MOSFETs. The PAs operate in class C from a 44-V supply and achieve a drain efficiency of 80 percent. Typically, about 6 percent of the output power is dissipated in the power combiners, harmonic-suppression filter, and lightning-protection circuit.

MF AM Broadcast Transmitters

Since the 1980s, AM broadcast transmitters (530 to 1710 kHz) have been made with class-D and -E RF-output stages. Amplitude modulation is produced by varying the supply voltage of the RF PA with a high-efficiency amplitude modulator.

Transmitters made by Harris (Mason, Ohio) produce peak-envelope output powers of 58, 86, 150, 300, and 550 kW (unmodulated carrier powers of 10, 15, 25, 50, and 100 kW). The 100-kW transmitter combines the output power from 1152 transistors. The output stages can use either bipolar or MOSFETs, typically operate in class DE from a 300-V supply, and achieve an efficiency of 98 percent. The output section of the Harris 3DX50 transmitter is shown in Figure 16.

Transmitters made by Broadcast Electronics (Quincy, IL) use class-E RF-output stages based upon APT6015LVR MOSFETs operating from 130-V maximum supply voltages. They achieve drain efficiencies of about 94 percent with peak-envelope output powers from 4.4 to 44 kW. The 44-kW AM-10A transmitter combines outputs from 40 individual output stages.

900-MHz Cellular-Telephone Handset

Most 900-MHz CDMA handsets use power-amplifier modules from vendors such as Conexant and RF Micro Devices. These modules typically contain a single GaAs-HBT RFIC that includes a single-ended class-AB PA. Recently developed PA modules also include a silicon control IC that provides the base-bias reference voltage and can be commanded to adjust the output-transistor base bias to optimize efficiency while maintaining acceptably low amplifier distortion over the full ranges of temperature and output power. A typical module (Figure 17) produces 28 dBm (631 mW) at full output with a PAE of 35 to 50 percent.

Cellular-Telephone Base Station Transmitter

The Spectrian MCPA 3060 cellular base-station transmitter for 1840-1870 MHz CDMA systems provides up to 60-W output while transmitting a signal that may include as many 9 modulated carriers. IMD is minimized by linearizing a class-AB main amplifier with both adaptive predistortion and adaptive feed-forward cancellation. The adaptive control

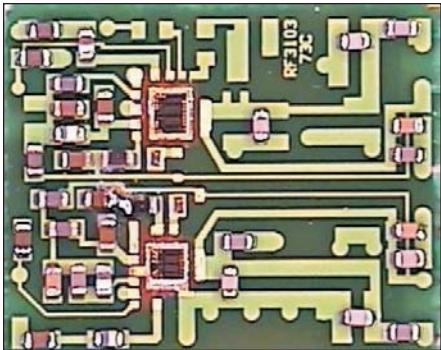


Figure 17 · Internal view of a dual-band (GSM/DCS) PA module for cellular telephone handsets. (Courtesy RF Micro Devices)

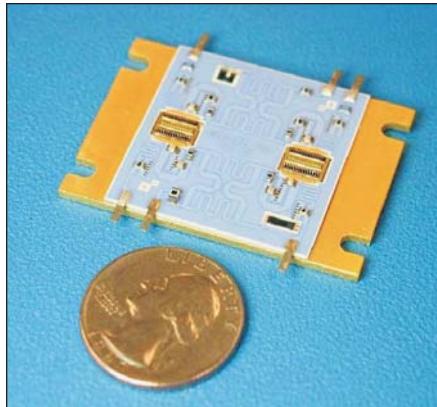


Figure 18 · Thick-film hybrid S-band PA module. (Courtesy UltraRF)

system adjusts operation as needed to compensate for changes due to temperature, time, and output power. The required adjustments are derived from continuous measurements of the system response to a spread-spectrum pilot test signal. The amplifier consumes a maximum of 810 W from a 27-V supply.

S-Band Hybrid Power Module

A thick-film-hybrid power-amplifier module made by UltraRF (now Cree Microwave) for 1805 to 1880 MHz DCS and 1930-1960 MHz PCS is shown in Figure 18. It uses four 140-mm LDMOS FETs operating from a 26-V drain supply. The individual PAs have 11-dB power gain and are quadrature-combined to produce a 100-W PEP output. The average output power is 40 W for EDGE and 7 W for CDMA, with an ACPR of -57 dBc for EDGE and -45 dBc for CDMA. The construction is based upon 0.02-in. thick film with silver metalization.

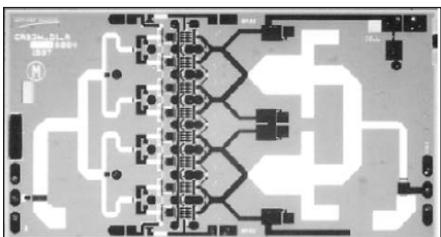


Figure 19 · MMIC PA for X- and K-bands.

GaAs MMIC Power Amplifier

A MMIC PA for use from 8 to 14 GHz is shown in Figure 19. This amplifier is fabricated with GaAs HBTs and intended for use in phased-array radar. It produces a 3-W output with a PAE of approximately 40 percent [59].

References

19. H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: Wiley, 1980.
20. R. Gupta and D. J. Allstot, "Fully monolithic CMOS RF power amplifiers: Recent advances," *IEEE Communications Mag.*, vol. 37, no. 4, pp. 94-98, April 1999.
21. F. H. Raab and D. J. Rupp, "HF power amplifier operates in both class B and class D," *Proc. RF Expo West '93*, San Jose, CA, pp. 114-124, March 17-19, 1993.
22. P. Asbeck, J. Mink, T. Itoh, and G. Haddad, "Device and circuit approaches for next-generation wireless communications," *Microwave J.*, vol. 42, no. 2, pp. 22-42, Feb. 1999.
23. N. O. Sokal and A. D. Sokal, "Class E—a new class of high efficiency tuned single-ended switching power amplifiers," *IEEE J. Solid-State Circuits*, vol. SC-10, no. 3, pp. 168-176, June 1975.
24. F. H. Raab, "Effects of circuit variations on the class E tuned power amplifier," *IEEE J. Solid State Circuits*, vol. SC-13, no. 2, pp. 239-247, April 1978.
25. F. H. Raab, "Effects of VSWR upon the class-E RF-power amplifier," *Proc. RF Expo East '88*, Philadelphia, PA, pp. 299-309, Oct. 25-27, 1988.
26. J. F. Davis and D. B. Rutledge, "A low-cost class-E power amplifier with sine-wave drive," *Int. Microwave Symp. Digest*, vol. 2, pp. 1113-1116, Baltimore, MD, June 7-11, 1998.
27. T. B. Mader and Z. B. Popovic, "The transmission-line high-efficiency class-E amplifier," *IEEE Microwave and Guided Wave Letters*, vol. 5, no. 9, pp. 290-292, Sept. 1995.
28. D. C. Hamill, "Class DE inverters and rectifiers for DC-DC conversion," *PESC96 Record*, vol. 1, pp. 854-860, June 1996.
29. F. H. Raab, "Maximum efficiency and output of class-F power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 6, pp. 1162-1166, June 2001.
30. F. H. Raab, "Class-E, -C, and -F power amplifiers based upon a finite number of harmonics," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 8, pp. 1462-1468, Aug. 2001.
31. S. C. Cripps, *RF Power Amplifiers for Wireless Communication*, Norwood, MA: Artech, 1999.
32. "A load pull system with harmonic tuning," *Microwave J.*, pp. 128-132, March 1986.
33. B. Hughes, A. Ferrero, and A. Cognata, "Accurate on-wafer power and harmonic measurements of microwave amplifiers and devices," *IEEE Int. Microwave Symp. Digest*, Albuquerque, NM, pp. 1019-1022, June 1-5, 1992.
34. V. J. Tyler, "A new high-efficiency

high power amplifier," *The Marconi Review*, vol. 21, no. 130, pp. 96-109, Fall 1958.

35. A. V. Grebennikov, "Circuit design technique for high-efficiency class-F amplifiers," *Int. Microwave Symp. Digest*, vol. 2, pp. 771-774, Boston, MA, June 13-15, 2000.

36. P. Colantonio, F. Giannini, G. Leuzzi, and E. Limiti, "On the class-F power-amplifier design," *RF and Microwave Computer-Aided Engineering*, vol. 32, no. 2, pp. 129-149, March 1999.

37. A. N. Rudiakova and V. G. Krizhanovski, "Driving waveforms for class-F power amplifiers," *Int. Microwave Symp. Digest*, vol. 1, pp. 473-476, Boston, MA, June 13-15, 2000.

38. A. Inoue, T. Heima, A. Ohta, R. Hattori, and Y. Mitsui, "Analysis of class-F and inverse class-F amplifiers," *Int. Microwave Symp. Digest*, vol. 2, pp. 775-778, Boston, MA, June 13-15, 2000.

39. F. van Rijs et al., "Influence of output impedance on power added efficiency of Si-bipolar power transistors," *Int. Microwave Symp. Digest*, vol. 3, pp. 1945-1948, Boston, MA, June 13-15, 2000.

40. F. Huin, C. Duvanaud, V. Serru, F. Robin, and E. Leclerc, "A single supply very high power and efficiency integrated PHEMT amplifier for GSM applications," *Proc. 2000 RFIC Symp.*, Boston, MA, CD-ROM, June 11-13, 2000.

41. B. Ingruber et al., "Rectangularly driven class-A harmonic-control amplifier," *IEEE Trans. Microwave Theory Tech.*, pt. 1, vol. 46, no. 11, pp. 1667-1672, Nov. 1998.

42. E. W. Bryerton, M. D. Weiss, and Z. Popovic, "Efficiency of chip-level versus external power combining," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 8, pp. 1482-1485, Aug. 1999.

43. S. Toyoda, "Push-pull power amplifiers in the X band," *Int. Microwave Symp. Digest*, vol. 3, pp. 1433-1436, Denver, CO, June 8-13, 1997.

44. T. B. Mader and Z. Popovic, "The transmission-line high-efficiency class-E amplifier," *IEEE Micro-wave and Guided Wave Lett.*, vol. 5, no. 9, pp. 290-292, Sept. 1995.

45. A. J. Wilkinson and J. K. A. Everard, "Transmission line load network topology for class E amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 6, pp. 1202-1210, June 2001.

46. F. H. Raab, "Suboptimum operation of class-E power amplifiers," *Proc. RF Technology Expo.*, Santa Clara, CA, pp. 85-98, Feb. 1989.

47. S. Li, "UHF and X-band class-E amplifiers," Ph.D. Thesis, California Institute of Technology, Pasadena, 1999.

48. F. J. Ortega-Gonzalez, J. L. Jimenez-Martin, A. Asensio-Lopez, G. Torregrosa-Penalva, "High-efficiency load-pull harmonic controled class-E power amplifier," *IEEE Microwave Guided Wave Lett.*, vol. 8, no. 10, pp. 348-350, Oct. 1998.

49. E. Bryerton, "High-efficiency switched-mode microwave circuits," Ph.D. dissertation, Univ. of Colorado, Boulder, June 1999.

50. T. B. Mader, E. W. Bryerton, M. Markovic, M. Forman, and Z. Popovic, "Switched-mode high-efficiency microwave power amplifiers in a free-space power-combiner array," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 10, pt. I, pp. 1391-1398, Oct. 1998.

51. M. D. Weiss and Z. Popovic, "A 10 GHz high-efficiency active antenna," *Int. Microwave Symp. Digest*, vol. 2, pp. 663-666, Anaheim, CA, June 14-17, 1999.

52. M. Weiss, M. Crites, E. Bryerton, J. Whitacker, and Z. Popovic, "Time domain optical sampling of nonlinear microwave amplifiers and multipliers," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 12, pp. 2599-2604, Dec. 1999.

53. J. J. Komiak, W. Kong, P. C. Chao, and K. Nichols, "Fully monolithic 4 watt high efficiency Ka-band power amplifier," *Int. Microwave Symp. Digest*, vol. 3, pp. 947-950, Anaheim, CA, June 14-17, 1999.

54. S.-W. Chen et al., "A 60-GHz high-efficiency monolithic power amplifier using 0.1- μ m pHEMTs," *IEEE Microwave and Guided Wave Lett.*, vol. 5, pp. 201-203, June 1995.

55. D. L. Ingram et al., "Compact W-band solid-state MMIC high power

sources," *Int. Microwave Symp. Digest*, vol. 2, pp. 955-958, Boston, MA, June 13-15, 2000.

56. H. Granberg, "Get 300 watts PEP linear across 2 to 30 MHz from this push-pull amplifier," Bulletin EB27A, Motorola Semiconductor Products, Phoenix, Feb. 1980.

57. H. Granberg, "A compact 1-kW 2-50 MHz solid-state linear amplifier," *QEX*, no. 101, pp. 3-8, July 1990. Also AR347, Motorola Semiconductor Products, Feb. Oct. 1990.

58. N. O. Sokal, "Class-E RF power amplifiers ...," *QEX*, No. 204, pp. 9-20, Jan./Feb. 2001.

59. M. Salib, A. Gupta, A. Ezis, M. Lee, and M. Murphy, "A robust 3W high efficiency 8-14 GHz GaAs/AlGaAs heterojunction bipolar transistor power amplifier," *Int. Microwave Symp. Digest*, vol. 2, pp. 581-584, Baltimore, MD, June 7-11, 1998.

Author Information

The authors of this series of articles are: Frederick H. Raab (lead author), Green Mountain Radio Research, e-mail: f.raab@ieee.org; Peter Asbeck, University of California at San Diego; Steve Cripps, Hywave Associates; Peter B. Kenington, Andrew Corporation; Zoya B. Popovic, University of Colorado; Nick Pothecary, Consultant; John F. Sevic, California Eastern Laboratories; and Nathan O. Sokal, Design Automation. Readers desiring more information should contact the lead author.

Notes

1. In Part 1 of this series (May 2003 issue), the references contained in Table 1 were not numbered correctly. The archived version has been corrected and may be downloaded from: www.hfdelectronics.com — click on "Archives," select "May 2003 — Vol. 2 No. 3" then click on the article title.

2. This series has been extended to five parts, to be published in successive issues through January 2004.

RF and Microwave Power Amplifier and Transmitter Technologies — Part 3

By Frederick H. Raab, Peter Asbeck, Steve Cripps, Peter B. Kenington, Zoya B. Popovich, Nick Pothecary, John F. Sevic and Nathan O. Sokal

Transmitter architectures is the subject of this installment of our continuing series on power amplifiers, with an emphasis on designs that can meet today's linearity and high efficiency requirements

The arrangement of building blocks is known as the architecture of a transmitter. The classic transmitter architecture is based upon linear PAs and power combiners. More recently, transmitters are being based upon a variety of different architectures including stage bypassing, Kahn, envelope tracking, outphasing, and Doherty. Many of these are actually fairly old techniques that have been recently made practical by the capabilities of DSP.

7a. LINEAR ARCHITECTURE

The conventional architecture for a linear microwave transmitter consists of a baseband or IF modulator, an up-converter, and a power-amplifier chain (Figure 20). The amplifier chain consists of cascaded gain stages with power gains in the range of 6 to 20 dB. If the transmitter must produce an amplitude-modulated or multi-carrier signal, each stage must have adequate linearity. This generally requires class-A amplifiers with substantial power back-off for all of the driver stages. The final amplifier (output stage) is always the most costly in terms of device size and current consumption, hence it is desirable to operate the output stage in class B. In applications requiring very high linearity, it is necessary to use class A in spite of the lower efficiency.

The building blocks used in transmitters are not only power amplifiers, but a variety of other circuit elements including oscillators, mixers, low-level amplifiers, filters, matching networks, combiners, and circulators. The

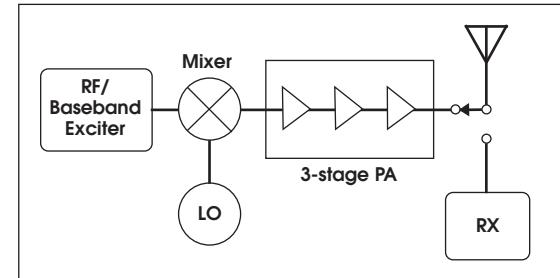


Figure 20 · A conventional transmitter.

The outputs of a driver stage must be matched to the input of the following stage much as the final amplifier is matched to the load. The matching tolerance for maintaining power level can be significantly lower than that for gain [60], hence the 1-dB load-pull contours are more tightly packed for power than for gain.

To obtain even modest bandwidths (e.g., above 5 percent), the use of quadrature balanced stages is advisable (Figure 21). The main benefit of the quadrature balanced configuration is that reflections from the transistors are cancelled by the action of the input and output couplers. An individual device can therefore be deliberately mismatched (e.g., to achieve a power match on the output), yet the quadrature-combined system appears to be well-matched. This configuration also acts as an effective power combiner, so that a given power rating can be achieved using a pair of devices having half of the required power performance. For moderate-bandwidth designs, the lower-power stages are typically designed using a simple single-ended cascade, which in some cases is available as an RFIC. Designs with bandwidths approaching an octave or

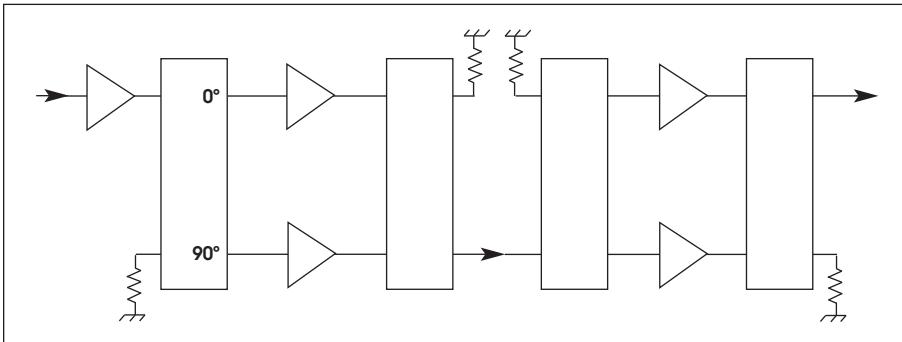


Figure 21 · Amplifier stages with quadrature combiners.

more require the use of quadrature-balanced stages throughout the entire chain.

Simple linear-amplifier chains of this kind have high linearity but only modest efficiency. Single-carrier applications usually operate the final amplifier to about the 1-dB compression point on amplitude modulation peaks. A thus-designed chain in which only the output stage exhibits compression can still deliver an ACPR in the range of about -25 dBc with 50-percent efficiency at PEP.

Two practical problems are frequently encountered in the design of linear PA chains: stability and low gain. Linear, class-A chains are actually more susceptible to oscillation due to their high gain, and single-path chains are especially prone to unstable behavior. Instability can be subdivided into the two distinct categories: Low-frequency oscillation and in-band instability. In-band instability is avoided by designing the individual gain stages to meet the criteria for unconditional stability; i.e., the Rollet k factor [61] must be greater than unity for both in-band and out-of-band frequencies. Meeting this criterion usually requires sacrificing some gain through the use of absorptive elements. Alternatively, the use of quadrature balanced stages provides much greater isolation between individual stages, and the broadband response of the quadrature couplers can eliminate the need to design the transistor

stage itself with $k > 1$. This is another reason for using quadrature coupled stages in the output of the chain.

Large RF power devices typically have very high transconductance, and this can produce low-frequency instability unless great care is taken to terminate both the input and output at low frequencies with impedances for unconditional stability. Because of large separation from the RF band, this is usually a simple matter requiring a few resistors and capacitors.

At X band and higher, the power gain of devices in the 10 W and above category can drop well below 10 dB. To maintain linearity, it may be necessary to use a similarly size device as a driver. Such an architecture clearly has a major negative impact upon the cost and efficiency of the whole chain. In the more extreme cases, it may be advantageous to consider a multi-way power combiner, where 4, 8, or an even greater number of smaller devices are combined. Such an approach also has other advantages, such as soft failure, better thermal management, and phase linearity. However, it typically consumes more board space.

7b. POWER COMBINERS

The need frequently arises to combine the outputs of several individual PAs to achieve the desired transmitter output. Whether to use a number of smaller PAs vs. a single larger PA is one of the most basic decisions in selection of an architec-

ture [60]. Even when larger devices are available, smaller devices often offer higher gain, a lower matching Q factor (wider bandwidth), better phase linearity, and lower cost. Heat dissipation is more readily accomplished with a number of small devices, and a soft-failure mode becomes possible. On the other hand, the increase in parts count, assembly time, and physical size are significant disadvantages to the use of multiple, smaller devices.

Direct connection of multiple PAs is generally impractical as the PAs interact, allowing changes in output from one to cause the load impedance seen by the other to vary. A constant load impedance, hence isolation of one PA from the other, is provided by a hybrid combiner. A hybrid combiner causes the difference between the two PA outputs to be routed to and dissipated in a balancing or "dump" resistor. In the event that one PA fails, the other continues to operate normally, with the transmitter output reduced to one fourth of nominal.

The most common power combiner is the quadrature-hybrid combiner. A 90° phase shift is introduced at input of one PA and also at the output of the other. The benefits of quadrature combining include constant input impedance in spite of variations of input impedances of the individual PAs, cancellation of odd harmonics, and cancellation of backward-IMD (IMD resulting from a signal entering the output port). In addition, the effect of load impedance upon the system output is greatly reduced (e.g., to 1.2 dB for a 3:1 SWR). Maintenance of a nearly constant output occurs because the load impedance presented to one PA decreases when that presented to the other PA increases. As a result, however, device ratings increase and efficiency decreases roughly in proportion to the SWR [65]. Because quadrature combiners are inherently two-terminal devices, they are used in a corporate combining architecture

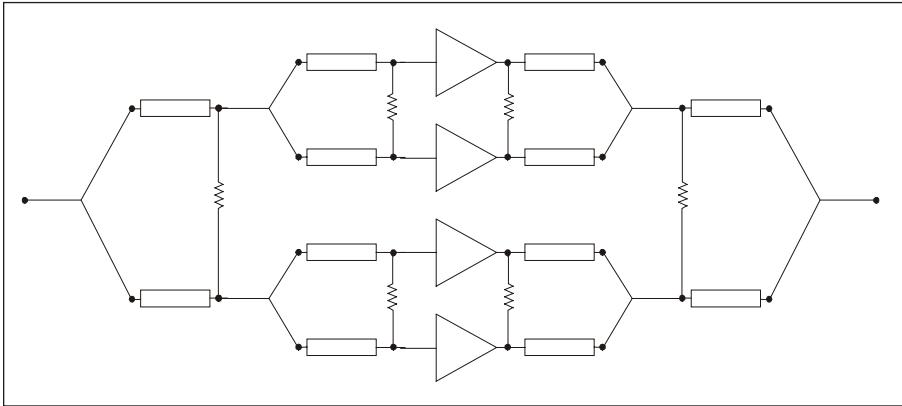


Figure 22 · Multi-section Wilkinson combining architecture.

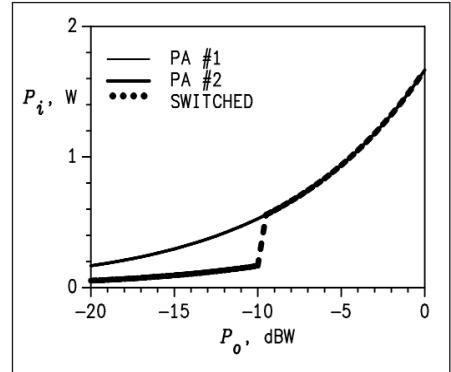


Figure 23 · Power consumption by PAs of different sizes.

(Figure 21). Unfortunately, the physical construction of such couplers poses some problems in a PC-board environment. The very tight coupling between the two quarter-wave transmission lines requires either very fine gaps or a three-dimensional structure. This problem is circumvented by the use of a miniature co-axial cable having a pair of precisely twisted wires to from the coupling section or ready-made, low-cost surface mount 3-dB couplers.

The Wilkinson or in-phase power combiner [62] is often more easily fabricated than a quadrature combiner. In the two-input form (as in each section in Figure 22), the outputs from two quarter-wavelength lines summed into load R_0 produce an apparent load impedance of $2R_0$, which is transformed through the lines into at the load impedances R_{PA} seen by the individual PAs. The difference between the two PA outputs is dissipated in a resistor connected across the two inputs. Proper choice of the balancing resistor ($2R_{PA}$) produces a hybrid combiner with good isolation between the two PAs. The Wilkinson concept can be extended to include more than two inputs [63].

Greater bandwidth can be obtained by increasing the number of transforming sections in each signal path. A single-section combiner can have a useful bandwidth of about 20 percent, whereas a two-section version can have a bandwidth close to an octave. In practice, escalating circuit losses generally preclude the use of more than two sections.

All power-combining techniques all suffer from circuit losses as well as mismatch losses. The losses in a simple two-way combiner are typically about 0.5 dB or 10 percent. For a four-way corporate structure, the interconnects typically result in higher losses. Simple open microstrip lines are too lossy for use in combining structures. One technique that offers a good compromise among cost, producibility, and performance, uses suspended stripline. The conductors are etched onto double-sided PC board, interconnected by vias, and then sus-

pended in a machined cavity. Structures of this kind allow high-power 8-way combiners with octave bandwidths and of 0.5 dB.

A wide variety of other approaches to power-combining circuits are possible [62, 64]. Microwave power can also be combined during radiation from multiple antennas through “quasi-optical” techniques [66].

7c. STAGE SWITCHING AND BYPASSING

The power amplifier in a portable transmitter generally operates well below PEP output, as discussed in Section 4 (Part 1). The size of the transistor, quiescent current, and supply voltage are, however, determined by the peak output of the PA. Consequently, a PA with a lower peak output produces low-amplitude signals more efficiently than does a PA with a larger peak output, as illustrated in Figure 23 for class-B PAs with PEP efficiencies of 60 percent. Stage-bypassing and gate-switching techniques [67, 68] reduce power consumption and increase efficiency by switching between large and small amplifiers according to signal level. This process is analogous to selection of supply voltage in a class-G PA, and the average efficiency can be similarly computed [69].

A typical stage-bypassing architecture is shown in Figure 24. For low-power operation, switches SA and SB route the drive signal around the final amplifier.

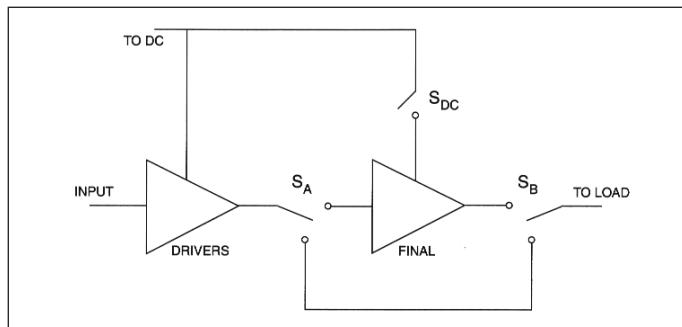


Figure 24 · Stage-bypassing architecture.

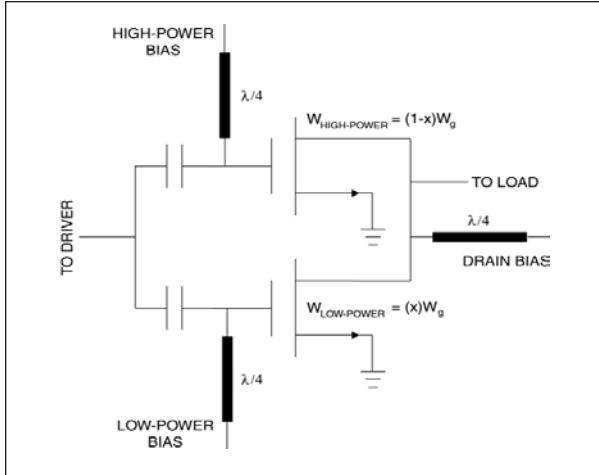


Figure 25 · Adaptive gate switching.

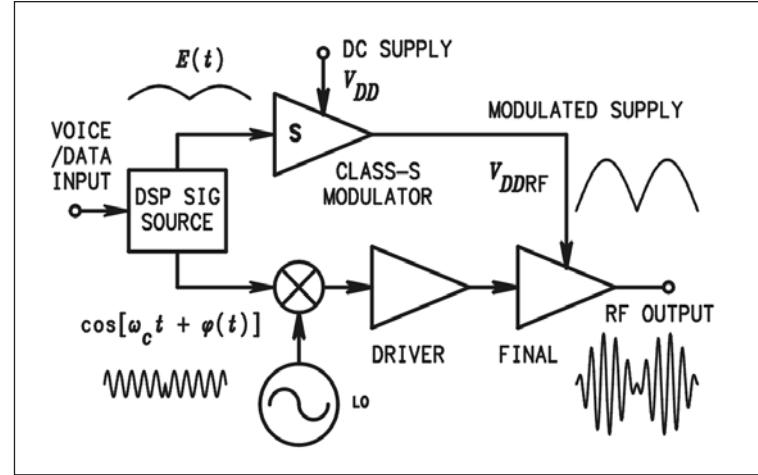


Figure 26 · Kahn-technique transmitter.

Simultaneously, switch SDC turns-off DC power to the final amplifier. The reduction in power consumption can improve the average efficiency significantly (e.g., from 2.1 to 9.5 percent in [70]). The control signal is based upon the signal envelope and power level (back-off). Avoiding hysteresis effects and distortion due to switching transients are critical issues in implementation.

A PA with adaptive gate switching is shown in Figure 25. The gate width (hence current and power capability) of the upper FET is typically ten to twenty times that of the lower FET. The gate bias for the high-power FET keeps it turned off unless it is needed to support a high-power output. Consequently, the quiescent drain current is reduced to low levels unless actually needed. The advantages of this technique are the absence of loss in the switches required by stage bypassing, and operation of the low-power FET in a more linear region (vs. varying the gate bias of a single large FET). The disadvantage is that the source and load impedances change as the upper FET is switched on and off.

7d. KAHN TECHNIQUE

The Kahn Envelope Elimination and Restoration (EER) technique (Figure 26) combines a highly effi-

cient but nonlinear RF power amplifier (PA) with a highly efficient envelope amplifier to implement a high-efficiency linear RF power amplifier. In its classic form [73], a limiter eliminates the envelope, allowing the constant-amplitude phase modulated carrier to be amplified efficiently by class-C, -D, -E, or -F RF PAs. Amplitude modulation of the final RF PA restores the envelope to the phase-modulated carrier creating an amplified replica of the input signal.

EER is based upon the equivalence of any narrowband signal to simultaneous amplitude (envelope) and phase modulations. In a modern implementation, both the envelope and the phase-modulated carrier are generated by a DSP. In contrast to linear amplifiers, a Kahn-technique transmitter operates with high efficiency over a wide dynamic range and therefore produces a high average efficiency for a wide range of signals and power (back-off) levels. Average efficiencies three to five times those of linear amplifiers have been demonstrated (Figure 27) from HF [74] to L band [75].

Transmitters based upon the Kahn technique generally have excellent linearity because linearity depends upon the modulator rather than RF power transistors. The two most important factors affecting the

linearity are the envelope bandwidth and alignment of the envelope and phase modulations. As a rule of thumb, the envelope bandwidth must be at least twice the RF bandwidth and the misalignment must not exceed one tenth of the inverse of the RF bandwidth [76]. In practice, the drive is not hard-limited as in the classical implementation. Drive power is conserved by allowing the drive to follow the envelope except at low levels. The use of a minimum drive level ensures proper operation of the RF PA at low signal levels where the gain is low [77]. At higher microwave frequencies, the RF power devices exhibit softer saturation characteristics and larger amounts of amplitude-to-phase conversion, necessitating the use of predistortion for good linearity [78].

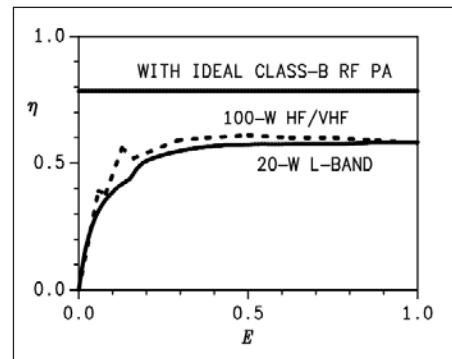


Figure 27 · Efficiency of Kahn-technique transmitters.

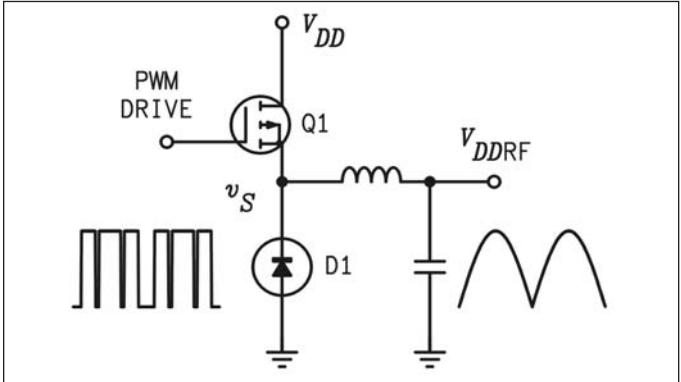


Figure 28 · Class-S modulator.

Class-S Modulator

A class-S modulator (Figure 28) uses a transistor and diode or a pair of transistors act as a two-pole switch to generate a rectangular waveform with a switching frequency several times that of the output signal. The width of pulses is varied in proportion to the instantaneous amplitude of the desired output signal, which is recovered by a low-pass filter. Class S is ideally 100 percent efficient and in practice can have high efficiency over a wide dynamic range. Class-S modulators are typically used as parts of a Kahn-technique transmitter, while class-S amplifiers are becoming popular for the efficient production of audio power in portable equipment. A class-S modulator can be driven by a digital (on/off) signal supplied directly from a DSP, eliminating the need for intermediate conversion to an analog signal.

Selection of the output filter is a compromise between passing the infinite-bandwidth envelope and rejecting FM-like spurious components that are inherent in the PWM process. Typically, the switching frequency must be six to seven times the RF bandwidth. Modulators with switching frequencies of 500 kHz are readily implemented using discrete MOSFETs and off-the-shelf ICs [74], while several MHz can be achieved using MOS ASICs or discrete GaAs devices [75].

Class-G Modulator

A class-G modulator (Figure 29) is a combination of linear series-pass (class-B) amplifiers that operate from different supply voltages. Power is conserved by selecting the one with the lowest useable supply voltage [69] so that the voltage drop across the active device is minimized.

Split-Band Modulator

Most of the power in the envelope resides at lower frequencies; typically 80 percent is in the DC component. The bandwidth of a class-S modulator can therefore be extended by combining it with a linear amplifier. While there are a number of approaches, the highest efficiency

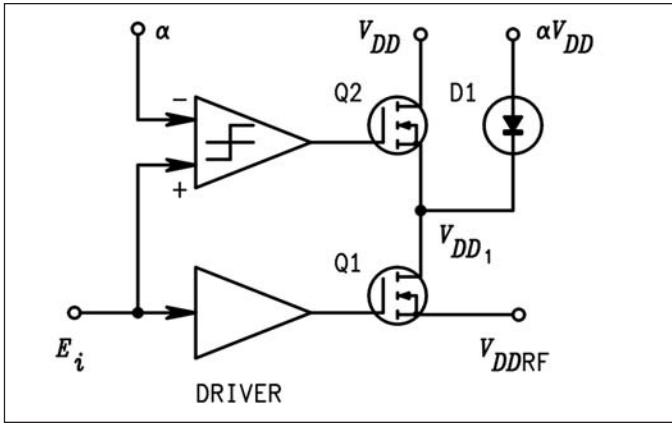


Figure 29 · Class-G modulator.

(typically 90 percent) is achieved by a diplexing combiner. Obtaining a flat frequency response and resistive loads for the two PAs is achieved by splitting the input signals in a DSP that acts as a pair of negative-component filters (Figure 30) [79]. The split-band modulator should make possible Kahn-technique transmitters with RF bandwidths of tens or even hundreds of MHz.

7e. ENVELOPE TRACKING

The envelope-tracking architecture (Figure 31) is similar to that of the Kahn technique. However, the final amplifier operates in a linear mode and the supply voltage is varied dynamically to conserve power [81, 82]. The RF drive contains both amplitude and phase information, and the burden of providing linear amplification lies entirely on the final RF PA. The role of the variable power supply is only to optimize efficiency.

Typically, the envelope is detected and used to control a DC-DC converter. While both buck (step-down) or boost (step-up) converters are used, the latter is more common as it allows operation of the RF PA from a supply voltage higher than the DC-supply voltage. This configuration is

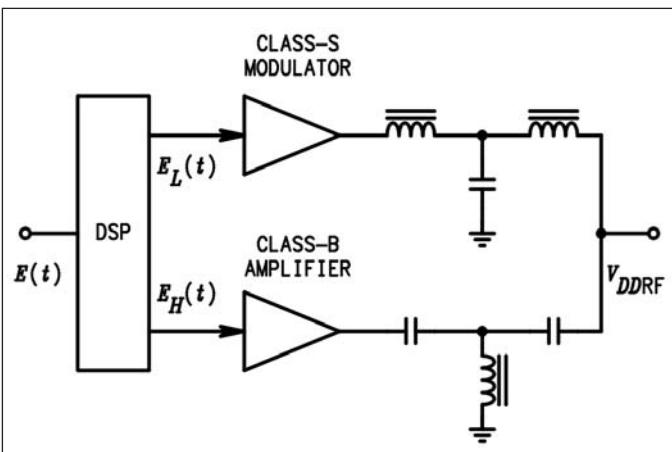


Figure 30 · Split-band modulator.

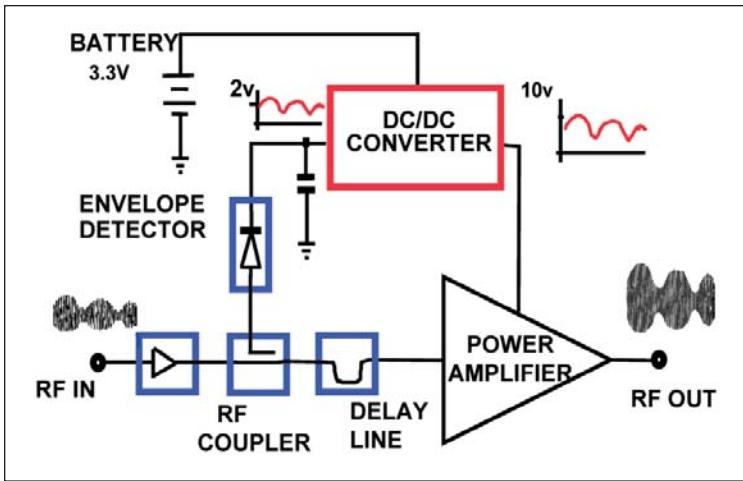


Figure 31 · Envelope-tracking architecture.

also more amenable to the use of *n*p_n or *n*-channel transistors for fast switching. The result is a minimum V_{DDRF} corresponding to the DC-supply voltage and tracking of larger envelopes with a fixed “headroom” to ensure linear operation of the RF PA. If the RF PA is operated in class A, its quiescent current can also be varied.

In general, excess power-supply voltage translates to reduced efficiency, rather than output distortion. In principle, perfect tracking of the envelope by the supply voltage preserves the peak efficiency of the RF PA for all output amplitudes, as in the Kahn technique. In practice, efficiency improvement is obtained over a limited range of output power.

A high switching frequency in the DC-DC converter allows both a high modulation bandwidth and the use of smaller inductors and capacitors. The switching devices in the converter can in fact be implemented using the same transistor technology used in the RF PA. Converters with switching frequencies of 10 to 20 MHz have recently been implemented using MOS ASICs [80], GaAs HBTs [83, 84] and RF power MOSFETs [85].

Representative results for an envelope-tracking transmitter based on a GaAs FET power amplifier are shown in Figure 32. The efficiency is lower at high power than that of the conventional amplifier with constant supply voltage due to the inefficiency of the DC-DC converter. However, the efficiency is much higher over a wide range of output power, with the average efficiency approximately 40 percent higher than that of the conventional linear amplifier.

Spurious outputs can be produced by supply-voltage ripple at the switching frequency. The effects of the ripple can be minimized by making the switching frequency sufficiently high or by using an appropriate filter. Variation of the RF PA gain with supply voltage can introduce distortion. Such distortion can, however, be countered by pre-distortion techniques [to be covered in Section 8 (Part 4)].

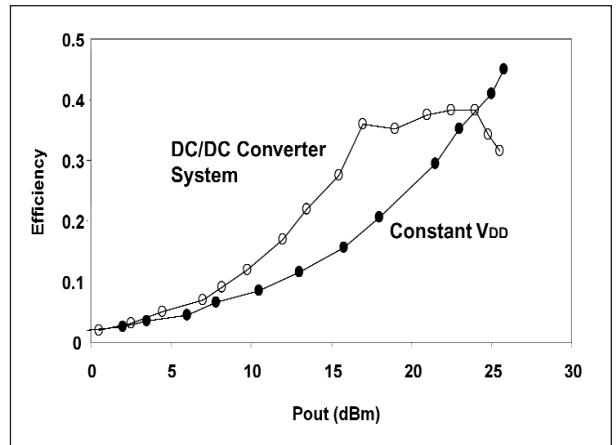


Figure 32 · Efficiency of a GaAs FET envelope-tracking transmitter.

7f. OUTPHASING

Outphasing was invented during the 1930s as a means of obtaining high-quality AM from vacuum tubes with poor linearity [86] and was used through about 1970 in RCA “Ampliphase” AM-broadcast transmitters. In the 1970s, it came into use at microwave frequencies under the name “LINC” (Linear Amplification using Nonlinear Components) [87].

An outphasing transmitter (Figure 33) produces an amplitude-modulated signal by combining the outputs of two PAs driven with signals of different time-varying phases. Basically, the phase modulation causes the instantaneous vector sum of the two PA outputs to follow the desired signal amplitude (Figure 34). The inverse sine of envelope E phase-modulates the driving signals for the two PAs to produce a transmitter output that is proportional to E . In a modern implementation, a DSP and synthesizer produce the inverse-sine modulations of the driving signals.

Hybrid combining (Figure 33) isolates the PAs from the reactive loads inherent in outphasing, allowing them to see resistive loads at all signal levels. However, both PAs deliver full power all of the time. Consequently, the efficiency of a hybrid-coupled outphasing transmitter varies with the output power (Figure 35), resulting in an average efficiency that is inversely proportional to peak-to-average ratio (as for class A). Recovery of the power from the dump port of the hybrid combiner offers some improvement in the efficiency [88].

The phase of the output current is that of the vector sum of the two PA-output voltages. Direct summation of the out-of-phase signals in a nonhybrid combiner inherently results in reactive load impedances for the power amplifiers [89]. If the reactances are not partially cancelled as in the Chireix technique, the current drawn from the PAs is proportional to the transmitter-output voltage.

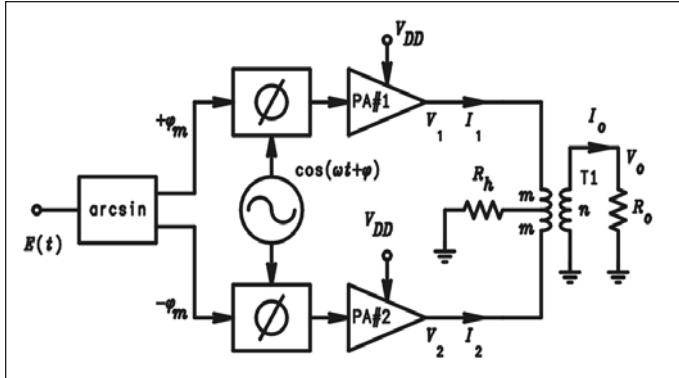


Figure 33 · Hybrid-combined outphasing transmitter.

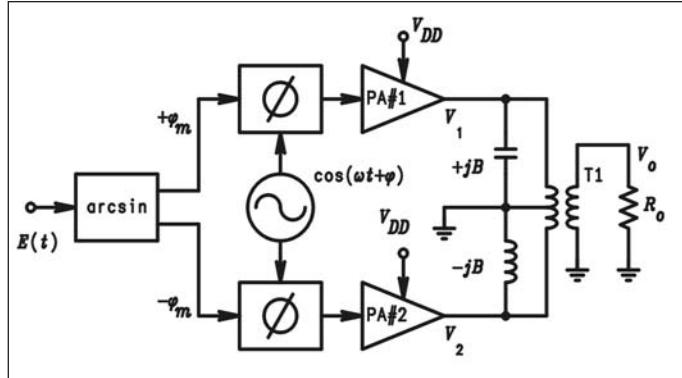


Figure 36 · Chireix-outphasing transmitter.

This results in an efficiency characteristic similar to that of a class-B PA.

The Chireix technique [86] uses shunt reactances on the inputs to the combiner (Figure 36) to tune-out the drain reactances at a particular amplitude, which in turn maximizes the efficiency in the vicinity of that amplitude. The efficiency at high and low amplitudes may be degraded. In the classic Chireix implementation, the shunt reactances maximize the efficiency at the level of the unmodulated carrier in an AM signal and produce good efficiency over the upper 6 dB of the output range. With judicious choice of the shunt susceptances, the average efficiency can be maximized for any given signal [89, 90]. For example, a normalized susceptance of 0.11 peaks the instantaneous efficiency at a somewhat lower amplitude, resulting in an average efficiency of 52.1 percent for an ideal class-B PA and a 10-dB Rayleigh-envelope signal (vs. 28 percent for lin-

ear amplification).

Virtually all microwave outphasing systems in use today are of the hybrid-coupled type. Use of the Chireix technique at microwave frequencies is difficult because microwave PAs do not behave as ideal voltage sources. Simulations suggest that direct (nonhybrid) combining increases both efficiency and distortion [91]. Since outphasing offers a wide bandwidth and the distortion can be mitigated by techniques such as predistortion, directly coupled and Chireix techniques should be fruitful areas for future investigation.

length lines or networks. The “carrier” (main) PA is biased in class B while the “peaking” (auxiliary) PA is biased in class C. Only the carrier PA is active when the signal amplitude is half or less of the PEP amplitude. Both PAs contribute output power when the signal amplitude is larger than half of the PEP amplitude

Operation of the Doherty system can be understood by dividing it into low-power, medium-power (load-modulation), and peak-power regions [96]. The current and voltage relationships are shown in Figure 38 for ideal transistors and lossless matching networks. In the low-power region, the instantaneous amplitude of the input signal is insufficient to overcome the class-C (negative) bias of the peaking PA, thus the peaking PA remains cut-off and appears as an open-circuit. With the example load impedances shown in Figure 37, the carrier PA sees a 100 ohm load and operates as an ordinary class-B amplifier. The drain voltage increases linearly with output until reaching supply voltage V_{DD} . The instantaneous efficiency at this point (-6 dB from PEP) is therefore the 78.5 percent of the ideal class-B PA.

As the signal amplitude increases into the medium-power region, the carrier PA saturates and the peaking PA becomes active. The additional current I_2 sent to the load by the peaking PA causes the apparent load impedance at V_L to increase above

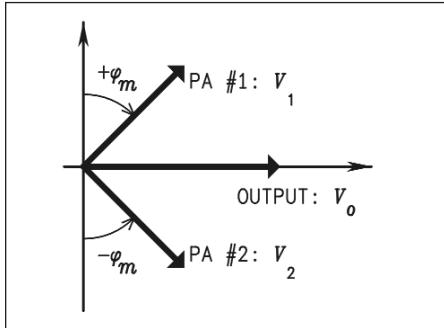


Figure 34 · Signal vectors in outphasing.

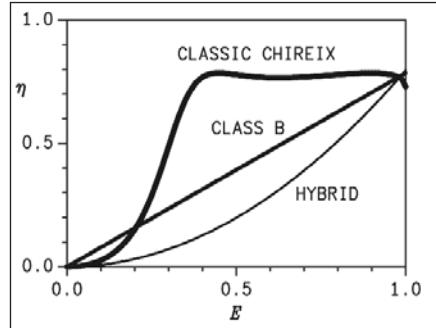


Figure 35 · Efficiency of outphasing transmitters with ideal class-B PAs.

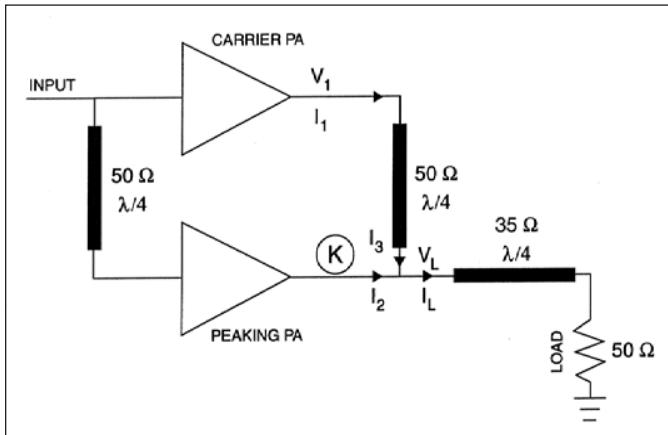


Figure 37 · Doherty transmitter.

the 25 ohms of the low-power region. Transformation through the quarter-wavelength line results in a decrease in the load presented to the carrier PA. The carrier PA remains in saturation and acts as a voltage source. It

39. The classical power division ($\alpha = 0.5$) approximately maximizes the average efficiency for full-carrier AM signals, as well as modern single-carrier digital signals. The use of other power-division ratios allows the lower

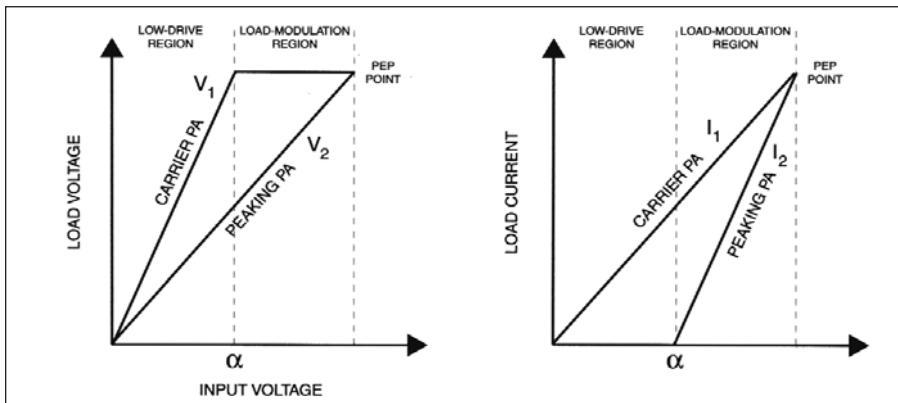


Figure 38 · Ideal voltage and current relationships in Doherty transmitter.

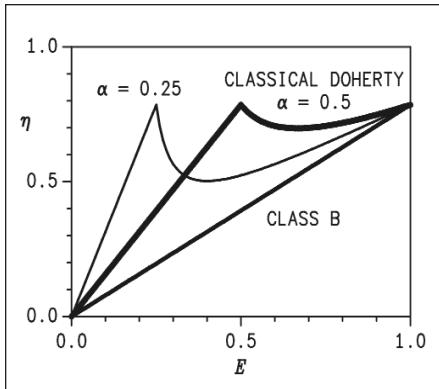


Figure 39 Instantaneous efficiency of the Doherty system with ideal class-B PAs.

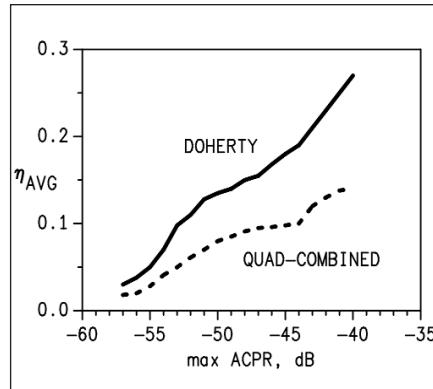


Figure 40 · Measured ACPR performance of an S-band Doherty transmitter.

operates at peak efficiency, but delivers an increasing amount of power. At PEP output, both PAs see 50-ohm loads and each delivers half of the system output power. The PEP efficiency is that of the class-B PAs.

The resulting instantaneous-efficiency curve is shown in Figure

efficiency peak to be shifted leftward so that the average efficiency is increased for signals with higher peak-to-average ratios. For example, $\alpha = 0.36$ results in a 60 percent average efficiency for a Rayleigh-envelope signal with a 10-dB peak-to-average ratio, which is a factor of 2.1 improvement over class B. Doherty transmitters with unequal power division can be implemented by using different PEP load impedances and different supply voltages in the two PAs [97].

Much recent effort has focused on accommodating non-ideal effects (e.g., nonlinearity, loss, phase shift) into a Doherty architecture [93, 94, 95]. The power consumed by the quiescent current of the peaking amplifier is also a concern. The measured ACPR characteristics of an S-band Doherty transmitter are compared to those of quadrature-combined class-B PAs in Figure 40. The signal is IS-95 forward link with pilot channel, paging channel, and sync-channel. The PAs are based upon 50-W LDMOS transistors. Back-off is varied to trade-off linearity against output. For the specified ACPR of -45 dBc, the average PAE is nearly twice that of the quadrature-combined PAs.

In a modern implementation, DSP can be used to control the drive and bias to the two PAs, for precise control and higher linearity. It is also possible to use three or more stages to keep the instantaneous efficiency relatively high over a larger dynamic range [96, 98]. For ideal class-B PAs, the average efficiency of a three-stage Doherty can be as high as 70 percent for a Rayleigh-envelope signal with 10-dB peak-to-average ratio.

References

60. S. C. Cripps, *RF Power Amplifiers for Wireless Communication*, Norwood, MA: Artech, 1999.
61. J. M. Rollett, "Stability and power-gain invariants of linear twoports," *IRE Trans. Circuit Theory*, pp. 29-32, March 1962.
62. S. Cohn, "A class of broadband three-port TEM modes hybrids," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-

16, no. 2, pp. 110-116, 1968.

63. J. Goel, "K band GaAsFET amplifier with 8.2 W output using a radial power combiner," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, no. 3, pp. 317-324, 1984.

64. A. Berr and D. Kaminsky, "The travelling wave power divider/combiner," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28, no. 12, pp. 1468-1473, 1980.

65. F. H. Raab, "Hybrid and quadrature splitters and combiners," Research Note RN97-22, Green Mountain Radio Research Company, Colchester, VT, Sept. 20, 1997.

66. R. A. York and Z. B. Popovic, *Active and Quasi-Optical Arrays for Solid-State Power Combining*, New York: Wiley, 1997.

67. S. Brozovich, "High efficiency multiple power level amplifier circuit," U.S. Patent 5,661,434, Aug. 1997.

68. J. Sevic, "Efficient parallel stage amplifier," U.S. Patent 5,872,481, Feb. 1999.

69. F. H. Raab, "Average efficiency of Class-G power amplifiers," *IEEE Trans. Consumer Electronics*, vol. CE-32, no. 2, pp. 145-150, May 1986.

70. J. Staudinger, "Applying switched gain stage concepts to improve efficiency and linearity for mobile CDMA power amplification," *Microwave J.*, vol. 43, no. 9, pp. 152-162, Sept. 2000.

71. "30 way radial combiner for miniature GaAsFET power applications"

72. T. C. Choinski, "Unequal power division using several couplers to split and recombine the output," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, no. 6, pp. 613-620, 1984.

73. L. R. Kahn, "Single sideband transmission by envelope elimination and restoration," *Proc. IRE*, vol. 40, no. 7, pp. 803-806, July 1952.

74. F. H. Raab and D. J. Rupp, "High-efficiency single-sideband HF/VHF transmitter based upon envelope elimination and restoration," *Proc. Sixth Int. Conf. HF Radio Systems and Techniques (HF '94)* (IEE CP 392), York, UK, pp. 21-25, July 4 - 7, 1994.

75. F. H. Raab, B. E. Sigmon, R. G. Myers, and R. M. Jackson, "L-band transmitter using Kahn EER technique," *IEEE Trans. Microwave Theory Tech.*, pt. 2, vol. 46, no. 12, pp. 2220-2225, Dec. 1998.

76. F. H. Raab, "Intermodulation distortion in Kahn-technique transmitters," *IEEE Trans. Microwave Theory Tech.*, vol. 44, no. 12, part 1, pp. 2273-2278, Dec. 1996.

77. F. H. Raab, "Drive modulation in Kahn-technique transmitters," *Int. Microwave Symp. Digest*, vol. 2, pp. 811-

Acronyms Used in Part 3

EER	Envelope Elimination and Restoration
AM	Amplitude Modulation
LINC	Linear Amplification with Nonlinear Components

814, Anaheim, CA, June 1999.

78. M. D. Weiss, F. H. Raab, and Z. B. Popovic, "Linearity characteristics of X-band power amplifiers in high-efficiency transmitters," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 6, pp. 1174-1179, June 2001.

79. F. H. Raab, "Technique for wide-band operation of power amplifiers and combiners," U. S. Patent 6,252,461, June 26, 2001.

80. J. Staudinger et al., "High efficiency CDMA RF power amplifier using dynamic envelope tracking technique," *Int. Microwave Symp. Digest*, vol. 2, pp. 873-876, Boston, MA, June 13-15, 2000.

81. A. A. M. Saleh and D. C. Cox, "Improving the power-added efficiency of FET amplifiers operating with varying-envelope signals," *IEEE Trans. Microwave Theory Tech.*, vol. 31, no. 1, pp. 51-55, Jan. 1983.

82. B. D. Geller, F. T. Assal, R. K. Gupta, and P. K. Cline, "A technique for the maintenance of FET power amplifier efficiency under backoff," *IEEE 1989 MTT-S Digest*, Long Beach, CA, pp. 949-952, June 1989.

83. G. Hanington, P.-F. Chen, P. M. Asbeck, and L. E. Larson, "High-efficiency power amplifier using dynamic power-supply voltage for CDMA applications," *IEEE Trans. Micro-wave Theory Tech.*, vol. 47, no. 8, pp. 1471-1476, Aug. 1999.

84. G. Hanington, A. Metzger, P. Asbeck, and H. Finlay, "Integrated dc-dc converter using GaAs HBT technology," *Electronics Letters*, vol. 35, no. 24, p.2110-2112, 1999.

85. D. R. Anderson and W. H. Cantrell, "High-efficiency inductor-coupled high-level modulator," *IMS '01 Digest*, Phoenix, AZ, May 2001.

86. H. Chireix, "High power outphasing modulation," *Proc. IRE*, vol. 23, no. 11, pp. 1370-1392, Nov. 1935.

87. D. C. Cox and R. P. Leck, "A VHF implementation of a LINC amplifier," *IEEE Trans. Commun.*, vol. COM-24, no. 9, pp. 1018-1022, Sept. 1976.

88. R. Langridge, T. Thornton, P. M. Asbeck, and L. E. Larson, "A power re-use technique for improving efficiency of out-phasing microwave power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 8, pp. 1467-1470 , Aug. 1999.

89. F. H. Raab, "Efficiency of outphasing power-amplifier systems," *IEEE Trans. Commun.*, vol. COM-33, no. 10, pp. 1094-1099, Oct. 1985.

90. B. Stengel and W. R. Eisenstat, "LINC power amplifier combiner efficiency optimization," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 229- 234, Jan. 2000.

91. C. P. Conradi, R. H. Johnston, and J. G. McRory, "Evaluation of a lossless combiner in a LINC transmitter," *Proc. 1999 IEEE Canadian Conf. Elec. and Comp. Engr.*, Edmonton, Alberta, Canada, pp. 105-109, May 1999.

92. W. H. Doherty, "A new high efficiency power amplifier for modulated waves," *Proc. IRE*, vol. 24, no. 9, pp. 1163-1182, Sept. 1936.

93. D. M. Upton and P. R. Maloney, "A new circuit topology to realize high efficiency, high linearity, and high power microwave amplifiers," *Proc. Radio and Wireless Conf. (RAWCON)*, Colorado Springs, pp. 317-320, Aug. 9-12, 1998.

94. J. Schuss et al, "Linear amplifier for high efficiency multi-carrier performance," U.S. Patent 5,568,086, Oct. 1996.

95. J. Long, "Apparatus and method for amplifying a signal," U.S. Patent 5,886,575, March 1999.

96. F. H. Raab, "Efficiency of Doherty RF-power amplifier systems," *IEEE Trans. Broadcasting*, vol. BC-33, no. 3, pp. 77-83, Sept. 1987.

97. M. Iwamoto et al., "An extended Doherty amplifier with high efficiency over a wide power range," *Int. Microwave Symp. Digest*, Phoenix, AZ, pp. 931-934, May 2001.

98. B. E. Sigmon, "Multistage high efficiency amplifier," U.S. Patent 5,786,938, July 28, 1998.

Author Information

The authors of this series of articles are: Frederick H. Raab (lead author), Green Mountain Radio Research, e-mail: f.raab@ieee.org; Peter Asbeck, University of California at San Diego; Steve Cripps, Hywave Associates; Peter B. Kenington, Andrew Corporation; Zoya B. Popovic, University of Colorado; Nick Pothecary, Consultant; John F. Sevic, California Eastern Laboratories; and Nathan O. Sokal, Design Automation. Readers desiring more information should contact the lead author.

RF and Microwave Power Amplifier and Transmitter Technologies — Part 4

By Frederick H. Raab, Peter Asbeck, Steve Cripps, Peter B. Kenington, Zoya B. Popovich, Nick Pothecary, John F. Sevic and Nathan O. Sokal

Linearization methods are the focus of Part 4 of our series on power amplifiers, which describes the basic architecture and performance capabilities of feedback, feedforward and predistortion techniques

vides a summary of the three main families of techniques have been developed: Feedback, feedforward, and predistortion.

8a. FEEDBACK

Feedback linearization can be applied either directly around the RF amplifier (RF feedback) or indirectly upon the modulation (envelope, phase, or I and Q components).

RF Feedback

The basis of this technique is similar to its audio-frequency counterpart. A portion of the RF-output signal from the amplifier is fed back to, and subtracted from, the RF-input signal without detection or down-conversion. Considerable care must be taken when using feedback at RF as the delays involved must be small to ensure stability. In addition, the loss of gain at RF is generally a more significant sacrifice than it is at audio frequencies. For these reasons, the use of RF feedback in discrete circuits is usually restricted to HF and lower VHF frequencies [99]. It can be applied within MMIC devices, however, well into the microwave region.

In an active RF feedback system, the voltage divider of a conventional passive-feedback system is replaced by an active (amplifier)

linearization techniques are incorporated into power amplifiers and transmitters for the dual purposes of improving linearity and for allowing operation with less back-off and therefore higher efficiency. This article pro-

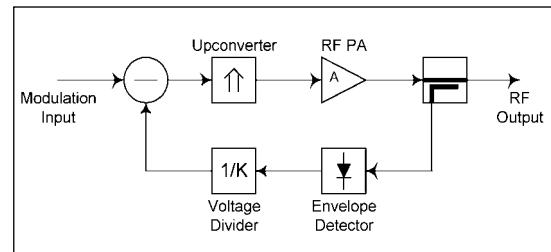


Fig 41 · Envelope feedback applied to a complete transmitter.

stage. The gain in the feedback path reduces the power dissipated in the feedback components. While such systems demonstrate IMD reduction [105], they tend to work best at a specific signal level.

Envelope Feedback

The problem of delay in RF feedback is alleviated to a large extent by utilizing the signal envelope as the feedback parameter. This approach takes care of in-band distortion products associated with amplitude nonlinearity. Harmonic distortion products, which are corrected by RF feedback, are generally not an issue as they can easily be removed by filtering in most applications. Envelope feedback is therefore a popular and simple technique.

Envelope feedback can be applied to either a complete transmitter (Figure 41) or a single power amplifier (Figure 42). The principles of operation are similar and both are described in detail in [100]. The RF input signal is sampled by a coupler and the envelope of the input sample is detected. The resulting envelope is then fed to one input of a differential amplifier, which subtracts it from a similarly

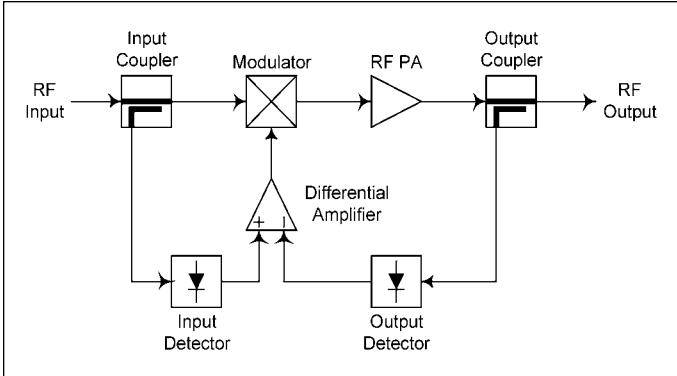


Figure 42 · Envelope feedback applied to an RF power amplifier.

obtained sample of the RF output. The difference signal, representing the error between the input and output envelopes, is used to drive a modulator in the main RF path. This modulator modifies the envelope of the RF signal which drives the RF PA. The envelope of the resulting output signal is therefore linearized to a degree determined by the loop gain of the feedback process. Examples of this type of system are reported in [101] and [102].

The degree of linearity improvement that can be obtained when using this technique depends upon the relative levels of the AM-AM and AM-PM conversion in the amplifier. For a VHF BJT amplifier, AM-AM distortion is dominant and two-tone IMD is typically reduced by 10 dB. Since AM-PM distortion is not corrected by envelope feedback, no linearity improvement is observed if phase distortion is the dominant form of nonlinearity. This is often the case in, for example, class-C and LDMOS PAs. The use of envelope feedback is therefore generally restricted to relatively linear class-A or AB amplifiers.

Polar-Loop Feedback

The polar-loop technique overcomes the fundamental inability of envelope feedback to correct for AM-PM distortion effects [103]. Essentially, a phase-locked loop is added to the envelope feedback system as shown in Figure 43. For a narrowband VHF PA, the improvement in two-tone IMD is typically around 30 dB.

The envelope- and phase-feedback functions operate essentially independently. In this case, envelope detection occurs at the intermediate frequency (IF), as the input signal is assumed to be a modulated carrier at IF. Likewise, phase detection takes place at the IF, with limiting being used to minimize the effects of signal amplitude upon the detected phase. Alternatively, it is possible to supply the envelope and phase modulating signals separately at baseband and to undertake the comparisons there.

The key disadvantage of polar feedback lies in the gen-

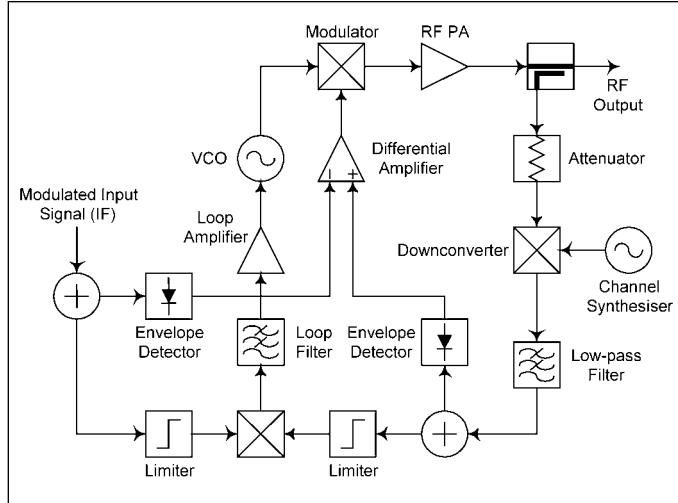


Figure 43 · Block diagram of a polar-loop transmitter.

erally different bandwidths required for the amplitude and phase feedback paths. Thus, differing levels of improvement of the AM-AM and AM-PM characteristics usually result, and this often leads to a poorer overall performance than that achievable from an equivalent Cartesian-loop transmitter. A good example of the difference occurs with a standard two-tone test, which causes the phase-feedback path to cope with a discontinuity at the envelope minima. In general, the phase bandwidth must be five to ten times the envelope bandwidth, which limits available loop gain for a given delay.

Cartesian Feedback

The Cartesian-feedback technique overcomes the problems associated with the wide bandwidth of the signal phase by applying modulation feedback in I and Q (Cartesian) components [104]. Since the I and Q components are the natural outputs of a modern DSP, the Cartesian loop is widely used in PMR and SMR systems.

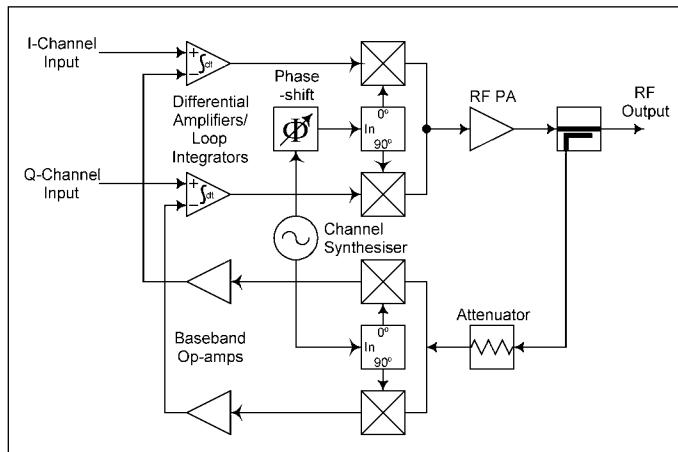


Figure 44 · Cartesian-loop transmitter configuration.

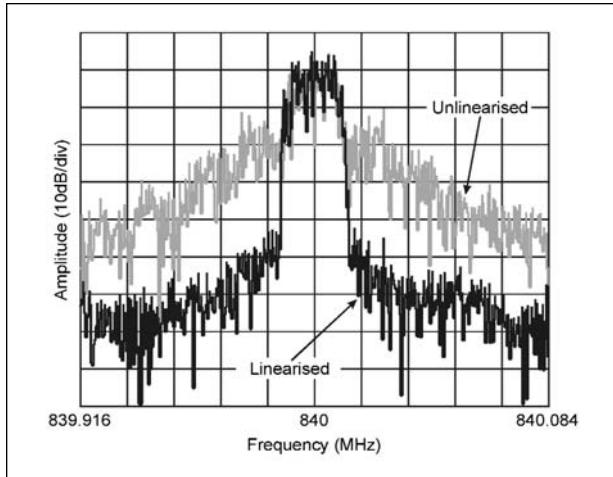


Figure 45 · Linearization of a class-C PA by Cartesian feedback (courtesy WSI).

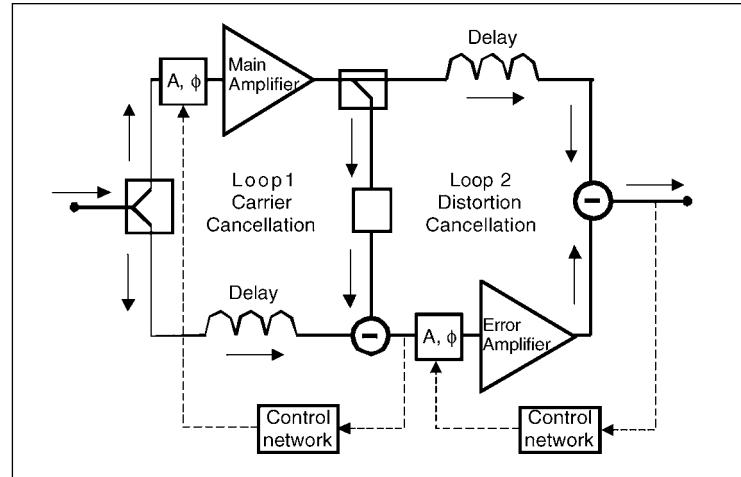


Figure 46 · Block diagram of a feed-forward transmitter in its basic form.

The basic Cartesian loop (Figure 44) consists of two identical feedback processes operating independently on the I and Q channels. The inputs are applied to differential integrators (in the case of a first-order loop) with the resulting difference (error) signals being modulated onto I and Q subcarriers and up-converted to drive the PA. A sample of the output from the PA is attenuated and quadrature-down-converted (synchronously with the up-conversion process). The resulting quadrature feedback signals then form the second inputs to the input differential integrators, completing the two feedback loops. The phase shifter shown in the up-converter local-oscillator path is used to align the phases of the up- and down-conversion processes, thereby ensuring that a negative feedback system is created and that the phase margin of the system is optimized.

The effects of applying Cartesian feedback to a highly nonlinear (class-C) PA amplifying an IS-136 (DAMPS) signal are shown in Figure 45. The first ACPR is improved by 35 dB and the signal is produced within specifications with an efficiency of 60 percent [100].

8b. FEEDFORWARD

The very wide bandwidths (10 to 100 MHz) required in multicarrier applications can render feedback and DSP impractical. In such cases, the feedforward technique can be used to achieve ultra-linear operation. In its basic configuration, feedforward typically gives improvements in distortion ranging from 20 to 40 dB.

Operation

In its basic form (Figure 46), a feedforward amplifier consists of two amplifiers (the main and error amplifiers), directional couplers, delay lines and loop control networks [110]. The directional couplers are used for power split-

ting/combining, and the delay lines ensure operation over a wide bandwidth. Loop-control networks, which consist of amplitude- and phase-shifting networks, maintain signal and distortion cancellation within the various feed-forward loops.

The input signal is first split into two paths, with one path going to the high-power main amplifier while the other signal path goes to a delay element. The output signal from the main amplifier contains both the desired signal and distortion. This signal is sampled and scaled using attenuators before being combined with the delayed portion of the input signal, which is regarded as distortion-free. The resulting "error signal" ideally contains only the distortion components in the output of the main amplifier. The error signal is then amplified by the low-power, high-linearity error amplifier, and then combined with a delayed version of the main amplifier output. This second combination ideally cancels the distortion components in the main-amplifier output while leaving the desired signal unaltered.

In practice, there is always some residual desired signal passing through the error amplifier. This is in general not a problem unless the additional power is sufficient in magnitude to degrade the linearity of the error amplifier and hence the linearity of the feedforward transmitter.

Signal Cancellation

Successful isolation of an error signal and the removal of distortion components depend upon precise signal cancellation over a band of frequencies. In practice, cancellation is achieved by the vector addition of signal voltages. The allowable amplitude and phase mismatches for different cancellation levels are shown in Figure 47. For manufactured equipment, realistic values of distortion

High Frequency Design

RF POWER AMPLIFIERS

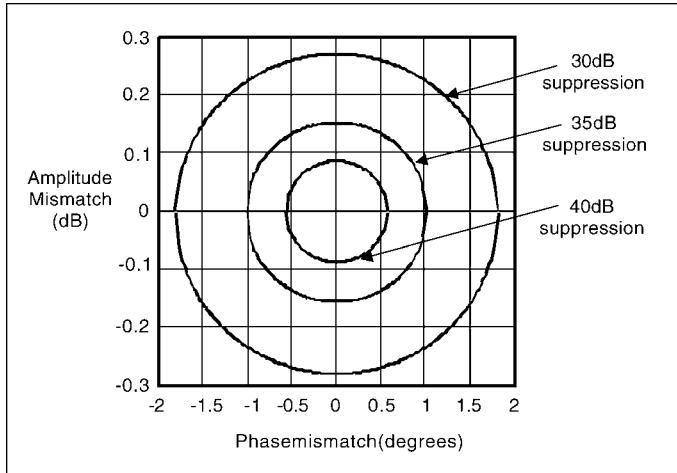


Figure 47 · Gain/phase matching requirements.

cancellation are around 25 to 30. The limiting factor is nearly always the bandwidth over which a given accuracy can be obtained.

Efficiency

The outputs of the main and error amplifiers are typically combined in a directional coupler that both isolates the PAs from each other and provides resistive input impedances. For a typical 10 dB coupling ratio, 90 percent of the power from the main PA reaches the output. For the same coupling ratio, only 10 percent of the power from the error amplifier reaches the load, thus the error amplifier must produce ten times the power of the distortion in the main amplifier. The peak-to-average ratio of the error signal is often much higher than that of the desired signal, making amplification of the error signal inherently much less efficient than that of the main signal. As a result, the power consumed by the error amplifier can be a significant fraction (e.g., one third) of that of the main amplifier. In addition, it may be necessary to operate one or both amplifiers well into back-off to improve linearity. The overall average efficiency of a feedforward transmitter may therefore be only 10 to 15 percent for typical multi-carrier signals.

Automatic Loop Control

Since feedforward is inherently an open-loop process,

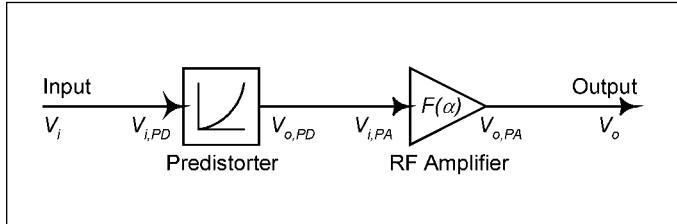


Figure 49 · Predistortion concept.

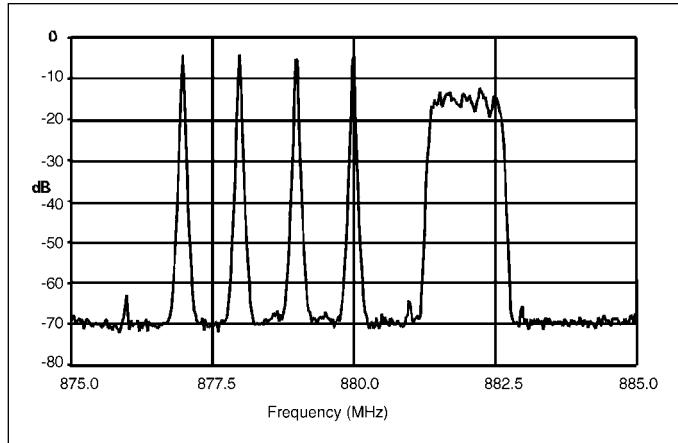


Figure 48 · Feedforward performance with mixed-mode modulation (TDMA and CDMA signals).

changes in device characteristics over time, temperature, voltage and signal level degrade the amplitude and phase matching and therefore increase distortion in the transmitter output. An automatic control scheme continuously adjusts the gain and phase to achieve the best signal cancellation and output linearity. The first step is to use FFT techniques, direct power measurement, or pilot signals to determine how well the loop is balanced. Both digital and analog techniques can be used for loop control and adjustment. Signal processing can be used to reduce the peaks in multi-carrier signals and to keep distortion products out of the nearby receiving band [111].

Performance

An example of the use of feedforward to improve linearity is shown in Figure 48. The signal consists of a mix of TDMA and CDMA carriers. The power amplifiers are based upon LDMOS transistors and have two-tone IMD levels in the range -30 to -35 dBc at nominal output power. The addition of feedforward reduces the level of distortion by approximately 30 dB to meet the required levels of better than -60 dBc. The average efficiency is typically about 10 percent.

8c. PREDISTORTION

The basic concept of a predistortion system (Figure 49) involves the insertion of a nonlinear element prior to the

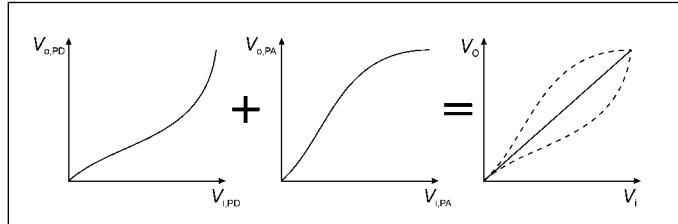


Figure 50 · Amplitude correction by predistortion.

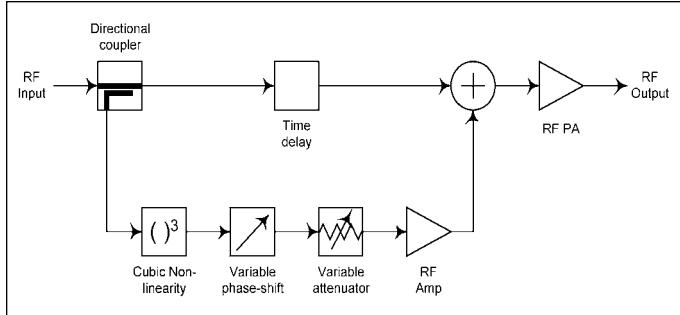


Figure 51 · An RF predistorter.

RF PA such that the combined transfer characteristic of both is linear (Figure 50). Predistortion can be accomplished at either RF or baseband.

RF Predistortion

The block diagram of a simple RF predistorter is shown in Figure 51. A compressive characteristic, created by the nonlinearity in the lower path (e.g., a diode) is subtracted from a linear characteristic (the upper path) to generate an expansive characteristic. The output of the linear path (typically just a time delay) is given by:

$$v_l(v_{in}) = a_1 v_{in} \quad (1)$$

and that of the compressive path is given by

$$v_c(v_{in}) = a_2 v_{in} - b v_{in}^3 \quad (2)$$

Subtracting the above equations gives

$$v_{pd}(v_{in}) = (a_2 - a_1)v_{in} - bv_{in}^3 \quad (3)$$

This is now an expansive characteristic with a linear gain of $a_1 - a_2$, and may be used to predistort a compressive amplifier characteristic (cubic in this example) by appropriate choice of a_1 , a_2 and b .

An example of the results from using a simple diode-based RF predistorter with a 120-W LDMOS PA amplifying an IS-95 CDMA signal is shown Figure 52. When applied to $\pi/4$ -DQPSK modulation in a satellite application, the same predistorter roughly halves the EVM, improves the efficiency from 22 to 29 percent, and doubles the available output power.

Predistortion bandwidths tend to be limited by similar factors to that of feedforward, namely gain and phase flatness of the predistorter itself and of the RF PA. In addition, memory effects in the PA and the predistorter limit the degree cancellation, and these tend to become poorer with increasing bandwidth.

Better performance can be achieved with more complex forms of RF predistortion such as Adaptive

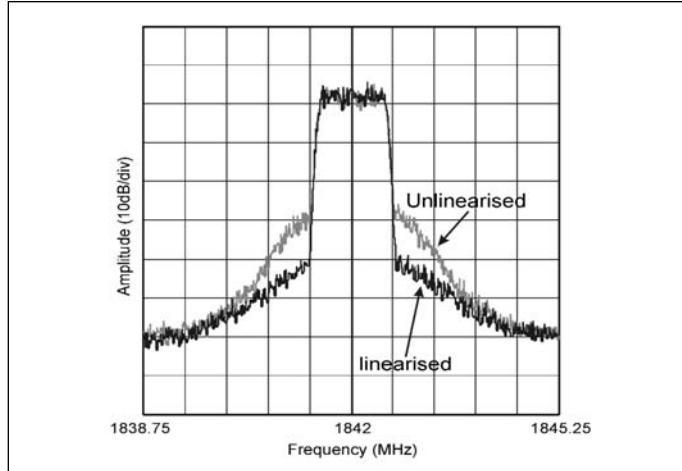


Figure 52 · Linearization by diode-based RF predistorter (courtesy WSI).

Parametric Linearization (APL®), which is capable of multi-order correction [106]. Most RF-predistortion techniques are capable of broadband operation with practical operational bandwidths similar to, or greater than, those of feedforward.

Digital Predistortion

Digital predistortion techniques exploit the considerable processing power now available from DSP devices, which allows them both to form and to update the required predistortion characteristic. They can operate with analog-baseband, digital-baseband, analog-IF, digital-IF, or analog-RF input signals. Digital-baseband and digital-IF processing are most common.

The two most common types of digital predistorter are termed mapping predistorters [107] and constant-gain

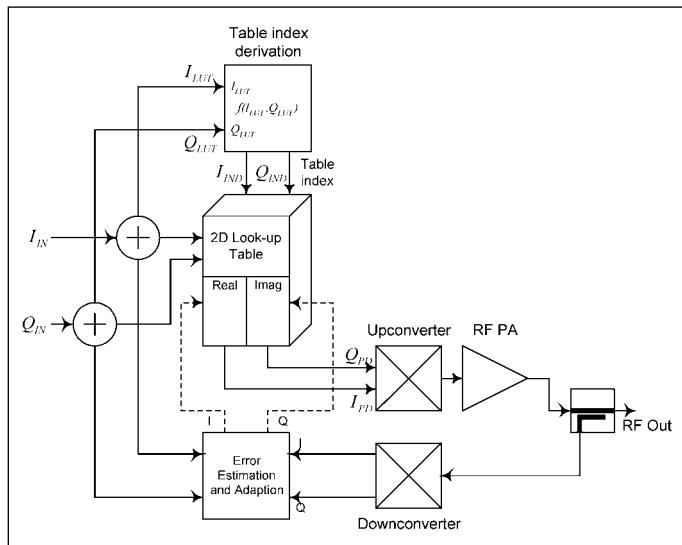


Figure 53 · Mapping predistorter.

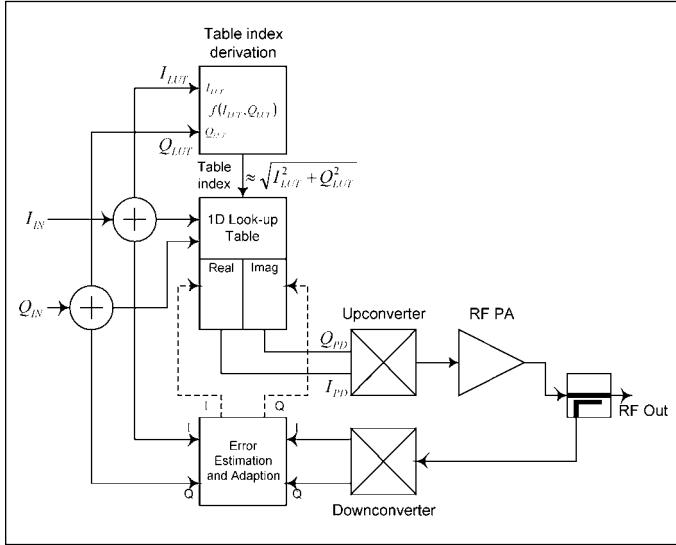


Figure 54 · Constant-gain predistorter.

predistorters [108]. A mapping predistorter utilizes two look-up tables, each of which is a function of two variables (I_{IN} and Q_{IN}), as shown in Figure 53. This type of predistorter is capable of excellent performance. However, it requires a significant storage and/or processing overhead for the look-up tables and their updating mechanism, and has a low speed of convergence. The low convergence speed results from the need to address all points in the I/Q complex plane before convergence can be completed.

A constant-gain predistorter (Figure 54) requires only a single-dimensional look-up table, indexed by the signal envelope. It is therefore a much simpler implementation and requires significantly less memory for a given level of performance and adaptation time. It uses the look-up table to force the predistorter and associated PA to exhibit a constant gain and phase at all envelope levels. The

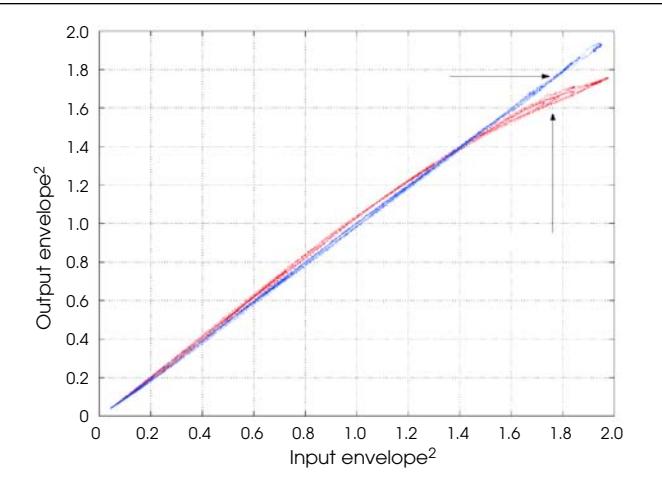


Figure 55 · Linearization of the amplitude transfer characteristic using an RF input/output digital predistorter (courtesy WSI).

overall transfer characteristic is then linear:

$$G_{PD}(I_{IN}(t), Q_{IN}(t)) \times G_{PA}(I_{PD}(t), Q_{PD}(t)) = k \quad (4)$$

An example of the improvement in the amplitude-transfer characteristic by an RF-input/output digital predistorter [109] is shown in Figure 55. The plot is based upon real-time using samples from a GSM-EDGE signal. Both the gain expansion and compression are improved by the linearizer. EVM is reduced from around 4.5 to 0.7 percent. The ACPR for IS-136 DAMPS modulation ($\pi/4$ -DQPSK) is reduced by nearly 20 dB (Figure 56). When generating mask-compliant EDGE modulation at full output power (850–900 MHz), the linearized PA has an efficiency of over 30 percent.

An example of linearization of a PA with two 3G W-

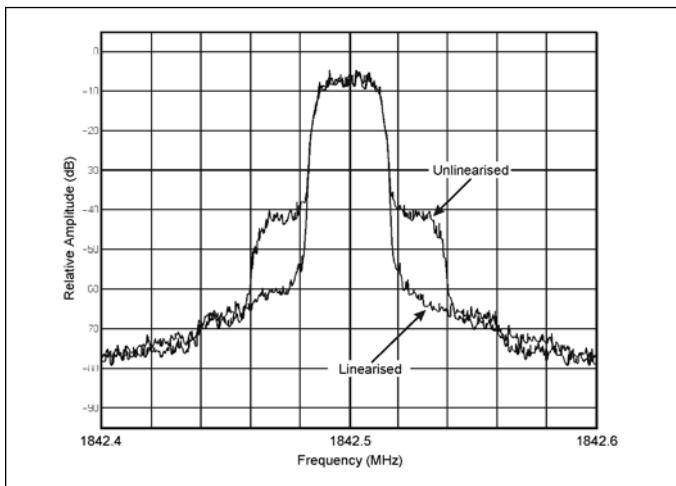


Figure 56 · Linearization of DAMPS PA by RF input/output predistorter (courtesy WSI).

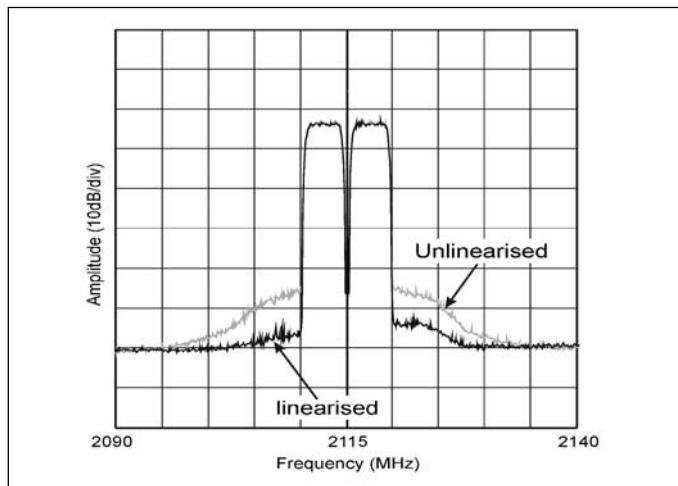


Figure 57 · Linearization of 3G W-CDMA PA signal by digital baseband input predistorter (courtesy WSI).



Figure 58 · A multi-carrier S-band transmitter with digital predistorter (courtesy WSI).

CDMA signals by a digital baseband-input predistorter is shown in Figure 57. The linearized amplifier meets the required spectral mask with a comfortable margin at all frequency offsets. The noise floor is set by the degree of clipping employed on the waveform, which limits the ACPR improvement obtained. It clearly demonstrates, however, that digital predistortion can be used in broadband as well as narrowband applications. Figure 58 shows a 3G transmitter that uses digital predistortion.

References

99. A. F. Mitchell, "A 135 MHz feedback amplifier," *IEE Colloq. Broadband High Frequency Amplifiers: Practice and Theory*, pp. 2/1-2/6, London, Nov. 22, 1979
100. P. B. Kenington, *High Linearity RF Amplifier Design*, Norwood, MA: Artech, 2000.
101. W. B. Bruene, "Distortion reducing means for single-sidedband transmitters," *Proc. IRE*, vol. 44, no. 12, pp. 1760-1765, Dec. 1956.
102. T. Arthanayake and H. B. Wood, "Linear amplification using envelope feedback," *Electronics Letters*, vol. 7, no. 7, pp. 145-146, April 8, 1971.
103. V. Petrovic and W. Gosling, "Polar-loop transmitter," *Electronics Letters*, vol. 15, no. 10, pp. 286-287, May 10, 1979.
104. V. Petrovic, "Reduction of spurious emission from radio transmitters by means of modulation feedback," *Proc. IEE Conf. No. 224 on Radio Spectrum Conservation Techniques*, UK., Sept. 6-8 1983.
105. E. Ballesteros, F. Perez, and J. Peres, "Analysis and design of microwave linearized amplifiers using active feedback," *IEEE Trans. Microwave Theory Tech.*, vol. 36, no. 3, pp. 499-504, March 1988.
106. P. B. Kenington, "Achieving high-efficiency in multi-carrier base-station power amplifiers," *Microwave Engr. Europe*, pp. 83-90. Sept. 1999.
107. Y. Nagata, "Linear amplification techniques for digital mobile communications," *Proc. IEEE Veh. Tech. Conf. (VTC '89)*, San Francisco, pp. 159-164, May 1-3, 1989.
108. J. K. Cavers, "Amplifier linearisation using a digital predistorter with fast adaptation and low memory requirements," *IEEE Trans. Veh. Tech.*, vol. 39, no. 4, pp. 374-382, Nov. 1990.
109. P. B. Kenington, M. Cope, R. M. Bennett, and J. Bishop, "GSM-EDGE high power amplifier utilising digital linearisation," *IMS'01 Digest*, Phoenix, AZ, May 20-25, 2001.
110. N. Pothecary, *Feedforward Linear Power Amplifiers*, Norwood, MA: Artech, 1999.
111. J. Tellado, *Multicarrier Modulation with Low PAR*, Boston: Kluwer, 2000.

Author Information

The authors of this series of articles are: Frederick H. Raab (lead author), Green Mountain Radio Research, e-mail: f.raab@ieee.org; Peter Asbeck, University of California at San Diego; Steve Cripps, Hywave Associates; Peter B. Kenington, Andrew Corporation; Zoya B. Popovic, University of Colorado; Nick Pothecary, Consultant; John F. Sevic, California Eastern Laboratories; and Nathan O. Sokal, Design Automation. Readers desiring more information should contact the lead author.

Acronyms Used in Part 4

ACPR	Adjacent Channel Power Ratio
APL	Adaptive Parametric Linearization
BER	Bit Error Rate
DAMPS	Digital American Mobile Phone System
EDGE	Enhanced Data for GSM Evolution
EVM	Error Vector Magnitude
IF	Intermediate Frequency
LDMOS	Laterally Diffused Metal Oxide Semiconductor
PA	Power Amplifier
PDF	Probability-Density Function
PMR	Private Mobile Radio
SMR	Specialized Mobile Radio
W-CDMA	Wideband Code-Division Multiple Access

RF and Microwave Power Amplifier and Transmitter Technologies — Part 5

By Frederick H. Raab, Peter Asbeck, Steve Cripps, Peter B. Kenington, Zoya B. Popovich, Nick Pothecary, John F. Sevic and Nathan O. Sokal

Emerging techniques are examined in this final installment of our series on power amplifier technologies, providing notes on new modulation methods and improvements in linearity and efficiency

signal, for example, taxes the capabilities of a Kahn-technique transmitter with a conventional class-S modulator. More acute are the problems in base-station and satellite transmitters, where multiple carriers must be amplified simultaneously, resulting in peak-to-average ratios of 10 to 13 dB and bandwidths of 30 to 100 MHz.

A number of the previously discussed techniques can be applied to this problem, including the Kahn EER with class-G modulator or split-band modulator, Chireix outphasing, and Doherty. This section presents some emerging technologies that may be applied to wideband, high efficiency amplification in the near future.

RF Pulse-Width Modulation

Variation of the duty ratio (pulse width) of a class-D RF PA [112] produces an amplitude-modulated carrier (Figure 59). The output envelope is proportional to the sine of the pulse width, hence the pulse width is varied in proportion to the inverse sine of the desired envelope. This can be accomplished in DSP, or by comparison of the desired envelope to a full-wave rectified sinusoid. The pulse timing

The ever-increasing demands for more bandwidth, coupled with requirements for both high linearity and high efficiency create ever-increasing challenges in the design of power amplifiers and transmitters. A single W-CDMA

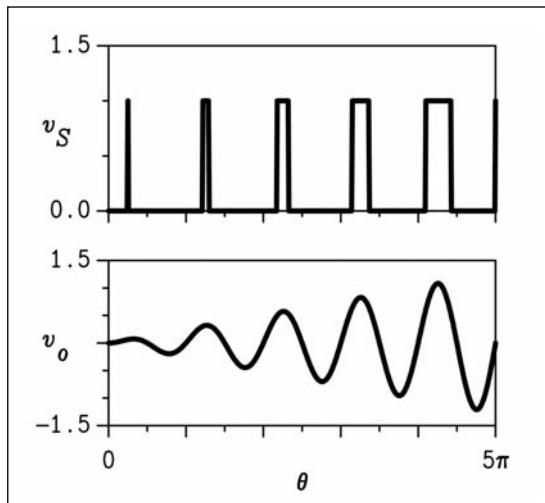


Figure 59 · RF pulse-width modulation.

conveys signal phase information as in the Kahn and other techniques.

Radio-frequency pulse-width modulation (RF PWM) eliminates the series-pass losses associated with the class-S modulator in a Kahn-technique transmitter. More importantly, the spurious products associated with PWM are located in the vicinity of the harmonics of the carrier [113] and therefore easily removed. Consequently, RF PWM can accommodate a significant RF bandwidth with only a simple, low-loss output filter.

Ideally, the efficiency is 100 percent. In practice, switching losses are increased over those in a class-D PA with a 50:50 duty ratio because drain current is nonzero during switching.

This series of articles is an expanded version of the paper, "Power Amplifiers and Transmitters for RF and Microwave" by the same authors, which appeared in the the 50th anniversary issue of the *IEEE Transactions on Microwave Theory and Techniques*, March 2002. © 2002 IEEE. Reprinted with permission.

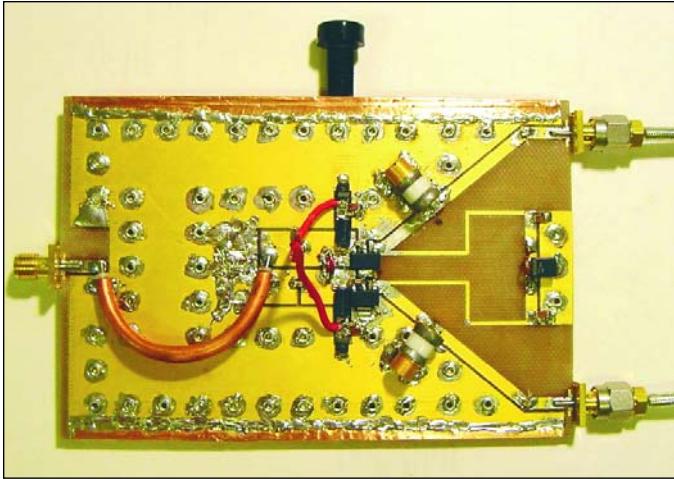


Figure 60 · Current-switching PA for 1 GHz (courtesy UCSD).

Previous applications of RF PWM have been limited to LF and MF transmitters (e.g., GWEN [114]). However, the recent development of class-D PAs for UHF and microwave frequencies (Figure 60) offers some interesting possibilities.

Delta-Sigma Modulation

Delta-sigma modulation is an alternative technique for directly modulating the carrier produced by a class-D RF PA (Figure 61) [PA8],[PA9]. In contrast to the basically analog operation of RF PWM, delta-sigma modulation drives the class-D PA at a fixed clock rate (hence fixed pulse width) that is generally higher than the carrier frequency (Figure 62). The polarity of the drive is toggled as necessary to create the desired output envelope from the

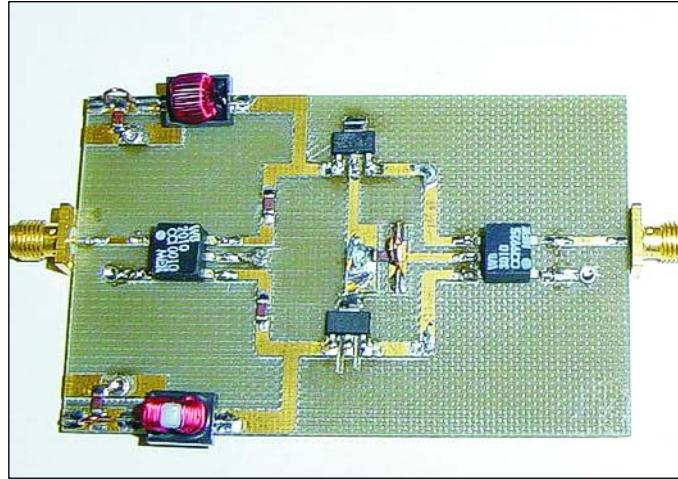


Figure 61 · Prototype class-D PA for delta-sigma modulation (courtesy UCSD).

average of the cycles in the PA. Phase is again conveyed in pulse timing.

The delta-sigma modulator employs an algorithm such as that shown in Figure 63. The signal is digitized by a quantizer (typically a single-bit comparator) whose output is subtracted from the input signal through a digital feedback loop, which acts as a band-pass filter. Basically, the output signal in the pass band is forced to track the desired input signal. The quantizing noise (associated with the averaging process necessary to obtain the desired instantaneous output amplitude) is forced outside of the pass band.

The degree of suppression of the quantization noise depends on the oversampling ratio; i.e., the ratio of the digital clock frequency to the RF bandwidth and is relatively independent of the RF center frequency. An example of the resultant spectrum for a single 900-MHz carrier and 3.6-GHz clock is shown in Figure 64. The quantization noise is reduced over a bandwidth of 50 MHz, which is sufficient for the entire cellular band. Out-of-band noise increases gradually and must be removed by a band-pass filter with sufficiently steep skirts.

As with RF PWM, the efficiency of a practical delta-sigma modulated class-D PA is reduced by switching losses associated with nonzero current at the times of switching. The narrow-band output filter may also introduce significant loss.

Carrier Pulse-Width Modulation

Carrier pulse-width modulation was first used in a UHF rescue radio at Cincinnati Electronics in the early 1970s. Basically, pulse-width modulation as in a class-S modulator gates the RF drive (hence RF drain current) on and off in bursts, as shown in Figure 65. The width of each burst is proportional to the instantaneous envelope of the

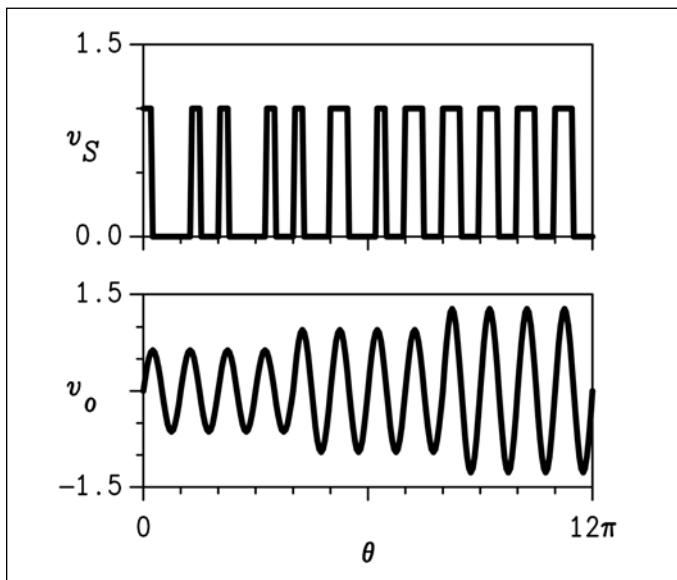


Figure 62 · Delta-sigma modulation.

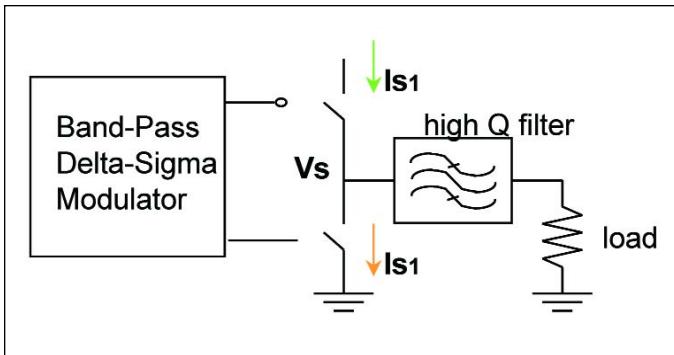


Figure 63 . Delta-sigma modulator.

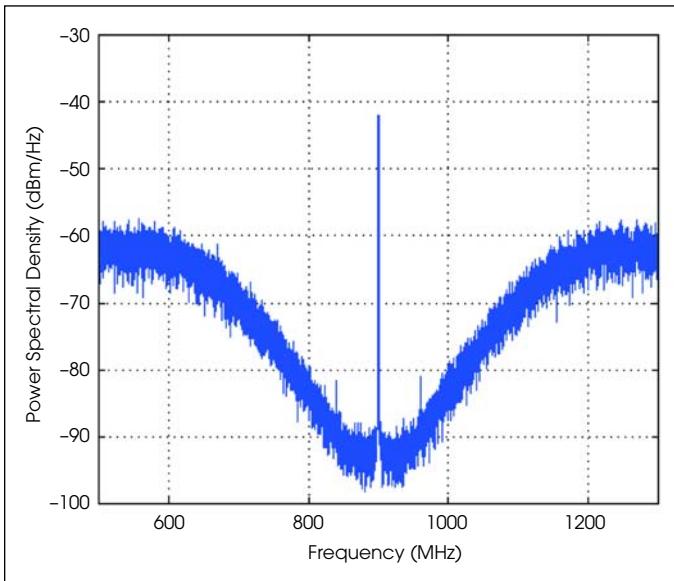


Figure 64 . Spectrum of delta-sigma modulation.

desired output. The amplitude-modulated output signal is recovered by a band-pass filter that removes the sidebands associated with the PWM switching frequency. The PWM signal can be generated by a comparator as in a class-S modulator or by delta-sigma techniques.

As with RF PWM and delta-sigma modulation, the series-pass losses and bandwidth limitations of the high-level modulator are eliminated. The switching frequency in carrier PWM is not limited by capabilities of power-switching devices and can therefore easily be tens of MHz, allowing large RF bandwidths. A second advantage is that carrier PWM can be applied to almost any type of RF PA. A disadvantage is that a narrow-band output filter with steep skirts is required to remove the switching-frequency sidebands, and such filters tend to have losses of 1 to 2 dB at microwave frequencies. Nonetheless, the losses in the filter may be more than offset by the improvement in efficiency for signals with high peak-to-average ratios.

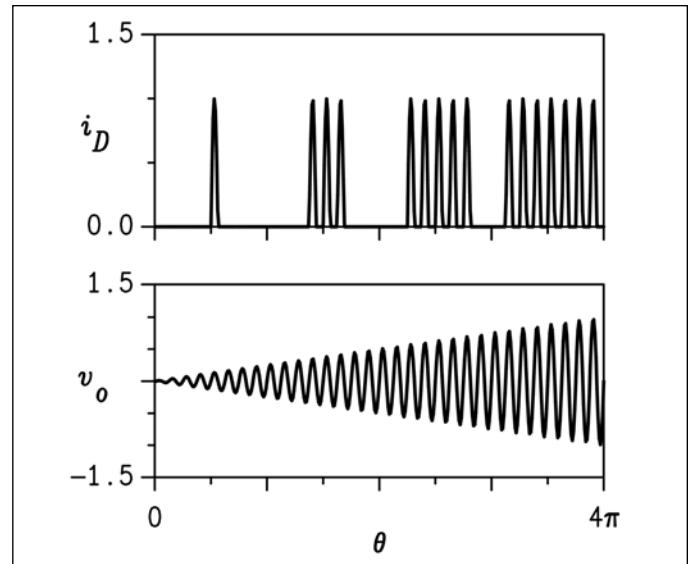


Figure 65 . Carrier pulse-width modulation.

Power Recovery

A number of RF processes result in significant RF power dissipated in “dump” resistors. Examples include power reflected from a mismatched load and dumped by a circulator and the difference between two inputs of hybrid combiner dumped to the balancing resistor. The notion of recovering and reusing wasted RF power was originally applied to the harmonics (18 percent of the output power) of an untuned LF class-D PA [117].

More recently, power recovery has been applied to out-phasing PAs with hybrid combiners [118, 119]. The instantaneous efficiency of such a system depends upon both the efficiency of the PA and that of the recovery system. Since the two PAs operate at full power regardless of the system output, inefficiency in the PA has a significant impact upon the system efficiency at the lower outputs. Nonetheless, a significant improvement over conventional hybrid-coupled outphasing is possible, and the PAs are presented with resistive loads that allow them to operate optimally. Typically, 50 percent of the dumped power can be recovered.

The power-recovery technology can also be used to implement miniature DC-DC converters. Basically, a high-efficiency RF-power amplifier (e.g., class-E) converts DC to RF and a high-efficiency rectifier circuit converts the RF to DC at the desired voltage. Implementation at microwave frequencies reduces the size of the tuning and filtering components, resulting in a very small physical size and high power density. In a prototype that operates at C band [120], the class-E PA uses a single MESFET to produce 120 mW with a PAE of 86 percent. The diode rectifier consists of a directional coupler with two Schottky

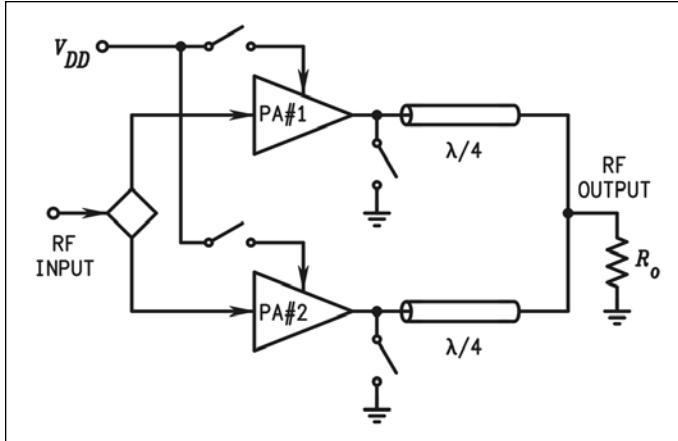


Figure 66 . Switched PAs with quarter-wavelength transmission line combiner.

diodes connected at the coupled and through ports and has a 98-percent conversion efficiency and an overall efficiency (including mismatch loss) of 83 percent. For a typical DC output of 3 V, the DC-DC conversion efficiency is 64 percent.

Switched PAs with Transmission-Line Combiners

RF-power amplifiers cannot simply be connected in series or parallel and switched on and off to make a transmitter module that adapts to variable peak envelope power. Attempting to do so generally produces either little effect or erratic variations in load impedance, sometimes leading to unstable operation and destruction of the transistors. Systems of microwave PAs that are toggled on and off are therefore connected through networks of quarter-wavelength transmission lines. The Doherty transmitter (discussed in part 4 of this series) is a classic example of this sort of technique.

An alternative topology (Figure 66) uses shorting switches and quarter-wavelength lines to decouple off-state PAs [121, 122]. The inactive PA is powered-down (by switching off its supply voltage), after which its output is shorted to ground. The quarter-wavelength line produces an open circuit at the opposite end where the outputs from multiple PAs are connected together to the load. This technique may be more easy to implement (especially for multiple PAs) than Doherty because a short is more readily realized than an open.

If PA #1 is the only PA active, its load is simply R_o . If both #1 and #2 are active, the combination produces an effective load impedance of $2R_o$ at the load ends of the lines. Inversion of this impedance through the lines places loads of $R_o/2$ on the RF PAs. Consequently, the peak power output for two active PAs is four times that with a single PA. As in discrete envelope tracking, the RF PAs operate as linear amplifiers. The number of PAs that

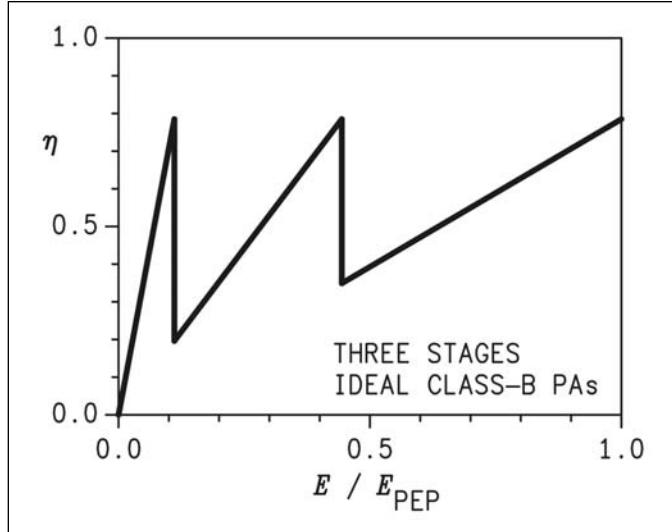


Figure 67 . Instantaneous efficiency of switched PAs.

are active is the minimum needed to produce the current output power. The peak power is thus kept relatively close to the saturated output, eliminating most of the effects of operating in back-off. The efficiency can therefore reach PEP efficiency at a number of different output levels, as shown in Figure 67.

The advantage of this technique is the ease in design associated with relying on short circuits rather than open circuits to isolate the off-state PAs. A possible disadvantage is operating individual PAs from multiple load impedances without retuning and a limited number of power steps available (e.g., 9/9, 4/9, 1/9 for a three-PA system).

Electronic Tuning

The performance of virtually all power amplifiers is degraded by load-impedance mismatch. Mismatched loads not only reduce efficiency, but also create higher stresses on the transistors. Because high-efficiency PAs generally require a specific set of harmonic impedances, their use is often restricted to narrow-band applications with well-defined loads.

Electronic tuning allows frequency agility, matching of unknown and variable loads, and amplitude modulation. Components for electronic tuning include pin-diode switches, MEMS switches, MEMS capacitors, semiconductor capacitors, ceramic capacitors (e.g., BST), and bias-controlled inductors. To date, electronic tuning has been applied mainly to small-signal circuits such as voltage-controlled oscillators. Recently demonstrated, however, are two electronically tuned power amplifiers. One operates in class E, produces 20 W with an efficiency of 60 to 70 percent, and can be tuned from 19 to 32 MHz (1.7:1 range) through the use of voltage-variable capacitors



Figure 68 . Electronically tunable class-D PA (courtesy GMRR).

[123, 124]. The second (Figure 68) operates in class D, produces 100 W with an efficiency of 60 to 70 percent, and can be tuned from 5 to 21 MHz (4.25:1 range) through the use of electronically tunable inductors and capacitors [125].

Load Modulation

The output of a power amplifier can be controlled by varying the drive, gate bias, DC supply voltage, or load impedance. “Load modulation” uses an electronically tuned output filter (Figure 69) to vary load impedance and thereby the instantaneous amplitude of the output signal. The modulation bandwidth can be quite wide, as it is limited only by the bias feeds to the tuning components.

A key aspect of load modulation is a diligent choice of the impedance locus so that it provides both good dynamic range and good efficiency. For ideal saturated PAs of classes A, B, C, and F, the optimum locus is the pure resistance line on the Smith chart that runs from the nominal load to an infinite load. For ideal class-E PAs with series inductance and shunt susceptance for optimum operation with the nominal load, the optimum locus is the unity-efficiency line running from the nominal load upward and rightward at an angle of 65° [126]. For real PAs, the opti-

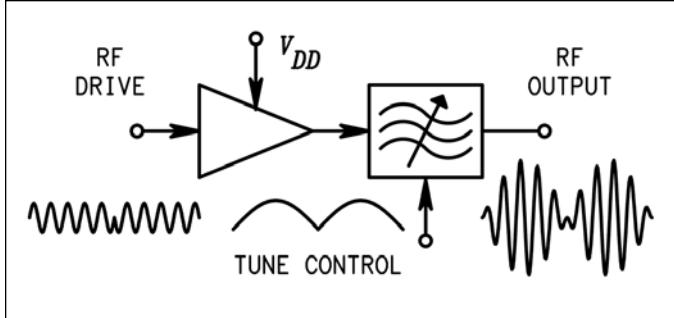


Figure 69 . Load modulation by electronic tuning.



Figure 70 . Load-modulated class-E PA (courtesy GMRR).

imum locus is found by examination of load-pull contours.

The simple T filter has a single electronically variable element, but provides an approximately optimum locus for class E over the top 12 dB of the dynamic range. The experimental 20-W, 30-MHz [124, 126] shown in Figure 70 achieves a 41-dB range of amplitude variation. The measured instantaneous-efficiency curve (Figure 71) corresponds to a factor of 2.1 improvement in the average efficiency for a Rayleigh-envelope signal with a 10-dB peak-to-average ratio.

A load-modulated PA for communications follows the electronically tuned filter with a passive filter to remove the harmonics associated with the nonlinear elements. Predistortion compensates for the incidental phase modulation inherent in dynamic tuning of the filter. Variation of the drive level can be used to conserve drive power and to extend the dynamic range.

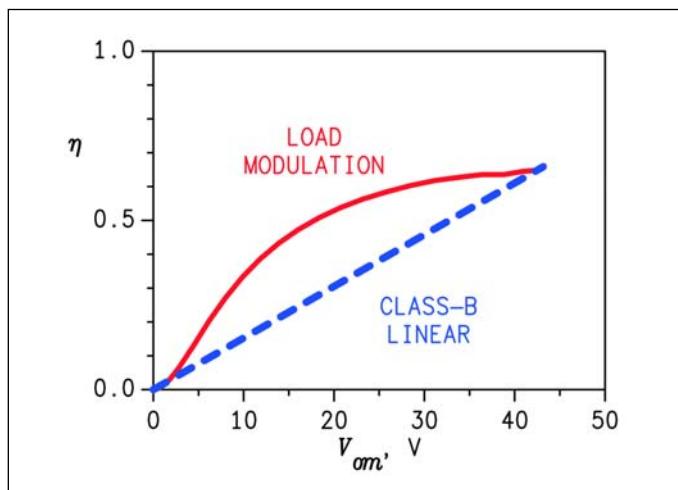


Figure 71 . Instantaneous efficiency of load modulation compared to class-B linear amplification.

References

112. Figures 8 and 9 in Part 2 of this series, *High Frequency Electronics*, July 2003.
113. F. H. Raab, "Radio frequency pulsedwidth modulation," *IEEE Trans. Commun.*, vol. COM-21, no. 8, pp. 958-966, Aug. 1973.
114. F. G. Tinta, "Direct single sideband modulation of transmitter output switcher stages," *Proc. RF Expo East '86*, Boston, MA, pp. 313-398, Nov. 10-12, 1986.
115. A. Jayaraman, P. F. Chen, G. Hanington, L. Larson, and P. Asbeck, "Linear high-efficiency microwave power amplifiers using bandpass delta-sigma modulators," *IEEE Microwave Guided Wave Lett.*, vol. 8, no. 3, pp. 121-123, March 1998.
116. J. Keyzer et al., "Generation of RF pulsedwidth modulated microwave signals using delta-sigma modulation," *2002 Int. Microwave Symp. Digest*, vol. 1, pp. 397-400, Seattle, WA, June 2-7, 2002.
117. J. D. Rogers and J. J. Wormser, "Solid-state high-power low frequency telemetry transmitters," *Proc. NEC*, vol. 22, pp. 171-176, Oct. 1966.
118. R. E. Stengel and S. A. Olson, "Method and apparatus for efficient signal power amplification," U.S. Patent 5,892,395, Apr. 6, 1999.
119. R. Langridge, T. Thornton, P. M. Asbeck, and L. E. Larson, "A power re-use technique for improving efficiency of outphasing microwave power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 8, pp. 1467-1470 , Aug. 1999.
120. S. Djukic, D. Maksimovic, and Z. Popovic, "A planar 4.5-GHz dc-dc power converter," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 8, pp. 1457-1460, Aug. 1999.
121. A. Shirvani, D. K. Su, and B. A. Wooley, "A CMOS RF power amplifier with parallel amplification for efficient power control," *IEEE J. Solid-State Circuits*, vol. 37, no. 6, pp. 684-693, June 2002.
122. C. Y. Hang, Y. Wang, and T. Itoh, "A new amplifier power combin-
- ing scheme with optimum efficiency under variable outputs," *2002 Int. Microwave Symp. Digest*, vol. 2, pp. 913-916, Seattle, WA, June 2-7, 2002.
123. F. H. Raab, "Electronically tunable class-E power amplifier," *Int. Microwave Symp. Digest*, Phoenix, AZ, pp. 1513-1516, May 20-25, 2001.
124. F. H. Raab, "Electronically tuned power amplifier," Patent pending.
125. F. H. Raab and D. Ruppe, "Frequency-agile class-D power amplifier," *Ninth Int. Conf. on HF Radio Systems and Techniques*, pp. 81-85, University of Bath, UK, June 23-26, 2003.
126. F. H. Raab, "High-efficiency linear amplification by dynamic load modulation," *Int. Microwave Symp. Digest*, vol. 3, pp. 1717-1720, Philadelphia, PA, June 8-13, 2003.

Correction

In Part 4 of this series (November 2003 issue), Figures 45, 52, 55, 56, 57 and 58 should have been credited as "Courtesy Andrew Corporation" instead of "Courtesy WSI."

Author Information

The authors of this series of articles are: Frederick H. Raab (lead author), Green Mountain Radio Research, e-mail: f.raab@ieee.org; Peter Asbeck, University of California at San Diego; Steve Cripps, Hywave Associates; Peter B. Kenington, Andrew Corporation; Zoya B. Popovic, University of Colorado; Nick Pothecary, Consultant; John F. Sevic, California Eastern Laboratories; and Nathan O. Sokal, Design Automation. Readers desiring more information should contact the lead author.

Acronyms Used in Part 5

BST	Barium Strontium Titanate
GWEN	Ground Wave Emergency Network
PWM	Pulse-Width Modulation

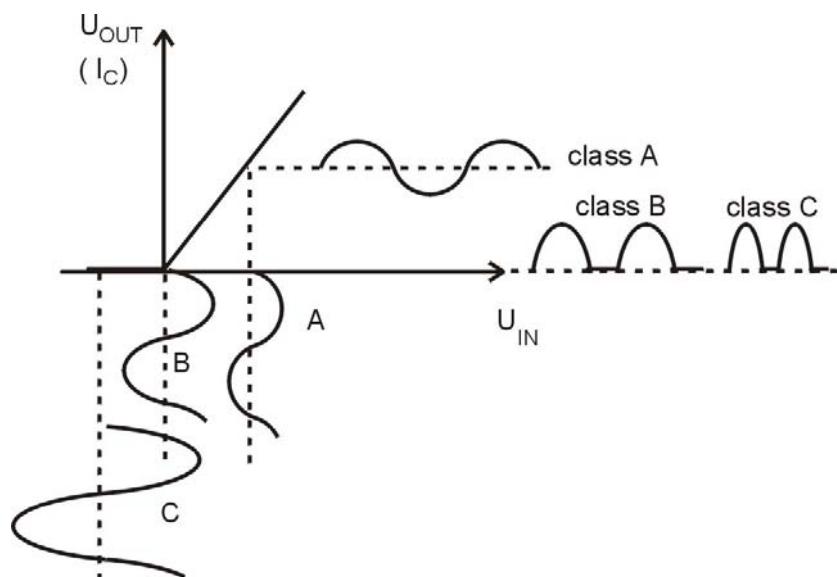
Power amplifiers

Amplifiers are used to increase the level of a signal and, depending on the increase required, stages are often cascaded to increase the gain. The last stage of the cascade may be required to drive same form of load, for example a loudspeaker, a servo mechanism or a coaxial cable for RF applications. In each case the load resistance is likely to be much lower than the output resistance of a typical amplifier stage and this can result in a considerable loss of signal (due to mismatch of resistance). In addition to providing a match to the load, the last stage is also usually required to generate a large signal to deliver sufficient power to the load.

The analysis of the output power stage is complicated by the fact that the signal levels are not small and the equivalent small – signal models for the transistors no longer apply. Large signal models are available for software simulation (e.g. Ebers – Moll model). For hand – based calculations linearized large – signal models are used to achieve rough estimates. However, an important parameter for an output stage, which is the amount of distortion resulting from nonlinearity in the output transistor characteristics, can only be obtained with the use of nonlinear models. The amount of distortion is usually specified as the total harmonic distortion (THD) which is the ratio of the RMS value of the harmonics in the output signal to the RMS of the fundamental.

An important design requirement for an output stage is the efficient transfer of power to the load, which may be hundreds of watts. This power is provided by the DC supply, which for low – voltage transistor circuits must be capable of supplying large currents. It is also necessary to minimize the power dissipation in the output transistors of about 150°C, there are also requirements to prolong the life of batteries in portable equipment by using the power efficiently as possible.

Power amplifiers are classified by the nature of the collector current waveform into class A, class B and class C. This classification is explained in the following figure, where a typical transfer characteristic of a transistor amplifier is used.



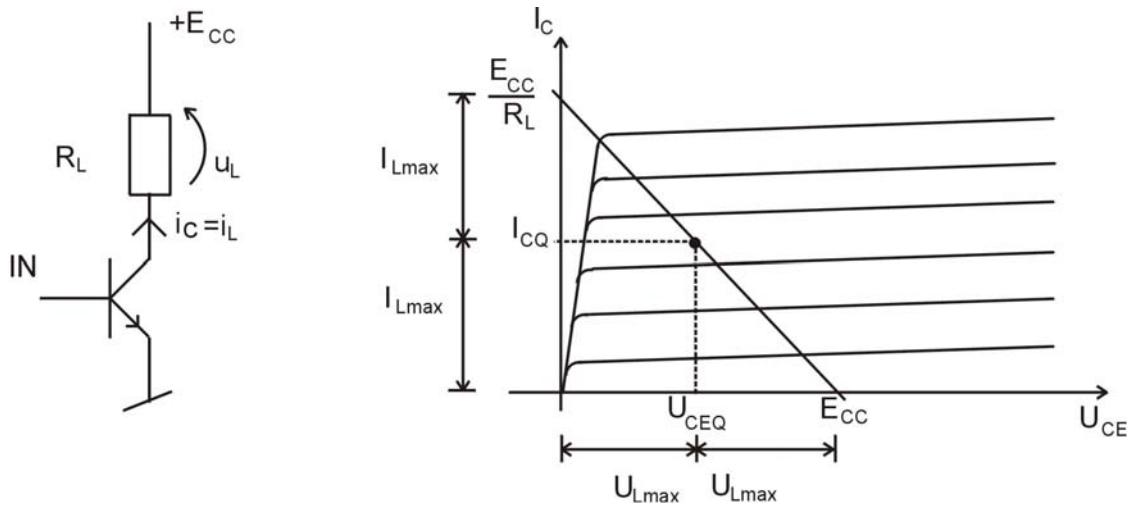
In the simplest class A output stage the transistor conducts for the complete cycle (the conduction angle θ is 2π). When the signal is removed a steady collector current flows in the load.

In the class B power amplifier the transistor only conducts when the input signal is present and is positive ($\theta=2\pi$). When the input signal is absent no collector current flows. Two transistors are required in order to reproduce the complete input signal for a complete cycle of a sinwave.

In the class C amplifier the current is only allowed to flow for a small fraction of each cycle ($\theta<\pi$). This pulsating collector current is used to generate a sine wave in a high Q turned LC circuit (radio frequency transmitter circuits).

Class A amplifier

The common – emitter (common – source) and emitter follower (source – follower) are often used as class A amplifiers. In both cases the output waveform is a faithful reproduction of the input wave, but both stages are very inefficient in the conversion of DC power from the supply to signal power in the load. However, the simplicity of the class A means that it is useful for low – power output stages of a few hundred miliwatts to drive a small loud speaker or earphones. The design involves the correct choice of DC biasing to ensure maximum symmetrical output swing.



The optimum Q – point for maximum symmetrical swing is obtained by allowing

$$I_{CQ} \approx \frac{E_{CC}}{2R_L}$$

$$U_{CEQ} \approx \frac{E_{CC}}{2}$$

An important factor in the design of a power amplifier is the efficient transfer of AC power from the DC source to the load. Efficiency is defined as:

$$\eta = \frac{\text{average signal power delivered to the load}}{\text{average power drawn from DC source}}$$

In further analyses the subscript “m” is used to denote the amplitude of the signal (with respect to the Q – point) and the subscript “max” means the maximum amplitude of the undistorted signal.

The average AC power delivered to the load is:

$$P_L = \frac{U_{Lm} I_{Lm}}{2} = \frac{U_{Lm}}{2R_L} = \frac{1}{2} I_{Lm}^2 R_L$$

The DC power taken from the supply is:

$$P_{Sup} = \frac{1}{2\pi} \int_0^{2\pi} E_{CC} (I_{CQ} + I_{Lm} \sin \omega t) d\omega t = E_{CC} I_{CQ}$$

$$I_{CQ} = \frac{1}{2} \frac{E_{CC}}{R_L} \quad P_{Sup} = \frac{1}{2} \frac{E_{CC}^2}{R_L}$$

The efficiency is:

$$\eta = \frac{P_L}{P_{Sup}} = \frac{\frac{U_{Lm}^2}{2R_L}}{\frac{E_{CC}^2}{2R_L}} = \left(\frac{U_{Lm}}{E_{CC}} \right)^2$$

and increases linearly with the amplitude of the signal.

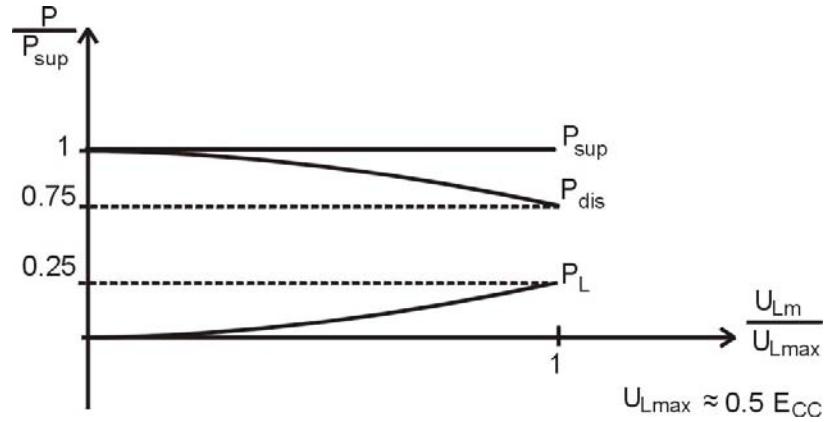
For the maximum amplitude of the undistorted signal:

$$U_{Lmax} \approx \frac{E_{CC}}{2}$$

the efficiency reaches its theoretical maximum:

$$\eta_{max} = \left(\frac{\frac{E_{CC}}{2}}{E_{CC}} \right)^2 = \frac{1}{4} \cong 25\%$$

The figure below shows the relationships between the power taken from the supply (P_{sup}), the power delivered to the load (P_L) and the power dissipated as heat ($P_{dis} = P_{sup} - P_L$)



With no signal the dissipated power reaches its maximum and is equally distributed between the load and the transistor:

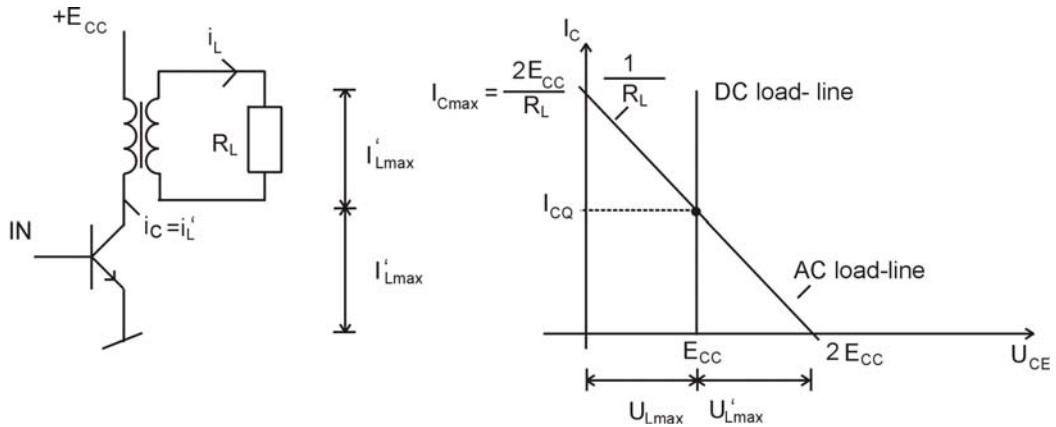
$$P_{disL} = P_{disT} = \frac{1}{2} P_{\text{sup}} = \frac{E_{CC}^2}{4R_L}$$

The maximum power delivered to the load is :

$$P_{L\max} = \frac{U_{L\max}^2}{2R_L} = \frac{\left(\frac{1}{2}E_{CC}\right)^2}{2R_L} = \frac{E_{CC}^2}{8R_L}$$

This means that the power rating for the transistor should be twice the maximum output power.

In the common – emitter amplifier there is a large mismatch between the load and the output resistance of the amplifier. To solve this problem, a transformer coupling can be used. The circuit and the corresponding load – lines are shown in the following figures.



With no signal applied the C-E voltage is E_{CC} and the load – line is vertical (if the resistance of the primary winding is ignored). When a sine wave signal is applied, the collector “sees” the resistance R_L transferred from the secondary to the primary:

$$R_L = \frac{R_L}{P^2} \quad p\text{-turns ratio}$$

To obtain maximum symmetrical signal swing the DC operating current should be set to

$$I_{CQ} = \frac{1}{2} I_{Cmax} = \frac{E_{CC}}{R_L}$$

The average AC power delivered to the load is:

$$P_L = \frac{U_{Lm} I_{Lm}}{2} \approx \frac{U_{Lm}^2}{2} = \frac{U_{Lm}^2}{2R_L} = \frac{1}{2} I_{Lm}^2 R_L$$

The superscript “ ’ ” refers to the primary.

The average power taken from the supply is:

$$P_{sup} = \frac{1}{2\pi} \int_0^{2\pi} E_{CC} (I_{CQ} + I_{Lm} \sin \omega t) d\omega t = E_{CC} I_{CQ}$$

$$I_{CQ} = \frac{E_{CC}}{R_L} \quad P_{sup} = \frac{E_{CC}^2}{R_L} \quad (\text{if } I_{CQ} = \frac{1}{2} I_{Cmax})$$

the efficiency is:

$$\eta = \frac{P_L}{P_{sup}} = \frac{\frac{U_{Lm}^2}{2R_L}}{\frac{E_{CC}^2}{R_L}} = \frac{1}{2} \left(\frac{U_{Lm}}{E_{CC}} \right)^2$$

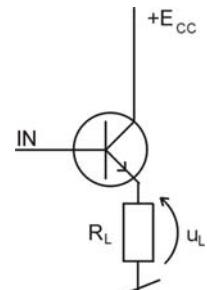
In this case the maximum amplitude of the undistorted collector – emitter voltage (and U_{Lmax} too) is E_{CC} , so

$$\eta_{max} = \frac{1}{2} \approx 50\%$$

Transformer coupling of the load ensures higher efficiency than the basic CE amplifier. nevertheless, this circuit is rarely used, because the performance of the transformer is poor due to DC component of the primary current and it is difficult to ensure a flat frequency response over a wide range of frequency.

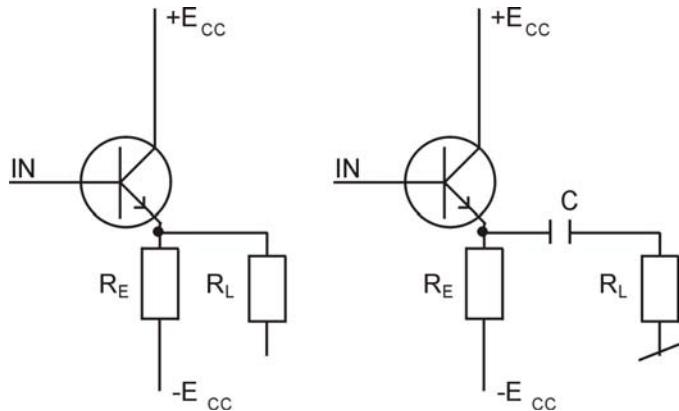
Notice that with no signal the DC power is totally dissipated in the transistor. Its the DC power is constant and the maximum efficiency is 50%, the power rating for the transistor should be twice the maximum output power.

For the common – collector (emitter follower) power amplifier, with the load connected directly between the emitter and the ground as in the figure, the maximum theoretical efficiency is the same as for the CE amplifier with direct – coupled load (25%).



Although its voltage gain is slightly smaller than 1, the low output resistance of the emitter follower is much better suited to driving low – resistance loads.

To avoid DC power dissipation in the load, the emitter follower can be arranged in such a way, that the voltage drop across the load is approximately zero with no signal applied. Two possible solutions are shown in the following figures.



In the first circuit the DC collector – emitter voltage is set to:

$$U_{CEQ} = E_{CC} \quad (U_{CEQ} = 0.5 E_{CC} \text{ in the second circuit})$$

and

$$I_{CQ} = \frac{E_{CC}}{R_E}$$

The power delivered by both supplies is:

$$P_{\text{sup}} = 2E_{CC} I_{CQ} = 2 \frac{E_{CC}^2}{R_E}$$

The average AC power delivered to the load is:

$$P_L = \frac{U_{Lm}^2}{2R_L}$$

Notice that in this case the maximum amplitude of the undistorted output signal results from the fact that for negative halfcycles of the input voltage the transistor can enter cut – off region and then

$$U_{L\max} = E_{CC} \frac{R_L}{R_E + R_L}$$

The maximum theoretical efficiency becomes

$$\eta_{\max} = \frac{P_{L\max}}{P_{\text{sup}}} = \frac{\frac{1}{2}R_L \left(E_{CC} \frac{R_L}{R_E + R_L} \right)^2}{\frac{2E_{CC}^2}{R_E}} = \frac{1}{4} \frac{R_E R_L}{(R_E + R_L)^2}$$

R_L is often chosen to be equal to R_E and then

$$\eta_{\max} = \frac{1}{16} \approx 6.25\%$$

This means that the power rating for the transistor should be at least 8 times the maximum output power.

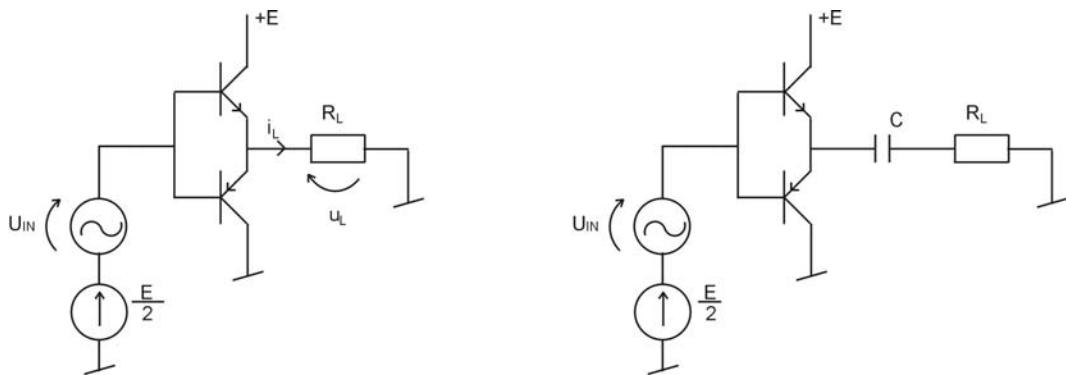
Class B amplifier

When the transistor conducts for only half of a sinusoidal cycle and is cut off for the remaining half, it is said to be a class B amplifier. This mode of operation is illustrated in the following figures for the both npn and pnp transistors.

$$I_{CQ} = 0 \quad U_{CEQ} = E_{CC}$$

$$I_{CQ} = 0 \quad U_{CEQ} = -E_{CC}$$

To create a complete sinusoidal cycle across the load, two such stages can be combined in a complementary push – pull amplifier as shown in the figure.



It is also possible to use a single power supply as in the figure on the right. In this circuit a large capacitor is charged to $0.5E$ and acts as an additional supply voltage source so that the equivalent supply voltage for each transistor is $0.5 E_{CC}$.

Let us consider the circuit on the left to determine the efficiency of the class B amplifier.

The average AC power delivered to the load is

$$P_L = \frac{U_{Lm}^2}{2R_L}$$

The average power taken from both power supplies becomes a function of the output signal amplitude now:

$$P_{\text{sup}} = 2 \cdot \frac{1}{2\Pi} \int_0^{\Pi} E_{CC} I_{Cm} \sin \omega t d\omega t = \frac{2}{\Pi} E_{CC} I_{Cm}$$

but

$$I_{Cm} \approx I_{Lm} = \frac{U_{Lm}}{R_L}$$

so:

$$P_{\text{sup}} = \frac{2}{\Pi} \frac{E_{CC} U_{Lm}}{R_L}$$

and

$$\eta = \frac{P_L}{P_{\text{sup}}} = \frac{\frac{U_{Lm}^2}{2R_L}}{\frac{2}{\Pi} \frac{E_{CC} U_{Lm}}{R_L}} = \frac{\Pi U_{Lm}}{4 E_{CC}}$$

The maximum amplitude of the undistorted signal is:

$$U_{L\max} \approx E_{CC}$$

and, finally,

$$\eta_{\max} = \frac{\Pi}{4} \approx 78,5\%$$

The power dissipated in the circuit is:

$$P_{\text{dis}} = P_{\text{sup}} - P_L = \frac{2}{\Pi} \frac{E_{CC} U_{Lm}}{R_L} - \frac{U_{Lm}^2}{2R_L}$$

and reaches a maximum value of

$$P_{dis\max} = \frac{2E_{CC}}{\Pi^2 R_L} \quad at \quad U_{Lm} = \frac{2}{\Pi} E_{CC} \approx 0.64 U_{Lmax}$$

This is the total power dissipated in both transistors. Thus each transistor must be capable of dissipating half this amount, that is:

$$P_{C\max} = \frac{1}{2} P_{dis\max} = \frac{E_{CC}^2}{\Pi^2 R_L}$$

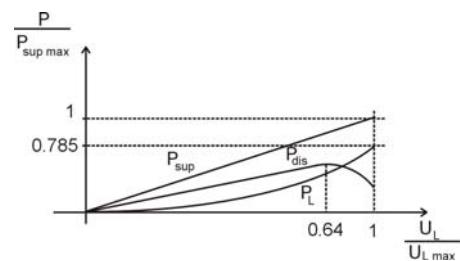
The ratio of collector power to maximum load power is:

$$\frac{P_{C\max}}{P_{L\max}} = \frac{\frac{E_{CC}^2}{\Pi^2 R_L}}{\frac{E_{CC}^2}{2R_L}} = \frac{2}{\Pi^2} \approx 0.2$$

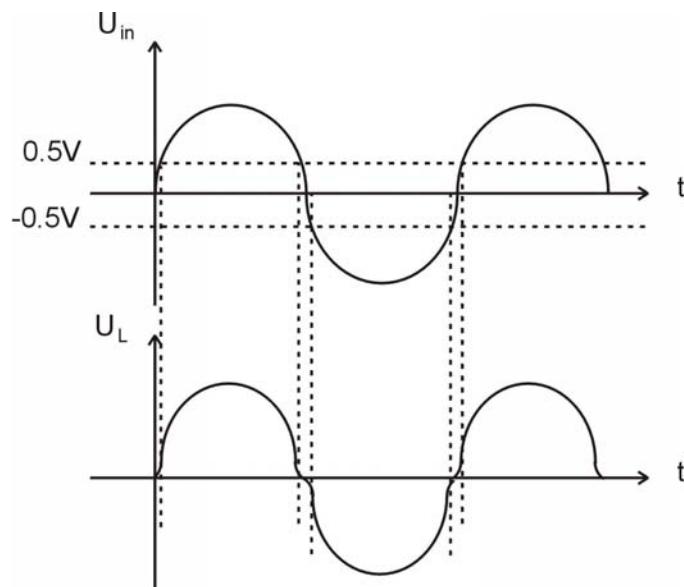
This represents an improvement by a factor of 10 over the same relationship for the class A amplifier. Thus for every watt delivered to the load only 0.2W is dissipated in each transistor.

The relationship between P_{sup} , P_L , P_{dis} and U_{Lm} are shown in the figure.

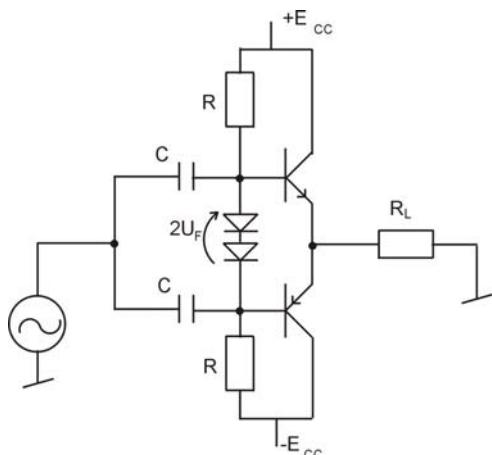
Thanks to the linear relationship between P_{sup} and U_{Lm} the use of energy is much more efficient which is especially important in battery – powered equipment.



The circuit shown previously for a class B amplifier is simple but it produces a large amount of distortion. In order that the transistors conduct when a signal is applied the input signal must exceed approximately 0.5 V. When the input signal drops below +/- 0.5V no transistor conducts and this results in the distortion of the output waveform that is known as cross – over distortion and shown in the figure below.



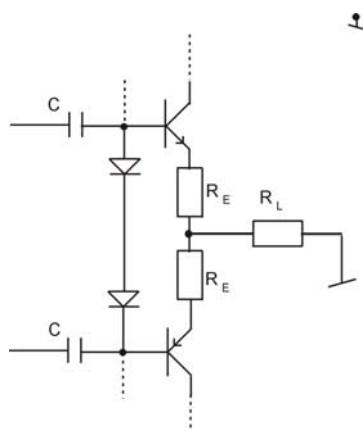
The cross – over distortion can be reduced by the application of negative feedback but a more satisfactory solution is to provide a small forward bias so that the transistors conduct when the input is reduced to zero, as shown in the figure.



This mode of operation is known as class AB.

The potential divider comprising two diodes provides a bias to the bases of the two transistors. This allows the transistors to conduct when the input signal is zero.

The bias diodes should operate at the same temperature as the transistors (mounted on the same heatsink). for a fixed base bias (e.g. resistors used in place of the diodes) a temperature rise would result in the collector current increasing exponentially. This would increase the power dissipation causing a further increase in the temperature, a further reduction in U_{BE} and a further rise in the collector currents. This is a form of positive feedback which can lead to thermal runaway.



A further improvement in thermal stability can be achieved by employing negative feedback, as shown in the figure, with the use of two emitter resistors. These are typically a fraction of Ω or a few ohms, depending on the output power required.

Transistor power dissipation.

One of the most important parameters that should be considered while choosing a transistor for a power amplifier is the maximum power rating. The load line should preferably be tangent to the maximum power hyperbola (at the mid – point for class A amplifiers).

The power dissipation heats the transistor and causes the temperature of the silicon chip to rise. for silicon transistors the maximum permitted temperature is between 150°C to 200°C. To dissipate this heat power transistors are mounted in metal or plastic cases used for small – signal transistors. The process is further improved by attaching the case of the transistor to a heat sink.

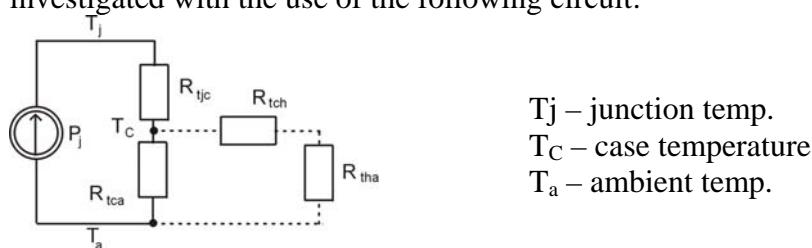
The heat sink may simply be the metal chassis of the instrument case or it may be a separate aluminium finned heat sink.

The analysis of heat flow makes use of a simple analogy of ohm's law in which current is replaced by power (heat), voltage by temperature and electrical resistance by thermal resistance:

$$\Delta T = P_j \cdot R_t \quad R_t \left[\frac{\text{deg}}{w} \right]$$

P_j is the power generated at the junction and ΔT is the temperature difference between the junction and the surroundings (ambient).

Heat flow across several boundaries, each with a different thermal resistance, can be investigated with the use of the following circuit:



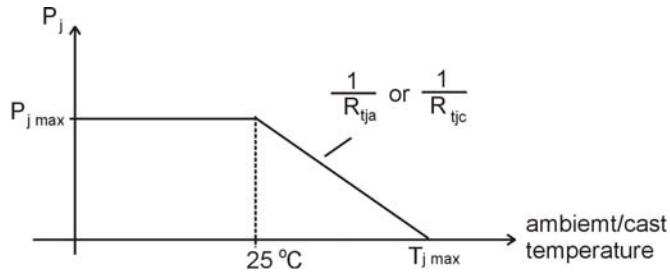
R_{tjc} – junction – to – case thermal resistance

R_{tca} – case – to ambient thermal resistance

R_{tch} – case – to – heat sink thermal resistance (usually a mica washer with silicone grease)

R_{tha} – heat sink – to – ambient thermal resistance

Manufacturers specify a maximum permissible power for their transistors at certain ambient/case temperatures, for example not exceeding 25°C. If the ambient/case temperature is above this value, then the power dissipation must be reduced according to a derating curve of power versus temperature, which is shown below.



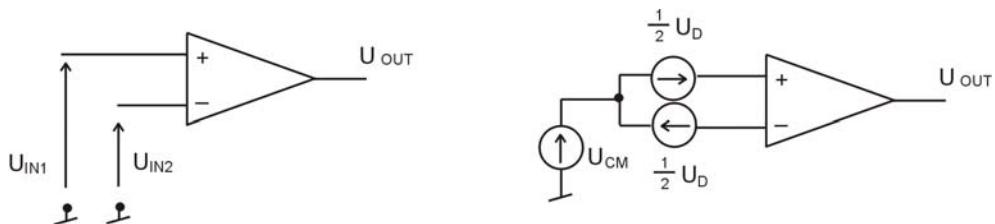
The slope of the curve beyond the critical temperature is the reciprocal of the thermal resistance for the junction to ambient/case.

DIFFERENTIAL (DIFFERENCE) AMPLIFIER

The single – ended amplifiers considered so far have involved one signal source followed by one or more transistor gain stages which produce an output which is proportional to the input.

The differential amplifier is an arrangement of transistors which allows the difference between two signal sources to be amplified and the output is proportional to the difference between these two inputs. The differential amplifier is used extensively in integrated circuit operational amplifiers.

A simple symbolic representation of the differential amplifier is shown in the figure below.



For any input signals they can always be represented as two components: (or difference-)

- the differential – mode signal

$$U_D = U_{IN1} - U_{IN2}$$

- the common – mode signal

$$U_{CM} = \frac{U_{IN1} + U_{IN2}}{2}$$

The differential signal is connected between the two inputs and the output may be expressed as:

$$U_{out} = k_D U_D$$

where k_D is the differential – mode gain.

The common – mode signal, which is common to each input, also produces an output which may be expressed as:

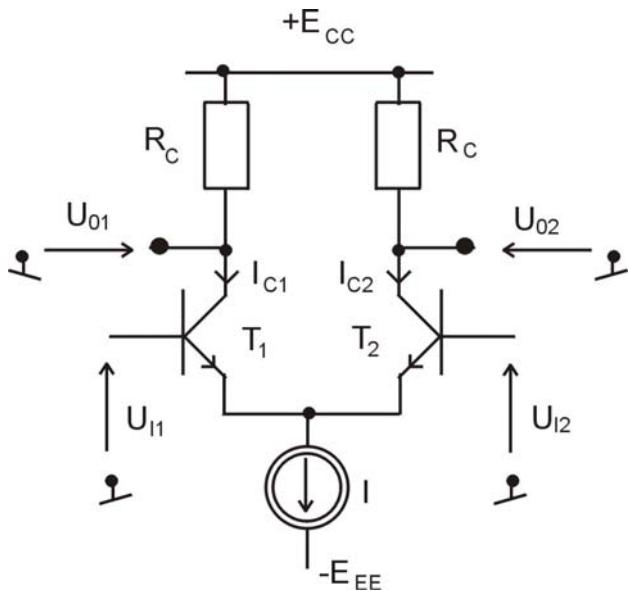
$$U_{out} = k_{CM} U_{CM}$$

The required output is $k_D U_D$ and the component $k_{CM} U_{CM}$ is an unwanted source of interference. To minimize the effect of the common – mode signal it is necessary for k_D to be much greater than k_{CM} .

The ratio of the two gain is an important parameter of differential amplifiers and is known as the Common – Mode Rejection Ratio (CMRR):

$$CMRR = \left| \frac{k_D}{k_{CM}} \right| \quad \text{or} \quad CMRR = 20 \log \left| \frac{k_D}{k_{CM}} \right| \quad [dB]$$

A simple version of the differential amplifier is the emitter coupled transistor pair shown below.



Large – signal analysis

If both inputs are zero, then the current I is equally distributed between both transistors:

$$I_{C1} = I_{C2} \approx \frac{1}{2}I \quad \text{if} \quad \beta \gg 1$$

assuming that the transistors are identical and operate at the same temperature.

When a nonzero common – mode signal is applied to both inputs the collector currents do not change and the common – mode gain is zero (for real current sources the current I slightly changes when the voltage drop across the source changes and $k_{CM} \neq 0$).

Notice that the temperature – induced variations of U_{BE} and β for both transistors have an effect similar to the common – mode signal, so they do not effect the output signals U_{01} and U_{02} .

If $U_{I1} > U_{I2}$, then $I_{C1} > I_{C2}$ but $\Delta I_{C1} = -\Delta I_{C2}$

To perform a large – signal analysis for the differential – mode signal let us assume that each transistor is described by the exponential function.

$$I_C = I_{CO} e^{\frac{U_{BE}}{U_T}}$$

in the active region (from the Ebers – Moll model).

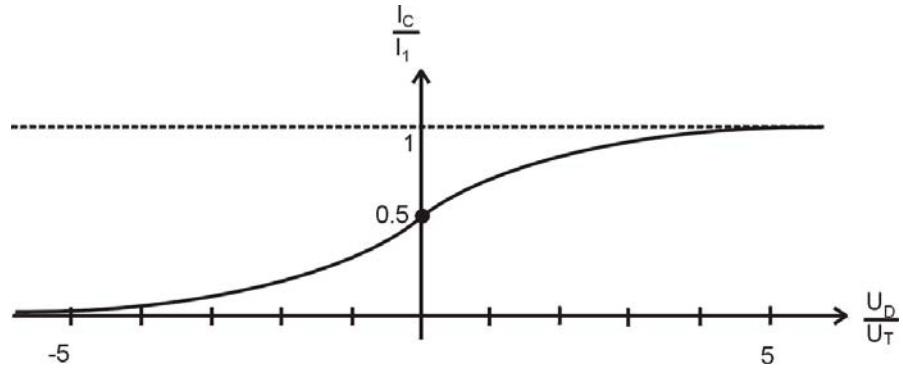
From the set of the following equations:

$$\begin{aligned} I_{C1} &= I_{CO} e^{\frac{U_{BE1}}{U_T}} & I_{C2} &= I_{CO} e^{\frac{U_{BE2}}{U_T}} \\ U_{I1} - U_{I2} &= U_{BE1} - U_{BE2} = U_D \\ I_{C1} + I_{C2} &\equiv I \end{aligned}$$

we get

$$\begin{aligned} I_{C1} &= I \frac{e^{\frac{U_D}{U_T}}}{1 + e^{\frac{U_D}{U_T}}} = \frac{I}{2} \left(1 + \tanh \frac{U_D}{2U_T} \right) \\ I_{C2} &= I \frac{1}{1 + e^{\frac{U_D}{U_T}}} = \frac{I}{2} \left(1 - \tanh \frac{U_D}{2U_T} \right) \end{aligned}$$

These relationships are represented graphically below.



When the differential voltage changes from $-5U_T$ to $+5U_T$ (that is bt 260mV at 300K), the current is completely switched from one transistor to the other.

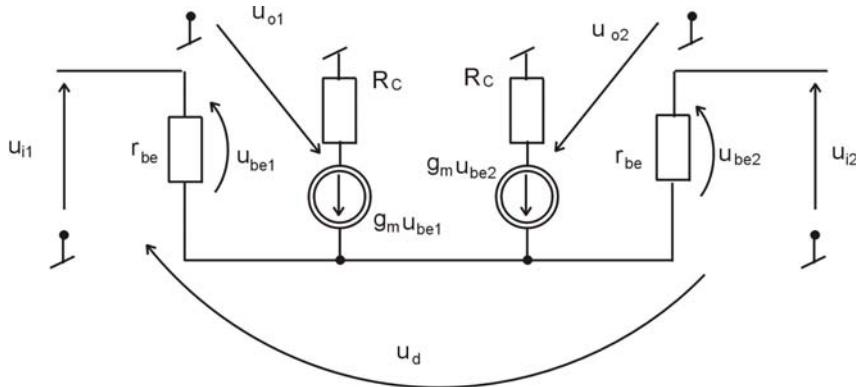
Small – signal analysis

We will perform the small – signal analysis for the differential – mode and common – mode signals separately.

First, let us assume that only a differential signal is applied to both inputs:

$$u_{cm} = \frac{u_{i1} + u_{i2}}{2} = 0 \quad u_{il} = -u_{i2} = \frac{1}{2}u_d$$

Signal level is sufficiently small to assume linearity of both transistors round their operating points. for this part of analysis the current source I can be assumed to be ideal, so the amplifier can be represented by its small signal equivalent circuit shown below (r_{ce} is ignored).



By applying KVL:

$$u_{i1} - u_{be1} + u_{be2} - u_{i2} = 0 \quad \text{and} \quad u_{il} = -u_{i2}$$

and KCL:

$$\frac{u_{be1}}{r_{be}} + g_m u_{be1} + g_m u_{be2} + \frac{u_{be2}}{r_{be}} = 0$$

$$(u_{be1} + u_{be2}) \left(\frac{1}{r_{be}} + g_m \right) = 0 \quad \rightarrow \quad u_{be1} = -u_{be2}$$

we get

$$u_{be1} = -u_{be2} = u_{i1} = -u_{i2} = \frac{1}{2} u_d$$

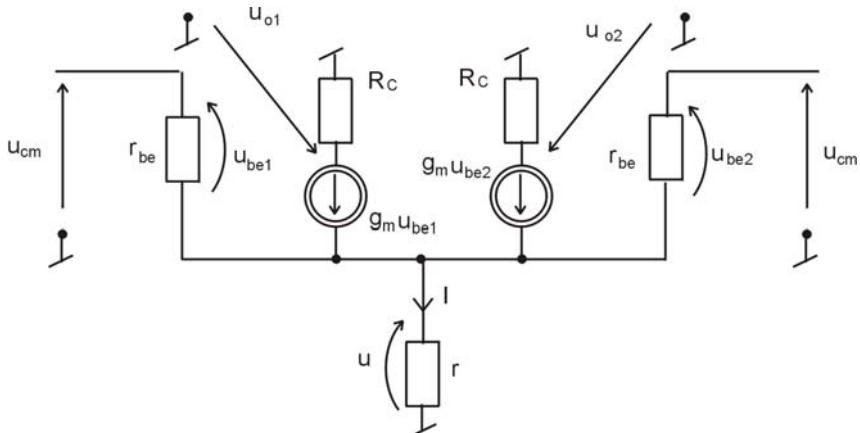
Now we can determine the differential gain for each output:

$$k_{d1} = \frac{u_{o1}}{u_d} = \frac{-g_m u_{be1} R_C}{u_d} = -\frac{1}{2} g_m R_C = k_d$$

$$k_{d2} = \frac{u_{o2}}{u_d} = \frac{-g_m u_{be2} R_C}{u_d} = +\frac{1}{2} g_m R_C = -k_d$$

The voltages at both outputs are half the voltage that would be obtained in a common – emitter amplifier with the same input voltage and they are out of phase with each other.

In the second part of analysis we assume that the same voltage is applied to both inputs. With an ideal current source I and perfectly symmetrical amplifier the common – mode gain would be zero , so we assume that the current source has a finite internal resistance r and the following small – signal equivalent circuit applies:



Applying KCL to the connection point of both emitters:

$$2 \frac{u_{be}}{r_{be}} + 2 g_m u_{be} = \frac{u}{r}$$

$$u = 2 r u_{be} \left(g_m + \frac{1}{r_{be}} \right) \approx 2 r g_m u_{be}$$

and substituting $u_{be} = u_{cm} - u$ leads to the following expression:

$$u = \frac{2rg_m u_{cm}}{1+2rg_m} 1 + 2rg_m \approx u_{cm} \quad \text{because} \quad rg_m \gg 1$$

For the common – mode signal the collector currents (increments) are equal and:

$$i_{c1} = i_{c2} = \frac{1}{2}i = \frac{1}{2}\frac{u}{r}$$

The common – mode gain is:

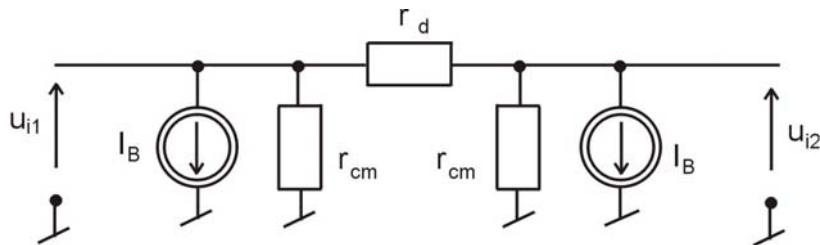
$$k_{cm} = \frac{u_{o1}}{u_{cm}} = \frac{u_{o2}}{u_{cm}} = -R_C i_C \approx -\frac{1}{2} \frac{R_C}{r}$$

The Common – Mode Rejection Ratio, which is a figure of merit for differential amplifiers, can be thus expressed as:

$$CMRR = \left| \frac{k_d}{k_{cm}} \right| \approx g_m r$$

According to the above analysis the common – mode gain results from a finite resistance of the current source. In real amplifiers this gain is additionally increased due to imperfect matching of the transistors or any other non symmetry.

Now let us determine the parameters that refer to the inputs of the differential amplifier. With two inputs there are two values of input resistance for the differential amplifier: the difference – mode resistance, which exists between the two inputs and the common – mode resistance, which exists between each input and ground, as shown in the figure below. Apart from incremental resistances this circuit also includes a DC parameter, the input bias current I_B which is the base current required to keep the transistors in the active region.



the input bias current is:

$$I_B = \frac{I_c}{\beta} = \frac{I}{2\beta}$$

The differential – mode resistance is:

$$r_d = \frac{u_d}{i_d} = \frac{2u_{be}}{i_b} = 2r_{be} = 2\frac{\beta}{g_m}$$

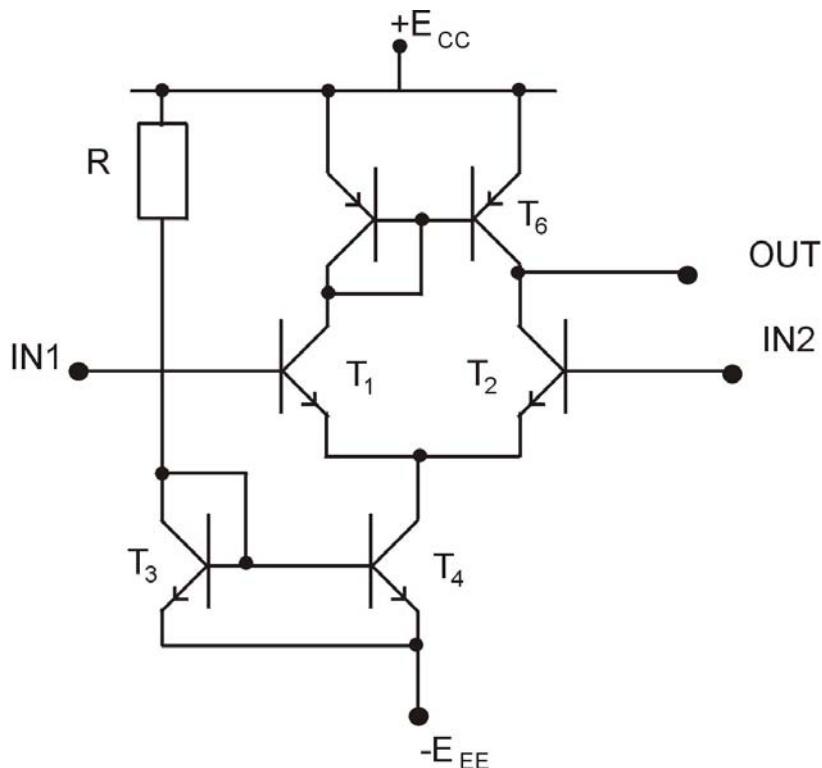
The common – mode resistance for each input is:

$$r_{cm} = \frac{u_{cm}}{i_b} \approx \frac{u_{cm}}{\frac{i_e}{\beta}} = \frac{u_{cm}}{2i_e} 2\beta = 2\beta r$$

The output resistance for each output (see the small – signal equivalent circuit of the amplifier) is:

$$R_{out1} = R_{out2} \approx R_C$$

The expression obtained in the small – signal analysis of the differential amplifier show that to obtain large values of k_d R_C needs to be large. To achieve large CMRR it is necessary to use a large value resistor in the emitter circuit. One of the difficulties in using large value resistors is that the DC voltage drops caused by the collector and emitter currents are large. These problems can be overcome by using currents sources and active loads. A typical differential amplifier in which such structures are used, is shown below.



The current source in the emitter circuit is the current mirror T₃T₄ whose current is set by R. The collector resistors are replaced by a current mirror T₅T₆ (active load). Since T₅ acts as a forward biased diode the resistance in the collector of T₁ is very low and the gain is small. For T₂ the load is the output resistance of T₆ which is very large, and the gain for T₂ is large. Because of this asymmetry in the voltage gains it is usual to operate this circuit with a single – ended output.

The main transistors T_1 and T_2 are often replaced with two Darlington pairs. Thanks to the increased value of equivalent β the differential – mode input resistance considerably increases and the input bias current significantly decreases.

One of the most important applications for the differential amplifier is as the input stage of the integrated operational amplifier.

High efficiency power amplifiers

Bo Berglund, Jan Johansson and Thomas Lejon

The never-ending quest for greater capacity and higher data rates in radio access networks puts increasing demands on the nominal output power from radio base stations. These demands must be met within the existing RBS footprint where cooling capacity and size are limited. At the same time, operators want to cut their radio network operating expenses (OPEX), of which energy consumption is a significant factor.

Linear radio frequency (RF) power amplifiers (PA) play a major role in terms of base station energy consumption and heat dissipation. More efficient base station power amplifiers are thus a crucial factor in the context of mobile system evolution. Reducing the energy consumed by radio base stations will also reduce the environmental impact of the radio access network.

The authors describe the critical aspects and limitations of highly efficient PA technology and indicate where new technology is needed to reach the industry's ambitious goals for greater efficiency.

Background

Linear RF power amplifiers consume large amounts of energy, dissipate heat, and take up space in base stations. Significantly more efficient PA technology will be instrumental to the evolution of mobile systems. The main requirements for future power amplifier technology are

- high linearity, to satisfy higher-order modulation schemes;
- greater average output power levels;

- broader operating bandwidths (more than twice today's typical 20MHz);
- reduced OPEX by decreasing RBS energy consumption; and
- reduced environmental impact by decreasing radio network energy consumption.

New technologies with greater PA drain efficiency were introduced in the past 12 months, significantly increasing radio unit (RU) power efficiency. Notwithstanding, new technologies with even greater drain efficiencies are needed to meet the challenges of the

next three to five years. (Drain efficiency is the ratio of delivered output power divided by applied DC power in the PA).¹

The main objective of PA research is to increase efficiency while maintaining linearity and broadening the operating bandwidth. Present-day power amplifiers with Doherty efficiency-enhancement technology will continue to evolve in coming years, but to increase efficiency significantly, other technologies must also be considered. Switch-mode PA technology has been identified as a way of achieving high PA efficiency. This technology is being used successfully with pulse width modulation (PWM) for audio and digital subscriber line (DSL) driver applications. In this role, switch-mode PAs are 80% to 90% efficient but their application is limited to operating frequencies of around a few megahertz. The challenge is thus to apply the technology to operating frequencies for mobile systems in the 1 - 4GHz range.

Because straightforward frequency scaling is not feasible due to the physical properties of the components, suitable methods must be found to mitigate the fundamental limitations of high-frequency operation. The major areas of research in this area are

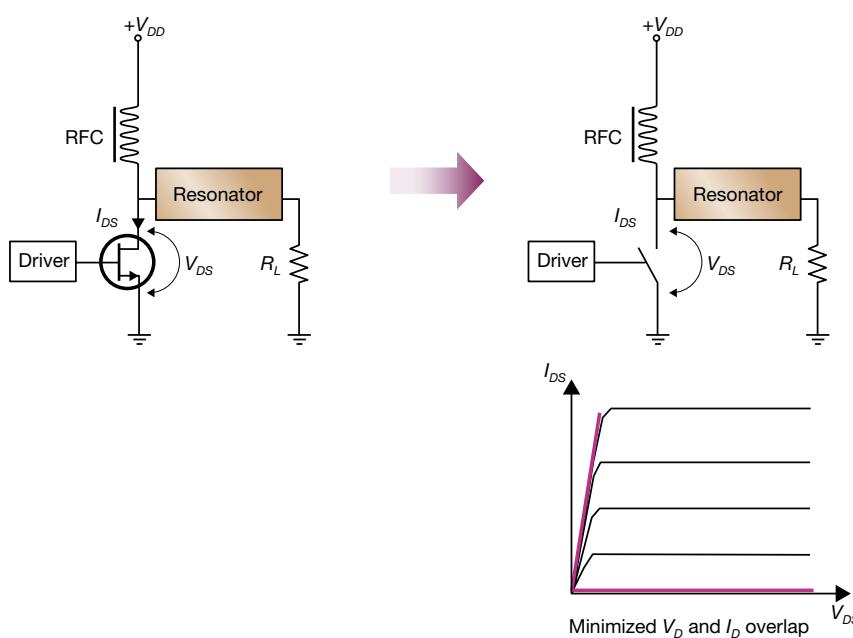
- switch-mode PA concepts with potentially high drain efficiency (greater than 60%);
- architectures that retain efficiency over a large, dynamic range of output power (typically 20dB) while meeting linearity requirements; and
- RF power transistor technology with performance that is suitable for switch-mode applications in the 1 - 4GHz frequency range.

Switch-mode PA technology

The main idea behind switch-mode PA technology is to operate the transistor in saturation, so that either voltage or current, depending on amplifier class, is switched on and off. Figure 1 shows a simplified block diagram of a switch-mode power amplifier. For our discussion of fundamental properties, the transistor can be replaced by a switch. When the switch is open, only voltage is present over the transistor. When closed, current flows through it. Since there is no overlap in time between voltage and current, power is not dissipated and one obtains 100% theoretical efficiency. In reality, a transistor is not a perfect switch (see trajectories, Figure 1) and overlap does, in fact, limit efficiency.

For comparison, Figure 2 illustrates a

Figure 1
Simplified view of switch-mode PA operation.



Class-AB power amplifier. Here, the transistor is biased to an operating point and follows the indicated load line, giving rise to power dissipation and loss.

In the switch-mode power amplifier, an output resonator helps shape the waveform by blocking harmonic components of the voltage and current – that is, it keeps these components from reaching the load. Consequently, only fundamental current is passed to the load and only fundamental voltage is generated over the resonator. A flywheel effect is created generating sinusoidal voltage and current in the load. The two necessary conditions for generating a single tone with 100% efficiency in the load are

- zero overlap between voltage over the transistor channel and current through the channel; and
- blocking of harmonic currents to the load. Figure 3 compares the simulated efficiency of almost ideal Class-AB and switch-mode Class-D amplifiers. In the Class-AB amplifier, there is significant overlap in time of voltage and current, which degrades efficiency. In real amplifiers, switching and component losses can significantly degrade efficiency. Therefore, a critical task when designing switch-mode power amplifiers is to minimize these losses. Examples of losses are

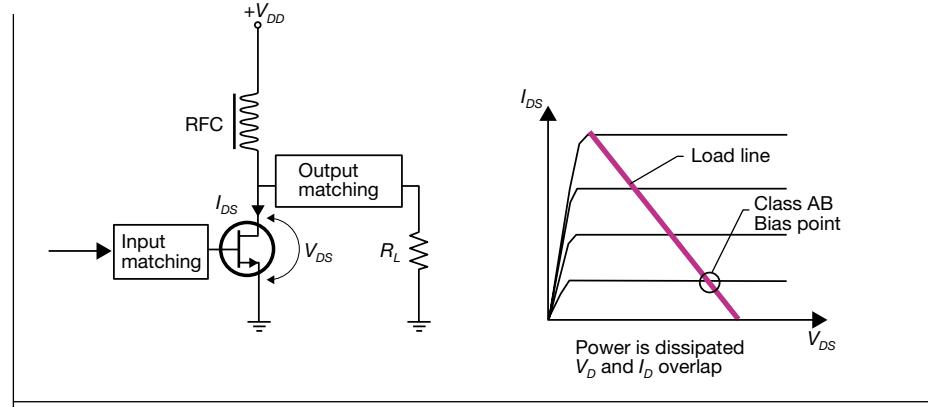


Figure 2
Class-AB power amplifier operation.

- parasitic capacitors, such as C_{ds} (drain to source capacitance). Parasitic capacitors cause loss when voltage is switched;
- R-ON (the drain-to-source resistance when the transistor is conducting);
- non-zero transition time. The square waveform requires a fast transistor (high f_t). If the switching frequency is close to f_t , then loss occurs due to overlap between voltage and current in the transistor; and
- PA implementation losses, including driv-

er power consumption, output circulator, and filtering.

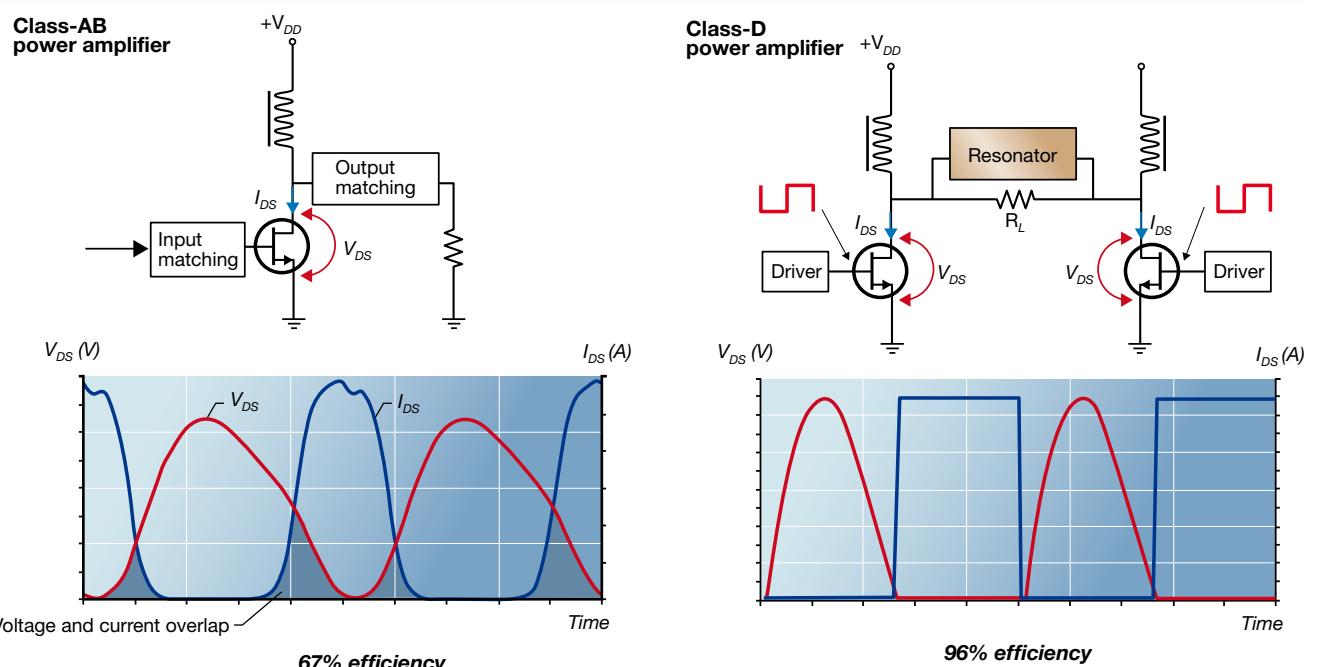
Taking all these loss mechanisms into account for PAs designed to operate in the gigahertz range, one can anticipate PA efficiencies of around 60% to 70% at best.

Switch-mode PA architectures

Numerous switch-mode PA classes of operation have been invented over the years. The main differences between these classes are topologies,

Figure 3

Left: Current and voltage waveforms in Class-AB power amplifiers. Right: Current and voltage waveforms in Class-D power amplifiers.



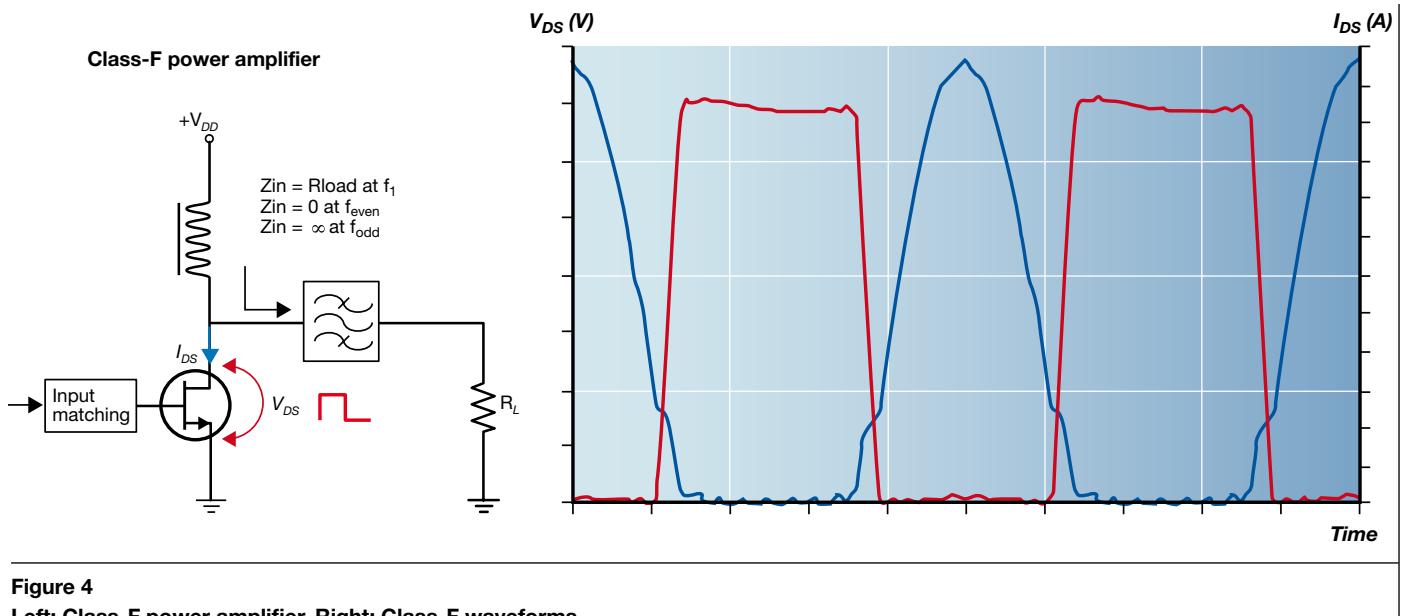
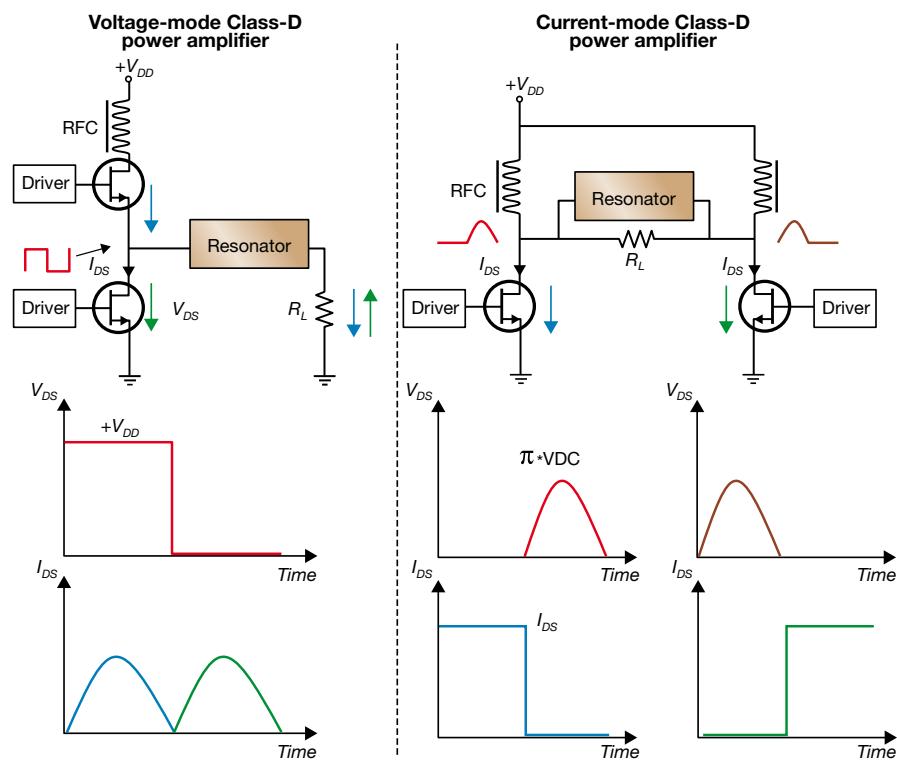


Figure 4

Left: Class-F power amplifier. **Right:** Class-F waveforms.

Figure 5
Voltage- and current-mode Class-D power amplifiers and their waveforms.



waveform shaping, and method of analysis.

Below follows a brief description of some common switch-mode classes of operation.

Class-F

Class-F power amplifiers use multiple resonators to control the harmonic contents of the drain voltage and current. In an ideal Class-F power amplifier, the drain voltage is square wave; current waveforms are half-sinusoidal (Figure 4).

The main obstacle to the Class-F design is the realization of harmonic terminations at high frequencies. Practical designs are typically limited to terminating the third harmonic, which limits the maximum theoretical efficiency to 75%. For a 2.2GHz design, this means the terminations must operate at 6.6GHz.

Class-D

There are two main realizations of Class-D power amplifiers:

- voltage-mode Class-D with serial resonator circuit; and
- current-mode Class-D using a parallel resonator circuit.

Each type has a topology with two transistors.

Figure 5 (left) shows a classic voltage-mode Class-D power amplifier where voltage is switched and the output resonator forces

the current to be sinusoidal. Voltage-mode Class-D power amplifiers and PWM technology make a highly efficient combination for audio applications. But because transistor output capacitances quickly become a dominant loss factor at higher frequencies, it is difficult to achieve the same good efficiency in the gigahertz frequency range.

In a current-mode Class-D power amplifier, current is switched. The short-circuit harmonic termination of the output resonator forces the voltage to be sinusoidal. The amplifier has an interesting balanced topology: both its transistors are grounded, and their output capacitances can be used in the output filter. The half-wave rectified sinusoidal waveform is created by the flywheel effect of the output network and the balanced configuration. In a balanced configuration, one needs only short-circuit the output at odd harmonics. Therefore this amplifier shows promise as a highly efficient performer at high power in the gigahertz range. A main drawback is high peak voltage, which calls for transistors with high breakdown voltage.

Class-E

The Class-E power amplifier, which is an interesting compromise between a linear Class-AB power amplifier and a switched power amplifier, has zero overlap between voltage and current over and through the transistor, giving 100% theoretical efficiency and potentially robust performance (Figure 6).²

The output network of a Class-E power amplifier starts with a shunt capacitor that absorbs the output capacitance of the transistor. Current passes through the capacitor when the transistor channel is closed. Compare I_s and I_c with I (Figure 6). The inductance and capacitance (LC) resonator ensures that only the fundamental frequency current can flow in the output network to load, giving a single tone in the load. The flywheel effect of the LC network drives the current through either the switch or the capacitor.

The waveforms of the Class-E power amplifiers are analog in shape without the ideal pulse-shaped form presented by other modes of operation. The Class-E mode can thus be supported by a transistor with slower switching characteristics and is better suited to high-frequency operation. As with Class-D mode, high peak voltage is a drawback.

Switch-mode PA performance

Research on switched PAs has produced promising results: Class-D, Class-E and

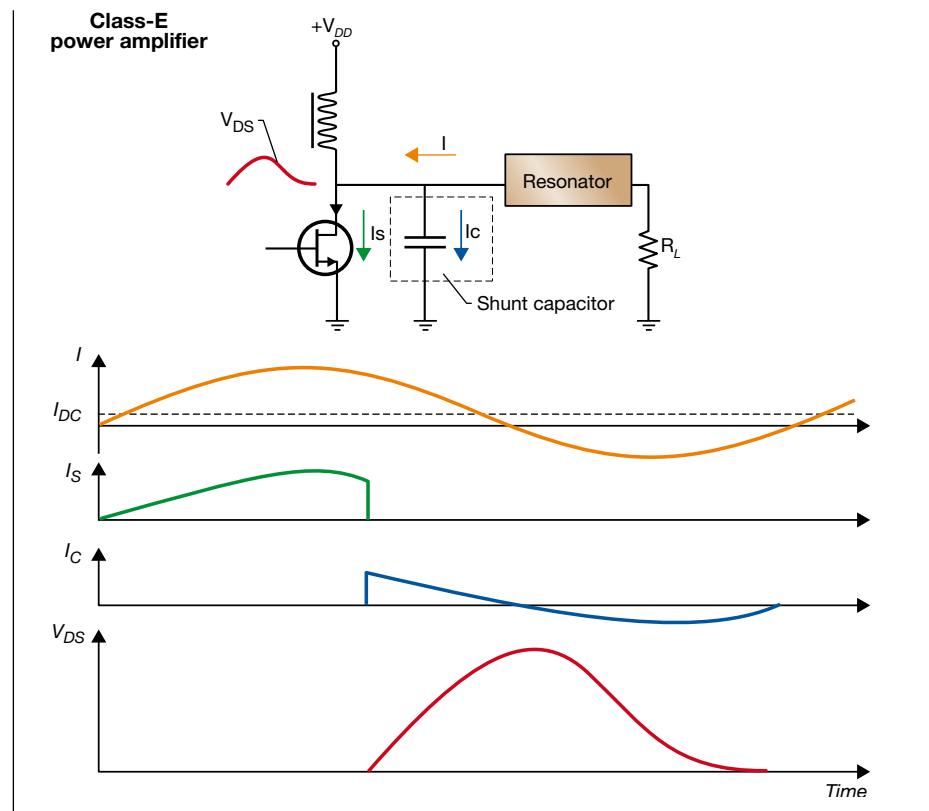


Figure 6
Class-E power amplifier and waveforms.

Class-F power amplifiers have reached high peak drain efficiencies in the gigahertz range at high power levels. In particular, note the results of Class-D and Class-E power amplifiers using gallium nitride (GaN) transistors (Table 1).

Modulated signals with high PAPR

The results in Table 1 show that it is possible to obtain high peak drain efficiencies even at microwave frequencies. The remaining challenge is to maintain high efficiency for signals with high peak-to-average power ratio (PAPR). To benefit from increased peak effi-

ciency for amplitude-modulated signals with high PAPR (typically 6 to 9dB for WCDMA and OFDM signals), the amplifier must work in saturation even when signal amplitude is backed off. Moreover, given that switch-mode amplifiers are inherently non-linear, the transmitter architecture must also provide a means of linearizing the amplifier. Several PA modulation technologies have the potential to work with high PAPR signals⁸:

- drain modulation (the drain voltage follows the signal envelope, keeping the power amplifier operating at maximum efficiency);

TABLE 1. PERFORMANCE OF SWITCH-MODE PAs.

Class	F(GHz)	Peak drain efficiency (%)	P _{out} (W)	Gain	Transistor technology	Reference
F-1	1	77.8	12.4	12.9	LDMOS	3
E	1	73	7.9	10	LDMOS	4
E	2	85(PAE)	10	12	GaN	5
CMCD	1	71	20.3	15.1	LDMOS	6
CMCD	1	78	51.1	10.6	GaN	7

TERMS AND ABBREVIATIONS

AlGaN	Aluminum GaN	OFDM	Orthogonal frequency-division multiplexing
CMCD	Current-mode Class-D	OPEX	Operating expenditure
CMOS	Complementary metal oxide semiconductor	PA	Power amplifier
DC	Direct current	PAE	Power-added efficiency
DSL	Digital subscriber line	PAPR	Peak-to-average power ratio
F_{max}	Frequency at which transistor unilateral power gain has rolled off to 0 dB	PWM	Pulsewidth modulation
F_t	Frequency at which the magnitude of the transistor short-circuit current gain, h_{21} , has rolled off to 0 dB	RBS	Radio base station
GaN	Gallium nitride	RF	Radio frequency
HEMT	High electron mobility	RFC	RF choke
LC	Lumped component	RL	Load resistance
LDMOS	Lateral double-diffused metal-oxide semiconductor	RU	Radio unit
		Si	Silicon
		V_{dd}	Drain voltage
		V_{ds}	Drain to source voltage
		WCDMA	Wideband CDMA

- load modulation (the load impedance is adapted to the signal envelope to keep the amplifier operating at peak efficiency); and
- RF pulse width modulation (this technique is mainly used at low frequencies but has the potential to work over a broad bandwidth with high efficiency).

More complex modulation and linearization schemes that require advanced signal processing might actually remove efficiency gains. Further research is needed to find the best compromise between efficiency, linear-

ity, and bandwidth throughout the entire transmitter chain.

RF power transistor technology

Silicon LDMOS

LDMOS transistors have been the dominating technology for high-power RF amplifiers for almost ten years and it is difficult to see any other technology competing in Class-AB applications at frequencies below 3GHz. However, for higher frequencies and emerging switch-mode architectures, fundamental

limitations, such as comparatively low f_t/f_{max} and high, lossy parasitic output capacitance, call for alternative technologies.

Gallium nitride

A comparison of intrinsic material properties shows that HEMTs (high electron mobility transistors) which utilize AlGaN/GaN heterostructures clearly stand out as the most promising of emerging technologies.

The energy gap of GaN is three times that of silicon (Si), resulting in reduced performance degradation at high temperatures. Similarly, breakdown at a six-fold electric field and two-and-a-half-fold carrier saturation velocity enable much greater power densities, resulting in the same output power capability at a much higher impedance level.

To reap the full benefit, devices must be optimized for a particular application. Therefore, the performance of GaN transistors in future switch-mode architectures might depend on what manufacturers see as the main application for these devices.

Compared with silicon LDMOS, GaN technology is still immature, hampered by basic manufacturability and reliability issues, and is far from being competitive in terms of cost.

Conclusion

The next generation of power amplifiers for mobile systems requires extensive research and component development to meet demanding requirements for efficiency, linearity output power, and bandwidth. Needed are advanced PA architectures, used in combination with complex signal processing and RF power transistors with very low parasitics.

Switch-mode PA architectures are promising candidates, provided RF power transistors with low parasitics and high breakdown voltage can be developed.

GaN transistors show potential for use in switch-mode applications but are still a long way from being a reliable and cost-effective component for commercial applications.

Silicon LDMOS technology and accompanying efficiency-enhancement technologies will continue to improve, mainly through the reduction of parasitic capacitors and further development of PA architectures, possibly to a point where the benefits of complex switch-mode architectures can be called into question. Either way, researchers anticipate significant increases in PA efficiency in coming years.

REFERENCES

1. Berglund, B., Englund, M. and Lundstedt, J: Third design release of Ericsson's WCDMA macro radio base stations. Ericsson Review, Vol. 82(2005):2, pp. 70-81
2. Cripps, S.: "RF Power Amplifiers for Wireless Communications", ISBN: 0-89006-989-1, 1999 ARTECH HOUSE, INC.
3. F. Lépine, A. Ådahl, H. Zirath, "L-band LDMOS power amplifier based on inverse Class-F architecture", IEEE Transactions on Microwave Theory and Techniques, Volume 53, no. 6, Part 2, pp. 2007 – 2012, June 2005
4. A. Ådahl, H. Zirath, "An 1 GHz Class-E LDMOS power amplifier," 33rd European Microwave Conference, Volume 1, pp. 285-288, Oct 2003
5. W.L. Pribble, J.M. Milligan, and R.S. Pengelly, "High Efficiency Class-E Amplifier Utilizing GaN HEMT Technology," 2006 IEEE Topical Workshop on Power Amplifiers for Wireless Communications, Technical Digest, Session 3, January 2006
6. H. Mashad Nemati, "Design, Implementation and Evaluation of a Current Mode Class-D Power Amplifier", Master's Thesis 2006, Chalmers University of technology
7. Ulf Gustavsson, "Design of an inverse Class-D amplifier using GaN HEMT technology, Master's Thesis 2006, Örebro University
8. F. Raab, P. Asbeck, S. Cripps, P.B. Kennington, Z. Popovic, N. Pothecary, J. F. Sevic and N. Sokal "RF and Microwave Power Amplifier and Transmitter Technologies – Part 5", High Frequency Electronics, May 2003

PA-Classes

AMPLIFIER CLASSES (adapted from Wikipedia)

Amplifier circuits are classified as A, B, AB and C for analog design, and class D, E, and F for high efficiency switching design. For the analog classes, each class defines what proportion of the input signal cycle (called the angle of flow) is used to actually switch on the amplifying device:

Class A

100% of the input signal is used (conduction angle $a = 360^\circ$ or 2π)

Class A amplifiers amplify over the whole of the input cycle such that the output signal is an exact scaled-up replica of the input with no clipping. They are not very efficient — a theoretical maximum of 50% is obtainable, but for small signals, this waste of power is still extremely small, and can be easily tolerated. Only when we need to create output powers with appreciable levels of voltage and current does Class A become problematic. In a Class A circuit, the amplifying element is biased such that the device is always conducting to some extent, and is operated over the most linear portion of its characteristic curve (known as its transfer characteristic or transconductance curve). Because the device is always conducting, even if there is no input at all, power is wasted. This is the reason for its inefficiency

Class AB

more than 50% but less than 100% is used. (181° to 359° , $\pi < a < 2\pi$)

- Class AB1 applies to tube or transistor amplifiers in class AB where the grid or base is more negatively biased than it is in class A.
- Class AB2 applies to tube or transistor amplifiers in class AB where the grid or base is often more negatively biased than in AB1, and the input signal is often larger. When the drive is high enough to make the grid or the base more positive, the grid or base current will increase. It is possible depending on the level of the signal input for the amplifier to move from class AB1 to AB2.

Class B

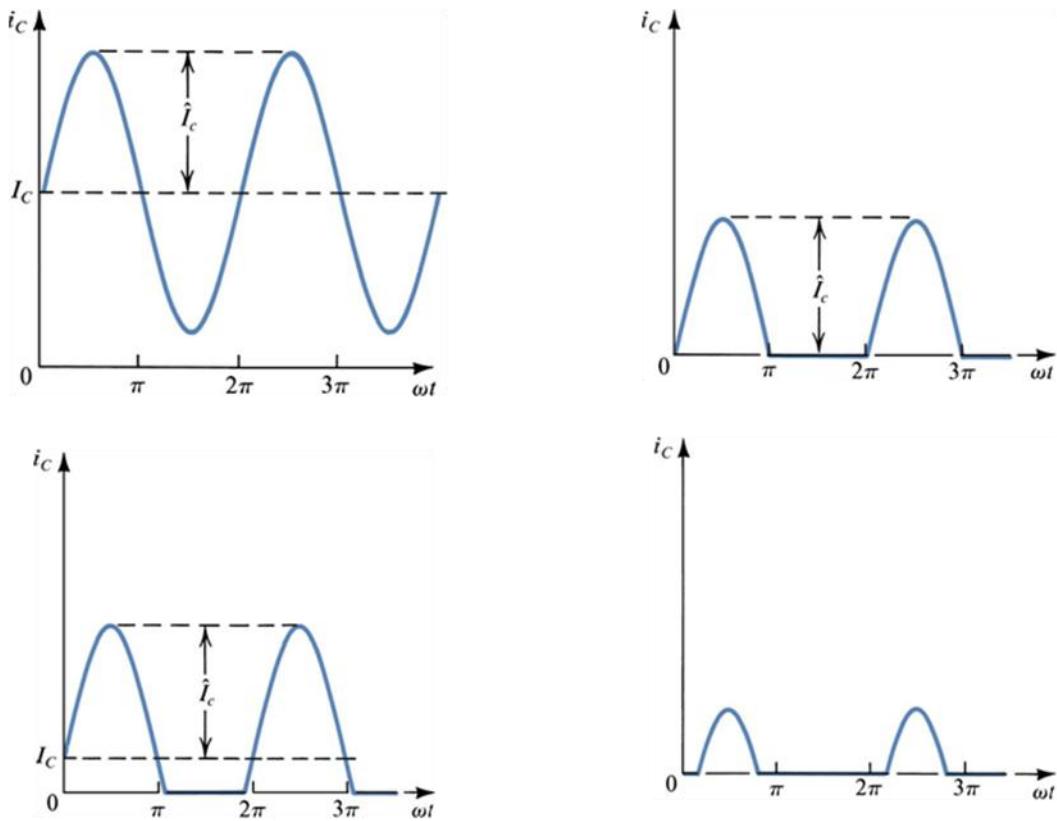
50% of the input signal is used ($a = 180^\circ$ or π)

Class B amplifiers only amplify half of the input wave cycle (maximum theoretical efficiency of 78.5%). This is because the amplifying element is switched off altogether half of the time, and so cannot dissipate power. A single Class B element is rarely found in practice, though it can be used in RF power amplifiers where the distortion is unimportant. However Class C is more commonly used for this.

Class C

less than 50% is used (0° to 179° , $a < \pi$)

Class C amplifiers conduct less than 50% of the input signal and the distortion at the output is high, but efficiencies of up to 90% can be reached. Some applications can tolerate the distortion, such as megaphones. A much more common application for Class C amplifiers is in RF transmitters, where the distortion can be vastly reduced by using tuned loads on the amplifier stage. The input signal is used to roughly switch the amplifying device on and off, which causes pulses of current to flow through a tuned circuit. The tuned circuit will only resonate at particular frequencies, and so the unwanted frequencies are dramatically suppressed, and the wanted full signal (sine wave) will be abstracted by the tuned load. Provided the transmitter is not required to operate over a very wide band of frequencies, this arrangement works extremely well. Other residual harmonics can be removed using a filter.



Collector current waveforms for transistors operating in (a) class A, (b) class B, (c) class AB, and (d) class C amplifier stages.

Class D

A class D amplifier is a power amplifier where all power devices are operated in on/off mode. Output stages such as those used in pulse generators are examples of class D amplifiers. Mostly though, the term applies to devices intended to reproduce signals with a bandwidth well below the switching frequency. These amplifiers use

pulse width modulation, pulse density modulation (sometimes referred to as pulse frequency modulation) or more advanced form of modulation such as Sigma delta. The input signal is converted to a sequence of pulses whose averaged value is directly proportional to the amplitude of the signal at that time. The frequency of the pulses is typically ten or more times the highest frequency of interest in the input signal. The output of such an amplifier contains unwanted spectral components (i.e.. the pulse frequency and its harmonics) that must be removed by a passive filter. The resulting filtered signal is then an amplified replica of the input.

The main advantage of a class D amplifier is power efficiency. Because the output pulses have fixed amplitude, the switching elements (transistors) are switched either on or off, rather than operated in linear mode. This means that very little power is dissipated by the transistors except during the very short interval between the on and off states. The wasted power is low because the instantaneous power dissipated in the transistor is the product of voltage and current, and one or the other is almost always close to zero. The lower losses permit the use of a smaller heat sink while the power supply requirements are lessened too.

Class D amplifiers were widely used to control small DC motors, but they are now also used as audio amplifiers, with some extra circuitry to allow analogue to be converted to a much higher frequency pulse width modulated signal.

Class E

The class E/F amplifier is a highly efficient switching power amplifier, typically used at such high frequencies that the switching time becomes comparable to the duty time. As said in the class-D amplifier the transistor is connected via a serial-LC-circuit to the load, and connected via a large L (inductivity) to the supply voltage. The supply voltage is connected to ground via a large capacitor to prevent any RF-signals leaking into the supply. The class-E amplifier adds a C between the transistor and ground and uses a defined L to connect to the supply voltage.

All previous designs use sharp edges to minimize the overlap. Class E uses a significant amount of second harmonic voltage. The second harmonic can be used to reduce the overlap with edges with finite sharpness. In reality the impedance is mostly reactive and the only reason for it is that class E is a class F amplifier with a very simplified load network and thus has to deal with imperfections. Note how in many amateur simulations of class E amplifiers sharp current edges are assumed nullifying the very motivation for class E and measurements near the transit frequency of the transistors show very symmetric curves, which look much similar to class F simulations.

Class F

The class F-amplifier takes the finite on resistance into account and tries to make the current touch the bottom at zero. This means the voltage and the current at the

transistor are symmetric with respect to time. The Fourier Transform allows an elegant formulation to generate the complicated LC-networks. It says that the first harmonic is passed into the load, all even harmonics are shorted and all higher odd harmonics are open.

In push-pull amplifiers and in CMOS the even harmonics of both transistors just cancel. Experiment tells that a square wave can be generated by those amplifiers and math tells that square wave do consist of odd harmonics only.

In a class D amplifier the output filter blocks all harmonics, which means the harmonics see an open load. So even small harmonic currents suffice to generate a voltage square wave. The current is in phase with the voltage applied to filter, but the voltage across the transistors is out of phase. Therefore there is a minimal overlap between current through the transistors and voltage across the transistors. The sharper the edges the lower the overlap.

While class D sees the transistors and the load as separate modules, the class F admits imperfections like the parasitics of the transistor and tries to optimize the global system to have a high impedance at the harmonics. Of course there has to be a finite voltage across the transistor to push the current across the on state resistance. Because the combined current through both transistors is mostly in the first harmonic it looks like a sine. That means that in the middle of the square the maximum of current has to flow, so it may make sense to have a dip in the square or in other words to allow some over swing of the voltage square wave. A class F load network by definition has to transmit below a cut off frequency and to reflect above. Any frequency lying below the cut off and having its second harmonic above the cut off can be amplified, that is an octave bandwidth. On the other hand a LC series circuit with a large L and a tunable C may be simpler to implement. By reducing the duty cycle below 0.5, the output amplitude can be modulated. The voltage square waveform will degrade, but any overheating is compensated by the lower overall power flowing. Any load mismatch behind the filter can only act on the first harmonic current waveform, clearly only a purely resistive load makes sense, then the lower the resistance the higher the current. Class F can be driven by sine or by a square wave, for a sine the input can be tuned by an L to increase gain. If class F is implemented with a single transistor the filter is complicated to short the even harmonics.

Notes

The main concept used in amplification is to model the active switching device, such as a transistor or MOSFET, as a linear combination of two parts: (1) a (theoretical) "perfect" switching element, and (2) a complex network of parasitic elements attached to it (capacitors, inductors and resistors). After the decomposition, it becomes trivial to eliminate the losses of each part:

(1) The "perfect" switching element should be switched ON during zero-voltage crossing, and should be switched OFF during zero-current crossing. Thus the switching element either conducts current, or has some non-zero voltage on it, but

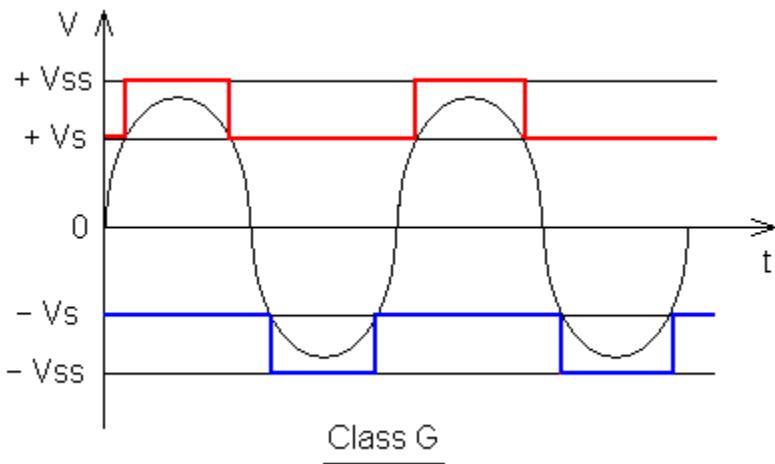
never both at the same time. Because the dissipated power is equal to current x voltage, it becomes zero. This can be arranged by adjusting the phase (capacitor) and DC bias (resistor) of the signal going into the transistor input.

(2) The imaginary part of the impedance of the parasitic elements can be tuned, one by one, by matching them to another passive element with the complex conjugate impedance, thus leaving only the real part of the complex impedance.

In theory, the only remaining loss is the real part of the impedance of the parasitic elements in the system, which cannot be avoided. This class of amplifier is unique to radio frequency ranges, where the amplifier analysis is usually done in the frequency domain and not in the voltage/current domain. This class is further divided to subclasses depending on which harmonics of the signal are taken into account during zero-voltage switching (ZVS) and zero-current switching (ZCS), with names such as Class E/F2,odd; inverse Class (F^{-1}); and so on. It is still an active area of research and new variants appear from time to time, usually with the letters "E" and "F" somewhere in class name.

Class G

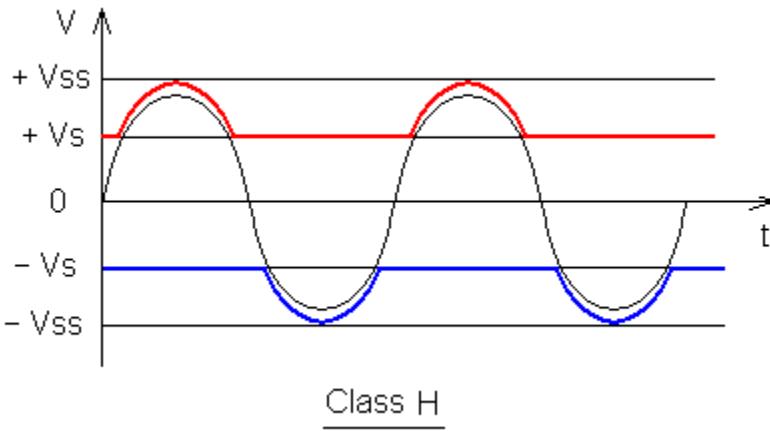
Class G amplifiers are a more efficient version of class AB amplifiers, which use "rail switching" to decrease power consumption and increase efficiency. The amplifier has several power rails at different voltages, and switches between rails as the signal output approaches each. Thus the amp increases efficiency by reducing the "wasted" power at the output transistors.



Class H

A Class H amplifier takes the idea of Class G one step further creating an infinite number of supply rails. This is done by modulating the supply rails so that the rails are only a few volts larger than the output signal at any given time. The output stage operates at its maximum efficiency all the time. Switched mode power supplies can be used to create the tracking rails. Significant efficiency gains can be achieved but

with the drawback of more complicated supply design and reduced THD performance.



Other classes

A hybrid configuration that is receiving new attention is the Doherty amplifier consisting of a class B main stage in parallel with a class C auxiliary (or "peaking") stage. The input signal is split evenly to drive the two amplifiers, and a Doherty combiner is used at the output to recombine the components. During periods of low signal level, the class B amplifier efficiently operates on the signal and the class C amplifier is inactive and consumes no power. During high signal peaks the class B amplifier saturates and the class C amplifier kicks in. The effect is that the Doherty amplifier maintains efficiency comparable to class B saturated performance even when operated at up to 6 dB output backoff. Interest in the Doherty configuration has been revived by cellular telephone and wireless internet applications where the sum of several constant envelope users creates an aggregate AM result. The main challenge of the Doherty amplifier is in aligning the two stages and getting the class C amplifier to turn on and off quickly.

Several audio amplifier manufacturers have started "inventing" new classes as a way to differentiate themselves. These class names usually do not reflect any revolutionary amplification technique, and are used mostly for marketing purposes. For example, Crown's K and I-Tech Series as well as several other models utilize Crown's patented **Class I** technology. Lab Gruppen use a form of class D amplifier called **class TD** or Tracked Class D which tracks the waveform to more accurately amplify it without the drawbacks of traditional class D amplifiers.

Class T is a trademark of TriPath company, which manufactures audio amplifier IC's. This new class "T" is a revision of the common class D amplifier, but with changes to ensure fidelity over the full audio spectrum, unlike traditional class D designs. It operates at a frequency of 650kHz, with a proprietary modulator.

Class Z is a trademark of Zetex semiconductor and is a direct digital feedback technology.



Institute of Electronics
National Chiao Tung Univ.

Power Amplifier Design

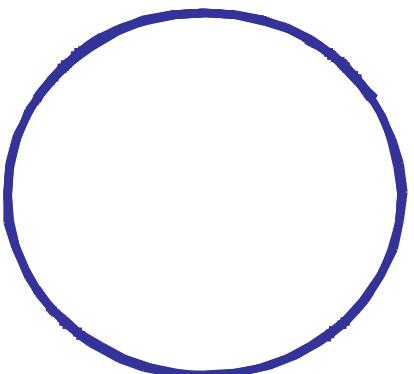
Design considerations

Power amplifier classes

Power amplifier examples

Linear/Nonlinear PA?

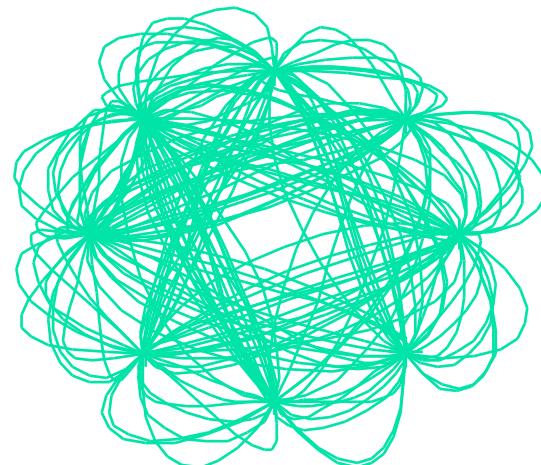
Constant envelop
modulation



GMSK
FSK

Nonlinear PA
High Efficiency

Nonconstant envelop
modulation

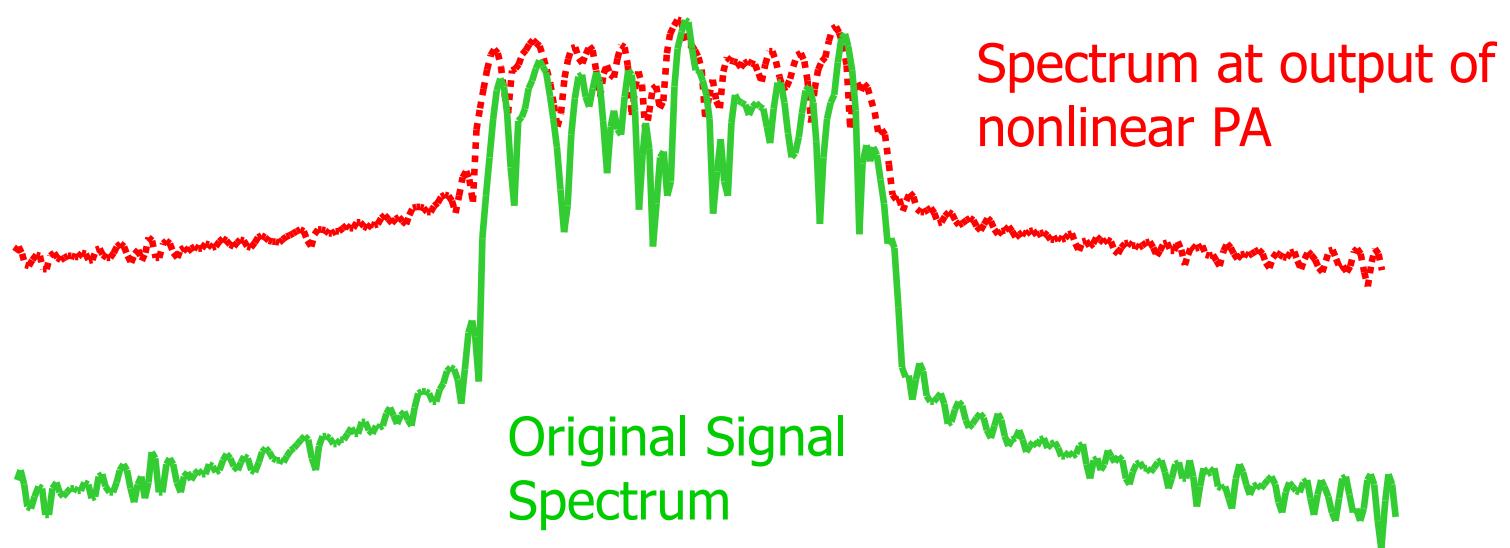


BPSK
QPSK
QAM

Linear PA
Low Efficiency

Spectral Regrowth

- Effect of nonlinear PA on nonconstant envelop signal



Power Amplifier Efficiency

- For ideal PA : $\eta = \frac{P_{out}}{P_{total}} = 1$
- The Drain Efficiency $\eta_{Drain} = \frac{P_{RFout}}{P_{DC}}$
- The Power Added Efficiency(PAE):
$$PAE = \frac{P_{RFout} - P_{RFin}}{P_{DC}}$$
- The overall efficiency:
$$\eta = \frac{P_{RFout}}{P_{DC} + P_{RFin}}$$



Basic Amplification:

Use RFC (RF Chock) to in a common source stage to drive the load

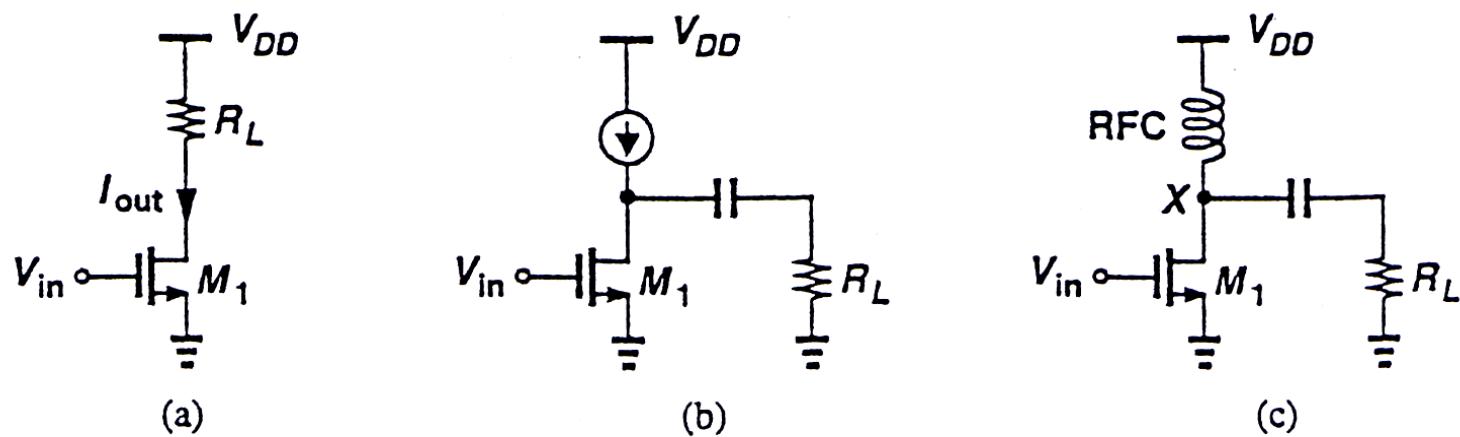
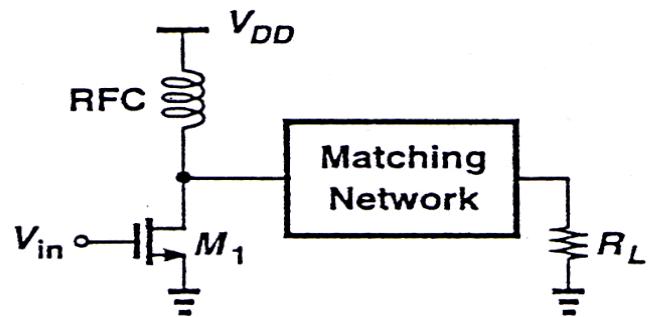
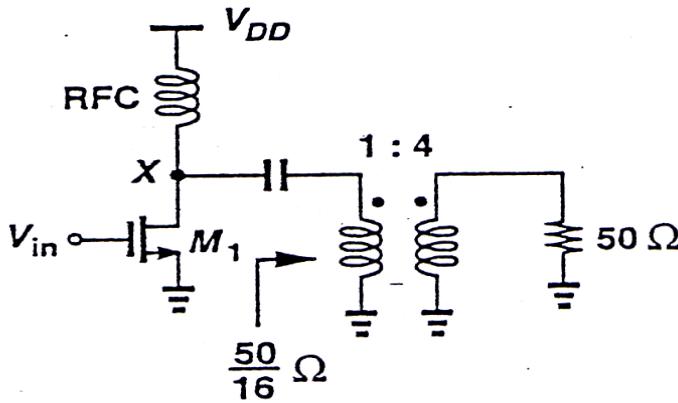


Figure 9.1 Common-source stages with different type of load connections.

Matching



(a)



(b)

Figure 9.2 (a) Matching network as voltage amplifier, (b) use of a transformer as a matching network.



Typical PA Performance

TABLE 9.1 Typical PA performance.

Output Power	+20 to +30 dBm
Efficiency	30% to 60%
IMD	-30 dBc
Supply Voltage	3.8 to 5.8 V
Gain	20 to 30 dB
Output Spurs and Harmonics	-50 to -70 dBc
Power Control	On-Off or 1-dB Steps
Stability Factor	> 1

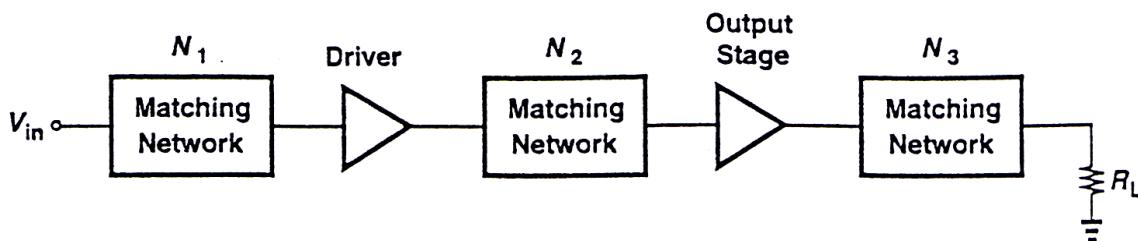


Figure 9.22 Typical PA system.

Linear and Nonlinear PA

- Linear/Nonlinear distinction
 - The fraction of the RF cycle for which the transistor conducts.

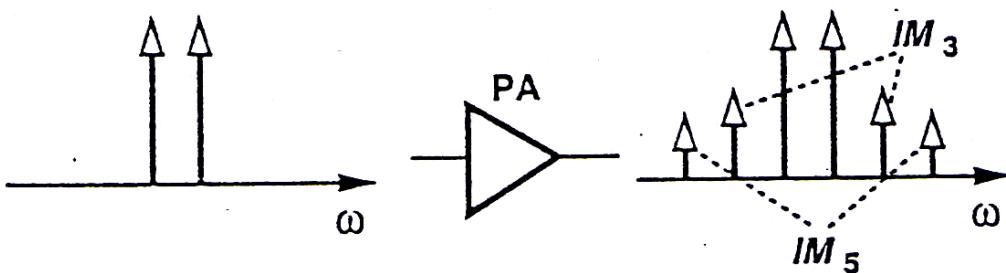
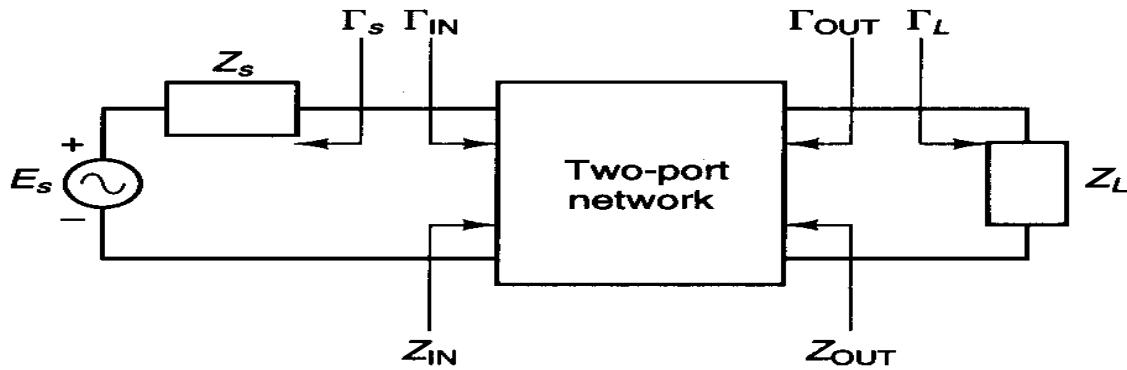


Figure 9.3 Two-tone test of a power amplifier.

Stability Consideration



$$|\Gamma_s| < 1, |\Gamma_L| < 1,$$

$$|\Gamma_{in}| = \left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| < 1,$$

$$|\Gamma_{Out}| = \left| S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \right| < 1$$

Unconditionally Stable :

$$K > 1, |\Delta| < 1, \text{ where}$$

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|},$$

$$\Delta = S_{11}S_{22} - S_{12}S_{21}$$

Ps: Stable circle on Smith chart is the general tool

Operating Power Gain

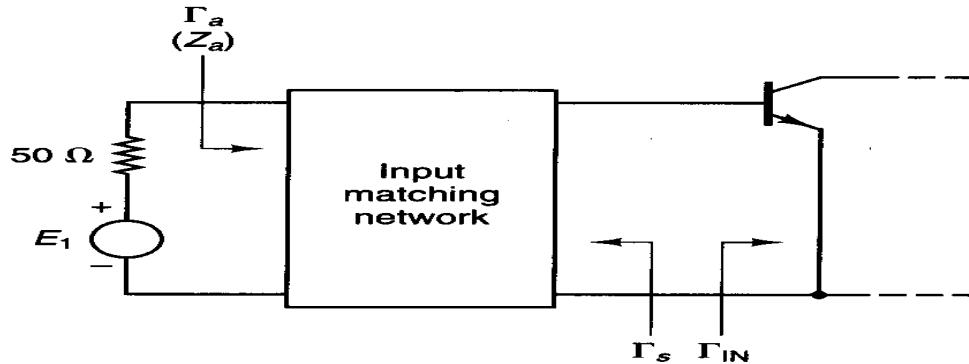
$$G_P = \frac{P_L}{P_{IN}} = \frac{\text{power delivered to load}}{\text{power input to network}},$$

$$G_P = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}, \text{ substitute } \Gamma_{IN},$$

$$G_P = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2)}{(1 - \left| \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} \right|)^2 |1 - S_{22}\Gamma_L|^2} = |S_{21}|^2 g_p$$

- Given the power gain, drawing the power gain circle, and select Γ_L in the stable region.
- Calculate Γ_{In} , determine if a conjugate match is in the stable region. If it's not stable, we can choose Γ_L arbitrarily, or according VSWR.

Constant VSWR Circle



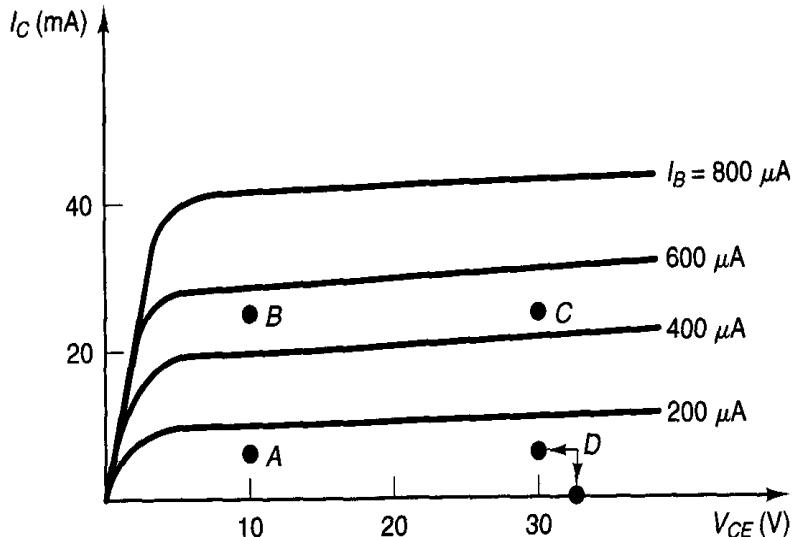
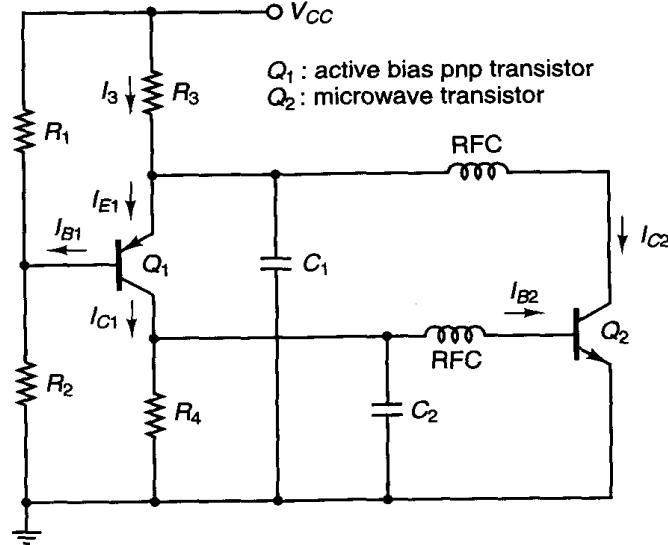
$$(VSWR)_{in} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} \geq 1, \text{ where}$$

$$|\Gamma_a| = \left| \frac{\Gamma_{IN} - \Gamma_s^*}{1 - \Gamma_{IN}\Gamma_s} \right|,$$

$$(VSWR)_{out} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} \geq 1, \text{ where}$$

$$|\Gamma_b| = \left| \frac{\Gamma_{out} - \Gamma_L^*}{1 - \Gamma_{out}\Gamma_L} \right|.$$

DC Bias Selection



Active bias network for a BJT

Low-noise,low-power : A

Low-noise,higher power-gain : B

High Output Power : C

Higher output power and higher efficiency : D

■ Power Amplifier Classes

- Class A: conduction angle 360
- Class B: conduction angle 180
- Class AB: conduction angle >180
- Class C: conduction angle <180
- Class F: an extension of class C
- Class E: switch mode

Class A Power Amplifiers

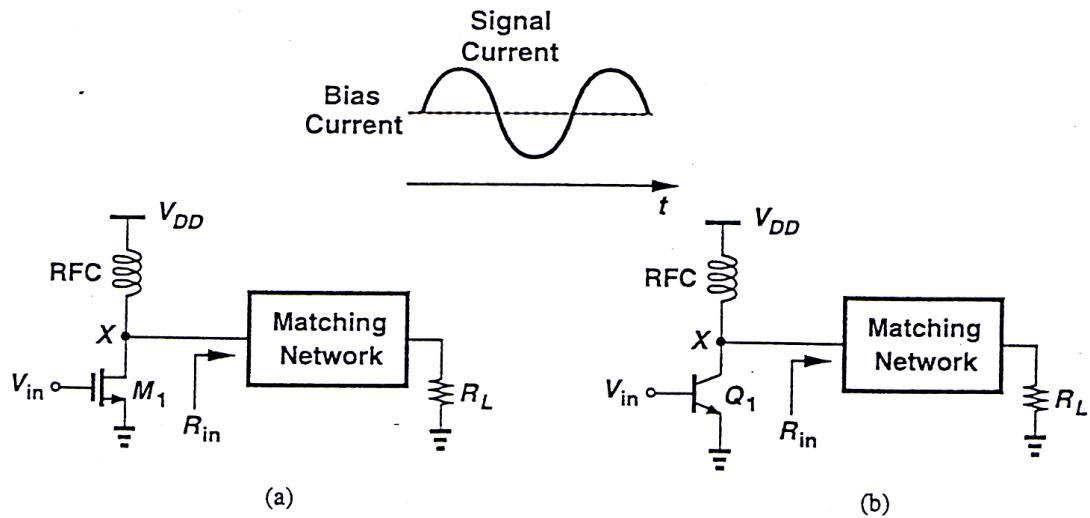
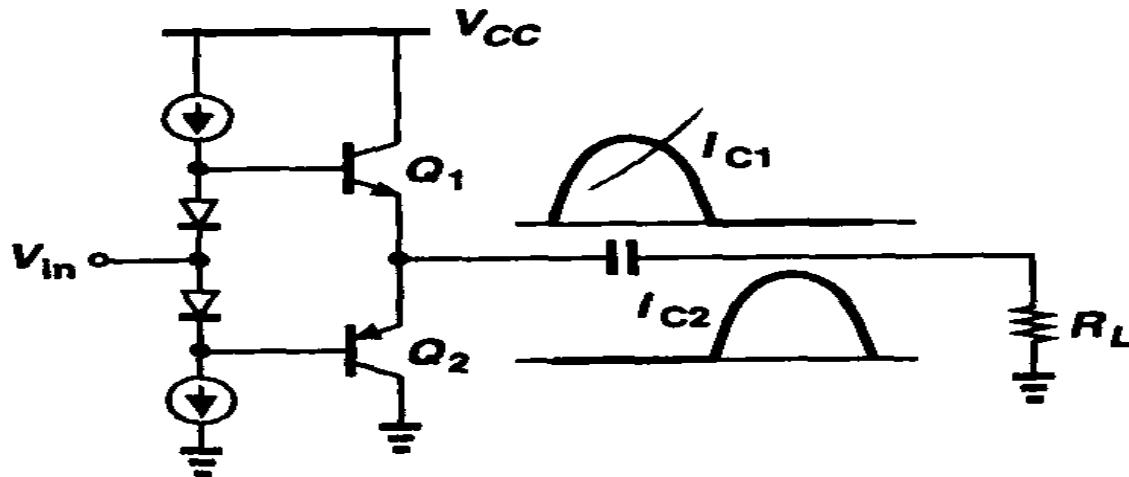


Figure 9.4 Class A stages using (a) MOS device, (b) bipolar transistor.

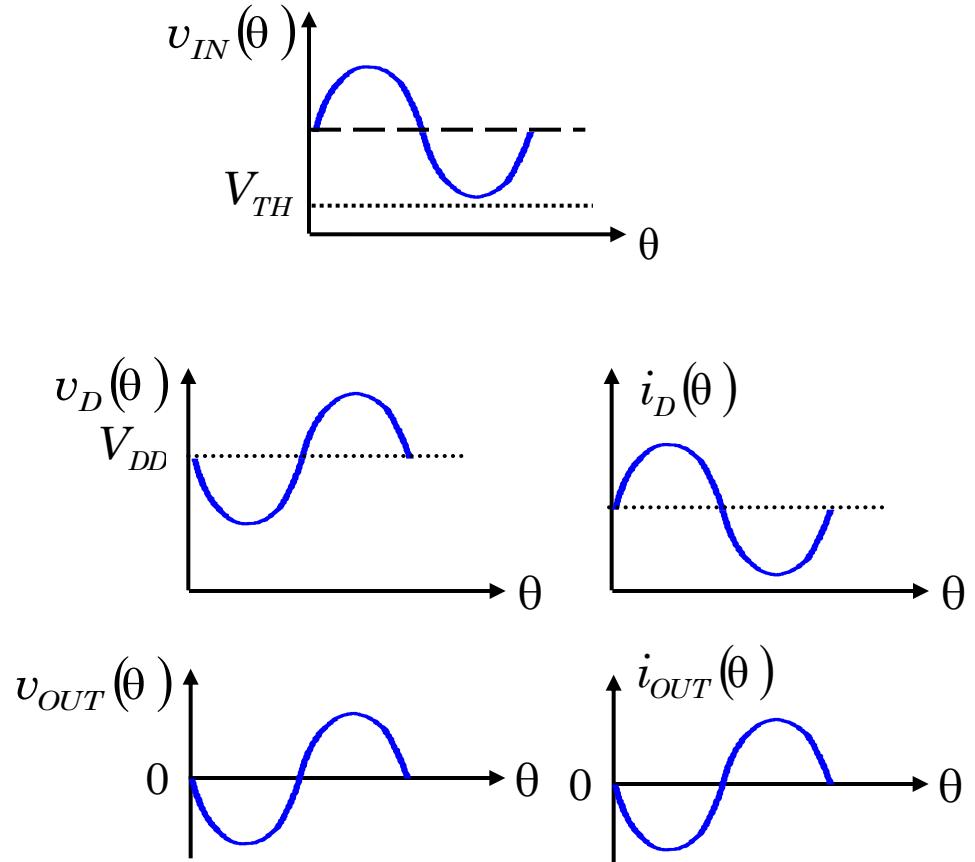
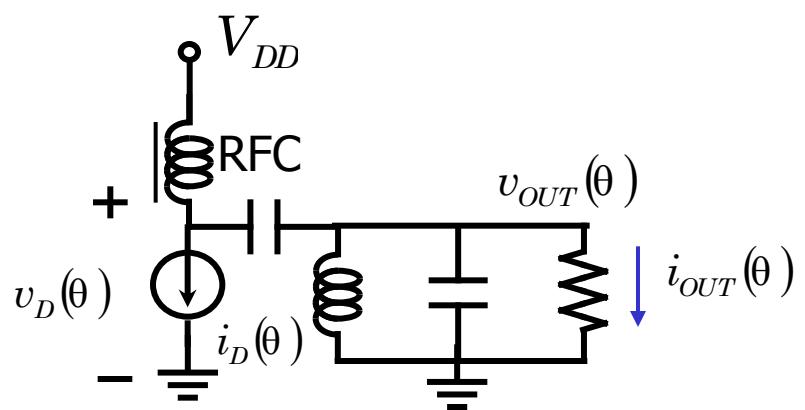
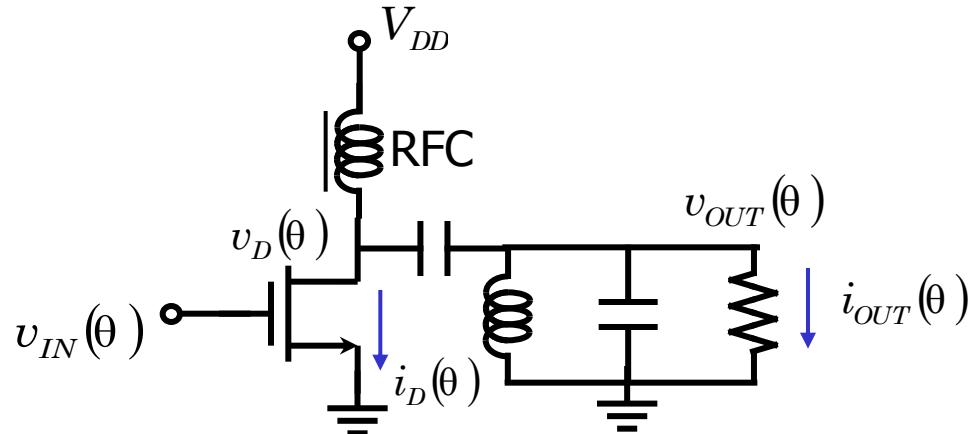
- Maximum efficiency of class A PA:
Assume drain(collector) voltage is a sinusoid having V_{pp} of $2V_{dd}$. The power delivered to matching network is $V_{DD}^2 / 2R_{in}$. And for V_x to reach $2V_{dd}$, the RFC must provide a current of V_{DD}^2 / R_{in} . Thus, the maximum efficiency is 50%.

Push-pull output stage



- The push-pull stage of above usually used in low-frequency power amplifier.
- The efficiency is better than class A PA.

Class A Power Amplifiers



Class A Power Amplifiers

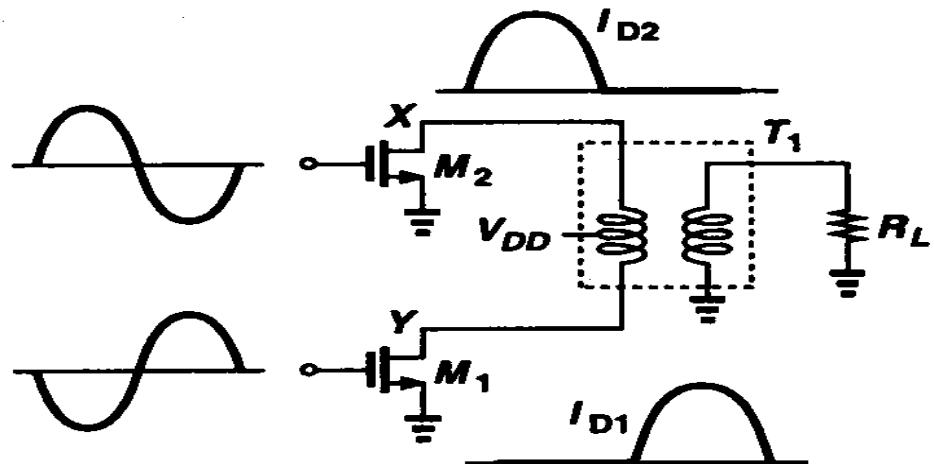
- Maximum output power

$$P_{RFout} = \frac{V_{om}^2}{2R} \leq \frac{V_{DD}^2}{2R}$$

- Efficiency

$$\eta_{Drain} = \frac{P_{RFout}}{P_{DC}} = \frac{\cancel{V_{om}^2 / 2R}}{\cancel{V_{DD}^2 / R}} = \frac{V_{om}^2}{2V_{DD}^2} \leq \frac{1}{2}$$

Class B Stage using a transform

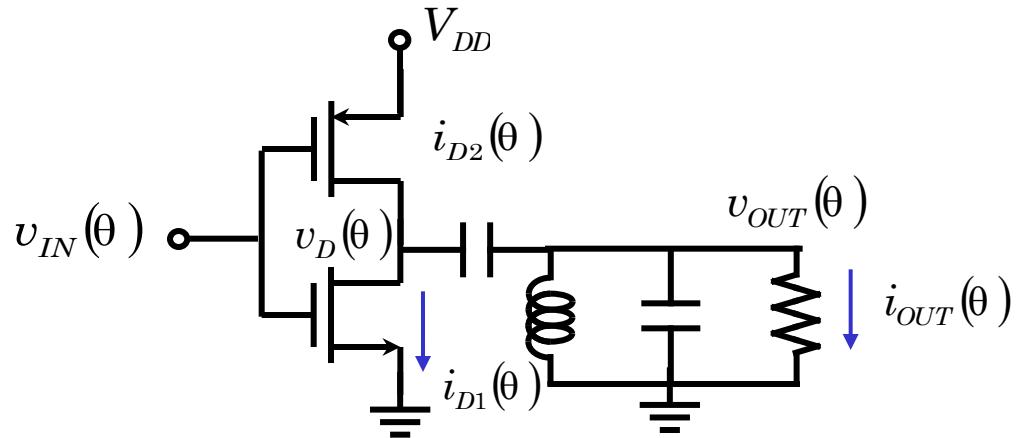


$$I_{DD,avg} = \frac{2}{T} \int_0^{T/2} \frac{V_{DD}}{n^2 R_L} \sin \omega t dt \quad (9.1)$$

- The maximum voltage swing at X and Y is $2V_{DD}$, And the equivalent resistance seen at each of X and Y is $n^2 R_L$
- The total input power of T1 is given by $P_{in} = V_{DD}^2 / 2 n^2 R_L$ and $P_{sup} = 2V_{DD}^2 / (\pi n^2 R_L)$.

$$\eta = P_{in} / P_{sup} = \pi / 4 \approx 79\%.$$

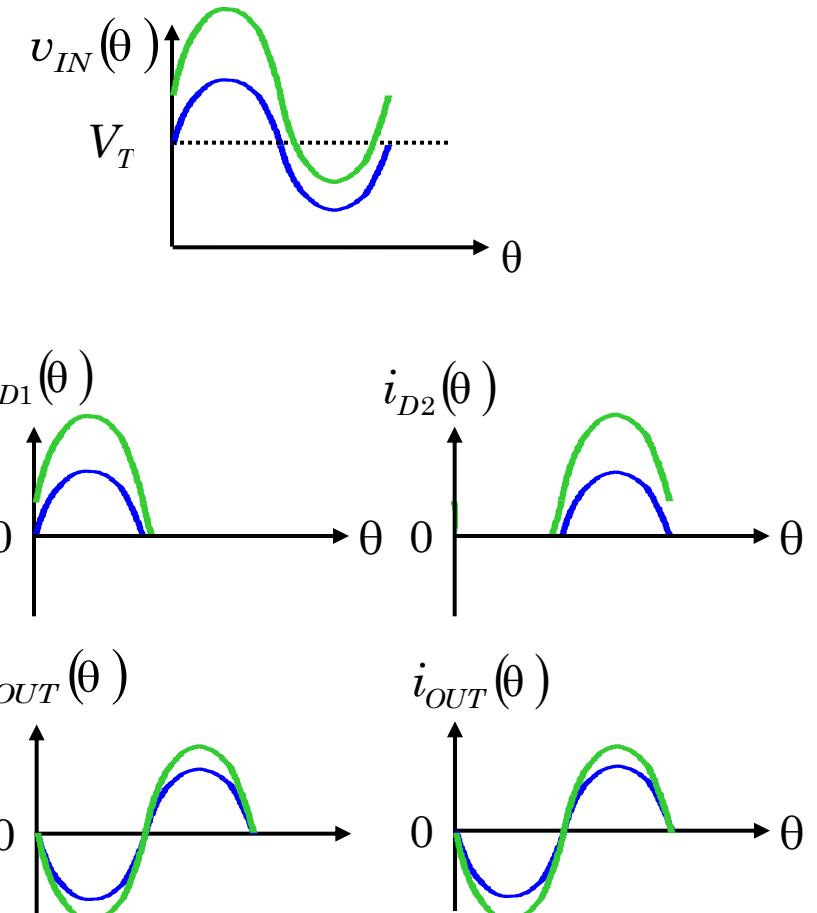
Class B, AB Power Amplifiers



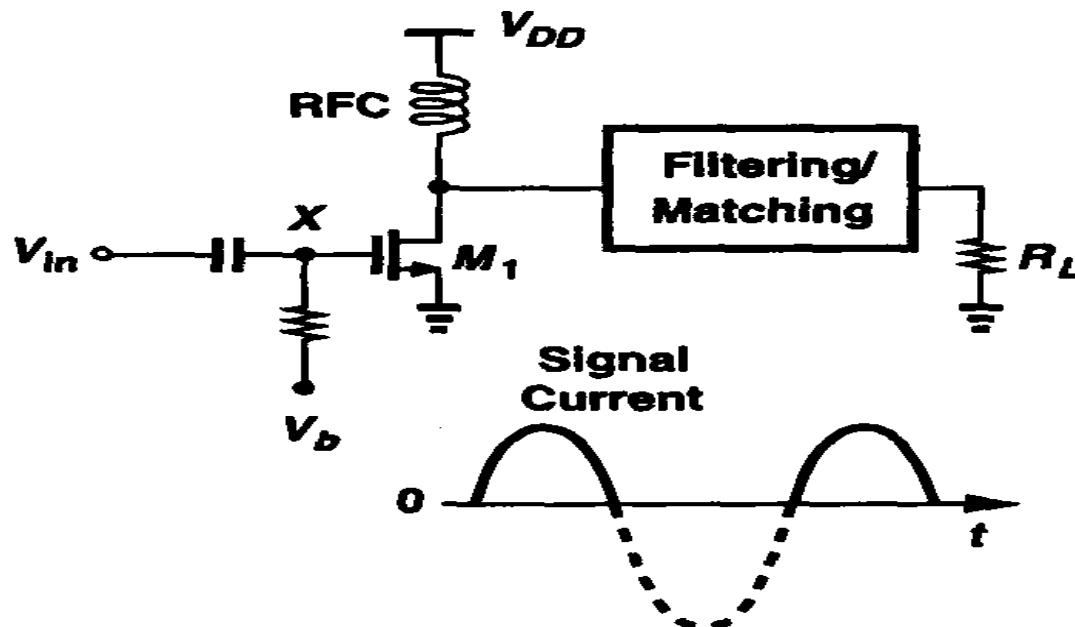
$$P_{RFout} = \frac{V_{om}^2}{2R} \leq \frac{V_{DD}^2}{2R}$$

$$I_{DC} = \frac{1}{2\pi} \int_0^{2\pi} I_D |\sin \theta| d\theta = \frac{2I_D}{\pi} = \frac{2}{\pi} \frac{V_{om}}{R}$$

$$\eta_{Drain} = \frac{P_{RFout}}{P_{DC}} = \frac{\frac{V_{om}^2}{2R}}{\frac{2}{\pi} \frac{V_{om}}{R} V_{DD}} = \frac{\pi}{4} \frac{V_{om}}{V_{DD}} \leq \frac{\pi}{4} \approx 0.785$$

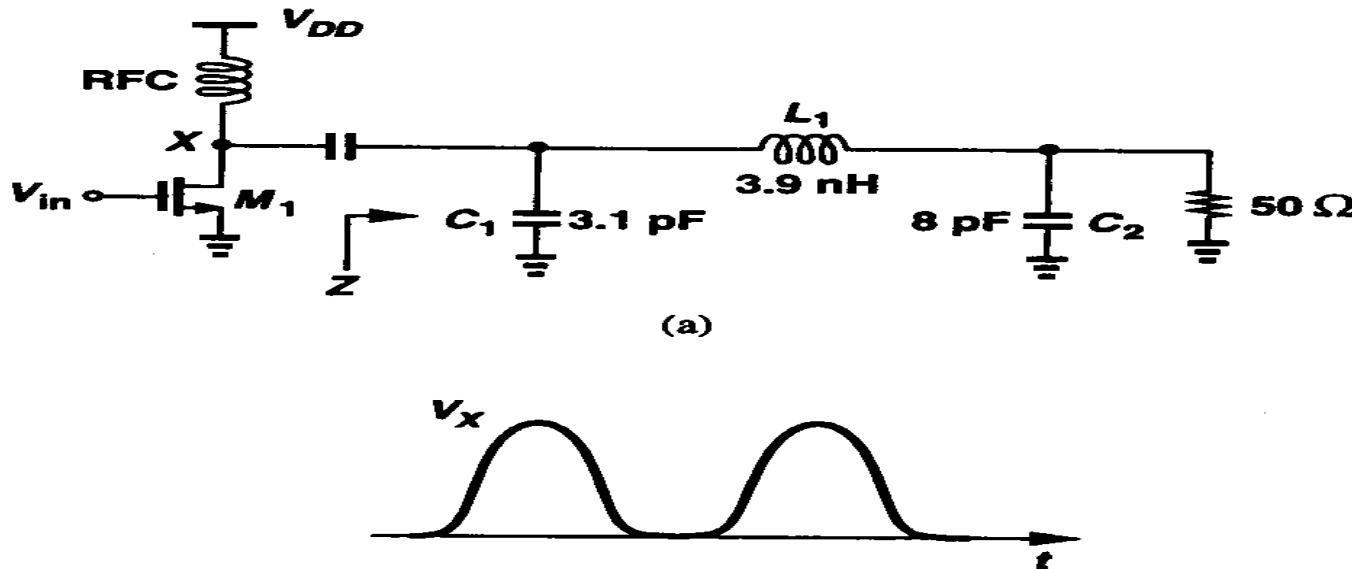


Class C PAs



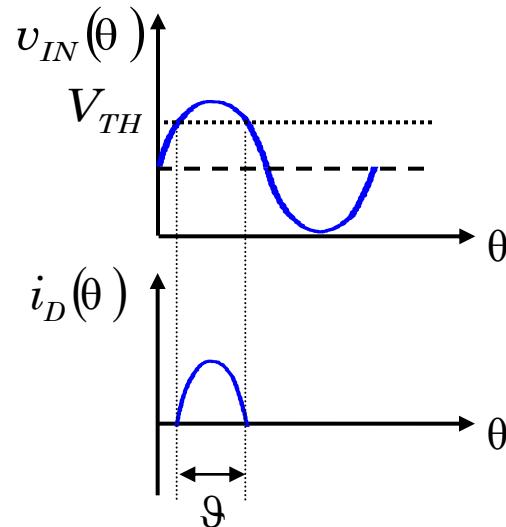
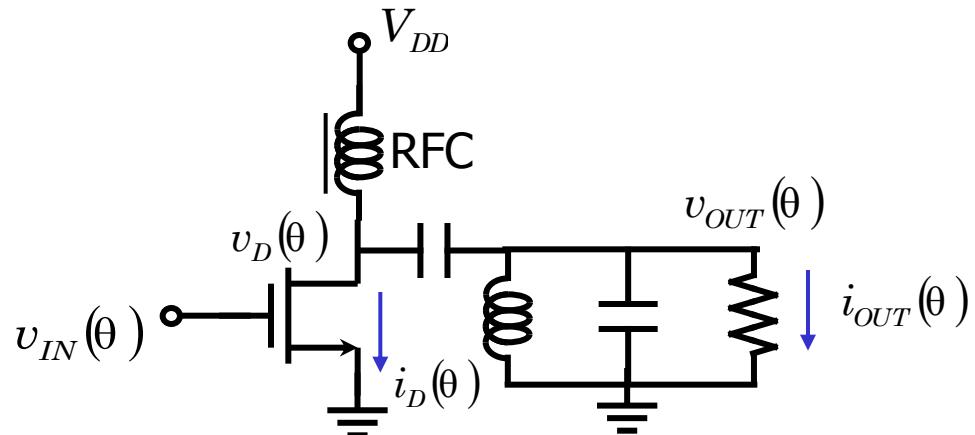
- M_1 turns On if $V_{in} > |V_b| + V_{TH}$
- The efficiency formula : $\eta = \frac{1}{4} \frac{\theta - \sin \theta}{\sin(\theta/2) - \theta/2 \cos(\theta/2)}$
- The power delivered to the load : $P_{out} \propto \frac{\theta - \sin \theta}{1 - \cos(\theta/2)}$

Ideas for Raising Efficiency



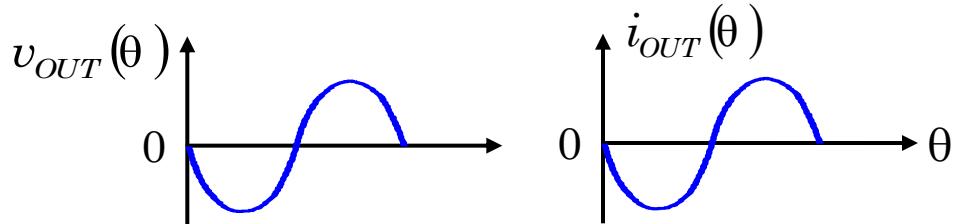
- Suppose the matching network is designed such that its input impedance is low at the fundamental frequency and quite high at the second harmonics. The drain voltage exhibits sharper edges than a sinusoid does, raising the efficiency.
- But the matching network becomes quite complex and lossy.

Class C Power Amplifier



$$P_{RFout} \propto \frac{9 - \sin 9}{1 - \cos(\frac{9}{2})}$$

$$\eta_{\text{Drain}} = \frac{P_{RFout}}{P_{DC}} = \frac{1}{4} \frac{9 - \sin 9}{\sin(\frac{9}{2}) - \frac{9}{2} \cos(\frac{9}{2})}$$



High Efficiency PA

Class A

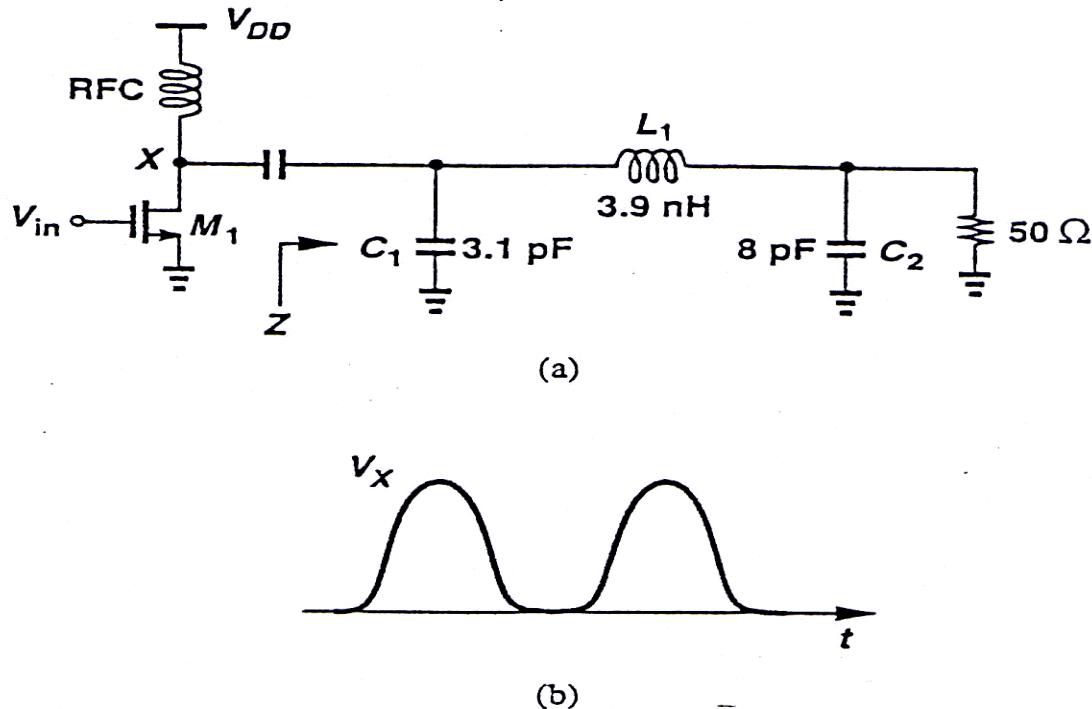
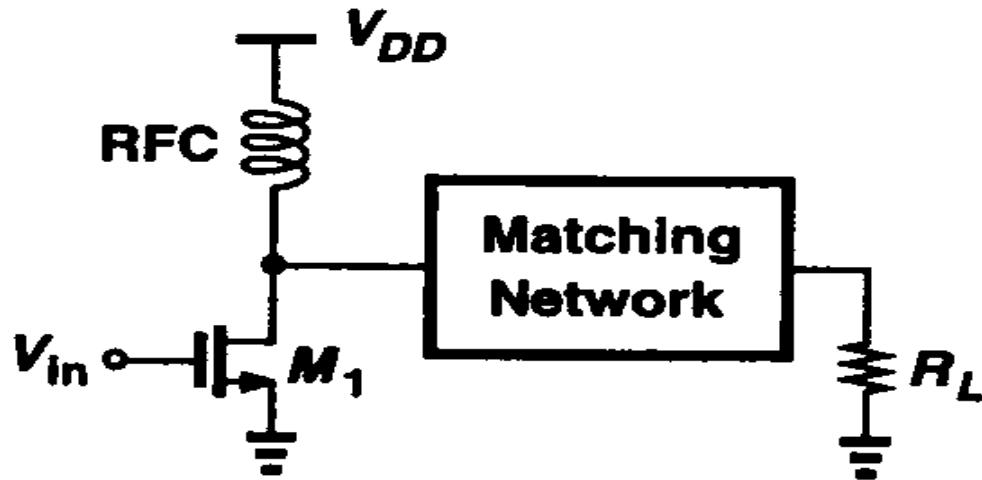


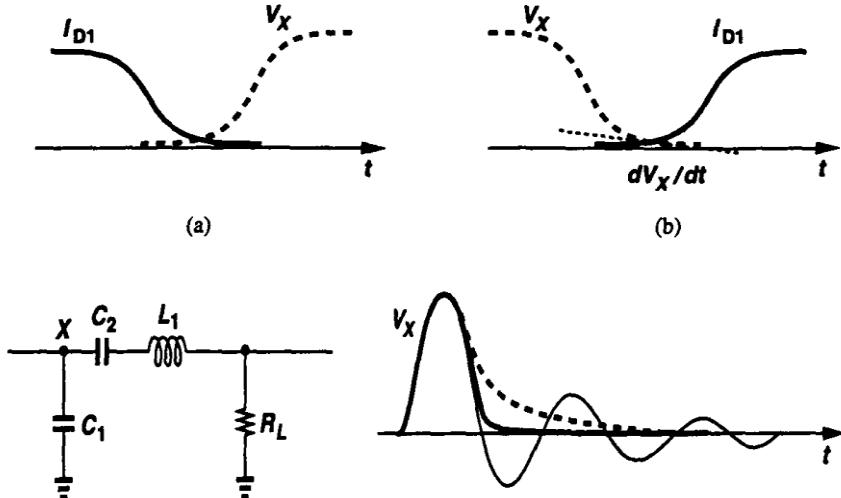
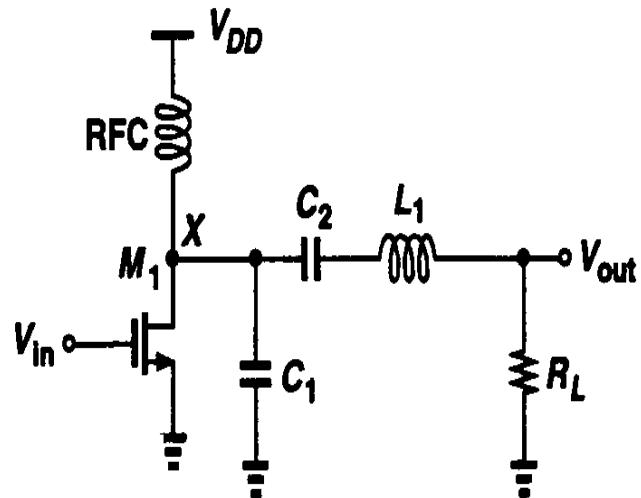
Figure 9.8 (a) PA with high harmonic termination, (b) drain voltage waveform.

Class E PAs



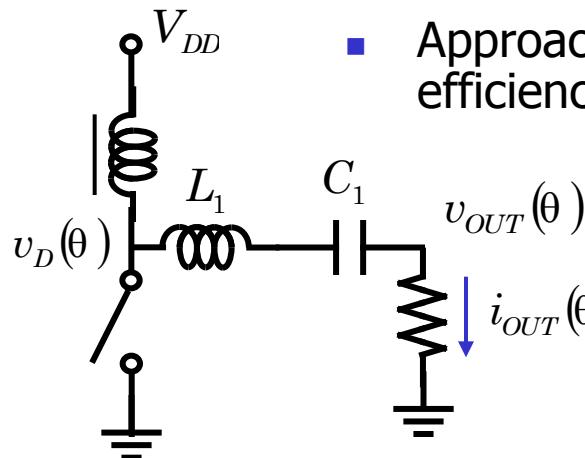
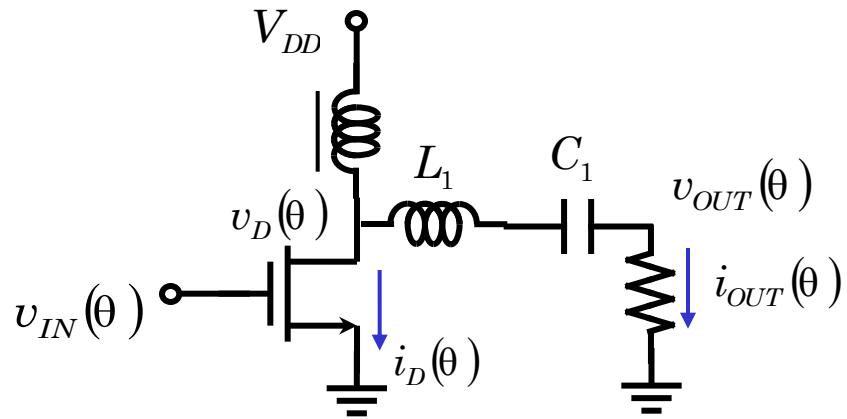
- Class E stages are nonlinear amplifiers that achieve efficiencies approaching 100% while delivering full power.
- It's a “switching power amplifier”.
- The voltage applied to the gate of M_1 must approximate a rectangular waveform. And the switch on-resistance must be low.

Class E Pas (Cont.)

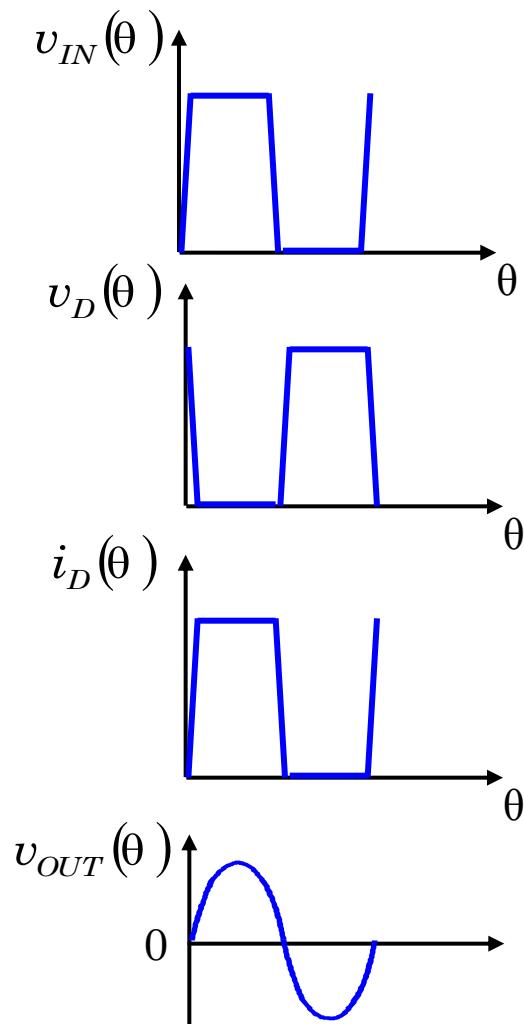


- As the switch turns off, V_x remains low long enough for the current to drop to zero.
- V_x reaches zero just before the switch turns on.
- dV_x/dt is also near zero when the switch turns on.
- After the switch turns off, the load network operates as a damped second-order system.

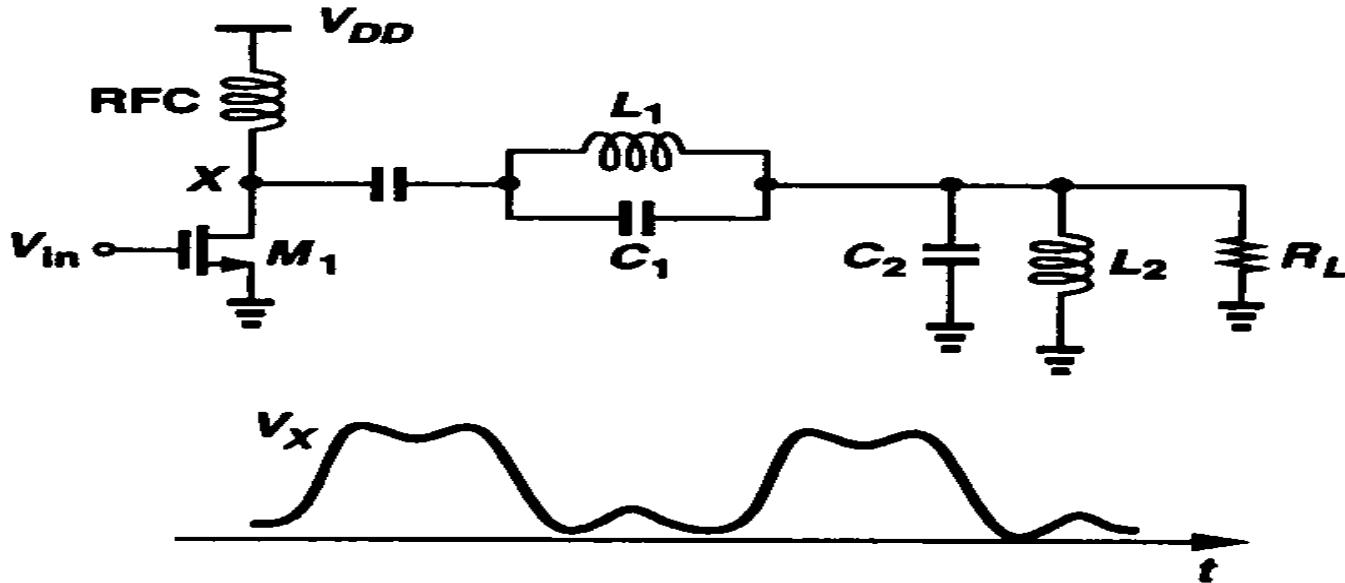
Class E Power Amplifiers



- Switch mode
- Approaching 100% efficiency

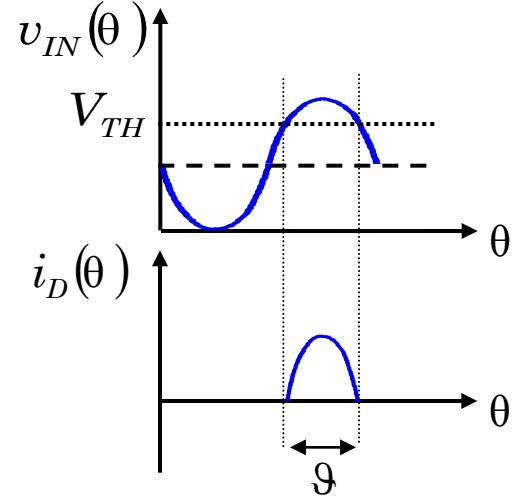
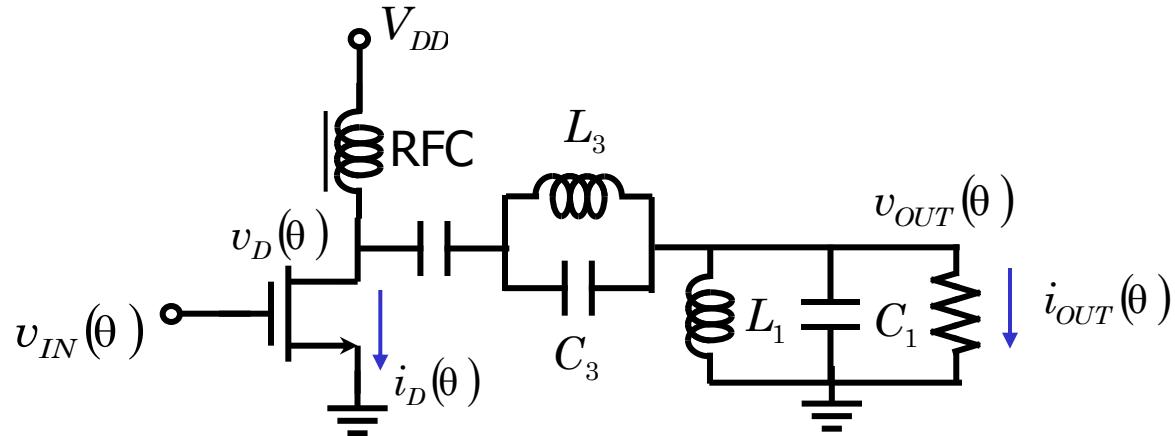


Class F PAs

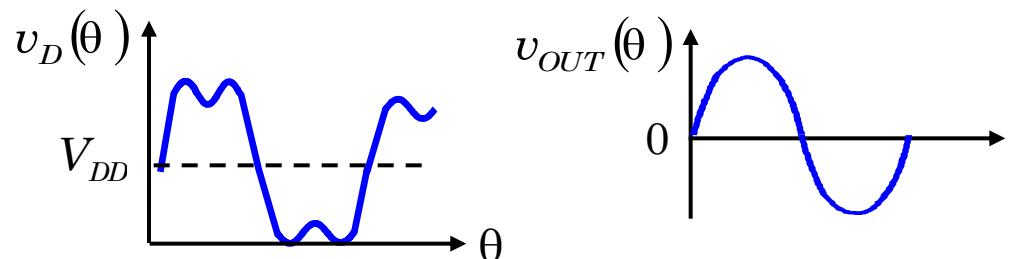


- The idea of harmonics termination for a class A stage can be extended to nonlinear amplifiers as well.
- It can be proved that the peak efficiency of class F amplifiers is equal to 88% for third-harmonics peaking and 85% for second-harmonics peaking.

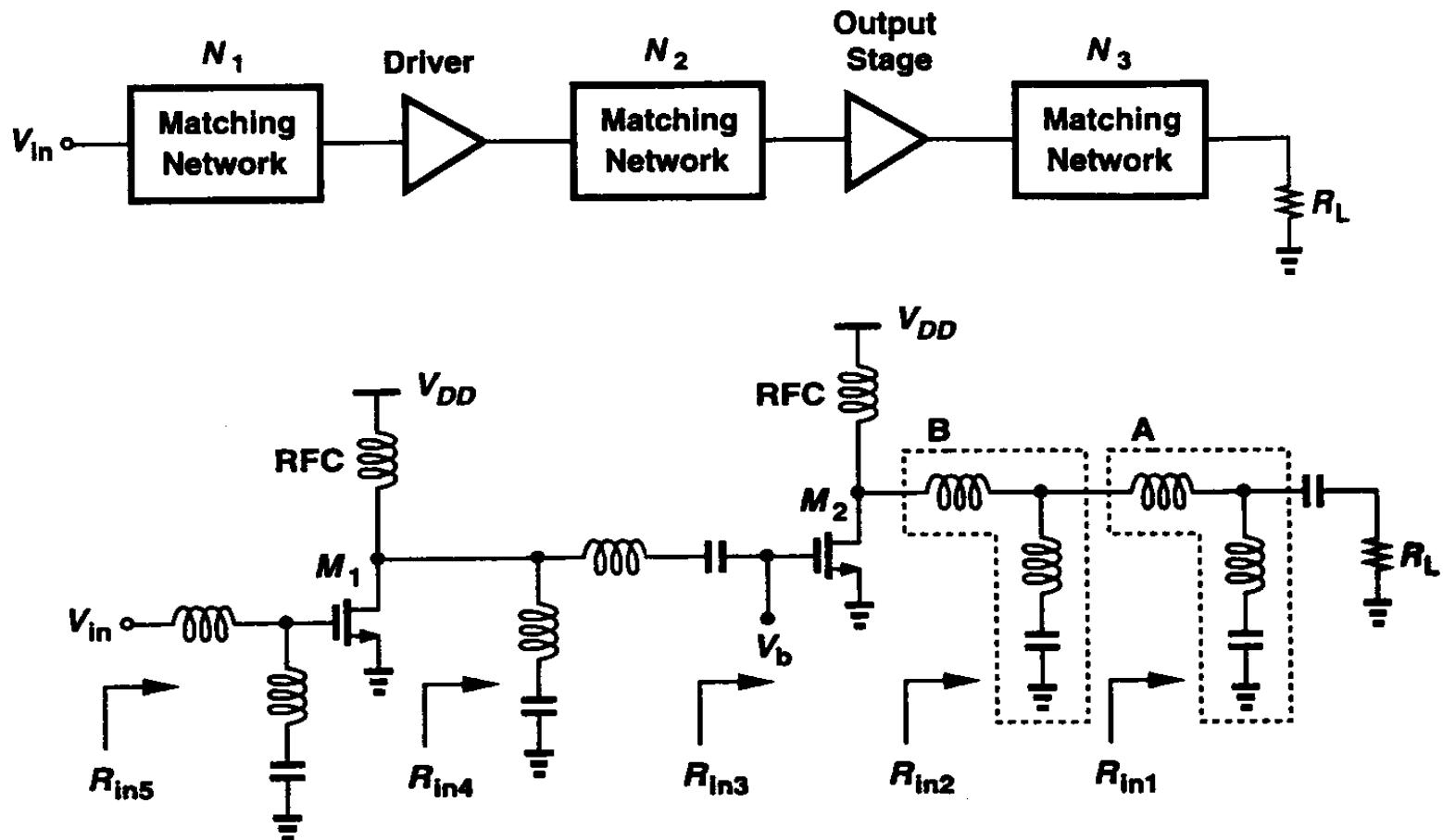
Class F Power Amplifiers



- L_3C_3 tuned to the 2nd or 3rd harmonics
- Peak efficiency
 - 88% for 3rd harmonics peaking
 - 85% for 2nd harmonics peaking.

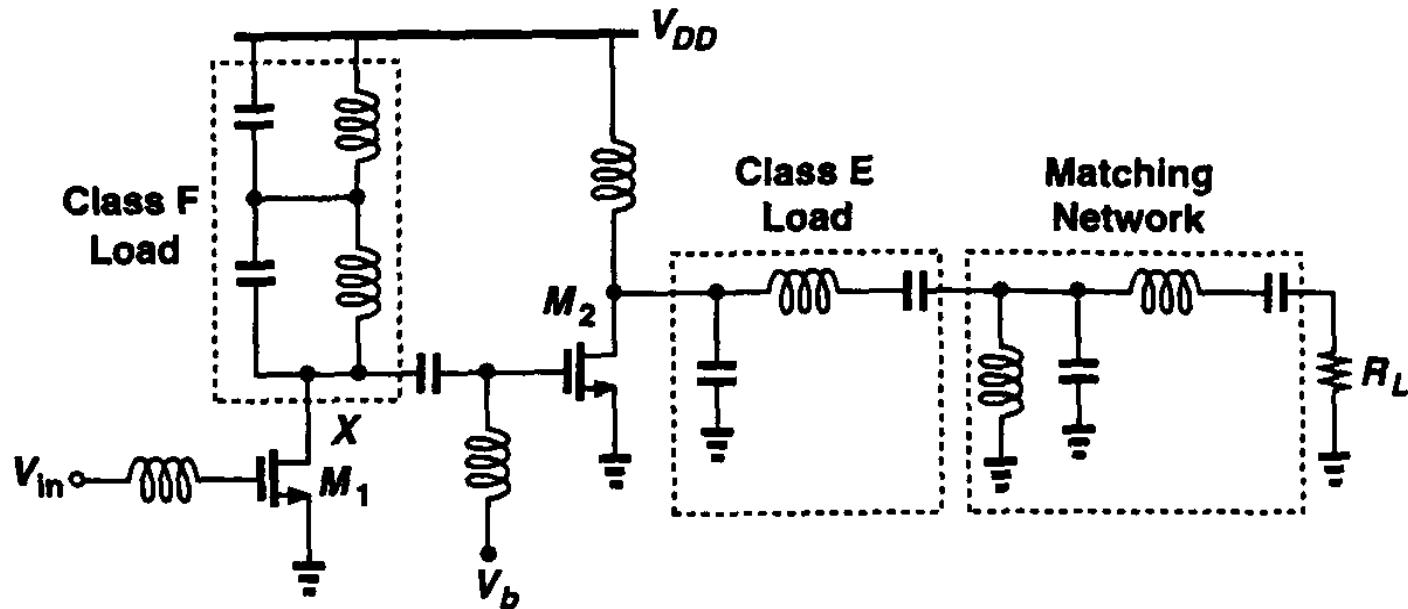


Power amplifier examples



[*] B. Razavi

Power amplifier examples

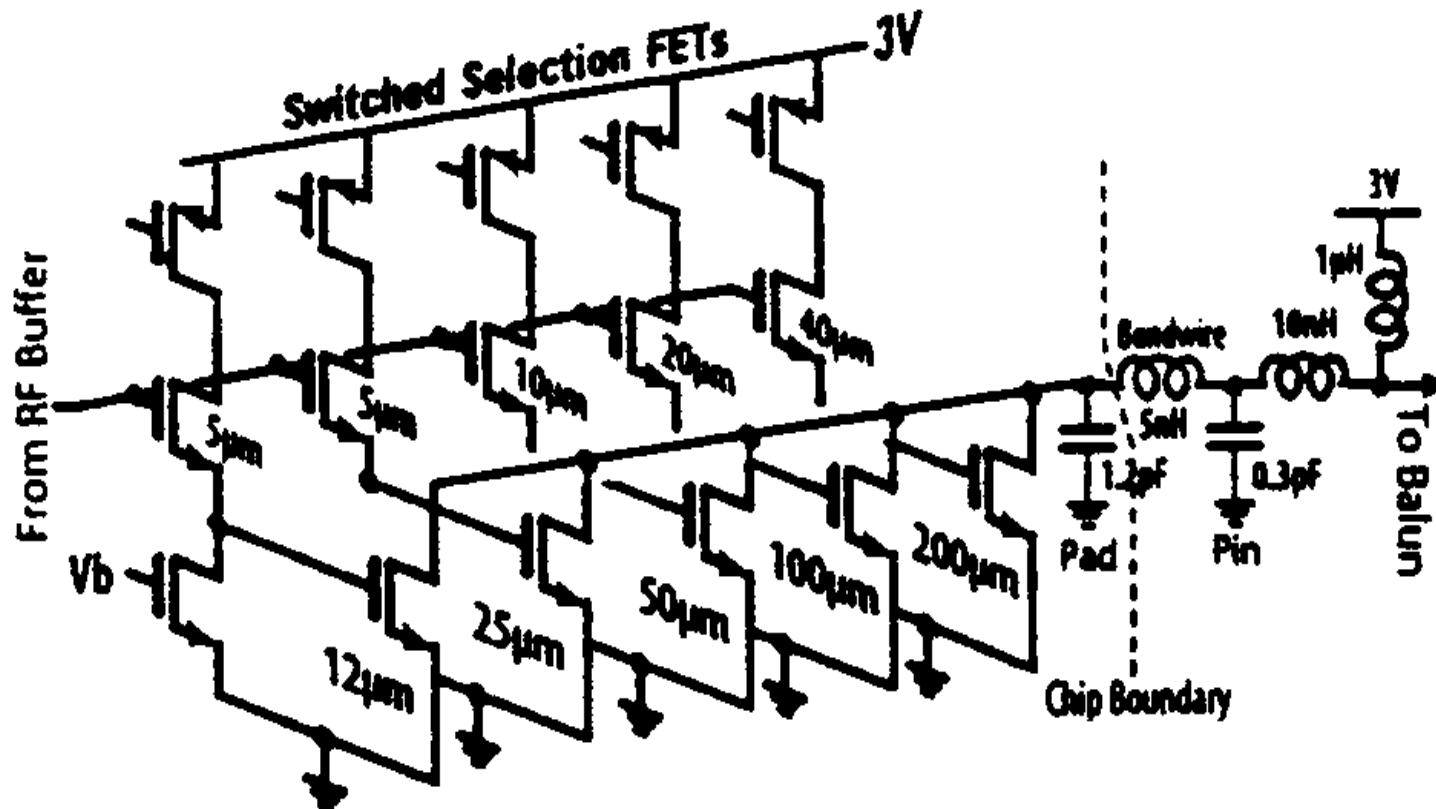


[*] B. Razavi

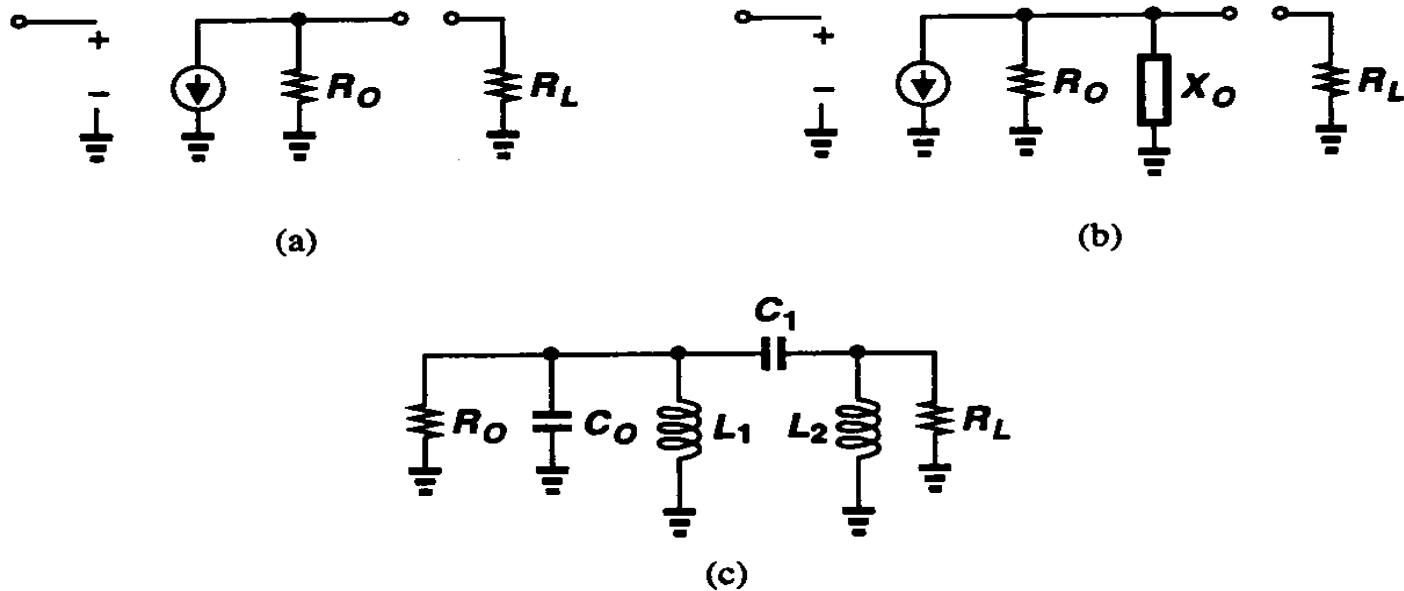
Kuei-Ann Wen National chiao Tung Univ.

30

Power amplifier examples



Nonlinear impedance matching



- Maximum power transfer does not correspond to maximum efficiency.
- The matching can be obtained roughly using small-signal approximation, but modifying these for maximum large-signal efficiency requires a great deal of trial and error.

Large-Signal Impedance Matching

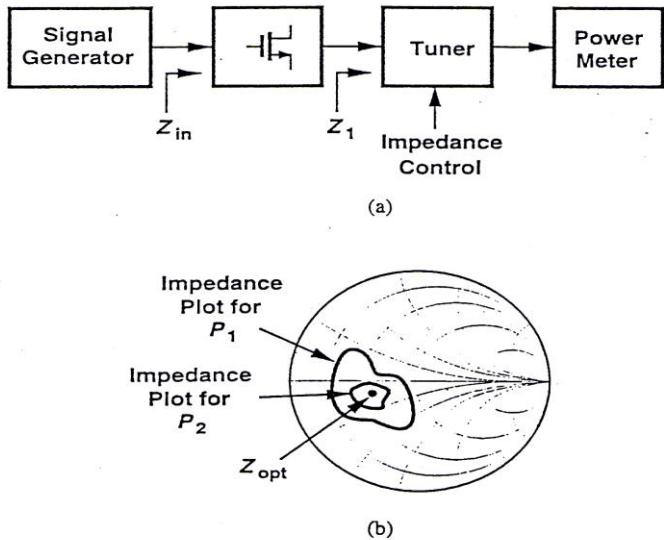
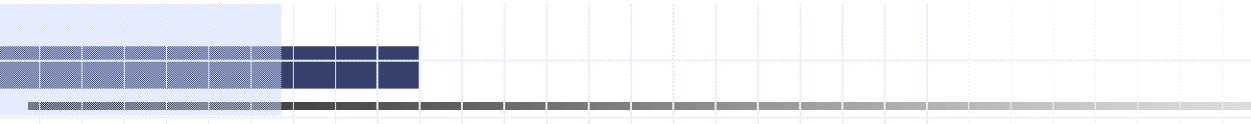


Figure 9.15 (a) Load-pull test, (b) power contours on a Smith chart.

- In a “load-pull” test, the output power is measured and plotted as a function of the complex load seen by the transistor.
- As Z_1 varies so does Z_{in} , a second tuner between the signal generator and the transistor is needed.

Linearization Techniques

- Most linear power Amp.
 - Class A of efficiency around %30 to %40 for portable devices.
- To improve efficiency
 - Linearization after nonlinear PAs.
- Linearization method:
 - feedford
 - feedback
 - envelope elimination and restoration
 - LINC



Liberalization Technology: Feedforward

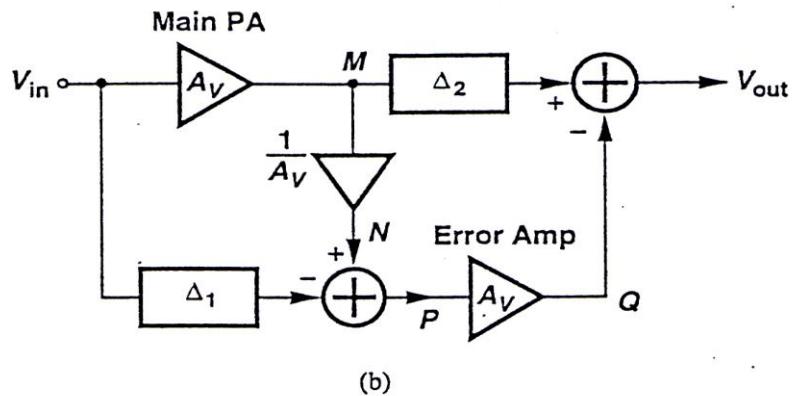
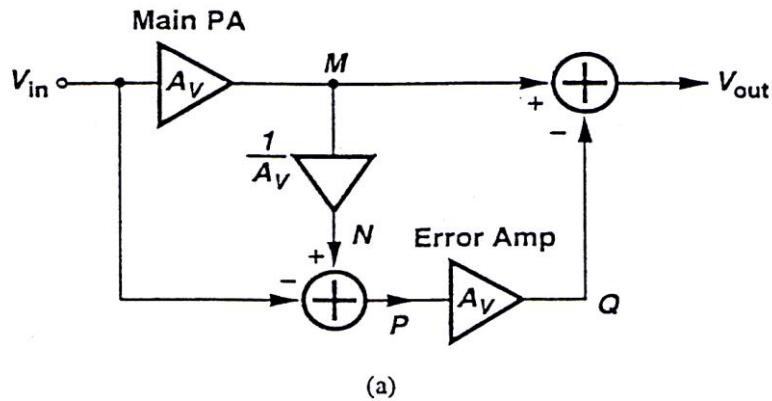


Figure 9.16 (a) Simple feedforward topology, (b) addition of delay elements.

$$V_M = A_V V_{in} + V_D$$

$$V_N = V_{in} + V_D / A_V$$

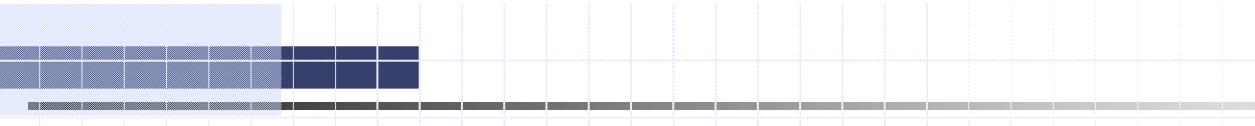
$$V_P = V_D / A_V, V_Q = V_D$$

$$V_{out} = A_V V_{in}$$

$$E = \sqrt{1 - 2(1 + \frac{\Delta A}{A}) \cos \Delta \phi + (1 + \frac{\Delta A}{A})^2} \quad (9.4)$$

The suppression of the magnitude of the IM products in V_{out} : E





Liberalization Technology: Feedback

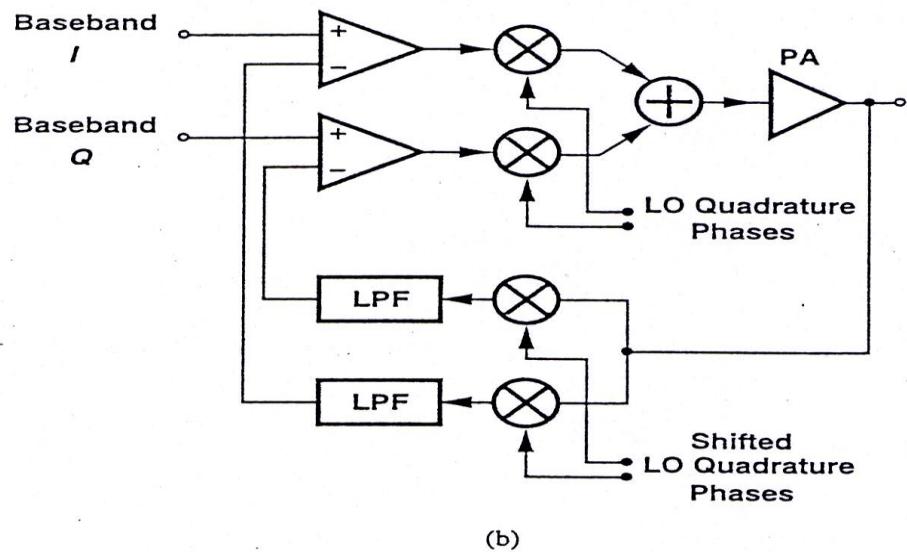
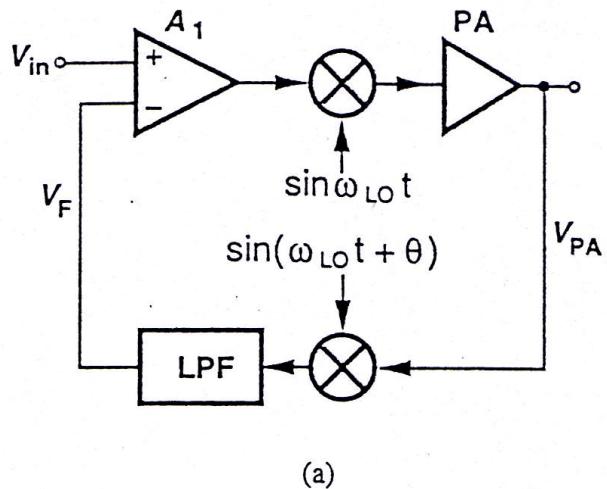
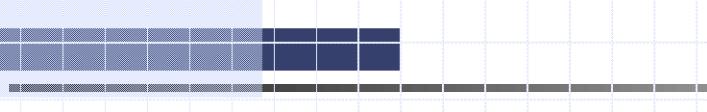


Figure 9.17 (a) Feedback by frequency translation, (b) Cartesian feedback.



Envelope Elimination and Restoration

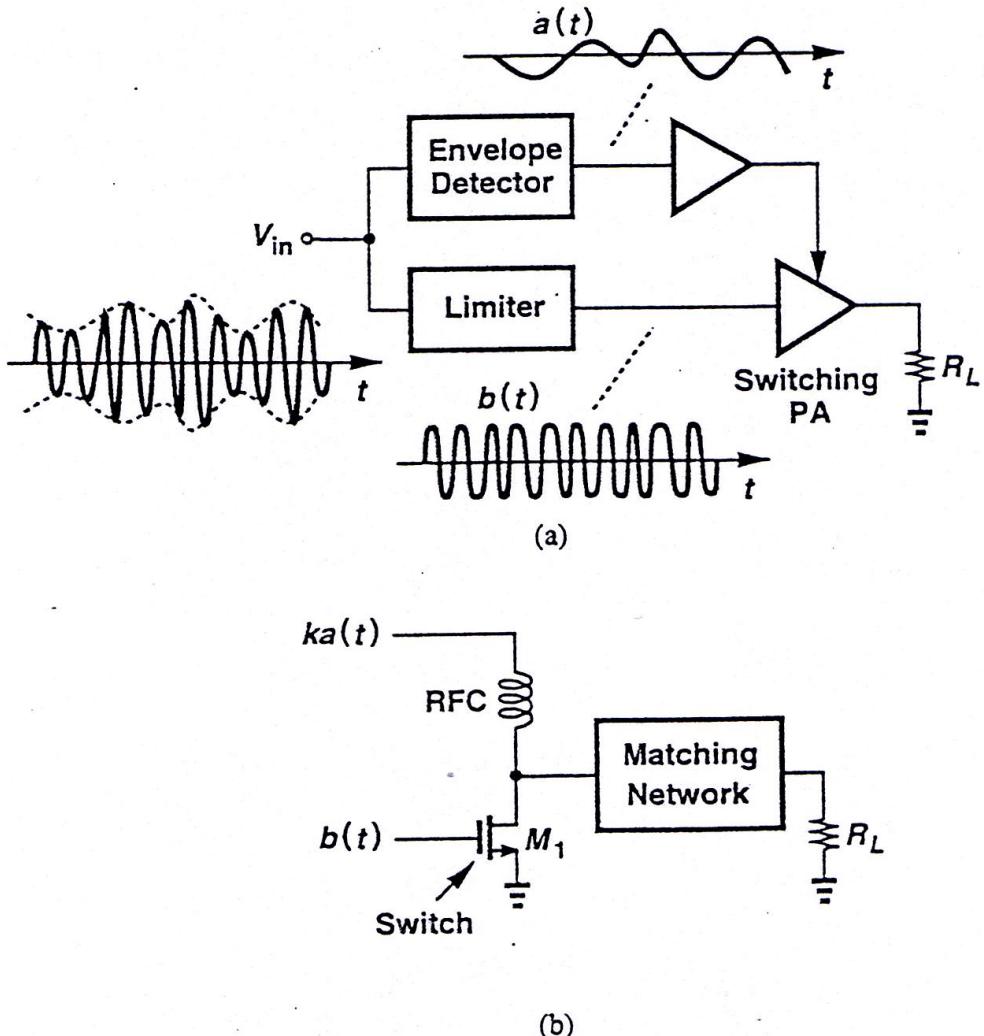
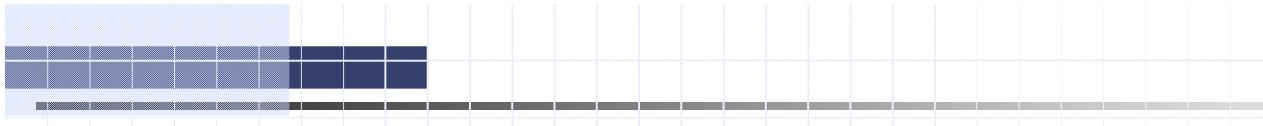


Figure 9.18 (a) Envelope elimination and restoration, (b) implementation of the output stage.



Linearization using non-linear circuits

LINC Technology

(1)

$$v_{in}(t) = a(t) \cos[\omega_c t + \phi(t)], \theta(t) = \sin^{-1}[a(t)/V_0]$$

$$v_1(t) = 0.5V_0 \sin[\omega_c t + \phi(t) + \theta(t)]$$

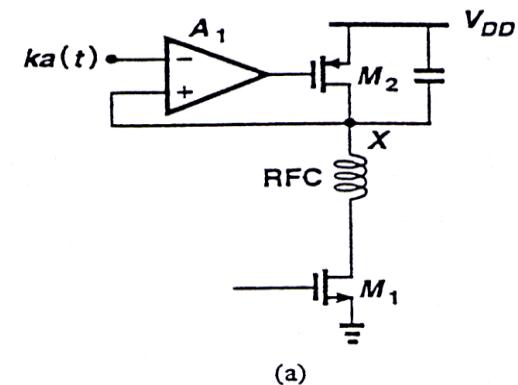
$$v_2(t) = -0.5V_0 \sin[\omega_c t + \phi(t) - \theta(t)]$$

(2)

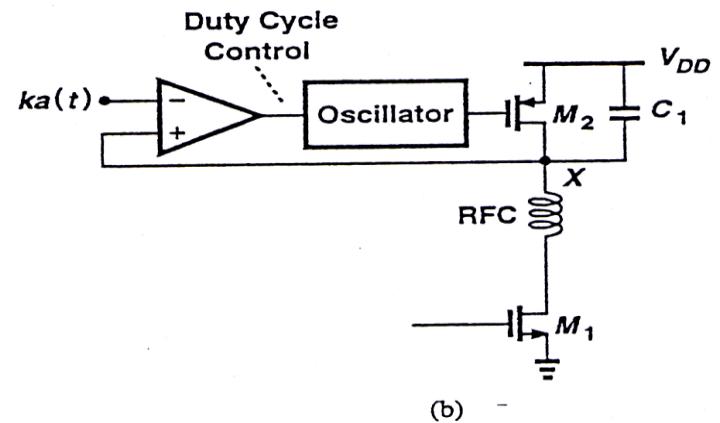
$$v_1(t) = v_I(t) \cos[\omega_c t + \phi(t)] + v_Q(t) \sin[\omega_c t + \phi(t)]$$

$$v_2(t) = -v_I(t) \cos[\omega_c t + \phi(t)] + v_Q(t) \sin[\omega_c t + \phi(t)]$$

$$v_I(t) = a(t)/2, v_Q(t) = \sqrt{V_0^2 - a(t)^2 / 2}$$



(a)



(b)

Figure 9.19 Modulation of the PA output by (a) a low-frequency feedback amplifier, (b) pulse-width modulation.

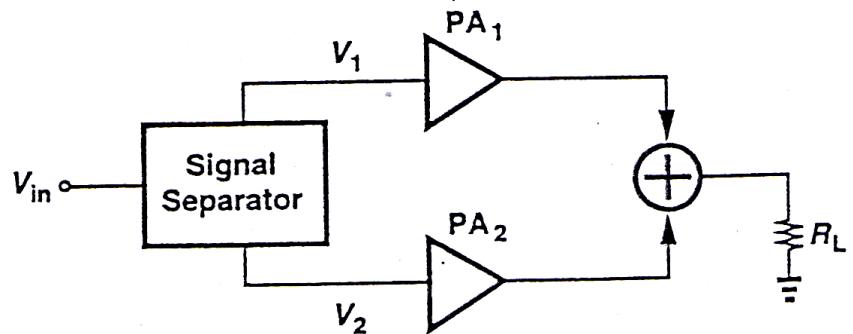


Figure 9.20 Linear amplification using nonlinear stages.

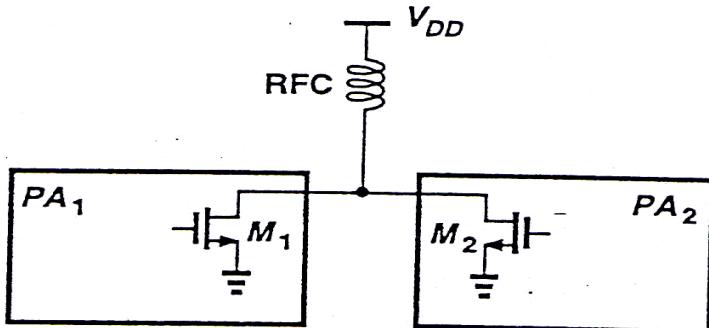


Figure 9.21 Addition of the outputs of two PAs.

$$v_1(t) = v_I(t) \cos(\omega_c t + \phi) + v_Q(t) \sin(\omega_c t + \phi) \quad (9.5)$$

$$v_2(t) = -v_I(t) \cos(\omega_c t + \phi) + v_Q(t) \sin(\omega_c t + \phi) \quad (9.6)$$

Limitations of integrated CMOS Power Amplifier

- Device Breakdown Voltage
 - Low voltage swing
 - Sub- μ CMOS process has low oxide breakdown
- Low current driving capabilities
 - Larger device required for a given current
- Larger Capacitances
 - Tuning is more difficult
- Substrate Coupling with the RF Blocks
 - PA injects more currents into substrate
- Lower Q passive elements

Conclusions

- CMOS Technology for RF is for applications
 - Integrated with significant digital circuits
 - Lowest cost
 - Moderate radio performance
- Accurate RF models are critical for RF CMOS circuit design
- Continuous process improvement enables CMOS RF capability

Chapter 9

Output Stages And Power Amplifiers

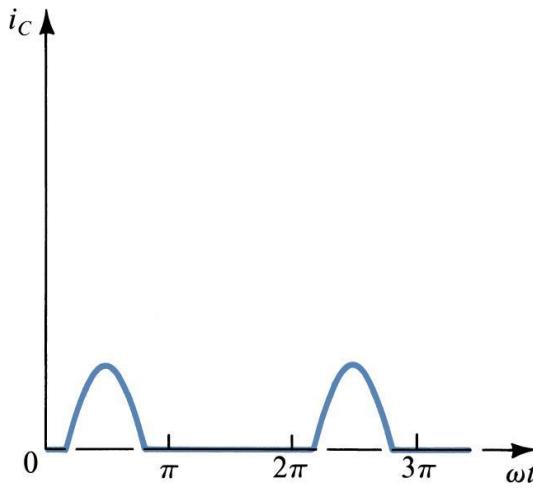
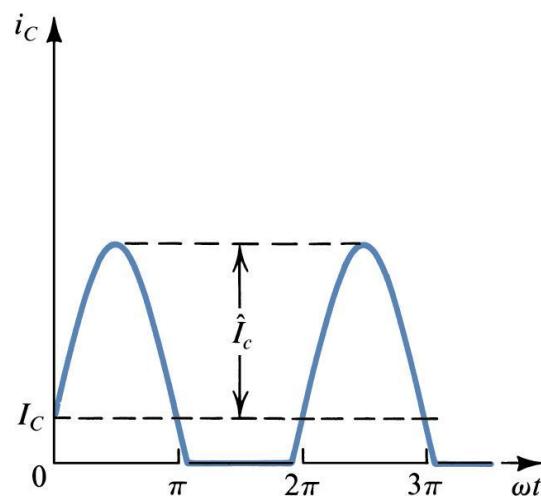
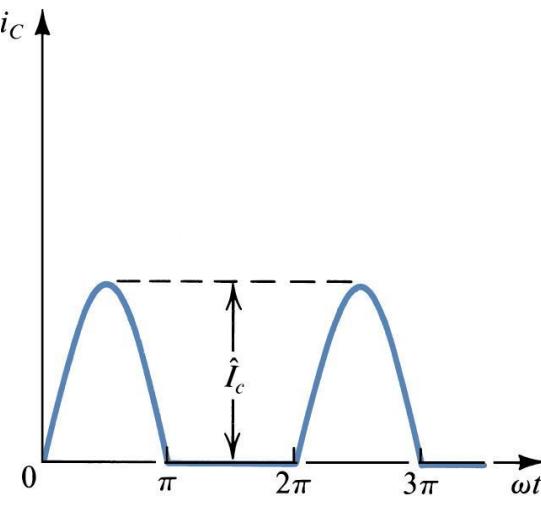
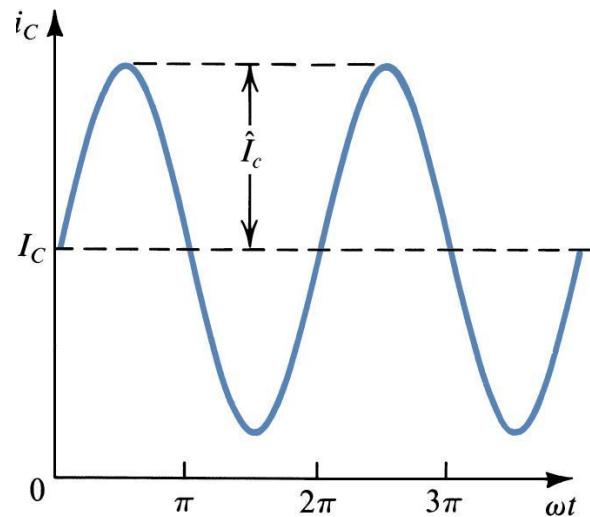
Low Output Resistance – no loss of gain

Small-Signal Not applicable

Total-Harmonic Distortion (fraction of %)

Efficiency

Temperature Requirements

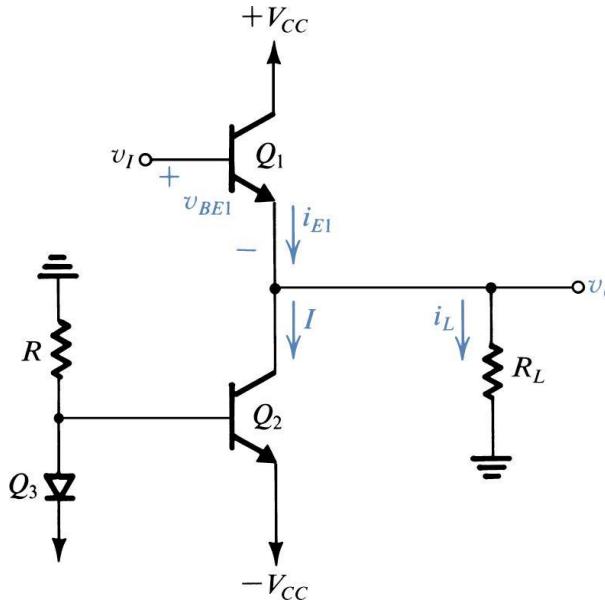
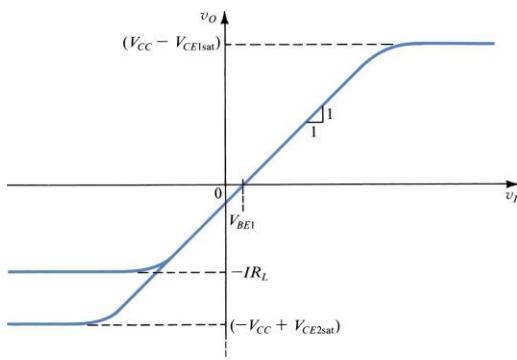
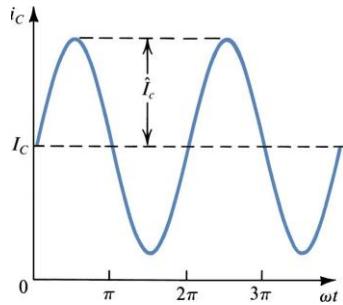


Collector current waveforms for transistors operating in (a) class A, (b) class B, (c) class AB, and (d) class C amplifier stages.

Class A

An emitter follower (Q_1) biased with a constant current I supplied by transistor Q_2 .

Transfer Characteristics



Transfer characteristic of the emitter follower. This linear characteristic is obtained by neglecting the change in v_{BE1} with i_L . The maximum positive output is determined by the saturation of Q_1 . In the negative direction, the limit of the linear region is determined either by Q_1 turning off or by Q_2 saturating, depending on the values of I and R_L .

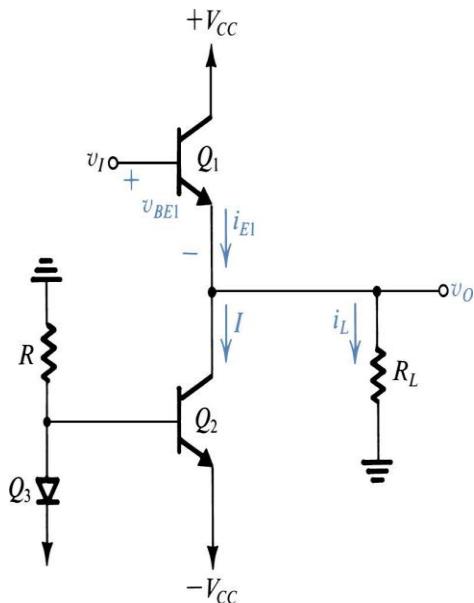
Class A

Transfer Characteristics

Crossover distortion can be eliminated by biasing the transistors at a small, non-zero current.

A bias Voltage V_{BB} is applied between Q_n and Q_p .

For $v_i = 0$, $v_o = 0$, and a voltage $V_{BB}/2$ appears across the base-emitter junction of each transistor.



$$i_N = i_P = I_Q = I_S \cdot e^{\frac{V_{BB}}{2 \cdot V_T}}$$

V_{BB} is selected to result the required quiescent current I_Q

$$v_o = v_i + \frac{V_{BB}}{2} - v_{BEN}$$

$$i_N = i_P + i_L$$

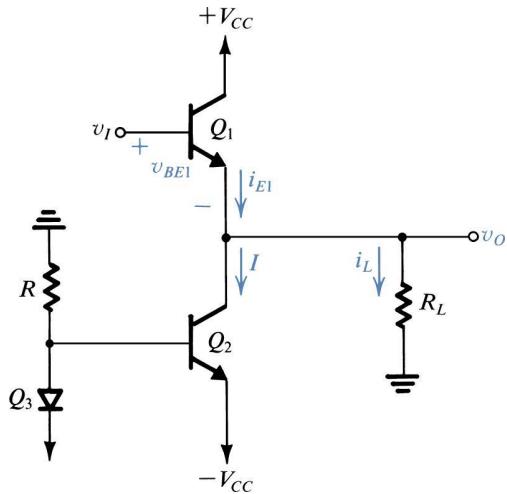
$$v_{BEN} + v_{EBP} = V_{BB} \quad V_T \cdot \ln\left(\frac{i_N}{I_S}\right) + V_T \cdot \ln\left(\frac{i_P}{I_S}\right) = 2 \cdot V_T \cdot \ln\left(\frac{i_Q}{I_S}\right)$$

$$i_N^2 = I_Q^2$$

$$i_N^2 - i_L \cdot i_N - I_Q^2 = 0$$

Class A

Transfer Characteristics



From figure 9.3 we can see that

$$v_{O\max} = V_{CC} - V_{CE1\text{sat}}$$

In the negative direction, the limit of the linear region is determined either by Q1 turning off

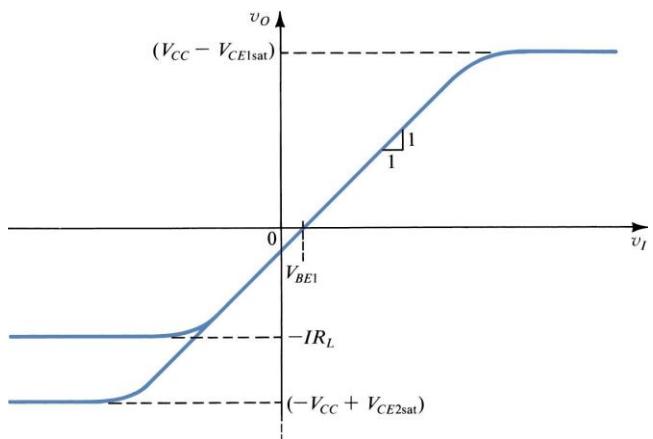
$$v_{O\min} = -I \cdot R_L$$

or by Q2 saturating

$$v_{O\min} = -V_{CC} + V_{CE2\text{sat}}$$

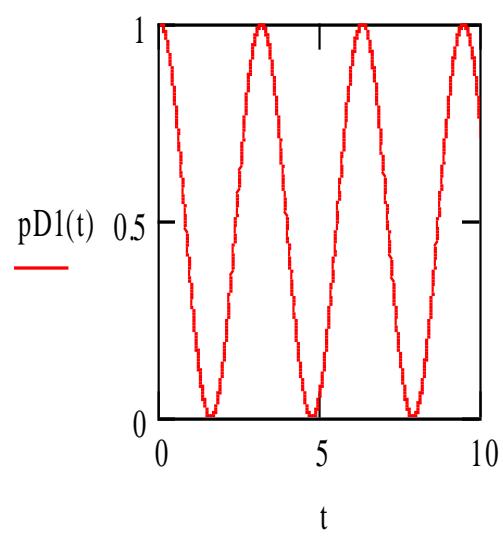
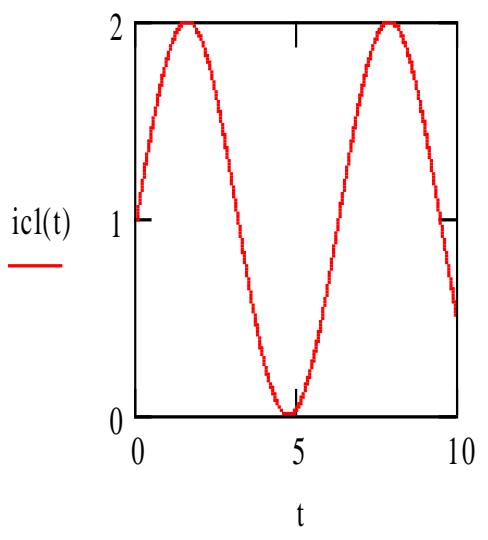
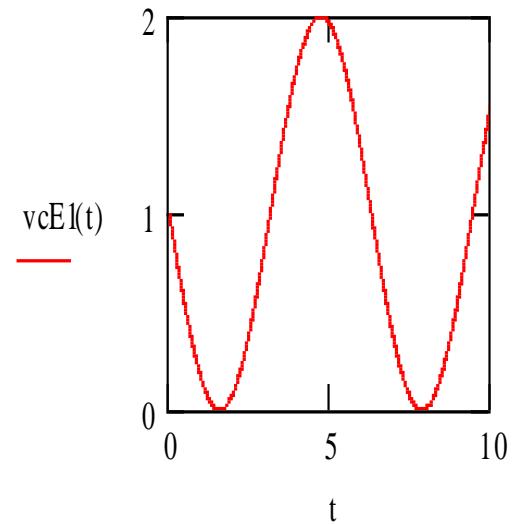
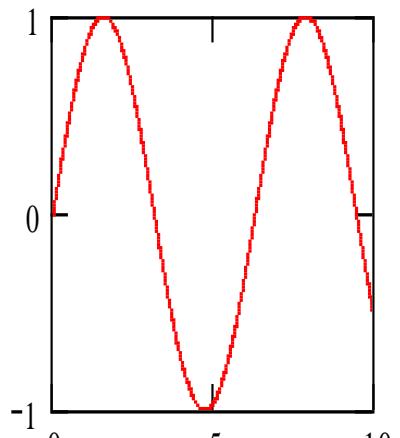
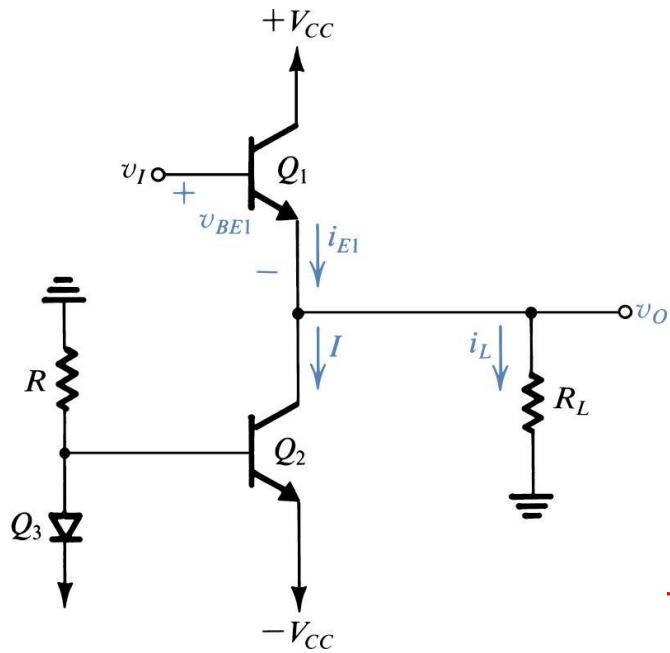
Depending on the values of I and R_L . The absolutely lowest output voltage is that given by the previous equation and is achieved provided that the bias current I is greater than the magnitude of the corresponding load current

$$I \geq \frac{|-V_{CC} + V_{CE2\text{sat}}|}{R_L}$$



Class A

Signal Waveforms



Class A

Power Dissipation

$$P = V_{CC} \cdot I$$

Largest Power Dissipation When $v_o = 0$

Q_1 must be able to withstand a continuous dissipation of $V_{CC} \cdot I$

The power dissipation of Q_1 depends on the value of R_L .

If R_L is infinite, $i_{C1} = I$ and the dissipation in Q_1 depends on v_o .

Maximum power dissipation will occur when $v_o = -V_{CC}$ since v_{CE1} will be $2V_{CC}$.
 $p_{D1} = 2V_{CC} \cdot I$. This condition would not normally persist for a prolonged interval, so
the design need not be that conservative. The average $p_{D1} = V_{CC} \cdot I$

When R_L is zero a positive voltage would result in a theoretically infinite current (large practical value) would flow through Q_1 . Short-circuit protection is necessary.

Class A

Power Conversion Efficiency

$$\eta = \frac{\text{load_power}(P_L)}{\text{supply_power}(P_S)}$$

$$P_L = \frac{1}{2} \cdot \frac{V_o^2}{R_L} \quad V_o \quad \text{average voltage}$$

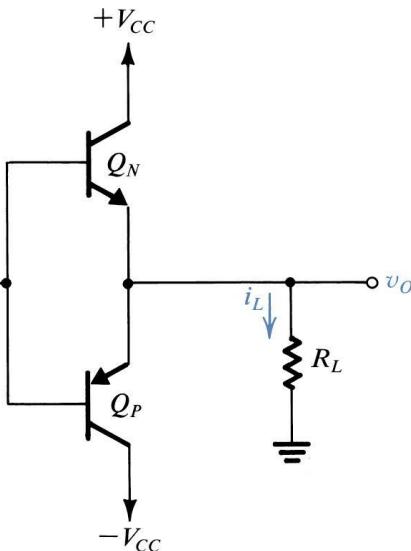
$$P_S = 2 \cdot V_{CC} \cdot I$$

$$\eta = \frac{1}{4} \cdot \frac{V_o^2}{I \cdot R_L \cdot V_{CC}} = \frac{1}{4} \cdot \left(\frac{V_o}{I \cdot R_L} \right) \cdot \left(\frac{V_o}{V_{CC}} \right)$$

$$V_o \leq V_{CC} \quad V_o \leq I \cdot R_L$$

maximum efficiency is obtained when

$$V_o = V_{CC} = I \cdot R_L$$



Biasing the Class B Output

- No DC current is used to bias this configuration.
- Activated when the input voltage is greater than the V_{be} for the transistors.
- npn Transistor operates when positive, pnp when negative.
- At a zero input voltage, we get no output voltage.

Class A

Power Conversion Efficiency

CLASS A

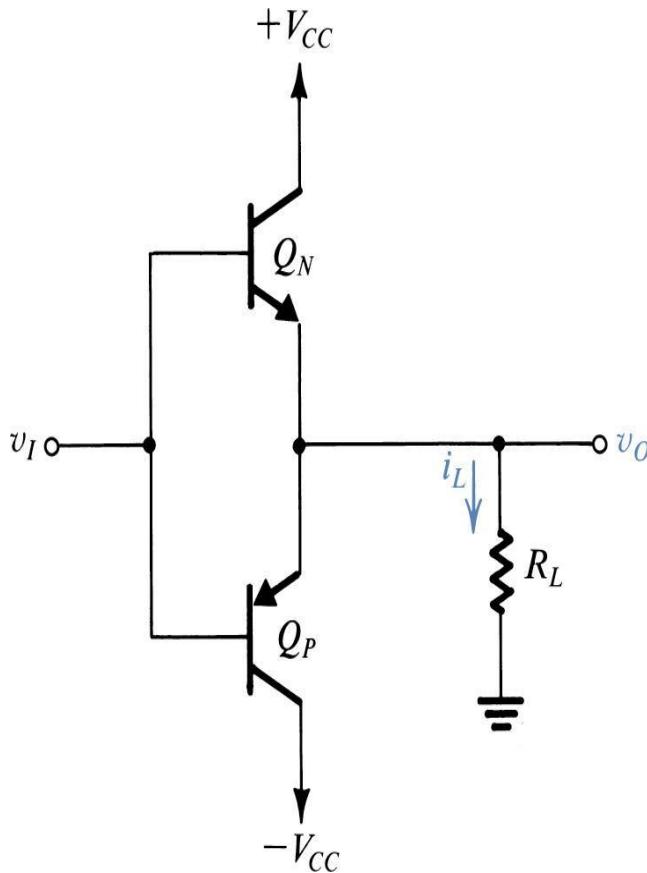
Many class A amplifiers use the same transistor(s) for both halves of the audio waveform. In this configuration, the output transistor(s) always has current flowing through it, even if it has no audio signal (the output transistors never 'turn off'). The current flowing through it is D.C.

A pure class 'A' amplifier is very inefficient and generally runs very hot even when there is no audio output. The current flowing through the output transistor(s) (with no audio signal) may be as much as the current which will be driven through the speaker load at FULL audio output power. Many people believe class 'A' amps to sound better than other configurations (and this may have been true at some point in time) but a well designed amplifier won't have any 'sound' and even the most critical 'ear' would be hard-pressed to tell one design from another.

NOTE: Some class A amplifiers use complimentary (separate transistors for positive and negative halves of the waveform) transistors for their output stage.

Class B

Circuit Operation



Class B output stage.

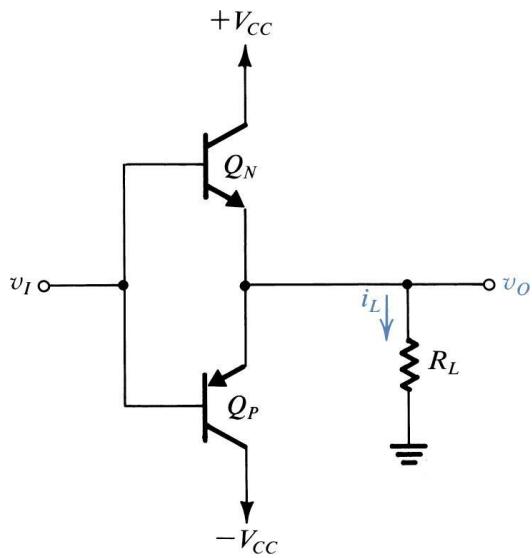
CLASS 'B'

A class 'B' amplifier uses complimentary transistors for each half of the waveform.

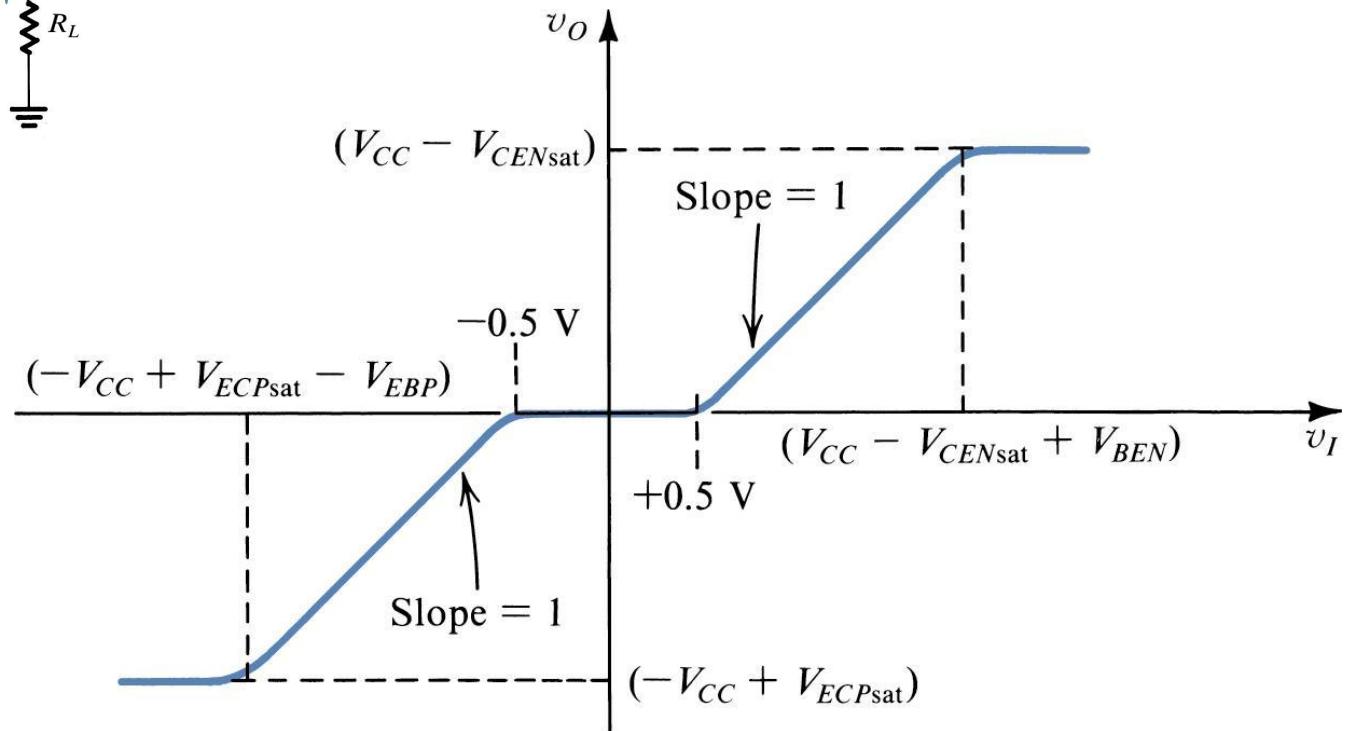
A true class 'B' amplifier is NOT generally used for audio. In a class 'B' amplifier, there is a small part of the waveform which will be distorted. You should remember that it takes approximately .6 volts (measured from base to emitter) to get a bipolar transistor to start conducting. In a pure class 'B' amplifier, the output transistors are not "biased" to an 'on' state of operation. This means that the the part of the waveform which falls within this .6 volt window will not be reproduced accurately.

The output transistors for each half of the waveform (positive and negative) will each have a .6 volt area in which they will not be conducting. The distorted part of the waveform is called 'crossover' or 'notch' distortion. Remember that distortion is any unwanted variation in a signal (compared to the original signal). The diagram below shows what crossover distortion looks like.

Class B

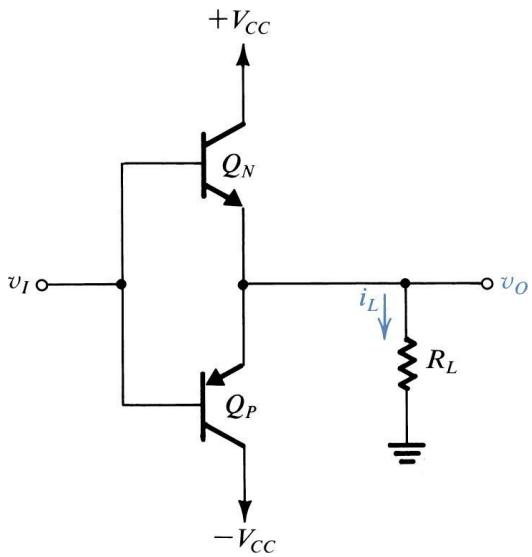


Circuit Operation



Transfer characteristic for the class B output stage in Fig. 9.5.

Operation



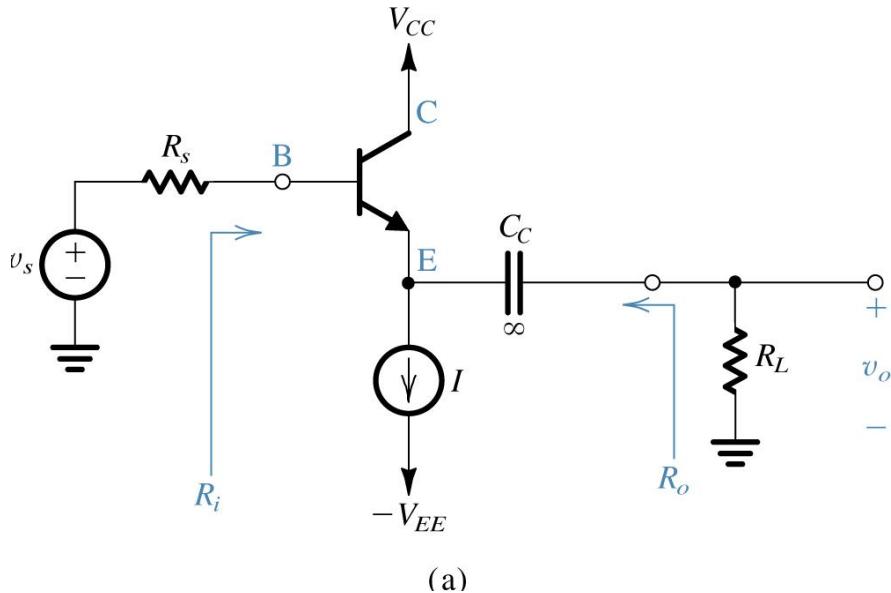
When the input voltage rises to be large enough to overcome the V_{be}, it will begin to cause an output voltage to appear. This occurs because Q_N begins to act like an emitter follower and Q_P shuts off. The input will be followed on the emitter until the transistor reaches saturation. The maximum input voltage is equal to the following:

$$v_{imax} = V_{CC} - V_{CENsat}$$

The same thing will begin to happen if the input voltage is negative by more than the V_{eb} of the transistor. This causes the Q_P to act like an emitter follower and Q_N turns off. This will continue to behave this way until saturation occurs at a minimum input voltage of:

$$v_{imin} = -V_{CC} + V_{ECPsat}$$

Emitter Follower Configuration (Chapter 4)



$$\frac{v_b}{v_s} = \frac{(\beta + 1)(r_e + par(R_L, r_o))}{R_S + (\beta + 1)(r_e + par(R_L, r_o))}$$

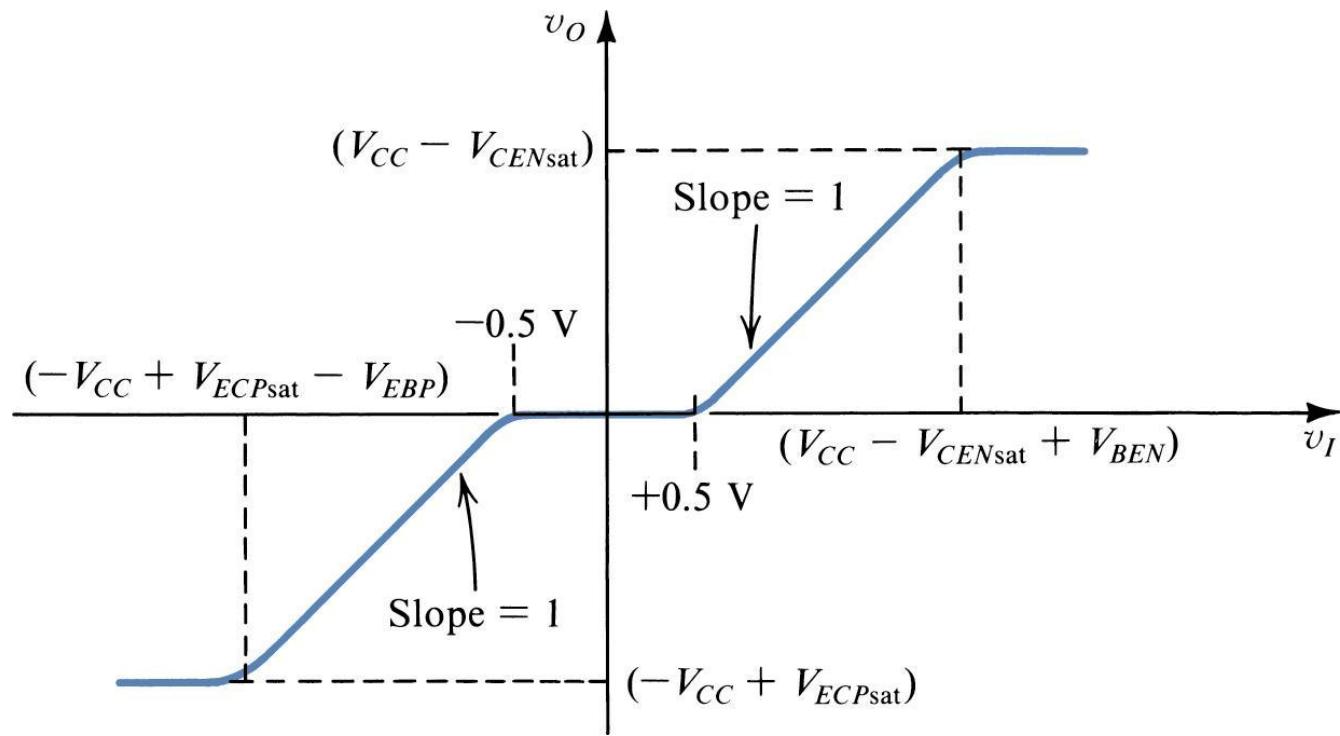
$$\frac{v_o}{v_b} = \frac{par(r_o, R_L)}{r_e + par(r_o, R_L)}$$

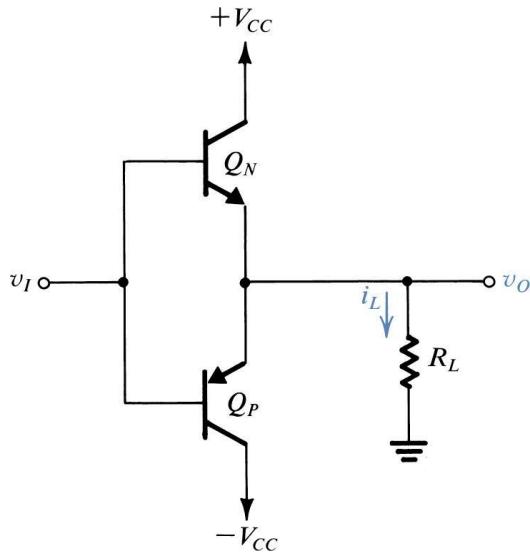
R_s will be small for most configurations, so the v_b/v_s will be a little less than unity. The same is true for r_e , so v_o/v_b will be a little less than unity making our v_o/v_s a little less than unity.

Characteristics of the Emitter Follower:

- High Input Resistance
- Low Output Resistance
- Near Unity Gain

Transfer Characteristic





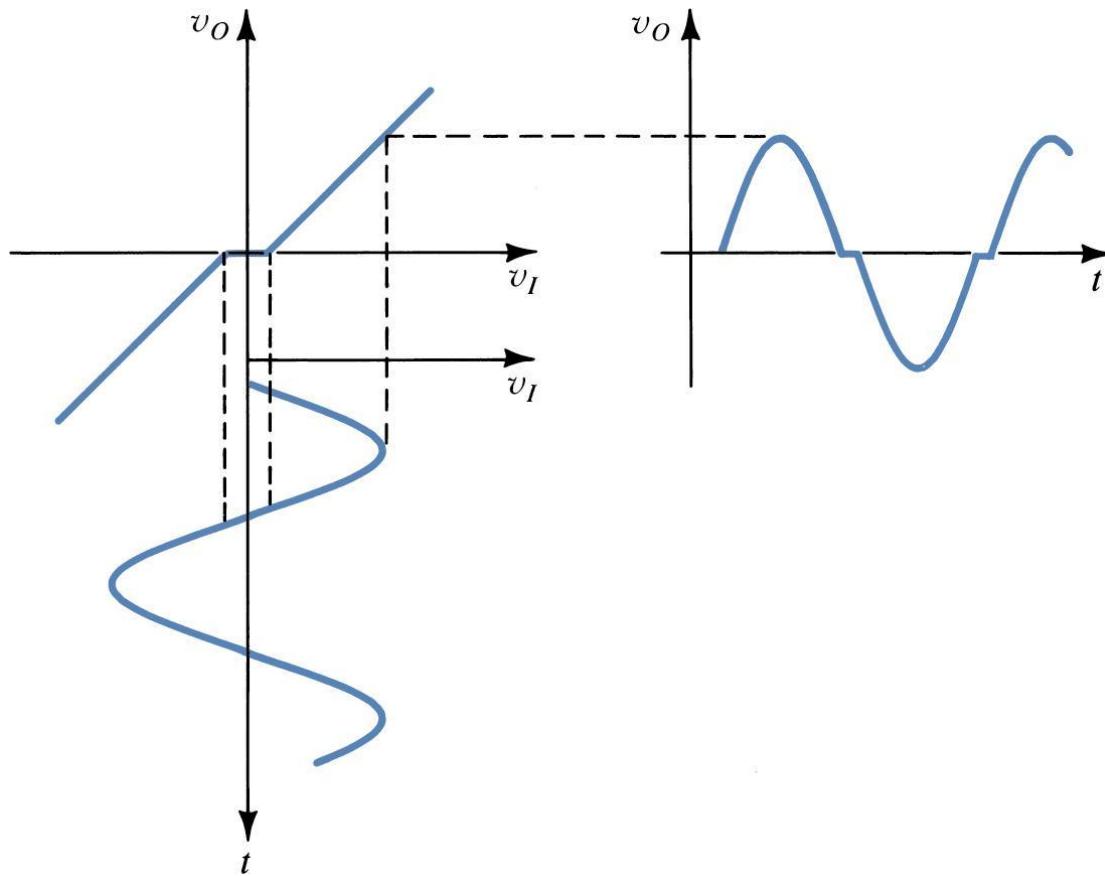
Push-Pull Nature of Class B

- Push: The npn transistor will push the current to ground when the input is positive.
- Pull: The pnp transistor will pull the current from the ground when the input is negative.

Crossover Distortion

The Crossover Distortion is due to the dead band of input voltages from $-.5V$ to $.5V$. This causes the Class B output stage to be a bad audio amplifier. For large input signals, the crossover distortion is limited, but at small input signals, it is most pronounced.

Graph of Crossover Distortion



Illustrating how the dead band in the class B **transfer** characteristic results in crossover distortion.

Power Efficiency

Load Power:

$$P_L = \frac{1}{2} \cdot \frac{V_{op}^2}{R_L}$$

Since each transistor is only conducting for one-half of the time, the power drawn from each source will be the same.

$$P_S = \frac{1}{\pi} \cdot \frac{V_{op}}{R_L} \cdot V_{CC}$$

This efficiency will be at a max when V_{op} is at a max. Since V_{op} cannot exceed V_{CC} , the maximum efficiency will occur at $\pi/4$.

$$\eta = \frac{P_L}{2 \cdot P_S} = \frac{\frac{1}{2} \cdot \frac{V_{op}^2}{R_L}}{2 \cdot \frac{1}{\pi} \cdot \frac{V_{op}}{R_L} \cdot V_{CC}}$$

$$\eta = \frac{\pi}{4} \cdot \frac{V_{op}}{V_{CC}}$$

$$\eta_{max} = \frac{\pi}{4}$$

This will be approximately 78.5%, much greater than the 25% for Class A.

Class AB

Circuit Operation

Crossover distortion can be eliminated by biasing the transistors at a small, non-zero current.

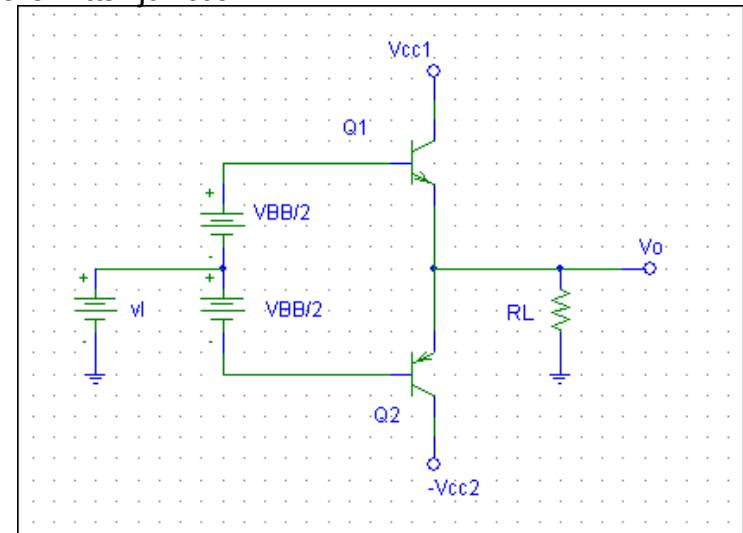
A bias Voltage V_{BB} is applied between Q_n and Q_p .

For $v_i = 0$, $v_o = 0$, and a voltage $V_{BB}/2$ appears across the base-emitter junction of each transistor.

$$i_N = i_P = I_Q = I_S \cdot e^{\frac{V_{BB}}{2 \cdot V_T}}$$

V_{BB} is selected to result the required quiescent current I_Q

$$v_o = v_i + \frac{V_{BB}}{2} - v_{BEN}$$



$$i_N = i_P + i_L$$

$$v_{BEN} + v_{EBP} = V_{BB}$$

$$V_T \cdot \ln\left(\frac{i_N}{I_S}\right) + V_T \cdot \ln\left(\frac{i_P}{I_S}\right) = 2 \cdot V_T \cdot \ln\left(\frac{i_Q}{I_S}\right)$$

$$i_N^2 = I_Q^2$$

$$i_N^2 - i_L \cdot i_N - I_Q^2 = 0$$

Class AB

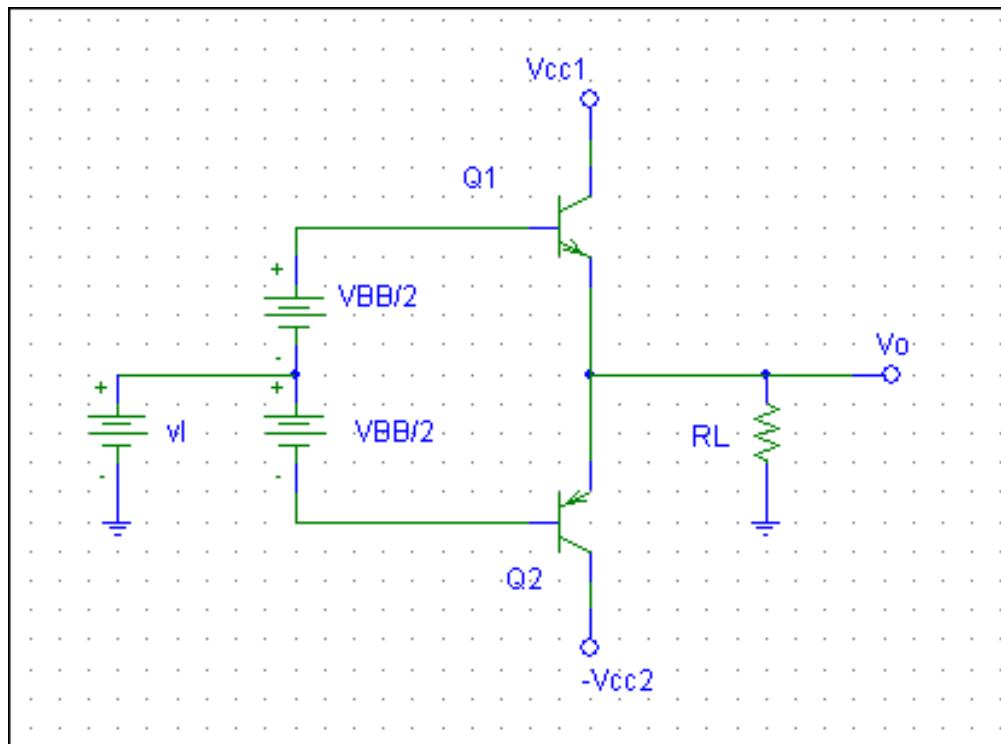
Exercise 9.6

Calvin College - ENGR 332 Class AB Output Stage Amplifier

Consider the class AB circuit (illustrated below) with $V_{cc} = 15 \text{ V}$, $I_Q = 2 \text{ mA}$, $R_L = 100 \text{ ohms}$. Determine V_{BB} . Determine the values of i_L , i_N , i_P , v_{BEN} , v_{EBP} , v_l , v_O/v_l , R_{out} , and v_o/v_i versus v_O for v_O varying from -10 to 10V.

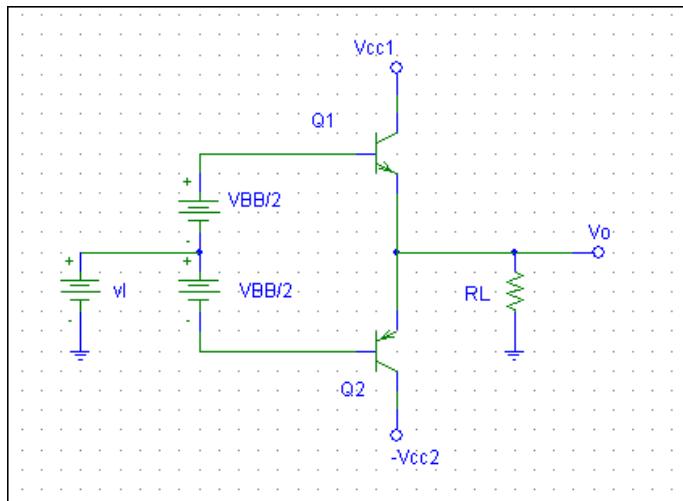
Note that v_O/v_l is the large signal voltage gain and v_o/v_i is the incremental gain obtained as $R_L/(R_L+R_{out})$. The incremental gain is equal to the slope of the transfer curve.

Assume QN and QP to be matched, with $I_S = 10E-13$.



Class AB

Exercise 9.6



under quiescent conditions $i_N = i_P = I_Q \quad v_O = v_I = 0$

Solving for V_{BB}

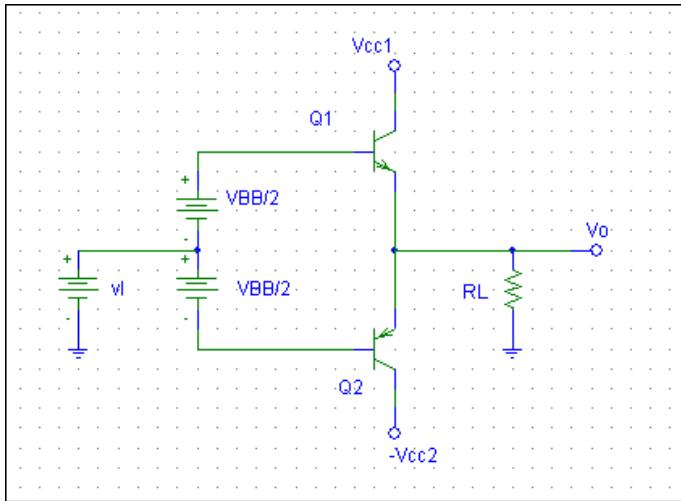
$$V_{BB} := 1 \quad IS := 10^{-13} \quad VT := 0.025 \quad I_Q := 2 \cdot 10^{-3} \quad RL := 100$$

Given
$$\frac{|V_{BB}|}{2} = \frac{VT}{I_Q} = IS \cdot e$$

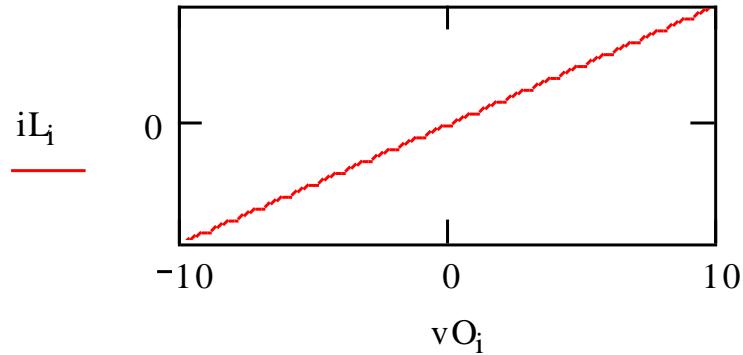
$$V_{BB} := \text{Find}(V_{BB}) \quad i := 0..100 \quad V_{BB} = 1.186$$

Class AB

Exercise 9.6



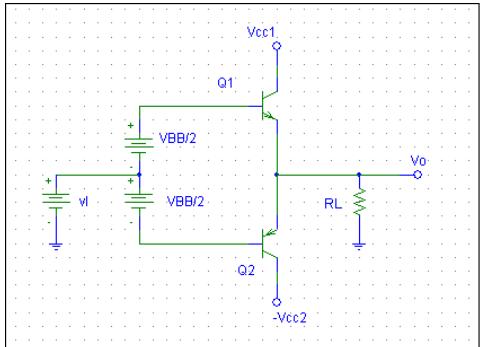
$$v_{O_i} := -10 + \frac{i}{5} \quad iL_1 := \frac{v_{O_i}}{RL}$$



Class AB

Solving for iN

Exercise 9.6



initial guesses $iN := 0.02$

$iLD := 0.02$

$IQ := 0.002$

Given

$$iN^2 - iLD \cdot iN - IQ^2 = 0$$

$$iNN(iQ, iLD) := \text{Find}(iN)$$

$$i := 0..100 \quad IQ_i := 0.002$$

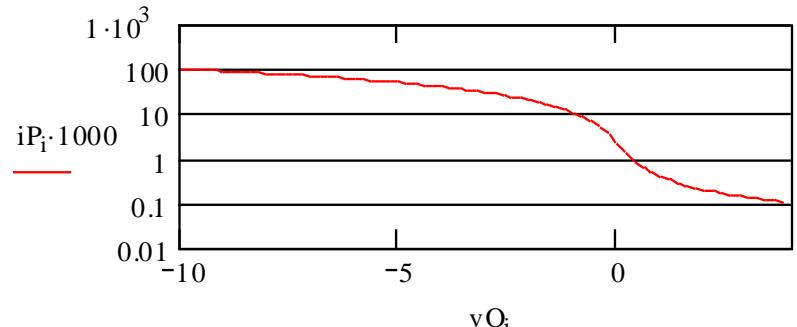
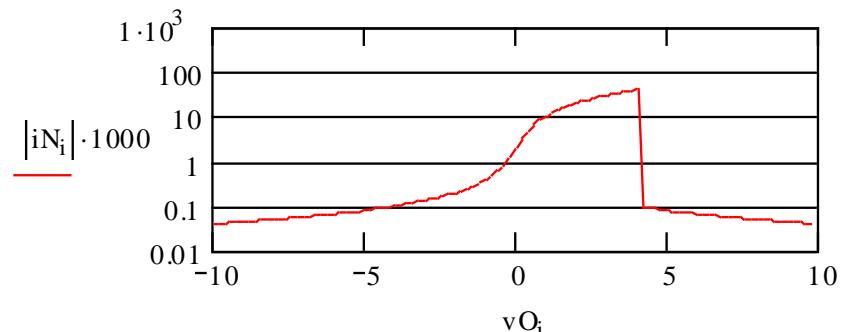
$$iLD_i := iL_i$$

$$iN_i := iNN(Q_i, L_i)$$

$$iN_{10} = 4.997 \times 10^{-5}$$

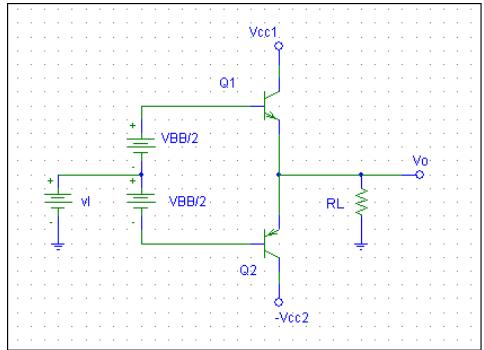
$$iP_i := iN_i - iLD_i$$

$$vBEN_i := VT \cdot \ln\left(\frac{iN_i}{IS}\right)$$



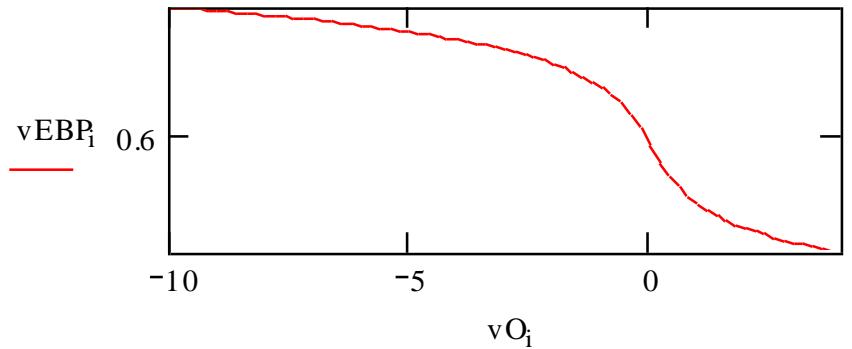
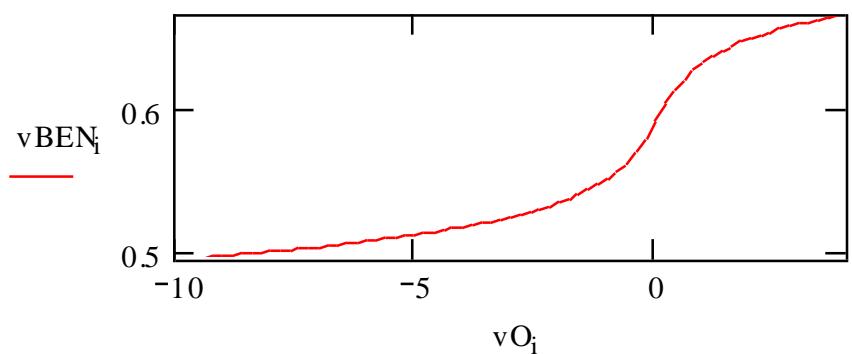
Class AB

Exercise 9.6



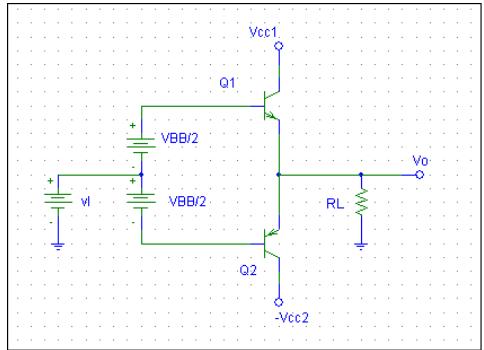
$$v_{BEN_i} := VT \cdot \ln \left(\frac{iN_i}{IS} \right)$$

$$v_{EBP_i} := VT \cdot \ln \left(\frac{iP_i}{IS} \right)$$



Class AB

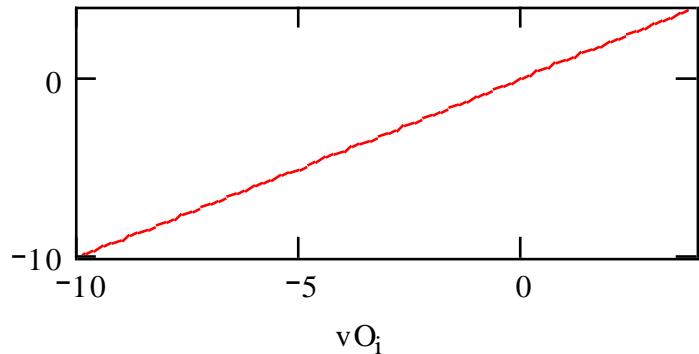
Exercise 9.6



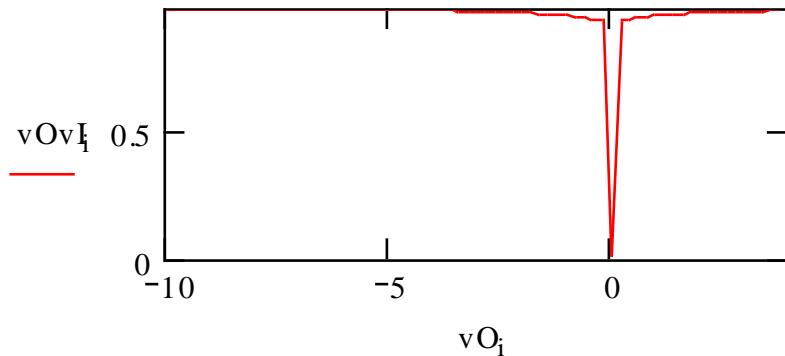
$$vI_i := vO_i + vBEN_i - \frac{VBB}{2}$$

$$vOvI_i := \frac{vO_i}{vI_i}$$

vI_i



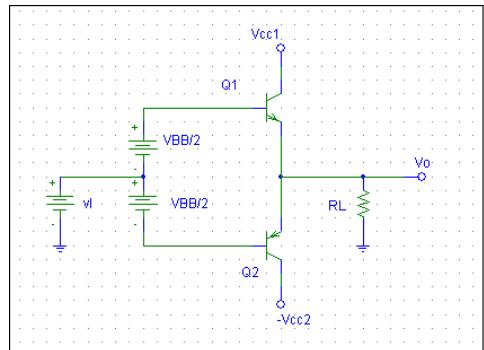
vOvI_i



Class AB

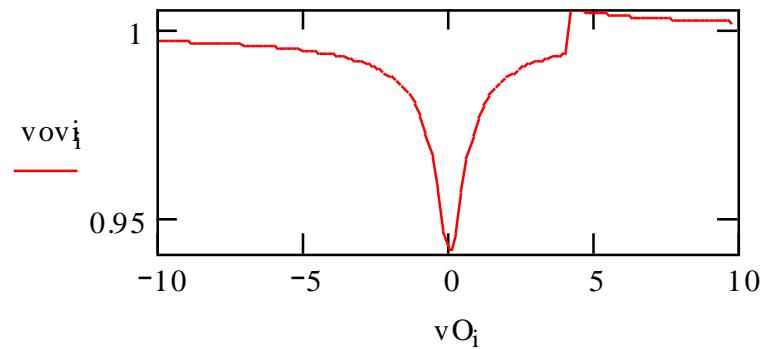
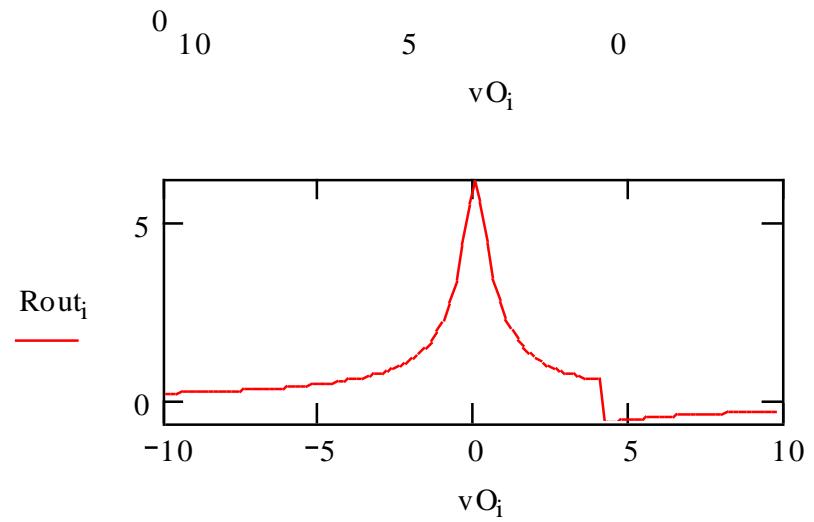
$v_{OvI} \approx 0.5$

Exercise 9.6



$$R_{out_i} := \frac{V_T}{iP_i + iN_i}$$

$$v_{ov_i} := \frac{RL}{RL + R_{out_i}}$$



Class E/F Amplifiers

Ali M Niknejad

Excerpts from the Ph.D. dissertation of Dr. Scott Kee

**The Class E/F Family of Harmonic-Tuned
Switching Power Amplifiers**

Normalized Output Power

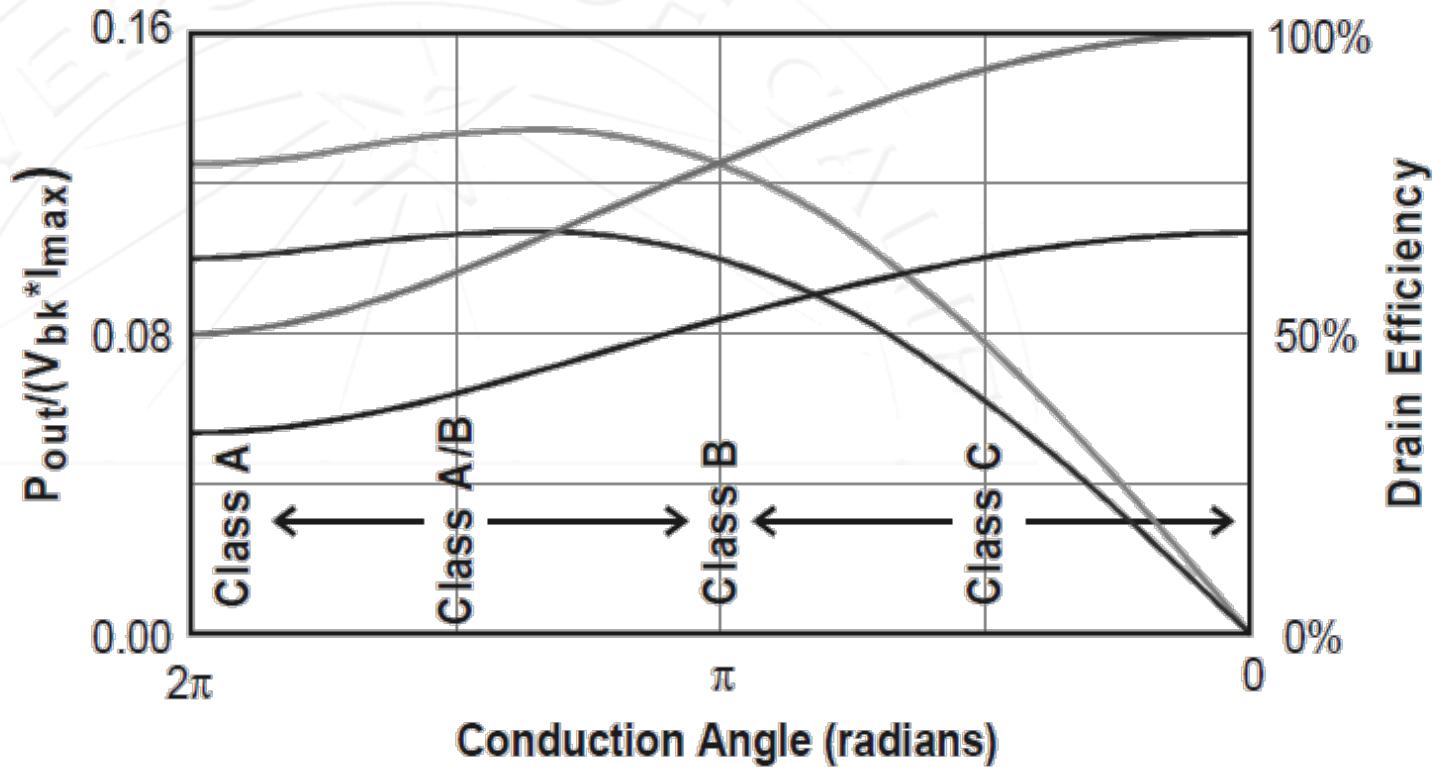
- It's easy to show that for Class A/B/C amplifiers, the efficiency and output power are given by:

$$\eta_D = \frac{1}{2} \cdot \left[\frac{1 - V_k/V_{bk}}{1 + V_k/V_{bk}} \right] \cdot \left[\frac{\alpha - \sin\alpha}{2\sin(\alpha/2) - \alpha\cos(\alpha/2)} \right]$$

$$P_{out} = \frac{1}{8\pi} \cdot \left[\frac{\alpha - \sin\alpha}{1 - \cos\alpha/2} \right] \cdot \left[1 - \frac{V_k}{V_{bk}} \right] \cdot V_{bk} \cdot I_{max}$$

- It's useful to normalize the output power versus the product of V_{bk} and I_{max} (I_{dc})

Class A/B/C

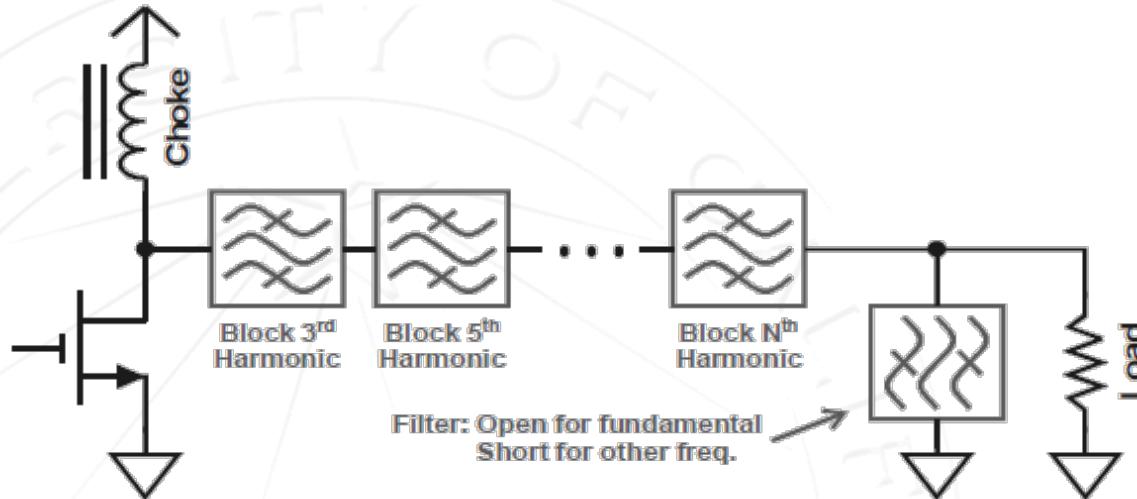


- As efficiency improves, the normalized output power drops from ~10% down to 0%

Class A/B/C Properties

- Keep voltage waveform sinusoidal → amplitude is limited to $V_{dd}/2$
- Only way to improve efficiency is to control current
- Require very large “on” current to deliver power

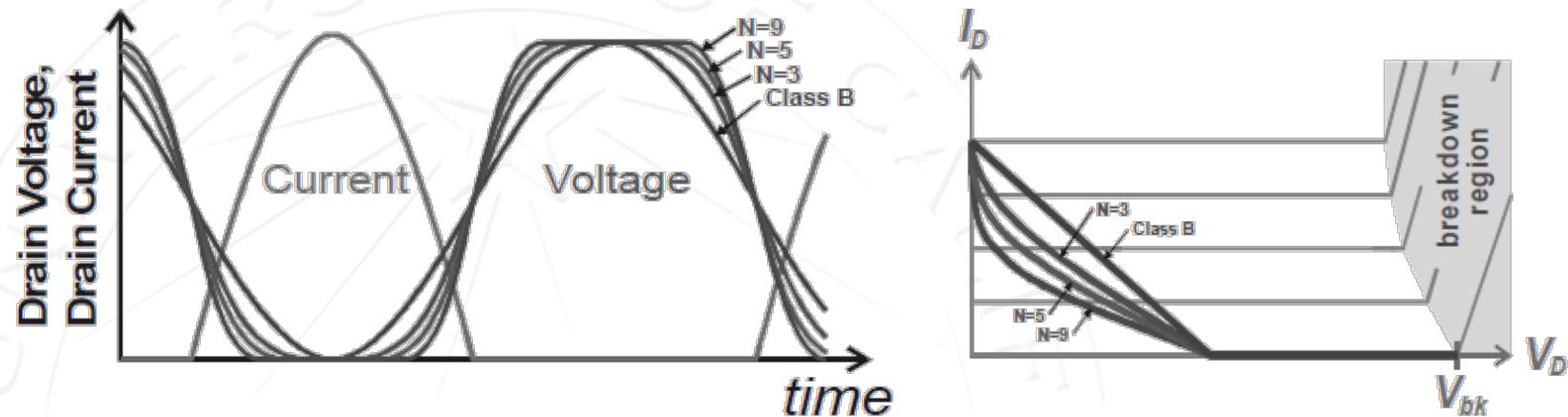
Class F



Class-F circuit conceptual implementation.

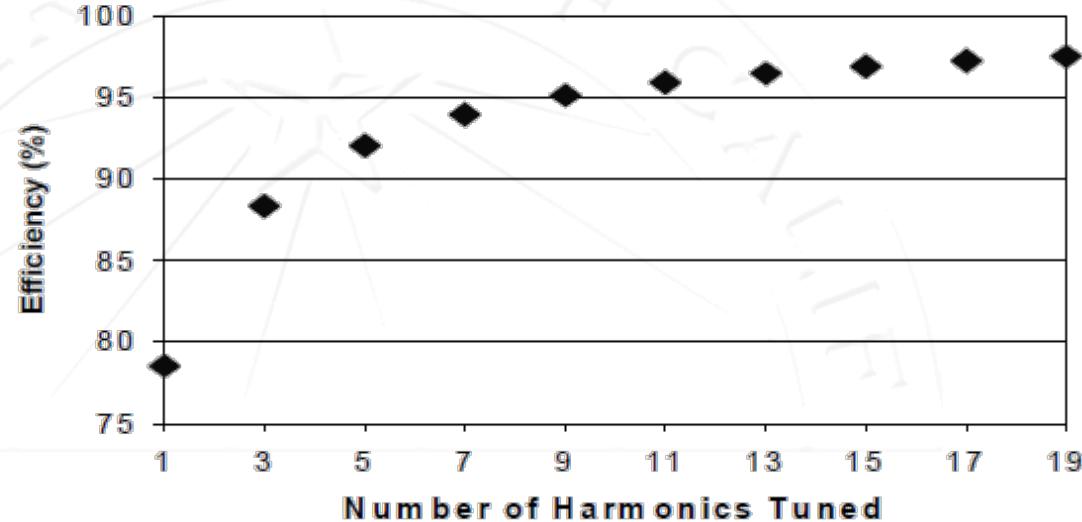
- Start with Class B current waveform → only odd harmonics
- Tune impedance at odd harmonics to be an open circuit to dissipate no harmonic power but allow odd harmonics in voltage waveform
- Tune even harmonics to short circuit to avoid dissipating power

Class F Waveforms



- Maximally flat Class F waveforms.
- An ideal Class F amplifier has a square voltage waveform and 100% efficiency.

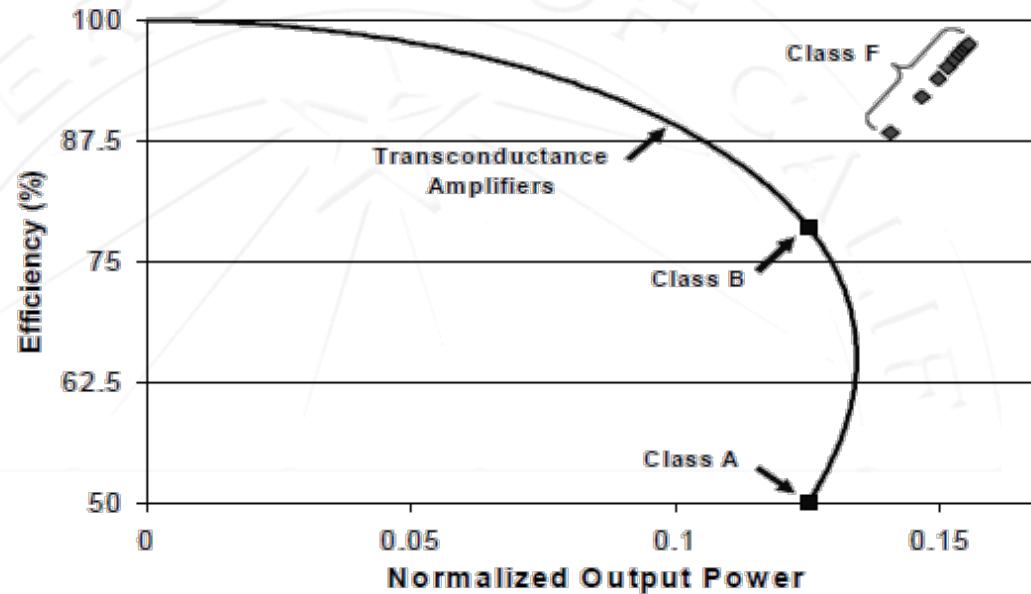
Class F Efficiency



Efficiency of maximally-flat voltage waveform class-F amplifiers with half-sinusoidal current for various numbers of voltage harmonics tuned.

- In theory, if you can control an infinite number of harmonics, efficiency approaches 100%

Class F Output Power



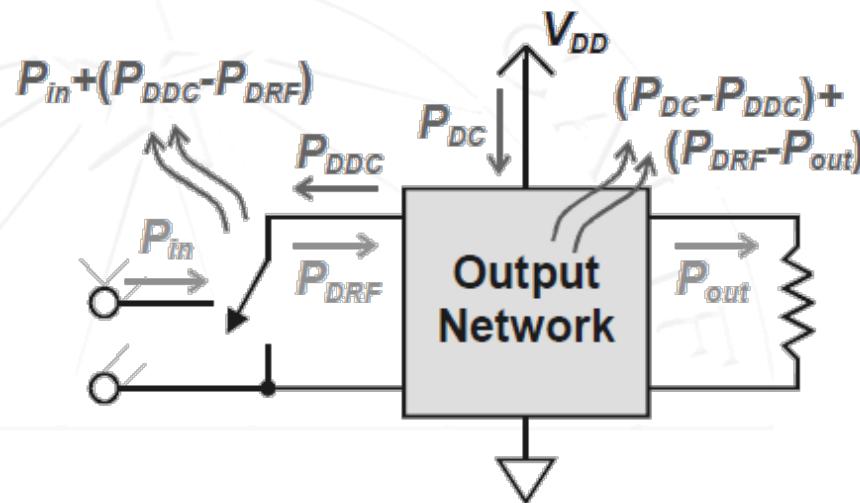
Efficiency vs. normalized output power for transconductance and class-F amplifiers. The output power is normalized to constant peak voltage and peak current.

- Square wave has a peak fundamental $4/\pi$ larger than the peak
→ 1 dB output power enhancement

Class F Disadvantages

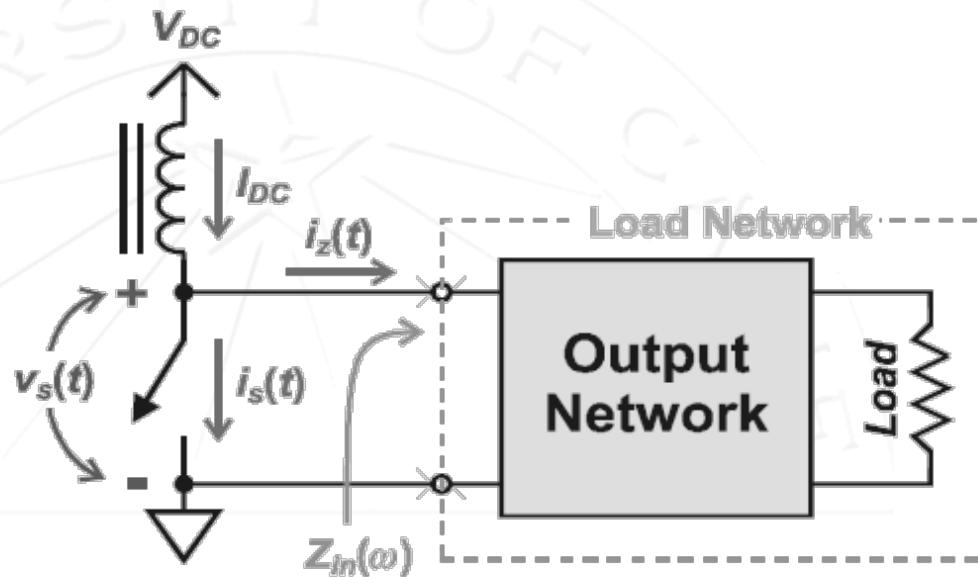
- Output capacitance of device not naturally absorbed into network → need inductor to tune it out
- Difficult to control more than 5th harmonic ... resonators are lossy and additional losses present diminishing returns on efficiency.

Switching Amplifiers



- Operate transistor in “triode” region where it acts like a switch.
- For an ideal switch the power dissipated in the switch is zero, right?
- Are all switching PA’s the same?

Linear Time-Varying Systems

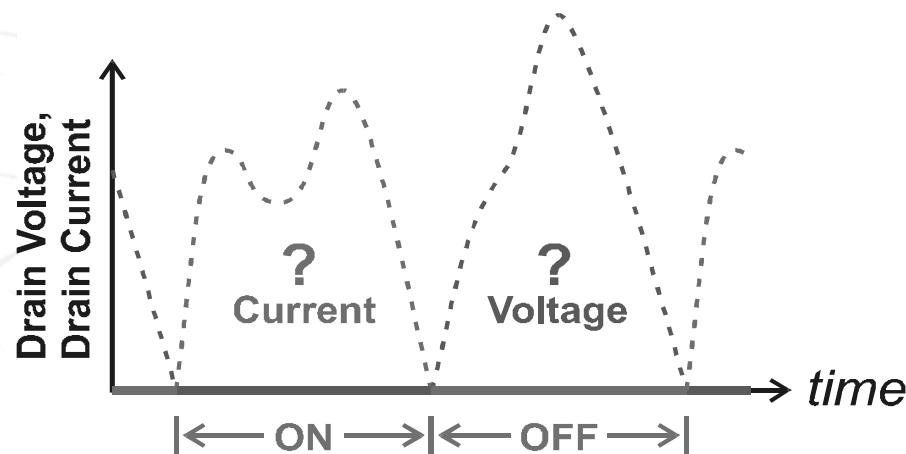


- Even though transistor is non-linear, the operation of the periodic switching action can be modeled as a linear time-varying (periodic) system. The design of the output network completely determines the behavior of the circuit.

I-V Solution for Switching Amps

$$(\theta \in D) \Rightarrow (v_s = 0)$$

$$(\theta \in \bar{D}) \Rightarrow (i_s = 0)$$



- For trans-conductance amplifiers, the current is known, so the voltage is determined by the load network.
- In a switching amplifier, when the switch is on, the voltage is forced to zero, and the current through the switch can take on any value. Likewise, when the switch is off, the switch current is zero, but the voltage can take on any value

Impedance at Harmonics

$$v_s(\theta) = V_{DC} + \sum_{k=1}^{\infty} v_k \cos(k\theta + \alpha_k)$$

$$i_s(\theta) = I_{DC} + \sum_{k=1}^{\infty} i_k \cos(k\theta + \beta_k)$$

$$\underbrace{(v_k/i_k)e^{j(\alpha_k - \beta_k)}}_{\forall k \in \{1, 2, 3, 4, \dots\}} = Z_{in}(k)$$

- The waveform shape, therefore, is completely determined by the load network impedance (it's a linear system viewed from this perspective)

Inverse Class of Operation

- By duality, any PA can be transformed into its dual (where the role of current/voltage are switched) by imposing the complementary admittance condition
- For instance a Class D voltage switching amplifier can be transformed into a current switching amp

$$i_s(\theta) = I_{DC} + \sum_{k=1}^{\infty} i_k \cos(k\theta + \alpha_k)$$

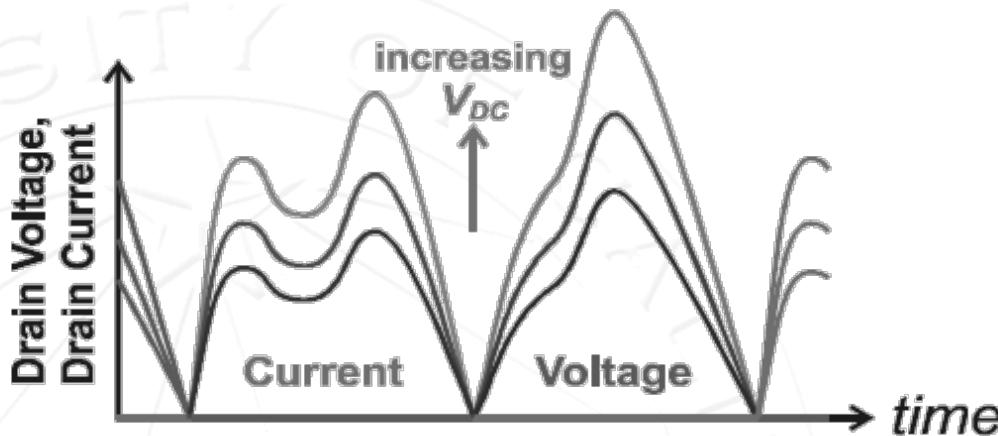
$$v_s(\theta) = V_{DC} + \sum_{k=1}^{\infty} v_k \cos(k\theta + \beta_k)$$

$$(\theta \in D) \Rightarrow (i_s = 0)$$

$$(\theta \in \bar{D}) \Rightarrow (v_s = 0)$$

$$\underbrace{(i_k/v_k)e^{j(\alpha_k - \beta_k)}}_{\forall k \in \{1, 2, 3, 4, \dots\}} = Y_{in}(k)$$

Bias Scaling



$$V_{DC} \rightarrow \lambda V_{DC} \Rightarrow \begin{cases} i_s(\theta) \rightarrow \lambda i_s(\theta) \\ v_s(\theta) \rightarrow \lambda v_s(\theta) \end{cases} \quad \hat{i}(\theta) = \lambda i_s(\theta) \quad \hat{v}_s(\theta) = \lambda v_s(\theta)$$

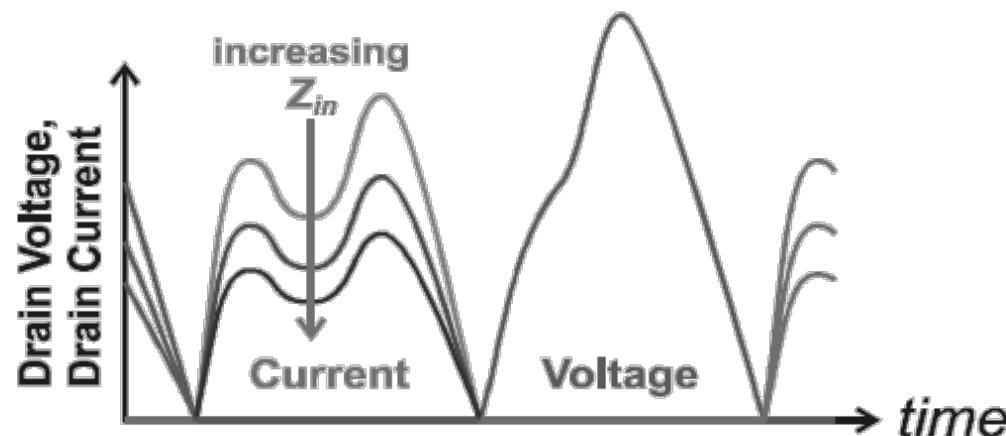
- Scaling supply changes voltage/current waveforms by the same scale factor.

Impedance Scaling

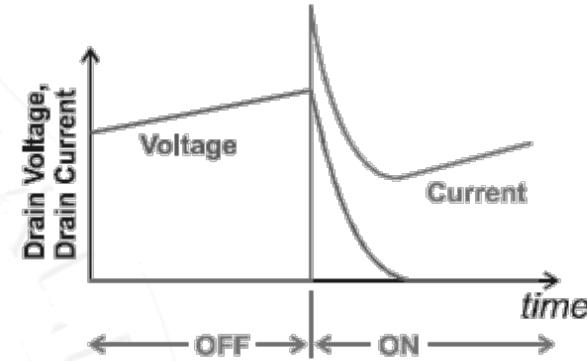
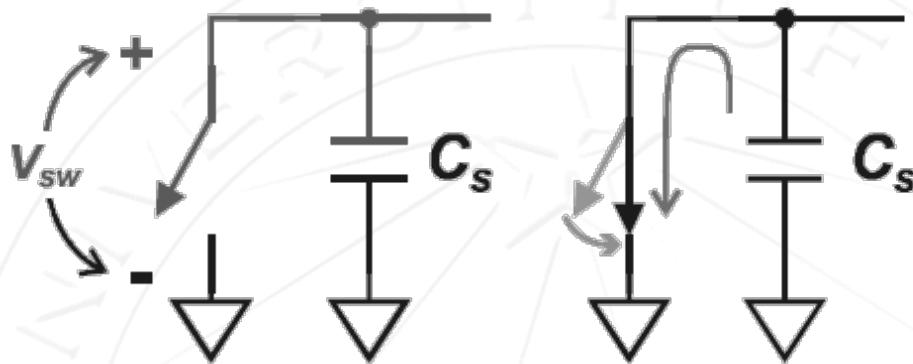
$$\hat{i}(\theta) = \hat{i}_s(\theta)/\lambda$$

$$\hat{v}_s(\theta) = v_s(\theta)$$

$$\underbrace{Z_{in}(k) \rightarrow \lambda \cdot Z_{in}(k)}_{\forall k \in \{1, 2, 3, \dots\}} \Rightarrow \begin{cases} i_s(\theta) \rightarrow i_s(\theta)/\lambda \\ v_s(\theta) \rightarrow v_s(\theta) \end{cases}$$



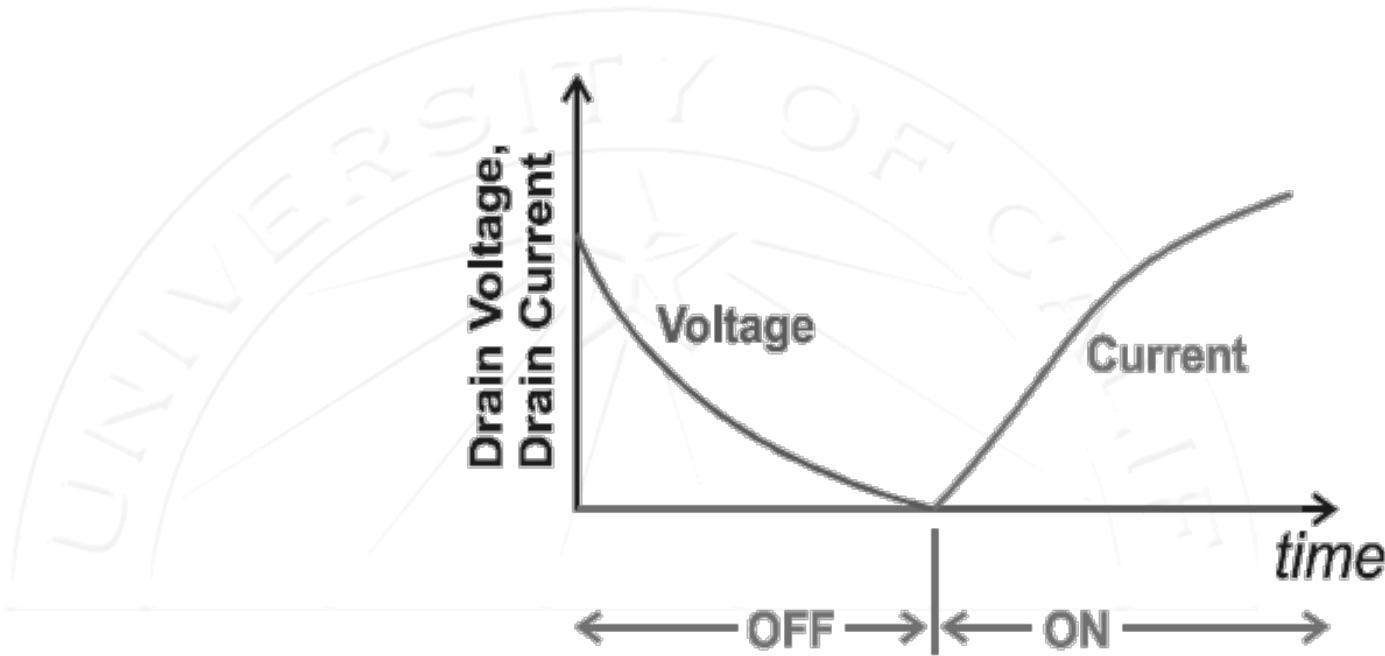
Switch Losses



$$P_{sw} = \frac{1}{2} C_s V_{sw}^2 f_0$$

- When a switch is closed across a capacitor, an impulse of current flows through the switch to discharge the capacitor. The energy stored in the capacitor is dissipated into heat through the switch. (ideal switch?)
- If you make a smaller switch, the on-resistance goes down so you have to live with finite capacitance.

ZVS



- A ZVS network will return the voltage to zero at the moment of switch turn-on. To make the circuit more robust, the derivative of the voltage can also be forced to zero (or n -derivatives ...) to obtain a maximally flat zero.

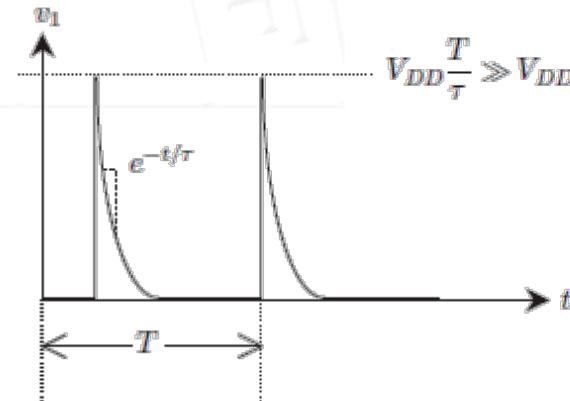
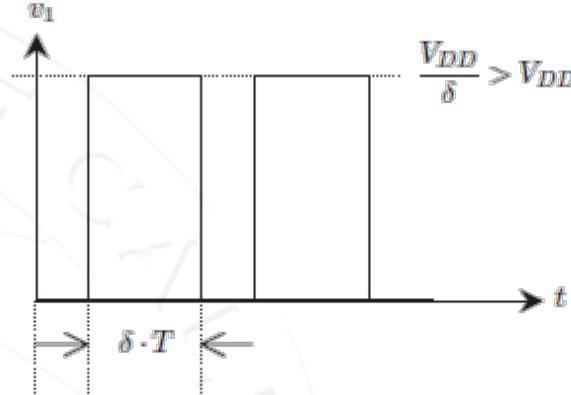
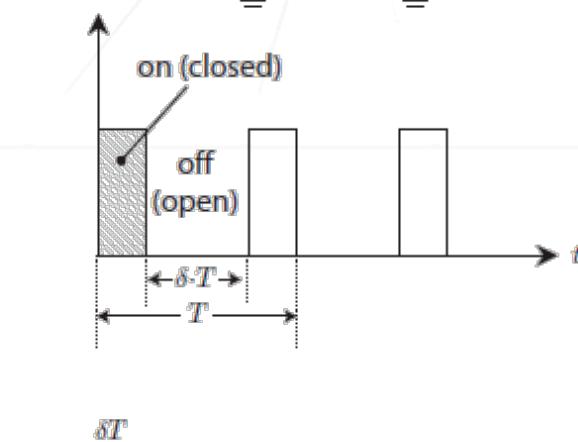
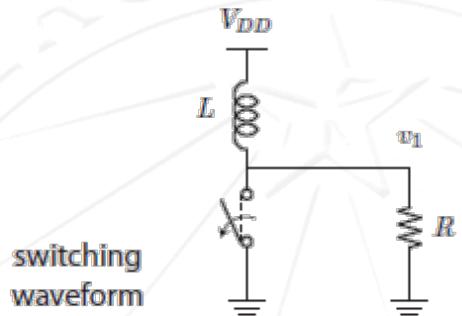
Switch Losses: ZCS Condition

- The dual of ZVS is ZCS.
- What happens if you open circuit an inductor with current (flux)? The energy stored in the magnetic flux is dissipated
- In practice the voltage “kick” produced by the inductor will break down the switch and conduct current.
- It’s also possible to design a load network that returns the current to zero just before the switch opens.

$$P_{lost} = f \cdot E_m = \frac{1}{2} L I_m^2 f$$

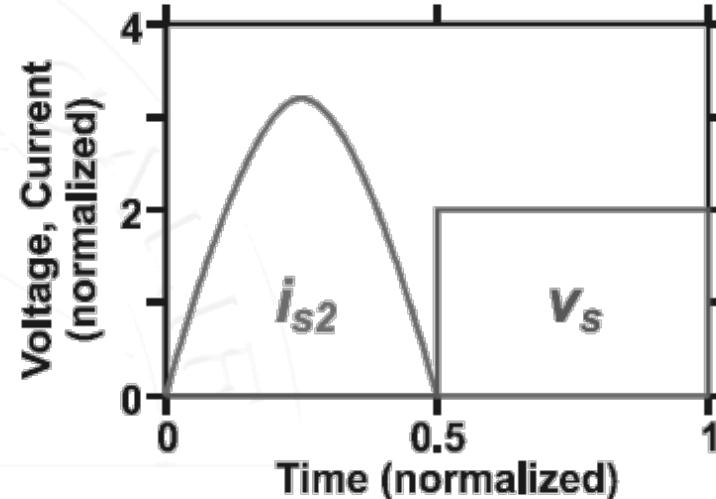
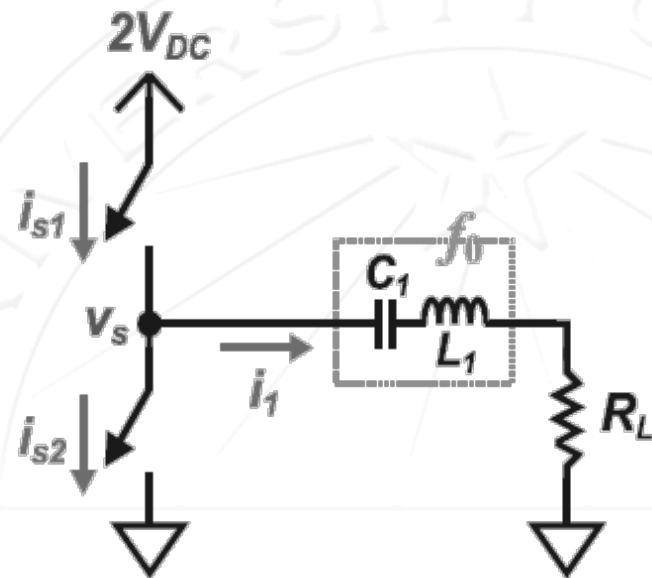
Switching Inductors

$$V_L(t) = \frac{d\psi}{dt} = L \frac{dI}{dt}$$



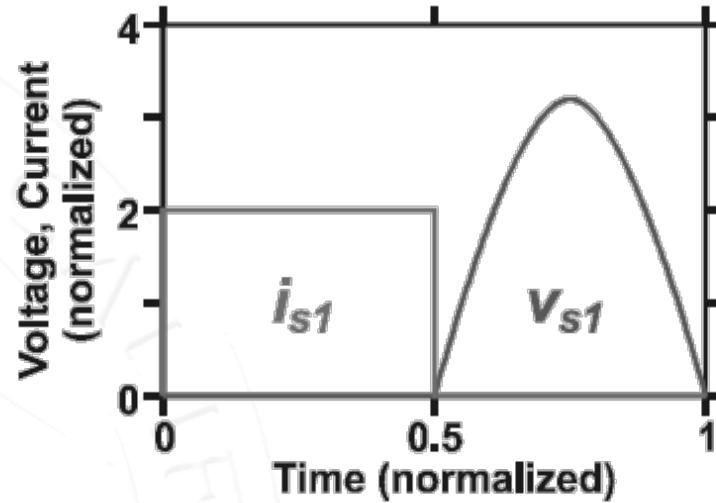
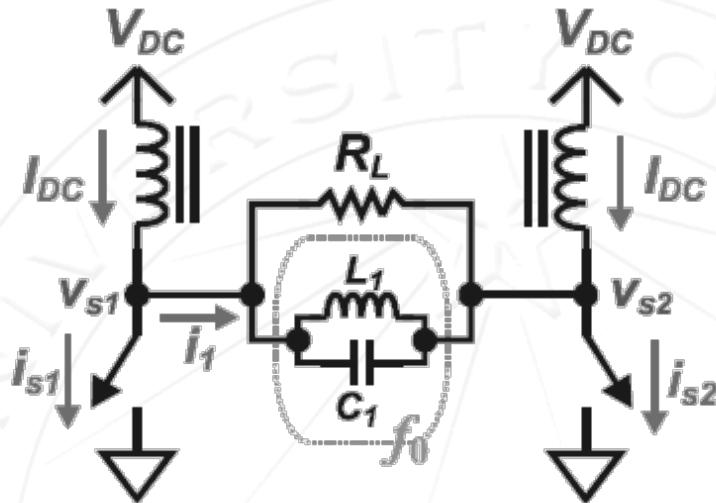
- If the inductor is large enough, its switching behavior can be idealized.

Class D



- Two switches used to realize square waveform.
- Series tank only allows fundamental current to flow into load.
- Switch capacitance limits efficiency in high frequency applications.

Class D-1

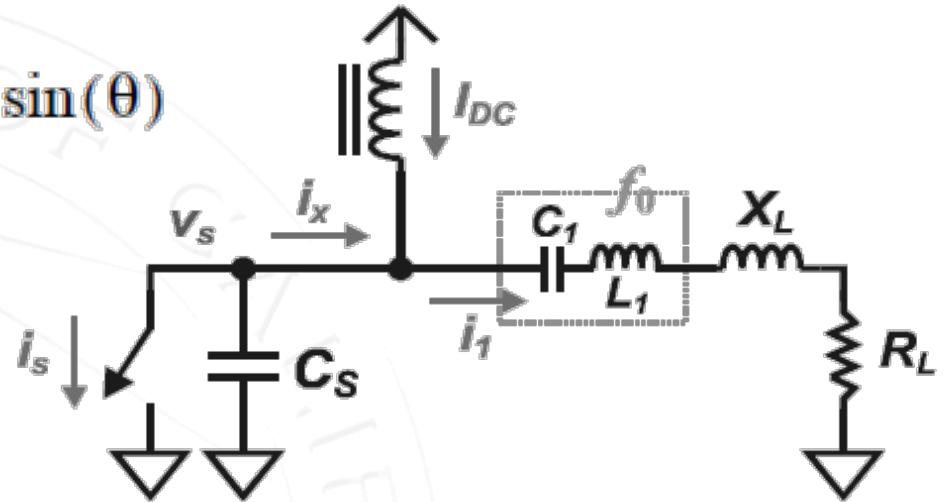


- The “Dual” Class D amplifier (interchange voltage/current → square wave current, sinusoidal voltage, parallel LCR filter)
- Chokes act like current sources. ZVS by “design” but only if there is no device capacitance to begin with.

$$v_{s1} = \begin{cases} 0 & 0 < \theta < \pi \\ -\frac{4}{\pi} I_{DC} R_L \sin(\theta) & \pi < \theta < 2\pi \end{cases}$$

Class E

$$i_x = -I_{DC} + \alpha \cos(\theta) + \beta \sin(\theta)$$



- Switch driven with 50% duty cycle. Device capacitance C_s absorbed into network.
- The current i_1 is sinusoidal and the current through the choke is DC. The sum of these currents flows through the switch + capacitor.

Class E Currents

$$i_s = \begin{cases} I_{DC} - \alpha \cos(\theta) - \beta \cos(2\theta) & 0 < \theta < \pi \\ 0 & \pi < \theta < 2\pi \end{cases}$$

$$i_{cs} = \begin{cases} 0 & 0 < \theta < \pi \\ I_{DC} - \alpha \cos(\theta) - \beta \cos(2\theta) & \pi < \theta < 2\pi \end{cases}$$

- When switch is closed, all the current flows through it. When open, this same current must flow through the capacitor. The voltage across the capacitor is given by the integral of the current since

$$\frac{dv_s}{dt} = \frac{1}{C_s} \cdot \begin{cases} 0 & 0 < \theta < \pi \\ I_{DC} - \alpha \cos(\theta) - \beta \cos(2\theta) & \pi < \theta < 2\pi \end{cases}$$

Class E Voltages

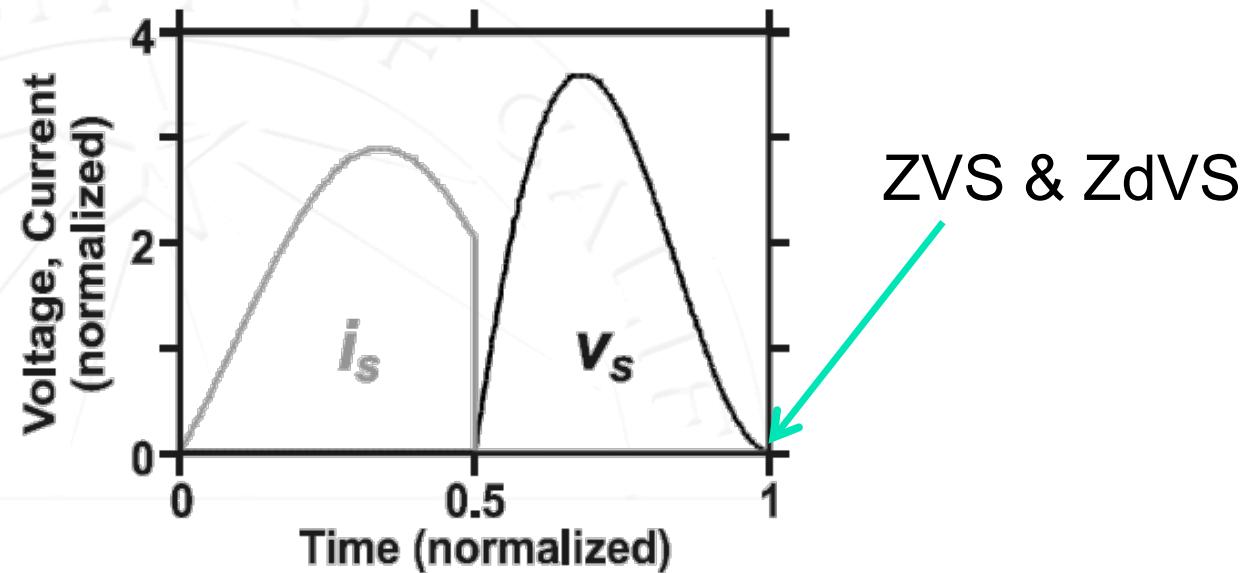
$$v_s = \frac{1}{2\pi f_0 C_s} \cdot \begin{cases} I_{DC} \cdot (\theta - \pi) - \alpha \sin(\theta) + \beta [\cos(\theta) + 1] & 0 < \theta < \pi \\ 0 & \pi < \theta < 2\pi \end{cases}$$

$$\alpha = I_{DC}$$

$$\beta = -\frac{\pi}{2} \cdot I_{DC}$$

- We can impose voltage continuity, so $\theta=\pi$. But we have two degrees of freedom, so we can also set the derivative of the voltage to zero (ZdVS). When both conditions are satisfied, we have a class E amplifier.

Class E Current/Voltage



$$v_s = \frac{I_{DC}}{2\pi f_0 C_s} \cdot \begin{cases} 0 & 0 < \theta < \pi \\ \theta - \sin(\theta) - (\pi/2)\cos(\theta) - 3\pi/2 & \pi < \theta < 2\pi \end{cases}$$

$$i_s = \begin{cases} I_{DC} - \cos(\theta) + (\pi/2)\cos(\theta) & 0 < \theta < \pi \\ 0 & \pi < \theta < 2\pi \end{cases}$$

Class E Load & Swing

- The load is given from Fourier analysis of the current/voltage.
- To realize a Class E amplifier requires an inductive load.
- One big disadvantage of the Class E amplifier is that the voltage swing across the device is very large (nearly $4 \times V_{DD}$).

Switching Amplifier Efficiency

$$P_{SW} = \frac{1}{2} \cdot C_S V_{SW}^2 f_0$$

$$P_{SW} = \frac{1}{4\pi} \cdot \frac{V_{SW}^2}{Z_{CS}}$$

$$Z_{CS} \equiv \frac{1}{2\pi f_0 C_S}$$

$$V_{DC} \equiv \frac{1}{2\pi} \cdot \int_0^{2\pi} v_s d\theta$$

$$P_{diss} \approx I_{RMS}^2 R_{on} + \frac{1}{4\pi} \cdot \frac{V_{SW}^2}{X_{CS}}$$

$$P_{DC} = V_{DC} I_{DC}$$

$$I_{DC} \equiv \frac{1}{2\pi} \cdot \int_0^{2\pi} i_s d\theta$$

$$\eta_D \equiv \frac{P_{out}}{P_{DC}} = 1 - \frac{P_{diss}}{P_{DC}}$$

$$\eta_D \approx 1 - \left(\frac{I_{RMS}^2 R_{on}}{V_{DC} I_{DC}} + \frac{V_{SW}^2}{4\pi X_{CS} V_{DC} I_{DC}} \right)$$

$$G = \frac{P_{DC} - P_{diss}}{P_{in}} = \frac{V_{DC} I_{DC} - I_{RMS}^2 R_{on}}{P_{in}}$$

- For ZVS, P_{sw} is zero

Switching Amplifier PAE

$$PAE \approx \left(1 - \frac{P_{in}}{V_{DC}I_{DC} - I_{RMS}^2 R_{on}}\right) \cdot \left[1 - \left(\frac{I_{RMS}^2 R_{on}}{V_{DC}I_{DC}} + \frac{V_{SW}^2}{4\pi X_{CS} V_{DC} I_{DC}}\right)\right]$$

$$PAE \approx \left(1 - \frac{P_{in}}{V_{DC}I_{DC}}\right) \cdot \left[1 - \left(\frac{I_{RMS}^2 R_{on}}{V_{DC}I_{DC}} + \frac{V_{SW}^2}{4\pi X_{CS} V_{DC} I_{DC}}\right)\right]$$

$$PAE \approx \left(1 - \frac{P_{in}}{V_{DC}I_{DC}}\right) \cdot \left(1 - \frac{I_{RMS}^2 R_{on}}{V_{DC}I_{DC}}\right)$$

- This result includes the gain of the amplifier. To arrive at the final result, we assume 100% drain efficiency and ZVS. Note that P_{in} is a function of the transistor size.

Switch FOM

$$P_{out} = \frac{1}{2} v_1 i_1 \cos(\alpha_1 - \beta_1)$$

- It's useful to relate the efficiency to peak current/voltage swings (stress) on the device. For a high efficiency PA we have

$$P_{out} \approx V_{DC} I_{DC}$$

$$P_{out} = (V_{pk} I_{pk}) \cdot \left(\frac{v_1 i_1 \cos(\alpha_1 - \beta_1)}{2 V_{pk} I_{pk}} \right) \quad P_{out} \approx (V_{pk} I_{pk}) \cdot \left(\frac{V_{DC} I_{DC}}{V_{pk} I_{pk}} \right)$$

$$E_P \equiv \frac{P_{out}}{V_{pk} I_{pk}} = \left(\frac{v_1}{V_{pk}} \right) \cdot \left(\frac{i_1}{I_{pk}} \right) \cdot \left[\frac{1}{2} \cos(\alpha_1 - \beta_1) \right]$$

$$E_P \equiv \frac{P_{out}}{V_{pk} I_{pk}} \approx \left(\frac{V_{DC}}{V_{pk}} \right) \cdot \left(\frac{I_{DC}}{I_{pk}} \right)$$

FOM (cont)

$$F_V \equiv V_{pk}/V_{DC}$$

$$E_P \equiv \frac{P_{out}}{V_{pk}I_{pk}} \approx \frac{1}{F_V F_{PI}}$$

$$F_{PI} \equiv I_{pk}/I_{DC}$$

$$F_C \equiv \frac{P_{out}}{V_{DC}^2/Z_C}$$

- Peak voltage versus DC
- Peak current versus DC
- RMS current versus DC
- Output power versus peak stress on transsistor
- Output power versus stored reactive power.
 - The smaller this ratio, the more the design can tolerate output capacitance, and hence a larger transistor with lower conductive losses.

Device Size Limited Maximum Drain Efficiency

$$\eta_D \approx 1 - \frac{I_{RMS}^2 R_{on}}{V_{DC} I_{DC}} \rightarrow \eta_D \approx 1 - \left[\left(\frac{I_{RMS}}{I_{DC}} \right)^2 \left(\frac{V_{pk}}{V_{DC}} \right)^2 \left(\frac{R_{on}}{V_{pk}^2} \right) (V_{DC} I_{DC}) \right]$$

$$\eta_D \approx 1 - \left[\left(\frac{I_{RMS}}{I_{DC}} \right)^2 \left(\frac{V_{pk}}{V_{DC}} \right)^2 \left(\frac{\bar{R}_{on}}{\bar{V}_{bk}^2} \right) (P_{out}) \right] \quad \eta_D \approx 1 - \left[(F_I^2 F_V^2) \frac{P_{out}}{(\bar{V}_{bk}^2 / \bar{R}_{on})} \right]$$

- All terms except the third are invariant and only depend on the tuning strategy.
- Minimize the third term by using the highest peak voltage possible (minimize current through device).

Capacitance Limited Drain Efficiency

$$\eta_D \approx 1 - \frac{I_{RMS}^2 R_{on}}{V_{DC} I_{DC}}$$

$\eta_D \approx 1 - \left[\left(\frac{I_{RMS}}{I_{DC}} \right)^2 \left(\frac{V_{DC} I_{DC}}{2\pi f_0 C_s V_{DC}^2} \right) (R_{on} C_{out}) (2\pi f_0) \left(\frac{C_s}{C_{out}} \right) \right]$

- Increase device size until the switch output capacitance equals the total output capacitance
- All terms except the last are invariant (bias, impedance scaling) and only depend on the switching network. Note the third term depends on technology but is independent of transistor size. Voltage waveform properties do not come into play.
- To minimize the final term, maximize C_{out} . Final efficiency only depends on technology RC time constant:

$$\eta_D \approx 1 - \left[\left(\frac{I_{RMS}}{I_{DC}} \right)^2 \left(\frac{V_{DC} I_{DC}}{2\pi f_0 C_s V_{DC}^2} \right) (\bar{R}_{on} \bar{C}_{out}) (2\pi f_0) \right]$$

Reactive Energy Term

- The second term needs further explanation: $\left(\frac{V_{DC}I_{DC}}{2\pi f_0 C_s V_{DC}^2} \right)$
- For a highly efficient amplifier, the numerator is equal to the output power, and the denominator has the switch capacitance admittance: $\left(\frac{P_{out}}{V_{DC}^2/Z_C} \right)$
- This is a ratio of the output power to the switch's stored reactive power. We wish to use a tuning strategy that maximizes the reactive energy of the switch and minimizes the RMS current through the switch. The voltage cannot be traded for current like the previous case.

$$\eta_D \approx 1 - \left[(F_I^2 F_C) \frac{2\pi(\bar{R}_{on} \bar{C}_{out})}{1/f_0} \right] \quad F_C \equiv \frac{P_{out}}{V_{DC}^2/Z_C}$$

Gain Limited PAE

$$PAE \approx \left(1 - \frac{\bar{P}_{in}}{P_{out}} \cdot \lambda\right) \cdot \left(1 - \frac{I_{RMS}^2 \bar{R}_{on}}{V_{DC} I_{DC}} \cdot \frac{1}{\lambda}\right)$$

$$\lambda = \sqrt{\frac{I_{RMS}^2 \bar{R}_{on} P_{out}}{V_{DC} I_{DC} \bar{P}_{in}}}$$

- For $C_s = C_{out}$, where λ is a scaling parameter. Clearly an optimal size exists since $R_{on} \sim \lambda$ whereas $P_{in} \sim 1/\lambda$.
- The optimal PAE is given by:

$$PAE \approx \left(1 - \sqrt{\frac{I_{RMS}^2 \bar{R}_{on} \bar{P}_{in}}{P_{out} V_{DC} I_{DC}}}\right)^2$$

$$PAE \approx \left[1 - \left(\frac{I_{RMS}}{I_{DC}}\right) \left(\frac{V_{pk}}{V_{DC}}\right) \left(\sqrt{\frac{\bar{P}_{in}}{V_{pk}^2 / \bar{R}_{on}}}\right)\right]^2$$

Under assumption of high drain efficiency

Gain Limited PAE (cont)

- Considering the breakdown limitations:

$$PAE \approx \left[1 - (F_I F_V) \left(\sqrt{\frac{\bar{P}_{in}}{V_{bk}^2 / \bar{R}_{on}}} \right) \right]^2$$

$$PAE \approx 1 - (2F_I F_V) \left(\sqrt{\frac{\bar{P}_{in}}{V_{bk}^2 / \bar{R}_{on}}} \right)$$

- Use a tuning network with low peak to DC current/voltage.
- Note that the final term is a scaling invariant property of transistor.

Capacitance Limited PAE

- If the optimal device is too large, it's output cap will be larger than C_s , and hence cannot be absorbed into the network. Must limit device size to C_s .

$$P_{in} = \frac{C_{out}}{\bar{C}_{out}} \cdot \bar{P}_{in} = \frac{C_s}{\bar{C}_{out}} \cdot \bar{P}_{in} \quad F_C \equiv \frac{P_{out}}{V_{DC}^2/Z_C}$$

$$PAE \approx \left(1 - \frac{C_s \bar{P}_{in}}{\bar{C}_{out} P_{out}}\right) \cdot \left[1 - (F_I^2 F_C) \left(\frac{2\pi \bar{R}_{on} \bar{C}_{out}}{1/f_0}\right)\right]$$

$$PAE \approx \left[1 - \left(\frac{\bar{P}_{in}}{\bar{C}_{out} V_{pk}^2}\right) \left(\frac{2\pi C_s V_{DC}^2}{P_{out}}\right) \left(\frac{V_{pk}}{V_{DC}}\right)^2 \left(\frac{1}{2\pi f_0}\right)\right] \cdot \left[1 - (F_I^2 F_C) \left(\frac{2\pi \bar{R}_{on} \bar{C}_{out}}{1/f_0}\right)\right]$$

Cap Limited PAE

$$PAE \approx \left[1 - \left(\frac{\bar{P}_{in}}{\bar{C}_{out} V_{pk}^2} \right) \left(\frac{F_V^2}{F_C} \right) \left(\frac{1}{2\pi f_0} \right) \right] \cdot \left[1 - (F_I^2 F_C) \left(\frac{2\pi \bar{R}_{on} \bar{C}_{out}}{1/f_0} \right) \right]$$

$$PAE \approx \left[1 - \left(\frac{\bar{P}_{in}}{\bar{C}_{out} V_{bk}^2} \right) \left(\frac{F_V^2}{F_C} \right) \left(\frac{1}{2\pi f_0} \right) \right] \cdot \left[1 - (F_I^2 F_C) \left(\frac{2\pi \bar{R}_{on} \bar{C}_{out}}{1/f_0} \right) \right]$$

- Make the peak voltage as large as possible and increase the gain.

$$G = \left(\frac{\bar{C}_{out} V_{bk}^2}{\bar{P}_{in}} \right) \left(\frac{F_C}{F_V^2} \right) (2\pi f_0)$$

Summary

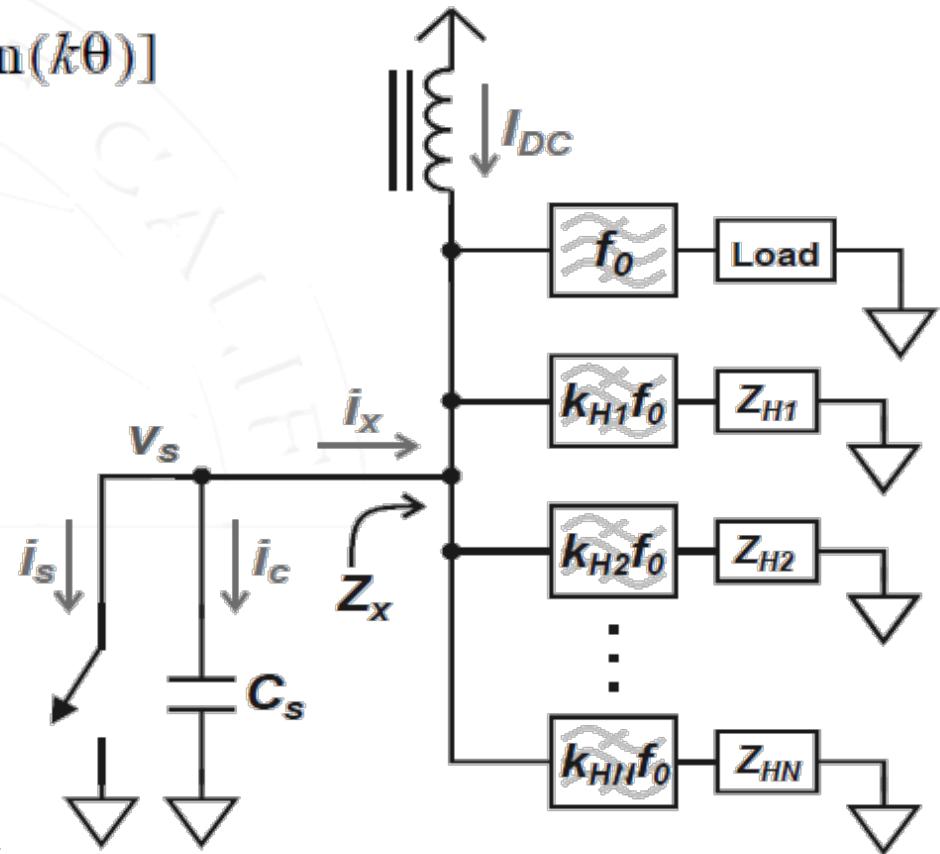
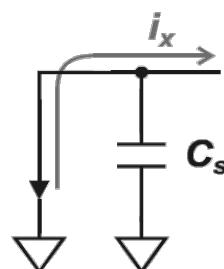
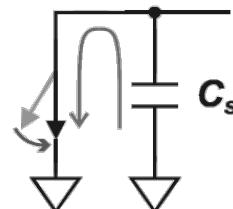
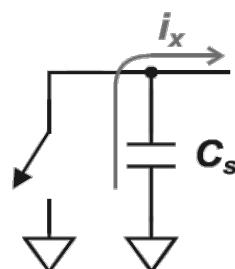
Design Constraint	Drain Loss WF Factor (W_D)	Gain WF Factor (W_G)	Drain Loss Device Factor (D_D)	Gain Device Factor (D_G)
device size limited η_D	$F_I^2 F_V^2$	N/A	$\frac{P_{out}}{(\bar{V}_{bk}^2 / \bar{R}_{on})}$	N/A
capacitance limited	$F_I^2 F_C$	F_V^2 / F_C	$\frac{2\pi \bar{R}_{on} \bar{C}_{out}}{1/f_0}$	$\frac{\bar{P}_{in}}{2\pi f_0 \bar{C}_{out} V_{bk}^2}$
gain limited PAE	$2F_I F_V$	$2F_I F_V$	$\sqrt{\frac{\bar{P}_{in}}{V_{bk}^2 / \bar{R}_{on}}}$	$\sqrt{\frac{\bar{P}_{in}}{V_{bk}^2 / \bar{R}_{on}}}$

F_V	F_I	F_C	F_{PI}
V_{pk}/V_{DC}	I_{RMS}/I_{DC}	$\frac{P_{out}}{V_{DC}^2/Z_C}$	I_{pk}/I_{DC}

General Class E/F Design

$$i_x = a_0 + \sum_{k \in T} [a_k \cos(k\theta) + b_k \sin(k\theta)]$$

- A switch with parallel capacitance, an ideal choke, and a possibly countable infinite number of harmonic impedances.



Switch/Cap Current

- The switch carries an impulsive current component due to cap discharge.

$$i_s = -s(\theta) \cdot i_x(\theta) + Q \cdot \delta(\theta)$$

$$\begin{aligned} I_s(k) &= \frac{1}{2\pi} Q - S_k \otimes I_x(k) \\ &= \frac{1}{2\pi} Q - \sum_{|l| \in \{0, T\}} S_{k-l} I_x(l) \end{aligned}$$

$$s(\theta) = \begin{cases} 1 & 0 < \theta < 2\pi D \\ 0 & 2\pi D < \theta < 2\pi \end{cases}$$

$$i_{cs} = -\bar{s}(\theta) \cdot i_x(\theta) - Q \cdot \delta(\theta)$$

$$S(\theta) = \begin{cases} D & k = 0 \\ \frac{\sin(2\pi Dk)}{2\pi k} - j \frac{\sin^2(\pi Dk)}{\pi k} & k \neq 0 \end{cases}$$

$$\begin{aligned} I_{cs}(k) &= -\frac{1}{2\pi} Q - \bar{S}_k \otimes I_x(k) \\ &= -\frac{1}{2\pi} Q - \sum_{|l| \in \{0, T\}} \bar{S}_{k-l} I_x(l) \end{aligned}$$

Cap Voltage

- The voltage across the capacitor is calculated from the current.
- The harmonic impedance constraint implies the following relations.

$$v_s(\theta) = \begin{cases} 0 & 0 < \theta < 2\pi D \\ Z_{cs} \int_0^{\theta} i_{cs}(\theta)d\theta & 2\pi D < \theta < 2\pi \\ 2\pi D & \end{cases}$$

$$V_s(k) = -j \frac{Z_{cs}}{k} I_{cs}(k)$$

$$Z_{cs} = 1/(2\pi f_0 C_s)$$

$$I_{cs}(0) = 0 \quad \underbrace{Z_k I_x(k)}_{\forall k \in T} = V_s(k)$$

$$\underbrace{jk \frac{Z_k}{Z_{cs}} I_x(k) - I_{cs}(k)}_{\forall k \in T} = 0$$

Constraint Equations

- There are $|T|$ complex valued equations and one real valued equation and $|T|+2$ unknowns.
- These equations can be solved for Q and i_x
- The ZVS solution has an additional constraint $Q=0$.

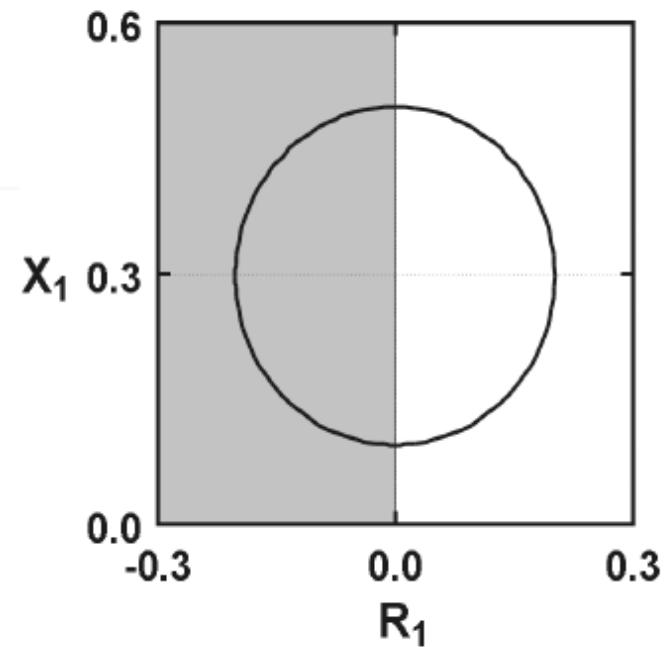
$$\underbrace{jk \frac{Z_k}{Z_{cs}} I_x(k) - I_{cs}(k)}_{\forall k \in \{0, T\}} = 0$$

$$\underbrace{jk \frac{Z_k}{Z_{cs}} I_x(k) + [\bar{S}_l \otimes I_x(l)] \Big|_{l=k} + \frac{1}{2\pi} Q}_{\forall k \in \{0, T\}} = 0$$

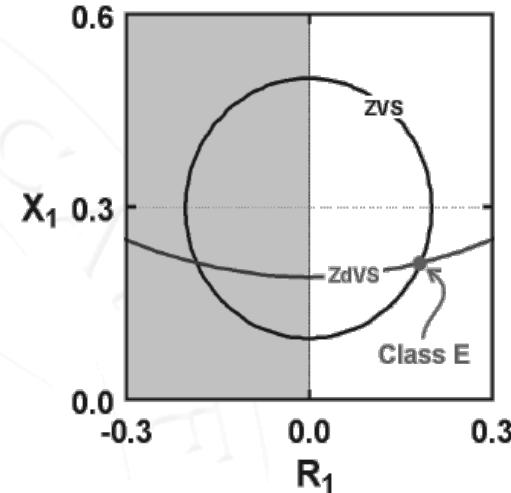
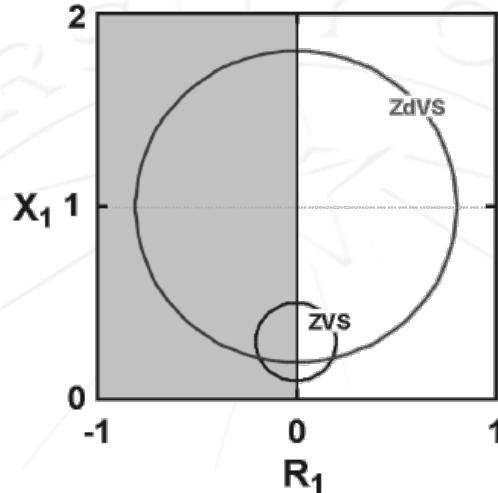
ZVS

- It can be shown that ZVS implies that:
- (R_1, X_1) is the fundamental load impedance.
- The center and radius of the circle is determined by the overtone network.

$$(X_1 - C_X)^2 + (R_1 - C_R)^2 - C_Z^2 = 0$$



ZdVS



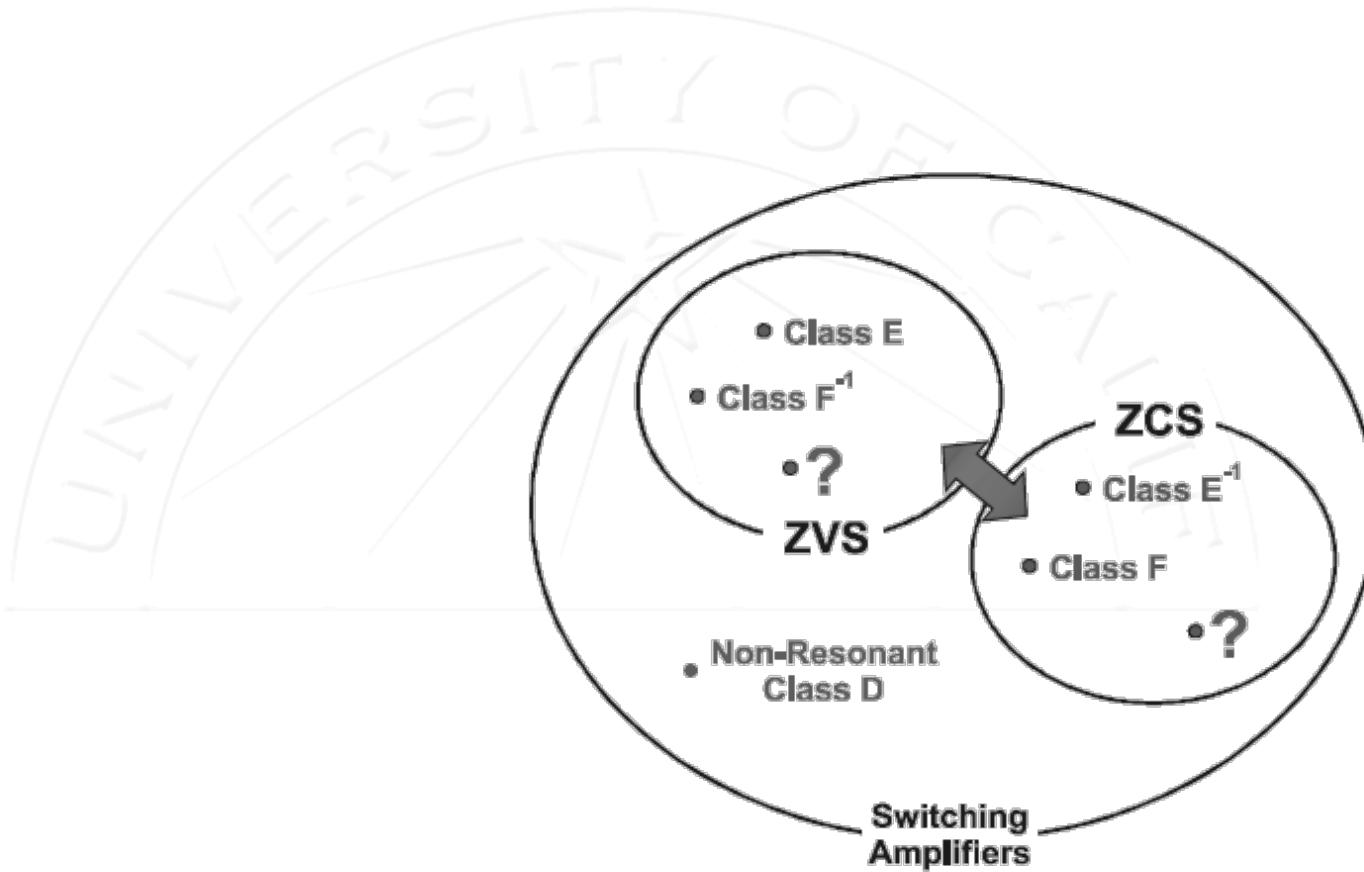
- The ZdVS conditions generate additional constraints.

$$\left. \frac{dV_s}{d\theta} \right|_{\theta=0} = Z_{cs} \sum_{k \in \{0, T\}} a_k$$

$$\frac{dV_s}{d\theta} = Z_{cs} i_{cs}$$

$$\sum_{k \in \{0, T\}} a_k = 0$$

Class E/F Amplifier Family

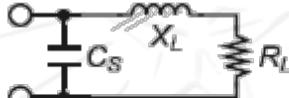


- Are there other interesting tuning networks besides the well known Class E and F?

Switching Amplifier Wish List

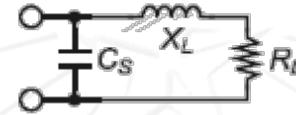
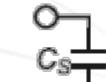
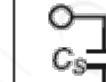
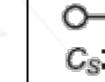
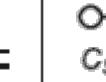
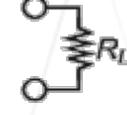
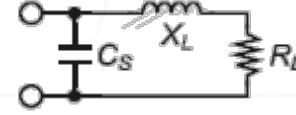
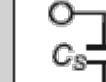
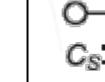
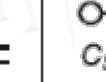
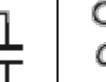
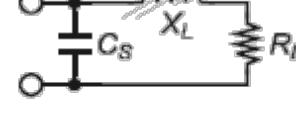
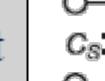
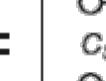
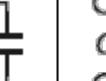
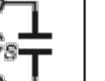
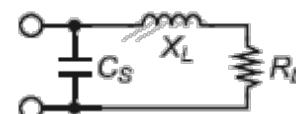
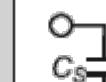
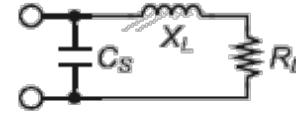
- ZVS Switching
- Inclusion of device output capacitance
- Simple circuit implementation
- Lower peak voltage (F_v)
- Lower RMS current (F_i)
- Capacitance Tolerance (F_c)

Class E versus F

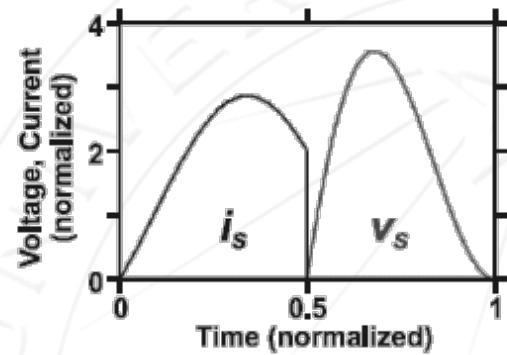
Tuning	f_0	$2f_0$	$3f_0$	$4f_0$	$5f_0$	$6f_0$	$7f_0$	$8f_0$
Class E								
Class F ⁻¹		open	short	open	short	open	short	open

- Consider a hybrid of Class E and F with desirable properties of both. Construct such a hybrid by choosing harmonics to either satisfy Class E or F conditions. Note that the fundamental load is set by Class E ZVS conditions.

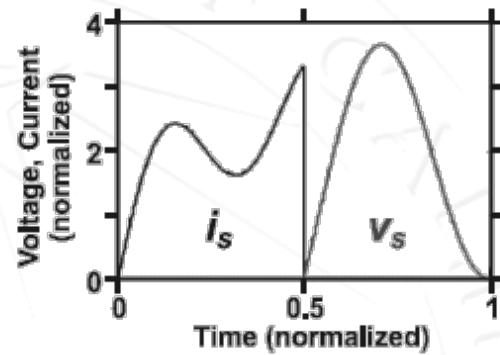
Class E/F Family

Tuning	f_0	$2f_0$	$3f_0$	$4f_0$	$5f_0$	$6f_0$	$7f_0$	$8f_0$
Class E								
Class F ⁻¹		open	short	open	short	open	short	open
Class E/F ₂		open						
Class E/F ₃			short					
Class E/F _{2,4,7}		open		open			short	
Class E/F _{2,3,4,5}		open	short	open	short			

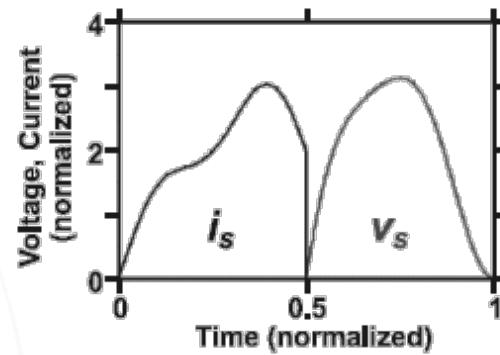
Example Class E/F Waveforms



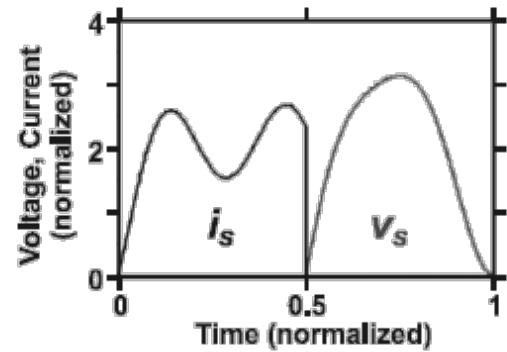
Class E



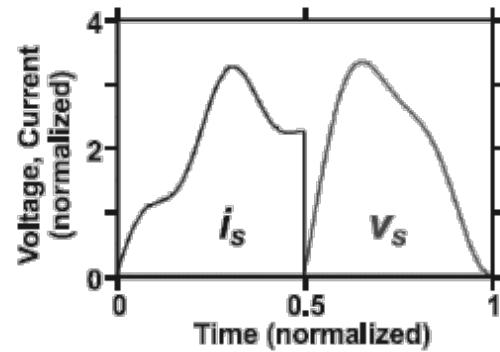
Class E/F₂



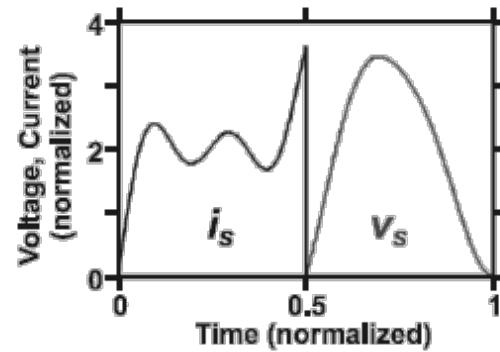
Class E/F₃



Class E/F_{2,3}

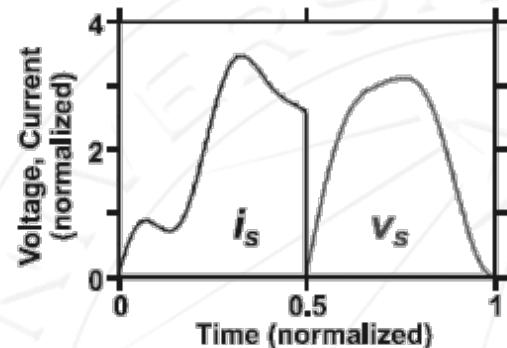


Class E/F₄

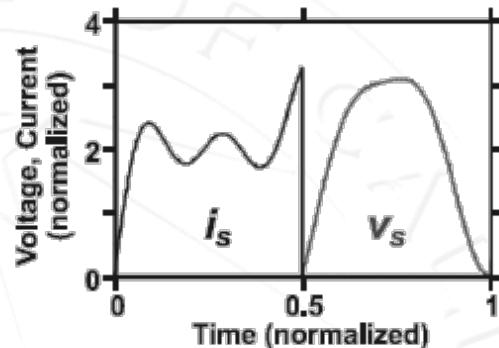


Class E/F_{2,4}

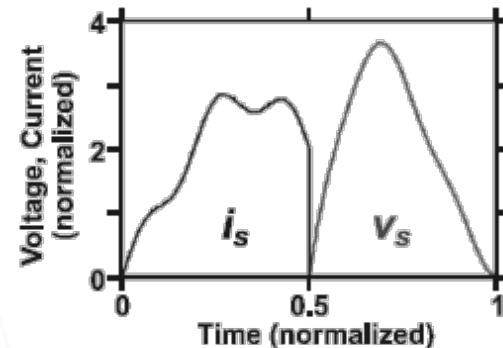
More Waveforms



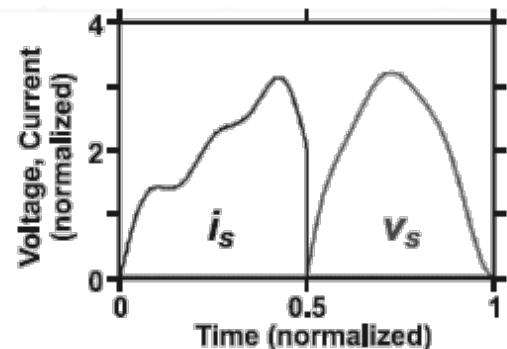
Class E/F_{3,4}



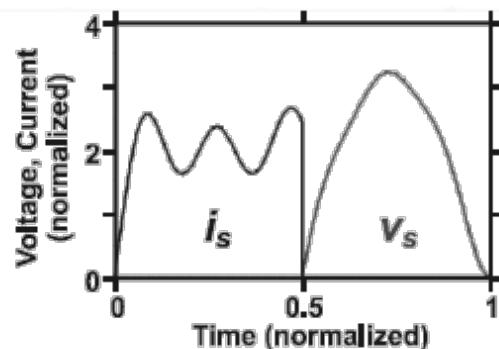
Class E/F_{2,3,4}



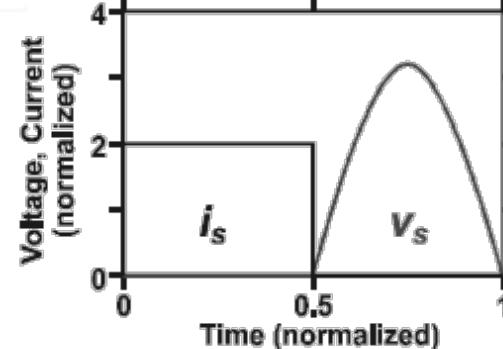
Class E/F₅



Class E/F_{3,5}

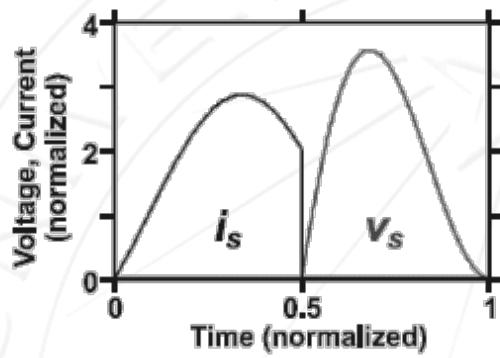


Class E/F_{2,3,4,5}

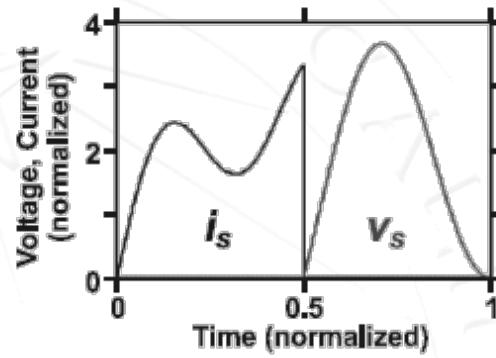


Class F⁻¹

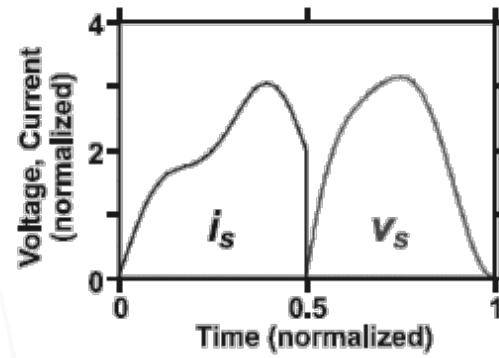
Single Harmonic Class E/F



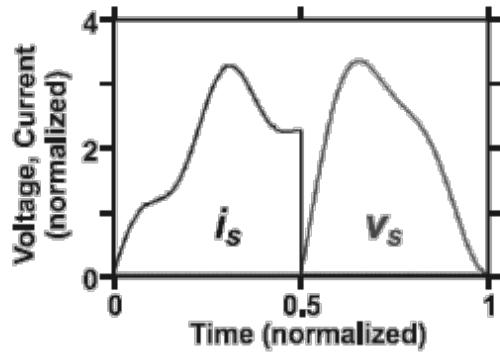
Class E



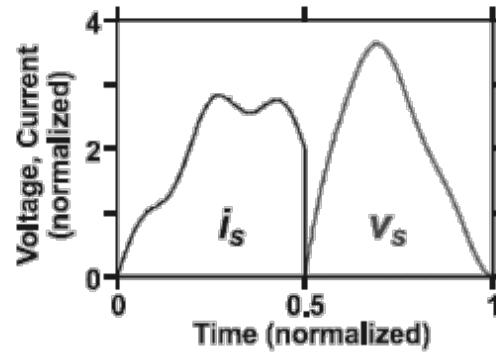
Class E/ F_2



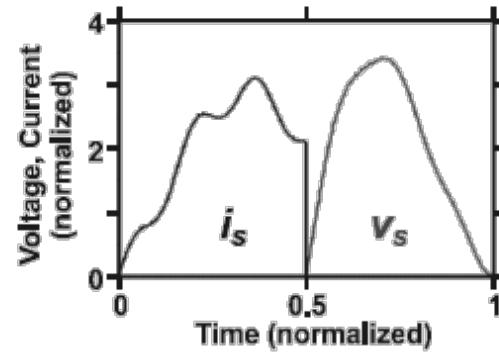
Class E/ F_3



Class E/ F_4

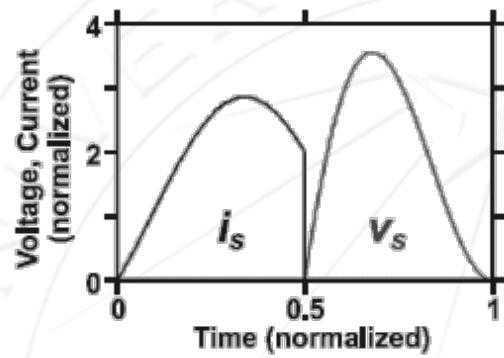


Class E/ F_5

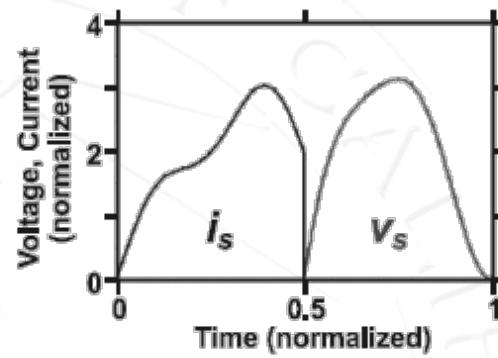


Class E/ F_6

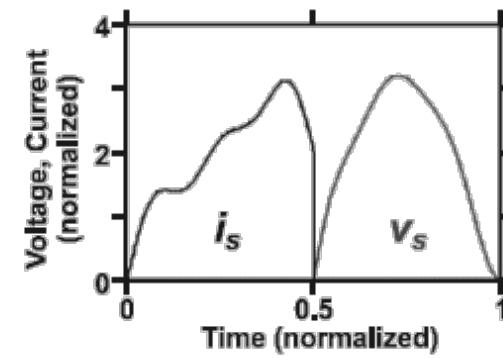
Odd Harmonic Class E/F



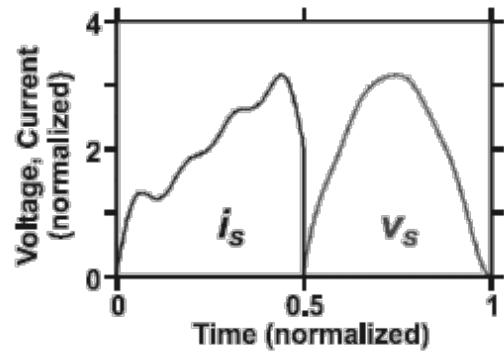
Class E



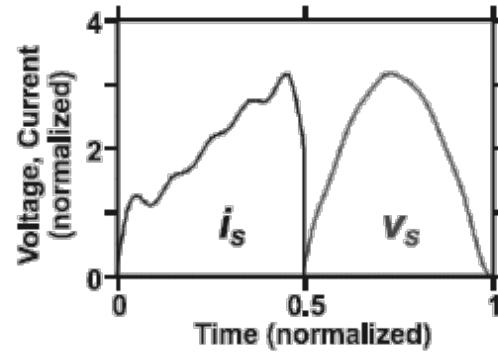
Class E/ F_3



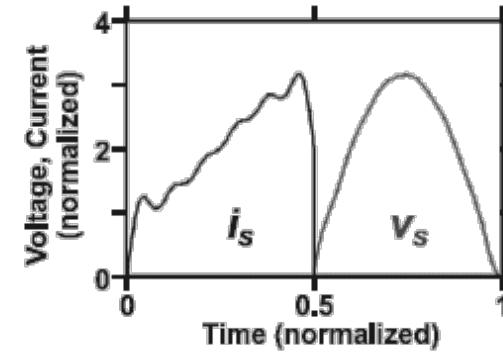
Class E/ $F_{3,5}$



Class E/ $F_{3,5,7}$

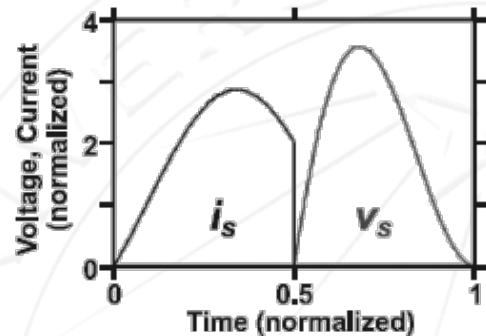


Class E/ $F_{3,5,7,9}$

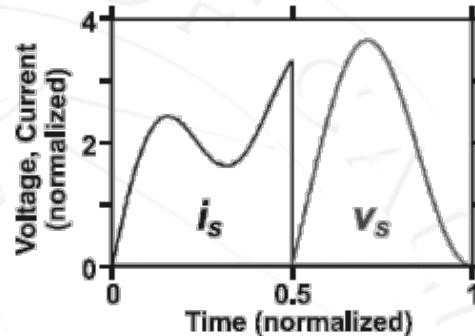


E/ $F_{3,5,7,9,11}$

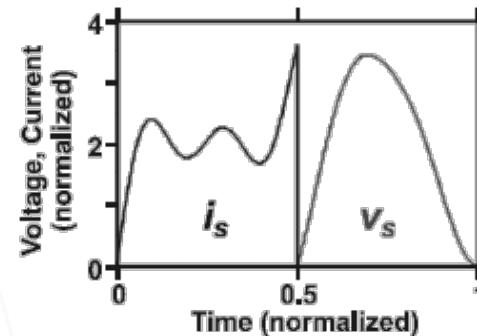
Even Harmonic Class E/F



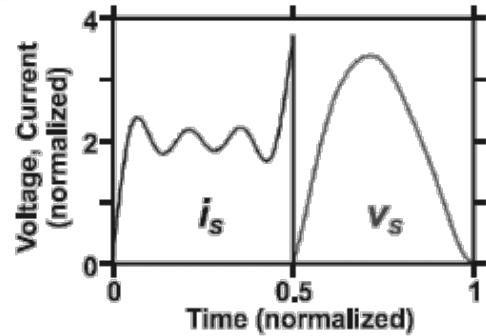
Class E



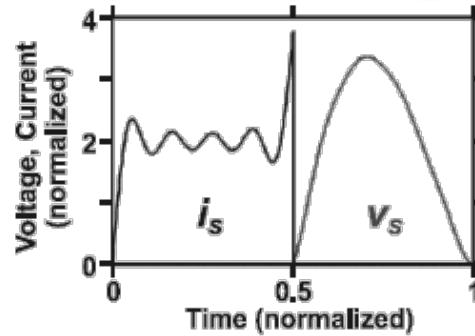
Class E/F₂



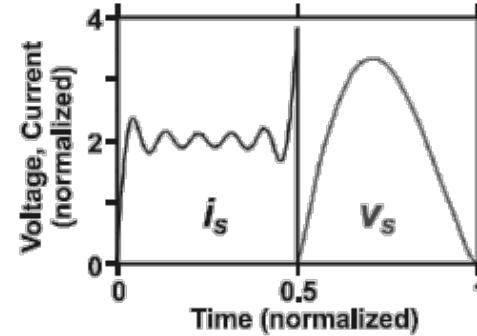
Class E/F_{2,4}



Class E/F_{2,4,6}

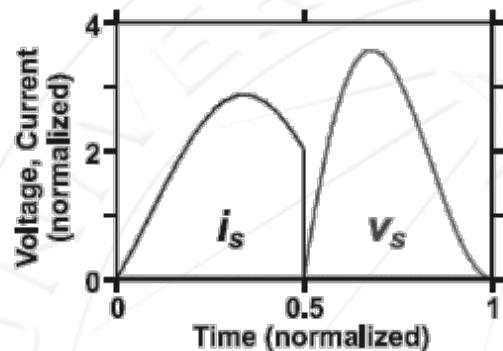


Class E/F_{2,4,6,8}

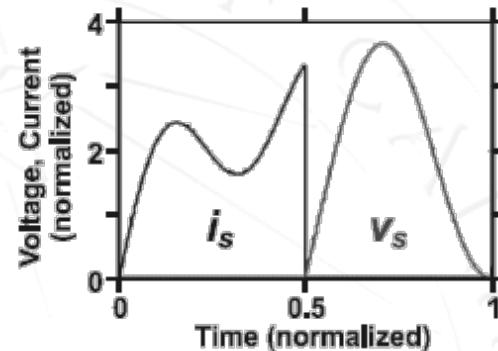


Class E/F_{2,4,6,8,10}

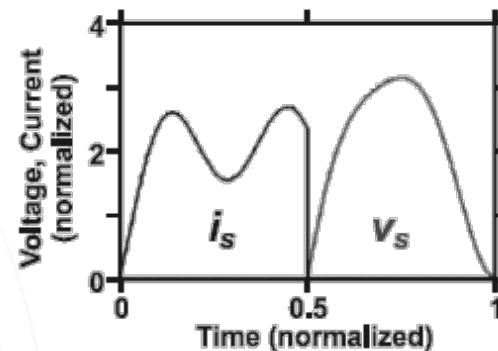
N-Harmonic Class E/F



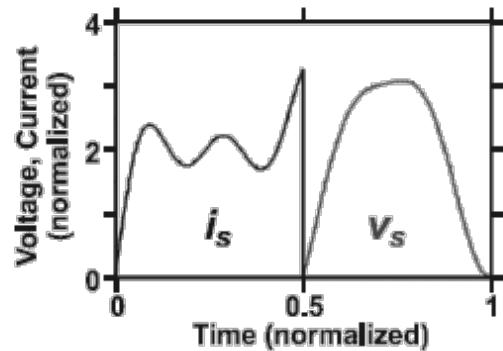
Class E



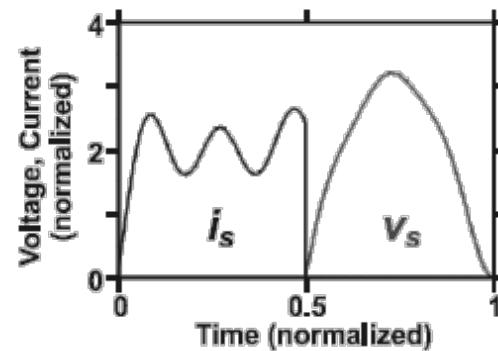
Class E/F₂



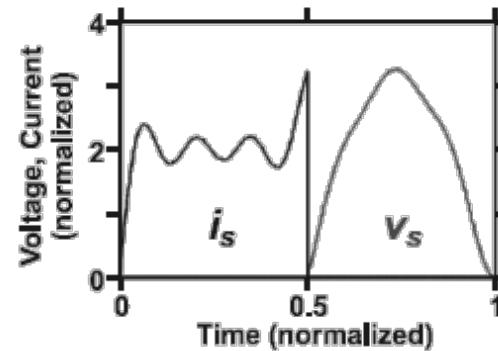
Class E/F_{2,3}



Class E/F_{2,3,4}



Class E/F_{2,3,4,5}

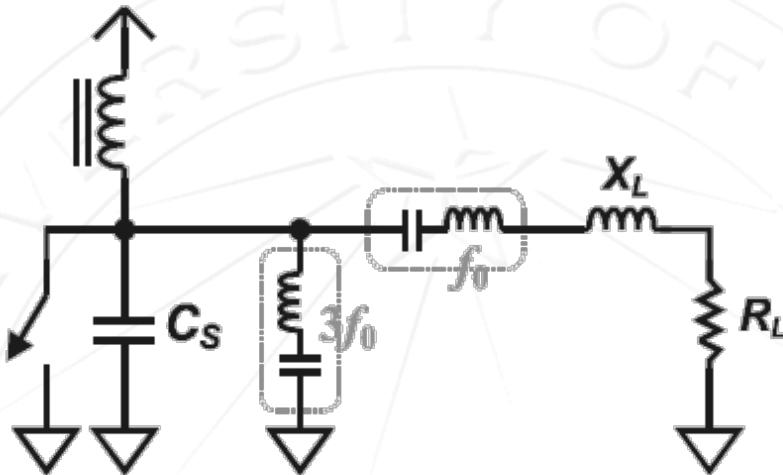


Class E/F_{2,3,4,5,6}

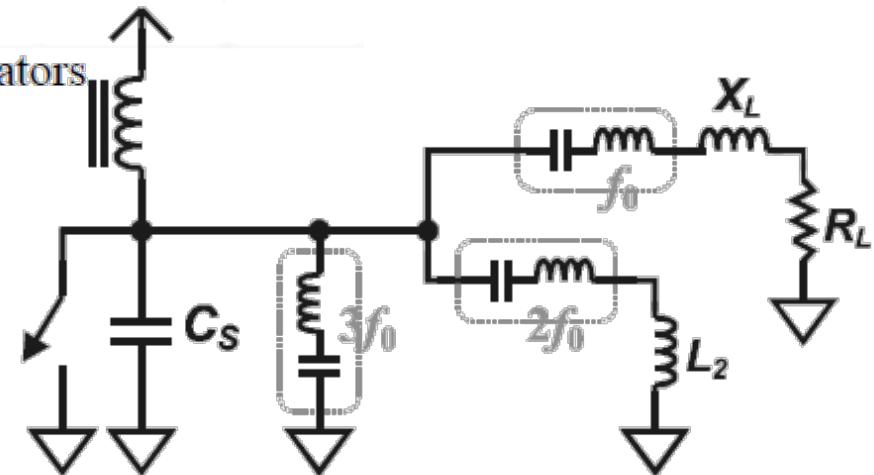
Overall Comparison

Tuning	Waveform Merit				Performance Merit (normalized to unity for Class E)				
	F_V	F_I	F_C	F_{PI}	$F_V^2 F_I$	$F_I^2 F_C$	F_V^2 / F_C	$2F_V F_I$	$F_V F_{PI}$
E	3.56	1.54	3.14	2.86	1.00	1.00	1.00	1.00	1.00
E/F ₂	3.67	1.48	1.13	3.33	0.98	0.33	2.94	0.99	1.20
E/F ₃	3.14	1.52	3.14	3.06	0.76	0.97	0.78	0.87	0.94
E/F _{2,3}	3.13	1.47	2.31	2.67	0.71	0.67	1.05	0.84	0.82
E/F ₄	3.34	1.55	2.45	3.27	0.89	0.79	1.13	0.94	1.07
E/F _{2,4}	3.43	1.46	0.97	3.60	0.84	0.28	3.00	0.91	1.21
E/F _{3,4}	3.10	1.62	1.93	3.45	0.84	0.68	1.23	0.91	1.05
E/F _{2,3,4}	3.08	1.45	1.18	3.26	0.67	0.34	1.99	0.82	0.99
E/F ₅	3.65	1.53	3.14	2.84	1.04	0.99	1.05	1.02	1.02
E/F _{3,5}	3.20	1.51	3.14	3.12	0.78	0.97	0.78	0.87	0.94
E/F _{2,3,4,5}	3.20	1.45	2.11	2.65	0.72	0.60	1.20	0.85	0.83
EECS 242 F ⁻¹	3.14	1.41	N/A	2.00	0.66	N/A	N/A	0.81	0.62

Direct Implementation

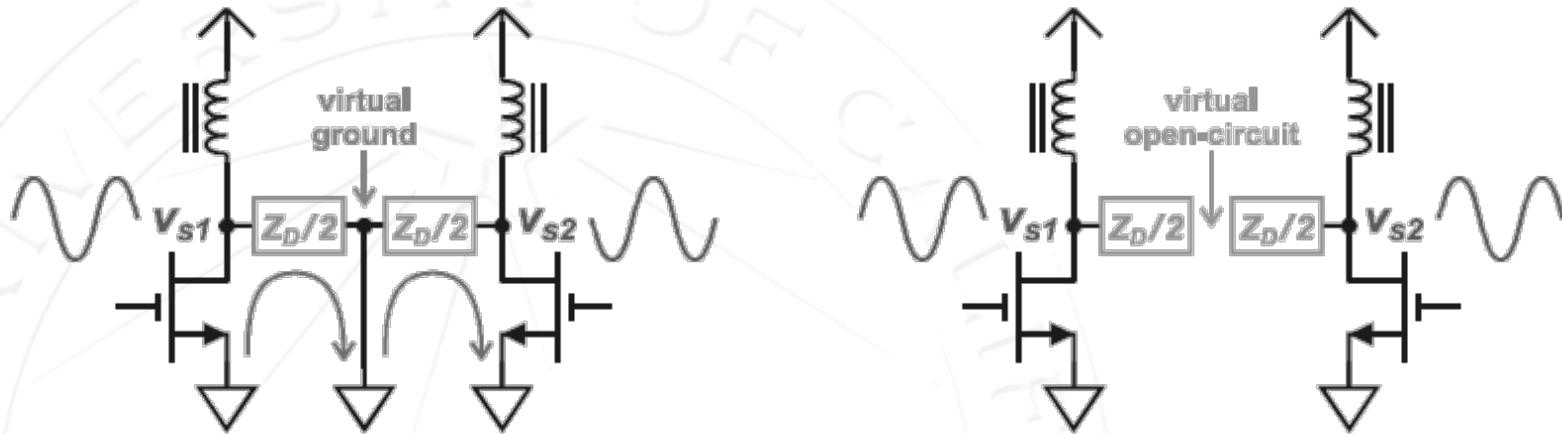


Direct E/F₃ implementation using high-Q resonators



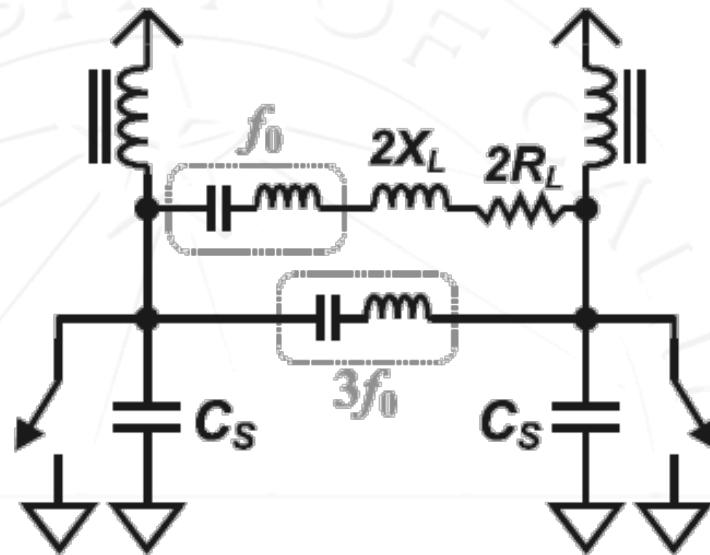
Direct E/F_{2,3} implementation using high-Q resonators.

Push-Pull Amplifiers



- Creation of virtual grounds at odd harmonics and open circuits at even harmonics is very handy for designing class E/F amplifiers.

Class E/F₃ Push-Pull

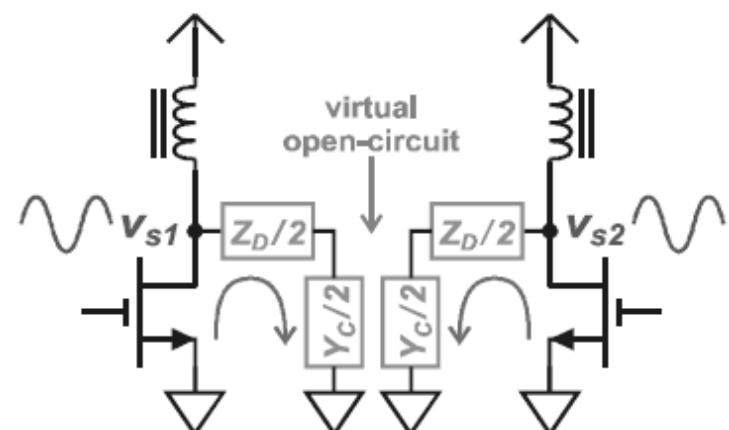
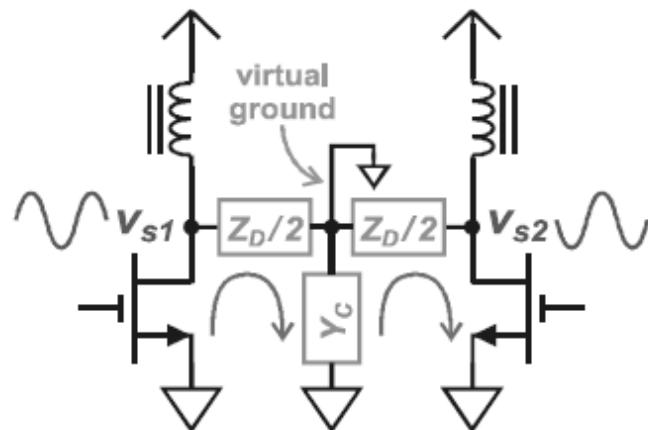
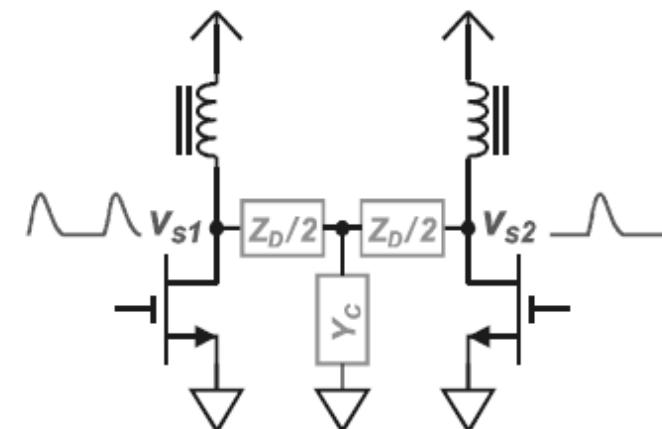


Class E/F₃ push/pull implementation allowing the use of low-Q resonators.

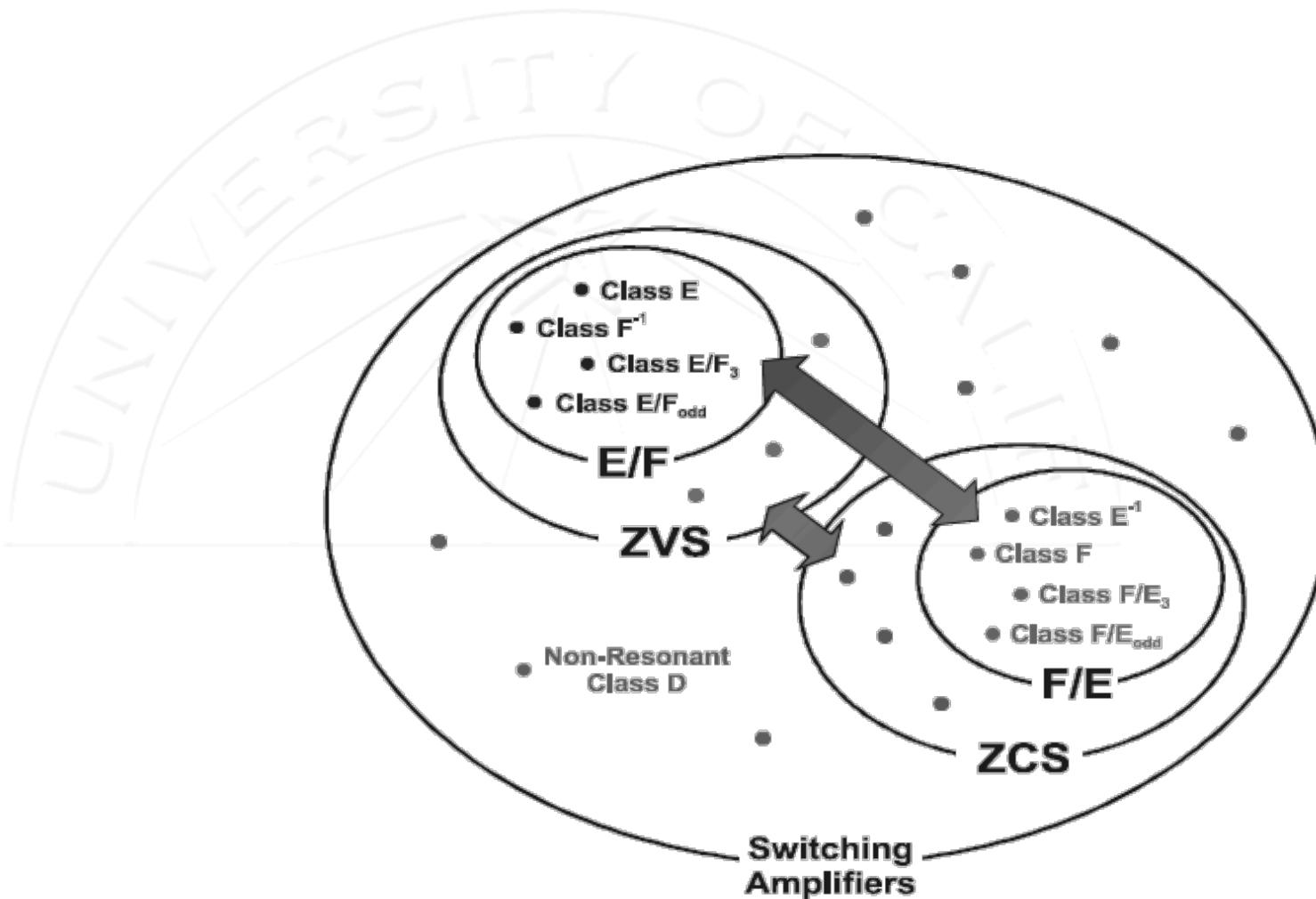
- Short at third harmonic. Does not need to be high Q since even harmonics don't "see it"

Even & Odd Harmonic Control

- Odd harmonics only see $Z_D/2$ since Y_C is shorted to ground.
- Even harmonics see Y_c to ground

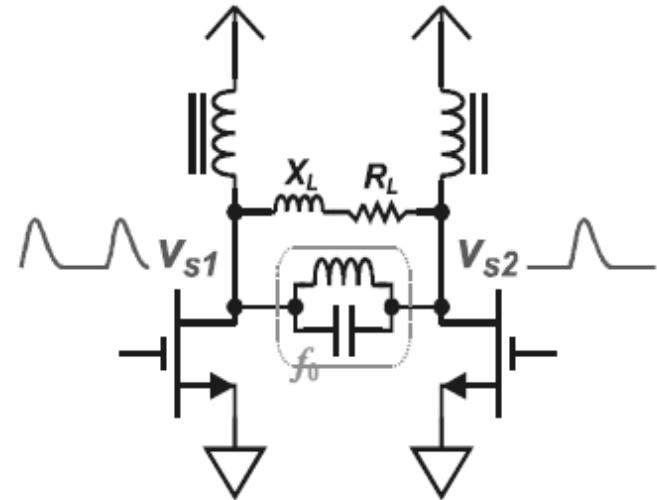


Switching Amplifier Landscape



Class E/F_{odd} Amplifier

- All odd harmonics see small impedance (ideally short) whereas even harmonics do not see shunt LC tank.



$$\begin{aligned}Y_1 &= \frac{I_1}{V_1} = \frac{4I_{DC}}{\pi^2 V_{DC}} - j \cdot \frac{1}{Z_C} \\&= \frac{1}{R_L} - j \cdot \frac{1}{Z_C}\end{aligned}$$

$$R_L \equiv \frac{\pi^2}{4} \cdot \frac{V_{DC}}{I_{DC}}$$

Prof. Ali M. Niknejad (C) 2009

AMPLIFIER CLASSES (adapted from Wikipedia)

Amplifier circuits are classified as A, B, AB and C for analog design, and class D, E, and F for high efficiency switching design. For the analog classes, each class defines what proportion of the input signal cycle (called the angle of flow) is used to actually switch on the amplifying device:

Class A

100% of the input signal is used (conduction angle $a = 360^\circ$ or 2π)

Class A amplifiers amplify over the whole of the input cycle such that the output signal is an exact scaled-up replica of the input with no clipping. They are not very efficient — a theoretical maximum of 50% is obtainable, but for small signals, this waste of power is still extremely small, and can be easily tolerated. Only when we need to create output powers with appreciable levels of voltage and current does Class A become problematic. In a Class A circuit, the amplifying element is biased such that the device is always conducting to some extent, and is operated over the most linear portion of its characteristic curve (known as its transfer characteristic or transconductance curve). Because the device is always conducting, even if there is no input at all, power is wasted. This is the reason for its inefficiency

Class AB

more than 50% but less than 100% is used. (181° to 359° , $\pi < a < 2\pi$)

- Class AB1 applies to tube or transistor amplifiers in class AB where the grid or base is more negatively biased than it is in class A.
- Class AB2 applies to tube or transistor amplifiers in class AB where the grid or base is often more negatively biased than in AB1, and the input signal is often larger. When the drive is high enough to make the grid or the base more positive, the grid or base current will increase. It is possible depending on the level of the signal input for the amplifier to move from class AB1 to AB2.

Class B

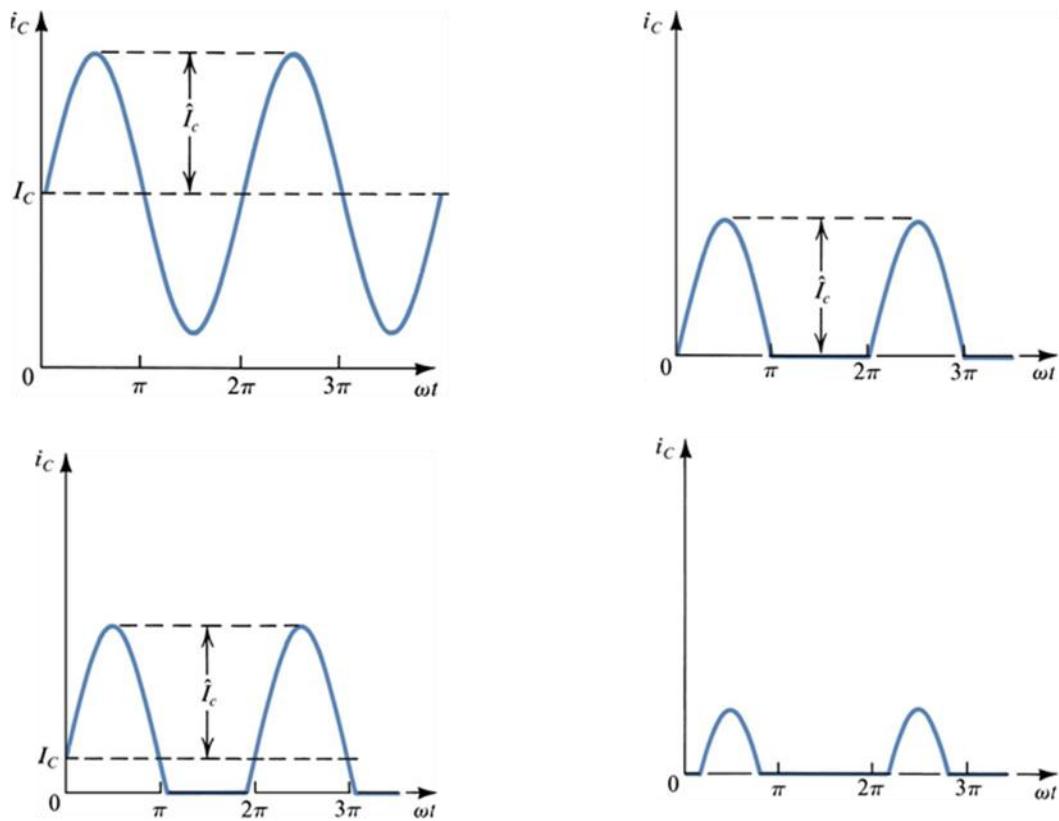
50% of the input signal is used ($a = 180^\circ$ or π)

Class B amplifiers only amplify half of the input wave cycle (maximum theoretical efficiency of 78.5%). This is because the amplifying element is switched off altogether half of the time, and so cannot dissipate power. A single Class B element is rarely found in practice, though it can be used in RF power amplifiers where the distortion is unimportant. However Class C is more commonly used for this.

Class C

less than 50% is used (0° to 179° , $a < \pi$)

Class C amplifiers conduct less than 50% of the input signal and the distortion at the output is high, but efficiencies of up to 90% can be reached. Some applications can tolerate the distortion, such as megaphones. A much more common application for Class C amplifiers is in RF transmitters, where the distortion can be vastly reduced by using tuned loads on the amplifier stage. The input signal is used to roughly switch the amplifying device on and off, which causes pulses of current to flow through a tuned circuit. The tuned circuit will only resonate at particular frequencies, and so the unwanted frequencies are dramatically suppressed, and the wanted full signal (sine wave) will be abstracted by the tuned load. Provided the transmitter is not required to operate over a very wide band of frequencies, this arrangement works extremely well. Other residual harmonics can be removed using a filter.



Collector current waveforms for transistors operating in (a) class A, (b) class B, (c) class AB, and (d) class C amplifier stages.

Class D

A class D amplifier is a power amplifier where all power devices are operated in on/off mode. Output stages such as those used in pulse generators are examples of class D amplifiers. Mostly though, the term applies to devices intended to reproduce signals with a bandwidth well below the switching frequency. These amplifiers use

pulse width modulation, pulse density modulation (sometimes referred to as pulse frequency modulation) or more advanced form of modulation such as Sigma delta. The input signal is converted to a sequence of pulses whose averaged value is directly proportional to the amplitude of the signal at that time. The frequency of the pulses is typically ten or more times the highest frequency of interest in the input signal. The output of such an amplifier contains unwanted spectral components (i.e.. the pulse frequency and its harmonics) that must be removed by a passive filter. The resulting filtered signal is then an amplified replica of the input.

The main advantage of a class D amplifier is power efficiency. Because the output pulses have fixed amplitude, the switching elements (transistors) are switched either on or off, rather than operated in linear mode. This means that very little power is dissipated by the transistors except during the very short interval between the on and off states. The wasted power is low because the instantaneous power dissipated in the transistor is the product of voltage and current, and one or the other is almost always close to zero. The lower losses permit the use of a smaller heat sink while the power supply requirements are lessened too.

Class D amplifiers were widely used to control small DC motors, but they are now also used as audio amplifiers, with some extra circuitry to allow analogue to be converted to a much higher frequency pulse width modulated signal.

Class E

The class E/F amplifier is a highly efficient switching power amplifier, typically used at such high frequencies that the switching time becomes comparable to the duty time. As said in the class-D amplifier the transistor is connected via a serial-LC-circuit to the load, and connected via a large L (inductivity) to the supply voltage. The supply voltage is connected to ground via a large capacitor to prevent any RF-signals leaking into the supply. The class-E amplifier adds a C between the transistor and ground and uses a defined L to connect to the supply voltage.

All previous designs use sharp edges to minimize the overlap. Class E uses a significant amount of second harmonic voltage. The second harmonic can be used to reduce the overlap with edges with finite sharpness. In reality the impedance is mostly reactive and the only reason for it is that class E is a class F amplifier with a very simplified load network and thus has to deal with imperfections. Note how in many amateur simulations of class E amplifiers sharp current edges are assumed nullifying the very motivation for class E and measurements near the transit frequency of the transistors show very symmetric curves, which look much similar to class F simulations.

Class F

The class F-amplifier takes the finite on resistance into account and tries to make the current touch the bottom at zero. This means the voltage and the current at the

transistor are symmetric with respect to time. The Fourier Transform allows an elegant formulation to generate the complicated LC-networks. It says that the first harmonic is passed into the load, all even harmonics are shorted and all higher odd harmonics are open.

In push-pull amplifiers and in CMOS the even harmonics of both transistors just cancel. Experiment tells that a square wave can be generated by those amplifiers and math tells that square wave do consist of odd harmonics only.

In a class D amplifier the output filter blocks all harmonics, which means the harmonics see an open load. So even small harmonic currents suffice to generate a voltage square wave. The current is in phase with the voltage applied to filter, but the voltage across the transistors is out of phase. Therefore there is a minimal overlap between current through the transistors and voltage across the transistors. The sharper the edges the lower the overlap.

While class D sees the transistors and the load as separate modules, the class F admits imperfections like the parasitics of the transistor and tries to optimize the global system to have a high impedance at the harmonics. Of course there has to be a finite voltage across the transistor to push the current across the on state resistance. Because the combined current through both transistors is mostly in the first harmonic it looks like a sine. That means that in the middle of the square the maximum of current has to flow, so it may make sense to have a dip in the square or in other words to allow some over swing of the voltage square wave. A class F load network by definition has to transmit below a cut off frequency and to reflect above. Any frequency lying below the cut off and having its second harmonic above the cut off can be amplified, that is an octave bandwidth. On the other hand a LC series circuit with a large L and a tunable C may be simpler to implement. By reducing the duty cycle below 0.5, the output amplitude can be modulated. The voltage square waveform will degrade, but any overheating is compensated by the lower overall power flowing. Any load mismatch behind the filter can only act on the first harmonic current waveform, clearly only a purely resistive load makes sense, then the lower the resistance the higher the current. Class F can be driven by sine or by a square wave, for a sine the input can be tuned by an L to increase gain. If class F is implemented with a single transistor the filter is complicated to short the even harmonics.

Notes

The main concept used in amplification is to model the active switching device, such as a transistor or MOSFET, as a linear combination of two parts: (1) a (theoretical) "perfect" switching element, and (2) a complex network of parasitic elements attached to it (capacitors, inductors and resistors). After the decomposition, it becomes trivial to eliminate the losses of each part:

(1) The "perfect" switching element should be switched ON during zero-voltage crossing, and should be switched OFF during zero-current crossing. Thus the switching element either conducts current, or has some non-zero voltage on it, but

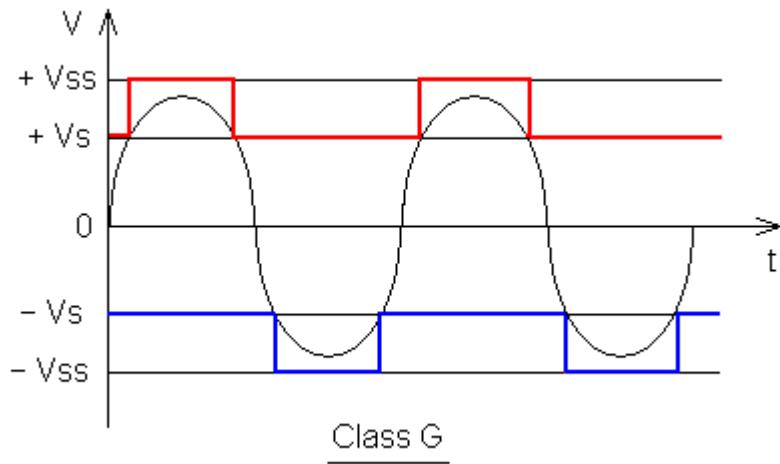
never both at the same time. Because the dissipated power is equal to current \times voltage, it becomes zero. This can be arranged by adjusting the phase (capacitor) and DC bias (resistor) of the signal going into the transistor input.

(2) The imaginary part of the impedance of the parasitic elements can be tuned, one by one, by matching them to another passive element with the complex conjugate impedance, thus leaving only the real part of the complex impedance.

In theory, the only remaining loss is the real part of the impedance of the parasitic elements in the system, which cannot be avoided. This class of amplifier is unique to radio frequency ranges, where the amplifier analysis is usually done in the frequency domain and not in the voltage/current domain. This class is further divided to subclasses depending on which harmonics of the signal are taken into account during zero-voltage switching (ZVS) and zero-current switching (ZCS), with names such as Class E/F2,odd; inverse Class (F^{-1}); and so on. It is still an active area of research and new variants appear from time to time, usually with the letters "E" and "F" somewhere in class name.

Class G

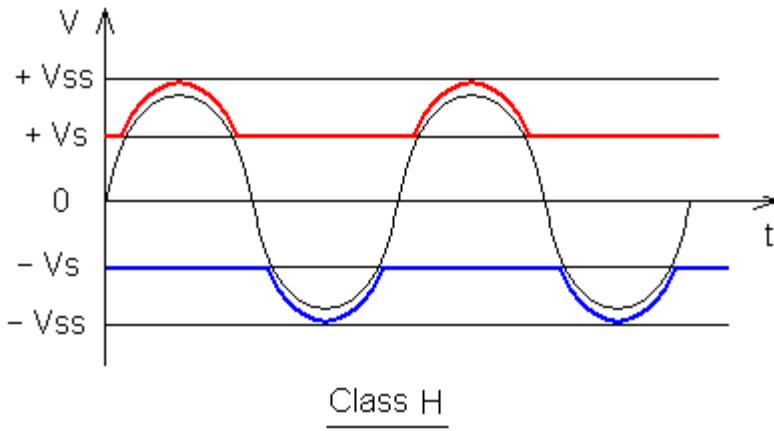
Class G amplifiers are a more efficient version of class AB amplifiers, which use "rail switching" to decrease power consumption and increase efficiency. The amplifier has several power rails at different voltages, and switches between rails as the signal output approaches each. Thus the amp increases efficiency by reducing the "wasted" power at the output transistors.



Class H

A Class H amplifier takes the idea of Class G one step further creating an infinite number of supply rails. This is done by modulating the supply rails so that the rails are only a few volts larger than the output signal at any given time. The output stage operates at its maximum efficiency all the time. Switched mode power supplies can be used to create the tracking rails. Significant efficiency gains can be achieved but

with the drawback of more complicated supply design and reduced THD performance.



Other classes

A hybrid configuration that is receiving new attention is the Doherty amplifier consisting of a class B main stage in parallel with a class C auxiliary (or "peaking") stage. The input signal is split evenly to drive the two amplifiers, and a Doherty combiner is used at the output to recombine the components. During periods of low signal level, the class B amplifier efficiently operates on the signal and the class C amplifier is inactive and consumes no power. During high signal peaks the class B amplifier saturates and the class C amplifier kicks in. The effect is that the Doherty amplifier maintains efficiency comparable to class B saturated performance even when operated at up to 6 dB output backoff. Interest in the Doherty configuration has been revived by cellular telephone and wireless internet applications where the sum of several constant envelope users creates an aggregate AM result. The main challenge of the Doherty amplifier is in aligning the two stages and getting the class C amplifier to turn on and off quickly.

Several audio amplifier manufacturers have started "inventing" new classes as a way to differentiate themselves. These class names usually do not reflect any revolutionary amplification technique, and are used mostly for marketing purposes. For example, Crowns K and I-Tech Series as well as several other models utilize Crowns patented **Class I** technology. Lab Gruppen use a form of class D amplifier called **class TD** or Tracked Class D which tracks the waveform to more accurately amplify it without the drawbacks of traditional class D amplifiers.

Class T is a trademark of TriPath company, which manufactures audio amplifier IC's. This new class "T" is a revision of the common class D amplifier, but with changes to ensure fidelity over the full audio spectrum, unlike traditional class D designs. It operates at a frequency of 650kHz, with a proprietary modulator.

Class Z is a trademark of Zetex semiconductor is a direct digital feedback technology.

Chapter 9

Output Stages And Power Amplifiers

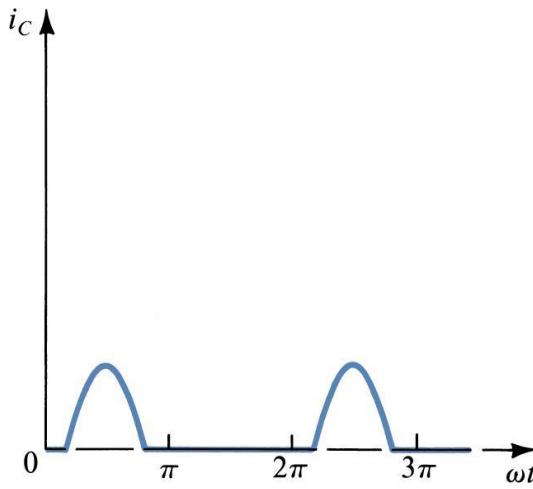
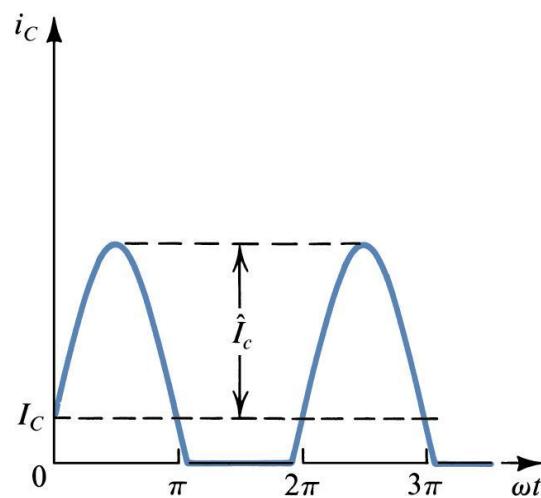
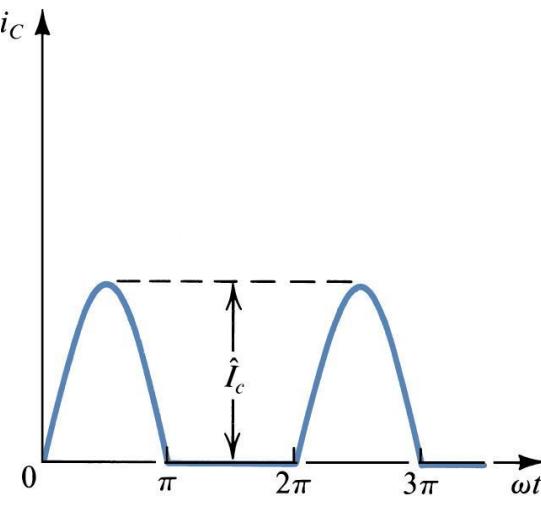
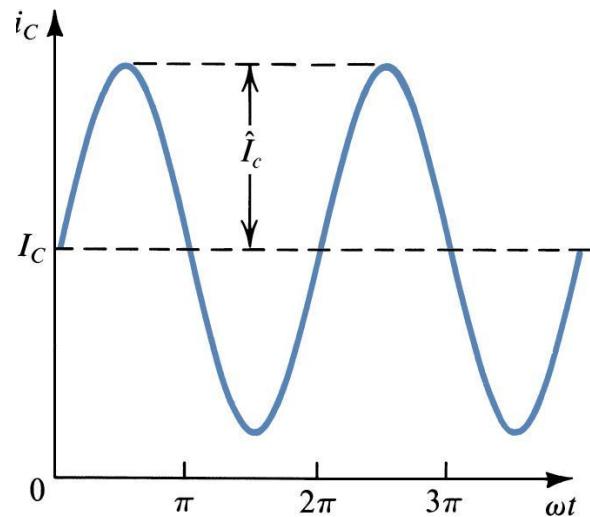
Low Output Resistance – no loss of gain

Small-Signal Not applicable

Total-Harmonic Distortion (fraction of %)

Efficiency

Temperature Requirements

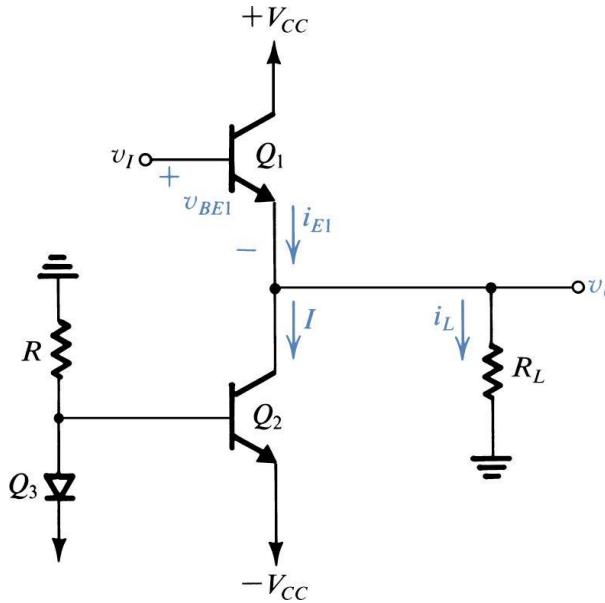
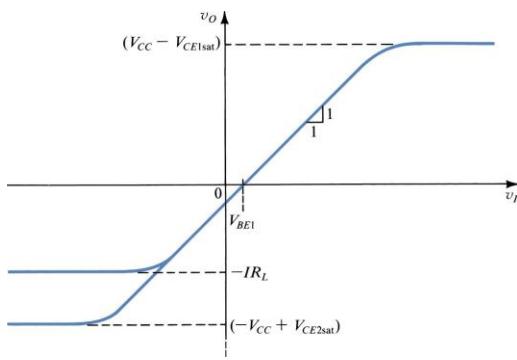
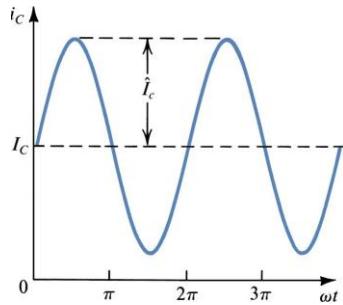


Collector current waveforms for transistors operating in (a) class A, (b) class B, (c) class AB, and (d) class C amplifier stages.

Class A

An emitter follower (Q_1) biased with a constant current I supplied by transistor Q_2 .

Transfer Characteristics



Transfer characteristic of the emitter follower. This linear characteristic is obtained by neglecting the change in v_{BE1} with i_L . The maximum positive output is determined by the saturation of Q_1 . In the negative direction, the limit of the linear region is determined either by Q_1 turning off or by Q_2 saturating, depending on the values of I and R_L .

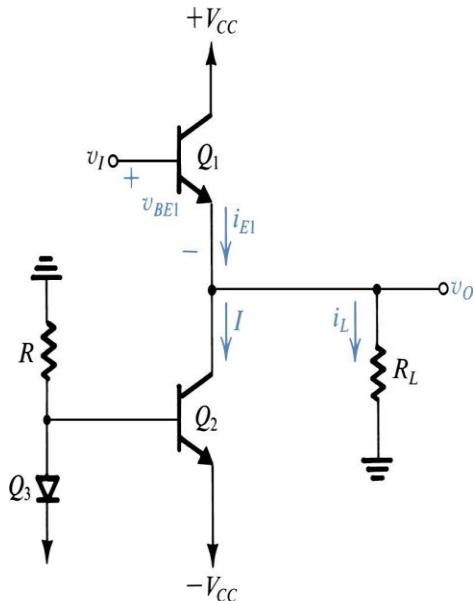
Class A

Transfer Characteristics

Crossover distortion can be eliminated by biasing the transistors at a small, non-zero current.

A bias Voltage V_{BB} is applied between Q_n and Q_p .

For $v_i = 0$, $v_o = 0$, and a voltage $V_{BB}/2$ appears across the base-emitter junction of each transistor.



$$i_N = i_P = I_Q = I_S \cdot e^{\frac{V_{BB}}{2 \cdot V_T}}$$

V_{BB} is selected to result the required quiescent current I_Q

$$v_o = v_i + \frac{V_{BB}}{2} - v_{BEN}$$

$$i_N = i_P + i_L$$

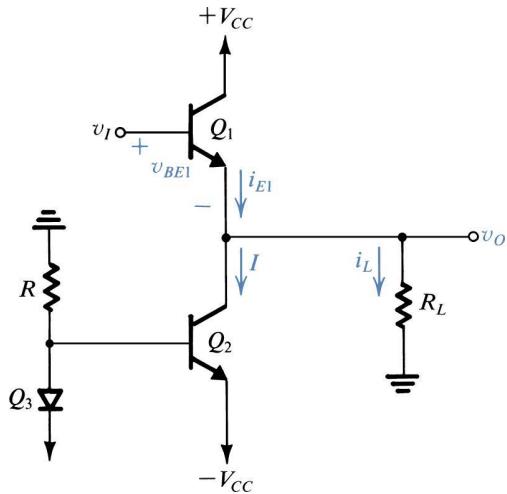
$$v_{BEN} + v_{EBP} = V_{BB} \quad V_T \cdot \ln\left(\frac{i_N}{I_S}\right) + V_T \cdot \ln\left(\frac{i_P}{I_S}\right) = 2 \cdot V_T \cdot \ln\left(\frac{i_Q}{I_S}\right)$$

$$i_N^2 = I_Q^2$$

$$i_N^2 - i_L \cdot i_N - I_Q^2 = 0$$

Class A

Transfer Characteristics



From figure 9.3 we can see that

$$v_{O\max} = V_{CC} - V_{CE1\text{sat}}$$

In the negative direction, the limit of the linear region is determined either by Q1 turning off

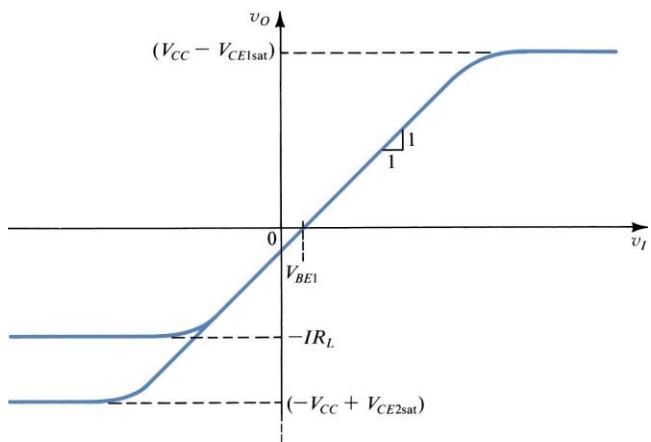
$$v_{O\min} = -I \cdot R_L$$

or by Q2 saturating

$$v_{O\min} = -V_{CC} + V_{CE2\text{sat}}$$

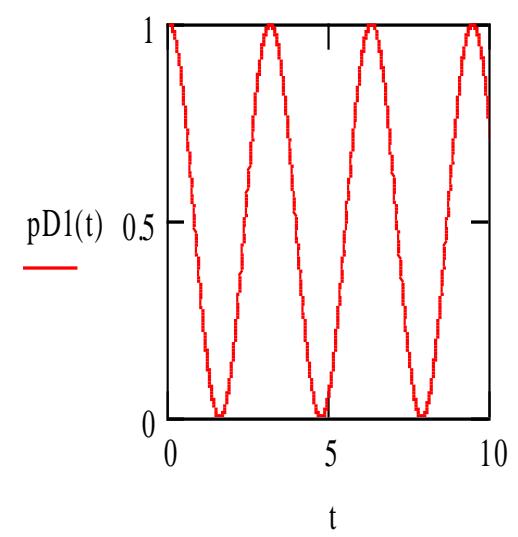
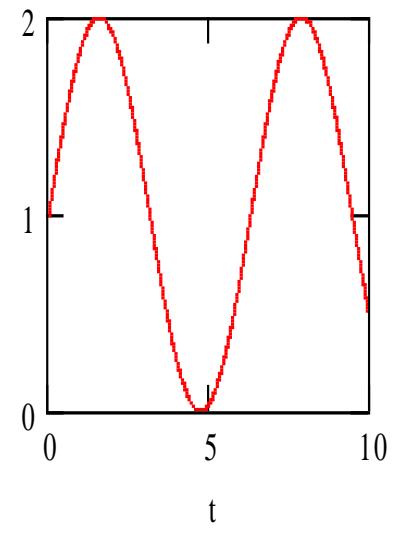
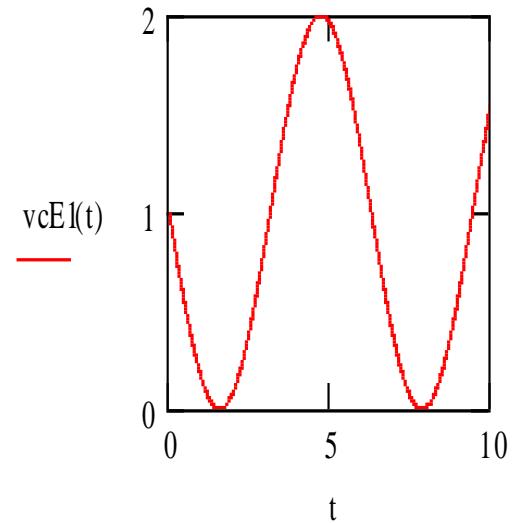
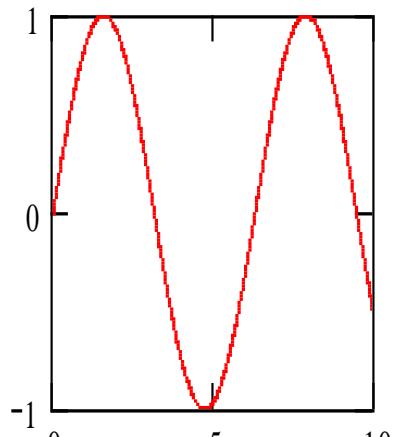
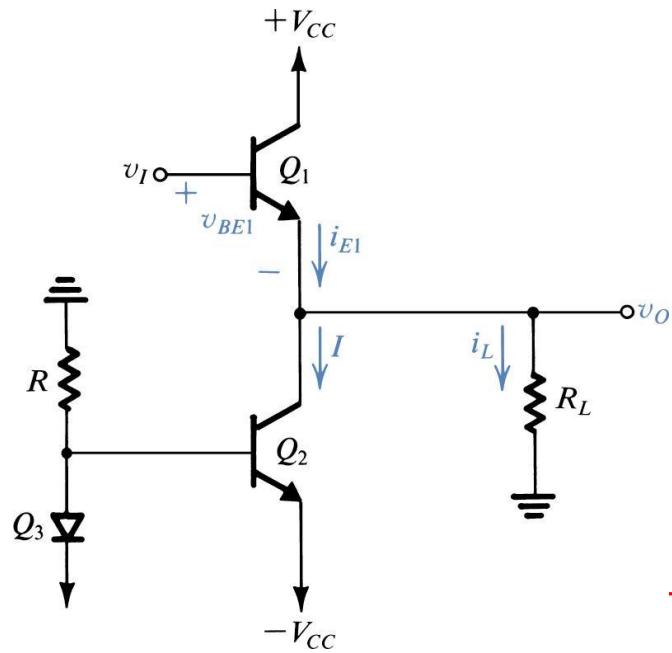
Depending on the values of I and R_L . The absolutely lowest output voltage is that given by the previous equation and is achieved provided that the bias current I is greater than the magnitude of the corresponding load current

$$I \geq \frac{|-V_{CC} + V_{CE2\text{sat}}|}{R_L}$$



Class A

Signal Waveforms



Class A

Power Dissipation

$$P = V_{CC} \cdot I$$

Largest Power Dissipation When $v_o = 0$

Q1 must be able to withstand a continuous dissipation of $V_{CC} \cdot I$

The power dissipation of Q1 depends on the value of R_L .

If R_L is infinite, $i_{C1} = I$ and the dissipation in Q1 depends on v_o .

Maximum power dissipation will occur when $v_o = -V_{CC}$ since v_{CE1} will be $2V_{CC}$
 $p_{D1} = 2V_{CC} \cdot I$. This condition would not normally persist for a prolonged interval.
the design need not be that conservative. The average $p_{D1} = V_{CC} \cdot I$

When R_L is zero a positive voltage would result in a theoretically infinite current (practical value) would flow through Q1. Short-circuit protection is necessary.

Class A

Power Conversion Efficiency

$$\eta = \frac{\text{load_power}(P_L)}{\text{supply_power}(P_S)}$$

$$P_L = \frac{1}{2} \cdot \frac{V_o^2}{R_L} \quad V_o \quad \text{average voltage}$$

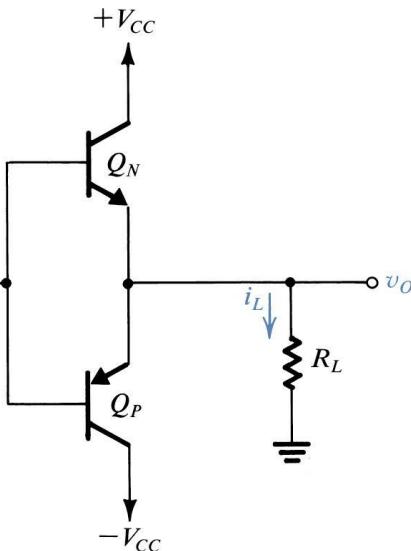
$$P_S = 2 \cdot V_{CC} \cdot I$$

$$\eta = \frac{1}{4} \cdot \frac{V_o^2}{I \cdot R_L \cdot V_{CC}} = \frac{1}{4} \cdot \left(\frac{V_o}{I \cdot R_L} \right) \cdot \left(\frac{V_o}{V_{CC}} \right)$$

$$V_o \leq V_{CC} \quad V_o \leq I \cdot R_L$$

maximum efficiency is obtained when

$$V_o = V_{CC} = I \cdot R_L$$



Biasing the Class B Output

- No DC current is used to bias this configuration.
- Activated when the input voltage is greater than the V_{be} for the transistors.
- npn Transistor operates when positive, pnp when negative.
- At a zero input voltage, we get no output voltage.

Class A

Power Conversion Efficiency

CLASS A

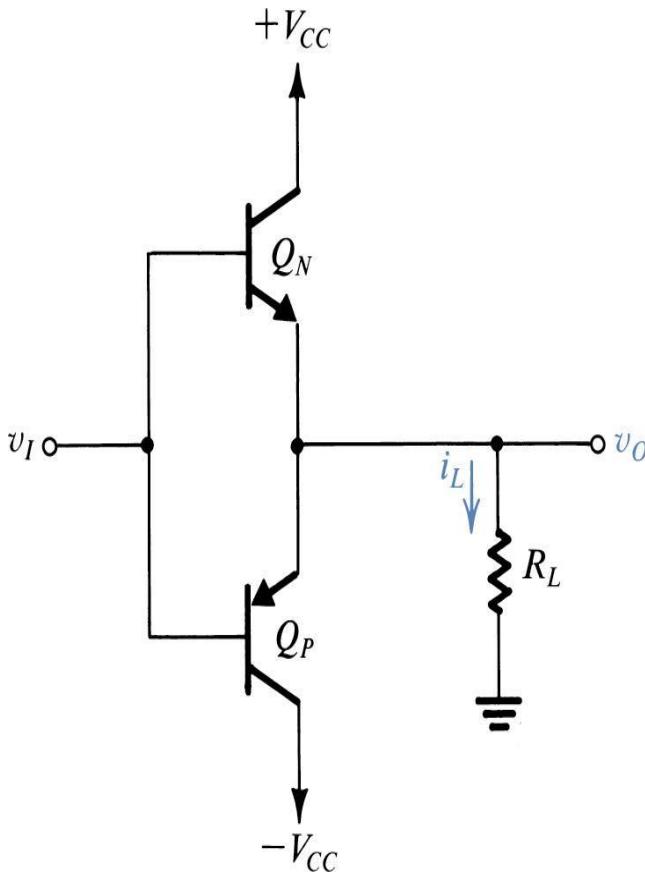
Many class A amplifiers use the same transistor(s) for both halves of the audio waveform. In this configuration, the output transistor(s) always has current flowing through it, even if it has no audio signal (the output transistors never 'turn off'). The current flowing through it is D.C.

A pure class 'A' amplifier is very inefficient and generally runs very hot even when there is no audio output. The current flowing through the output transistor(s) (with no audio signal) may be as much as the current which will be driven through the speaker load at FULL audio output power. Many people believe class 'A' amps to sound better than other configurations (and this may have been true at some point in time) but a well designed amplifier won't have any 'sound' and even the most critical 'ear' would be hard-pressed to tell one design from another.

NOTE: Some class A amplifiers use complimentary (separate transistors for positive and negative halves of the waveform) transistors for their output stage.

Class B

Circuit Operation



Class B output stage.

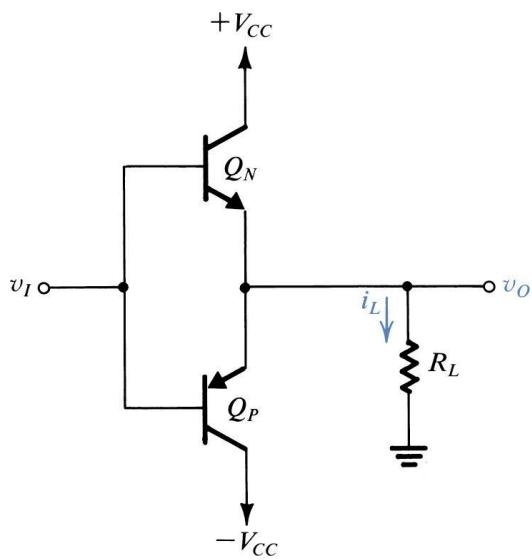
CLASS 'B'

A class 'B' amplifier uses complimentary transistors for each half of the waveform.

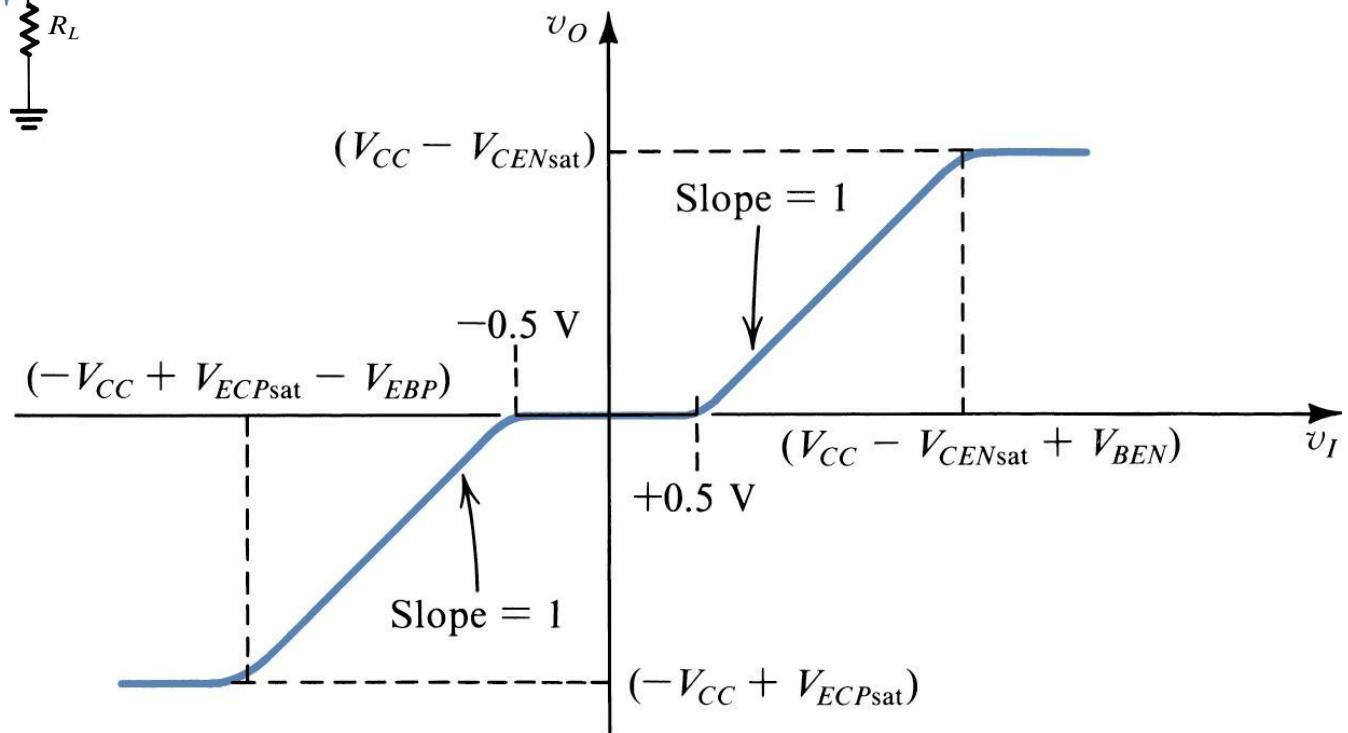
A true class 'B' amplifier is NOT generally used for audio. In a class 'B' amplifier, there is a small part of the waveform which will be distorted. You should remember that it takes approximately .6 volts (measured from base to emitter) to get a bipolar transistor to start conducting. In a pure class 'B' amplifier, the output transistors are not "biased" to an 'on' state of operation. This means that the the part of the waveform which falls within this .6 volt window will not be reproduced accurately.

The output transistors for each half of the waveform (positive and negative) will each have a .6 volt area in which they will not be conducting. The distorted part of the waveform is called 'crossover' or 'notch' distortion. Remember that distortion is any unwanted variation in a signal (compared to the original signal). The diagram below shows what crossover distortion looks like.

Class B

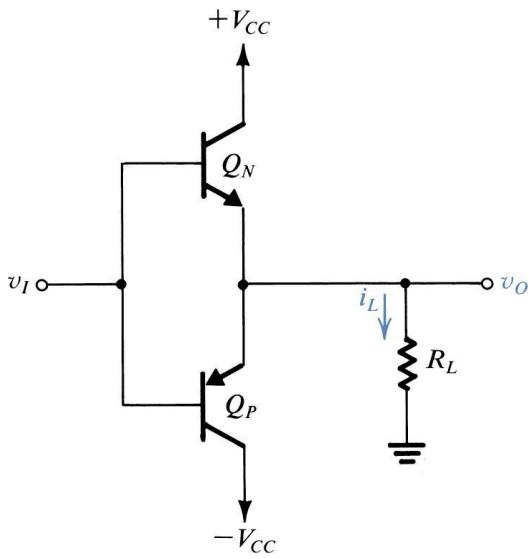


Circuit Operation



Transfer characteristic for the class B output stage in Fig. 9.5.

Operation



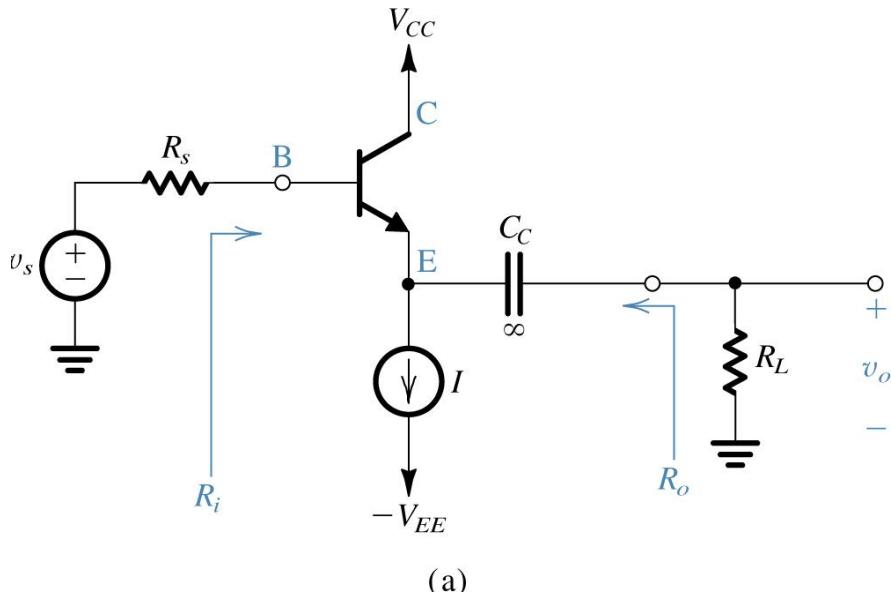
When the input voltage rises to be large enough to overcome the V_{be} , it will begin to cause an output voltage to appear. This occurs because Q_N begins to act like an emitter follower and Q_P shuts off. The input will be followed on the emitter until the transistor reaches saturation. The maximum input voltage is equal to the following:

$$v_{imax} = V_{CC} - V_{CENsat}$$

The same thing will begin to happen if the input voltage is negative by more than the V_{eb} of the transistor. This causes the Q_P to act like an emitter follower and Q_N turns off. This will continue to behave this way until saturation occurs at a minimum input voltage of:

$$v_{imin} = -V_{CC} + V_{ECPsat}$$

Emitter Follower Configuration (Chapter 4)



$$\frac{v_b}{v_s} = \frac{(\beta + 1)(r_e + par(R_L, r_o))}{R_S + (\beta + 1)(r_e + par(R_L, r_o))}$$

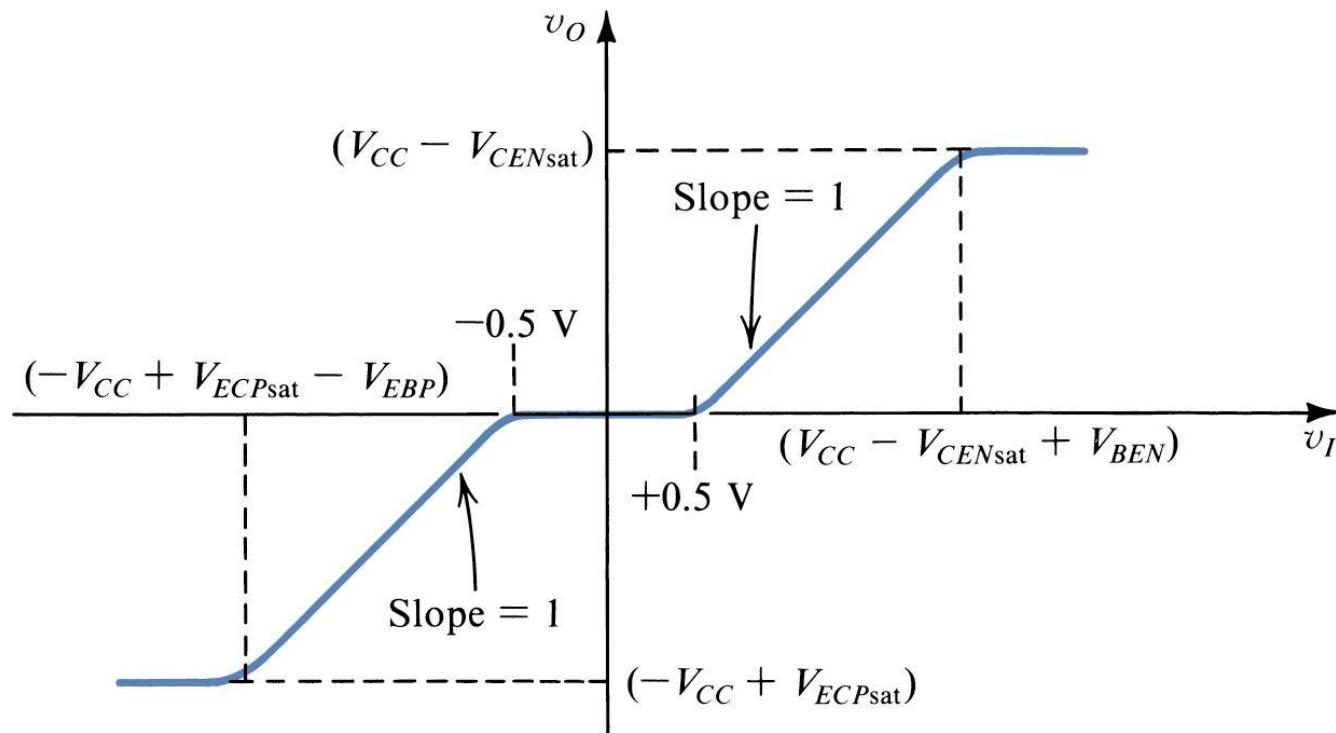
$$\frac{v_o}{v_b} = \frac{par(r_o, R_L)}{r_e + par(r_o, R_L)}$$

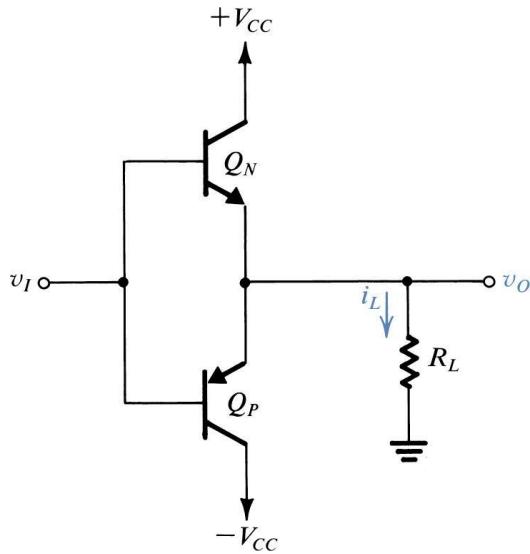
R_s will be small for most configurations, so the v_b/v_s will be a little less than unity. The same is true for r_e , so v_o/v_b will be a little less than unity making our v_o/v_s a little less than unity.

Characteristics of the Emitter Follower:

- High Input Resistance
- Low Output Resistance
- Near Unity Gain

Transfer Characteristic





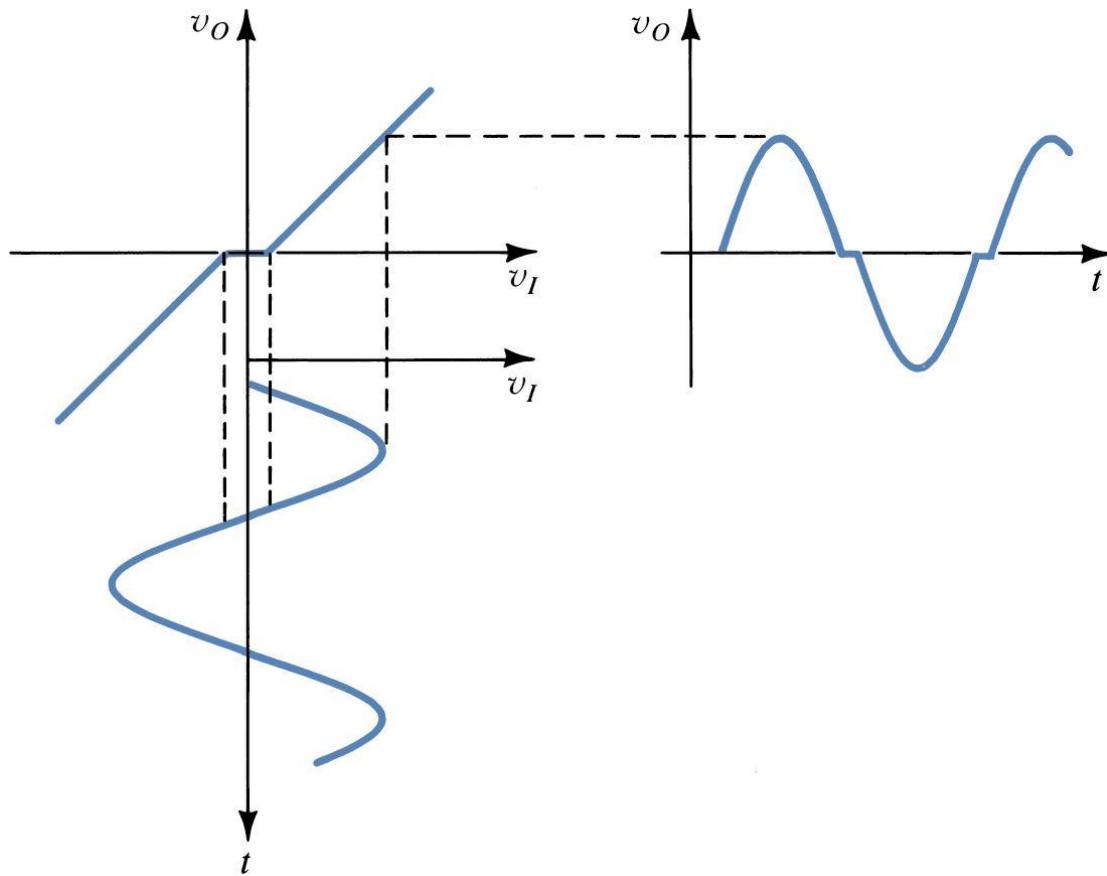
Push-Pull Nature of Class B

- Push: The npn transistor will push the current to ground when the input is positive.
- Pull: The pnp transistor will pull the current from the ground when the input is negative.

Crossover Distortion

The Crossover Distortion is due to the dead band of input voltages from $-.5V$ to $.5V$. This causes the Class B output stage to be a bad audio amplifier. For large input signals, the crossover distortion is limited, but at small input signals, it is most pronounced.

Graph of Crossover Distortion



Illustrating how the dead band in the class B **transfer** characteristic results in crossover distortion.

Power Efficiency

Load Power:

$$P_L = \frac{1}{2} \cdot \frac{V_{op}^2}{R_L}$$

Since each transistor is only conducting for one-half of the time, the power drawn from each source will be the same.

$$P_S = \frac{1}{\pi} \cdot \frac{V_{op}}{R_L} \cdot V_{CC}$$

This efficiency will be at a max when V_{op} is at a max. Since V_{op} cannot exceed V_{CC} , the maximum efficiency will occur at $\pi/4$.

$$\eta = \frac{P_L}{2 \cdot P_S} = \frac{\frac{1}{2} \cdot \frac{V_{op}^2}{R_L}}{2 \cdot \frac{1}{\pi} \cdot \frac{V_{op}}{R_L} \cdot V_{CC}}$$

$$\eta = \frac{\pi}{4} \cdot \frac{V_{op}}{V_{CC}}$$

$$\eta_{max} = \frac{\pi}{4}$$

This will be approximately 78.5%, much greater than the 25% for Class A.

Class AB

Circuit Operation

Crossover distortion can be eliminated by biasing the transistors at a small, non-zero current.

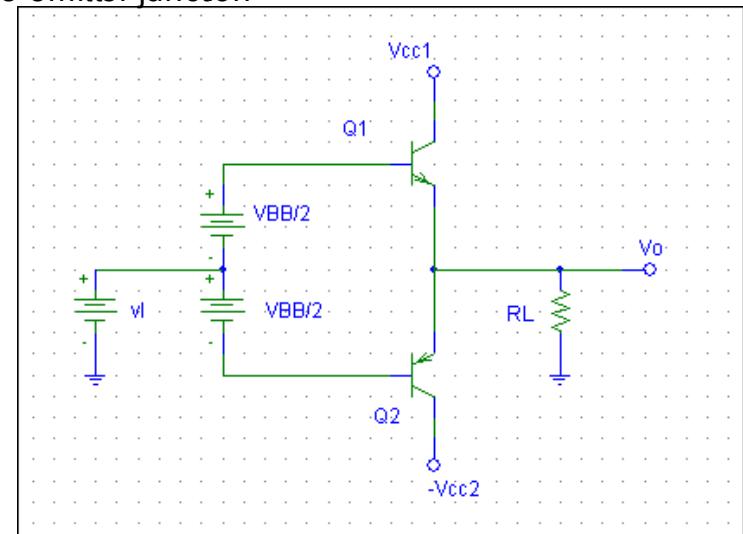
A bias Voltage V_{BB} is applied between Q_n and Q_p .

For $v_i = 0$, $v_o = 0$, and a voltage $V_{BB}/2$ appears across the base-emitter junction of each transistor.

$$i_N = i_P = I_Q = I_S \cdot e^{\frac{V_{BB}}{2 \cdot V_T}}$$

V_{BB} is selected to result the required quiescent current I_Q

$$v_o = v_i + \frac{V_{BB}}{2} - v_{BEN}$$



$$i_N = i_P + i_L$$

$$v_{BEN} + v_{EBP} = V_{BB}$$

$$V_T \cdot \ln\left(\frac{i_N}{I_S}\right) + V_T \cdot \ln\left(\frac{i_P}{I_S}\right) = 2 \cdot V_T \cdot \ln\left(\frac{i_Q}{I_S}\right)$$

$$i_N^2 = I_Q^2$$

$$i_N^2 - i_L \cdot i_N - I_Q^2 = 0$$

Class AB

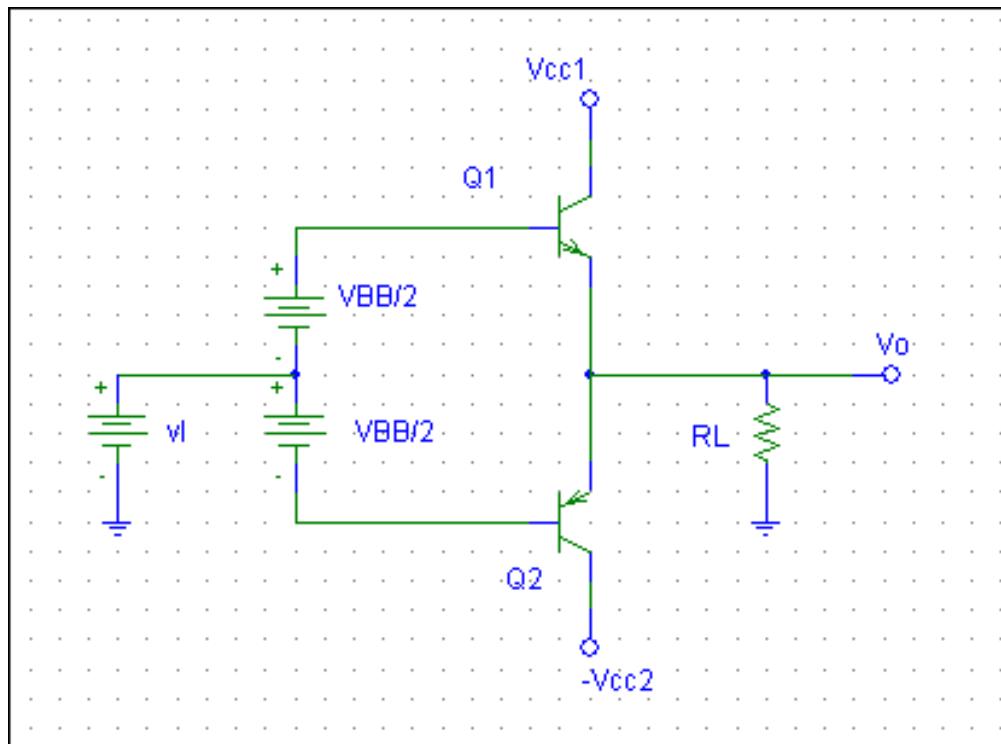
Exercise 9.6

Calvin College - ENGR 332 Class AB Output Stage Amplifier

Consider the class AB circuit (illustrated below) with $V_{cc} = 15 \text{ V}$, $I_Q = 2 \text{ mA}$, $R_L = 100 \text{ ohms}$. Determine V_{BB} . Determine the values of i_L , i_N , i_P , v_{BEN} , v_{EBP} , v_l , v_O/v_l , R_{out} , and v_o/v_i for v_O varying from -10 to 10V.

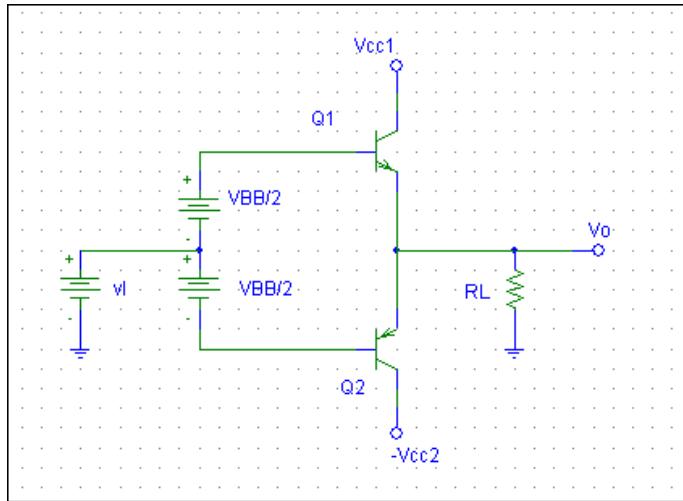
Note that v_O/v_l is the large signal voltage gain and v_o/v_i is the incremental gain obtained at $R_L/(R_L+R_{out})$. The incremental gain is equal to the slope of the transfer curve.

Assume QN and QP to be matched, with $I_S = 10E-13$.



Class AB

Exercise 9.6



under quiescent conditions $i_N = i_P = I_Q \quad v_O = v_I = 0$

Solving for V_{BB}

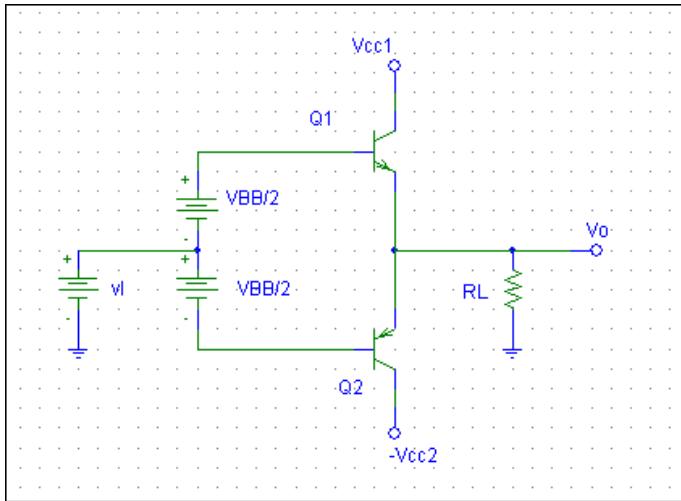
$$V_{BB} := 1 \quad IS := 10^{-13} \quad VT := 0.025 \quad I_Q := 2 \cdot 10^{-3} \quad RL := 10\Omega$$

Given
$$\frac{|V_{BB}|}{2} = \frac{VT}{I_Q} = IS \cdot e$$

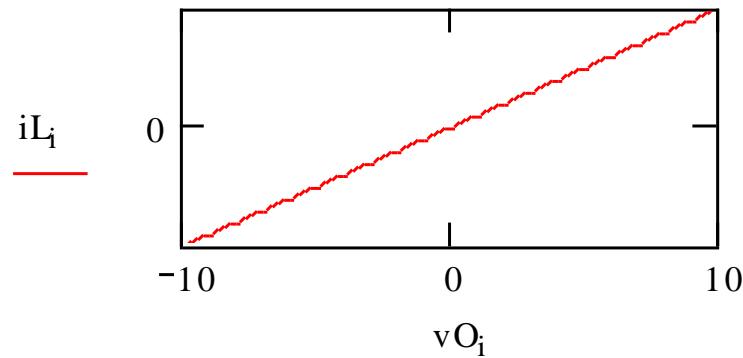
$$V_{BB} := \text{Find}(V_{BB}) \quad i := 0..100 \quad V_{BB} = 1.186$$

Class AB

Exercise 9.6



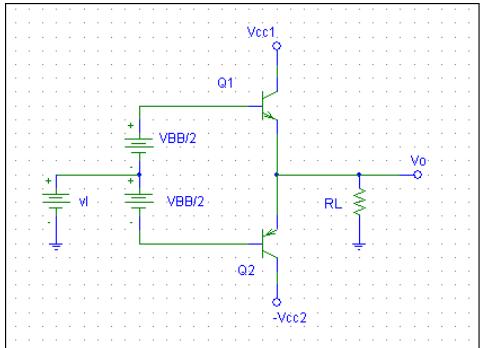
$$v_{O_i} := -10 + \frac{i}{5} \quad iL_1 := \frac{v_{O_i}}{RL}$$



Class AB

Solving for iN

Exercise 9.6



initial guesses $iN := 0.02$

$iLD := 0.02$

$IQ := 0.002$

Given

$$iN^2 - iLD \cdot iN - IQ^2 = 0$$

$$iNN(iQ, iLD) := \text{Find}(iN)$$

$$i := 0..100 \quad IQ_i := 0.002$$

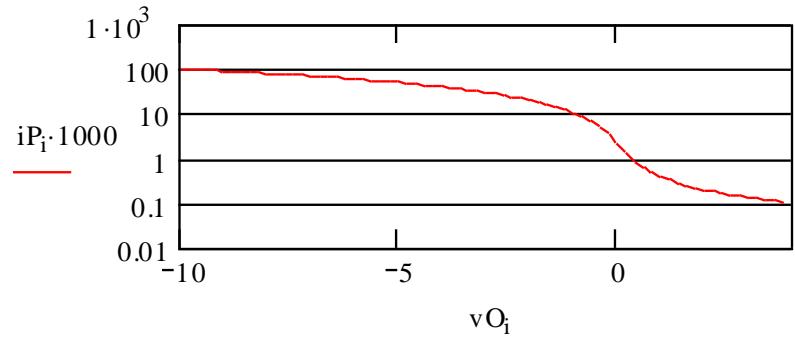
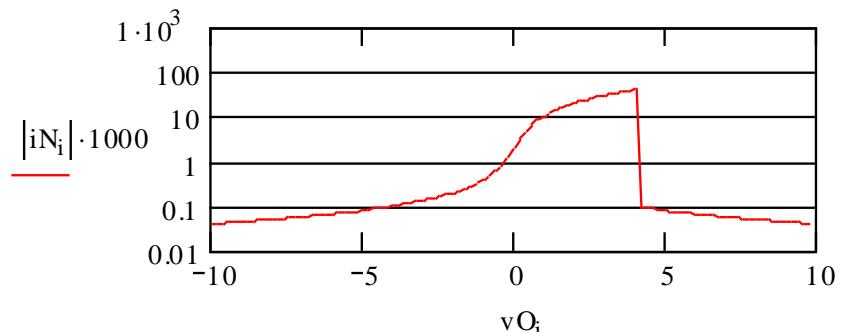
$$iLD_i := iL_i$$

$$iN_i := iNN(Q_i, L_i)$$

$$iN_{10} = 4.997 \times 10^{-5}$$

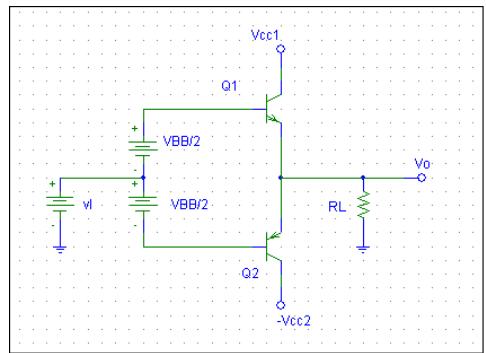
$$iP_i := iN_i - iLD_i$$

$$\left(iN_i \right)$$



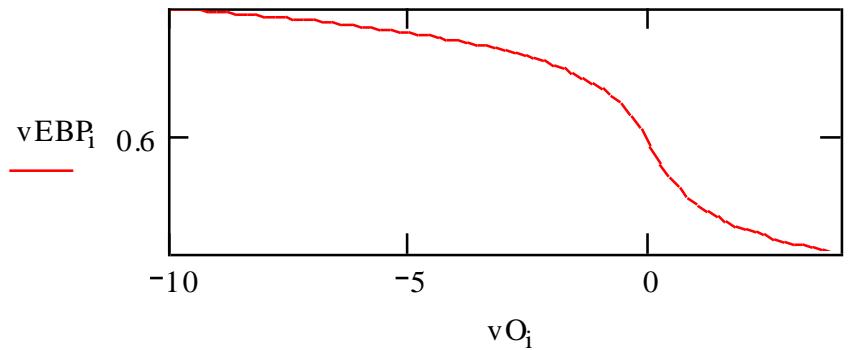
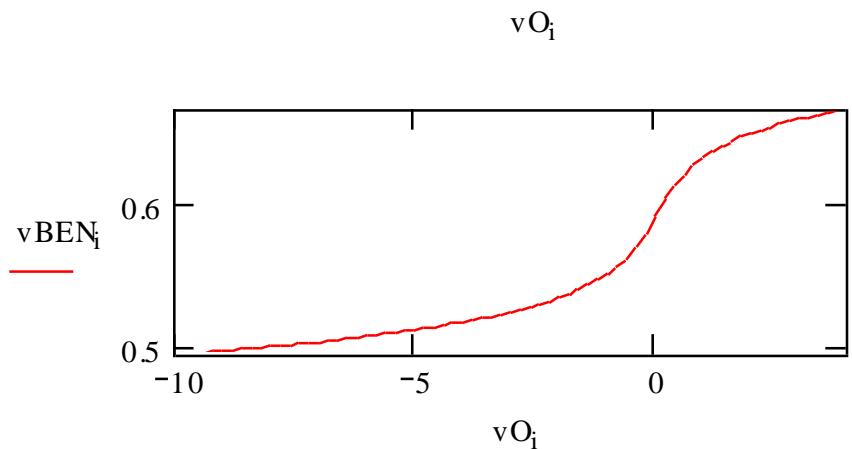
Class AB

Exercise 9.6



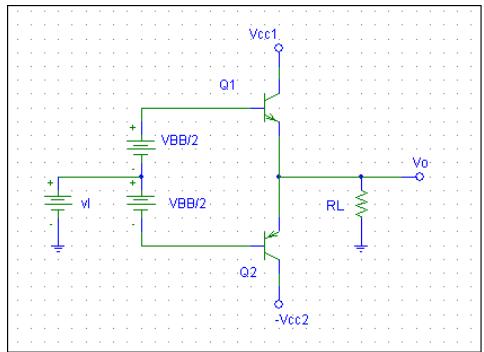
$$v_{BEN_i} := V_T \cdot \ln \left(\frac{iN_i}{IS} \right)$$

$$v_{EBP_i} := V_T \cdot \ln \left(\frac{iP_i}{IS} \right)$$



Class AB

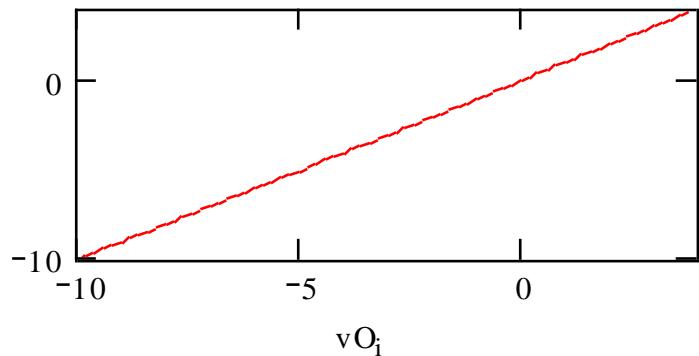
Exercise 9.6



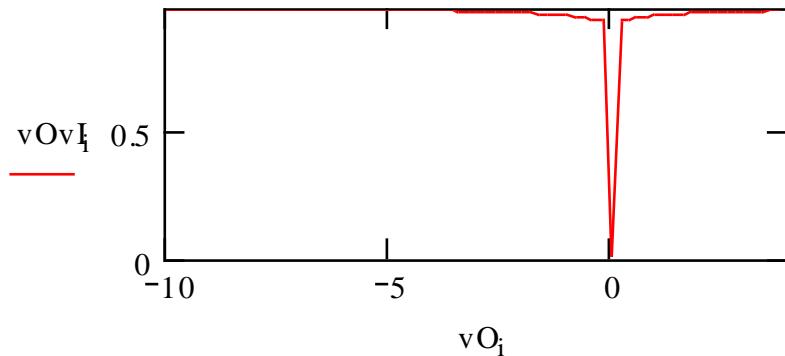
$$vI_i := vO_i + vBEN_i - \frac{VBB}{2}$$

$$vOvI_i := \frac{vO_i}{vI_i}$$

vI_i

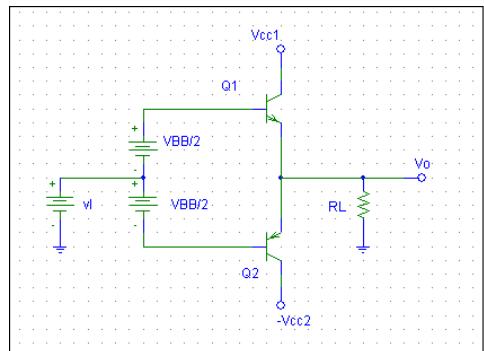


vOvI_i



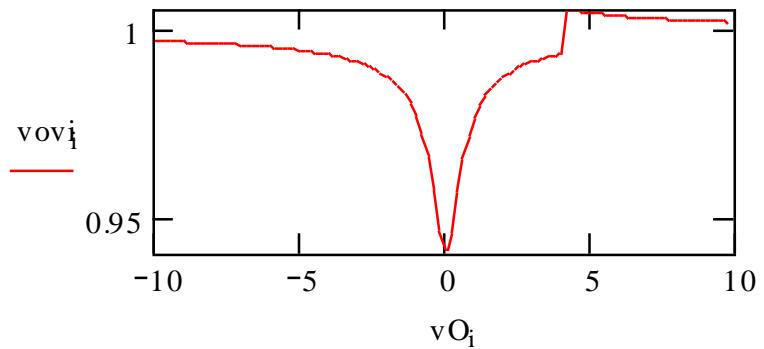
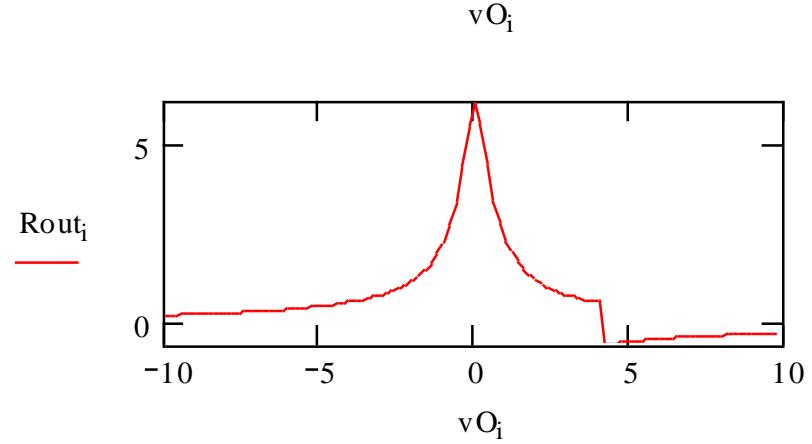
Class AB

Exercise 9.6



$$R_{out_i} := \frac{V_T}{iP_i + iN_i}$$

$$v_{ov_i} := \frac{RL}{RL + R_{out_i}}$$





Institute of Electronics
National Chiao Tung Univ.

Power Amplifier Design

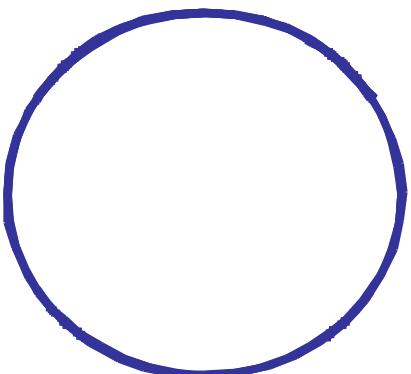
Design considerations

Power amplifier classes

Power amplifier examples

Linear/Nonlinear PA?

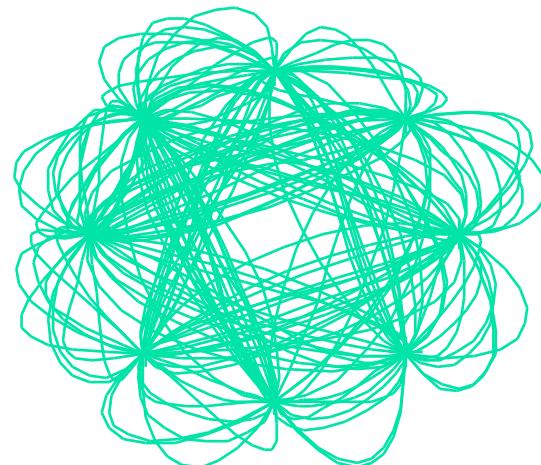
Constant envelop
modulation



GMSK
FSK

Nonlinear PA
High Efficiency

Nonconstant envelop
modulation

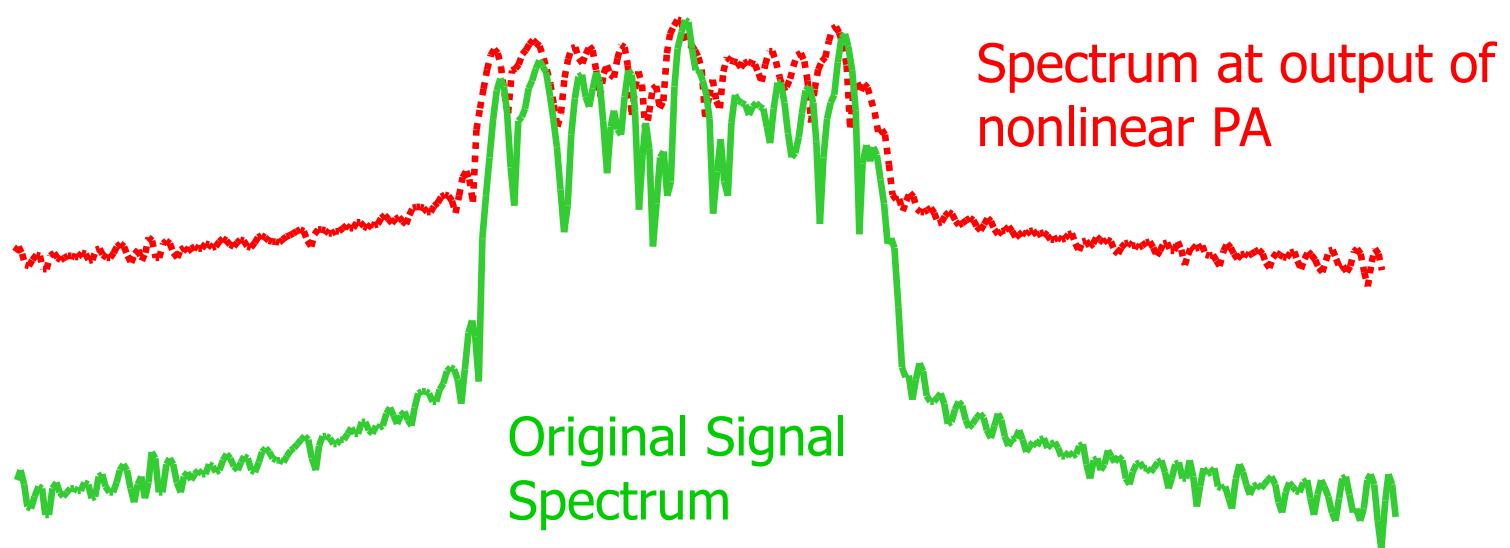


BPSK
QPSK
QAM

Linear PA
Low Efficiency

Spectral Regrowth

- Effect of nonlinear PA on nonconstant envelop signal



Power Amplifier Efficiency

- For ideal PA : $\eta = \frac{P_{out}}{P_{total}} = 1$
- The Drain Efficiency $\eta_{Drain} = \frac{P_{RFout}}{P_{DC}}$
- The Power Added Efficiency(PAE):
$$PAE = \frac{P_{RFout} - P_{RFin}}{P_{DC}}$$
- The overall efficiency:
$$\eta = \frac{P_{RFout}}{P_{DC} + P_{RFin}}$$



Basic Amplification:

Use RFC (RF Chock) to in a common source stage to drive the load

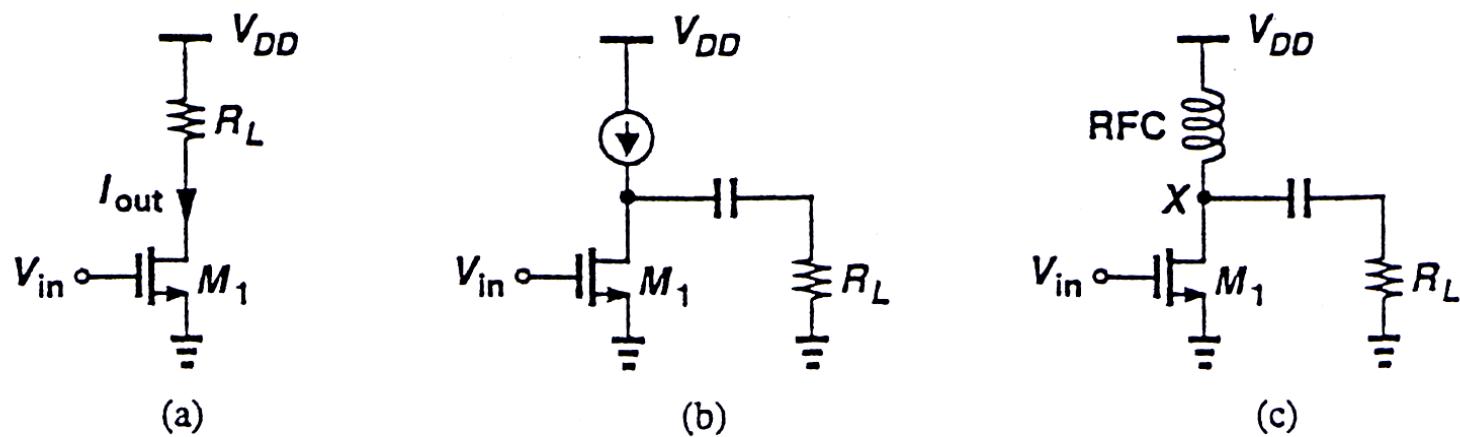
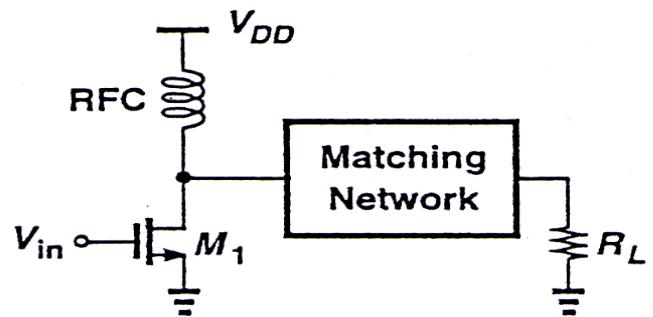
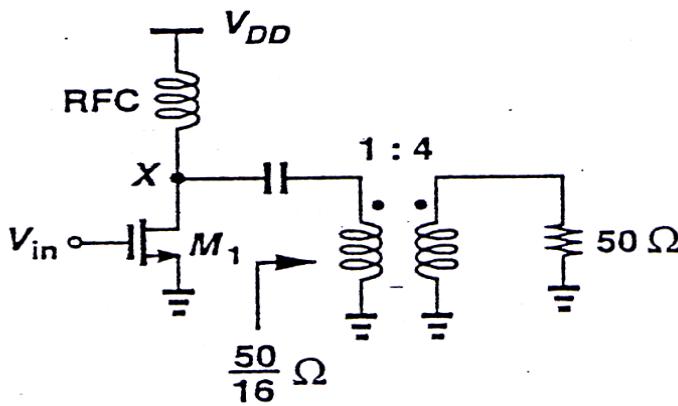


Figure 9.1 Common-source stages with different type of load connections.

Matching



(a)



(b)

Figure 9.2 (a) Matching network as voltage amplifier, (b) use of a transformer as a matching network.



Typical PA Performance

TABLE 9.1 Typical PA performance.

Output Power	+20 to +30 dBm
Efficiency	30% to 60%
IMD	-30 dBc
Supply Voltage	3.8 to 5.8 V
Gain	20 to 30 dB
Output Spurs and Harmonics	-50 to -70 dBc
Power Control	On-Off or 1-dB Steps
Stability Factor	> 1

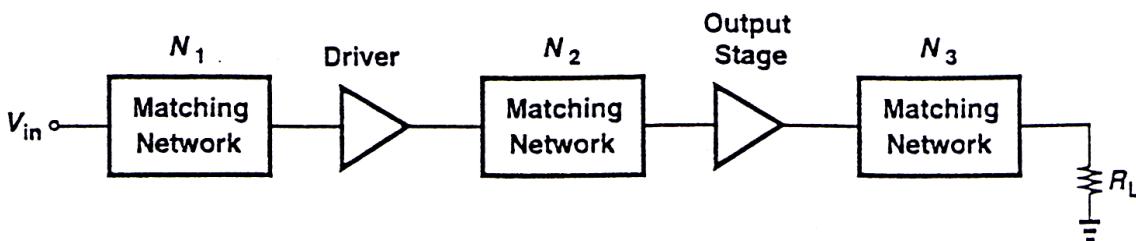


Figure 9.22 Typical PA system.

Linear and Nonlinear PA

- Linear/Nonlinear distinction
 - The fraction of the RF cycle for which the transistor conducts.

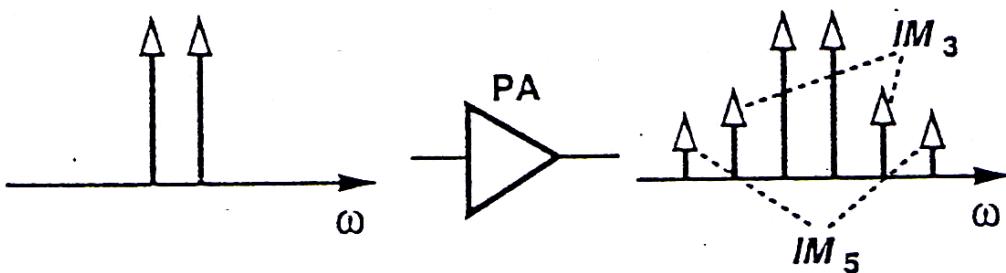
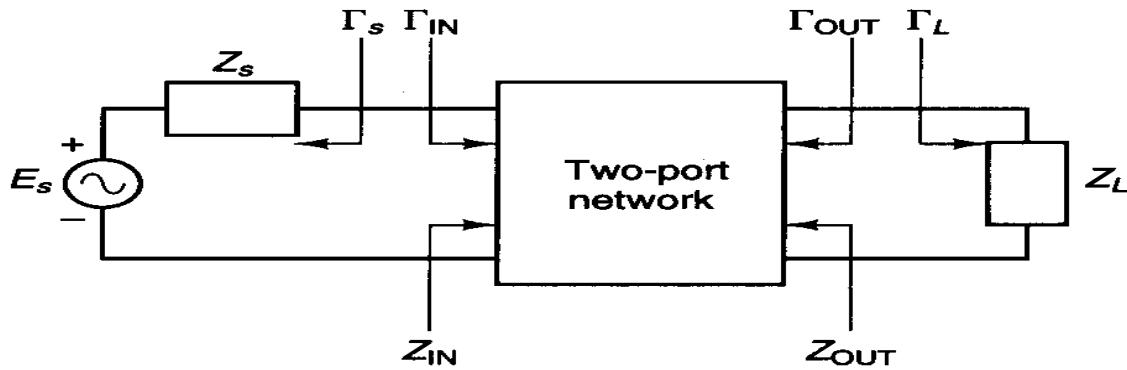


Figure 9.3 Two-tone test of a power amplifier.

Stability Consideration



$$|\Gamma_s| < 1, |\Gamma_L| < 1,$$

$$|\Gamma_{in}| = \left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| < 1,$$

$$|\Gamma_{out}| = \left| S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \right| < 1$$

Unconditionally Stable :

$$K > 1, |\Delta| < 1, \text{ where}$$

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|},$$

$$\Delta = S_{11}S_{22} - S_{12}S_{21}$$

Ps: Stable circle on Smith chart is the general tool

Operating Power Gain

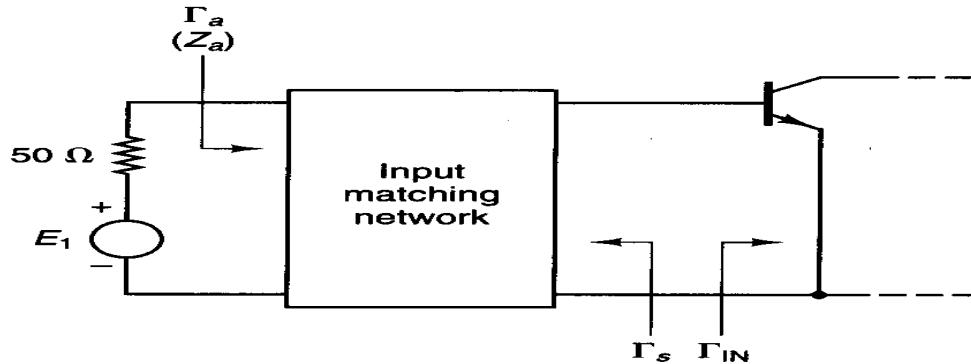
$$G_P = \frac{P_L}{P_{IN}} = \frac{\text{power delivered to load}}{\text{power input to network}},$$

$$G_P = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}, \text{ substitute } \Gamma_{IN},$$

$$G_P = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2)}{(1 - \left| \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} \right|)^2 |1 - S_{22}\Gamma_L|^2} = |S_{21}|^2 g_p$$

- Given the power gain, drawing the power gain circle, and select Γ_L in the stable region.
- Calculate Γ_{In} , determine if a conjugate match is in the stable region. If it's not stable, we can choose Γ arbitrarily, or according VSWR.

Constant VSWR Circle



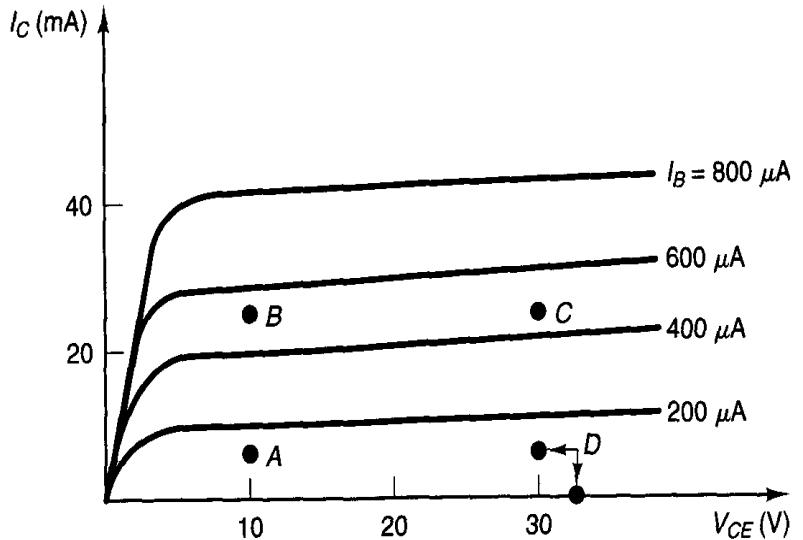
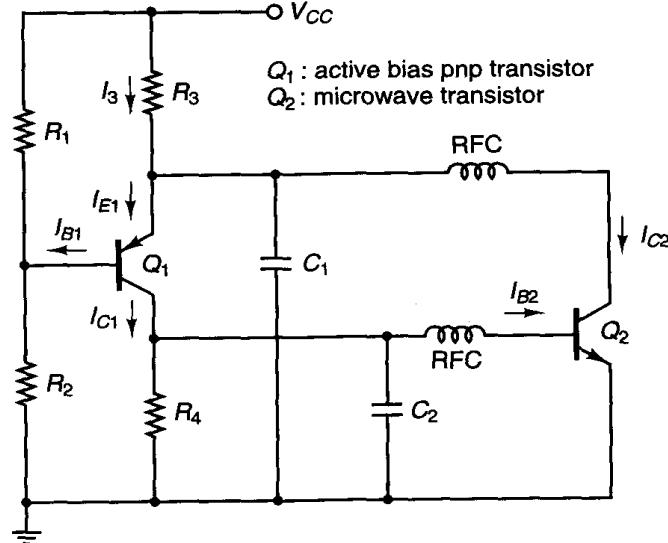
$$(VSWR)_{in} = \frac{1 + |\Gamma_a|}{1 - |\Gamma_a|} \geq 1, \text{ where}$$

$$|\Gamma_a| = \left| \frac{\Gamma_{IN} - \Gamma_s^*}{1 - \Gamma_{IN}\Gamma_s} \right|,$$

$$(VSWR)_{out} = \frac{1 + |\Gamma_b|}{1 - |\Gamma_b|} \geq 1, \text{ where}$$

$$|\Gamma_b| = \left| \frac{\Gamma_{out} - \Gamma_L^*}{1 - \Gamma_{out}\Gamma_L} \right|.$$

DC Bias Selection



Active bias network for a BJT

Low-noise,low-power : A

Low-noise,higher power-gain : B

High Output Power : C

Higher output power and higher efficiency : D

■ Power Amplifier Classes

- Class A: conduction angle 360
- Class B: conduction angle 180
- Class AB: conduction angle >180
- Class C: conduction angle <180
- Class F: an extension of class C
- Class E: switch mode

Class A Power Amplifiers

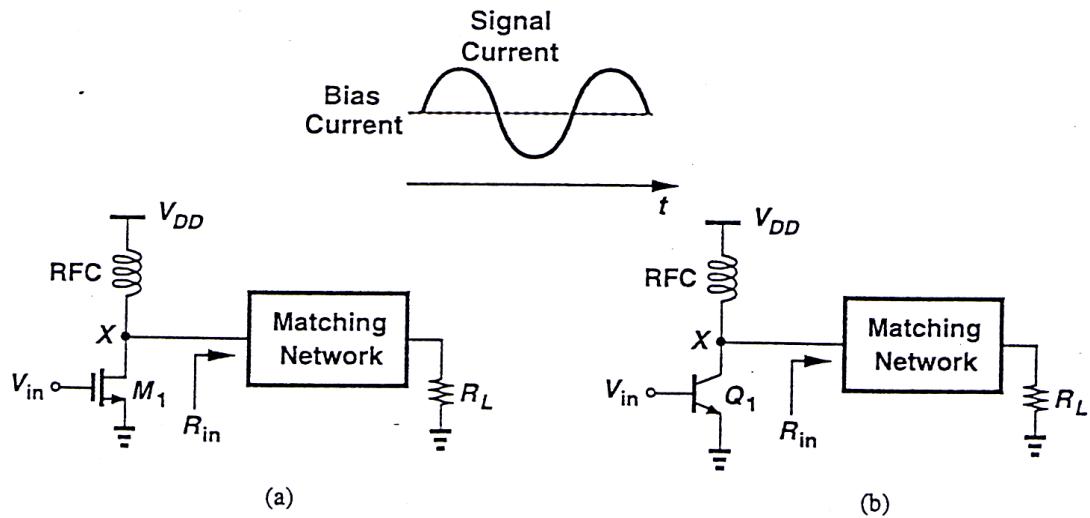
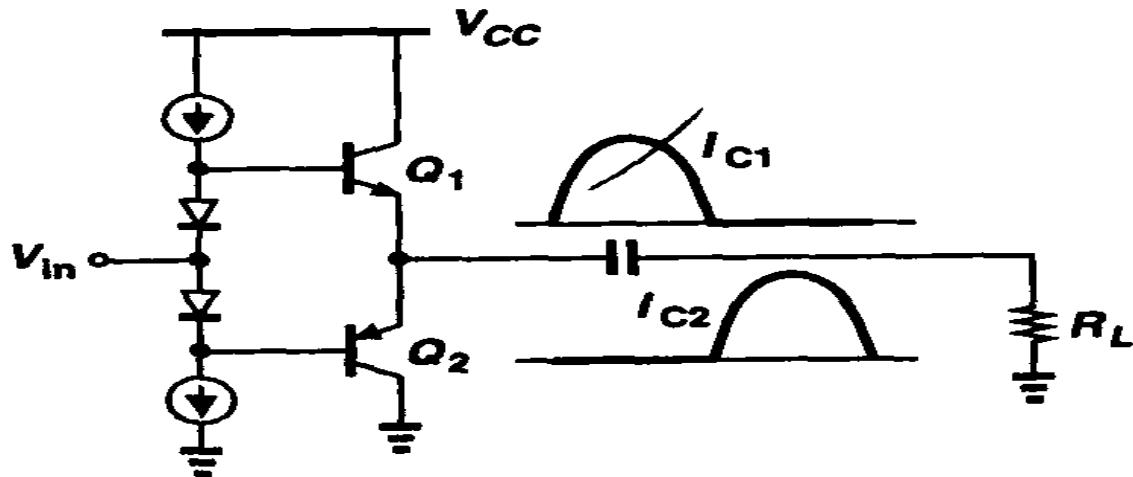


Figure 9.4 Class A stages using (a) MOS device, (b) bipolar transistor.

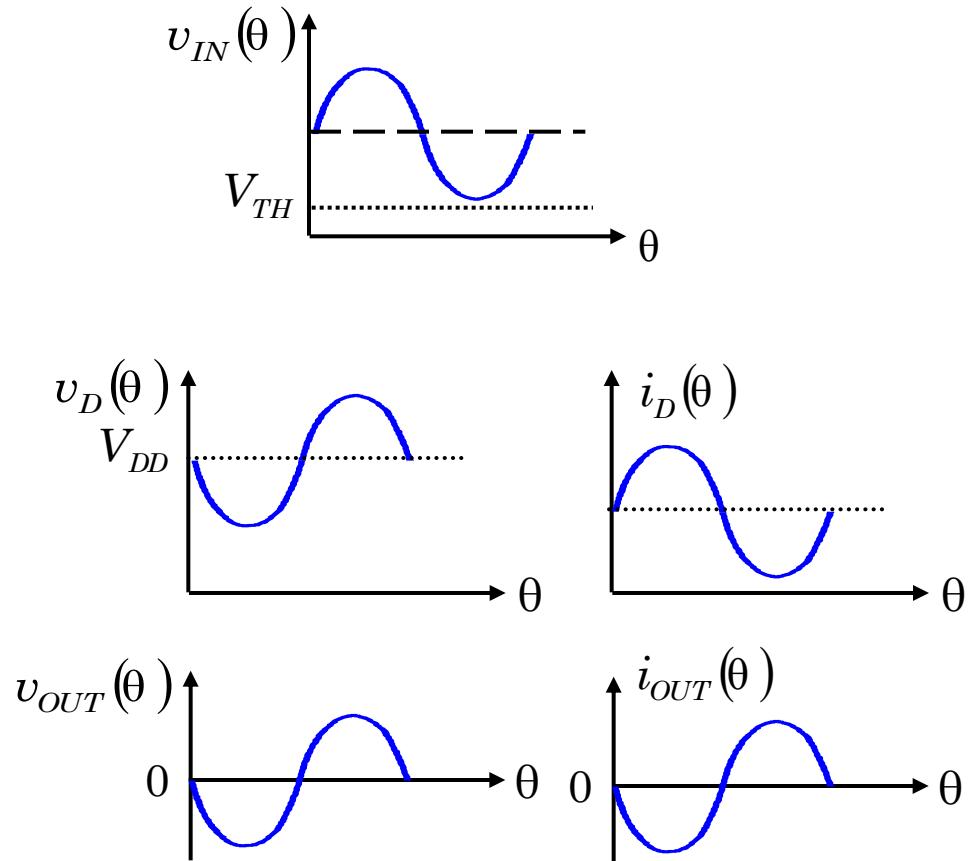
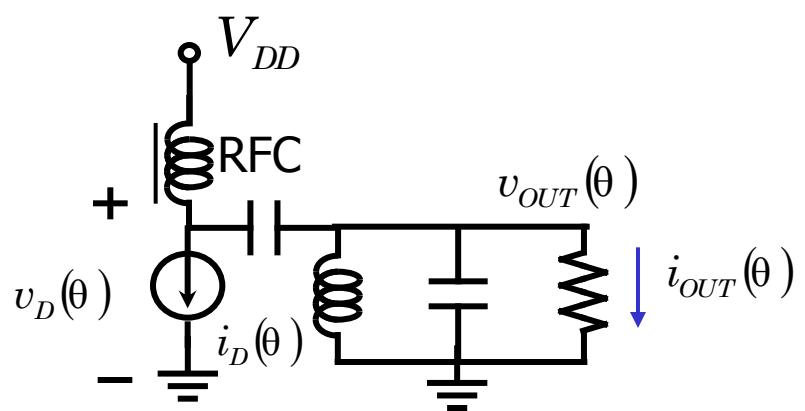
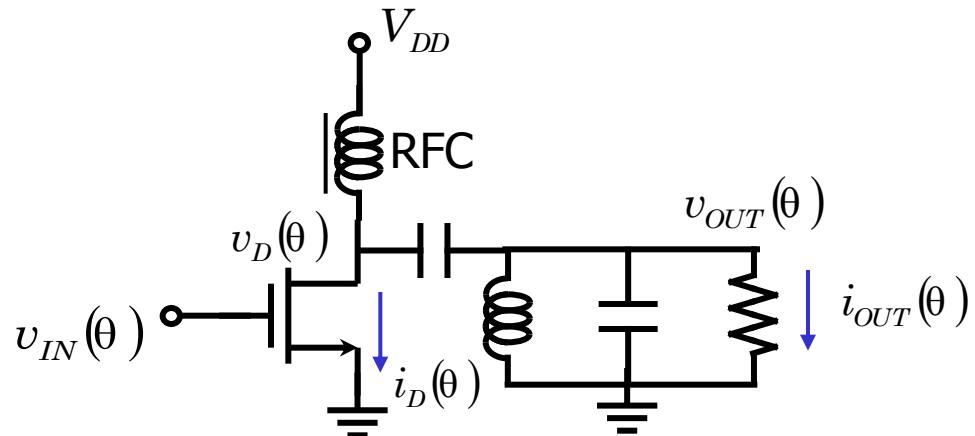
- Maximum efficiency of class A PA:
Assume drain(collector) voltage is a sinusoid having V_{pp} of $2V_{dd}$. The power delivered to matching network is $V_{DD}^2 / 2R_{in}$. And for V_x to reach $2V_{dd}$, the RFC must provide a current of V_{DD}^2 / R_{in} . Thus, the maximum efficiency is 50%.

Push-pull output stage



- The push-pull stage of above usually used in low-frequency power amplifier.
- The efficiency is better than class A PA.

Class A Power Amplifiers



Class A Power Amplifiers

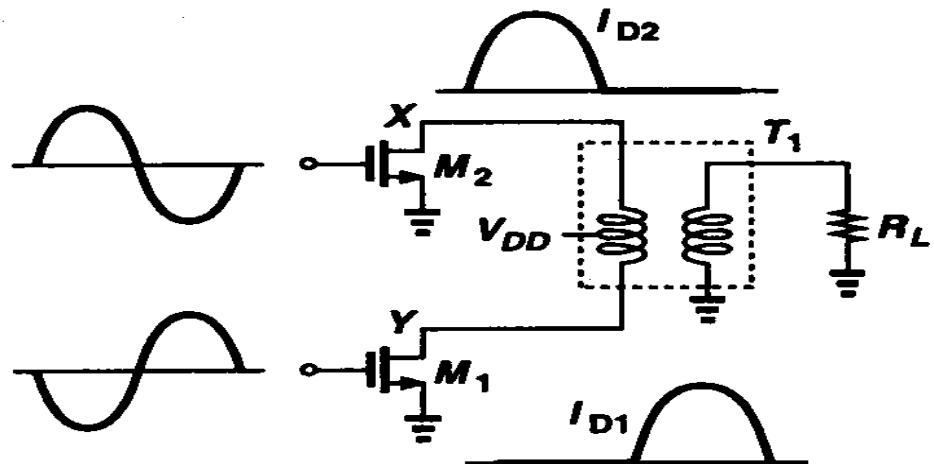
- Maximum output power

$$P_{RFout} = \frac{V_{om}^2}{2R} \leq \frac{V_{DD}^2}{2R}$$

- Efficiency

$$\eta_{Drain} = \frac{P_{RFout}}{P_{DC}} = \frac{\cancel{V_{om}^2 / 2R}}{\cancel{V_{DD}^2 / R}} = \frac{V_{om}^2}{2V_{DD}^2} \leq \frac{1}{2}$$

Class B Stage using a transform

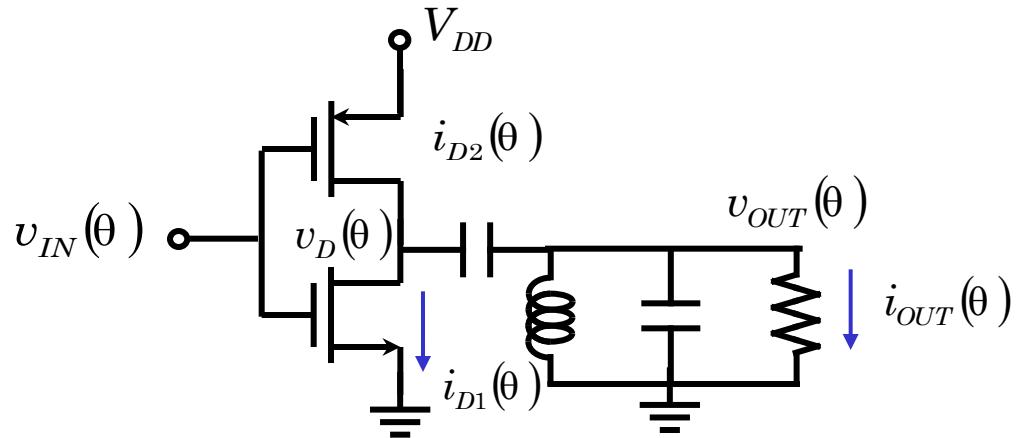


$$I_{DD,avg} = \frac{2}{T} \int_0^{T/2} \frac{V_{DD}}{n^2 R_L} \sin \omega t dt \quad (9.1)$$

- The maximum voltage swing at X and Y is $2V_{DD}$, And the equivalent resistance seen at each of X and Y is $n^2 R_L$
- The total input power of T1 is given by $P_{in} = V_{DD}^2 / 2 n^2 R_L$ and $P_{sup} = 2V_{DD}^2 / (\pi n^2 R_L)$.

$$\eta = P_{in} / P_{sup} = \pi / 4 \approx 79\%.$$

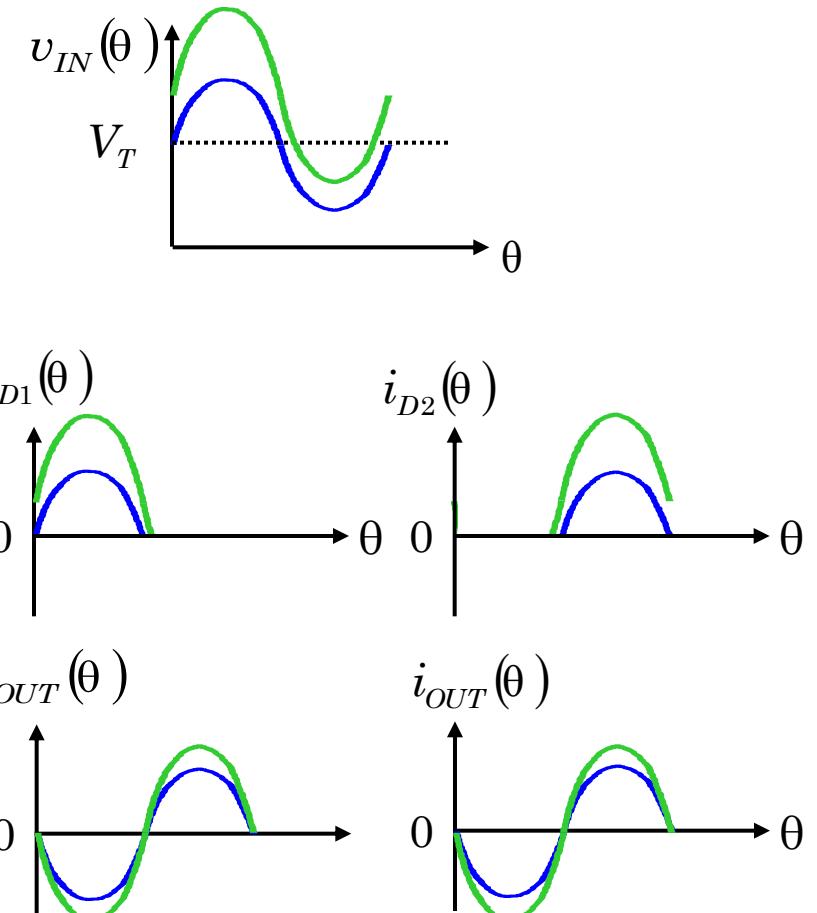
Class B, AB Power Amplifiers



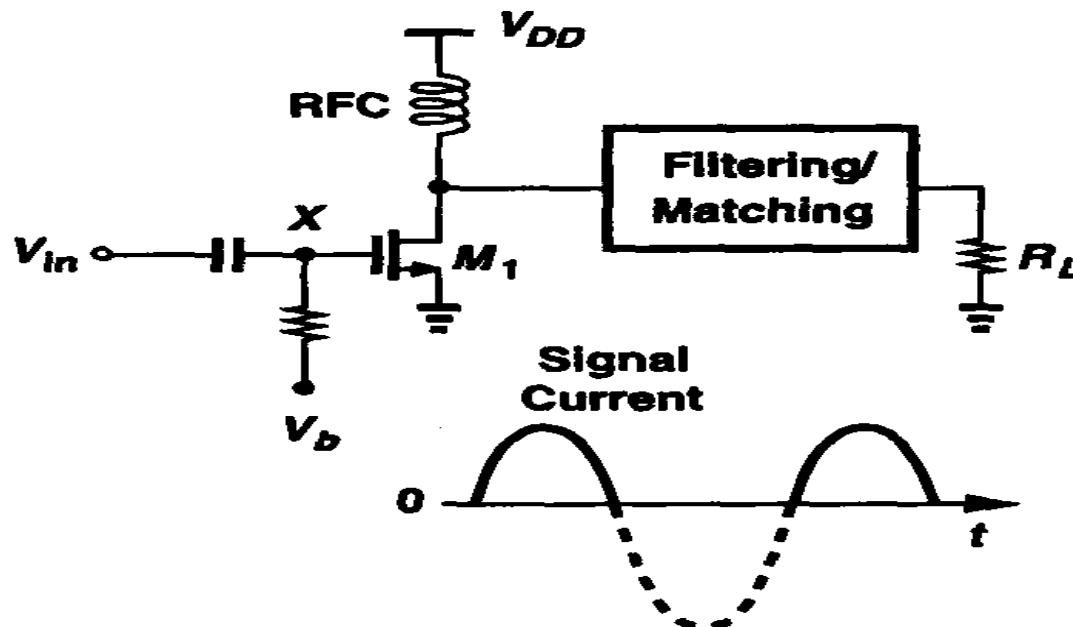
$$P_{RFout} = \frac{V_{om}^2}{2R} \leq \frac{V_{DD}^2}{2R}$$

$$I_{DC} = \frac{1}{2\pi} \int_0^{2\pi} I_D |\sin \theta| d\theta = \frac{2I_D}{\pi} = \frac{2}{\pi} \frac{V_{om}}{R}$$

$$\eta_{Drain} = \frac{P_{RFout}}{P_{DC}} = \frac{\frac{V_{om}^2}{2R}}{\frac{2}{\pi} \frac{V_{om}}{R} V_{DD}} = \frac{\pi}{4} \frac{V_{om}}{V_{DD}} \leq \frac{\pi}{4} \approx 0.785$$

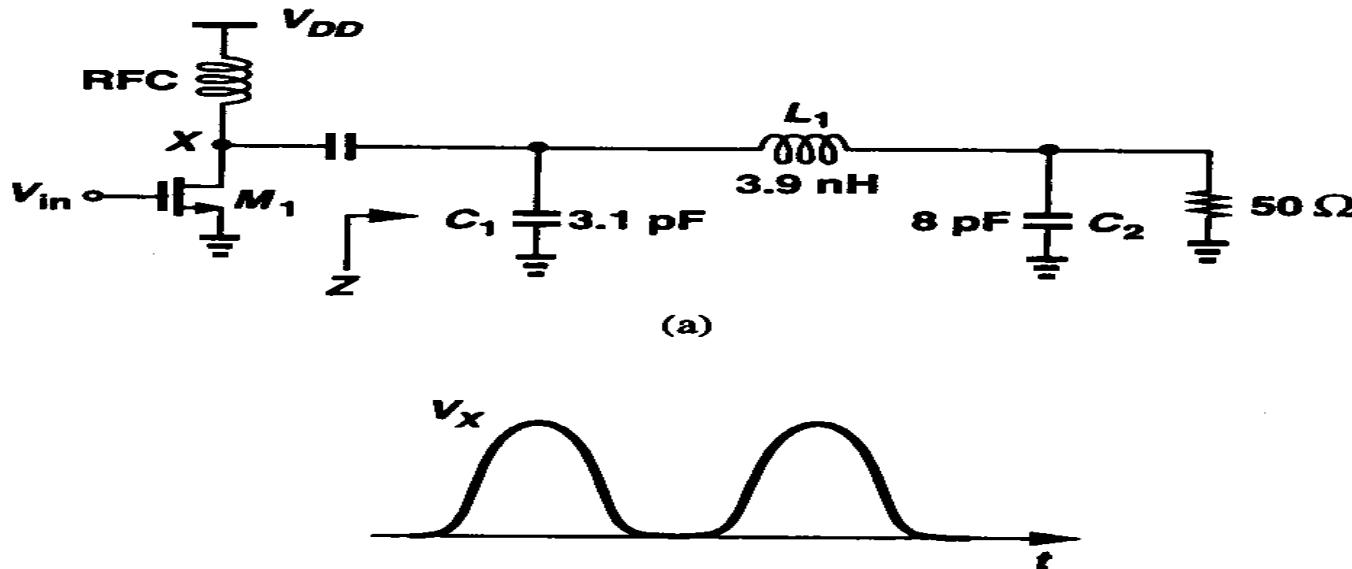


Class C PAs



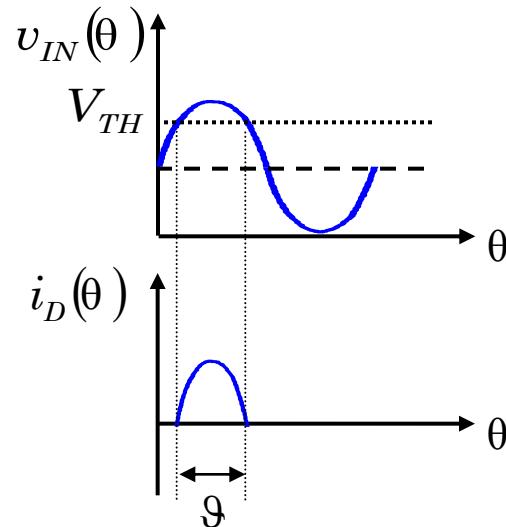
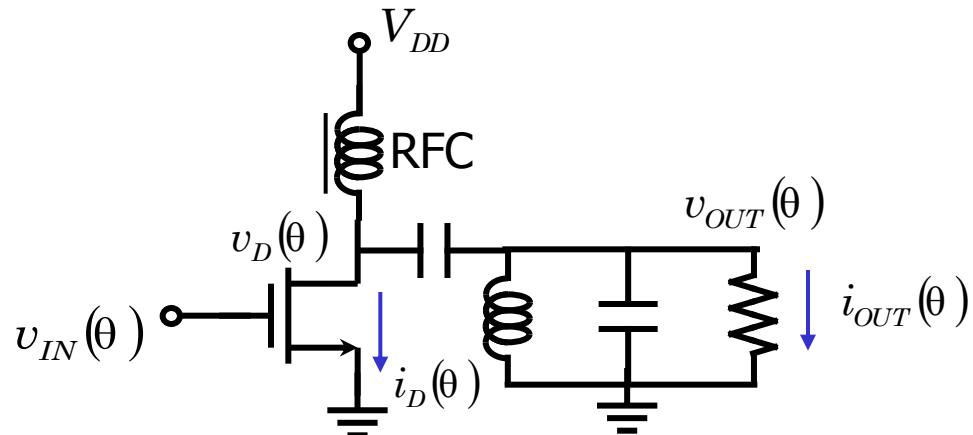
- M_1 turns On if $V_{in} > |V_b| + V_{TH}$
- The efficiency formula : $\eta = \frac{1}{4} \frac{\theta - \sin \theta}{\sin(\theta/2) - \theta/2 \cos(\theta/2)}$
- The power delivered to the load : $P_{out} \propto \frac{\theta - \sin \theta}{1 - \cos(\theta/2)}$

Ideas for Raising Efficiency



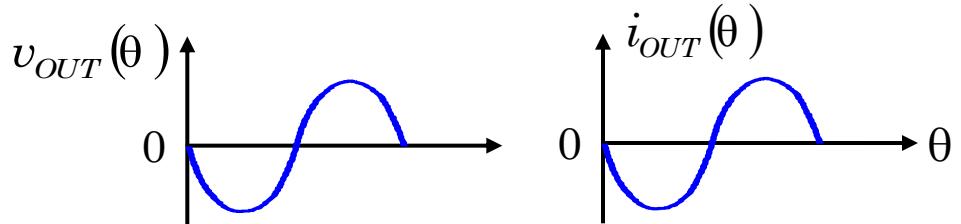
- Suppose the matching network is designed such that its input impedance is low at the fundamental frequency and quite high at the second harmonics. The drain voltage exhibits sharper edges than a sinusoid does, raising the efficiency.
- But the matching network becomes quite complex and lossy.

Class C Power Amplifier



$$P_{RFout} \propto \frac{\vartheta - \sin \vartheta}{1 - \cos(\vartheta/2)}$$

$$\eta_{\text{Drain}} = \frac{P_{RFout}}{P_{DC}} = \frac{1}{4} \frac{\vartheta - \sin \vartheta}{\sin(\vartheta/2) - \vartheta/2 \cos(\vartheta/2)}$$



High Efficiency PA

Class A

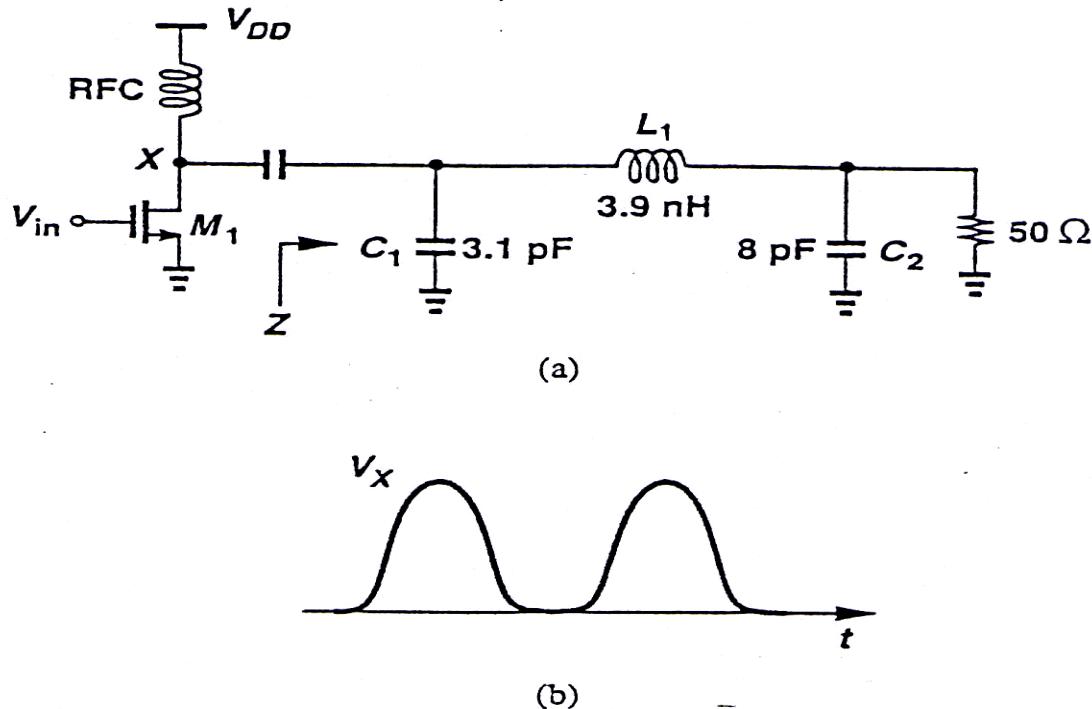
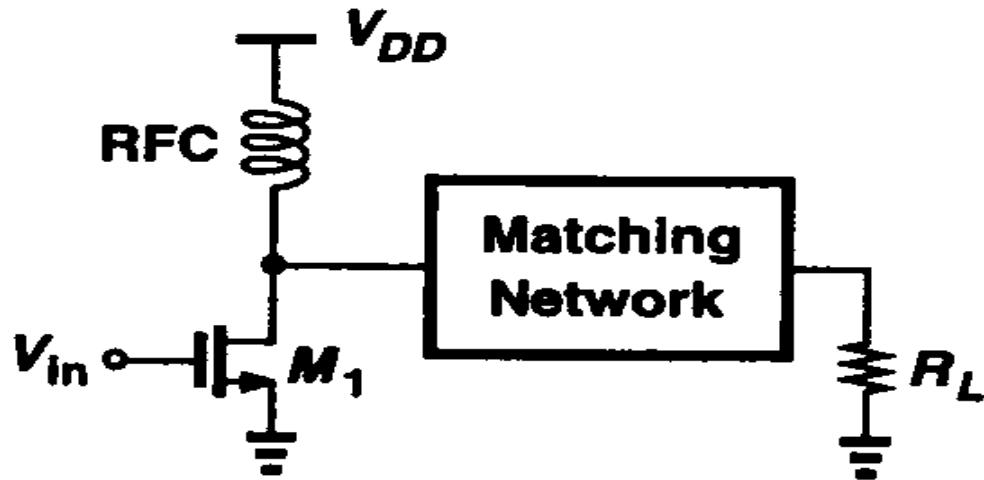


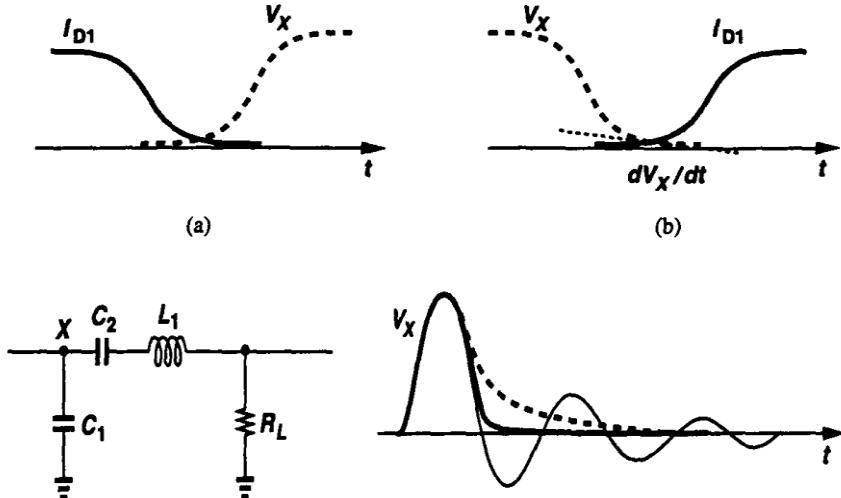
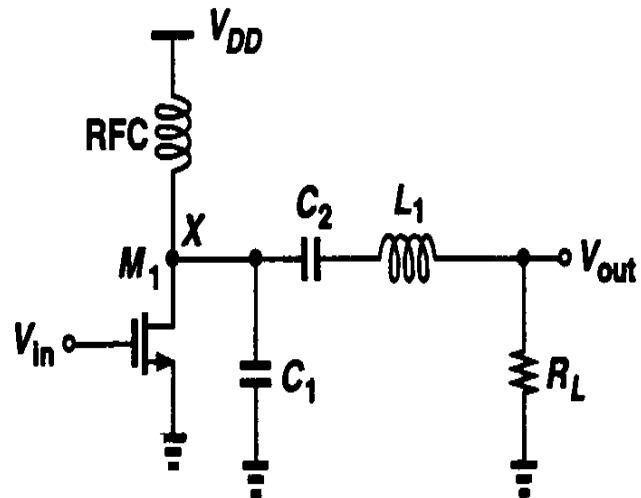
Figure 9.8 (a) PA with high harmonic termination, (b) drain voltage waveform.

Class E PAs



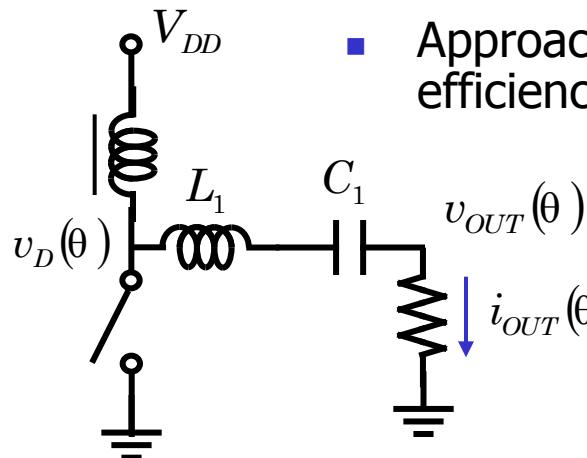
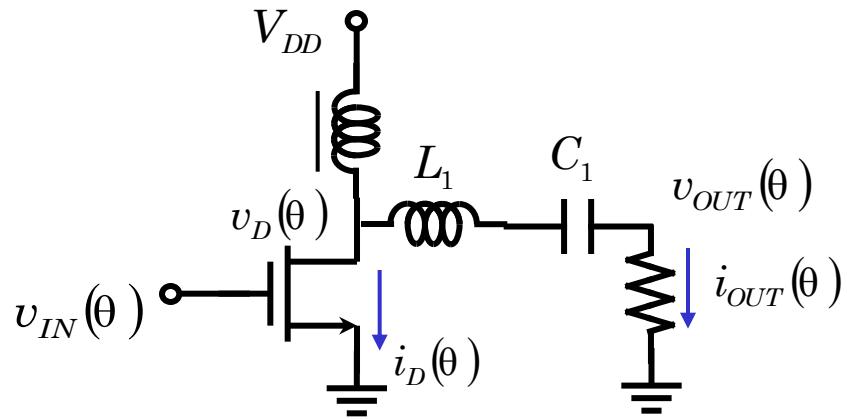
- Class E stages are nonlinear amplifiers that achieve efficiencies approaching 100% while delivering full power.
- It's a “switching power amplifier”.
- The voltage applied to the gate of M_1 must approximate a rectangular waveform. And the switch on-resistance must be low.

Class E Pas (Cont.)

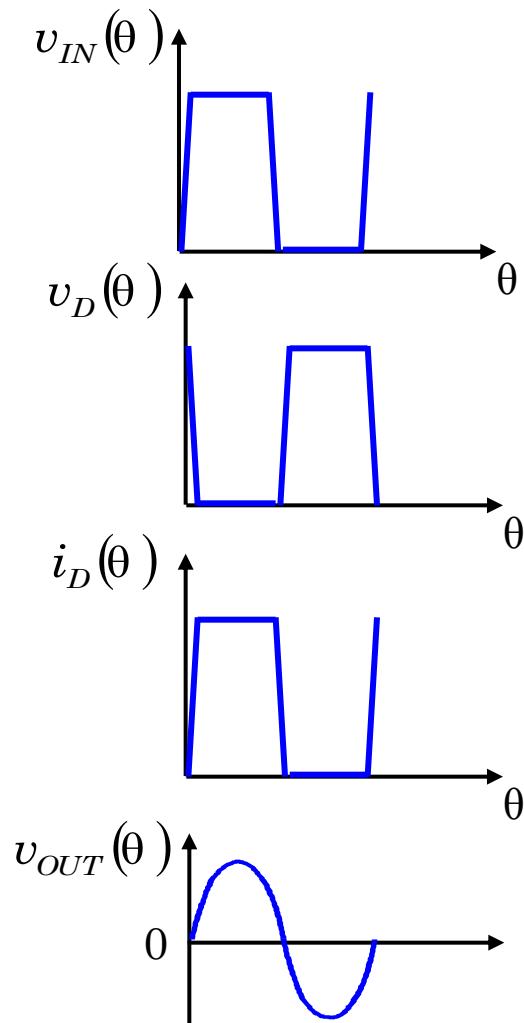


- As the switch turns off, V_x remains low long enough for the current to drop to zero.
- V_x reaches zero just before the switch turns on.
- dV_x/dt is also near zero when the switch turns on.
- After the switch turns off, the load network operates as a damped second-order system.

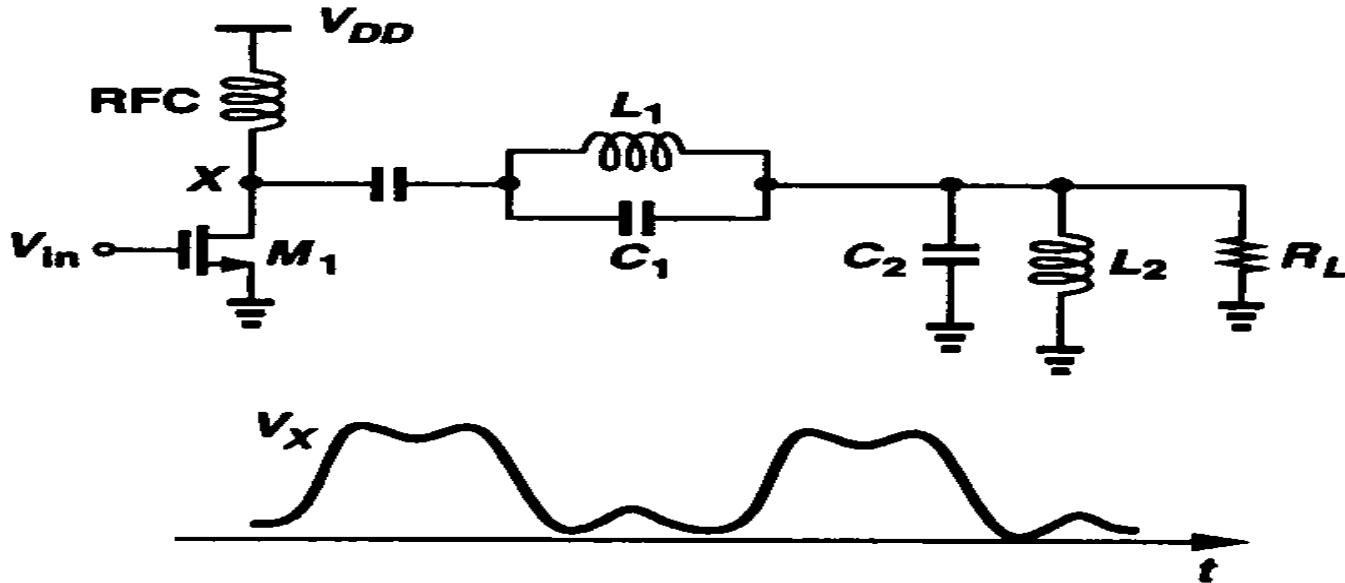
Class E Power Amplifiers



- Switch mode
- Approaching 100% efficiency

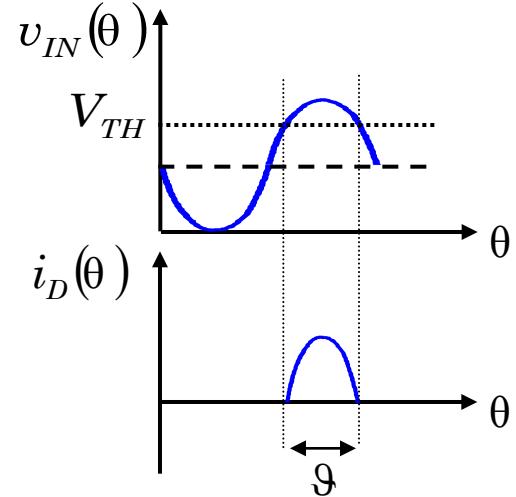
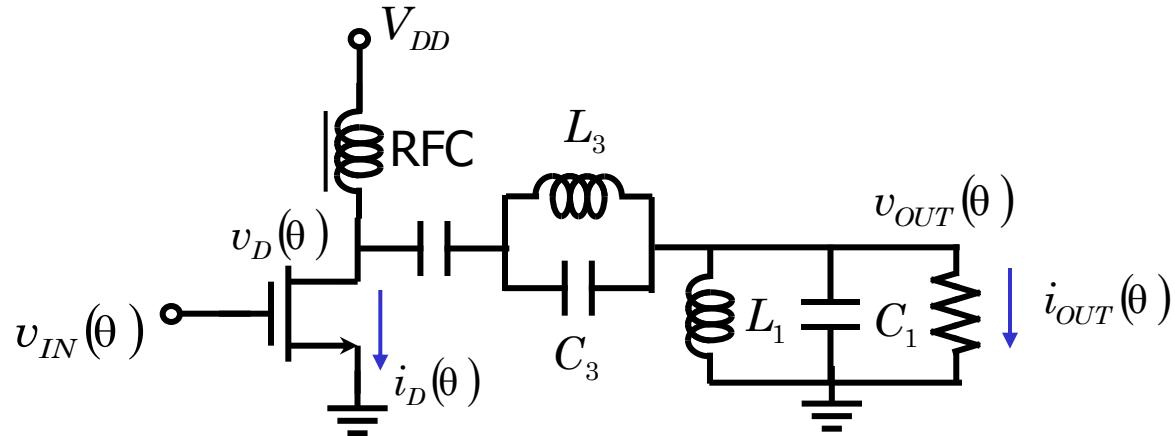


Class F PAs

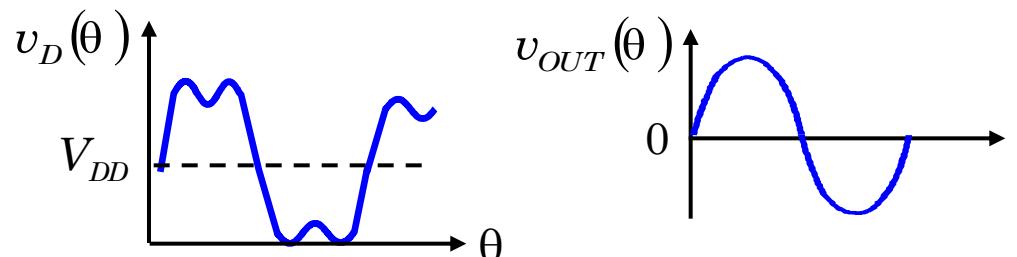


- The idea of harmonics termination for a class A stage can be extended to nonlinear amplifiers as well.
- It can be proved that the peak efficiency of class F amplifiers is equal to 88% for third-harmonics peaking and 85% for second-harmonics peaking.

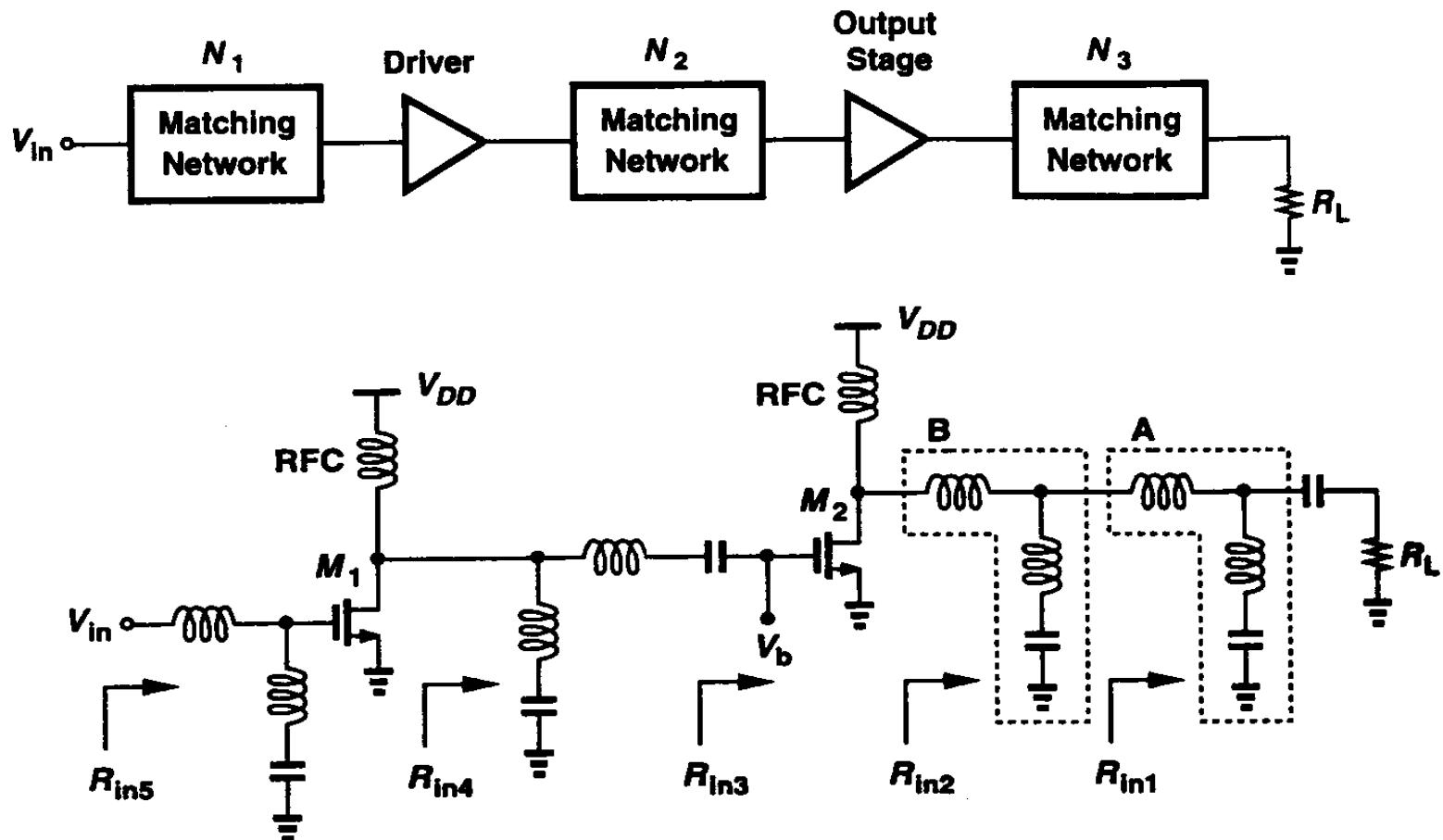
Class F Power Amplifiers



- L_3C_3 tuned to the 2nd or 3rd harmonics
- Peak efficiency
 - 88% for 3rd harmonics peaking
 - 85% for 2nd harmonics peaking.

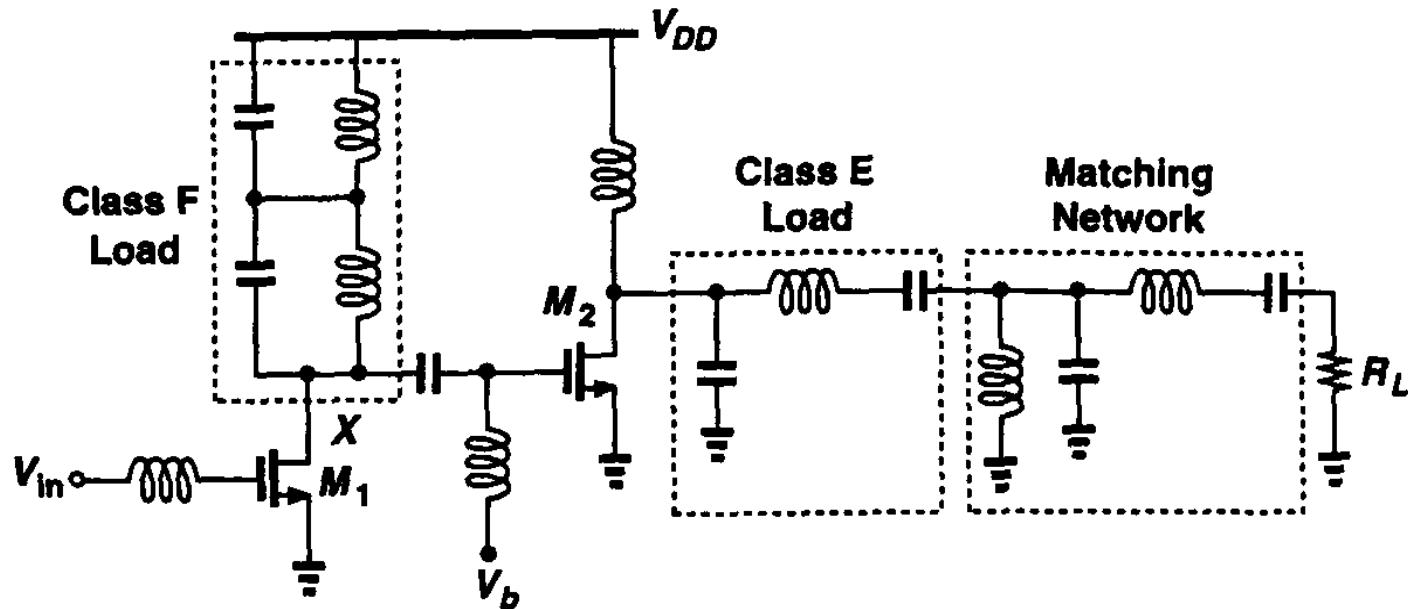


Power amplifier examples



[*] B. Razavi

Power amplifier examples

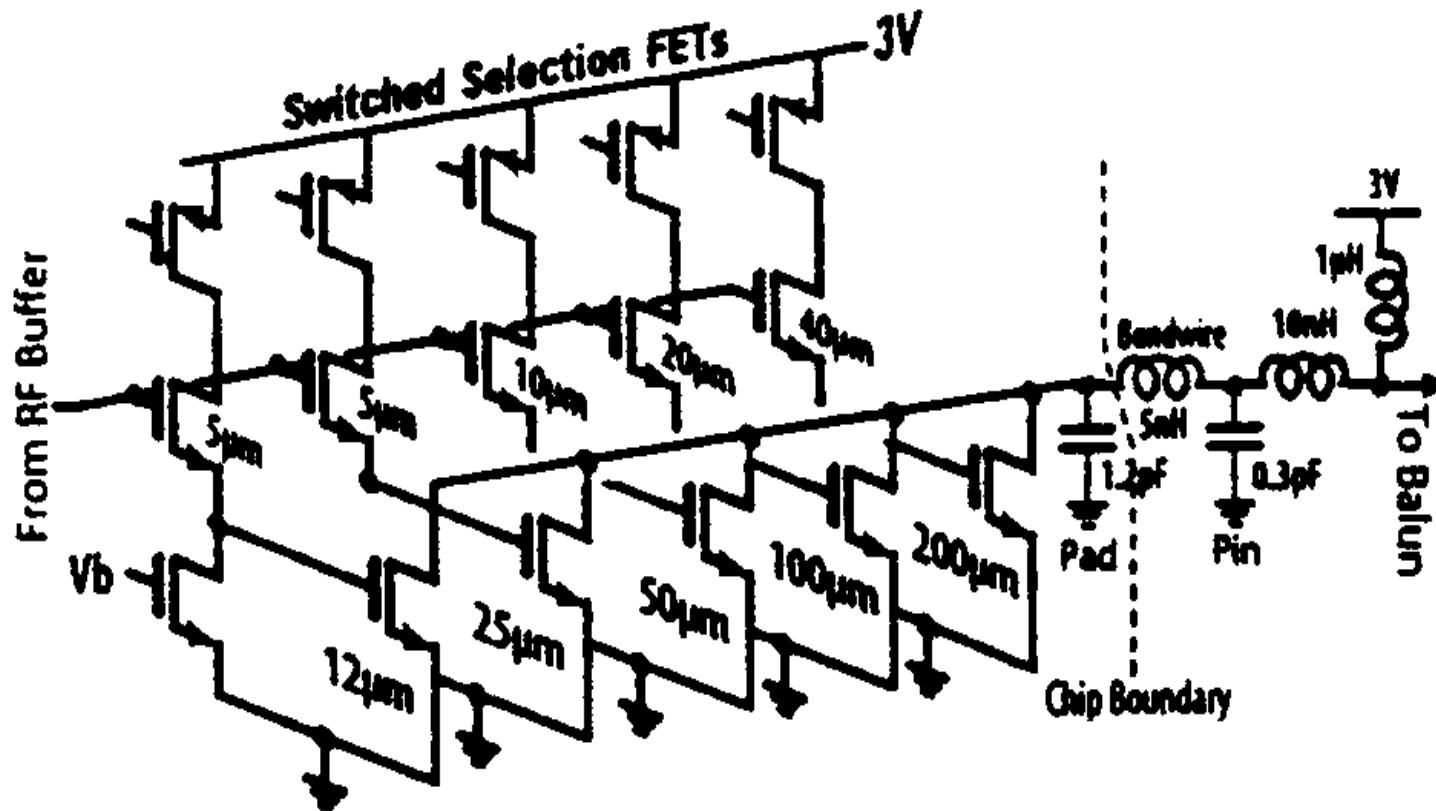


[*] B. Razavi

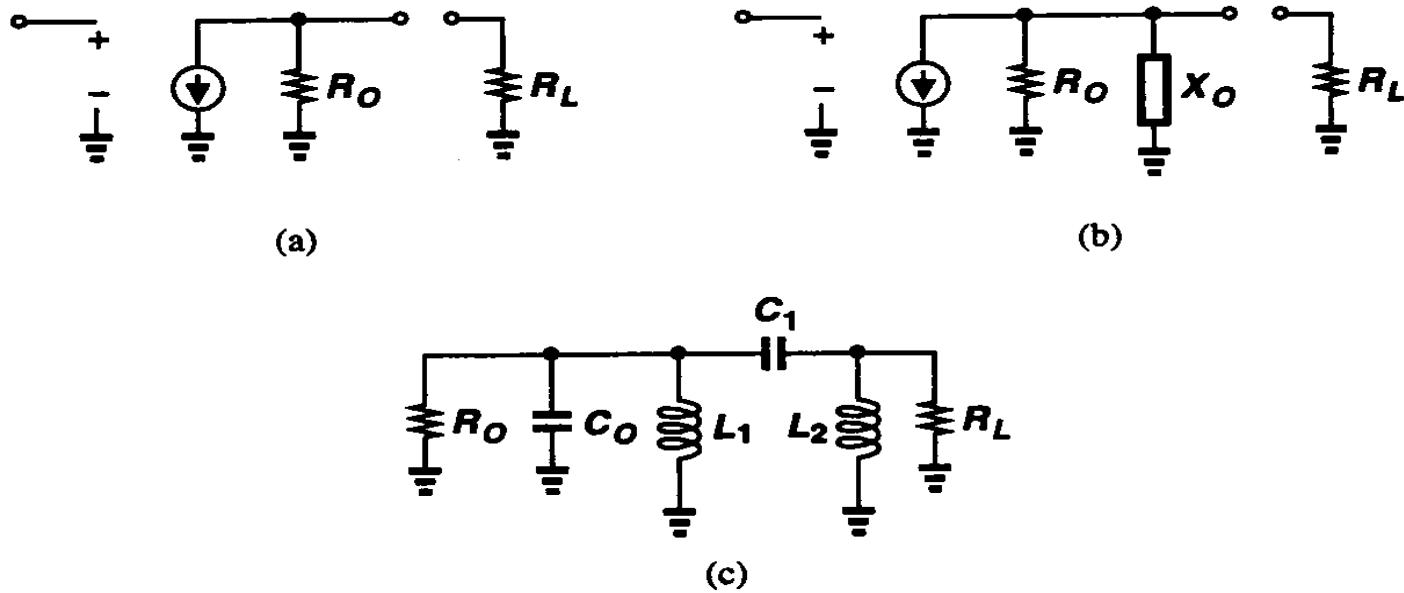
Kuei-Ann Wen National chiao Tung Univ.

30

Power amplifier examples



Nonlinear impedance matching



- Maximum power transfer does not correspond to maximum efficiency.
- The matching can be obtained roughly using small-signal approximation, but modifying these for maximum large-signal efficiency requires a great deal of trial and error.

Large-Signal Impedance Matching

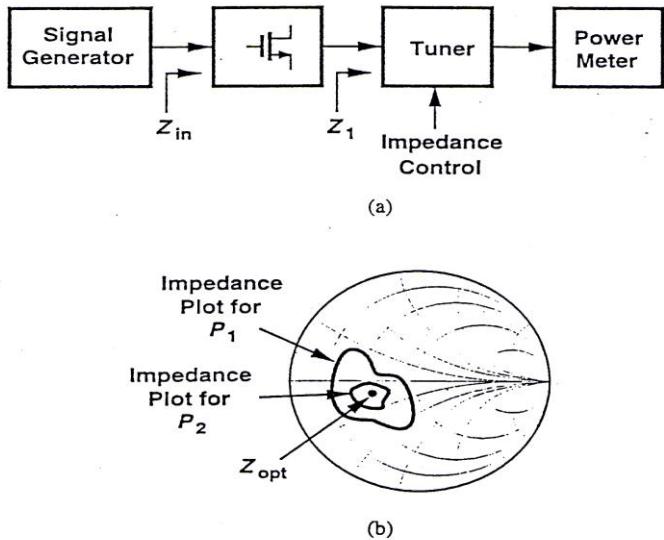
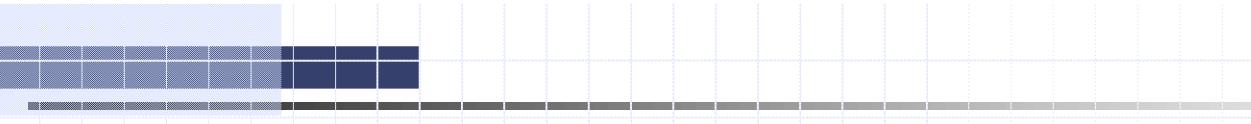


Figure 9.15 (a) Load-pull test, (b) power contours on a Smith chart.

- In a “load-pull” test, the output power is measured and plotted as a function of the complex load seen by the transistor.
- As Z_1 varies so does Z_{in} , a second tuner between the signal generator and the transistor is needed.

Linearization Techniques

- Most linear power Amp.
 - Class A of efficiency around %30 to %40 for portable devices.
- To improve efficiency
 - Linearization after nonlinear PAs.
- Linearization method:
 - feedford
 - feedback
 - envelope elimination and restoration
 - LINC



Liberalization Technology: Feedforward

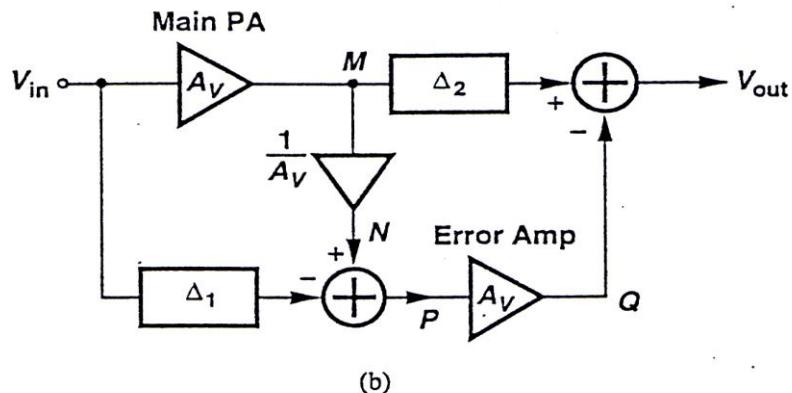
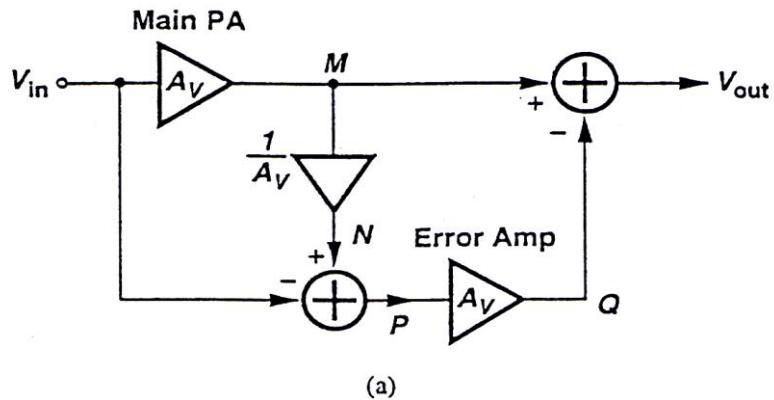


Figure 9.16 (a) Simple feedforward topology, (b) addition of delay elements.

$$V_M = A_V V_{in} + V_D$$

$$V_N = V_{in} + V_D / A_V$$

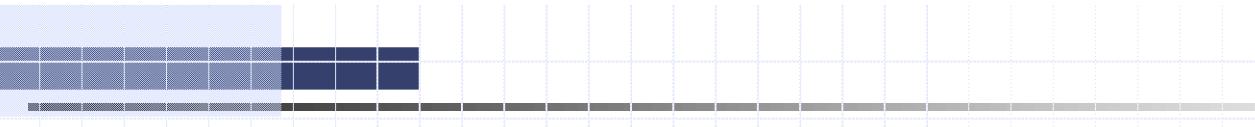
$$V_P = V_D / A_V, V_Q = V_D$$

$$V_{out} = A_V V_{in}$$

$$E = \sqrt{1 - 2(1 + \frac{\Delta A}{A}) \cos \Delta \phi + (1 + \frac{\Delta A}{A})^2} \quad (9.4)$$

The suppression of the magnitude of the IM products in V_{out} : E





Liberalization Technology: Feedback

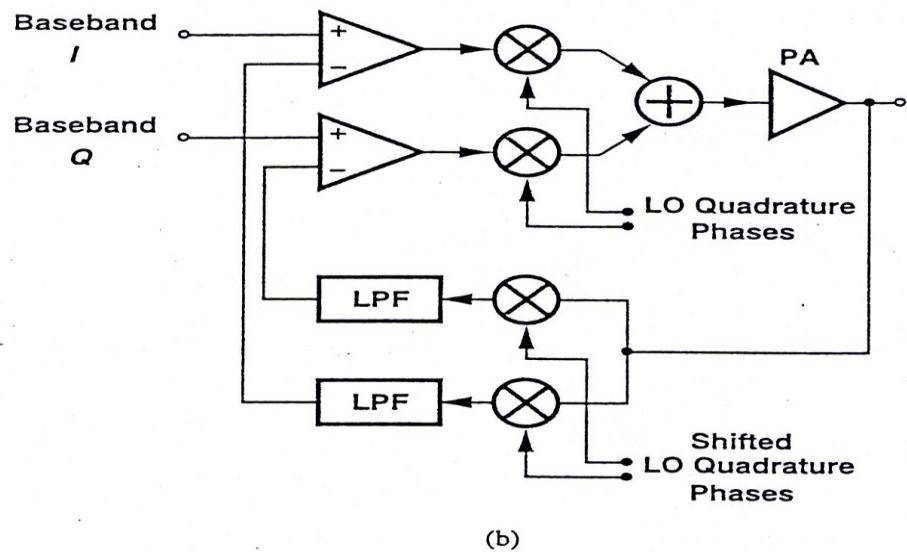
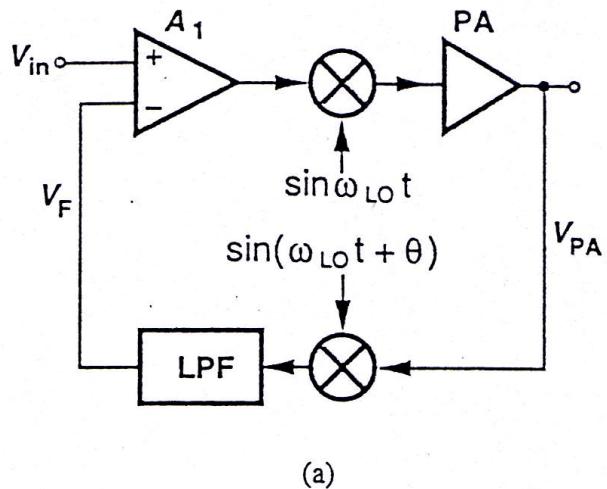
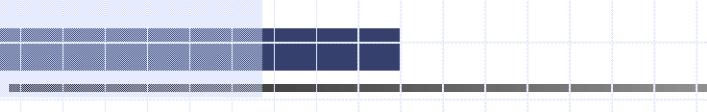


Figure 9.17 (a) Feedback by frequency translation, (b) Cartesian feedback.



Envelope Elimination and Restoration

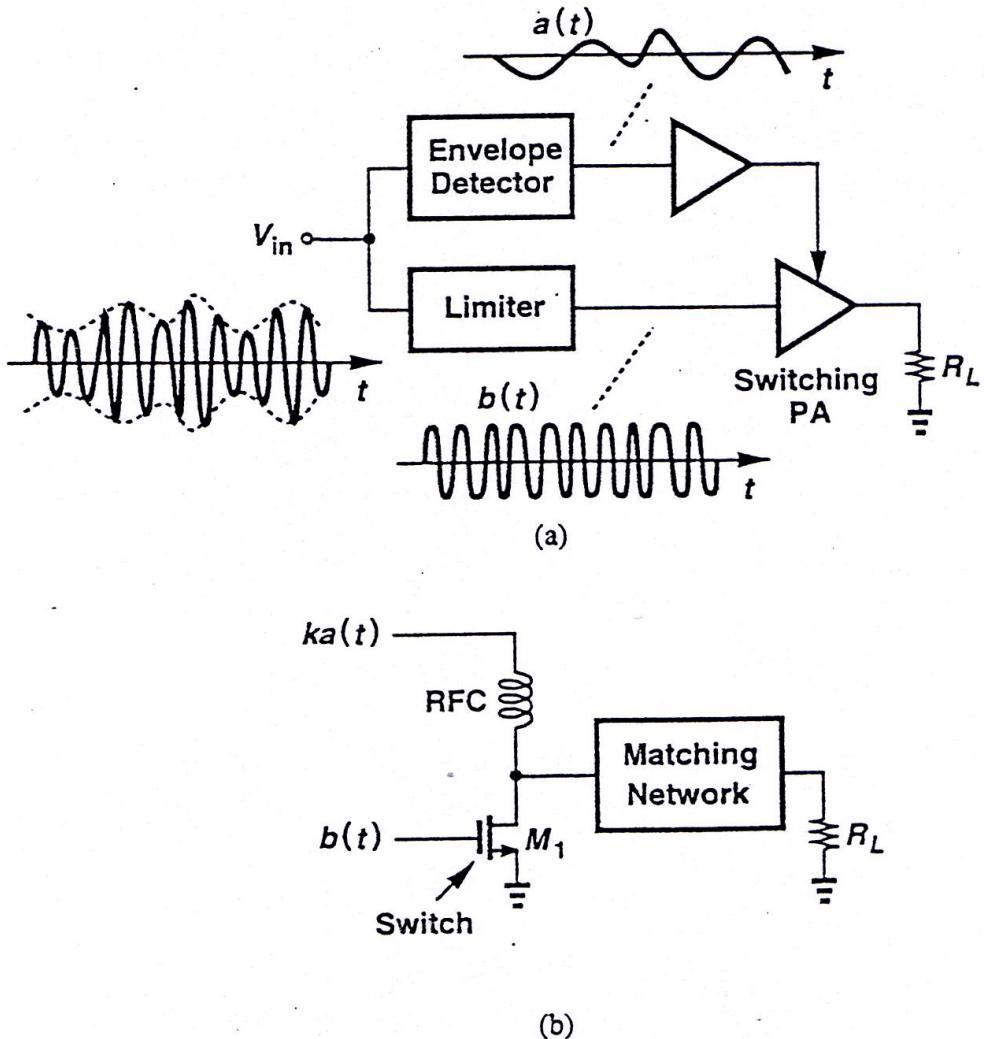
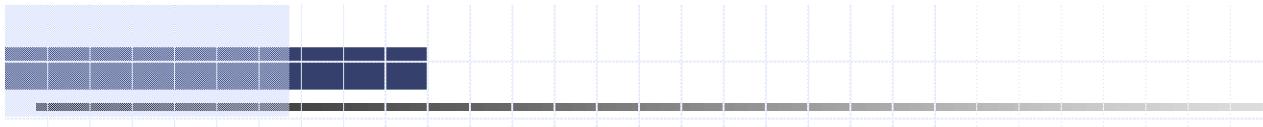


Figure 9.18 (a) Envelope elimination and restoration, (b) implementation of the output stage.



Linearization using non-linear circuits

LINC Technology

(1)

$$v_{in}(t) = a(t) \cos[\omega_c t + \phi(t)], \theta(t) = \sin^{-1}[a(t)/V_0]$$

$$v_1(t) = 0.5V_0 \sin[\omega_c t + \phi(t) + \theta(t)]$$

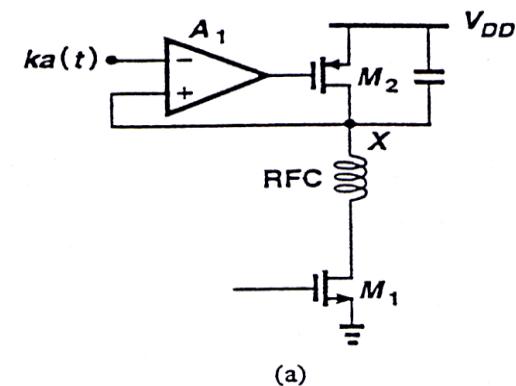
$$v_2(t) = -0.5V_0 \sin[\omega_c t + \phi(t) - \theta(t)]$$

(2)

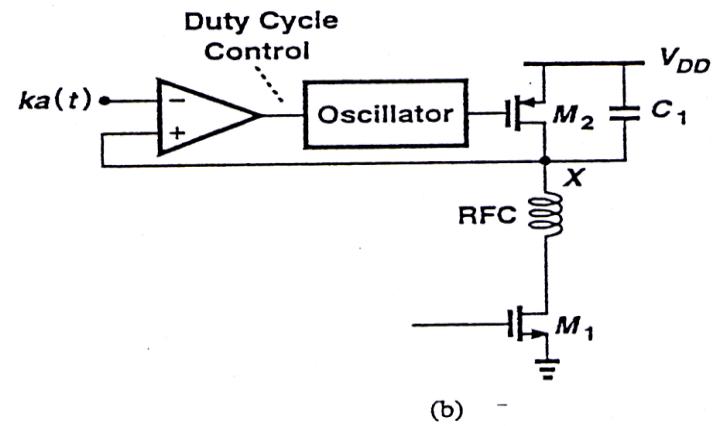
$$v_1(t) = v_I(t) \cos[\omega_c t + \phi(t)] + v_Q(t) \sin[\omega_c t + \phi(t)]$$

$$v_2(t) = -v_I(t) \cos[\omega_c t + \phi(t)] + v_Q(t) \sin[\omega_c t + \phi(t)]$$

$$v_I(t) = a(t)/2, v_Q(t) = \sqrt{V_0^2 - a(t)^2 / 2}$$



(a)



(b)

Figure 9.19 Modulation of the PA output by (a) a low-frequency feedback amplifier, (b) pulse-width modulation.

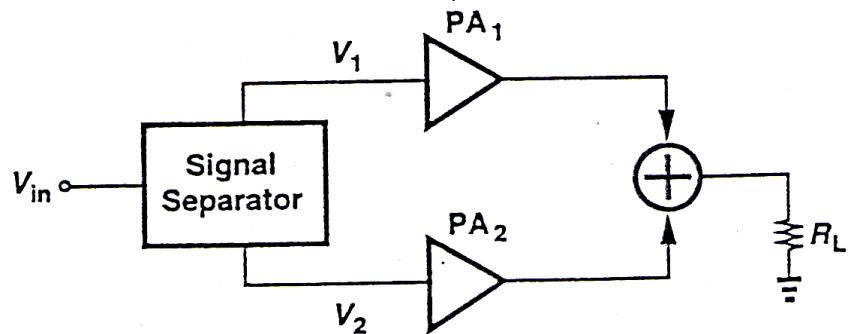


Figure 9.20 Linear amplification using nonlinear stages.

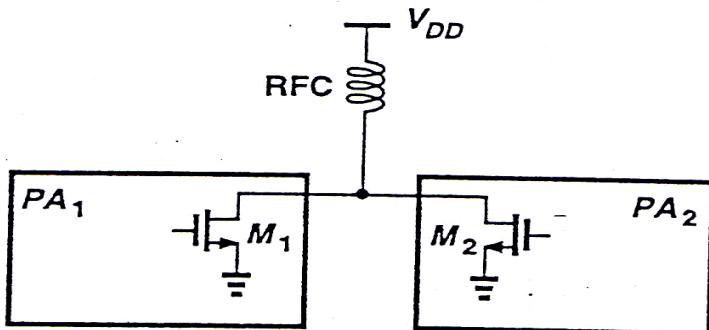


Figure 9.21 Addition of the outputs of two PAs.

$$v_1(t) = v_I(t) \cos(\omega_c t + \phi) + v_Q(t) \sin(\omega_c t + \phi) \quad (9.5)$$

$$v_2(t) = -v_I(t) \cos(\omega_c t + \phi) + v_Q(t) \sin(\omega_c t + \phi) \quad (9.6)$$

Limitations of integrated CMOS Power Amplifier

- Device Breakdown Voltage
 - Low voltage swing
 - Sub- μ CMOS process has low oxide breakdown
- Low current driving capabilities
 - Larger device required for a given current
- Larger Capacitances
 - Tuning is more difficult
- Substrate Coupling with the RF Blocks
 - PA injects more currents into substrate
- Lower Q passive elements



Conclusions

- CMOS Technology for RF is for applications
 - Integrated with significant digital circuits
 - Lowest cost
 - Moderate radio performance
- Accurate RF models are critical for RF CMOS circuit design
- Continuous process improvement enables CMOS RF capability

AMPLIFIER CLASSES (adapted from Wikipedia)

Amplifier circuits are classified as A, B, AB and C for analog design, and class D, E, and F for high efficiency switching design. For the analog classes, each class defines what proportion of the input signal cycle (called the angle of flow) is used to actually switch on the amplifying device:

Class A

100% of the input signal is used (conduction angle $a = 360^\circ$ or 2π)

Class A amplifiers amplify over the whole of the input cycle such that the output signal is an exact scaled-up replica of the input with no clipping. They are not very efficient — a theoretical maximum of 50% is obtainable, but for small signals, this waste of power is still extremely small, and can be easily tolerated. Only when we need to create output powers with appreciable levels of voltage and current does Class A become problematic. In a Class A circuit, the amplifying element is biased such that the device is always conducting to some extent, and is operated over the most linear portion of its characteristic curve (known as its transfer characteristic or transconductance curve). Because the device is always conducting, even if there is no input at all, power is wasted. This is the reason for its inefficiency

Class AB

more than 50% but less than 100% is used. (181° to 359° , $\pi < a < 2\pi$)

- Class AB1 applies to tube or transistor amplifiers in class AB where the grid or base is more negatively biased than it is in class A.
- Class AB2 applies to tube or transistor amplifiers in class AB where the grid or base is often more negatively biased than in AB1, and the input signal is often larger. When the drive is high enough to make the grid or the base more positive, the grid or base current will increase. It is possible depending on the level of the signal input for the amplifier to move from class AB1 to AB2.

Class B

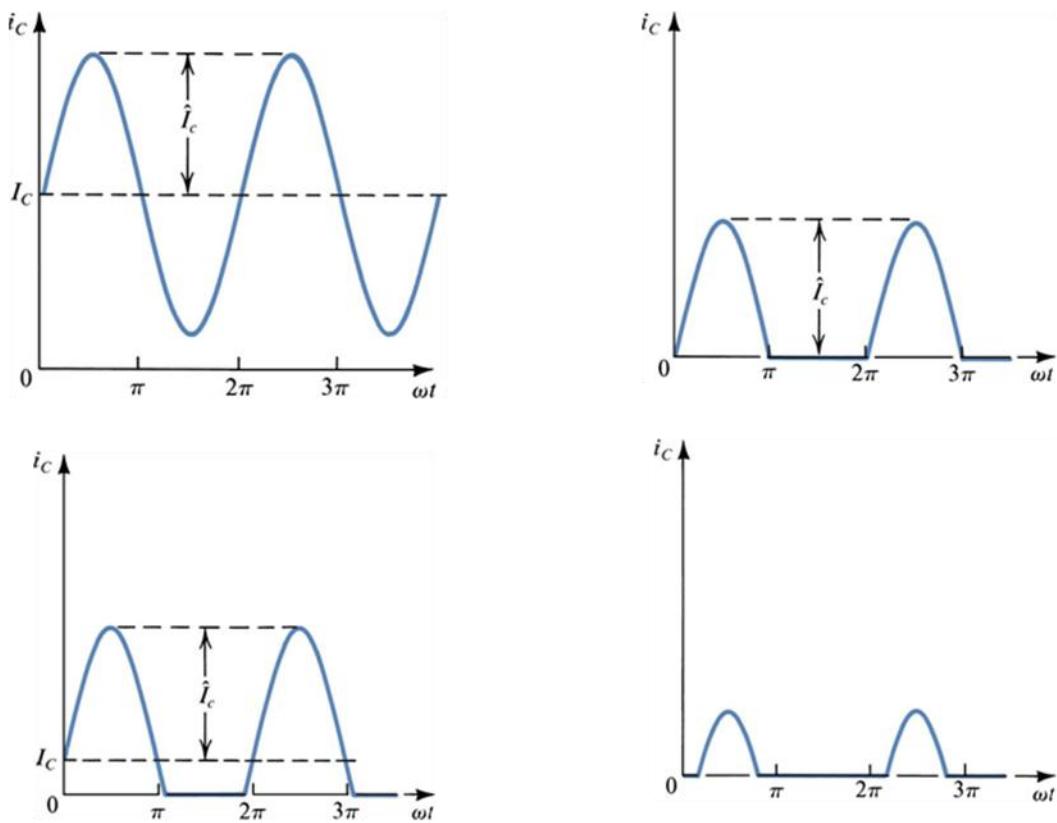
50% of the input signal is used ($a = 180^\circ$ or π)

Class B amplifiers only amplify half of the input wave cycle (maximum theoretical efficiency of 78.5%). This is because the amplifying element is switched off altogether half of the time, and so cannot dissipate power. A single Class B element is rarely found in practice, though it can be used in RF power amplifiers where the distortion is unimportant. However Class C is more commonly used for this.

Class C

less than 50% is used (0° to 179° , $a < \pi$)

Class C amplifiers conduct less than 50% of the input signal and the distortion at the output is high, but efficiencies of up to 90% can be reached. Some applications can tolerate the distortion, such as megaphones. A much more common application for Class C amplifiers is in RF transmitters, where the distortion can be vastly reduced by using tuned loads on the amplifier stage. The input signal is used to roughly switch the amplifying device on and off, which causes pulses of current to flow through a tuned circuit. The tuned circuit will only resonate at particular frequencies, and so the unwanted frequencies are dramatically suppressed, and the wanted full signal (sine wave) will be abstracted by the tuned load. Provided the transmitter is not required to operate over a very wide band of frequencies, this arrangement works extremely well. Other residual harmonics can be removed using a filter.



Collector current waveforms for transistors operating in (a) class A, (b) class B, (c) class AB, and (d) class C amplifier stages.

Class D

A class D amplifier is a power amplifier where all power devices are operated in on/off mode. Output stages such as those used in pulse generators are examples of class D amplifiers. Mostly though, the term applies to devices intended to reproduce signals with a bandwidth well below the switching frequency. These amplifiers use

pulse width modulation, pulse density modulation (sometimes referred to as pulse frequency modulation) or more advanced form of modulation such as Sigma delta. The input signal is converted to a sequence of pulses whose averaged value is directly proportional to the amplitude of the signal at that time. The frequency of the pulses is typically ten or more times the highest frequency of interest in the input signal. The output of such an amplifier contains unwanted spectral components (i.e.. the pulse frequency and its harmonics) that must be removed by a passive filter. The resulting filtered signal is then an amplified replica of the input.

The main advantage of a class D amplifier is power efficiency. Because the output pulses have fixed amplitude, the switching elements (transistors) are switched either on or off, rather than operated in linear mode. This means that very little power is dissipated by the transistors except during the very short interval between the on and off states. The wasted power is low because the instantaneous power dissipated in the transistor is the product of voltage and current, and one or the other is almost always close to zero. The lower losses permit the use of a smaller heat sink while the power supply requirements are lessened too.

Class D amplifiers were widely used to control small DC motors, but they are now also used as audio amplifiers, with some extra circuitry to allow analogue to be converted to a much higher frequency pulse width modulated signal.

Class E

The class E/F amplifier is a highly efficient switching power amplifier, typically used at such high frequencies that the switching time becomes comparable to the duty time. As said in the class-D amplifier the transistor is connected via a serial-LC-circuit to the load, and connected via a large L (inductivity) to the supply voltage. The supply voltage is connected to ground via a large capacitor to prevent any RF-signals leaking into the supply. The class-E amplifier adds a C between the transistor and ground and uses a defined L to connect to the supply voltage.

All previous designs use sharp edges to minimize the overlap. Class E uses a significant amount of second harmonic voltage. The second harmonic can be used to reduce the overlap with edges with finite sharpness. In reality the impedance is mostly reactive and the only reason for it is that class E is a class F amplifier with a very simplified load network and thus has to deal with imperfections. Note how in many amateur simulations of class E amplifiers sharp current edges are assumed nullifying the very motivation for class E and measurements near the transit frequency of the transistors show very symmetric curves, which look much similar to class F simulations.

Class F

The class F-amplifier takes the finite on resistance into account and tries to make the current touch the bottom at zero. This means the voltage and the current at the

transistor are symmetric with respect to time. The Fourier Transform allows an elegant formulation to generate the complicated LC-networks. It says that the first harmonic is passed into the load, all even harmonics are shorted and all higher odd harmonics are open.

In push-pull amplifiers and in CMOS the even harmonics of both transistors just cancel. Experiment tells that a square wave can be generated by those amplifiers and math tells that square wave do consist of odd harmonics only.

In a class D amplifier the output filter blocks all harmonics, which means the harmonics see an open load. So even small harmonic currents suffice to generate a voltage square wave. The current is in phase with the voltage applied to filter, but the voltage across the transistors is out of phase. Therefore there is a minimal overlap between current through the transistors and voltage across the transistors. The sharper the edges the lower the overlap.

While class D sees the transistors and the load as separate modules, the class F admits imperfections like the parasitics of the transistor and tries to optimize the global system to have a high impedance at the harmonics. Of course there has to be a finite voltage across the transistor to push the current across the on state resistance. Because the combined current through both transistors is mostly in the first harmonic it looks like a sine. That means that in the middle of the square the maximum of current has to flow, so it may make sense to have a dip in the square or in other words to allow some over swing of the voltage square wave. A class F load network by definition has to transmit below a cut off frequency and to reflect above. Any frequency lying below the cut off and having its second harmonic above the cut off can be amplified, that is an octave bandwidth. On the other hand a LC series circuit with a large L and a tunable C may be simpler to implement. By reducing the duty cycle below 0.5, the output amplitude can be modulated. The voltage square waveform will degrade, but any overheating is compensated by the lower overall power flowing. Any load mismatch behind the filter can only act on the first harmonic current waveform, clearly only a purely resistive load makes sense, then the lower the resistance the higher the current. Class F can be driven by sine or by a square wave, for a sine the input can be tuned by an L to increase gain. If class F is implemented with a single transistor the filter is complicated to short the even harmonics.

Notes

The main concept used in amplification is to model the active switching device, such as a transistor or MOSFET, as a linear combination of two parts: (1) a (theoretical) "perfect" switching element, and (2) a complex network of parasitic elements attached to it (capacitors, inductors and resistors). After the decomposition, it becomes trivial to eliminate the losses of each part:

(1) The "perfect" switching element should be switched ON during zero-voltage crossing, and should be switched OFF during zero-current crossing. Thus the switching element either conducts current, or has some non-zero voltage on it, but

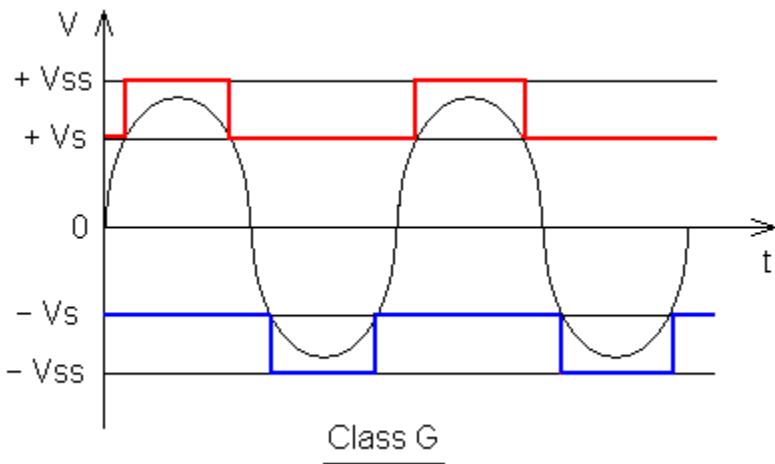
never both at the same time. Because the dissipated power is equal to current x voltage, it becomes zero. This can be arranged by adjusting the phase (capacitor) and DC bias (resistor) of the signal going into the transistor input.

(2) The imaginary part of the impedance of the parasitic elements can be tuned, one by one, by matching them to another passive element with the complex conjugate impedance, thus leaving only the real part of the complex impedance.

In theory, the only remaining loss is the real part of the impedance of the parasitic elements in the system, which cannot be avoided. This class of amplifier is unique to radio frequency ranges, where the amplifier analysis is usually done in the frequency domain and not in the voltage/current domain. This class is further divided to subclasses depending on which harmonics of the signal are taken into account during zero-voltage switching (ZVS) and zero-current switching (ZCS), with names such as Class E/F2,odd; inverse Class (F^{-1}); and so on. It is still an active area of research and new variants appear from time to time, usually with the letters "E" and "F" somewhere in class name.

Class G

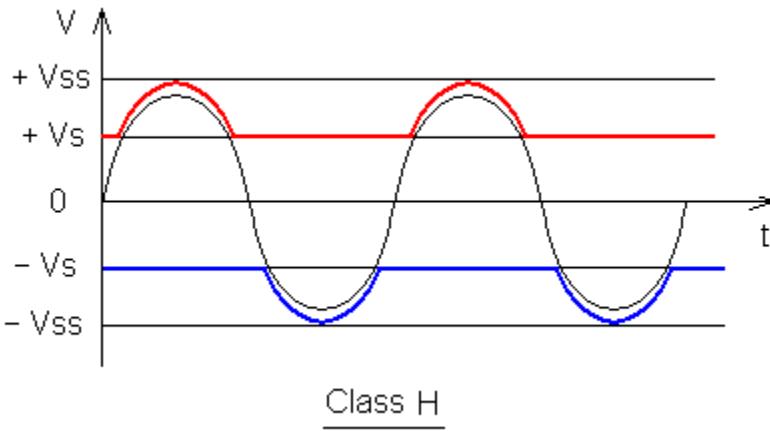
Class G amplifiers are a more efficient version of class AB amplifiers, which use "rail switching" to decrease power consumption and increase efficiency. The amplifier has several power rails at different voltages, and switches between rails as the signal output approaches each. Thus the amp increases efficiency by reducing the "wasted" power at the output transistors.



Class H

A Class H amplifier takes the idea of Class G one step further creating an infinite number of supply rails. This is done by modulating the supply rails so that the rails are only a few volts larger than the output signal at any given time. The output stage operates at its maximum efficiency all the time. Switched mode power supplies can be used to create the tracking rails. Significant efficiency gains can be achieved but

with the drawback of more complicated supply design and reduced THD performance.



Other classes

A hybrid configuration that is receiving new attention is the Doherty amplifier consisting of a class B main stage in parallel with a class C auxiliary (or "peaking") stage. The input signal is split evenly to drive the two amplifiers, and a Doherty combiner is used at the output to recombine the components. During periods of low signal level, the class B amplifier efficiently operates on the signal and the class C amplifier is inactive and consumes no power. During high signal peaks the class B amplifier saturates and the class C amplifier kicks in. The effect is that the Doherty amplifier maintains efficiency comparable to class B saturated performance even when operated at up to 6 dB output backoff. Interest in the Doherty configuration has been revived by cellular telephone and wireless internet applications where the sum of several constant envelope users creates an aggregate AM result. The main challenge of the Doherty amplifier is in aligning the two stages and getting the class C amplifier to turn on and off quickly.

Several audio amplifier manufacturers have started "inventing" new classes as a way to differentiate themselves. These class names usually do not reflect any revolutionary amplification technique, and are used mostly for marketing purposes. For example, Crown's K and I-Tech Series as well as several other models utilize Crown's patented ***Class I*** technology. Lab Gruppen use a form of class D amplifier called ***class TD*** or Tracked Class D which tracks the waveform to more accurately amplify it without the drawbacks of traditional class D amplifiers.

Class T is a trademark of TriPath company, which manufactures audio amplifier IC's. This new class "T" is a revision of the common class D amplifier, but with changes to ensure fidelity over the full audio spectrum, unlike traditional class D designs. It operates at a frequency of 650kHz, with a proprietary modulator.

Class Z is a trademark of Zetex semiconductor and is a direct digital feedback technology.