Class Notes, 31415 RF-Communication Circuits

Chapter IV

NOISE and DISTORTION

Jens Vidkjær

NB 232

Contents

IV Noise and Distortion

Linear amplification for a wide range of signal levels is crucial in many RF applications. Electronic circuits may be designed to show linearity over large operating ranges, but eventually the condition ceases at both high and low drive levels as demonstrated by the typical amplifier characteristic in Fig.1.

When at low levels the output flattens to constancy regardless of input, we have reached the so-called noise floor. Its major causes are thermal motions and quantizised current flows in the circuit components, which add a random fluctuation to the deterministic output signals. When noise and signal become equal in size, we are close to the borderline below which the output is useless. It is a common design challenge to push the noise floor to an acceptable minimum in a given application, and it is one of the major objectives of the present section to provide the background and present techniques for doing so.

In the opposite range of input drives the amplifier performance degenerates due to nonlinearities in the circuit, commonly introduced by the large signal operation of transistors or other electron devices. At the ultimate edge the output may saturate to constancy but in most application the limit for acceptable performance is reached long before that. There are many occasions where we highly benefit from nonlinear operation of electronic circuits but presently we focus on failures that arise when linearity is expected. Transmission of a signal through a nonlinear device is accompanied by formation of deterministic distortion components at frequencies which differs from the input signal frequencies. Filtering may remove some of these new components, known as spurious responses, but others are so close to the original ones that they seriously influence the signal processing that succeeds amplification, detection for instance. We shall consider various forms of nonlinear distortions and their characterization in the last section of this chapter.

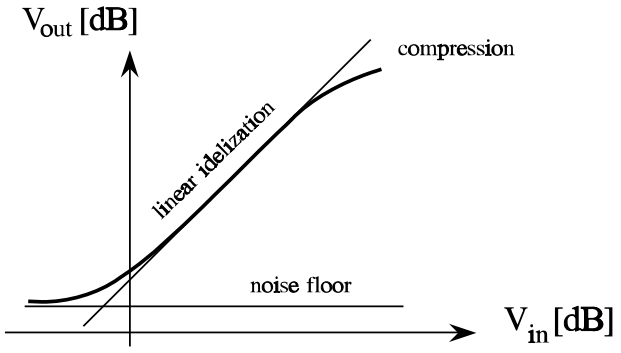


Fig.1 Typical input-output characteristic of an amplifier designed for linear applications.

IV-1 Sources and Basic Properties of Electrical Noise

A voltage or current carrying a signal includes in practice both the desired deterministic signal and disturbing components, which again may be either deterministic or random. The latter encompasses terms that are completely unavoidable on basis of fundamental physical principles, and they are always referred to as noise. Solving practical problem, moreover, other non-ideal contributions may be included under the noise term, particularly if the statistical means and properties that apply to physical noise remain usable. The basic origins and properties of electrical noise are presented in this first section with the aim of providing a foundation for our subsequent discussion of noise in electronic devices and circuits.

Thermal Noise

Thermal motion of the electrons in a resistor causes a fluctuating noise voltage across the terminals. The voltage has Gaussian distribution around a mean value of zero. If the resistor is in thermal equilibrium with its surroundings at temperature T [K], the two-sided spectral density of the mean squared noise voltage is given by,

$$S_v(\omega) = 2R\,\frac{h\omega/2\pi}{e^{h\omega/2\pi kT}-1} \underset{\omega \ll 2\pi kT/h}{\approx} 2kTR \quad - \textit{ where}$$

$$\underline{\textit{Planck's constant}}:\ h = 6.6260755\cdot 10^{-34}\,[Js], \tag{1}$$

$$\underline{\textit{Boltzman's constant}}:\ k = 1.380658\cdot 10^{-23}\,[J/K].$$

The first frequency dependent expression is included to indicate that there is a natural upper frequency roll-off in spectral density, even though it suffices for most practical purposes to use the last - frequency independent - approximation. Around room temperature 300 K we have kT/h=6.44 THz, so here it is the bandwidth of circuit embodying the resistor that impose restrictions on the resultant noise. In a frequency band of Δf Hz around any frequency $\omega_0=2\pi f_0$ fairly below the kT/h bound, thermal noise provides a fluctuation voltage v_n that still has zero mean value and a variance, i.e. mean squared value, given by[1]

$$\overline{v_n^2} = 2\times\frac{1}{2\pi}\int_{\omega_0-\pi\Delta f}^{\omega_0+\pi\Delta f} S_v(\omega)\,d\omega = 2\,S_v(\omega_0)\Delta f = 4kTR\Delta f \quad [V^2]\ . \tag{2}$$

---

1) The overline notation has a long tradition in noise literature for both time averaging like $\langle v_n^2\rangle$ in common notation or statistical averaging - expectation - like $E[v_n^2]$. Most noise processes we consider are ergodic with equal time and ensemble averages and the overline notation is maintained below unless it leads to ambiguity.

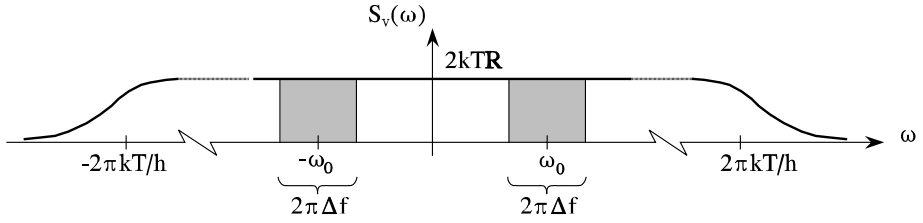


Fig.2 Double-sided power spectral density of thermal noise from a resistor. The mean squared voltage from frequency interval Δf is the hatched area divided by 2π.

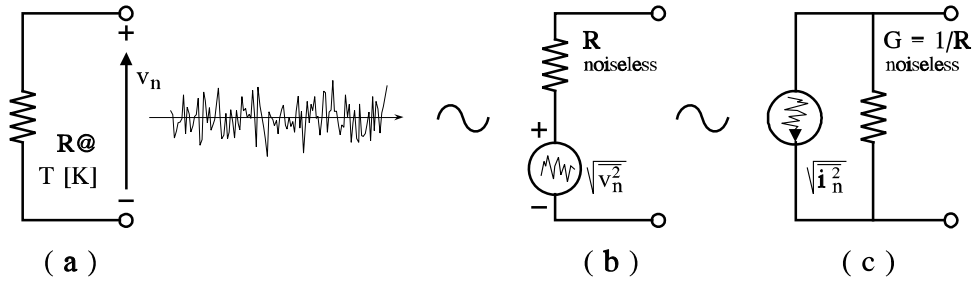


Fig.3 Circuit schematic representations of thermal noise from a resistor. The square roots are often omitted for clarity although the noise generators still represent RMS (root mean square) voltages or currents.

Incorporating thermal noise from a resistor in a circuit diagram is often done as shown by Fig.3(b). The noise is accounted for separately by the RMS voltage generator and the resistor R is assumed noiseless. Translated to the Norton equivalent form in Fig.3(c), the corresponding short-circuit noise current is given by

$$\overline{i_n^2} = \frac{\overline{v_n^2}}{R^2} = 4kTG\Delta f \quad [A^2], \tag{3}$$

where conductance G equals 1/R. It is superfluous in most cases to orient the noise voltage or current generators as it is done in the figures. On the other hand this makes no harm, so the orientations are maintained since we later shall consider situations that requires signs. Regarding equivalent circuits that incorporates noise sources of the types above, there is a tacit assumption that if we are going to measure a noise voltage, ideal filtering must precede the instrument. Fig.4 shows the simplest case, where a resistor is the only circuit component.

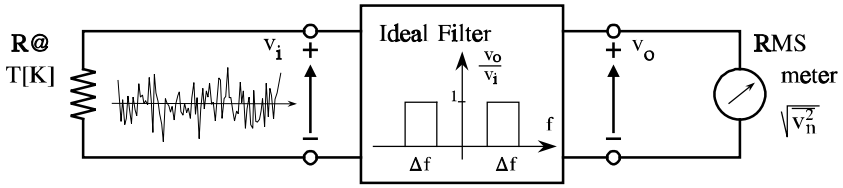


Fig.4 Hypothetical setup for noise voltage measurement corresponding to Fig.3. Substituting with practical filters Δf represents the so-called noise bandwidth.

The available noise power N_{av} from the equivalent circuits in Fig.3(b) and (c) is independent of the resistance value and given by,

$$N_{av} = \frac{\overline{v_n^2}}{4R} = kT\Delta f \quad [W]. \tag{4}$$

Thermal noise is sometimes called Johnson noise after its discoverer. The physical and mathematical descriptions were provided by Nyquist, who also proved that any electrical system in thermal equilibrium has an open circuit thermal noise voltage corresponding to the real part of the impedance, equivalently a short-circuit current corresponding to the real part of the admittance. Both cases agree with the available power expression from Eq.(4).

$$Z(\omega_0) = R(\omega_0) + jX(\omega_0): \quad \overline{v_{nZ}^2} = 4kTR(\omega_0)\Delta f \quad \Rightarrow \quad N_{av} = \frac{\overline{v_{nZ}^2}}{4Re\{Z(\omega_0)\}} = kT\Delta f, \tag{5}$$

$$Y(\omega_0) = G(\omega_0) + jB(\omega_0): \quad \overline{i_{nY}^2} = 4kTG(\omega_0)\Delta f \quad \Rightarrow \quad N_{av} = \frac{\overline{i_{nY}^2}}{4Re\{Y(\omega_0)\}} = kT\Delta f. \tag{6}$$

It is assumed here that the frequency interval Δf is small enough to assume constancy of the impedance or admittance across the interval, alternatively that Δf represents the so-called noise bandwidth, which is exemplified below. Cases that can be described by the method above include noise from magnetization losses in inductors and transformers, noise due to dielectrical losses in transmission lines, noise from the radiation and reception resistance in antennas, microphones or other transducers. For a more genuine discussion of these and many other basic topics, the reader should consult ref's [1] and [2], which are collections of fundamental papers on noise in electrical circuits.

Example IV-1-1 (noise in a RC circuit, noise bandwidth)

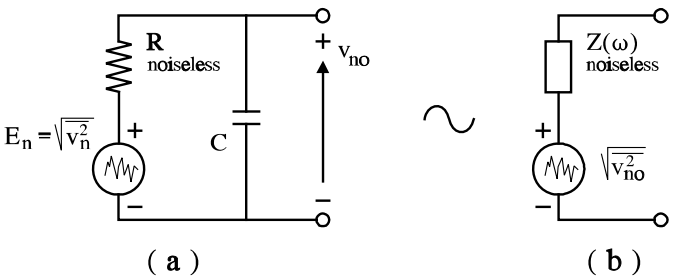


Fig.5 Calculation of RC circuit noise from (a), transfer function $H(j\omega)=v_{no}/E_n$, and (b) directly from Eq.(5).

Consistency of the above results is demonstrated below by considering the RC network in Fig.5(a). It has the output impedance

$$Z_o(\omega) = \boldsymbol{R}(\omega) + \boldsymbol{j}\,X(\omega) = \frac{\boldsymbol{R}/\boldsymbol{j}\omega C}{\boldsymbol{R} + 1/\boldsymbol{j}\omega C} = \frac{\boldsymbol{R}}{1+\omega^2\tau^2} - \frac{\boldsymbol{j}\,\boldsymbol{R}\omega\tau}{1+\omega^2\tau^2}, \quad \textit{where } \tau = \boldsymbol{R}C. \tag{7}$$

The mean-squared output noise voltage per unit bandwidth follows directly from Eq.(5),

$$\left.\frac{\overline{v_{no}^2}}{\Delta f}\right|_{\omega_0} = 2\,S_{vo}(\omega_0) = \frac{4\,k\,T\,\boldsymbol{R}}{1+\omega_0^2\tau^2} \tag{8}$$

A leading factor of two is required in the equation above to let $S_{vo}(\omega)$ be a double-sided spectrum. We shall see that the same result follows from the resistor noise representation in Fig.5(a) where $H(j\omega)$ is the transfer function from voltages E_n to voltage v_{no}. This is a usual voltage division, which in absolutely squared form provides

$$\begin{aligned} H(\boldsymbol{j}\omega) &= \frac{v_{no}}{E_n} = \frac{1/\boldsymbol{j}\omega C}{\boldsymbol{R} + 1/\boldsymbol{j}\omega C} = \frac{1}{1+\boldsymbol{j}\omega\tau} \quad \Rightarrow \quad |H(\boldsymbol{j}\omega)|^2 = \frac{1}{1+\omega^2\tau^2} \\ \left.\frac{\overline{v_{no}^2}}{\Delta f}\right|_{\omega_0} &= 2\,S_{no}(\omega_0) = 2\,|H(\boldsymbol{j}\omega_0)|^2\,S_n(\omega_0) = \frac{4\,k\,T\,\boldsymbol{R}}{1+\omega_0^2\tau^2} \end{aligned} \tag{9}$$

i.e. the same result as the one obtained by Eq.(8). Including all frequencies, it is furthermore seen, that the total mean-squared voltage from a RC circuit is independent of the resistance value,

$$\overline{v_{no,tot}^2} = \frac{1}{2\pi}\int_0^\infty 2\,S_{no}(\omega)\,\boldsymbol{d}\omega \underset{u=\omega\tau}{=} \frac{4kT\boldsymbol{R}}{2\pi\tau}\int_0^\infty \frac{du}{1+u^2} = \frac{kT}{C}, \tag{10}$$

since the last definite integral is $\pi/2$. The noise output voltage is also the voltage across the capacitor, which therefore holds a thermal noise mean energy of size

$$\overline{E_C} = \frac{1}{2}\,C\,\overline{v_{no,tot}^2} = \frac{1}{2}\,kT\,. \tag{11}$$

This result shows agreement with classic statistical thermodynamics. A system in thermal equilibrium with its surroundings at an absolute temperature T has an energy of ½kT per degree of freedom, i.e. number of state variables required to describe the system. The RC circuit needs only one - the capacitor voltage.

The so-called noise bandwidth is illustrated by Fig.6 for the RC circuit. It is the bandwidth that provides the same available noise power from a uniform distribution at peak level as the power actually available from the entire, frequency dependent distribution. The noise-bandwidth W_N spans the interval from $-W_N$ to W_N, so the hatched rectangle has an area equal to the area beneath the actual spectral density curve. Using the result from Eq.(10), the noise-bandwidth in radians per second becomes

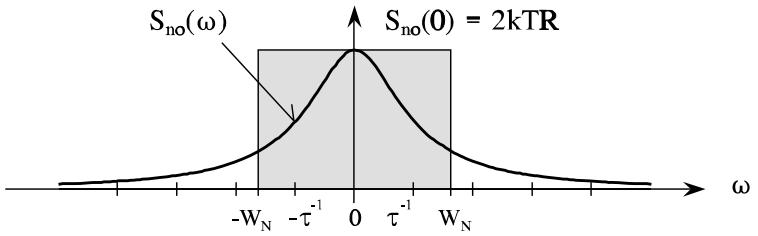


Fig.6 Noise bandwidth W_N in a RC-circuit. The hatched area corresponds to the total area beneath the spectral density curve, $S_{no}(\omega)$.

$$2\,W_N\,S_{no}(0) = 2\int_0^\infty S_{no}(\omega)\,d\omega \quad \Rightarrow \quad W_N = \frac{2\pi kT/C}{4kTR} = \frac{\pi}{2}\,\frac{1}{\tau}\,. \tag{12}$$

In an RC circuit the noise bandwidth is seen to be ½π = 1.56 times greater than the usual 3dB (half power) bandwidth, which is the inverse of the time constant τ = RC.

Example IV-1-1 end

Shot Noise

If the current through a device is the net effect of many single events that occur randomly, but with a mean rate $\overline{N_I}$ events per second, the double-sided power spectrum of current fluctuations around the mean current is given by Carson's theorem, [4] sec.2.2.c,[2]

$$S_I(\omega) = \overline{N_I}\;E\left[\;|p_I(\omega)|^2\;\right], \tag{13}$$

Here, $p_I(\omega)$ represents Fourier transforms of pulse shapes that contributes to the current. Taking expectation of the absolutely squared transform conducts averaging over the ensemble in cases that includes different pulse shapes. As a simple illustration we consider one electron, which is injected into the depletion region in a pn-junction, where it moves across the region at mean velocity v_e[m/s]. The width of the region is d, so T_d=d/v_e [s] is the mean time of traversing the region. While in move the electron causes a current pulse of height q/T_d [A] through the bulk-regions and circuitry that connects the two sides of the depletion region. Being rectangular each current pulse has the sinx/x shaped Fourier transform

$$p_I(t) = \begin{cases} \dfrac{q}{T_d}\;, & -\dfrac{T_d}{2} \le t \le \dfrac{T_d}{2} \\ 0\;, & otherwise \end{cases} \qquad \Rightarrow \qquad p_I(\omega) = q\;\frac{\sin(\omega T_d/2)}{\omega T_d/2} \tag{14}$$

$$\underline{electron\ charge}: \quad q = 1.60207\cdot 10^{-19}\ [C]$$

---

2) Note that to find the spectrum of a random binary signal in section I-3, a similar expression was used in slightly weaker form due to the constant bit-rate.

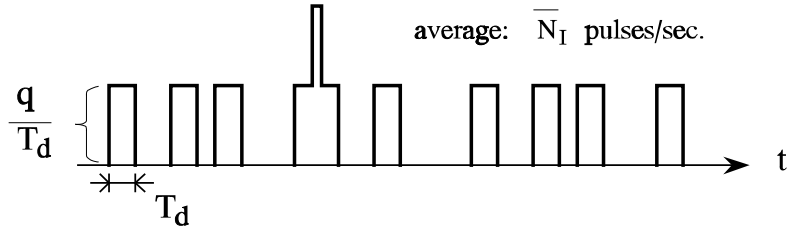


Fig.7 Train of pulses, each representing the transition of one electron across the depletion region in a pn-junction.

A train of electron pulses sums up to the DC current $I=q\overline{N_I}$. Substituting I into Eq(13), the doublesided spectrum for current fluctuations becomes

$$S_I(\omega) = qI\left[\frac{\sin(\omega T_d/2)}{\omega T_d/2}\right]^2 \underset{f \ll 1/T_d}{\approx} qI \,. \tag{15}$$

Thus, the total mean squared noise current from the fluctuations in a frequency band Δf around $\omega_0 = 2\pi f_0$ is given by

$$\overline{i_I^2} = 2 \times \frac{1}{2\pi} \int_{\omega_0 - \pi\Delta f}^{\omega_0 + \pi\Delta f} S_I(\omega)\, d\omega \underset{f_0 \ll 1/T_d}{\approx} 2qI\Delta f \,. \tag{16}$$

The last, approximated result is called Schottky's theorem. It represents the limit where $T_d \to 0$, so each current contribution is an impulse - a shot - of weight q and a flat spectrum in concordance with the $1/T_d$ frequency bound for the approximation. The frequency independent simplification suffices for most practical application of devices that exhibits shot noise, for instance Schottky barrier and pn junction diodes or bipolar transistors where the time carriers are in drift across depletion regions commonly are considerably shorter than the transit time through neutral bulk regions [3].

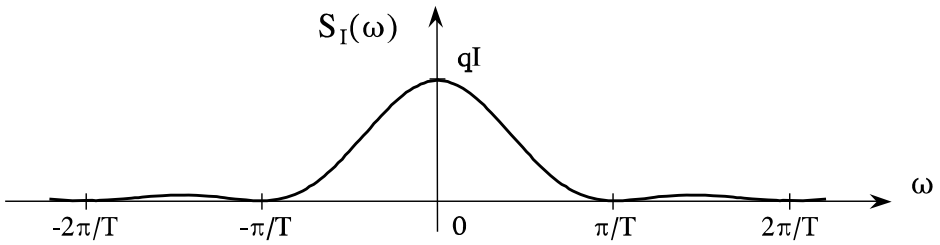


Fig.8 Spectrum of a mean squared shot noise current. In a pn-junction T is the transit time of carriers across the depletion region.

Flicker or 1/f Noise

While the types of noise considered above were ascribed to specific and well understood noise mechanism, the term flicker or 1/f noise is distinguished solely from the shape of the spectral density,

$$S_f(\omega) = \frac{C}{|\omega|^{\alpha}}, \qquad \textit{where} \quad \alpha \approx 1 \tag{17}$$

The numerator C depends commonly upon the device biasing, but it is independent of frequency. This type of noise is observed in most electronic devices, passive as well as active, although the extent may vary by orders of magnitude between different types. Among transistors, MOSFET's are in the high flicker noise end and BJT's are in the low end. Considering resistors, bulk carbon and carbon film types are the noisiest while metal film or wirewound resistors have low flicker noise.

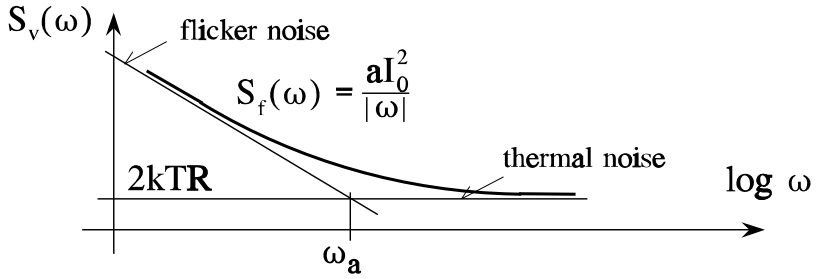


Fig.9 Spectrum including flicker noise for the mean squared noise voltage across a carbon resistor. I_0 is the DC current through the resistor.

To exemplify flicker noise we start considering the noise voltage across a carbon resistor. In addition to the flat thermal noise spectrum we get an 1/f shaped contribution as sketched in Fig.9. Observations show that the level of flicker noise is proportional to the square of the DC current through the resistor, an influence that may be understood if we associate the 1/f noise with random resistance fluctuations δR around a mean value R_0,

$$R = R_0 + \delta R \tag{18}$$

Applying DC current I_0, the voltage across the resistor and especially the mean squared noise voltage are given by

$$v = v_0 + \delta v = R_0 I_0 + \delta R\, I_0 \quad \Rightarrow \quad \overline{\delta v^2} = \overline{\delta R^2}\, I_0^2 . \tag{19}$$

The spectrum of the mean squared voltage fluctuations originates here from a 1/f spectrum in the mean squared resistance fluctuations. Note that the DC current only is required to observe the noise but not for controlling the underlying mechanisms of resistance fluctuations.

Although flicker noise basically is a low frequency phenomenon, low frequency parameter fluctuations might get substantial influence on RF circuit performances. Applying a sinusoidal current $I_c\cos\omega_c t$ in the carbon resistor case gives the voltage components

$$v \;=\; v_1 + \delta v \;=\; R_0\, I_c \cos\omega_c t + \delta R\, I_c \cos\omega_c t\,. \tag{20}$$

The last term holds the noise, but it is no longer a stationary random variable but a cyclo-stationary one of period $T_c=2\pi/\omega_c$. The resultant power spectrum for the mean squared voltage is given by

$$S_v(\omega) \;=\; \frac{R_0^2 I_c^2}{4}\,\delta(\omega_c) + \frac{R_0^2 I_c^2}{4}\,\delta(-\omega_c) + \frac{I_c^2}{4}\,S_{Rf}(\omega+\omega_c) + \frac{I_c^2}{4}\,S_{Rf}(\omega-\omega_c)\,. \tag{21}$$

The two first terms account for the sinusoidal carrier in v_1. If $S_{Rf}(\omega)$ denotes the mean square 1/f shaped spectrum of the stationary random variable δR, the noise portion of the resultant spectrum is shifted in frequency to fringe the carrier. Suppose we have two signals represented by carriers I_{c1} and the nearby but much smaller I_{c2}, which is the one we subsequently want to separate by filtering and detect. From the above relationship and the sketch in Fig.10, we see that the flicker noise induced by the large carrier may overwhelm the signal to be found and the succeeding filtering becomes difficult. So despite its basic low-frequency outset, flicker noise may be critical also at RF frequencies.

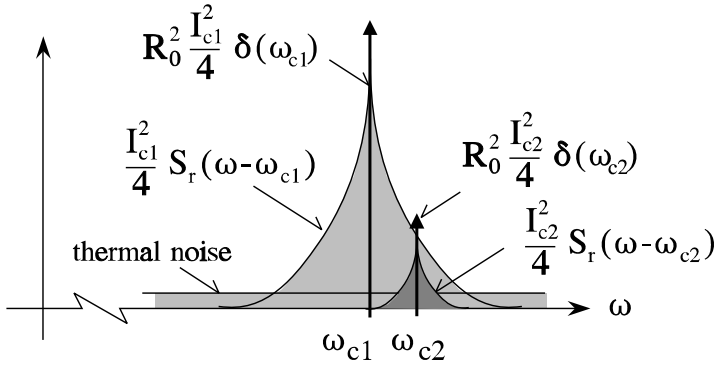


Fig.10 Noise and signal components when two RF signals are inflicted by parameter flicker noise like resistance fluctuations in carbon resistors. Only the positive frequency parts from Eq.(21) are shown.

Flicker noise contribution to the mean squared noise voltage in a frequency interval from f_l to f_u by the 1/f relationship gives

$$\overline{v_f^2} \;=\; 2\times\frac{C}{2\pi}\int_{2\pi f_l}^{2\pi f_u}\frac{d\omega}{\omega} \;=\; 2\,C\,\log f_u - 2\,C\,\log f_l\,, \tag{22}$$

Both limits, $f_u \to \infty$ or $f_l \to 0$, make unrealistic advancements towards infinity. A closer look on details in the spectrum must therefore depart from 1/f in the limits. In the high end the spectrum must roll off faster than 1/f, for example like $1/f^2$ above a certain frequency f_l. Correspondingly the spectrum must grow slower than 1/f when f goes towards zero. A constant spectrum below a certain frequency f_l would be appropriate. For most practical noise calculations we have commonly no detailed knowledge about these matters. If the bandwidth of interest spans the whole frequency range of flicker noise, a safe choice on the upper bound is to use f_u above the limit frequency f_a where thermal noise becomes dominant as indicated by Fig.9. The low bound f_l could be chosen so low that we are no longer willing to spent time observing consequences of the noise, alternatively $1/f_l$ could be taken as the length of the longest message that is processed separately in a communication system.

There are many physical origins of flicker noise in electronic components and devices. The cause in resistors might be fluctuations in the mobility of the carriers that convey the current. In semiconductors, flicker noise also accompanies the generation and recombination process of free carriers. Often phenomena that deteriorate normal device performance, for instance uncontrolled oxide traps and surface states in transistors, contribute dominantly to flicker noise. In new device types this noise may be excessive, but commonly it levels off when the corresponding technology matures. Flicker noise in semiconductor devices manifests itself both directly as noise currents or voltages, but it also contributes to parameter fluctuations. The transconductance, for instance, might get an 1/f spectrum, so the modulating property that was illuminated for bulk resistors by Eq.(21) and Fig.10 applies to transistors as well. Consult refs. [1] and [4] chap.8 to gain more insight into the physics behind 1/f noise. A discussion of common basic aspects of the flicker noise may be found in ref.[5].

Other Noise Sources in Electronics

The three noise types we have considered are far the most prominent in RF circuits. They are also typical from the point of view, that we have covered three basic forms, the unavoidable flat spectrum thermal noise, the bias current dependent shot noise with approximately flat spectrum, and finally flicker noise whose dominant low-frequency spectrum transforms to RF and smear out signal spectra. There are other types and sources of noise. Most of them can hardly be distinguished from the types above in experimental data. For the same reason, noise models in circuit simulator programs are concentrated around our basic types. If other sources are present in the circuit, they are taken into account by parameter adjustments. However, some specific noise contributions are so frequently referred to in literature that they need a few comments.

Generation-recombination noise accompanies the random processes of generating and recombining carriers in electronic devices. If one mechanism dominates the process, for instance a well defined trap, the corresponding noise gets a spectrum of shape,

$$S_{gr}(\omega) = K \frac{\tau}{1 + \tau^2 \omega^2} \tag{23}$$

Here τ is the lifetime of carriers. Factor K depends typically on bias currents. It is seen that the high frequency asymptote decays with the square of frequency, so it is much steeper than in flicker noise. However, many different processes of widely separated lifetimes are usually engaged in generation and recombination of carriers in semiconductor devices. Their joint effect approximates the 1/f spectral density of flicker noise [4].

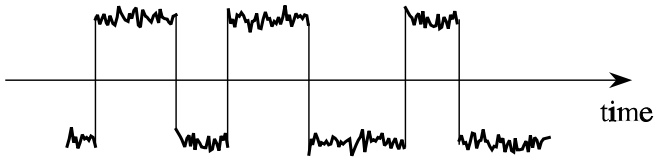


Fig.11 Typical time-domain burst noise waveshape

Burst noise is a special form of generation-recombination noise with spectral density like Eq.(23). This noise is sorted out due to its characteristic wave shape like the one in Fig.11, but the basic mechanisms of the pronounced jumps are not completely understood [4]. Burst noise is sometime called popcorn noise after the sound the waveshape produces in a loudspeaker. Like flicker and other noise types with low-frequency spectral dominance, burst noise gets significance in RF if it controls device parameter variations.

Avalanche noise follows the avalanche effect in reverse biased pn-junctions. If the field is sufficiently high, free carriers may gain the energy that is necessary to produce new electron hole pairs by impact ionization. The total current is still made by a sequence of random events, so the noise is a shot noise of the type in Eq.(16). However, the current in the noise expression under reverse bias is not the saturation current I_S, which we might extract from the characteristic of the forward biased junction, but instead the actual and commonly much greater experimental reverse current that includes avalanching.

Characterizing and Combining Noise Sources

The power spectral density functions are the direct way of specifying a noise source, either $S_v(\omega)$ [V^2/Hz] for a voltage source or $S_i(\omega)$ [A^2/Hz] for a current source. Data sheets give often the square root of twice the densities that are shown, for instance the functions

$$V_n(f) = \sqrt{2\,S_v(2\pi f)}\ \left[\frac{V}{\sqrt{Hz}}\right] \quad or \quad I_n(f) = \sqrt{2\,S_i(2\pi f)}\ \left[\frac{A}{\sqrt{Hz}}\right] \tag{24}$$

Assuming flat spectrum, we get the RMS voltage or current from the noise source by multiplying the $V_n(f)$ or the $I_n(f)$ function by the square root of the frequency interval or the noise bandwidth in the measurements.

Instead of the direct specification above, noise sources of any kind are sometimes compared to the idealized flat spectrum thermal noise from resistors, which was defined by

either Eq.(2) or Eq.(3) for voltage and current respectively. There are two approaches here. The first one requires that the noise source has an output impedance or admittance with positive real part. An effective noise temperature T_{eff} for the one-port is now defined as the temperature that an impedance equal to the one-port impedance must have to give thermal noise equal to the observed noise. If a one-port has impedance $Z_1(\omega_0)$ and mean squared noise voltage or current equal to either $\overline{v^2}_1$ or $\overline{i^2}_1$ in the frequency interval Δf around ω_0, the noise temperature becomes

$$T_{eff} \equiv \frac{\overline{v_{o1}^2}}{4\ k\ Re\{Z_1(\omega_0)\}\Delta f} \qquad or \qquad T_{eff} \equiv \frac{\overline{i_{s1}^2}}{4\ k\ Re\{Y_1(\omega_0)\}\Delta f}\ . \tag{25}$$

Another way of characterizing a noise source is to specify the resistance R_n or the conductance G_n that provide thermal noise equal to the observed source. This is particularly useful when the source is given solely in the form of either a RMS voltage or a RMS current generator at a given temperature. R_n and G_n are called the noise resistance and the noise conductance respectively. If the observed noise in frequency interval Δf is either $\overline{v^2}_1$ or $\overline{i^2}_1$, the corresponding noise resistance or conductance are

$$R_n \equiv \frac{\overline{v_{o1}^2}}{4\ k\ T\ \Delta f} \qquad or \qquad G_n \equiv \frac{\overline{i_{s1}^2}}{4\ k\ T\ \Delta f}\ . \tag{26}$$

It should be emphasized that R_n and G_n are noise level parameters, which are not required to be recognizable as resistors or conductances in an equivalent circuit, although this will be the case, if the noise is thermal and the temperature has been agreed upon. The temperature is often the reference temperature of 290 [K], which is a part of the noise figure definition for two-ports to be discussed below in section IV-3 on page 25.

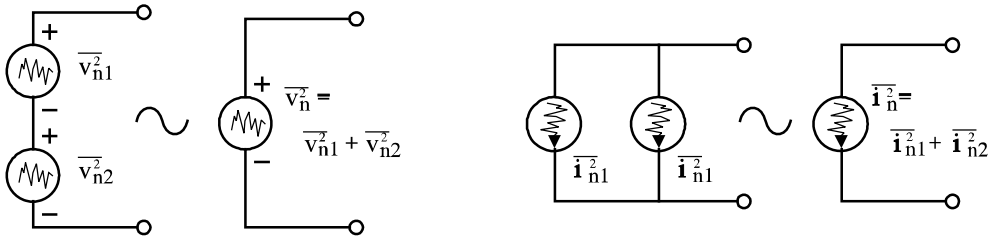


Fig.12 Combining uncorrelated noise sources. The mean square of the resultant voltage or current is the sum of the two original mean squared quantities.

The combined effect of two series connected noise voltage sources - two paralleled noise current sources behave similarly - must be calculated on a mean-squared basis to yield

$$\overline{v_n^2} = \overline{|\ v_{n1} + v_{n2}\ |^2} = \overline{v_{n1}^2} + \overline{v_{n2}^2} + 2\ \overline{v_{n1}v_{n2}} = \overline{v_{n1}^2} + \overline{v_{n2}^2} + 2\ \rho\ \sqrt{\overline{v_{n1}^2}\ \overline{v_{n2}^2}}. \tag{27}$$

Here ρ is a correlation coefficient between the two sources. With time averaging it is defined

$$\rho \equiv \frac{< v_1 v_2 >}{\sqrt{\overline{v_{n1}^2}\,\overline{v_{n2}^2}}} \tag{28}$$

This coefficient is confined to the interval [-1 ≤ρ≤1].

Uncorrelated noise sources have a zero-valued correlation coefficient, so the mean square of the result becomes the sum of mean squares from the two original sources as indicated by Fig.12. Imagining that the two original sources are thermal noise from resistors R_1 and R_2 at the same temperature, the result above makes sense since their series connection is the sum of the resistors and the mean squared noise is proportional to the resistance value. If the sources are characterized by noise resistances, the combined effect in Eq.(27) corresponds to the sum of the noise resistances. Finally, if the noise sources refer to the same real part impedance, the combined result also corresponds to adding noise temperatures.

While noise contributions from independent origins are uncorrelated, correlated noise generators often result when noise from a common physical origin is sensed at different places in a network. We shall see below that the equivalent input and output noise sources in device models may be correlated. In network calculation it is often convenient to use cross spectra, which might be complex even with real valued noise signals. To make circuit computations in frequency domain a corresponding complex correlation coefficient is

$$c \equiv \frac{S_{21}(\omega_0)}{\sqrt{S_1(\omega_0)\,S_2(\omega_0)}} = \frac{\overline{v_1 v_2^*}}{\sqrt{\overline{v_1^2}\,\overline{v_2^2}}} \tag{29}$$

The last form is often seen in literature. Like deterministic signal whose components may be represented by complex phasors in frequency response calculations, noise contributions from a narrow frequency band around a given frequency may be represented by random phasors for transmission computations. In analogy with deterministic phasors we should use ρ = Re{c} when translating back to the time domain version in Eq.(27).

Man-Made Noise

Man-made noise is often encountered in the reception of radio signals, which may be equivalenced as sketched in Fig.13. The antenna has impedance Z_a which resembles a bandpass characteristic around a center frequency. Besides radio signals, which are indicated by the v_s voltage source, noise contributions of various nature are received. They are included in the circuit by the mean square voltage $\overline{v_n^2}$. The level of noise depends very much upon the application that guided the antenna design.

In mobile communication antennas are often omnidirectional and a considerable amount of noise from all directions may be picked up. Fig.14 - from ref.[6] through

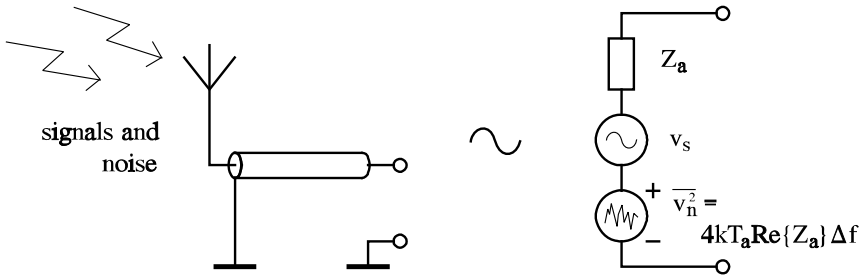


Fig.13 Equivalent circuit for radio signal reception

ref.[7] - shows the nature and relative significance of the received noise quantities as function of frequency. The left scale gives the noise temperature of the antenna impedance. As seen, both types of urban area man-made noise - from city center and suburbs - dominate the entire frequency range of practical interest for mobile communications (approx. 10 Mhz to 2 GHz). Man-made noise includes terms like ignition noise from cars or radiation due to switching in industrial controllers and power regulators. At much lower levels we have the natural sources antenna noise, i.e. noise from galaxies, sun activity, and atmospheric motions. Finally, the curve termed "typical receiver" shows how much the electronic circuitry of a conventional receiver adds to the noise temperature of the antenna. It should be clear from this figure that in mobile communications, there is no need to spend a lot of efforts in squeezing receiver noise towards ultimately low limits when man-made noise from the environment is the most significant part of input noise.

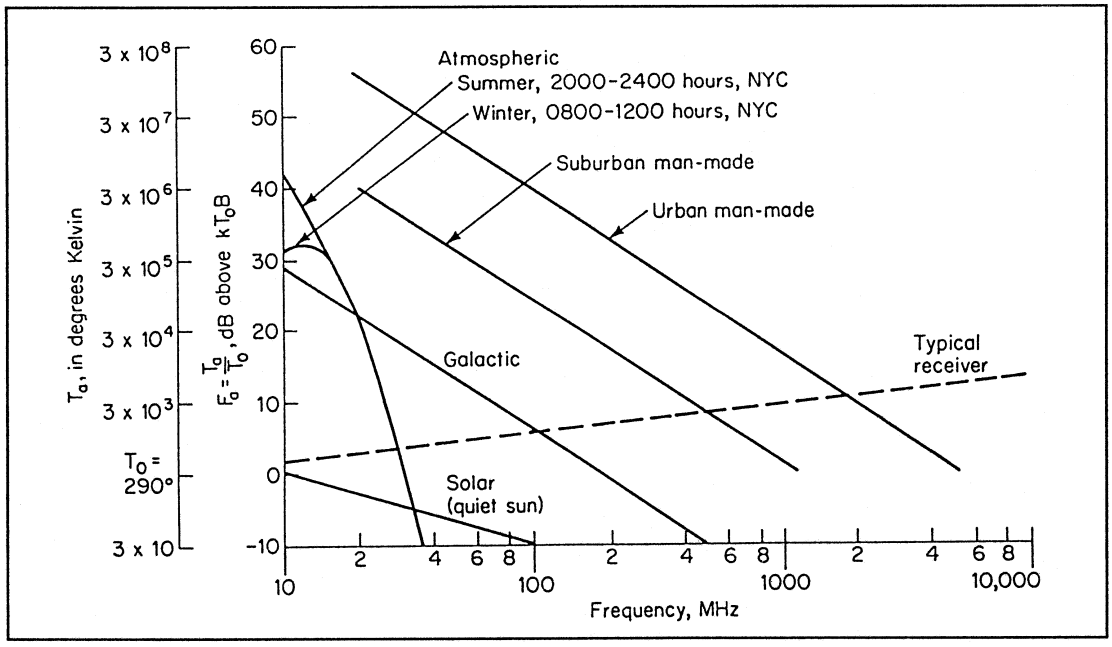


Fig.14 Typical noise components at the antenna terminal in mobile communications (omnidirectional reception). From ref.[6].

IV-2 Noise in Semiconductor Devices

Two ingredients are required to keep control with the noise level in electronic designs. The first one is noise models for the components and devices, the second is systematic approaches for characterizing and constructing circuits that include noise. We address the models in this section, which gives a survey of the most important noise contributions in common RF devices. Beyond this limited scope, the reader should consult the literature on devices, noise, and modeling, for instance [4],[8],[9],[10], or [11].

Noise in Diodes

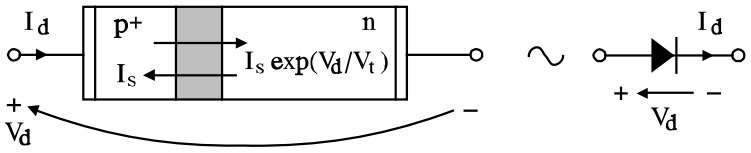


Fig.15 Short p+n-junction. Hole injections across the depletion regions are the only contributions that are included in the idealized model.

A short p^+n diode is considered. The p region is doped more heavily than the n region, so the current is dominated by holes. Assuming ideal conditions, the current through the diode is given by

$$I_d = I_s \left(e^{V_d/V_t} - 1 \right) = I_s e^{V_d/V_t} - I_s, \qquad V_t = kT/q. \tag{30}$$

The first bias dependent term represent flow of holes injected from the p+ region. The second opposing term is a current of holes injected from the n region. At zero bias, i.e. V_d=0, the two terms balance. Under forward bias the first term dominates and gives the exponential voltage to current relationship. Under reverse bias the resultant current approaches the constant $-I_s$, which for small signal silicon diodes typically is in the femto ampere range. Injection from the two sides of the depletion region are statistically independent and implies both full shot noise. The mean squared noise current is therefore

$$\overline{i_{dn}^2} = 2qI_s e^{V_d/V_t}\Delta f + 2qI_s\Delta f \tag{31}$$

A large signal model for an ideal diode is shown in Fig.16(a). Linearized small-signal descriptions includes the differential conductance of the diode,[3]

$$g_d = \frac{dI_d}{dV_d} = \frac{I_s}{V_t} e^{V_d/V_t} \tag{32}$$

---

3) At high frequencies, diode capacitances should be paralleled. They are left out here for clarity since they introduce no new noise aspects.

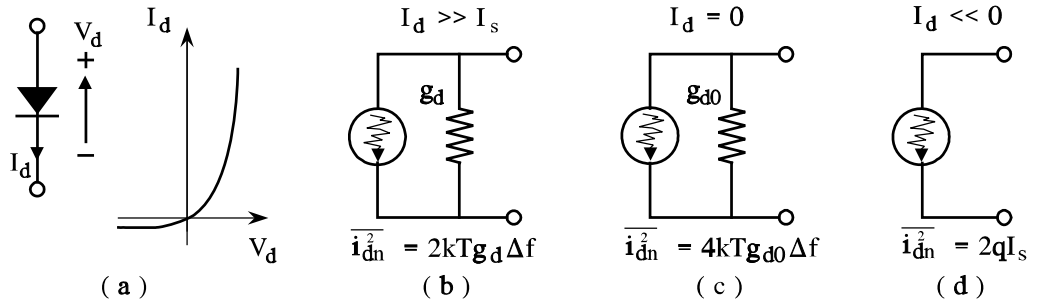


Fig.16 Small-signal noise equivalent circuits for diode in large forward (b), zero (c), and reverse bias (d). $I_d(V_d)$ from Eq.(35) may replace I_s in reverse breakdown.

Using the common assumption that the exponential term dominates the current under forward bias, the conductance and the noise expressions above approximate to,

$$\underline{V_d > 0:}$$
$$I_d(V_d) \approx I_s e^{V_d/V_t} \quad \Rightarrow \quad g_d \approx \frac{I_d(V_d)}{V_t} \quad \Rightarrow \quad \overline{i_{dn}^2} \approx 2qI_d(V_d)\Delta f = 2kTg_d\Delta f \,. \tag{33}$$

To get the last expression, parameter V_t was rewritten kT/q. Comparing with thermal noise from Eq.(3), a conducting diode shows only half the noise that a conductance of similar size provides in thermal equilibrium. In terminology of noise temperatures, the diode conductance has as noise temperature that is half the physical temperature. The difference between the conductor and the diode regarding noise is that a forward biased diode is not in thermal equilibrium. To illustrate this point further it is seen that at zero bias, where diode is in thermal equilibrium, we avoid approximations and get directly from Eq.(31) a result that is in agreement with the thermal noise equation,

$$\underline{V_d = 0:} \qquad g_d = g_{d0} = \frac{I_s}{V_t}\,, \qquad \overline{i_{dn}^2} = 4qI_s\Delta f = 4kTg_{d0}\Delta f \,. \tag{34}$$

Under moderate reverse bias the ideal pn-diode model tends to have no significant conductance, and the noise corresponds to the shot noise associated with the saturation current I_s as sketched in Fig.16(d). At greater reverse biases this description may be too simple for practical purposes, as the reverse current raises significantly beyond I_s. Although it still may be small, the excessive reverse current includes start of Zener and avalanche breakdowns, phenomena that are accompanied by shot noise processes themselves. To complete a noise calculation, an expression including all reverse currents must be provided, for instance using the so-called Miller or avalanche multiplier, M_{av}. It is often stated empirically, [12]. Keeping orientations from Fig.16(a), reverse DC and noise characteristics are now written

$$\underline{V_d < 0:}$$
$$I_d(V_d) = I_s\left[\,1 + M_{av}(V_d)\,\right], \qquad M_{av}(V_d) = \left[\,1 - \left(\frac{-V_d}{V_{av}}\right)\right]^{-N_{av}}, \qquad \overline{i_n^2} = 2qI_d(V_d)\Delta f. \tag{35}$$

The multiplier rises towards infinity when $-V_d$ is approaching the breakdown voltage V_{av}, which is a positive parameter. The exponent N_{av} is another positive parameter that falls in the range from approximately 2 to 7.

Noise in Bipolar Transistors

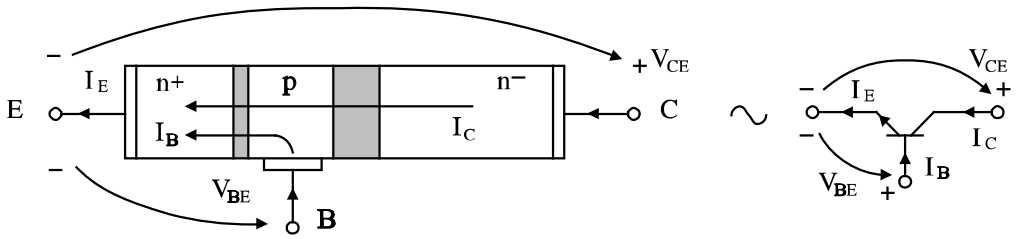


Fig.17 Structure of npn bipolar junction transistor (BJT). The dominant part of the bias currents I_C and I_B are carried by electrons.

Under normal, forward biased operation, the most significant current in a npn transistor consists of electrons that are emitted into the base which they traverse before they eventually are collected after drift through the collector-basis junction. This is the collector terminal current I_C in the sketch of Fig.17. Its noise component is dominated completely by shot noise. The base current I_B is more composite, but in most cases, it is dominated by a flow of holes injected from the base into the emitter, i.e. the mechanism that reduces the so-called emitter efficiency that controls the current gain β. It is a process completely equivalent to emitting electrons to I_C, so that part of the base current shows shot noise. In addition, the base current may contain contributions from carrier recombinations in the base, in the depletion regions, and possibly at surface traps, so the base current may contain flicker and burst noise components in the low frequency part of the spectrum. Noise components, which are associated with I_C and I_B, are included in a transistor equivalent circuit like Fig.18 by the mean squared currents $\overline{i^2_{nc}}$ and $\overline{i^2_{nb}}$. Without flicker and burst noise, the noise sources in this diagram are given by

$$\overline{i_{nc}^2} = 2\,q\,I_C\,\Delta f\,, \qquad \overline{i_{nb}^2} = 2\,q\,I_B\,\Delta f\,, \qquad \overline{v_{nbb}^2} = 4\,k\,T\,R_{BB}\,\Delta f\,, \tag{36}$$

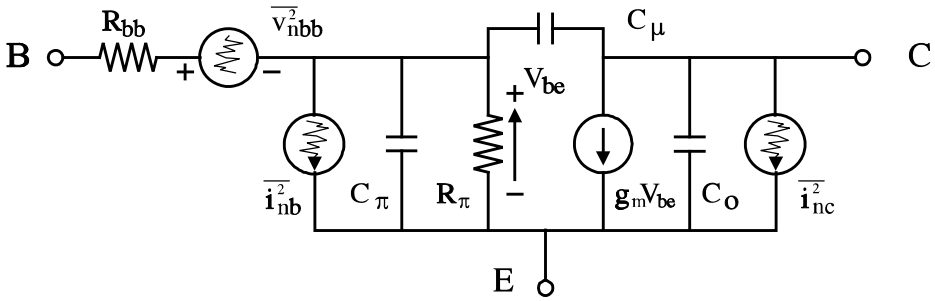


Fig.18 Dominant noise sources in a BJT under forward bias. $\overline{v^2_{nbb}}$ is thermal while $\overline{i^2_{nb}}$ and $\overline{i^2_{nc}}$ are shot noise contributions.

where the noise voltage $\overline{v^2}_{nbb}$ is thermal noise from the base series resistance R_{bb}. This noise may be the dominant noise contribution, so a major objective in RF transistor design is to reduce the base resistance, both for the sake of reducing noise, but also to keep the bandwidth large.

If flicker and burst noise are important in an application, a standard way of inclusion is the following extension of $\overline{i^2}_{nb}$, cf. ref.[13],p.271,

$$\overline{i_{nb}^2} = 2\,q\,I_B\,\Delta f + K_f\,\frac{I_B^{a_f}}{f}\,\Delta f + K_b\,\frac{I_B}{1+(f/f_c)^2}\,\Delta f\,, \tag{37}$$

where K_f and a_f are parameters for flicker noise and K_b, f_c describes burst noise separately, cf. Eq.(23). Note, this model assumes an observation bandwidth Δf, which is small compared to the operating frequency f. Then it is unnecessary to integrate the frequency dependent denominators over the Δf interval as it was done in Eq.(22).

Example IV-2-1 (minimum noise in bipolar transistors)

The relative significance of the different noise sources in a transistor coupling may be studied by their contributions to the output mean square short circuit current. In this example, we consider noise in the mid-frequency range of the transistor, which means that we are sufficiently low in frequency to disregard all the capacitors in the equivalent circuit, Fig.19, but high enough in frequency to ignore possible flicker and burst noise terms. Besides the internal transistor noise generators, the noise from the generator resistance, R_g, is included by $\overline{v^2}_{ng}$. The total short circuit output noise current gets a term from each source,

$$\overline{i_{no}^2}\Big|_{total} = \overline{i_{no}^2}\Big|_{v_{ng}} + \overline{i_{no}^2}\Big|_{v_{nbb}} + \overline{i_{no}^2}\Big|_{i_{nb}} + \overline{i_{no}^2}\Big|_{i_{nc}}\,, \tag{38}$$

As the noise sources are uncorrelated, the different terms are found from the diagram applying the corresponding source one by one

$$\begin{aligned}
\overline{i_{no}^2}\Big|_{v_{ng}} &= \overline{v_{ng}^2}\,g_m^2\left[\frac{R_\pi}{R_\pi + R_{bb} + R_g}\right]^2, \qquad \overline{i_{no}^2}\Big|_{v_{nbb}} = \overline{v_{nbb}^2}\,g_m^2\left[\frac{R_\pi}{R_\pi + R_{bb} + R_g}\right]^2, \\
\overline{i_{no}^2}\Big|_{i_{nb}} &= \overline{i_{nb}^2}\,g_m^2\left[\frac{R_\pi\,(R_{bb}+R_g)}{R_\pi + R_{bb} + R_g}\right]^2, \qquad \overline{i_{no}^2}\Big|_{i_{nc}} = \overline{i_{nc}^2}\,.
\end{aligned} \tag{39}$$

The ratio of the total mean square short circuit output current over the component that originates from the generator resistor alone is a common measure of noise performance, the so-called noise figure F. We shall discuss its interpretation intensively in next section. For now we just calculate to get

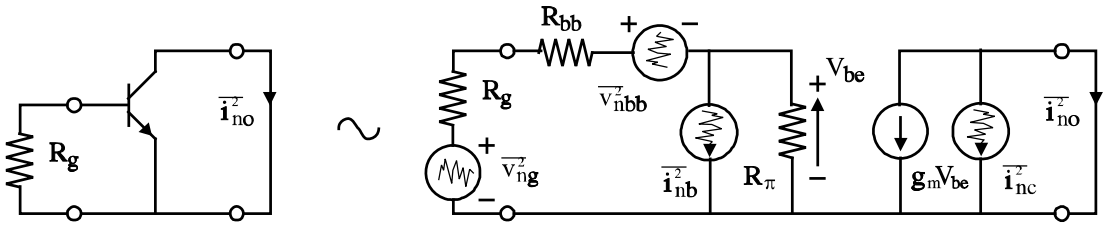


Fig.19 Equivalent circuit for calculating mid-frequency output mean square short circuit current in a bipolar transistor.

$$F \equiv \frac{\overline{i_{no}^2}\Big|_{total}}{\overline{i_{no}^2}\Big|_{v_{ng}}} = 1 + \frac{\overline{i_{no}^2}\Big|_{v_{nbb}} + \overline{i_{no}^2}\Big|_{i_{nb}} + \overline{i_{no}^2}\Big|_{i_{nc}}}{\overline{i_{no}^2}\Big|_{v_{ng}}} \tag{40}$$

$$= 1 + \frac{\overline{v_{nbb}^2}}{\overline{v_{ng}^2}} + \frac{\overline{i_{nb}^2}}{\overline{v_{ng}^2}}\left(R_{bb} + R_g\right)^2 + \frac{\overline{i_{nc}^2}}{\overline{v_{ng}^2}}\frac{1}{g_m^2}\left[1 + \frac{R_{bb} + R_g}{R_\pi}\right]^2$$

To express bias dependencies, we introduce the bipolar transistor parameter relationships

$$g_m = \frac{I_C}{V_t}, \qquad R_\pi = \frac{\beta}{g_m} = \frac{\beta\, V_t}{I_C}, \qquad V_t = \frac{k\,T}{q}. \tag{41}$$

where β is the common emitter DC current gain, which is supposed to be constant in the calculations. It is furthermore assumed that possible recombination currents are a small part of the base current, so β is also the scaling between the shot noises from I_C and I_B, i.e.

$$\overline{i_{nc}^2} = 2\,q\,I_C\Delta f, \qquad \overline{i_{nb}^2} = 2\,q\,I_B\Delta f = \frac{1}{\beta}\,2\,q\,I_C\Delta f = \frac{\overline{i_{nc}^2}}{\beta}. \tag{42}$$

Finally we introduce the two thermal noise terms from the generator and the base series resistors respectively,

$$\overline{v_{ng}^2} = 4\,k\,T\,R_g\Delta f, \qquad \overline{v_{nbb}^2} = 4\,k\,T\,R_{bb}\Delta f. \tag{43}$$

Inserting the relationships above into Eq.(40), the noise figure may be expressed as a function of the collector bias current I_C to read

$$F = \left(1 + \frac{1}{\beta}\right)\left(1 + \frac{R_{bb}}{R_g}\right) + \frac{1}{2\,R_g}\left[\frac{V_t}{I_C} + \frac{\left(R_{bb} + R_g\right)^2}{\beta\,V_t}\left(1 + \frac{1}{\beta}\right)I_C\right]. \tag{44}$$

The factor in brackets has terms that are inversely and directly proportional to I_C. For a given generator resistance R_g, we seek the collector current that minimizes noise figure F,

$$\frac{\partial F}{\partial I_C} = \frac{1}{2R_g}\left[-\frac{V_t}{I_C^2} + \frac{\left(R_{bb} + R_g \right)^2}{\beta\, V_t}\left(1 + \frac{1}{\beta} \right) \right] = 0 \ , \quad \Rightarrow \tag{45}$$

$$I_{C,minF} = \frac{V_t}{R_{bb} + R_g}\,\frac{\beta}{\sqrt{\beta + 1}} \approx \frac{V_t\,\sqrt{\beta}}{R_{bb} + R_g} \ . \tag{46}$$

At the minimizing $I_{C,minF}$ current, the noise figure becomes

$$F_{I,min} = \left(1 + \frac{1 + \sqrt{\beta + 1}}{\beta} \right)\left(1 + \frac{R_{bb}}{R_g} \right) . \tag{47}$$

Considered isolated, the last result suggests use of a high generator resistance to keep the noise figure down. Simultaneously, however, Eq.(46) dictates a low collector current. The transconductance $g_m = \mathrm{Re}\{y_{21}\}$ is proportional to the collector current, so a small I_C implies a small power gain in an amplifier. This is our first mentioning of a common problem in design considerations including noise, namely that conditions for low noise figure and high gain differ, so a trade-off between the two has to be made.

Keeping the collector current fixed, there must be a generator resistance that minimizes the noise figure. The first term in Eq.(44) reduces wile its second term eventually will grow with increasing R_g. Rearranging the equation gives the noise figure expression

$$F = \left(1 + \frac{1}{\beta} \right)\left[1 + \frac{I_C R_{bb}}{\beta\, V_t} + \left(R_{bb} + \frac{I_C R_{bb}^2}{2\, V_t\, \beta} + \frac{V_t\, \beta}{2\, I_C (1 + \beta)} \right)\frac{1}{R_g} + \frac{I_C}{2\, \beta\, V_t} R_g \right] . \tag{48}$$

Minimum in the noise figure requires

$$\frac{\partial F}{\partial R_g} = \left(1 + \frac{1}{\beta} \right)\left[-\left(R_{bb} + \frac{I_C R_{bb}^2}{2\, V_t\, \beta} + \frac{V_t\, \beta}{2\, I_C (1 + \beta)} \right)\frac{1}{R_g^2} + \frac{I_C}{2\, \beta\, V_t} \right] = 0 \ , \quad \Rightarrow \tag{49}$$

$$R_{g,minF} = \sqrt{R_{bb}^2 + R_{bb}\,\frac{2\,\beta\, V_t}{I_C} + \frac{\beta^2 V_t^2}{I_C^2 (1 + \beta)}} \ , \tag{50}$$

which finally provides

$$F_{R,min} = \left(1 + \frac{1}{\beta} \right)\left[1 + \frac{I_C}{\beta\, V_T}\left(R_{bb} + R_{g,minF} \right) \right] . \tag{51}$$

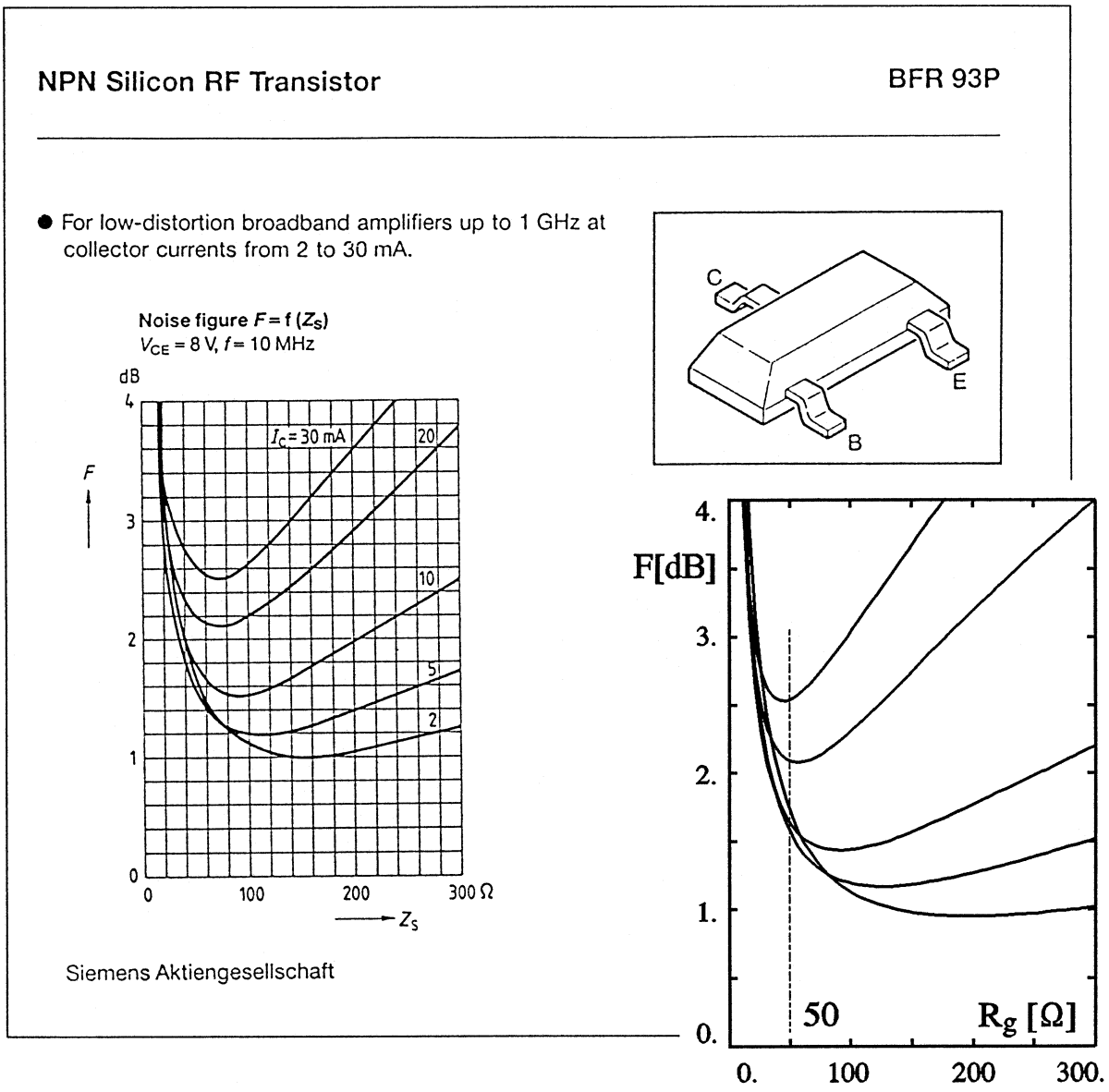


Fig.20 Extract from Siemens RF and Microwave Transistors and Diodes Data Book 1990/91 exemplifying low-frequency noise figures. The insert demonstrates similar theoretical results.

A practical example of the two types of noise figure minima may be observed in the data book extract in Fig.20. With a fixed generator resistor R_g of 50Ω (Z_s in data), there is a minimum in the noise figure between the I_C = 2mA and 5mA curves. Keeping the collector currents fixed, the noise figure curves indicate the minimizing generator resistance very distinctly. The example here is a RF transistor usable to 1GHz. The noise figure in the extract apply to 10MHz, and we should expect that the simple low to middle frequency equivalent circuit from Fig.19 apply to the data. To investigate the question we search pairs of β,R_{bb} that satisfy corresponding $F_{R,min}$, $R_{g,minF}$ readings from the curves. The process was conducted numerically requiring Eqs.(48) and (51) to fit the data. Table I summarizes the results and, as seen, neither β nor R_{bb} are completely constant. However they follow expectable variations,

Table I Parameters for transistor data in Fig.20. β and R_{bb} are fitted to reproduce $F_{R,min}$ and $R_{g,minF}$ in the data.

I_C[mA]	$F_{I,min}$[dB]	$R_{g,minF}$[Ω]	β	R_{bb}[Ω]
30.	2.5	70.	85.	11.5
20.	2.1	75.	90.	12.0
10.	1.5	90.	110.	13.0
5.	1.2	120.	95.	13.5
2.	1.0	155.	75.	14.0

cf.ref´s [13] p.70 or [9] p.56, where the current gain often has a maximum before the high current rating of the transistor (50 mA for BFR93) due to high injection across the emitter base junction. High injection also increases conductivity in the base beneath the emitter area, which reduces the base series resistance. The third but less direct consequence of high injection is the fact that despite β and R_{bb} are adjusted to fit the minimum conditions, complete coincidence between data and theoretical curves are not achieved. The reason is the transconductance g_m is reduced compared to the ideal value of I_C/V_t by a factor that approaches one half. Even with this precautions, the simplified theoretical results are qualitatively correct and accurate enough for initial designs, especially taking into account that the base series resistance R_{bb} is a parameter that commonly not is accurately known. As a matter of fact, noise measurements and identification like the example here is suggested as a method to determine the base series resistance experimentally, cf. ref.[9] p.291.

Example IV-2-1 end

Noise in Field Effect Transistors

The basic operation of junction and metal gate field effect transistors, JFET's or MESFET's - are that the resultant conductance of a semiconductor region in a channel between the drain and source terminals is modulated primarily by the gate to source voltage but also by the drain to source voltage. The mechanism of modulation is that a region of the channel beneath the gate junction or electrode is depleted from mobile carriers under control of the gate to channel voltage. The depleted region - sketched by the hatched area in Fig.21 - cannot conduct current. Due to the voltage drop along the channel it narrows to practically zero in the drain end under normal biasing to pinched operation, which is the only condition we consider. JFETs and MESFETs are our vehicles for presentation. The physics of MOSFET´s is quite different and more involved regarding the formation of a conducting channel under the isolated gate, but in small-signal and noise respects, there are no needs to make major distinctions between MOSFET's and other FET types, so the discussion below applies to all types of field effect transistors.

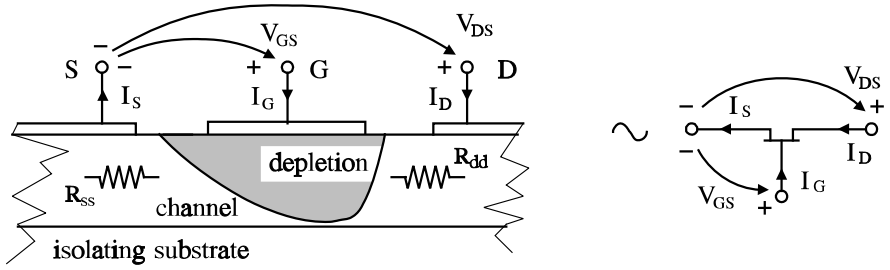


Fig.21 Structure of field effect transistor (MESFET or junction FET). Channel noise originates from thermal and flicker noise in the undepleted part of the channel.

Being a voltage controlled conductance the major noise mechanisms in field effect transistor are thermal noise and flicker noise of various origins. An equivalent circuit that includes most of the contributions that are suggested in literature is shown by Fig.22. The series noise voltages are associated with bulk regions outside the channel. They are commonly assumed independent of bias and obey simple thermal noise rules, i.e.

$$\overline{v_{ngg}^2} = 4kTR_{gg}\Delta f\,, \qquad \overline{v_{ndd}^2} = 4kTR_{dd}\Delta f\,, \qquad \overline{v_{nss}^2} = 4kTR_{ss}\Delta f\,. \tag{52}$$

The series resistances to the channel may include flicker noise in proportion to the square of the drain current. It is, however, difficult to identify the series resistances explicitly. The effect of a small source resistor R_{ss} is a reduction of the transconductance that is seen from the device terminals by a denominator equal to $(1+g_mR_{ss})$. At the drain side the small resistance R_{dd} is series connected to a high output impedance from the active part of the transistor. Without detailed knowledge about processing of the transistor, the effects of channel series resistances may be accounted for by the remaining part of the equivalent circuit.

Thermal and flicker noise from the undepleted part of the channel is described by the $\overline{i^2}_{nd}$ noise current in the equivalent circuit. There is a close connection between transconductance and channel conductance. If the transistor is biased for a given g_m, the thermal noise and flicker noise that arise from channel conductance may be included by the two terms,

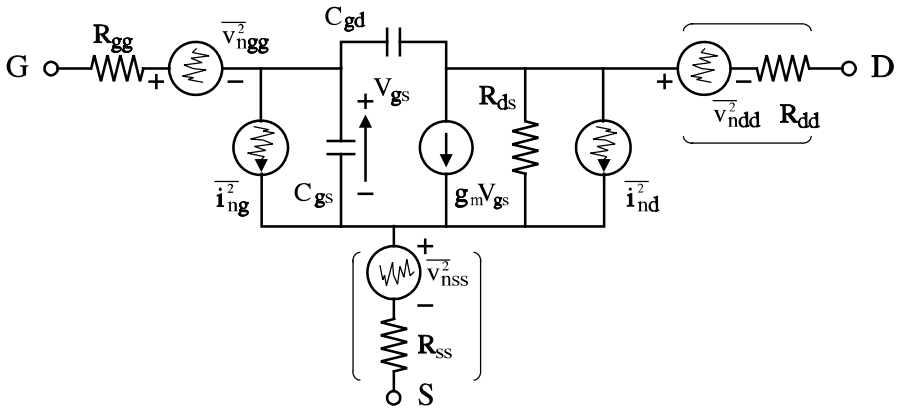


Fig.22 FET noise. $\overline{i^2}_{nd}$, channel noise incl. flicker. $\overline{i^2}_{ng}$, leakage shot and induced noise. Series resistances and noise sources in parentheses are often omitted.

$$\overline{i_{nd}^2} = \overline{i_{nd,thermal}^2} + \overline{i_{nd,flick}^2} = \frac{2}{3} \times 4kTg_m\Delta f + K_{ch}\frac{I_D^{a_{ch}}}{f}\Delta f . \qquad (53)$$

The weight factor of 2/3 in front of the thermal noise term is a result of averaging noise contributions along the channel. The last term holds flicker noise characterized by parameters K_{ch} and the drain current exponent, a_{ch}, which may be close to two. Like the similar modeling in bipolar transistors, it is assumed that the Δf bandwidth is small compared to frequency f. Otherwise we must again conduct an integration over the observation bandwidth as it was done in Eq.(22).

There are two highly different contributions to the gate noise current, $\overline{i^2}_{ng}$. The first contributions arises because voltage fluctuations from the channel noise also appear as fluctuations in voltage across the depletion region. It forces a capacitively induced gate current through the circuit that is connected to the gate electrode. Secondly, a possible small gate DC leakage current I_G implies shot-noise like the noise from a reverse biased diode given through Eq.(35) above.

$$\overline{i_{ng}^2} = \overline{i_{ng,induced}^2} + \overline{i_{ng,shot}^2} = 0.25 \times 4kT\frac{(\omega C_{gs})^2}{g_m}\Delta f + 2q\left|I_G\right|\Delta f . \qquad (54)$$

Like in drain noise, the leading factor in the channel induced term represents an averaging of the coupling effect along the channel. Its actual value differs somewhat in literature, partly according to the FET type under investigation, partly due to different ways of defining and including terms, cf. ref´s [4]p.91, [9]p.245, and [14]. The value of 0.25 is chosen halfway between the figures in the first two references. The capacitive nature of the term is emphasized by proportionality between the mean squared induced noise current and the squared susceptance of the gate capacitance, in turns to the squared frequency. Originating from the same noise sources in the channel that formerly gave the first term in Eq.(53), the induced gate noise and the drain noise are correlated having the correlation coefficient

$$c = \frac{\overline{i_{ng,induced} i_{nd,thermal}^*}}{\sqrt{\overline{i_{ng,induced}^2}\;\overline{i_{nd,thermal}^2}}} = j\,0.4 . \qquad (55)$$

We shall see below that it is important to keep track on correlations between noise sources when the noise properties of composite circuits are determined.

IV-3 Noise Characterization of Two-Ports

Small-signal parameters like y or s-parameters provide concentrated characterization of building blocks in electronic circuit that may contain many components. The parameters are both measurable under well defined conditions and directly suitable in design of deterministic signal handling. The number of independent noise sources inside electronic components and subcircuits requires equally simple and concentrated approaches with respect to noise characterization and design for low noise properties. The noise figure fulfill this need and is the most common way of specifying electronic devices with respect to noise. Noise temperature is an alternative or supplement to the noise figure and both concepts are considered in details below. Like the small-signal parameters presentation in the foregoing chapter, the discussion is concentrated on the important two-port case. The corresponding multiport properties may be found elsewhere in literature, for instance ref´s [15] or [16].

Noise Figure

At a given frequency the noise figure concentrates the effect of all internal noise sources of a two-port that is driven by a signal generator into a single figure F, the so-called noise figure[4]. The basic definition is, [17],

$$F = \frac{N_{out,tot}/\Delta f}{N_{out,g0}/\Delta f} \equiv \frac{\textit{total available output noise power per unit bandwidth}}{\textit{available output noise power per unit bandwidth caused by the generator adtmittance or impedance @ } 290\,[K]} \tag{56}$$

It is assumed in measurements that the noise powers $N_{out,tot}$ and $N_{out,g0}$ are determined in a bandwidth, which is sufficiently narrow to take the underlying noise spectra constant, so the noise figure becomes independent of bandwidth. To emphasize this assumption, F above is sometimes referred to as the spot noise figure to be distinguished from the bandwidth dependent so-called average noise figure [17]. The available output powers are independent of the actual loading of the two-port and this property conveys to the noise figure. However, the generator impedance or admittance is significant to the way by which internal noise sources contribute to the output. In consequence, a noise figure F for a two-port is meaningless without a specification of the corresponding generator impedance or admittance. In this

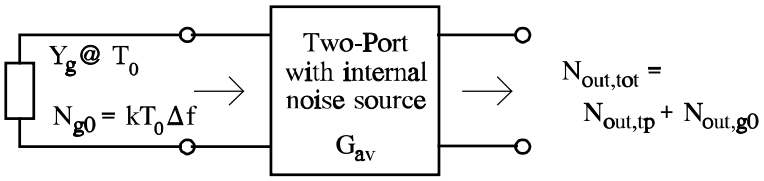


Fig.23 Two-Port with internal noise sources. Y_g is kept at the reference temperature T_0=290K. $N_{out,tp}$ is the effect of the internal two-port noise sources.

---

4) Some authors distinguish between noise factor F and the db-scaled noise figure NF=10.log(F). No such distinctions between terms are made below but, clearly, it is always pointed out whether we are using dB scale or not.

context the generator impedance or admittance is a remedy for noise characterization of the two-port. To get unambiguous two-port noise figures, the generator must be settled at a reference temperature, which by definition is set to T_0= 290 K.

Subdividing the total output noise into terms stemming from sources internal to the two-port and from the generator admittance, as indicated in Fig.23, yields

$$\begin{aligned} N_{out,tot} &= F\,N_{out,g0} = N_{out,tp} + N_{out,g0} \quad \Rightarrow \\ N_{out,tp} &= (F-1)\,N_{out,g0} = (F-1)\,G_{av}N_{g0} = (F-1)\,G_{av}\,kT_0\Delta f \end{aligned} \tag{57}$$

Comparing the upper and lower part shows that the two-ports own contribution to the output noise is a fraction (F-1)/F of the total output noise. In the last sequence above, the generator contribution is referred back to the input through the available power gain G_{av}, which - like the noise figure - is independent of the two-port load and therefore particularly useful in noise calculations. If also the two-port noise is referred to the input side and called $N_{in,tp}$, we get

$$N_{out,tp} = G_{av}\,N_{in,tp} \quad \Rightarrow \quad N_{in,tp} = (F-1)\,kT_0\Delta f \tag{58}$$

If there are no significant internal noise sources, the minimum value of a two-port noise figure is one. This is the case with circuits build from ideal lossless, passive components that do not radiate or receive radio signals.

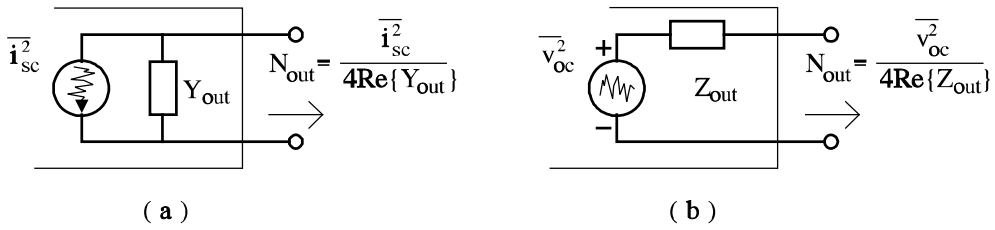


Fig.24 Noise equivalent circuit for the output side of a two-port in a) Norton and b) Thevenin equivalent forms.

The output from the two-port may be described by either a Norton or a Thevenin equivalent circuit as shown in Fig.24. In the first case the available output noise power is given by the real part of the output admittance in conjunction with the squared noise short-circuit current $\overline{i^2}_{sc}$. The output admittance depends on the two-port parameters and the generator admittance as we saw in Chap.III, p.6. To calculate the noise figure according to the definition, the short circuit noise current must be available in two versions, $\overline{i^2}_{scc}|_{tot}$ representing the joint effect of all noise contributions in the two-port and the generator network, and $\overline{i^2}_{sc}|_{g0}$ holding the output noise short-circuit current caused by generator admittance or impedance alone. In duality, the Thevenin equivalent in Fig.24(b) expresses available output noise powers in terms of the two-port output impedance and the squared noise open-circuit voltage, either the total,$\overline{v^2}_{oc}|_{tot}$, or the contribution from the generator alone, $\overline{v^2}_{oc}|_{g0}$. Output admittances or impedances cancel when taking ratios of available powers in the two equivalent representations, so the noise figure is expresses solely in ratios of squared noise short-circuit currents or open-circuit voltages,

$$F = \frac{\overline{i_{sc}^2}\Big|_{tot}}{\overline{i_{sc}^2}\Big|_{Y_g}} = \frac{\overline{v_{oc}^2}\Big|_{tot}}{\overline{v_{oc}^2}\Big|_{Z_g}} \tag{59}$$

These results - sometimes called Van der Ziels equations [18] - are highly useful for computing noise figures in composite networks, for instance device models as it was anticipated by example IV-2-1 on page 18. The effects of separate noise sources are combined by accumulating their contributions to the short-circuit current or open-circuit voltage. This way of calculating noise figures free us from the crucial problem that the concept of available power at the output of the two-port implicitly assumes positive real parts of the output admittance or impedance. If the two-port is potentially unstable, this condition may not be satisfied. Nevertheless the two-port may still be useful in practice if it is properly loaded. By the equations above the noise figure of the two-port may also be calculated and it is still independent of the actual load.[5]

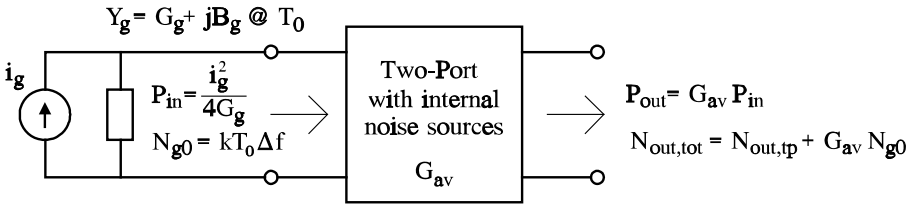


Fig.25 Noisy two-port connected to a signal source. P_{in}, P_{out} are available signal powers, and N_{Yg}, $N_{out,tot}$ are the available noise powers at either side of the two-port.

Noise figures may also be expressed in terms of signal to noise ratios, SNR's, at the input and output side of two-ports. Applying a signal source of admittance Y_g and available signal power P_{in} to the input side of a two-port as shown by Fig.25, we get a development from the formalized definition in Eq.(56) that reads

$$F = \frac{N_{out,tot}}{N_{out,g0}} = \frac{P_{in}N_{out,tot}}{P_{in}G_{av}N_{g0}} = \frac{P_{in}/N_{g0}}{P_{out}/N_{out,tot}} = \frac{SNR_{in,0}}{SNR_{out}} \ . \tag{60}$$

Here G_{av} is the available power gain of the two-port that is driven by a generator of admittance Y_g and P_{out} is the available signal output power. Since the noise figure cannot be smaller than one, the noise figure may be considered as the divisor by which a signal to noise ratio is deteriorated in passage through the two-port

$$SNR_{out} = \frac{1}{F} SNR_{in,0} \ . \tag{61}$$

---

5) The more formalistic approach of extending noise figure definitions to encompass two-ports that get negative output conductance or resistance is given in [15] or [16].

This way of interpreting noise figures may be useful when it is adequate to have the generator impedance at reference temperature T_0=290K as assumed by the noise figure definition. If this is not the case, a so-called effective noise figures satisfying the SNR conditions above at a given temperature, is sometimes introduced. However, in such cases it makes more sense to use the concept of two-port noise temperatures, which are described in the next paragraph.

Example IV-3-1 (noise figure measurement)

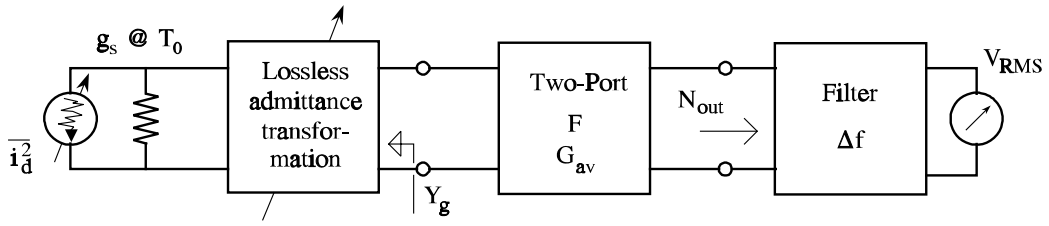


Fig.26 Setup for noise figure measurement. The adjustable noise current source is typically a diode where the noise is directly controlled by the dc-current.

A principle for measuring two-port noise figures is shown in Fig.26. The generator admittance for the noise figure to be measured is provided by an impedance transforming network, which is assumed to be lossless. The available noise power out of the transformer equals the available power from the source. The source includes a conductance g_s held at the reference temperature and an additional, adjustable noise current generator $\overline{i^2_d}$. The output noise power N_{out} is the portion of the available output power within the noise-bandwidth Δf of the filter that connects to and RMS voltmeter. The squared meter reading is proportional to N_{out}. Two readings are required to measure a noise figure. First the noise generator $\overline{i^2_d}$ is kept at zero and we get the output reading $V_{RMS,0}$. Second, the noise generator is adjusted to double the output power, i.e. to a reading 3dB or $\sqrt{2}$ above the first one, i.e.

$$\begin{aligned} &\underline{step\ 1}: && V_{RMS0} \sim N_{out,0} = G_{av} F k T_0 \Delta f \\ &\underline{step\ 2}: && V_{RMS} = \sqrt{2}\, V_{RMS0} \sim 2N_{out,0} = G_{av} F k T_0 \Delta f + G_{av} \frac{\overline{i_d^2}}{4 g_s} \end{aligned} \qquad (62)$$

By this procedure the two terms in the last reading are equal. The adjustable noise source is often a diode where the dc-current controls the squared noise current, cf.Eq.(33), so we get

$$F = \frac{\overline{i_d^2}}{4 k T_0 g_s \Delta f} \underset{diode}{=} \frac{I_D}{2 g_s V_t}, \quad where \quad V_t = \frac{k T_0}{q} \approx 25{,}0\ mV \qquad (63)$$

The last noise figure expression shows that a DC-meter, which measure the diode current I_D, may be calibrated directly in noise figures.

Example IV-3-1 end

Noise Temperature of Two-Ports

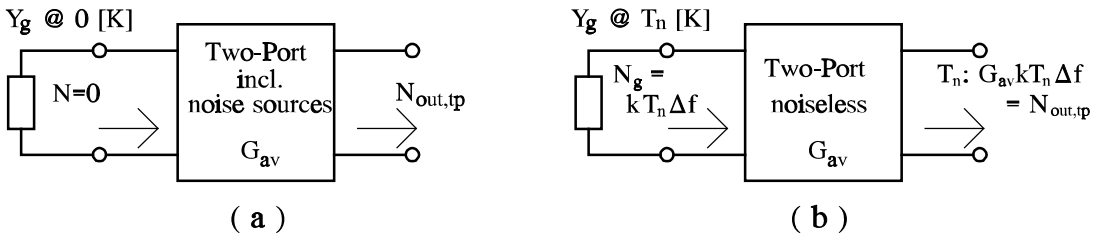


Fig.27 Noise temperature of two-port, T_n. The noise generated inside the two-port in (a) is ascribed to the generator admittance at temperature T_n in (b).

An effective noise temperature was introduced for one-ports by Eq.(25) as a method of specifying its available noise power. Without regards to the actual noise sources, the noise temperature was the temperature in Kelvin of the one-port, had all noise been of thermal origin. In extension, the noise temperature of a two-port driven from a generator is defined as the temperature of the generator if the entire noise contribution from inside the two-port is ascribed to thermal noise from the generator impedance or admittance, cf.[17]. The conditions are depicted in Fig.27, where in (a) the source hypothetically is kept at zero Kelvin, so the available output noise power comes solely from internal sources of the two-port. Assuming no internal noise sources, the noise temperature T_n provides the same output noise power if the temperature of the generator admittance is raised to T_n [K]. Like noise figures, the noise temperature of a two-port needs a specification of the pertinent generator admittance or impedance. Using Eq.(58), the relationship between the two noise characterizations becomes,

$$N_{out,tp} = G_{av}(F-1)\,k\,T_0\,\Delta f = G_{av}k\,T_n\,\Delta f \;\Rightarrow\; \begin{cases} T_n = (F-1)\,T_0\,, & (a) \\ F = 1 + \dfrac{T_n}{T_0}\,. & (b) \end{cases} \tag{64}$$

where T_0 still is the reference temperature 290 [K].

If the effective noise temperature of the generator admittance in Fig.27 is T_g before any connection to the two-port, the generator itself will contributes to the output available power by an amount we call $N_{out,g}$. Including noise from the two-port, the total output available noise power is expressed

$$N_{out,tot} = N_{out,tp} + N_{out,g} = G_{av}k\,T_n\,\Delta f + G_{av}k\,T_g\,\Delta f = G_{av}k\,(T_n + T_g)\,\Delta f\,. \tag{65}$$

Thus, the two noise temperatures are added to express the total, resultant noise. It is this additive property that makes noise temperatures a convenient tool for noise calculations handling other temperatures than the noise figure reference T_0.

Noise in Cascaded Two-Ports

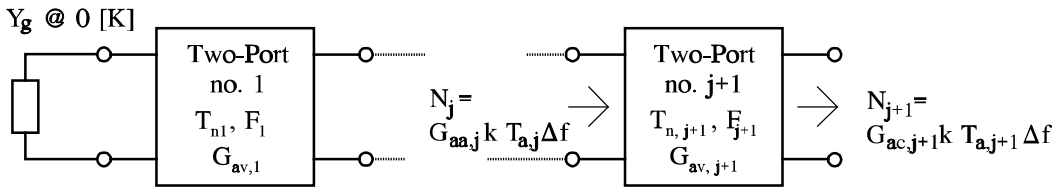


Fig.28 Input and output available noise powers around port no. j+1 in a chain of noisy two-ports. $G_{av,j}$ is the accumulated available gain and $T_{a,j}$ is the accumulated noise temperature.

If more two-ports are cascaded, the first one will usually be the most significant with respect to noise figure or noise temperature for the whole chain. To see this, we calculate the contribution from the j+1'th two-port in the chain. The preceding j ports present an available noise power N_j at the input to port number j+1. Like the first expression in Eq.(64), N_j is related to the accumulated noise temperature $T_{a,j}$ and accumulated available gain $G_{aa,j}$ through

$$N_j = G_{aa,j}\, k\, T_{a,j}\, \Delta f, \qquad \textit{where} \qquad G_{aa,j} = G_{av,1}\, G_{av,2} \cdots G_{av,j} \ . \tag{66}$$

Seen from the j+1'th two-port, the product $G_{aa,j}\ T_{a,j}$ is also the effective generator noise temperature. To include the noise that is generated by this port, a noise temperature $T_{n,j+1}$ must be added to yield

$$N_{j+1} = G_{av,j+1}\, k \left[G_{aa,j} T_{a,j} + T_{n,j+1} \right] \Delta f = G_{av,j+1}\, k \left[G_{aa,j} (T_{a,j} + \Delta T_{a,j}) \right] \Delta f \ . \tag{67}$$

By the last rewriting the noise contribution from two-port number j+1 is considered as an increment of the accumulated noise temperature from the j'th to the j+1'th step. Equating terms gives

$$\Delta T_{a,j} = \frac{T_{n,j+1}}{G_{aa,j}} \ . \tag{68}$$

Including the recursion for the gains, the development in noise temperature when two-ports are cascaded is given by

$$T_{n,chain} = T_{n,1} + \frac{T_{n,2}}{G_{av,1}} + \frac{T_{n,3}}{G_{av,1}\, G_{av,2}} + \frac{T_{n,4}}{G_{av,1}\, G_{av,2}\, G_{av,3}} \cdots + \frac{T_{n,j+1}}{G_{av,1}\, G_{av,2} \cdots G_{av,j}} \cdots \ . \tag{69}$$

Translating to noise figures, repeated use of Eq.(64) provides

$$T_{n,chain} = T_0 \ (\ F_{chain} - 1 \ ) \ , \qquad T_{n,j} = T_0 \ (\ F_j - 1 \ ) \qquad \Rightarrow \tag{70}$$

$$F_{chain} = F_1 + \frac{F_2 - 1}{G_{av,1}} + \frac{F_3 - 1}{G_{av,1}\,G_{av,2}} + \frac{F_4 - 1}{G_{av,1}\,G_{av,2}\,G_{av,3}} \cdots + \frac{F_{j+1} - 1}{G_{av,1}\,G_{av,2} \cdots G_{av,j}} \cdots . \quad (71)$$

The last expression is called Friis' formula for cascaded two-ports [19]. Equations (69) and (71) show that if the first two-port in a chain has sufficient gain, the noise temperature or noise figure of the first stage dominate the whole chain.

Example IV-3-2 (link budget)

The radio link system in Fig.29 operates in QPSK modulation and transmits the bit sequence b_k at rate R_b=8.2 Mbps. The total attenuation from the transmitter output to the RF amplifier input is 137 dB. All components are supposed to be matched. The antenna has noise temperature T_{ant}=195 K and the RF amplifier has noise figure F=6.5 dB. Find the transmitter output power P_{tr} that gives the bit error rate BER=10^{-6} after demodulation.

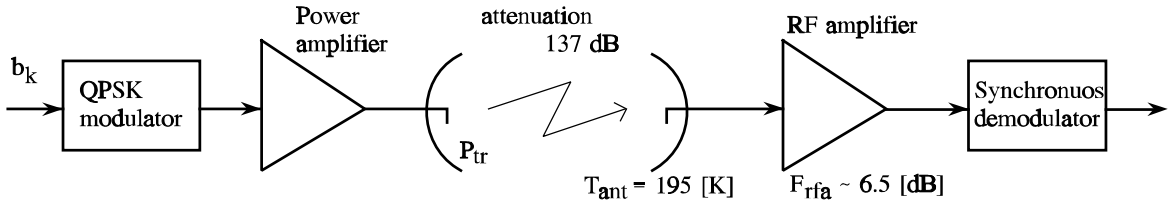


Fig.29 Block diagram for a radio link.

Bit error rates in QPSK-modulation corresponds to PSK-PRK modulation, which was discussed in chap.I, where Fig.I-29 provides the signal to noise ratio per bit, BER,

$$BER = Q\left(\sqrt{\frac{2\,E_b}{\eta}} \right) = 10^{-6} \quad \Rightarrow \quad (72)$$

$$\frac{E_b}{\eta} = \frac{E_b}{k\,T_{eff}} = 10.5\,[dB] \sim 11.2 \; . \quad (73)$$

The noise power spectral density is held in $\eta = kT_{eff}$ with T_{eff} being the total effective noise temperature at the RF-amplifier input. To find T_{eff}, the amplifier noise figure is converted to noise temperature T_{rfa} by Eq.(64)(a). We get

$$T_{eff} = T_{ant} + T_{rfa} \; , \quad T_{rfa} = T_0 \; (\; F_{rfa} - 1 \;) \; , \quad F_{rfa} \sim 6.5\,[dB] \sim 4.47 \Rightarrow$$
$$T_{rfa} = 290 \; (\; 4.47 - 1 \;) = 1006.\,[K] \; , \quad T_{eff} = 195 + 1006 = 1201\,[K] \; . \quad (74)$$

From Eq.(73), the required energy per bit at the receiver input is

$$E_b = 11.2 \cdot k\ T_{eff} = 11.2 \cdot 1.381 \cdot 10^{-23}\ [Ws/K] \cdot 1201\ [K] = 1.858 \cdot 10^{-19}\ [Ws]\ . \quad (75)$$

To calculate the corresponding input power, we recall that in QPSK the symbol time is twice the input bit period. However, two bits are simultaneously transmitted in quadrature, so the necessary input power is the ratio of the bit energy over the input bit period, i.e.

$$P_{in} = E_b\ R_b = 1.858 \cdot 10^{-19} \cdot 8.2 \cdot 10^{6} = 1.523 \cdot 10^{-12}\ [W]. \quad (76)$$

Taking attenuation from transmitter to receiver into account, the transmitter power becomes

$$D \sim 137\,[dB] \sim 5.01 \cdot 10^{13} \quad \Rightarrow$$
$$\underline{P_{tr}} = P_{in}\ D = 1.523 \cdot 10^{-12}\ 5.01 \cdot 10^{13} = \underline{76.30\,[W]}\ . \quad (77)$$

Suppose the transmitter amplifier has a maximum undistorted output of P_{tr}=40[W], and the receiver is improved by inserting a low noise amplifier, LNA, of 6 dB gain in front of the receiver as shown in Fig.30. Find the noise figure requirement F_{LNA} to the low-noise amplifier, if the resultant BER should remain unchanged.

Unaffected BER requires unaffected signal to noise ratio at the detector or equivalently, that the total effective noise temperature at the receiver input is reduced by the same factor as is the input power, i.e.

$$T_{eff,new} = T_{eff}\ \frac{P_{tr,new}}{P_{tr}} = 1201.\ \frac{40.}{76.30} = 629.6\ [K]\ . \quad (78)$$

Subtracting antenna noise temperature from $T_{eff,new}$ gives the noise temperature T_{cas} of the two cascaded amplifiers, which is expressed through Eq.(69). Therefore,

$$T_{cas} = T_{LNA} + \frac{T_{amp}}{G_{LNA}} = T_{eff,new} - T_{ant} = 629.6 - 195 = 434.6[K]\ ,$$
$$G_{LNA} \sim 6\,[dB] \sim 3.98 \quad \Rightarrow \quad T_{LNA} = 434.6 - 1006/3.98 = 181.8\,[K]\ . \quad (79)$$

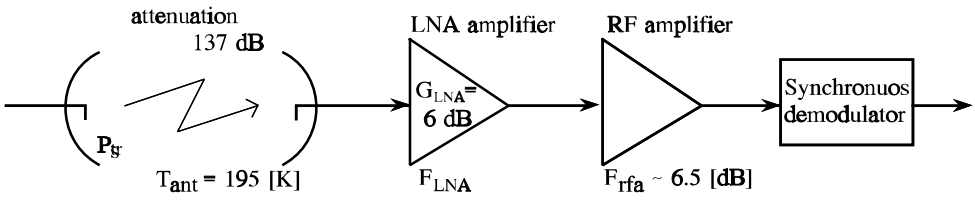


Fig.30 Inclusion of a low-noise amplifier, LNA, in front of the receiver from Fig.29.

Calculating back to noise figure by Eq.(64)(b) yields the LNA noise figure requirement,

$$\underline{F_{LNA}} = 1 + \frac{T_{LNA}}{T_0} = 1 + \frac{181.8}{290.} = \underline{1.627} \sim \underline{2.11\,[dB]}\ . \tag{80}$$

Example IV-3-2 end

Noise Representation in Two-Ports

Noise figures and temperatures are suitable for calculating the type of noise performance required in system design, for instance signal-to-noise ratio degradations and bit error rates. Like gain functions, noise figure and temperature depends on the operating conditions of the two-port in the form of either admittance or impedance of the driving generator. It is desirable, therefore, to have a more basic noise characterization that is self-contained and from which questions about the best way of using a given two-port regarding noise may be investigated.

Terminal currents are the dependent variables in conventional y-parameter small-signal characterization of two-ports, so in this representation it is natural to include noise properties by paralleling noise currents across the two ports as shown in Fig.31. The two noise generators represent the joint contributions from all noise sources inside the two-port to the short-circuit terminal noise currents. An internal noise source may contribute to the short-circuit currents at both the input and the output side of the two-port, so the two equivalent representations $\overline{i^2_{n1}}$ and $\overline{i^2_{n2}}$ may be correlated. The noise current i_{n1} and i_{n2} have real power spectra $S_1(\omega)$ and $S_2(\omega)$ respectively. They are normalized to 1Ω. In a bandwidth Δf around center frequency f_0 the two noise sources are characterized by the noise conductances $G_{n1}(f_0)$ and $G_{n2}(f_0)$ defined through the spectra

$$\begin{aligned} G_{n1}(f_0)\ :\ \overline{i_{n1}^2} &= 2S_1(\omega_0)\Delta f = 4kT_0 G_{n1}(f_0)\Delta f\ , \\ G_{n2}(f_0)\ :\ \overline{i_{n2}^2} &= 2S_2(\omega_0)\Delta f = 4kT_0 G_{n2}(f_0)\Delta f\ . \end{aligned} \tag{81}$$

Using noise conductances G_{n1} and G_{n2} is a convenient method of quantifying noise levels, but observe that the figures are not ascribed to any particular circuit elements. The correlation coefficient between the input and the output noise is defined by

$$c = \frac{\overline{i_{n1}\, i_{n2}^*}}{\sqrt{\overline{i_{n1}^2}\ \overline{i_{n2}^2}}} = \frac{S_{21}(\omega_0)}{\sqrt{S_1(\omega_0)\ S_2(\omega_0)}}\ , \tag{82}$$

where $S_{21}(\omega)$ is the cross spectrum between the input and the output noise currents. Even when the power spectra for the two currents are real, their cross spectrum may be complex,

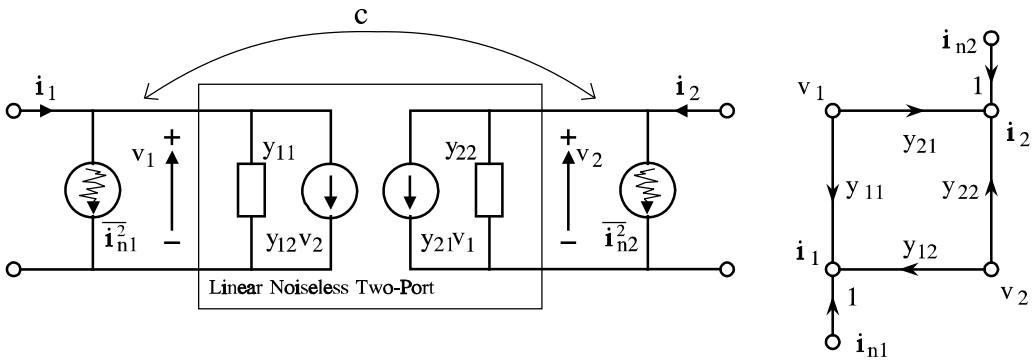


Fig.31 Noise representation in a linear two-port by two short-circuit noise currents. The possible correlation is expressed through the complex coefficient c.

and so does the correlation coefficient c. Four parameters are therefore required to describe the noise properties of the two-port at a given frequency, G_{n1}, G_{n2}, Re{c}, and Im{c}.

The representation in Fig.31 may be useful in basic noise analysis, for instance of device models where the short-circuit noise currents at either port and their possible correlation often are obtained easily. In signal calculations, however, it is desirable to let all noise contributions refer to the input side of the two port. A common representation is here to use a series noise voltage and a shunt noise current as shown by Fig.32, where γ denotes the correlation coefficient between the noise generators [20]. We may establish the connection between the two representations by setting their terminal short-circuit noise currents equal to each other. The flow graphs illuminate this process and give

$$\left.\begin{array}{ll} \textit{input port, } i_1: & i_{n1} = i_{n,tot} - y_{11} v_n\,, \quad (a) \\ \textit{output port, } i_2: & i_{n2} = -y_{21} v_n\,, \quad (b) \end{array}\right\} \Rightarrow \left\{\begin{array}{ll} i_{n,tot} = i_{n1} - \dfrac{y_{11}}{y_{21}} i_{n2}\,, & (c) \\ v_n = -\dfrac{1}{y_{21}} i_{n2}\,. & (d) \end{array}\right. \tag{83}$$

Before setting up the relationship between correlations in the two representations, the shunt current $i_{n,tot}$ is split into a part i_n that is completely uncorrelated with v_n and a part that is

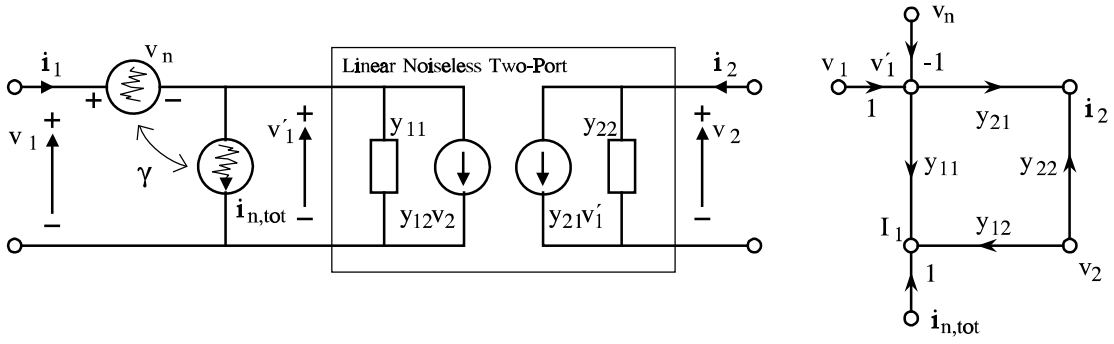


Fig.32 Linear two-port noise representation referred to the input port. Series noise voltage and shunt noise current may be dependent with complex correlation coefficient γ.

fully correlated. The last one follows v_n by a constant, complex scale factor that has dimension of admittance and therefore is denoted Y_{cor}.[6] We get

$$\boldsymbol{i}_{n,tot} = \boldsymbol{i}_n + Y_{cor}\boldsymbol{v}_n \quad \textit{where} \quad \overline{\boldsymbol{i}_n \boldsymbol{v}_n^*} = 0 \quad \Rightarrow \quad \gamma \equiv \frac{\overline{\boldsymbol{i}_{ntot}\boldsymbol{v}_n^*}}{\sqrt{\overline{v_n^2}\;\overline{\boldsymbol{i}_{ntot}^2}}} = Y_{cor}\sqrt{\frac{\overline{v_n^2}}{\overline{\boldsymbol{i}_{ntot}^2}}}\,, \tag{84}$$

so Y_{cor} may substitute for the γ correlation coefficient. Equations (82) and (83)(a-b) provide

$$\overline{\boldsymbol{i}_{n2}^*\boldsymbol{i}_{n1}} = -y_{21}^*\left(\overline{\boldsymbol{v}_n^*\boldsymbol{i}_{n,tot}} - y_{11}\overline{v_n^2}\right) = -y_{21}^*\left(Y_{cor} - y_{11}\right)\overline{v_n^2} = \frac{-1}{y_{21}}\left(Y_{cor} - y_{11}\right)\overline{\boldsymbol{i}_{n2}^2} \;\Rightarrow$$

$$Y_{cor} = y_{11} - y_{21}\frac{\overline{\boldsymbol{i}_{n1}\boldsymbol{i}_{n2}^*}}{\overline{\boldsymbol{i}_{n2}^2}} = y_{11} - y_{21}\,c\sqrt{\frac{\overline{\boldsymbol{i}_{n1}^2}}{\overline{\boldsymbol{i}_{n2}^2}}}\,. \tag{85}$$

Like the first equivalent circuit, the one with noise referred to the input port may be characterized by equivalent noise components, resistance R_n to describe the series noise voltage and conductance G_n to describe the uncorrelated part of the shunt current. They are defined and - by comparing equivalent circuits - related to the previous parameters in Eq.(81) through

$$\boldsymbol{R}_n: \ \overline{v_n^2} = 4kT_0\boldsymbol{R}_n(f_0)\Delta f\,, \qquad G_n: \ \overline{\boldsymbol{i}_n^2} = 4kT_0G_n(f_0)\Delta f\,,$$

$$\textit{where} \quad \boldsymbol{R}_n = \frac{G_{n2}}{|y_{21}|^2}. \qquad \textit{where} \quad G_n = G_{n1} + G_{n2}\frac{|y_{11} - Y_{cor}|^2}{|y_{21}|^2}\,. \tag{86}$$

Four parameters are again used to characterize the noise performance of the two-port, R_n, G_n, and the two components of Y_{cor}. A direct representation of these noise parameters is shown in Fig.33(b). Due to the parallelling of two admittances of opposite sign, the port gives no loading to signal transmissions, but a short-circuit on either port will carry the full squared noise current, $\overline{i^2}_{n,tot}$, while the opposite open port shows the squared noise voltage $\overline{v^2}_n$.

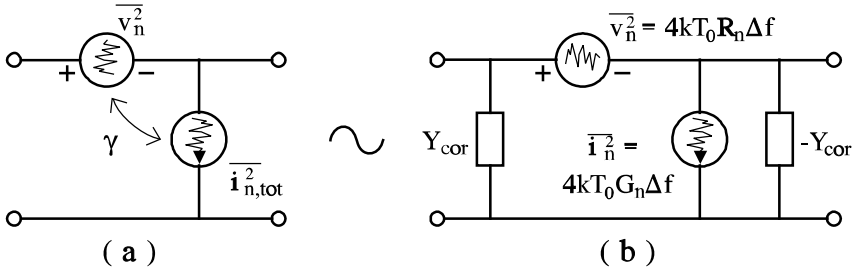


Fig.33 Equivalent representation of a noisy two-port specified by parameters R_n, G_n, and the noiseless correlation admittance Y_{cor}

---

6) Be aware that the correlation admittance Y_{cor} is not uniquely defined in the literature. Here we follow the definition in [20].

Minimum Noise Conditions

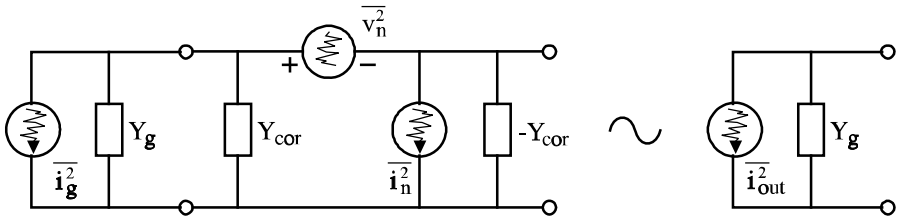


Fig.34 Equivalent circuit for calculating the noise figure of a two-port connected to a source, which by definition must have the reference temperature T_0=290[K].

Concentrating the noise characterization to the input side of a two-port means that the rest of the port is noiseless as indicated in Fig.32. A noiseless two-port has noise figure equal to one and, according to the cascading rule in Eq.(71), this will not influence the noise figure when the two port is connected to a generator. To investigate the noise figure of the complete two-port, it suffices to consider the generator and the noise input equivalent parts that are shown in Fig.34. The definition of the noise figures assumes that the generator admittance $Y_g=G_g+jB_g$ has settled at the reference temperature T_0, so its squared noise current becomes

$$\overline{i_g^2} = 4kT_0G_g(f_0)\Delta f \,. \tag{87}$$

Now, Eq.(59) can be used to calculate the noise figure by taking ratios between short-circuit contributions. We get

$$\begin{aligned} F &= \frac{\overline{i_{out}^2}|_{tot}}{\overline{i_{out}^2}|_{Y_g}} = \frac{\overline{i_g^2} + \overline{i_n^2} + \overline{v_n^2}\,|Y_g + Y_{cor}|^2}{\overline{i_g^2}} \\ &= 1 + \frac{G_n}{G_g} + \frac{R_n}{G_g}\left[(G_g + G_{cor})^2 + (B_g + B_{cor})^2\right], \end{aligned} \tag{88}$$

where the last equation includes the noise parameters from Eq.(86) and the components of the complex correlation admittance $Y_{cor}=G_{cor}+jB_{cor}$. To find the generator admittance Y_{nfo} that minimizes the noise figure, it is readily seen, that one requirement toward minimizing the noise figure is to use a generator susceptance, which cancels the correlation susceptance. This is called noise tuning, i.e.

$$\underline{\textit{Noise Tuning}}: \quad B_g = B_{nfo} = -B_{cor}. \tag{89}$$

The conductance that simultaneously minimizes the noise figure is found by setting the partial derivative of Eq.(88) with respect to generator conductance equal zero,

$$\left.\frac{\partial F}{\partial G_g}\right|_{B_g=-B_{cor}} = \frac{1}{G_g^2}\left(G_g^2R_n - G_{cor}^2R_n - G_n\right) = 0 \qquad \Rightarrow \qquad G_g = G_{nfo} = \sqrt{\frac{G_n}{R_n} + G_{cor}^2}\,. \tag{90}$$

The 2nd order derivative becomes $2R_nG_{nfo}^2/G_g^3$, which is positive if G_g is positive, so the result above gives a minimum in noise figure. Simultaneous use of the two conditions is called noise matching, where

$$\underline{\textit{Noise Matching}}: \quad Y_g = Y_{nfo} \equiv G_{nfo} + jB_{nfo} = \sqrt{\frac{G_n}{R_n} + G_{cor}^2} - jB_{cor}\,. \tag{91}$$

The corresponding minimum noise figure is found by inserting Y_{nfo} into Eq.(88),

$$\begin{aligned} F_{min} &= 1 + \frac{R_n}{G_{nfo}}\left[\frac{G_n}{R_n} + \left(G_{nfo}+G_{cor}\right)^2\right] = 1 + \frac{R_n}{G_{nfo}}\left[2G_{nfo}^2 + 2G_{nfo}G_{cor}\right] \\ &= 1 + 2R_n\left(G_{cor}+G_{nfo}\right). \end{aligned} \tag{92}$$

Substituting this result back into Eq.(88), the noise figure with an arbitrary generator admittance is written

$$\begin{aligned} F &= 1 + \frac{R_n}{G_g}\left[\frac{G_n}{R_n} + G_g^2 + G_{cor}^2 + 2G_gG_{cor} + \left(B_g - B_{nfo}\right)^2 + 2G_gG_{nfo} - 2G_gG_{nfo}\right] \\ &= 1 + \frac{R_n}{G_g}\left[\,2G_gG_{cor} + 2G_gG_{nfo}\,\right] + \frac{R_n}{G_g}\left[\left(G_g - G_{nfo}\right)^2 + \left(B_g - B_{cor}\right)^2\right] \\ &= F_{min} + \frac{R_n}{G_g}\,|\,Y_g - Y_{nfo}\,|^2. \end{aligned} \tag{93}$$

Note, in this form four parameters are still needed to characterize the noise properties of the two-port, namely F_{min}, R_n, and the two components G_{nfo}, B_{nfo} of the generator admittance Y_{nfo} that minimizes the noise figure.

Constant Noise Figure Circles

Commonly the optimum generator admittance with respect to noise for a transistor differs from the choice that would be optimal regarding gain and stability considerations alone. The designer must compromise between the two concerns. An aid to decisions is to map contours of constant noise figures. They constitute a system of circles with centers in y_{gcn} and radii g_{cn} , where

$$y_{gcn} = Y_{nfo} + \frac{F - F_{min}}{2R_n}, \qquad g_{cn} = \sqrt{\left(\frac{F - F_{min}}{2R_n}\right)^2 + \frac{G_{nfo}}{R_n}\left(F - F_{min}\right)}\,. \tag{94}$$

These expressions are verified by observing that a circle of center y_{gcn} and radius g_{cn} in the complex Y_g-plane is described by

$$| Y_g - y_{gcn} |^2 = Y_g Y_g^* + y_{gcn} y_{gcn}^* - Y_g^* y_{gcn} - Y_g y_{gcn}^* = g_{cn}^2 . \tag{95}$$

Expanding the noise figure in Eq.(93) with the circle expression in mind, using $G_g = \frac{1}{2}(Y_g + Y_g^*)$, gives temporarily

$$\begin{aligned} \left| Y_g - Y_{nfo} \right|^2 &= Y_g Y_g^* + Y_{nfo} Y_{nfo}^* - Y_g Y_{nfo}^* - Y_g^* Y_{nfo} = \frac{F - F_{\min}}{2R_n} \left(Y_g + Y_g^* \right) \Rightarrow \\ Y_g Y_g^* &- Y_g \left(Y_{nfo}^* + \frac{F - F_{\min}}{2R_n} \right) - Y_g^* \left(Y_{nfo} + \frac{F - F_{\min}}{2R_n} \right) = - Y_{nfo} Y_{nfo}^* . \end{aligned} \tag{96}$$

By this step the center y_{gcn} in Eq.(94) is identified from the coefficients to either Y_g or Y_g^*. To complete the circle requirement and get the radius, we add $y_{gcn} y_{gcn}^*$ to both sides of the last equation, i.e.

$$\begin{aligned} g_{cn}^2 &= Y_g Y_g^* - Y_g y_{gcn}^* - Y_g^* y_{gcn} + y_{gcn} y_{gcn}^* = y_{gcn} y_{gcn}^* - Y_{nfo} Y_{nfo}^* \\ &= \left(Y_{nfo} + \frac{F - F_{\min}}{2R_n} \right) \left(Y_{nfo}^* + \frac{F - F_{\min}}{2R_n} \right) - Y_{nfo} Y_{nfo}^* = \left(\frac{F - F_{\min}}{2R_n} \right)^2 + \frac{G_{nfo}}{R_n} (F - F_{\min}) . \end{aligned} \tag{97}$$

The set of parameters containing the minimum noise figure F_{min}, the series noise equivalent resistance R_n, and the optimal generator admittance Y_{nfo}, or the equivalent sets based on the generator reflection coefficient, are becoming more and more common in device data sheets. The relationship between the two forms is straightforward to establish, albeit it needs a little algebra. Reflection coefficients for impedances and admittances were introduced in Chap.II sec.8. With reference impedance Z_0, which is real, the generator and optimum admittances are expressed by their reflection coefficients through

$$Y_g = \frac{1}{Z_0} \frac{1 - \Gamma_g}{1 + \Gamma_g} , \qquad Y_{nfo} = \frac{1}{Z_0} \frac{1 - \Gamma_{nfo}}{1 + \Gamma_{nfo}} \quad \Rightarrow \tag{98}$$

$$\begin{cases} \left| Y_g - Y_{nfo} \right|^2 = \dfrac{1}{Z_0^2} \left| \dfrac{(1 - \Gamma_g)(1 + \Gamma_{nfo}) - (1 - \Gamma_{nfo})(1 + \Gamma_g)}{(1 + \Gamma_g)(1 + \Gamma_{nfo})} \right|^2 = \dfrac{4}{Z_0^2} \dfrac{\left| \Gamma_g - \Gamma_{nfo} \right|^2}{\left| 1 + \Gamma_g \right|^2 \left| 1 + \Gamma_{nfo} \right|^2} , \\ G_g = \dfrac{\left| Y_g + Y_g^* \right|}{2} = \dfrac{1}{2Z_0} \left| \dfrac{(1 - \Gamma_g)(1 + \Gamma_g^*) + (1 - \Gamma_g^*)(1 + \Gamma_g)}{(1 + \Gamma_g)(1 + \Gamma_g^*)} \right| = \dfrac{1}{Z_0} \dfrac{1 - \left| \Gamma_g \right|^2}{\left| 1 + \Gamma_g \right|^2} . \end{cases} \tag{99}$$

Inserting into the last expression from Eq.(93) yields

$$F = F_{\min} + \frac{4R_n}{Z_0} \frac{\left| \Gamma_g - \Gamma_{nfo} \right|^2}{\left(1 + |\Gamma_g|^2 \right) \left| 1 + \Gamma_{nfo} \right|^2} \ . \tag{100}$$

Moving from admittance plane to reflection plane is represented by the Smith chart. The underlying transformation maps circles into circles, so contours of constant noise figures in an admittance plane become also a system of circles in a reflection plane or a Smith chart. To get the corresponding centers and radii we rewrite Eq.(100) to read

$$\Gamma_g \Gamma_g^* + \Gamma_{nfo} \Gamma_{nfo}^* - \Gamma_g \Gamma_{nfo}^* - \Gamma_g^* \Gamma_{nfo} = \left(1 + \Gamma_g \Gamma_g^* \right)^2 N \ , \qquad \textit{where} \quad N = \frac{F - F_{\min}}{4R_n / Z_0} \left| 1 + \Gamma_{nfo} \right|^2 \quad \Rightarrow \tag{101}$$

$$\Gamma_g \Gamma_g^* + \frac{\Gamma_{nfo} \Gamma_{nfo}^*}{1+N} - \Gamma_g \frac{\Gamma_{nfo}^*}{1+N} - \Gamma_g^* \frac{\Gamma_{nfo}}{1+N} = N + \frac{\Gamma_{nfo} \Gamma_{nfo}^*}{(1+N)^2} - \frac{\Gamma_{nfo} \Gamma_{nfo}^*}{(1+N)^2} \ . \tag{102}$$

Like the previous computation in Eq.(96), it is the Γ_g and Γ_g^* coefficients that determine the center of the circle. The radius is adjusted accordingly afterwards. Above in Eq.(102) this is anticipated by adding and subtracting the upcoming center term in absolutely squared form at the right hand side. Rearranging this equation gives the constant noise figure circles,

$$\left| \Gamma_g - \Gamma_{gcn} \right|^2 = \rho_{cn}^2 : \qquad \Gamma_{gcn} = \frac{\Gamma_{nfo}}{1+N} \ ,$$
$$\rho_{cn} = \sqrt{N + \frac{|\Gamma_{nfo}|^2}{(1+N)^2} - \frac{|\Gamma_{nfo}|^2}{1+N}} = \frac{\sqrt{N^2 + N\left(1 - |\Gamma_{nfo}|^2\right)}}{1+N} \ . \tag{103}$$

Here centers and radii are denoted Γ_{gcn} and ρ_{cn} respectively. The noise figures for the circles are expressed through the quantity N that was defined in Eq.(101).

Example IV-3-3 (noise figure circles in Y-plane)

Fig.35 shows constant noise figure circles for two collector bias currents in a generator admittance plane for a bipolar transistor, BFR91A. All centers of the circle system in each figure have equal imaginary parts, which is in agreement with the first equation in (94), where the second non-constant term is real.

Philips Semiconductors — Product specification

NPN 6 GHz wideband transistor — BFR91A

FEATURES

- Low noise
- Low intermodulation distortion
- High power gain
- Gold metallization.

DESCRIPTION

NPN transistor in a plastic SOT37 envelope primarily intended for use in RF wideband amplifiers.

A SOT54 (TO-92) version (ref: ON4185) is available on request.

PNP complement is BFQ23.

PINNING

PIN	DESCRIPTION
Code: BFR91A/02	
1	base
2	emitter
3	collector

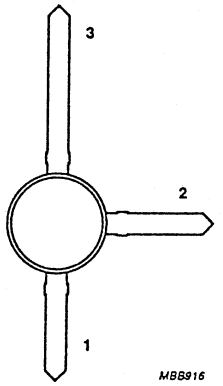
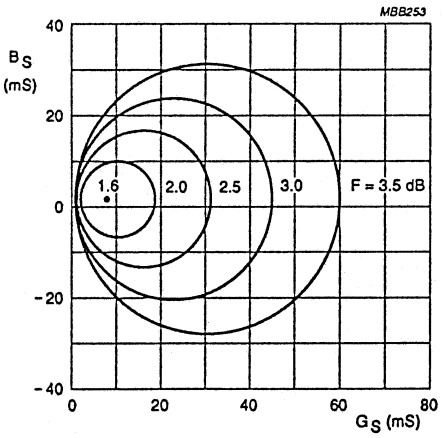
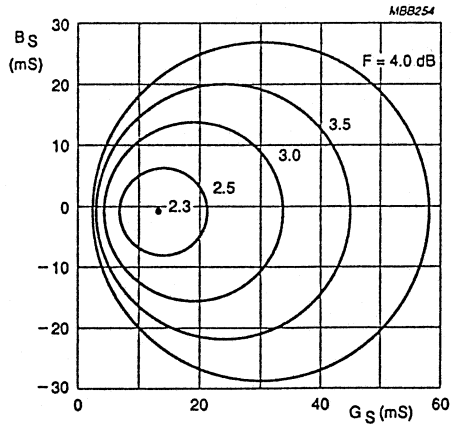


Fig.1 SOT37.



I_C = 4 mA; V_{CE} = 8 V; f = 800 MHz; T_{amb} = 25 °C.

Fig.12 Noise circle figure.



I_C = 30 mA; V_{CE} = 8 V; f = 800 MHz; T_{amb} = 25 °C.

Fig.13 Noise circle figure.

Fig.35 Extract from Philips Data Handbook, SC14 RF Wideband Transistors, Video Transistors and Modules, 1993.

In design considerations it is sometimes convenient to use the primary equivalent two-port input noise parameters R_n, G_n, and $Y_{cor}=G_{cor}+jB_{cor}$ from Fig.33(b). Calculating backwards from F_{min} and $Y_{nfo}=G_{nfo}+jB_{nfo}$, keeping R_n and using Eqs.(92),(91), provides

$$G_{cor} = (F_{min} - 1)/2R_n - G_{nfo}\ , \quad B_{cor} = -B_{nfo}\ , \quad G_n = R_n\,(G_{nfo}^2 - G_{cor}^2)\ . \tag{104}$$

J.Vidkjær

Table II Extraction of noise equivalent circuit parameters for BFR91A from data in Fig.35

I_C, V_{CE}	4 mA, 8 V	30 mA, 8 V
F_{min}	1.6 dB ~ 1.45	2.3 dB ~ 1.70
G_{nfo}	8.0 mS	13.5 mS
B_{nfo}	1.7 mS	-0.75 mS
g_{cn}	29.4 mS	28.1 mS
F @ g_{cn}	3.5 dB ~ 2.24	4.0 dB ~ 2.51
R_n	17.4 Ω	22.6 Ω
G_{cor}	4.96 mS	1.97 mS
B_{cor}	-1.7 mS	0.75 mS
G_n	0.68 mS	4.04 mS
$g_m/2\beta=I_C/0.026/180$	0.85 mS	6.41 mS
$G_{n,tot}=G_n+\|Y_{cor}\|^2R_n$	1.16 mS	4.14 mS

While F_{min} and Y_{cor} are directly recognized in the above figures, R_n must be calculated from a circle radius g_{cn} and the corresponding noise figure. By Eq.(94) we get

$$R_n = \frac{G_{nfo}(F - F_{min})}{2\,g_{cn}^2}\left[1 + \sqrt{1 + \left(\frac{g_{cn}}{G_{nfo}}\right)^2}\right]. \qquad (105)$$

Table II summarizes data that are deduced from the two systems of circles in Fig.35 by the expressions above. The first five rows contain direct readings and the next four hold parameters of the input equivalent circuit. The two bottom rows are included to illuminate the results where first the $g_m/2\beta$ value is the idealized value of G_n. To see this we recall from Eqs.(83),(84) that G_n controls an uncorrelated source, which according to the discussions on page 17 is the shot noise associated with the base current in the transistor. Using $I_B=I_C/\beta$ we get from Eq.(36)

$$\overline{i_{nb}^2} = 2\,q\,\frac{I_C}{\beta}\,\Delta f = 4\,k\,T\,G_n\,\Delta f \quad \Rightarrow \quad G_n = \frac{I_C}{2\,\beta\,V_t} = \frac{g_m}{2\,\beta} \qquad (106)$$

Table entries correspond to a typical β=90 value from the data sheets and V_t=26mV. They are relatively close to the extracted G_n´s. It is supposed that the observed overestimation in the

check values are caused by reductions of the actual transistor conductances due to high injections. The last row indicate the significance of correlation by showing the value of G_n that correspond to the total input noise shunt current. Especially in the low current case the difference and thereby the correlation is pronounced. This is in agreement with the fact that the transistor input impedance is relatively high at low currents where the input noise voltage source encompasses the effect of the collector noise current if the generator impedance is low.
Example IV-3-3 end

Example IV-3-4 (noise figure circles in Smith chart)

Fig.36 shows noise circles in Smith chart representing generator impedances. There are two system of circles. The closed ones at the right hand side are the noise figure circles corresponding to Eq.(103). The four corresponding noise parameters are also given. The other system includes circular arcs,[7] where the input impedance provide constant gain if the transistor is matched conjugatedly at the output port, i.e. curves of constant available power gain. Transforming to admittance or impedance form, the noise parameters are

$$F_{\min} = 2.[dB] \sim 1.58, \qquad R_n = 50 \cdot 2.6 = 130\ \Omega, \qquad \Gamma_{nfo} = 0.74 \angle 8° = 0.733 + j0.103 \Rightarrow \tag{107}$$

$$Y_{nfo} = \frac{1}{50}\,\frac{1-\Gamma_{nfo}}{1+\Gamma_{nfo}} = 3.00 - j\,1.37\ [mS] \qquad \sim \qquad Z_{nfo} = 1/Y_{nfo} = 276. + j126.\ [\Omega]\ .$$

At the particular bias and frequency, the transistor is potentially unstable and cannot be simultaneously matched at both ports with respect to power. This is the reason why the Smith chart is divided by a so-called stability circle. Generator impedances map into output impedances of negative real parts in the unstable region. The arc of highest gain, which is shown in the figure, is 11.2dB. It is called MSG and equals is the maximum stable gain $G_{ms}=|y_{21}/y_{12}|=|s_{21}/s_{12}|$ that was introduced in chap.III, p.13.

The Smith chart in the example shows that it is possible to construct a stable amplifier of gain equal to G_{ms} with a noise figure slightly below 3dB. Sacrificing gain, the optimum noise figure of 2dB may be achieved with an available power gain of approximately 8dB. To see the further consequences of such a choice we need small-signal parameters for the transistor. In the present case they are transformed from data sheet s-parameters to read

$$\boldsymbol{Y}_{tr} = \begin{Vmatrix} g_{11} + jb_{11} & g_{12} + jb_{12} \\ g_{21} + jb_{21} & g_{22} + jb_{22} \end{Vmatrix} = \begin{Vmatrix} 2.33 + j4.73\ mS & -0.014 - j2.09\ mS \\ 21.3 - j17.9\ mS & 0.154 + j3.46\ mS \end{Vmatrix}. \tag{108}$$

---

7) It is proven in most literature on RF amplifiers, for instance [8] or chap.III ref's [5], [6], that curves of constant power gain constitute a systems of circles in either Smith charts or impedance and admittance planes.

Philips Semiconductors — Product specification

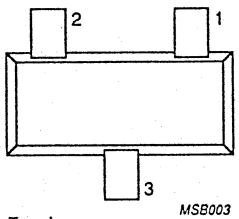
NPN 5 GHz wideband transistor — BFT25A

FEATURES

- Low current consumption (100 µA - 1 mA)
- Low noise figure
- Gold metallization ensures excellent reliability.

PINNING

PIN	DESCRIPTION
Code: V10	
1	base
2	emitter
3	collector



Top view

Fig.1 SOT23.

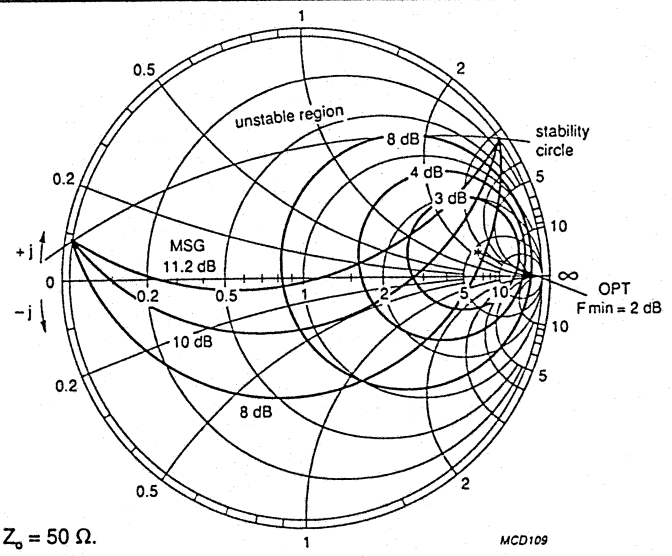
DESCRIPTION

The BFT25A is a silicon npn transistor, primarily intended for use in RF low power amplifiers, such as pocket telephones and paging systems with signal frequencies up to 2 GHz.

f (MHz)	V_{CE} (V)	I_C (mA)
1000	1	1

Noise Parameters

F_{min} (dB)	Gamma (opt)		$R_n/50$
	(mag)	(ang)	
2	0.74	8	2.6



Z_o = 50 Ω.

Average gain parameter: MSG = 11.2 dB.

Fig.13 Noise circle figure.

Fig.36 Extract from Philips Data Handbook, SC14 RF Wideband Transistors, Video Transistors and Modules, 1993.

Highest gain with minimum noise figure is achieved when the load y_L equals the complex conjugated of the transistor output admittance using the optimal noise figure generator admittance Y_{nfo} at the input port. We get

$$y_{12}y_{21} = (-0.014 - j2.09)(21.3 - j17.6) = -37.7 - j44.3\,[mS]^2 = 58.2\,[mS]^2 \angle -130.^\circ . \quad (109)$$

$$y_L^* = y_{out} = y_{22} - \frac{y_{12}y_{21}}{y_{11}+Y_{nfo}} = (0.154+j3.46) - \frac{(-37.7-j44.3)}{(2.23+j4.73)+(3.00-j1.37)} \qquad (110)$$

$$= 8.96 + j6.21\,[mS] \quad \sim \quad z_L = 1/y_L^* = 75.4 + j52.3\,[\Omega]\,.$$

The corresponding available gain and the stability factor in augmented estimation are

$$G_{av} = \left|\frac{y_{21}}{y_{11}+y_{nfo}}\right|^2 \frac{g_{nfo}}{g_{out}} = \frac{21.3^2+17.9^2}{(2.33+3.00)^2+(4.73-1.37)^2}\,\frac{3.00}{8.96} = 6.53 \sim 8.15\,[dB]\,, \qquad (111)$$

$$K_{aug} = \frac{2(g_{11}+g_{nfo})(g_{22}+g_L) - Re\{y_{12}y_{21}\}}{|y_{12}y_{21}|} = \frac{2(2.33+3.00)(0.154+8.96)+37.7}{58.2} = 2.32.$$

The gain becomes a little more than 8 dB in agreement with Fig.36, where the point of optimum noise figure is slightly above this gain arch. The K-factor is far below the ideal of five that makes a parameter insensitive amplifier, so one of the penalties of getting minimum noise figure with most gain from this transistor is that the resultant amplifier may be difficult to tune. Another drawback is that the input port becomes highly mismatched. The input admittance and impedance of the amplifier are

$$y_{in} = y_{11} - \frac{y_{12}y_{21}}{y_{22}+y_L} = (2.33+j4.73) - \frac{(-37.7-j44.3)}{(.154+j3.46)+(8.96-j6.21)} \qquad (112)$$

$$= 4.78 + j10.33\,[mS] \quad \sim \quad z_{in} = 1/y_{in} = 36.9 - j79.7\,[\Omega]\,,$$

which implies a mismatching of

$$M_{mch} = \frac{4\,G_{nfo}\,g_{in}}{|Y_{nfo}+y_{in}|^2} = \frac{4\cdot 3.00\cdot 4.78}{(3.00+4.78)^2+(-.1.37+10.3)^2} = 0.408\,, \qquad (113)$$

$$|\Gamma| = \sqrt{1-M_{mch}} = 0.770 \quad \Rightarrow \quad SWR = 7.68\,.$$

According to the previous discussion in chap.III,p.15 ff., the mismatch pertains at either side of any lossless input matching network. Therefore, it will also be present at the input port to any amplifier designed to minimum noise figure with most gain around this transistor. The mismatch is rather high and may be impractical. To lower it we may either reduce gain, sacrifice noise figure, or both. As an example, applying an ohmic load of 50Ω in parallel with a capacitor of 15.9pF (y_L=20mS+j10mS) provides gain - transducer gain as the output port is no longer matched - and stability factor that are calculated using

$$(y_{11}+y_{nfo})(y_{22}+y_L) = [(2.33+3.00)+j(4.73-1.37)][(0.154+20.0)+j(3.56+10)] \qquad (114)$$

$$= 62.2+j140.$$

$$G_{tr} = \left| \frac{y_{21}}{(y_{11}+y_{nfo})(y_{22}+y_L) - y_{12}y_{21}} \right|^2 4\,g_{nfo}\,g_L = \frac{(21.3^2+17.9^2)4\cdot 3.00\cdot 20.0}{(62.2+37.7)^2+(140.+44.3)^2}$$

$$= 4.25 \sim 6.21\,[dB], \quad (115)$$

$$K_{aug} = \frac{2(g_{11}+g_{nfo})(g_{22}+g_L) - Re\{y_{12}y_{21}\}}{|y_{12}y_{21}|} = \frac{2(2.33+3.00)(0.154+20.0)+37.7}{58.2} = 4.34.$$

The corresponding input admittance and mismatching properties are

$$y_{in} = y_{11} - \frac{y_{12}y_{21}}{y_{22}+y_L} = 2.33+j4.73 - \frac{-37.7-j44.3}{20.154+j13.46} = 4.64 + j5.38\,[mS]\ . \quad (116)$$

$$M_{mch} = \frac{4\,g_{nfo}\,g_{in}}{|y_{nfo}+y_{in}|^2} = \frac{4\cdot 3.00\cdot 4.64}{(3.00+4.64)^2+(-1.37+5.38)^2} = 0.748\ , \quad (117)$$

$$|\Gamma| = \sqrt{1-M_{mch}} = 0.502 \quad \Rightarrow \quad SWR = 3.01\ .$$

In this case we maintain the minimum noise figure and obtain an amplifier that has improved the input mismatch at the expense of gain. Another choice with the same ohmic load could be to optimize gain letting an input matching network provide conjugated matching to the input admittance from Eq.(116). In that case the input power equals the available power, so it is relevant to consider the operational power gain, which yields

$$G_p = \left| \frac{y_{21}}{y_{22}+y_L} \right|^2 \frac{g_L}{g_{in}} = \frac{21.3^2+17.9^2}{(0.154+20.0)^2+(3.46+10)^2}\ \frac{20.}{4.64} = 5.68 \sim 7.55\,[dB]\ , \quad (118)$$

$$K_{aug} = \frac{2(g_{11}+g_{in})(g_{22}+g_L) - Re\{y_{12}y_{21}\}}{|y_{12}y_{21}|} = \frac{2(2.33+4.64)(0.154+20.0)+37.7}{58.2} = 5.48.$$

To find the corresponding noise figure, Eq.(93) is employed with generator admittance equal to the complex conjugated input admittance, i.e.

$$F = F_{\min} + \frac{R_n}{g_{in}} |y_{in}^* - y_{nfo}|^2 = 1.58 + \frac{0.130}{4.64}\left[(4.64-3.00)^2+(5.38+1.37)^2\right]$$

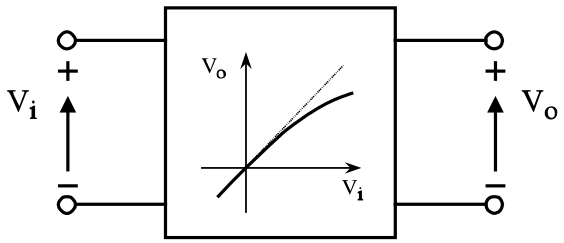
$$= 2.11 \sim 3.25\,[dB] \quad (119)$$

Clearly, the noise figure increases when the generator admittance differs from the optimal one, instead we achieve higher gain, better stability and no mismatch at the input port.

Example IV-3-4 end

IV-4 Distortion in Almost Linear Circuits

The type of distortion we consider in this section is the so-called nonlinear distortion, i.e. departures from linear input-output relationships of electronic circuits and components. Nonlinear distortion accompanies large drive signals and sets an upper bound on the operation level for proper performance of many intentionally linear circuits.



$$V_o = f(V_i) \ , \qquad (120)$$

Fig.37 Nonlinear two-port example, a voltage controlled voltage generator circuit.

To fix ideas of a circuit that eventually exhibits nonlinear distortion, the voltage controlled voltage source in Fig.37 is considered. It could be a DC-coupled voltage amplifier with a simple, memoryless input-output relationship like Eq.(120). Taken as a function of the entire input range in V_i, this general form is a large-signal expression, which is commonly not linear. Linear amplification is established in restricted input intervals around the DC biasing point and expressed mathematically by the first order term of a Taylor expansion in the deviation v_i from the bias V_{i0},

$$V_o = f(V_{i0} + v_i) = f(V_{i0}) + \left.\frac{df}{dV_i}\right|_{V_{i0}} v_i + \frac{1}{2}\left.\frac{d^2f}{dV_i^2}\right|_{V_{i0}} v_i^2 + \frac{1}{6}\left.\frac{d^3f}{dV_i^3}\right|_{V_{i0}} v_i^3 + \cdots \qquad (121)$$

Subdividing output into bias and a small-signal part yields

$$V_o = V_{o0} + v_o \ , \quad \begin{cases} \textit{bias term}: & V_{o0} = f(V_{i0}) \ , & (a) \\ \textit{small-signal terms}: & v_o = a\,v_i + b\,v_i^2 + c\,v_i^3 + \ldots & (b) \end{cases} \qquad (122)$$

where the a, b, c, and possibly more coefficients of the small-signal polynomial are shorthand notation for the scaled derivatives of the original, large-signal expression,

$$a = \left.\frac{df}{dV_i}\right|_{V_{i0}} , \qquad b = \frac{1}{2}\left.\frac{d^2f}{dV_i^2}\right|_{V_{i0}} , \qquad c = \frac{1}{6}\left.\frac{d^3f}{dV_i^3}\right|_{V_{i0}} . \qquad (123)$$

In an amplifier the a coefficient would be the small signal voltage gain and the higher order terms b and c introduce distortion. To keep distortion low we should select components with small second, third and even higher order expansion terms or hold the signal level low, so powers in v_i stay insignificant. The nonlinear expansion coefficients b, c, etc. refer to DC and

instant voltages or currents. If DC bias is unimportant, a polynomial expansion may express distortion directly in RMS scale, i.e.

$$v_{o,rms} = \frac{v_o}{\sqrt{2}} = a_{rms} v_{i,rms} + b_{rms} v_{i,rms}^2 + c_{rms} v_{i,rms}^3 ,$$
$$\text{where} \quad a_{rms} = a, \quad b_{rms} = \sqrt{2}\, b, \quad c_{rms} = 2c , \quad v_{i,rms} = \frac{v_i}{\sqrt{2}} \tag{124}$$

One of the most prominent consequences of the nonlinear distortion is the occurrence of deterministic signal components at frequencies that are not present in the input signal. No strictly linear circuits are able to do that and there are several technical utilizations of the property, mixing for instance. In this section we shall, however, concentrate on the different ways in which nonlinear effects may degrade the performance of RF communication circuits and on methods of quantifying this in data sheets.

A precautionary word on the validity of our results is in place. To let the response of a nonlinear circuit be represented by a Taylor series expansion requires that all nonlinearities combine memoryless. This will be the case with algebraic characteristics in ohmic networks, but typically not if the problem includes combinations of linear and nonlinear conductances and reactances. In such cases more powerful but also significantly more laborious methods are required to conduct genuine analytical treatments. Volterra series expansions - introduced in ref´s [21] or [22] chap.4 - are suggestions for that purpose. Using these methods we get expansion coefficients corresponding to the a, b, and c's above, which become frequency dependent. However, no more frequency components arise with combined nonlinearities compared to the ones that are demonstrated by Taylor series expansions, which therefore suffice for the present purpose of exploring basic matters. They also suffice in design if the polynomial coefficients are deduced from data that apply to the actual operating frequencies.

Simplifying concerns are also the reason why expansions mostly are limited to encompass third order terms only. Going higher would provide no really new phenomena but a lot more messy algebra.

Single Signal Distortions

Fig.38 shows the single sided frequency components in the amplifier voltages when a single tone input signal drives the circuit that was described by Eq.(122),

$$v_o = a v_i + b v_i^2 + c v_i^3 + \ldots \qquad \text{where} \qquad v_i = A \cos \omega_a t . \tag{125}$$

The frequency components in the output emerge when the trigonometric identities,

$$\cos^2 x = \frac{1}{2} + \frac{1}{2} \cos 2x , \qquad \cos^3 x = \frac{3}{4} \cos x + \frac{1}{4} \cos 3x , \tag{126}$$

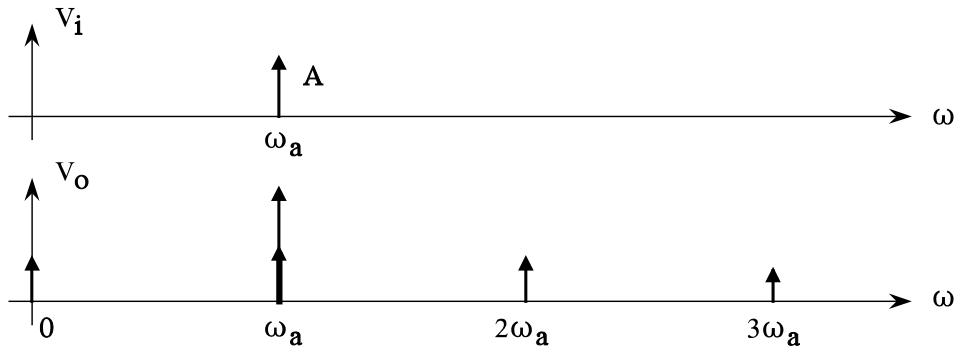


Fig.38 Signal and distortion components when a single tone input signal is applied to a circuit which includes nonlinear components.

are applied to the polynomial input-output relationship. The different terms become

Taylor terms	v_o *components*	
1*st order* :	$aA\cos\omega_a t$	(a)
2*nd order* :	$+\frac{1}{2}bA^2 + \frac{1}{2}bA^2\cos 2\omega_a t$	(b)
3*rd order* :	$+\frac{3}{4}cA^3\cos\omega_a t + \frac{1}{4}cA^3\cos 3\omega_a t$	(c)

(127)

The first order term is the conventional linear small-signal gain term. The second order term in the Taylor expansion causes a DC contribution and a second harmonic component. Care must be taken, both in circuit analysis and in practical design, to ensure that proper biasing of the component persists even when the quadratic term adds to the DC voltage or current. The third order term produces both a component at the input frequency and a third harmonic component. It is possible in many application to discard the effects of higher harmonic components by filtering. But it is impossible to remove the distortion component from the signal at the fundamental frequency, so the third order term will always cause distortion. Therefore, third order terms often draw most attention in specifications and data sheets. Actually all odd ordered Taylor expansion terms would contribute to the fundamental signal component, but their significance decrease rapidly with raising order in electronic devices that are intended for linear operation and driven accordingly, so we stay with the third order maximum below.

Including all frequency components, except DC, gives a measure called the total harmonic distortion, K_{THD}, in an amplifier. It includes all components, i.e.

$$V_{o,1} = aA + \frac{3}{4}cA^3 , \qquad V_{o,2} = \frac{1}{2}bA^2 , \qquad V_{o,3} = \frac{1}{4}cA^3 . \tag{128}$$

$$K_{THD} = \sqrt{\frac{V_{o,2}^2 + V_{o,3}^2 + \cdots}{V_{o,1}^2}} = \sqrt{K_2^2 + K_3^2 + \ldots} , \quad \textit{where} \tag{129}$$

$$K_2 = \frac{V_{o,2}}{V_{o,1}} \approx \frac{1}{2}\left|\frac{b}{a}\right| A , \qquad K_3 = \frac{V_{o,3}}{V_{o,1}} \approx \frac{1}{4}\left|\frac{c}{a}\right| A^2 , \quad \ldots .$$

The approximations in the distinct second, third, or possible higher harmonic terms like K_2 and K_3 require that the third order term at the signal frequency is small compared to the first order term, i.e. ¾c/a $A^2 \ll 1$. The assumption is necessary for calculating the total distortion from experimental data, because it is difficult to distinguish between fundamental frequency outputs arising from the first and the third order terms in measurements.

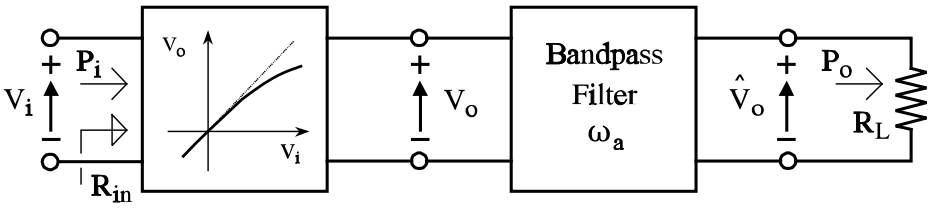


Fig.39 Amplifier or other electronic component followed by a bandpass filter that removes all but signal components around the fundamental frequency.

Removing all but the signal frequency components from the output, as indicated by the setup of Fig.39, leaves the first third order term in Eq.(127)(c) as the only disturbing component in observation. Fig.40 shows a typical shape of the input to filtered output characteristics of amplifiers or transistors, where the output signal falls below the linear extrapolation at higher signal levels. This is called compression and may be characterized by the RMS input drive level, $V_{i,1dB}$, where the compression corresponds to 1dB output signal reduction or, equivalently, an output amplitude reduction by at factor of 0.891 compared to the assumed linear value. In terms of the Taylor series expansion, the compression level becomes

$$\hat{V}_o \Big|_{\omega_a} = aA + \frac{3}{4}cA^3 \quad \underset{1dB\,cmprss.}{\Rightarrow} \quad \frac{3}{4}|c|A_{1dB}^3 = 0.109\,|a|\,A_{1dB} \quad \Rightarrow$$

$$\underline{1\ dB\ compression}: \qquad V_{i,1dB} = \frac{A_{1dB}}{\sqrt{2}} = 0.269\sqrt{\left|\frac{a}{c}\right|}\ \left[V_{rms}\right] . \tag{130}$$

Absolute values of the Taylor coefficients are required above to account for the fact that *a* and *c* must be of opposite signs if the resultant output amplitude falls below the linear term. The result is scaled to RMS values, that are common in data sheets and instrument readings.

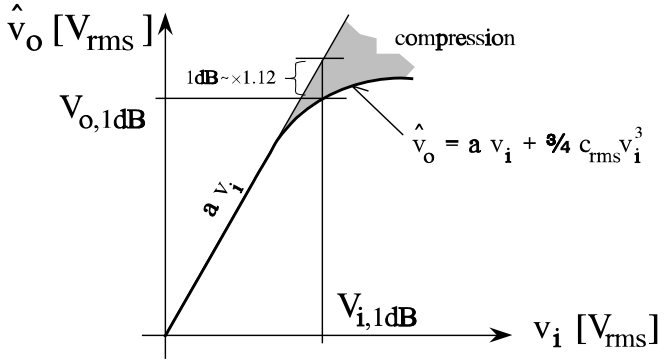


Fig.40 Amplifier characteristics with compression. Coefficients a and c_{rms} must be of opposite signs. Input and output hold fundamental frequency components only.

Instead of referring to the input side, compression specifications refer typically to the output side by giving an output power level and a gain, either the stipulated linear G_{lin} or the actual gain at 1dB compression, G_{1dB}. Backtracking from powers to voltages requires that input and output impedances are defined. Dealing with amplifier blocks this is commonly not a problem whereas transistor data often require careful review before interpretation. For the setup in Fig.39 the powers and the power gains are

$$P_i = \frac{V_{i,rms}^2}{R_{in}}\ , \qquad P_o = \frac{V_{o,rms}^2}{R_L} = G_{lin}\, P_i \quad \Rightarrow \quad G_{lin} = a^2\, \frac{R_{in}}{R_L}\ . \tag{131}$$

Absolute power indications may by in dBm, decibels relative to 1mW. Translated to dB scales, the linear region of the input-output plot in Fig.41 shows

$$P_i[dBm] = 10 \log_{10}\left[\, V_{i,rms}^2 / R_{in}\, 10^{-3} \,\right] = 20 \log_{10} V_{i,rms} - 10 \log_{10} R_{in} + 30\ , \tag{132}$$

$$P_o[dBm] = P_i[dBm] + G_{lin}[dB]\ , \qquad \textit{where} \qquad G_{lin}[dB] = 10 \log_{10}\left[\, a^2 R_{in}/R_L \,\right]\ .$$

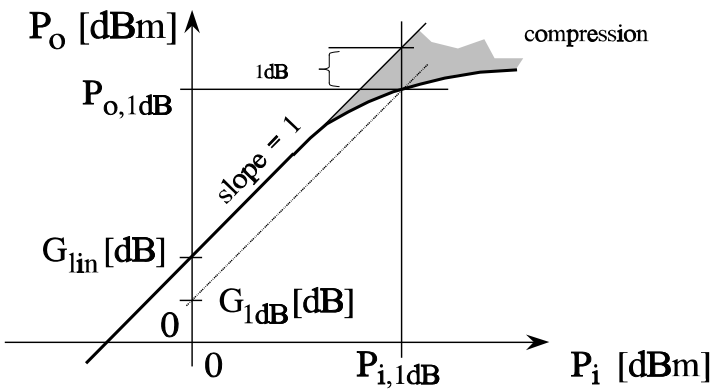


Fig.41 Amplifier characteristics in dB scale. Gains are represented by lines of slope equal to one with values given by the vertical positions.

Inserting the voltage for 1dB compression from Eq.(130), the corresponding input and output powers are expressed,

$$P_{i,1dB}[dBm] = 10\log_{10}[V_{i,1dB}^2/R_{in}10^{-3}] = 10\log_{10}|a/c| - 10\log_{10}R_{in} + 18.60, \quad (133)$$

$$P_{o,1dB}[dBm] = P_{i,1dB}[dBm] + G_{lin}[dB] - 1 = 10\log_{10}|a/c| + 20\log_{10}a - 10\log_{10}R_L + 17.60.$$

Distortions with Two Signals

Applying two sinusoids of sufficient amplitudes to an amplifier that eventually becomes nonlinear may produce a multitude of frequency components in the output spectrum. With two tones input to the Taylor expansion,

$$v_o = a\,v_i + b\,v_i^2 + c\,v_i^3 + \ldots \qquad where \qquad v_i = A\cos\omega_a t + B\cos\omega_b t, \quad (134)$$

we get all the components that are shown below in Fig.42 and Eq.(136). They are calculated from the trigonometric identities in Eq.(126) with supplement of the product formula,

$$\cos x\cos y = \frac{1}{2}\cos(x+y) + \frac{1}{2}\cos(x-y)\ . \quad (135)$$

For illustrating purposes the two input signal components in Fig.42 are chosen with close frequency spacing, which is an important situation in practice. In that case we still refer to the fundamental input frequency in the meaning of the range around ω_a and ω_b. Regarding output components in Eq.(136), the first observation is that all terms generated from a single tone in Eq.(127) are present with unaltered powers of amplitudes for both the ω_a and the ω_b input components. It is also observed that the second order Taylor term causes components around DC and around twice the fundamental input frequency. This is equivalent to the single tone case. There is also resemblance in the third order term, which produces a complex of components in the fundamental frequency range and a complex around third harmonics.

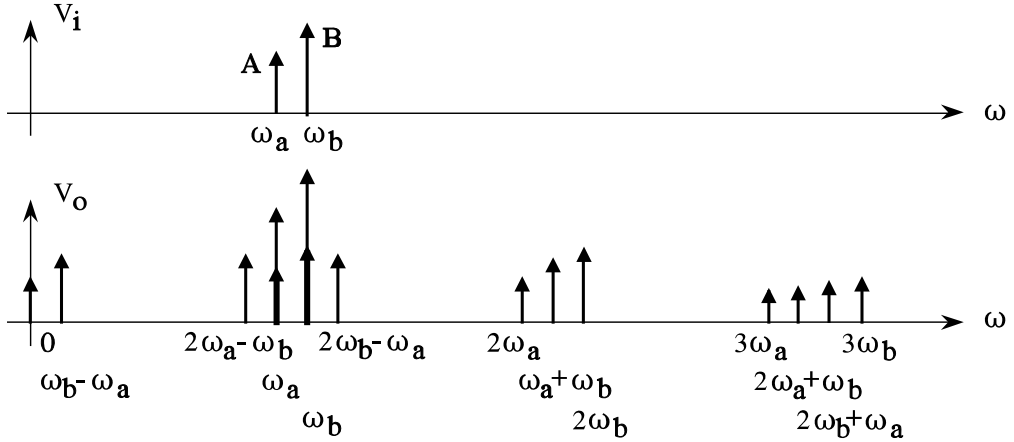


Fig.42 Output signal and distortion components with two-tone signal input to a nonlinear component.

Taylor terms	v_o *components*	
1*st order* :	$aA\cos\omega_a t + aB\cos\omega_b t$	(a)
2*nd order* :	$+\frac{1}{2}bA^2 + \frac{1}{2}bB^2$	$(b1)$
	$+\frac{1}{2}bA^2\cos 2\omega_a t + \frac{1}{2}bB^2\cos 2\omega_b t$	$(b1)$
	$+ bAB\cos(\omega_a - \omega_b)t + bAB\cos(\omega_a + \omega_b)t$	$(b2)$ (136)
3*rd order* :	$+\left[\frac{3}{4}cA^3 + \frac{3}{2}cAB^2\right]\cos\omega_a t + \left[\frac{3}{4}cB^3 + \frac{3}{2}cA^2B\right]\cos\omega_b t$	$(c1)$
	$+\frac{1}{4}cA^3\cos 3\omega_a t + \frac{1}{4}cB^3\cos 3\omega_b t$	$(c2)$
	$+\frac{3}{4}cA^2B\cos(2\omega_a + \omega_b)t + \frac{3}{4}cAB^2\cos(\omega_a + 2\omega_b)t$	$(c3)$
	$+\frac{3}{4}cA^2B\cos(2\omega_a - \omega_b)t + \frac{3}{4}cAB^2\cos(\omega_a - 2\omega_b)t$	$(c4)$

News with two tones are the different possibilities for interaction between the two. The second order terms in Eq.(136)(b2) and the third order terms in (c3) and (c4) contain so-called intermodulation products, where especially the last ones may cause troubles if the resultant frequencies hit the desired fundamental frequency range. Moreover, there is a direct cross-coupling between the two signals at their fundamental frequencies through the third order distortions terms in Eq.(136)(c1). The latter have great practical importance in radio receiver design causing blocking and cross-modulation effects, which we shall consider below. Keep in mind here that the cosine functions are indifferent with respect to sign of the arguments, so in comparisons the terms follow signs of the coefficients a, b, and c.

Blocking is a condition that occurs when two signals are present at the input to an amplifier which has a narrow channel selecting bandpass filter in the output port. As sketched in Fig.43(a), the input holds a week signal, which is the one we are supposed to receive, so its frequency ω_a lies in the passband. The other signal is much stronger. Albeit close in frequency to the desired signal, its frequency ω_b falls outside the passband of the filter. The total output from the passband gets the amplitude,

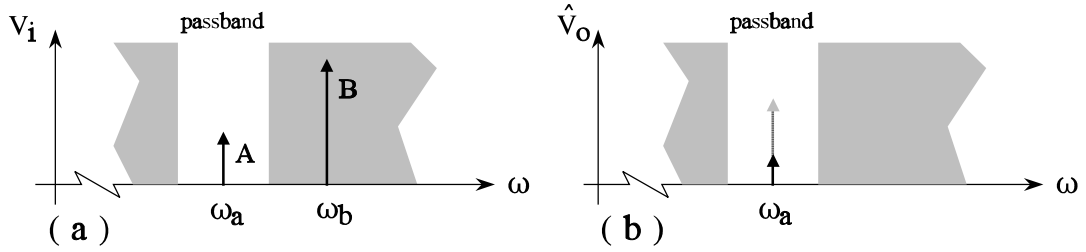


Fig.43 Input and output components illustrating signal desensitization towards blocking.

$$\hat{V}_o\Big|_{\omega_a} = a\,A + \frac{3}{4}\,c\,A^3 + \frac{3}{2}\,c\,A\,B^2 \underset{small\ A}{\approx} a\,A\left(1 - \frac{3}{2}\left|\frac{c}{a}\right|B^2\right). \qquad (137)$$

It is assumed that the A signal is so small, that its own third order distortion term may be ignored. When it is also supposed that the nonlinearity is of compression type, so the a and *c* coefficients have opposite signs, the last expression clearly shows that the undisirable B signal controls the size of the received signal. Blocking is the condition where the received signal disappears, i.e. $\hat{V}_o = 0$ in Eq.(137). The corresponding RMS value of the B signal is

$$\underline{Blocking}: \quad B_{block} = \sqrt{\frac{2}{3}\left|\frac{a}{c}\right|} \quad \Rightarrow \quad V_{block} = \frac{B_{block}}{\sqrt{2}} = 0.577\sqrt{\left|\frac{a}{c}\right|} \quad \left[V_{rms}\right]. \qquad (138)$$

The situation where the received signal still is present, but significantly reduced by the neighboring signal - like Fig.43(b) - is sometime called desensitization.

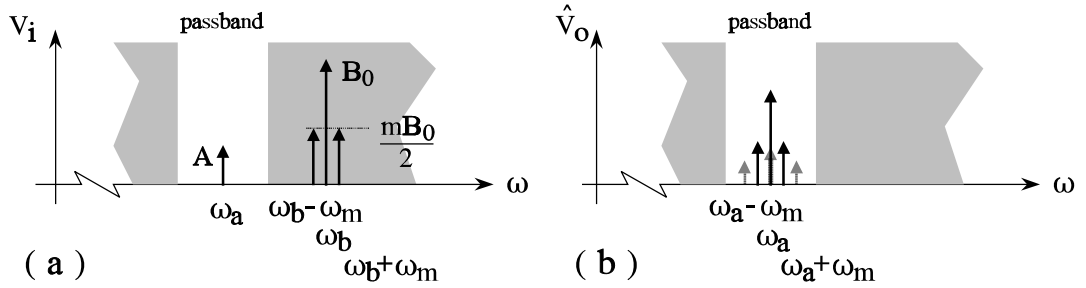


Fig.44 Input and output components illustrating cross-modulation. The selected ω_a signal is inflicted by the modulation of a neighboring but undesired channel around ω_b.

Cross modulation is the situation where the modulation in a large out-of-band signal transfers to a smaller signal in reception. For clarity the latter signal is represented by the unmodulated carrier at frequency ω_a. We suppose that the B signal is amplitude modulated by a low frequency sinusoid at ω_m with modulation index m, which is smaller than one,

$$B = B_0\,(1 + m\cos\omega_m t)\,. \qquad (139)$$

The squared value of disturbing signal is approximated,

$$B^2 = B_o^2 \left(1 + \frac{m^2}{2} + 2m\cos\omega_m t + \frac{m^2}{2}\cos 2\omega_m t \right) \underset{m\ll 1}{\approx} B_0^2 \left(1 + 2m\cos\omega_m t \right). \quad (140)$$

where it is assumed again that the signal to be received is small enough to let its own distortion be ignored on insertion into Eq.(136)(c1). The resultant passband output gets thereby a form that clearly shows transfer of modulation from the B signal to the A signal with a resultant modulation index m̂,

$$V_o\Big|_{passb.} = aA + \frac{3}{2}cAB^2 = aA\left(1 + 3mB_0^2\frac{c}{a}\cos\omega_m t \right) = aA\left(1 + \hat{m}\cos\omega_m t \right). \quad (141)$$

A common way of specifying cross modulation is to give the RMS voltage of the disturbing out-of-band carrier, which corresponds to 1% overcoupled modulation, i.e.

$$\frac{\hat{m}}{m} = \frac{1}{100} \quad \Rightarrow \quad B_{0,cm}^2 = \frac{a}{300\,c}, \quad (142)$$

$$\underline{1\%\,cross-modulation}: \qquad V_{1\%\,cm} = \frac{B_{0,cm}}{\sqrt{2}} = 0.0408\sqrt{\left|\frac{a}{c}\right|}\ \left[V_{rms}\right]. \quad (143)$$

Neighboring input signals interacts through the third order terms since they can produce frequency components close or equal to the original ones. Interaction between widely separated frequency components may occur by other terms. An example is interaction between a RF signal of frequency ω_a and a LF signal at ω_b, which is illustrated in Fig.45. Examples are low-frequency components from insufficient smoothing and decoupling of the bias in sensitive amplifiers. The second order terms in Eq.(136)(b2) may be rejoined to give,

$$v_o\Big|_{passb.} = aA\cos\omega_a t + 2bAB\cos\omega_a t\cos\omega_b t = aA\cos\omega_a t\left(1 + 2\frac{b}{a}B\cos\omega_b t \right). \quad (144)$$

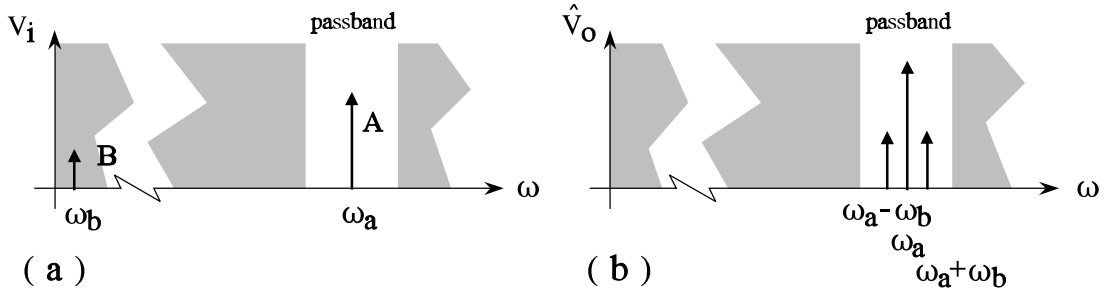


Fig.45 Input and output components illustrating hum-modulation. The low-frequency ω_b signal translates to an amplitude modulation of the RF signal at ω_a.

The last expression demonstrates that the LF signal amplitude modulates the RF signal with an AM modulation index given by

$$\underline{Hum-modulation}: \qquad \hat{m} = 2\left|\frac{b}{a}\right| B \ . \tag{145}$$

This phenomenon is sometimes called hum-modulation because it could be heard in many older receivers where it stemmed from the AC supply.

Common in all the examples above is that a signal in the passband and a signal outside enter a device that is nonlinear at large drive levels. The interaction leaves distortion components that cannot be removed by subsequent filtering. To anticipate the thought that we could apply filtering at the input side and avoid all the troubles, it should be recalled that in tunable receivers, bandpass filtering is done in the IF section after the RF signal is transferred to the intermediate frequency by a mixer, cf. Chap.I, Sec.5. In such receivers, the passband indications above must be interpreted as the effect of IF filtering transferred in frequency to the RF band. It should also be mentioned at this point that mixers introduce distortion defects like compression and cross-modulation. They are characterized and quantified by the same concepts we are introducing here for components and amplifiers, the only exception being that input and output fundamental frequencies do not coincide but differ by the frequency of the local oscillator. We shall return to the matters below when mixers are considered explicitly.

Intercept Points

Distortion specifications may be directly related to the practical consequences of intermodulation, for instance by cross-modulation as exemplified by Eq.(143). Recall, however, that all the implications of distortion we have considered express basically the absolute ratio of a higher order expansion term over the fundamental frequency term. A simple and easily interpretable way to get this information experimentally is by so-called intercept points. Below we start concentrating on the most important third order distortion and the associated 3rd order interception point. Afterwards the less frequent second order points, which develop equivalently, cf.[23] chap.10, is summarized briefly.

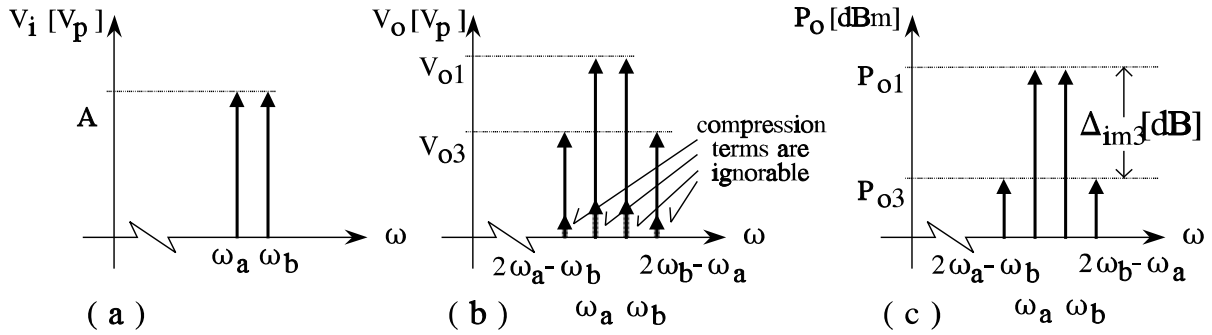


Fig.46 Third order intercept point conditions, (a) equal inputs, sufficiently low to ignore (b) compression terms in outputs. Experimental data are often dBm scaled like (c).

Intercept points are recorded from two-tone setups where the two input amplitudes are equal. Using Eq.(136), B is set equal to A while the frequencies ω_a and ω_b still differ. A picture of this situation preparing third order intercept point determination is shown in Fig.46. The two output components corresponding to the two-tone input are equal if ω_a is sufficiently close to ω_b. We call the corresponding output amplitude V_{o1}. The two intermodulation products at nearby frequencies $2\omega_a$-ω_b and $2\omega_b$-ω_a are also practically equal in size each of amplitude V_{o3}. With common input amplitudes, V_{i1}=A=B, one linear gain term from Eq.(136)(a) and one intermodulation term from Eq.(136)(c4) are

$$V_{o1} = aA \ , \qquad V_{o3} = \frac{3}{4} c \, A^3 \ . \tag{146}$$

Corresponding powers of input and output components with input and load resistances R_{in} and R_L respectively

$$P_i = \frac{A^2}{2 R_{in}}, \qquad P_{o1} = \frac{V_{01}^2}{2 R_L} = a^2 \frac{R_{in}}{R_L} P_i, \qquad P_{o3} = \frac{V_{03}^2}{2 R_L} = \left(\frac{3}{2} c \right)^2 \frac{R_{in}^3}{R_L} P_i^3 . \tag{147}$$

Considered as function of input power P_{in}, the intermodulation term P_{o3} rises much faster than the linear term P_{o1}. Their ratio, often called the intermodulation ratio or - in dB scale - the intermodulation distortion, IMD, is

$$\Delta_{im3}(P_i) = \frac{P_{o1}}{P_{o3}} = \frac{4}{9 R_{in}^2} \left(\frac{a}{c} \right)^2 \frac{1}{P_i^2} \ . \tag{148}$$

The third order intercept point is the stipulated crossing point between P_{o1} and P_{o3} where the intermodulation ratio is one, and input and output powers $P_{i,IP3}$, $P_{o,IP3}$ are

$$\Delta_{im3}(P_{i,IP3}) = 1: \qquad P_{i,IP3} = \frac{2}{3 R_{in}} \left| \frac{a}{c} \right| , \qquad P_{o,IP3} = \frac{2 a^2}{3 R_L} \left| \frac{a}{c} \right| . \tag{149}$$

Translating the above results to dB scales provide a simple and easily interpretable graphical presentation as seen in Fig.47 where,

$$\begin{aligned} P_{o1}[dBm] &= P_i[dBm] + 10 \log_{10}\left[a^2 R_{in} / R_L \right] , \\ P_{o3}[dBm] &= 3 \, P_i[dBm] + 10 \log_{10}\left[9 c^2 R_{in}^3 / 4 R_L \right] . \end{aligned} \tag{150}$$

When assumptions of no compressions are met, the fundamental frequency term P_{o1}[dBm] grows in direct proportion to P_i[dBm] - slope equal to one - while the third order intermodulation term P_{o3}[dBm] grows linearly, but faster with a slope of three. Above certain drive levels, the assumptions are no longer met and experimental data start falling below the assumptions. P_{o1} is lowered by compression as described formerly on page 50 while lowering

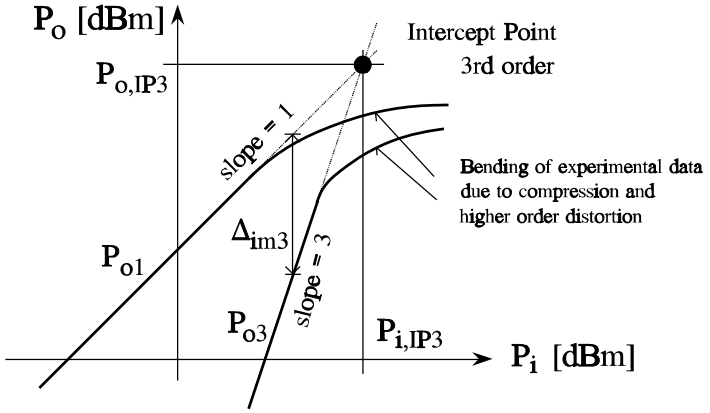


Fig.47 Third order intercept point: Crossing of a fundamental frequency output P_{o1} and a 3rd order intermodulation product, P_{o3}, both in straight line stipulations.

of P_{o3} is caused by odd-numbered higher order terms which we commonly disregard under small signal excitations. Dealing with experimental data that bend as sketched in the figure, it is still easy to draw linear extrapolations and find the intercept point. In dB scale we get

$$\begin{aligned} P_{i,IP3}[dBm] &= 10\log_{10}|a/c| - 10\log_{10}R_{in} + 28.24 \\ P_{o,IP3}[dBm] &= 10\log_{10}|a/c| + 20\log_{10}a - 10\log_{10}R_L + 28.24 \end{aligned} \tag{151}$$

Comparing the last result with the 1dB compression output power, $P_{o,1dB}$ from Eq.(133), we must expect that under similar conditions, i.e. same frequency, same bias i.e. same a and c coefficients, and same load impedances, the output power at the intercept points should lie 10.6 dB above the 1dB compression level, if a simple Taylor expansion is quantitatively adequate for the circuit in consideration[8].

The second order intercept point concerns the intermodulation product of frequency equal to either the sum or the difference of the two input signals, i.e. the first or the second term in Eq.(136)(b2). In analogy with Eqs.(146) to (148), keeping in mind that the input power expressions for P_i and P_{01} stay unaltered from the third order case, we get,

$$V_{o1} = aA\ , \qquad V_{o2} = b\,A^2 \qquad \Rightarrow \qquad P_{02} = 2b^2\,\frac{R_{in}^2}{R_L}\,P_{in}^2\ . \tag{152}$$

$$\Delta_{im2}(P_i) = \frac{P_{o1}}{P_{o2}} = \frac{1}{2R_{in}}\left(\frac{a}{b}\right)^2\frac{1}{P_i} \quad \Rightarrow \quad P_{i,IP2} = \frac{1}{2R_{in}}\left(\frac{a}{b}\right)^2, \quad P_{o,IP2} = \frac{a^2}{2R_L}\left(\frac{a}{b}\right)^2 \tag{153}$$

---

8) A common violation of the assumption occurs when the output swing is limited by power supply, so the amplifier or device is driven into heavy saturation and/or cut-off. Taylor expansions with three terms are not adequate for characterizing such abrupt events that might include relaxation effects.

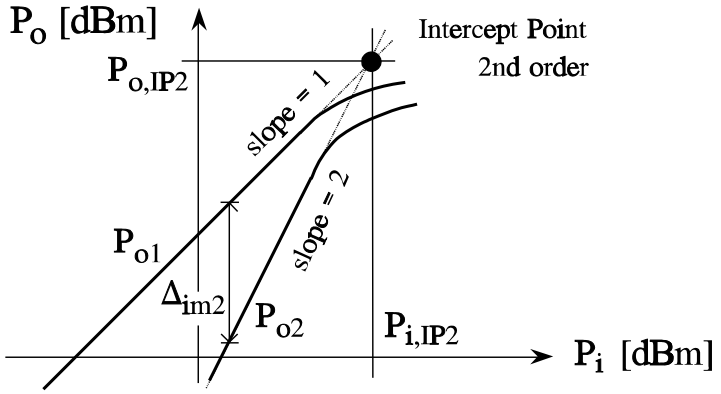


Fig.48 Second order intercept point: Crossing of a fundamental frequency output P_{o1} and a 2nd order intermodulation product, P_{o2}, both in straight line stipulations.

The dB-scaled asymptotes that determines the second order intercept point as shown in Fig.48 are given by

$$\begin{aligned} P_{o1}[dBm] &= P_i[dBm] + 10\log_{10}\left[a^2 R_{in}/R_L \right], \\ P_{o2}[dBm] &= 2\,P_i[dBm] + 10\log_{10}\left[2\,b^2 R_{in}^2/R_L \right], \end{aligned} \tag{154}$$

and the intercept point gets the coordinates

$$\begin{aligned} P_{i,IP2}[dBm] &= 20\log_{10}|a/b| - 10\log_{10}R_{in} + 26.99, \\ P_{o,IP2}[dBm] &= 20\log_{10}|a/b| + 20\log_{10}a - 10\log_{10}R_L + 26.99. \end{aligned} \tag{155}$$

Second order intermodulation products occur at the sum and the difference frequency of the incoming signal. The distance in frequencies between the signal components might be large, so if the a and b Taylor expansion coefficients are frequency dependent it is important to ensure that the second order specifications applies to the problem at hand.

Sensitivity and Dynamic Range

Noise determines a lower boundary for the signal level that may be processed successfully in a communication system. In complete systems it is often the first amplifier stage in the receiver that limits performance. The minimum required signal-to-noise ratio at the amplifier output is called $SNR_{o,mds}$, where mds stands for minimum discernible signal. Its level depends on system architecture including data rates, modulation, coding, interleaving schemes etc. The corresponding minimum input signal to the amplifier, $P_{i,mds}$, is sometimes called the sensitivity. Besides the required signal-to-noise ratio, it depends on noise bandwidth B_N Hz, on noise temperature of the source impedance T_g, and on noise temperature T_n of the amplifier, equivalently on the noise figure $F=1+T_n/T_0$ from Eq.(64). The level is

$$\frac{P_{o,mds}}{N_o} = \frac{G_{av} P_{i,mds}}{G_{av} k (T_g + T_n) B_N} = SNR_{o,mds} \Rightarrow$$
$$P_{i,mds} = SNR_{o,mds} k (T_g + T_n) B_N = SNR_{o,mds} k T_0 \left(F - 1 + \frac{T_g}{T_0} \right) B_N , \quad (156)$$

where the temporarily included G_{av} is the available power gain of the amplifier. If there is no association to specific system requirements, the acceptable output signal level from an amplifier may be taken equal to its noise output, i.e. using $SNR_{o,mds}=1$. If, furthermore, the source impedance is supposed to be at the reference temperature, $T_g=T_0=290$ K, the minimum discernible level gets a simpler form that sometimes is seen in literature,

$$P_{i,mds0} = P_{i,mds} \Big|_{T_g = T_0,\, SNR_o = 1} = k T_0 F B_N . \quad (157)$$

The dynamic range of an amplifier indicates the span of useful input levels. It is commonly the distance - in decibels - from the minimum discernible signal to the upper bound. Several upper bound criteria are possible, a simple one being the so-called spurious free dynamic range. The upper bound is the input level where the third order intermodulation term P_{o3} passes the noise floor at the amplifier output. The noise floor is taken as the output corresponding to the minimum discernible signal from Eq.(156). Triangle considerations in Fig.49 provide a simple expression for the spurious-free dynamic range,

$$D_{spf}[dB] = P_{i,spf}[dBm] - P_{i,mds}[dBm] = \frac{2}{3} \left\{ P_{i,IP3}[dBm] - P_{i,mds}[dBm] \right\} , \quad (158)$$

where third order intermodulation is parameterized through the third order intercept point input level $P_{i,IP3}$.

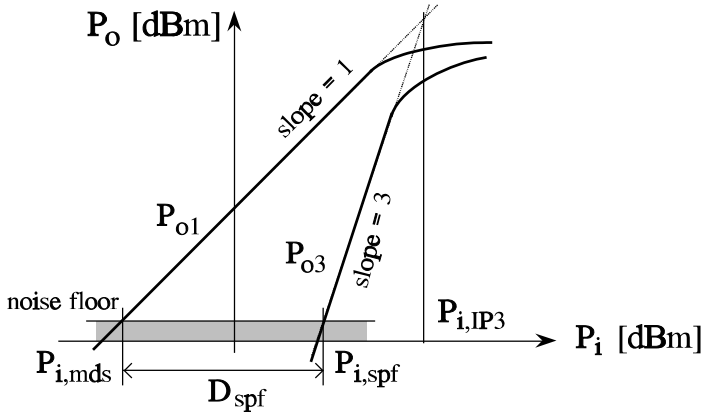


Fig.49 Spurious-free dynamic range, D_{spf} dB, the distance in input power levels where the third order intermodulation and the signal exceed the noise floor respectively.

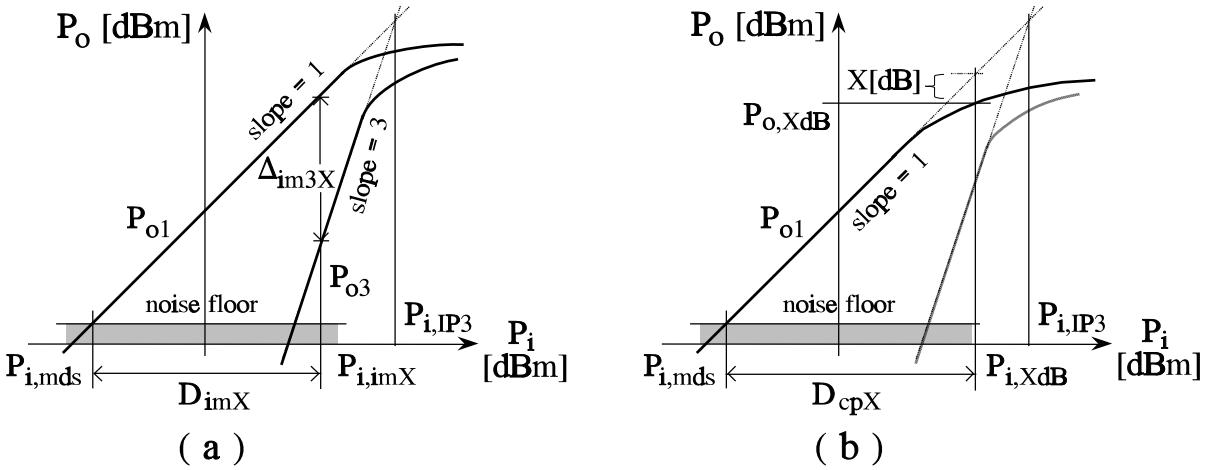


Fig.50 Dynamic range specifications with upper bound set by (a) intermodulation product or (b) compression.

Alternative dynamic range specifications use upper bound input that gives a minimum intermodulation ratio, for instance 40dB, or a certain amount of gain compression, for instance .1 or 1 dB. These two cases are summarized by Fig.50. To express the intermodulation bounded dynamic range, D_{imX}, Eqs.(148),(149) are combined to give the input power for a required intermodulation ratio Δ_{im3X} in terms of the third order intercept point,

$$\Delta_{im3} = \frac{P_{i,IP3}^2}{P_i^2} \qquad \Rightarrow \qquad P_{i,im3X} = \frac{P_{i,IP3}}{\sqrt{\Delta_{im3X}}} \; . \tag{159}$$

Now the intermodulation bounded dynamic range is expressed

$$D_{imX}[dB] = 10 \log \left(\frac{P_{i,imx}}{P_{i,mds}} \right) = P_{i,IP3}[dBm] - \frac{1}{2} \Delta_{im3X}[dB] - P_{i,mds}[dBm] \; . \tag{160}$$

Compression is commonly specified by a single tone test and expressed separately. When the compression levels of specification equals the upper dynamic range bound, the corresponding dynamic is given by

$$D_{cpX}[dB] = P_{i,XdB}[dBm] - P_{i,mds}[dBm] = P_{o,XdB}[dBm] - G_{lin}[dB] + 1 - P_{i,mds}[dBm] \; . \tag{161}$$

The last expression uses data that refer to the output side, i.e. the output power level at XdB compression and the linear gain G_{lin} prior to compression.

Problems

P.IV-1

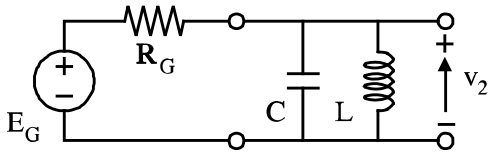


Fig.51

Assume narrowband conditions and show that the noise-bandwidth of the transmission from signal source E_G to v_2 in Fig.51 becomes $\pi/2$ times the 3dB bandwidth like the RC circuit in Example IV-1-1. What is the RMS noise voltage at port 2 if R_G=50Ω @ 290[K] and the circuit has center frequency f_0=100 MHz with 3dB bandwidth BW_{3dB}=10 MHz?

P.IV-2

An amplifier, which has ohmic optimal noise admittance, $y_{nfo}=g_{nfo}$, and ignorable input and correlation admittances, y_{in}, $y_{cor}\rightarrow 0$, is driven from a source with ohmic generator admittance at 290[K] where $y_G=g_G<g_{nfo}$. Show and explain the fact that if the input port of the amplifier is shunted by a resistor to provide noise matching, the resultant output SNR decreases compared to the case with no shunting resistance.

P.IV-3 [9]

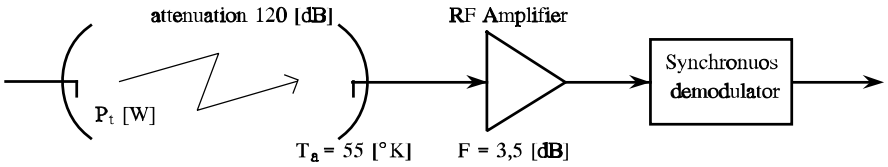


Fig.52

A QPSK signal is transmitted at 45 Mbps rate and is received as shown in Fig.52. The attenuation from the transmitter to the receiver is 120dB and the antenna has the noise temperature T_a=55°K. The noise figure of the RF amplifier corresponding to the antenna impedance is F=3.5 dB.

Calculate the transmitter power P_t[W] corresponding to the resultant bit error rates of 10^{-5}, 10^{-6}, and 10^{-7} respectively.

P.IV-4 [9,10]

A satellite transmits from a distance of 36000 kM a 1.2 Mbps QPSK modulated signal on a 12.2 GHz carrier. The transmitter power is 100 W and the satellite

---

9) Review chap.I pp 24-27 on bit error rates in digital modulations.

10) Review Friis' transmission formula and antenna areas

antenna gain is 27 dB. At the receiver site, the antenna has the noise temperature 52°K and the noise figure of the input amplifier is 2.5 dB at the antenna impedance of 50Ω.

Find the receiver antenna gain and the antenna area that are required to get BER < 10^{-4}. What is the corresponding signal voltage at the receiver amplifier input ?

P.IV-5 [9]

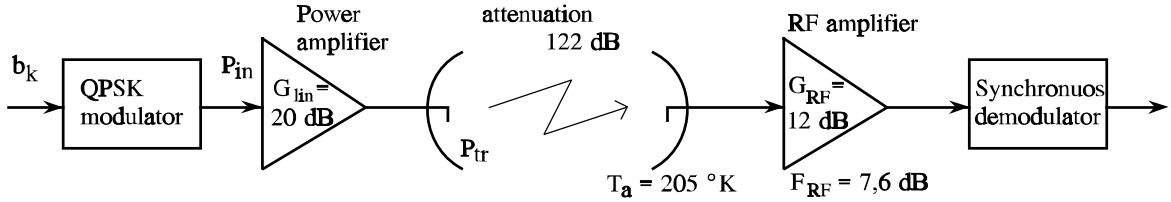


Fig.53

The transmission system in Fig.53 operates in QPSK modulation at rate R_b = 102 Mbps. The carrier frequency is f_0=14.2 GHz and the 122 dB attenuation is from the transmitter output power to the RF amplifier input. It is supposed that all components are matched at impedance level Z_o=50Ω.

What is the RF amplifier input signal to noise ratio per bit E_b/η if a bit error rate of BER = 10^{-6} is required ? Find the corresponding transmitter output power.

What are the 1dB compression input and output levels for the transmitter power amplifier if it is supposed that 0.2 dB compression or less has no significant influence on the resultant bit error rate ?

Repeat the above calculations when a low-noise front amplifier is inserted as shown by Fig.54.

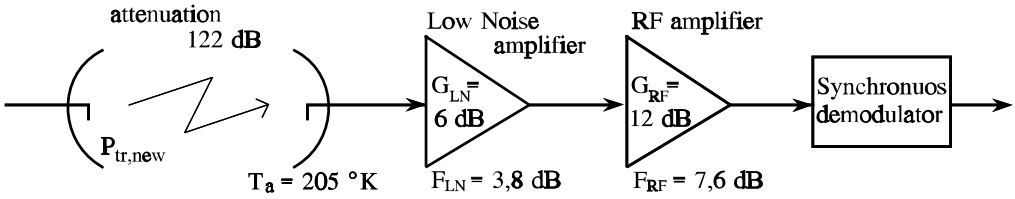


Fig.54

P.IV-6

The transistor in Example IV-3-3, biased by I_C=30mA and V_{CE}=8V, has at 800 MHz the following y-parameters

$$\boldsymbol{Y}_{tr} = \begin{Bmatrix} 12.2 + j4.34\ mS & -0.262 - j2.71\ mS \\ -5.42 - j114.\ mS & 0.600 + j7.63\ mS \end{Bmatrix}. \qquad (162)$$

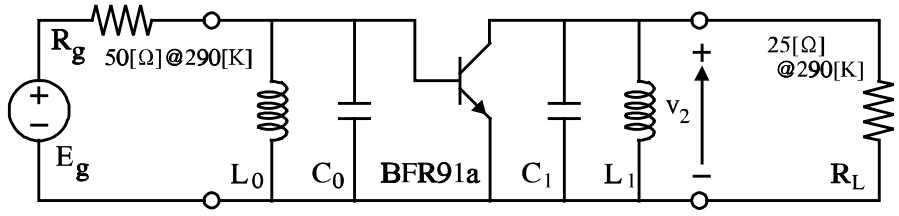


Fig.55

The transistor is driven by a 50Ω impedance and loaded by 25Ω as shown in Fig.6. The input and output resonance circuits are designed independently neglecting feed-back to approximate synchronous tuning around 800 MHz with 3dB bandwidth equal to 80 MHz. It is assumed that the generator and load impedances are at the reference temperature of 290K. Data sheets specify 1dB compression at 36 dBm output power if the transistor is loaded with 75Ω. It is supposed that this level reduces to 28 dBm when the load is 25Ω.

- Verify that it is legal to disregard feed-back effects when the tuning conditions are calculated. Find the operational, the transducer and the available power gains for the amplifiers and explain their differences.
- What is the noise figure of the amplifier ?
- Find the noise bandwidth of the amplifier if it assumed that it corresponds to the synchronously tuned characteristic. What is the total RMS output noise voltage at the amplifier output port ?
- Calculate the amplifier sensitivity in terms of input available power and find the corresponding 50dB 3rd order intermodulation bounded dynamic range.

P.IV-7

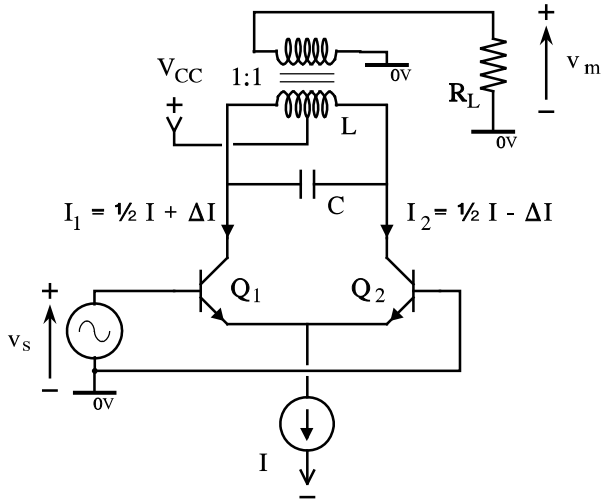


Fig.56

Fig.56 shows a single-tuned differential amplifier that is tuned to center frequency f_o. The differential current is expressed

$$\Delta I = \frac{I}{2} \tanh\left(\frac{v_s}{2 V_t} \right), \quad \textit{where} \quad V_t = \frac{k T}{q} \approx 25 \ [mV]. \tag{163}$$

The bias current is I=5mA. The input signal is amplitude modulated,

$$v_s = V_s \cos \omega_0 t \left(1 + m \cos \omega_m t \right), \qquad \omega_0 = 2 \pi f_0, \qquad \omega_m = 2 \pi f_m \ . \tag{164}$$

The modulation index is m=1, and it is assumed that the modulating frequency is so low, that the sideband transistor load impedance is the same as the center frequency impedance. The nonlinear characteristic in Eq.(163) causes distortion in the low-frequency envelope at the output from the amplifier. Derive an expression for the third harmonic contribution K_3 to the total harmonic distortion of the output envelope, the so-called third order modulation distortion. Find the values with V_s = 1mV and 10mV. Assume that the voltage bias and R_L are chosen to prevent transistor saturation.

References and Further Reading

[1] M.S.Gupta,ed., Electrical Noise: Fundamentals & Sources, IEEE press, 1977 Includes 22 papers on basic noise properties and extensive bibliographies.

[2] M.S.Gupta,ed., Selected Papers on Noise in Circuits and Systems, IEEE press, 1988. Includes 13 more papers on noise.

[3] H.F.Cooke,"Microwave Transistors: Theory and Design", Proc.IEEE, vol.59, pp.1163-1181,Aug.1971, reprinted in [8].

[4] A.van der Ziel, Noise in Solid State Devices and Circuits, Wiley, NY 1987.

[5] M.S.Keshner,"1/f Noise",Proc.IEEE, Vol.70, No.3, pp.212-218, March 1982.

[6] ITT Reference Data for Radio Engineers, Sams, NY, 1968

[7] W.C.E.Lee, Mobile Communications Engineering, McGraw-Hill NY, 1982

[8] H.Fukui,ed., Low-Noise Microwave Transistors and Amplifiers, IEEE press, 1981. Includes 60 papers on device noise and low-noise design.

[9] H.C.de Graff, F.M.Klaassen, Compact Transistor Modelling for Circuit Design, Springer, Wien-N.Y, 1990.

[10] Y.P.Tsividis, Operation and Modeling of the MOS Transistor, McGraw-Hill, N.Y. 1988.

[11] A.van der Ziel, "Noise in Solid-State Devices and Lasers", Proc.IEEE, vol.58,- pp.1178-1206, Aug.1970. Reprinted in [1].

[12] C.D.Bulucea, D.C.Prisecaru,,"The Calculation of the Avalanche Multiplication Factor in Silicon P-N Junctions Taking into Account the Carrier Generation (Thermal or Optical) in the Space Charge Layer",IEEE trans.,Vol.ED-20,pp.693-701,Aug.1973.

[13] P.Antognetti, G.Massobrio, eds., Semiconductor Device Modeling with SPICE, McGraw Hill,1988.

[14] H.Statz,H.A.Hauss,R.A.Pucel,"Noise Characterization of Gallium Arsenide Field-Effect Transistors",IEEE trans. Electron Devices, vol. ED-21,pp.549-562, Sept.1974., reprinted in [8].

[15] H.A.Haus, R.B.Adler, Circuit Theory of Linear Noisy Networks, Technology Press, Wiley, 1959.

[16] J.Engberg, S.T.Larsen, Noise Theory of Linear and Nonlinear Circuits, Wiley 1996.

[17] IRE Standards on Methods of Measuring Noise in Linear Twoports, 1959, Proc.IRE, vol.48, pp.60-68, Jan.1960, included in [8].

[18] A.van der Ziel, Noise: Sources, Characterization, Measurement, Prentice-Hall N.Y. 1970.

[19] H.T.Friis,"Noise Figures of Radio Receivers", Proc.IRE vol.32, pp.419-422, July 1944. Reprinted in [8].

[20] H.Rothe, W.Dahlke, "Theory of Noisy Fourpoles" Proc.IRE, vol.44, pp.811-818, June 1956, reprinted in [8].

[21] Y.L.Kuo,"Frequency-Domain Analysis of Weakly Nonlinear Networks, "Canned" Volterra Analysis", IEEE Circuit and Systems Magazine, aug.1977, pp.2-8 (part 1), and Oct.1977, pp.2-6 (part 2).

[22] S.A.Maas, Nonlinear Microwave Circuits, Artech, Boston 1988.

[23] R.S.Carson, Radio Communications Concepts: Analog, Wiley NY. 1990.

Index