RADIOMETRY

by

Torleiv Orhaug

National Radio Astronomy Observatory
Green Bank, West Virginia

LECTURE NOTES

June, 1961

RADIOMETRY

by

Torleiv Orhaug

I. INTRODUCTION

The receiving equipment used to detect astronomical radio waves is usually called a radiometer. In this chapter, we will be dealing with the principles of radiometry. Specificly, we will be interested in the ultimate limit of every measuring system, the limit set by the statistical noise fluctuations, and we shall see how to calculate the minimum measurable or detectable signal from the overall system parameters. We shall also investigate another important limitation in a receiving system, such as the effects of variations in the receiving system gain, and we will see how the noise fluctuations and the gain fluctuations together determines the overall sensitivity of the measuring system.

The useful range of frequencies in radio astronomy is bounded on the low frequency side by the transparency of the terrestrial ionosphere, with 20 Mc. being a good average figure for the lower limit. On the high frequency side, the attenuation in the lower atmosphere determines an upper limit for about 30 kMc. We will therefore be dealing with radiometers useful in this frequency range.

In practice, the following sources of "noise" limits the minimum detectable signal:

1. Interference
   a. Man-made (civilization interference)
   b. Natural
2. Fluctuations caused by transmission medium
   a. Ionospheric influence: absorption, irregularities
   b. Atmospheric influence: absorption, irregularities
3. Instability in receiver system gain (variation in input coupling losses, connectors losses, and amplifier gain)

4. Total receiver system noise
   a. "Signal noise"
   b. Receiver noise
   c. Antenna noise
   d. Input losses

In order to get a qualitative picture of the problems involved in measuring radio astronomical sources, we will calculate the power from the most powerful point source, Cassiopeia. Within the bandwidth B, the total power collected on a plane area which is perpendicular to the rays from the source, A is

$$P = \tfrac{1}{2} SA \cdot B \text{ watts} \qquad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

If we observe at a frequency of 100 Mc, then $S \approx 200 \cdot 10^{-24} m^{-2} (c/s)^{-1}$. We assume an observing bandwidth B = 10 Mc, and compute the total power falling onto the earth's surface (Diameter = 6400 km). This gives

$$P = 0.12 \text{ watts}$$

So, even for an antenna system having the same diameter as the earth, a 10 Mc bandwidth receiver only intercepts about 1/10 of one watt. Any reasonable antenna system only intercepts about 10^{-10} to 10^{-11} of this supply.

We know that a source giving flux density S and occupying a solid angle Ω, the corresponding brightness b, is

$$b = {}^{S}/_{\Omega}$$

and according to Raleigh-Jean's law, the brightness temperature T_b is

$$T_b = \frac{b\lambda^2}{2k} \ ^\circ K$$

of this temperature, an antenna, subtending a solid angle Ω_A, receives an amount of power, corresponding to a noise temperature T_A, given by

$$T_A = \frac{\Omega}{\Omega_A} T_b \qquad \ldots\ldots\ldots\ldots\ldots \quad (2)$$

or

$$T_A = \frac{S\lambda^2}{\Omega_A 2k}$$

and for $\Omega_A = \frac{\lambda^2}{A}$, we obtain

$$T_A = \frac{SA}{2k} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2a)$$

If we use the same example as before, then $S = 200 \cdot 10^{-24}$ watt m^{-2} $(c/s)^{-1}$, $k = 1.38 \cdot 10^{-23}$, and for an antenna having a collecting area of $A = 20^2 m^2$, we obtain the following antenna temperature:

$$T_A \simeq 200^\circ \text{ K.}$$

This antenna temperature must now be observed with a receiver having an equivalent temperature of the order of 500° K or more. Although this observation seems difficult, we shall later see how it is possible not only to measure this 200° K signal using the 500° K receiver, but even noise signals less than 1° K, or smaller than 1/1000 of the receiver noise background. The method which enables this relatively high degree of sensitivity is <u>integration</u>. We shall later see how the integration may be accomplished either by having a large rf-bandwidth (from several kc to hundreds of Mc) or by using a long integration time at the output. The first method has the disadvantage of reducing the frequency resolution, and in case of spectrum observations (hydrogen line for instance) the rf-bandwidth should not exceed 5-50 kc. For continued observations much larger bandwidth can be used, because the spectra are slowly varying with frequency. The second method, on the other hand, has the disadvantage of using much observing time; for large and expensive antenna installations this is a factor which must be taken into account.

II. THE CONCEPT OF NOISE TEMPERATURE AND NOISE FIGURE

Before we go into a detailed discussion of a radiometer system, we have to consider the concept of noise temperature (Ref. 1, 2, and 3). We know that a thermal noise source at the absolute temperature T°K generates an available noise power of

$$P = kTB \text{ watts}$$

within the bandwidth B. We notice that this power is independent of the electrical characteristics of the sources. If we, for instance, consider a resistor of this temperature, we have the corresponding rms noise voltage V, given by

$$\frac{\overline{V^2}}{4R} = kTB$$

$$\text{or} \quad \overline{V^2} = 4RkTB$$

$$\text{and} \quad \sqrt{\overline{V^2}} = \sqrt{4\ RkTB} \qquad \ldots\ldots\ldots\ldots\ (3)$$

For T = 300°K, B = 1 Mc., R = 50 ohm, we have

$$V = \sqrt{\overline{V^2}} = \sqrt{4 \cdot 50 \cdot 1.38 \cdot 10^{-23} \cdot 300 \cdot 10^{6}}$$

$$V \simeq 0.9\ \mu \text{ Volt}$$

This is the input voltage for a 300°K noise source. In order to enable the recording of this small voltage, we have to amplify this voltage roughly 10^6 times (or 120 db).

Let us connect two resistors R_1 and R_2 in series, each of which has the temperature T_1 and T_2, respectively. We now wish to find an equivalent circuit for the series combination for these resistors (that is, a single resistor at a uniform temperature). Because the noise voltages in the two resistors are uncorrelated, we have to add the corresponding noise powers,

$$\overline{V_{eq}}^2 = 4\ kT_1R_1B + 4\ kT_2R_2B$$

$$= 4\ k\ T_{eq}\ (R_1+R_2)B$$

$$\text{where } T_{eq} = T_1 \frac{R_1}{R_1+R_2} + T_2 \frac{R_2}{R_1+R_2} \qquad \ldots\ldots\ldots\ (4)$$

The original circuit now produces exactly as much noise power as if it has been at the uniform temperature T_{eq}.

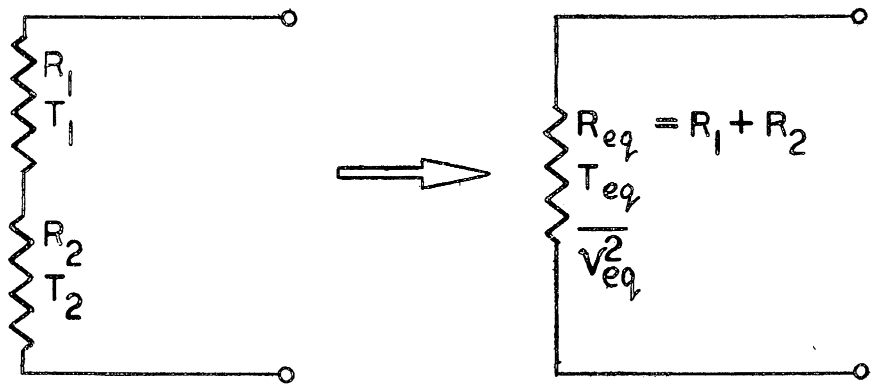


Figure 1. Equivalent circuit of two series connected resistors.

Next, let us consider the transmission of noise power through a passive fourterminal device (a chain of resistors for instance).

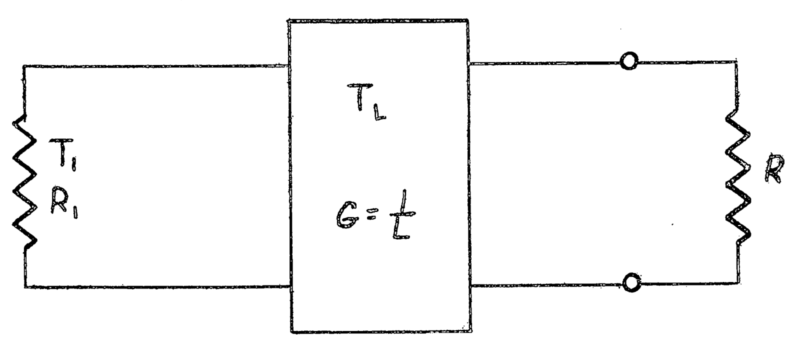


Figure 2. Noise source connected to a passive four-terminal device.

The power delivered to the output load R_ℓ consists of $\frac{1}{L} kT_1B$, due to the input source and an additional $m\, kT_LB$ from the network itself. That is,

$$P_{output} = \frac{1}{L} kT_1B + mkT_LB$$

If $T_L = T_1$, we know that $P_{out} = kT_1B$, which means that $m = 1 - 1/L$, giving

$$P_{out} = \frac{1}{L} kT_1B + (1 - 1/L)kT_2B$$

or, in terms of equivalent noise temperature

$$\boxed{T_{eq} = \frac{T_1}{L} + (1 - 1/L)T_L} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

It is interesting to note that if $L = \infty$, $T_{eq} = T_L$.

A problem which often occurs is that the front end of the receiver often contains lossy elements before the signal is amplified. In this case, $T_L = T_o$ and

$$T_{out} = \frac{T_{in}}{L} + (1 - 1/L)T_o$$

provided that the lossy device has the temperature T_o (where T_o = 290°K).

Let us now measure all noise temperatures relative to the temperature T_o = 290°K, and we introduce the temperature ratio

$$t = T/T_o$$

We can now write

$$t_{out} = \frac{t_1}{L} + (1 - 1/L)t_L$$

or

$$t_{out} - 1 = \frac{t_1}{L} - \frac{1}{L} + \frac{1}{L} + (1 - 1/L)t_L - 1$$

$$= \frac{t_1 - 1}{L} - (1 - 1/L) + (1 - 1/L)t_L$$

$$t_{out} - 1 = \frac{t_1 - 1}{L} + (1 - 1/L)(t_L - 1) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

But, $t - 1 = \frac{T - T_o}{T_o}$ which is a measure of the excess noise. The above equation therefore states that for the transmission of noise through a lossy network excess noises are added.

We shall now introduce the ideas of receiver noise temperature and noise figure by means of two different approaches. First, let us state that the concept of noise temperature is a convenient way of looking at the system with regard to the output result. The reason for conceiving this concept is to introduce an equivalent receiving system, having the same parameters (G, B, and output fluctuations) and having all noise sources concentrated at the input. The noise temperature is defined as the temperature which a resistor at the input (having the same impedance as the actual input circuit, i.e. the antenna) should have in order to produce the same output noise level and fluctuations with a noise-free receiver. By using this temperature (T_R), we can define the noise figure of the receiver as

$$F - 1 = \frac{T_R}{T_o}$$

or

$$F = \frac{T_R + T_o}{T_o}$$

In other words, the noise figure is the ratio of the resistor noise temperature at room temperature increased in temperature by the noise temperature of the receiver.

The reason behind this special definition may be made clearer if we use the historical approach. Friis (4) defined F as

$$F = \frac{\dfrac{\text{available sign power from the generator}}{\text{available noise power from the generator}}}{\dfrac{\text{available sign. power from the amplifier}}{\text{available noise power from the amplifier}}} = \frac{(S/N)\text{gen.}}{(S/N)_{out}}$$

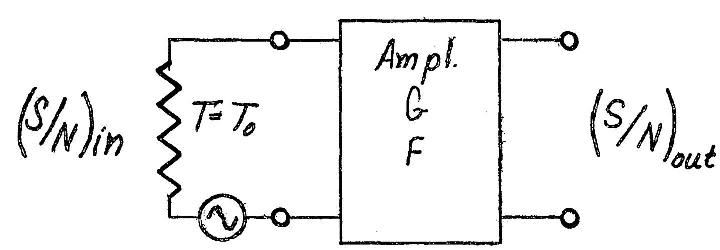


Figure 3. System for definition of noise figure.

We can also state that F is the degeneration of the signal to noise ratio because of the passage and through the action of the amplifier. We therefore have

$$F = \frac{S_g / kT_oB}{G\,S_g / N_{out}}$$

or

$$N_{out} = F\,GkT_oB = G\left\{kT_oB + (F-1)kT_oB\right\}$$

$$= GkT_{tot}\,B$$

Therefore, the total input noise temperature is

$$T_{tot} = \underset{\text{input noise from gen.}}{T_o} + \underset{\text{input noise of ampl.}}{\underbrace{(F-1)T_o}} = T_o + T_R$$

Rearranged, we have

$$F = \frac{T_o + T_R}{T_o}$$

and

$$\boxed{T_R = (F-1)T_o} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

as before.

<u>Example</u>: Find the noise figure of a passive, lossy network

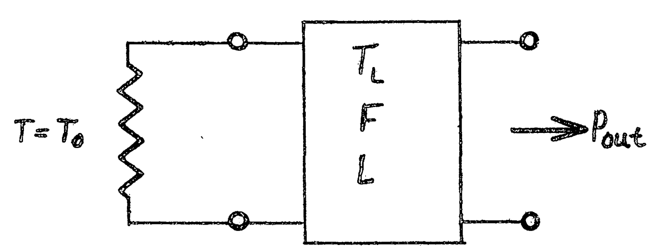


Figure 4. Noise figure of a lossy device of temp. T_L.

$$P_{out} = \frac{1}{L} kT_oB + (1 - 1/L)kT_LB$$

$$= \frac{1}{L} kT_oB \cdot F$$

$$\boxed{F = 1 + (L - 1)\frac{T_L}{T_o}} \quad \ldots\ldots\ldots\ldots\ldots \quad (8)$$

Special case, $T_L = T_o$

$$\boxed{\begin{array}{l} F = L \\ \text{for } T_L = T_o \end{array}} \quad \ldots\ldots\ldots\ldots\ldots \quad (9)$$

Next, we consider the noise result from cascading a number of noisy stages:

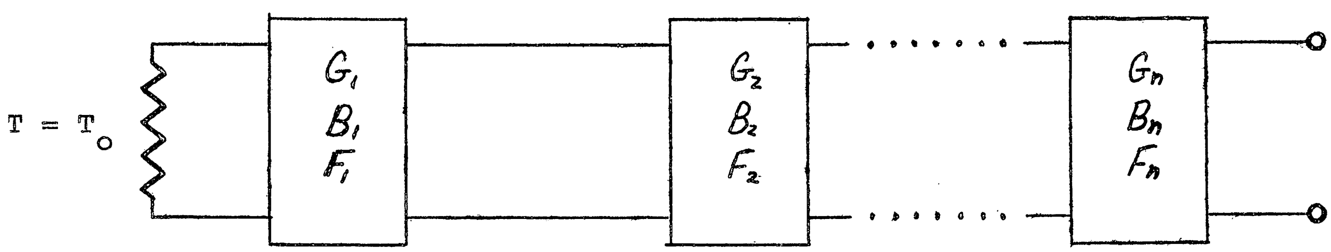


Figure 5. Cascading of n noisy stages.

$$N_{out} = kT_oB_{1,n}G_1\cdots G_n + (F_1-1)kT_oB_{1,n}G_1\cdots G_n + (F_2-1)kT_oB_{2,n}G_2\cdots G_n$$

$$= F_{tot}G_1\cdots G_nB_{1,n}kT_o$$

$$\boxed{F_{tot} - 1 = F - 1 + \frac{F_2 - 1}{G_1}\,\frac{B_{2,n}}{B_{1,n}} + } \quad \ldots\ldots \quad (10)$$

This formula is only valid for the spot noise figure, and not for the integrated noise figure. The spot noise figure is the noise figure we obtain by investigating the noise performance in a small frequency band (one c/s), and the integrated noise figure F is obtained from the spot noise figure F by means of the relation

$$\overline{F} = \frac{\int F \cdot G(f)\,df}{\int G(f)\,df}$$

where G(f) is the gain-function of the network.

Example: Compute the total noise figure of a system consisting of a lossy pad (G = 1/L) at temperature T_L followed by a receiver with noise figure F_R and corresponding noise temperature T_R.

$$F_L = 1 + (L - 1)\frac{T_L}{T_o}$$

$$F_{tot} = 1 + (L - 1)\frac{T_L}{T_o} + \frac{F_R - 1}{1/L}$$

$$F_{tot} = 1 + L \left\{ (1 - 1/L)\frac{T_L}{T_o} + F_R - 1 \right\} \quad \ldots \quad (11)$$

Special case

$$T_L = T_o \qquad \boxed{F_{tot} = L\,F_R} \quad \ldots \quad (12)$$

Referred to the input:

$$T_{tot} = L\,T_R + (L - 1)T_L$$

The increase in noise temperature because of L at temperature T_L:

$$T_{tot} - T_R = LT_R - T_R + (L - 1)T_L = (L - 1)T_R + (L - 1)T_L$$

$$\boxed{T_{tot} - T_R = (L - 1)(T_R + T_L)} \quad \ldots \quad (13)$$

Equation (13) shows that the system suffers from an increase in noise system temperature, even when $T_L = 0°K$. This is, of course, due to the attenuation of the signal in L.

Because noise powers are additive, the total receiver system noise temperature is

$$T_{syst} = T_R + T_A$$

where

$$T_A = T_{source} + T_{gal} + T_{spillover} + T_{atm}.$$

Here T_{source} is antenna temperature caused by the source, T_{gal} is the antenna temperature caused by the galactic background, $T_{spillover}$ is the temperature caused by the part of the ground ($T = T_o$) seen by the primary feed, and $T_{atm.}$ is the antenna temperature caused by the atmospheric attenuation. On the lower frequencies T_{gal} makes an important contribution (for frequencies lower than approximately 100 Mc) while in the microwave range $T_{atm.}$ plays an important role.

One feature of radio astronomical observations is that the S/N ratio is independent upon receiver bandwidth. In communication theory this is often <u>not</u> the case, because the signal has a very limited range in frequency. In this case, the S/N ratio <u>increases</u> with <u>decreasing</u> bandwidth B, or

$$\frac{S}{N} = \frac{S}{kT_{syst}} \; \frac{1}{B}$$

III. NOISE TEMPERATURE AND SENSITIVITY OF A RADIOMETER

Now that we have obtained all necessary formulas for noise calculations, we will attack the more practical problem of calculating the noise temperature of a crystal mixer receiver. We do this as an exercise because this classical receiver is still the most common type of radiometer design.

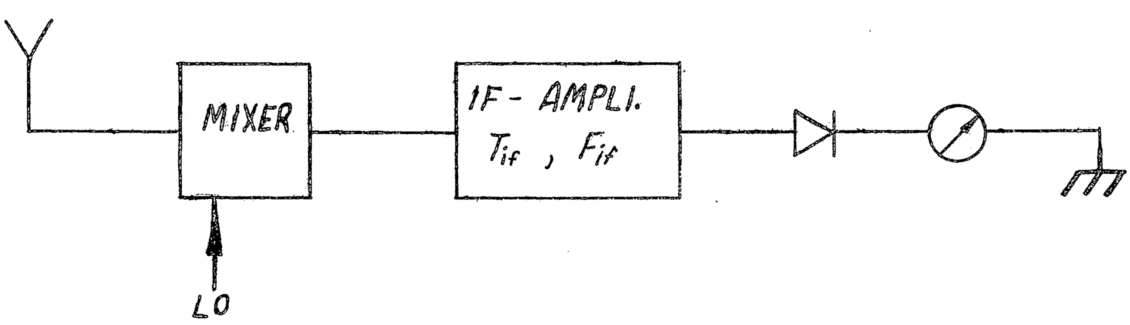


Figure 6. Principle of a crystal mixer receiver.

The output noise power of the mixer is,

$$P_{mx} = kT_o F_{mx} \frac{1}{L_{mx}} B = kT_c B$$

Where T_o is the temperature of the input termination and T_c is the noise ratio of the crystal. The mixer noise figure is then

$$F_{mx} = \frac{T_c}{T_o} L_{mx}$$

and

$$F_{tot} = F_{mx} + L(F_{if} - 1)$$

$$\boxed{F_{tot} = L_{mx} \left(\frac{T_c}{T_o} + F_{if} - 1\right)} \quad \ldots\ldots\ldots\ldots (14)$$

L_{mx} = conversion loss (3-5 times)

$\frac{T_c}{T_o} = t$ = temperature ratio of crystal (1.1 - 1.6)

F_{if} = 1.2 - 2.0 for F_{if} = 30 Mc.

Example:

$F_{if} = 1.5$

$t_x = 1.3$

L_{mx} = 4 times

$F_{tot} = 4\ (1.3 + 0.5) = 4 \cdot 1.8 = 7.2$

giving

$$T_{tot} = (F_{tot} - 1)T_o = 1800^\circ\ K$$

Let us now consider the signal through the crystal mixer-radiometer with respect to both time and frequency. (The signals in Figure 7 are depicted as the function of time and frequency in different stages of the receiver.)

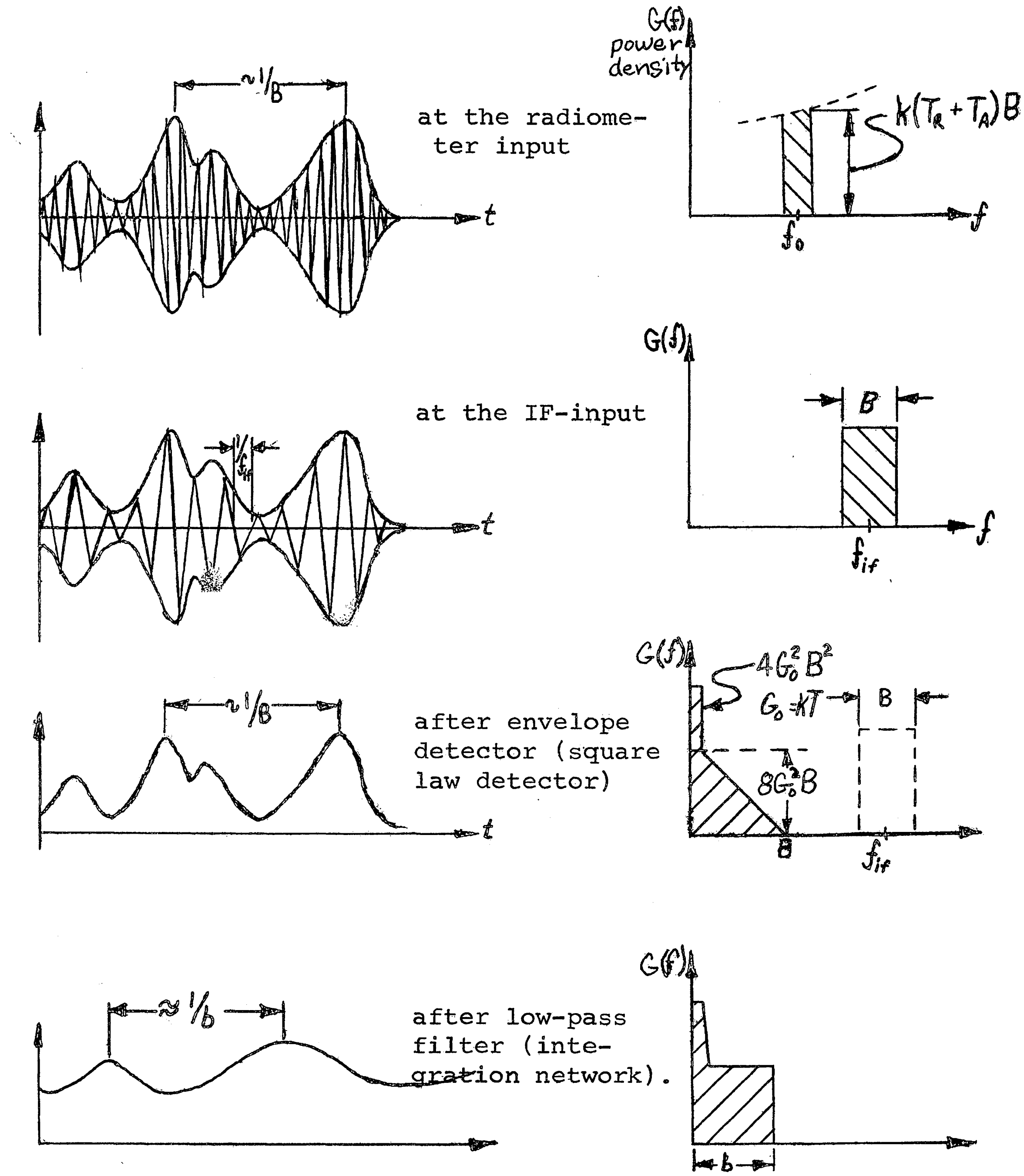


Figure 7. Time signals and power spectra for noise at different stages in a radiometer.

The incident signal consists of broadband (white) noise. The mixer accepts only components at and around the signal frequency f_o, and converts this spectrum to the if-frequency (f_{if}). At the if-level, the amplifiers have a bandwidth B. The different noise component within this spectrum beat with each other, resulting in a randomly modulated wave at the if-frequency (f_{if}). The mean modulation frequency is B c/s, while the modulation amplitude has the celebrated Rayleigh probability distribution. In other words we can say that the high frequency components are correlated over a time of the order of $\frac{1}{B}$ secs, so the finite bandwidth has processed an integration of the high frequency noise components. In most superheterodyne systems, the main amplification is done at the if-level. From the if-amplifier, the signal is fed into an envelope detector giving a DC-(current) output equal to the envelope of the if-signal (the modulation component). The output consists of a DC-component (proportional to KTB for a square law detector) and noise like components superimposed with the highest frequency component being B. The after-detector power spectrum is depicted in Figure 7. This spectrum is then filtered through a low-pass filter (usually a RC-network) which passes the DC-component and the lower frequencies of the AC-spectrum . In this process the noise fluctuations are smoothed, and the waveform is correlated within a time of the order of the time constant τ.

We are now interested in finding the AC-fluctuations on the output record relative to the DC-level. Since the DC-level is directly proportional to the input system temperature T, we can directly get the output fluctuations into terms of an equivalent noise temperature. If the noise bandwidth of the output low-pass filter is b c/s, the RMS-value of the output fluctuation is according to Figure 7 (5)

$$\sqrt{\overline{V_n}^2} = \sqrt{8G_o^2 Bb}$$

and relative to the DC-level, the fluctuation level is

$$\sqrt{\frac{\overline{V_n}^2}{V_o^2}} = \sqrt{\frac{8G_o^2 Bb}{4G_o^2 B^2}} = \sqrt{2\frac{b}{B}}$$

But the DC-level is now

$$V_o = C_1 \cdot kT_{tot}B$$

$$= C_1 \cdot k(T_A + T_R)B$$

(C_1 being a constant) and a change in antenna temperature, ΔT_A corresponds to a change in output DC-level, equal to

$$\Delta V_o = \left| \frac{\partial V_o}{\partial T_A} \right| \cdot \Delta T_A = C_1 \cdot kB \cdot \Delta T_A$$

giving the relative noise fluctuation equalized to input temperature

$$\boxed{\Delta T_A = T_{tot} \sqrt{2\frac{b}{B}}} \quad \text{. . . . . (15)}$$

The frequency response for an RC-integration network is

$$G(\omega) = \frac{1}{1 + \omega^2 \tau^2}$$

where

$$\tau = RC$$

The corresponding noise bandwidth b, is now obtained by integrating the frequency response over all frequencies

$$2\pi b = \frac{d\omega}{1 + \omega^2 \tau^2} = \frac{1}{\tau} \cdot tg^{-1} (\omega\tau) = \frac{\pi}{2}$$

or

$$b = \frac{1}{4\tau}$$

The noise bandwidth b should now be compared with the 3-DB bandwidth of the low-pass filter

$$b_{3db} = \frac{1}{2\pi\tau}$$

and the ratio between these two bandwidths is

$$\frac{b}{b_{3db}} = \frac{2\pi}{4} \simeq 1.57$$

In terms of the integration time τ, the relative fluctuation is equal to

$$\boxed{\Delta T_A = \frac{T_{tot}}{\sqrt{2B\tau}}} \quad \text{. . . . . (16)}$$

This is the RMS-value of the output fluctuations measured in terms of input temperature, and this value is often referred to as the sensitivity of the "DC-radiometers" or total-power radiometer. In practice, however, the minimum detectable signal is many times greater than the RMS-noise level.

We can also qualitatively derive Equation (17) by means of physical reasoning. As we have seen earlier, the predetection noise signal is correlated over a period of time of the order of $\frac{1}{B}$ secs. This means that there are B effectively independent noise contributions per second producing detector output. These contributions are averaged in the post detection filter over a time given by the time constant. During this interval there are $B\tau$ independent contributions, and therefore, the probable root-means-square deviation is $\sqrt{\frac{1}{B\tau}}$ times that of the mean (DC-) output level.

Let us now consider possible causes of variations in the output DC-level other than the thermal noise itself. We will assume that there are instabilities both in the total receiver gain and in the receiver noise figure, and we assume that the relative (RMS) variations are $\frac{\Delta G}{G}$ and $\frac{\Delta F}{F}$ respectively. For the square law detector, the DC-output is

$$V_o = C_2 \cdot Gk \left[T_A + (F - 1) T_o \right]$$

where C_2 is a constant. If we equalize the output variations to an apparent input signal variation T_A, we have

$$\left| \frac{\partial V_o}{\partial T_A} \right| \Delta T_A = \left| \frac{\partial V_o}{\partial G} \right| \Delta G + \left| \frac{\partial V_o}{\partial F} \right| \Delta F$$

giving

$$\Delta T_A = (T_A + T_R) \frac{\Delta G}{G} + T_o \Delta F$$

or

$$\boxed{\Delta T_A = (T_A + T_R) \frac{\Delta G}{G} + \Delta T_R} \qquad \ldots\ldots\ldots\ldots (17)$$

Equation (17) summarizes the disadvantages from which the total power radiometer suffers; the gain-fluctuation term is proportional to the <u>total</u> system temperature and also, a change in receiver temperature shows directly on the output record. The far more dangerous term in the Equation (17) is the gain instability term, and a good approximation to the output fluctuations, measured in terms of input temperature, is therefore

$$\Delta T_{tot} \simeq c \frac{T_{tot}}{\sqrt{B\tau}} + T_{tot} \frac{\Delta G}{G}$$

where c is a constant determined by the detector law ($c = 2^{-\frac{1}{2}}$ for a square law device). Rewriting the above results we have

$$\boxed{\Delta T_{tot} \simeq T_{tot} \left[\frac{c}{\sqrt{B\tau}} + \frac{\Delta G}{G} \right]} \quad \ldots\ldots\ldots\ldots \quad (18a)$$

Up to this point we have assumed that the various fluctuation terms are added in amplitude. In practice, however, the noise fluctuations and the gain-instability fluctuations are absolutely uncorrelated, and the fluctuation terms should therefore be added as follows

$$\Delta T_{tot} \simeq T_{tot} \left[\frac{c}{\sqrt{B\tau}}^{2} + \frac{\Delta G}{G}^{2} \right]^{\frac{1}{2}} \quad \ldots\ldots \quad (18b)$$

The constant fight in total power radiometer design is to make

$$\frac{\Delta G}{G} < \frac{c}{\sqrt{B\tau}}$$

so that the system sensitivity is governed by the noise fluctuations.

IV. IMPROVEMENT OF RECEIVER STABILITY

Dicke (6) introduced the comparison radiometer which greatly improved the stability of the system, and at the cost of more complex electronic circuitry the receiver stability problem was more or less solved. The idea which Dicke introduced was to switch the input rapidly between the signal source (the antenna) and a comparison source. If the two sources present the same impedances to the receiver they should give the same output provided that they are at the same temperature. If the input circuits are at different temperatures, and the output is switched between two integrating circuits which are connected in opposition, one will read a DC-value proportional to the difference in temperature between the two input sources. A phase detector, switched with a reference voltage in synchronism with the input switching, reads the amplitude of a signal component in phase with the reference voltage and is therefore used as an element to detect the value of the switch-component.

The switch-comparison system has, however, the disadvantage that it lowers the sensitivity by a factor of 2. First, the signal (antenna) is only present half of the time, and therefore, lowers the available integration time to half the value for the total power radiometer. Also, the comparison is made against a noise signal which also is present for half the time. Each of these factors individually lower the sensitivity of a factor $\sqrt{2}$, giving a factor of 2 together. (Ref. 7, 8, 9, 10).

Let us next consider what happens to the stability. The output of the envelope detector for the two switch positions is

$$V_1 = C_1 k \ (T_A + T_R) \ BG$$

$$V_2 = C_1 k \ (T_C + T_R) \ BG$$

where T_A and T_C are the temperatures of the antenna (signal) and comparison positions respectively. The output from the phase sensitive detector is therefore

$$v = C_2 (V_1 - V_2)$$

$$= C_1 C_2 kBG \left[T_A - T_C \right]$$

and the instability term becomes

$$\Delta T = (T_A - T_C) \frac{\Delta G}{G}$$

showing that the stability with respect to gain changes has been improved by a factor of

$$\mathcal{K} = \frac{T_A + T_R}{T_A - T_C}$$

or

$$\mathcal{K} = \frac{1 + {}^{T_R}/_{T_A}}{1 - {}^{T_C}/_{T_A}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (19)$$

compared to the total power receiver.

It is obvious from equation (19) that the gain instability term vanished (at least in theory) when $T_A = T_C$. There are two ways available to fultill this condition. In the first system, the channel having the lowest temperature can have noise added to it, so that $T_A = T_C$. In order for the other system to work, the gain of the receiver is switched synchrounisly with the front end switching so that

$$(T_A + T_R)G_A = (T_C + T_R)G_C$$

where G_A and G_C are the gain in the signal and comparison positions respectively. This system is called "gain-modulation."

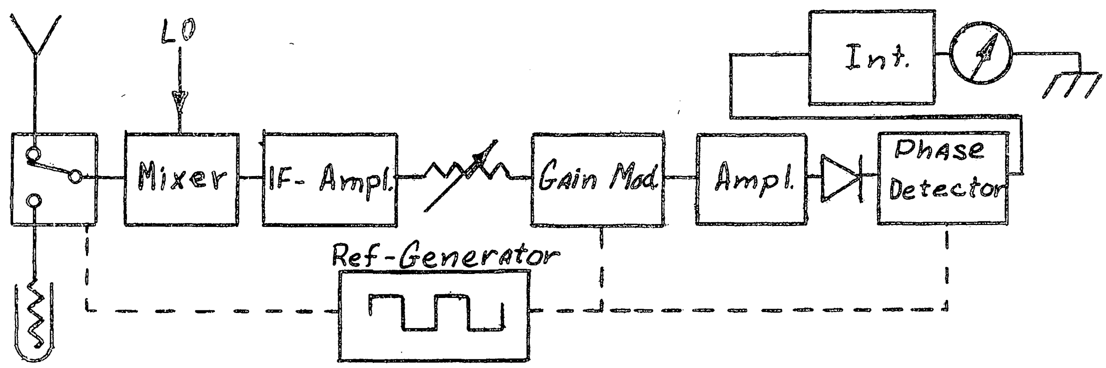


Figure 8. Principle of a switched receiver.

We have briefly discussed the principles of the switched receiver, but we have not discussed how the switching is done. One form of switching uses a resistive load at a constant temperature for the comparison channel (switch-load-radiometer). The front end (microwave) switch may either be a crystal switch or a circulator, while the comparison channel may be another antenna slightly set off focus. In this case, the signal from the radio source is compared with the signal from a point near the source. This system is often used in interferometry and in array-antenna systems. If the signal has a narrow frequency spectrum as the hydrogen line, the comparison channel may be a frequency lying outside the line. In this case, the switching is done by switching the local oscillator between two convenient spaced frequencies.

It should be pointed out that although the switched-radiometer has a lower sensitivity with regard to thermal noise as compared to the total power receiver, the total sensitivity is far better when the gain fluctuation are taken into consideration.

The switch receiver is not the only way to decrease the effects of instabilities. Another method is the use of two receivers and multiplying the video output signals. If these two receivers are connected to the same antenna, only the correlated noise coming from the antenna produces an output since the noise generated in the receivers are not correlated, and so produce no dc-output. This type of radiometer is commonly used in interferometer work where one already has two receivers (at least two front ends) at the different antennas. For single antennas the system is little used, mainly because of the difficulties of connection of the two receivers to the single antenna.

The switching principle has more or less solved the stability problem in modern radiometry; the next problem is to lower the noise term. In principle, this can be done in three ways: (1) Increasing the time constant, (2) increasing the bandwidth B, (3) decreasing the system noise temperature T_{tot}. The last way will be considered in the next chapter. The integration time can not usually be increase unlimited in practice, so a value of τ of more than a couple of minutes is usually not used. For ordinary mixer and cascode-preamplifier-combinations a bandwidth of more than 10 Mc is difficult to obtain. There exists, however, a microwave amplifier with a reasonable noise figure, and extremely large bandwidth. This is the traveling wave tube, (TWT). For comparing different systems from the noise-fluctuations point of view, we can use the parameter M (figure of merit).

$$M = \frac{T_{tot}}{\sqrt{B}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (20)$$

giving the output fluctuations for a one second time constant. The figure of merit shows that the noise figure or noise temperature is <u>not</u> the only parameter of importance. For spectrum work however, the noise temperature is the only important parameter since one wants to have B small in order to get a high frequency resolution. Consequently, B can not be increased in order to decrease the output noise fluctuations. The only way to lower the noise fluctuations is to decrease the system noise temperature (keeping the integration time constant.)

Using the broad band technique (TWT) causes two problems. One is the increased probability of getting interference in the receiving system, and the other is the requirement for a smaller value of the gain instability term, $(T_A - T_C)\frac{\Delta G}{G}$. The broad band technique has, however, been successfully used and a sensitivity of the order of 0.01°K has been obtained (Ref. 11).

IV. SYSTEM NOISE CONSIDERATIONS AND LOW NOISE TECHNIQUES

In the foregoing chapter we have seen how to compute the overall system sensitivity if we are given the system parameters. It, therefore, seems natural to discuss the importance of the different noise parameters. For a moment do not consider modern low noise techniques, but look at the receiver temperature as a function of frequency. As depicted in Figure 9, we see that for frequencies below approximately 500 Mc. UHF-triodes are superior to mixers from a receiver temperature point of view.

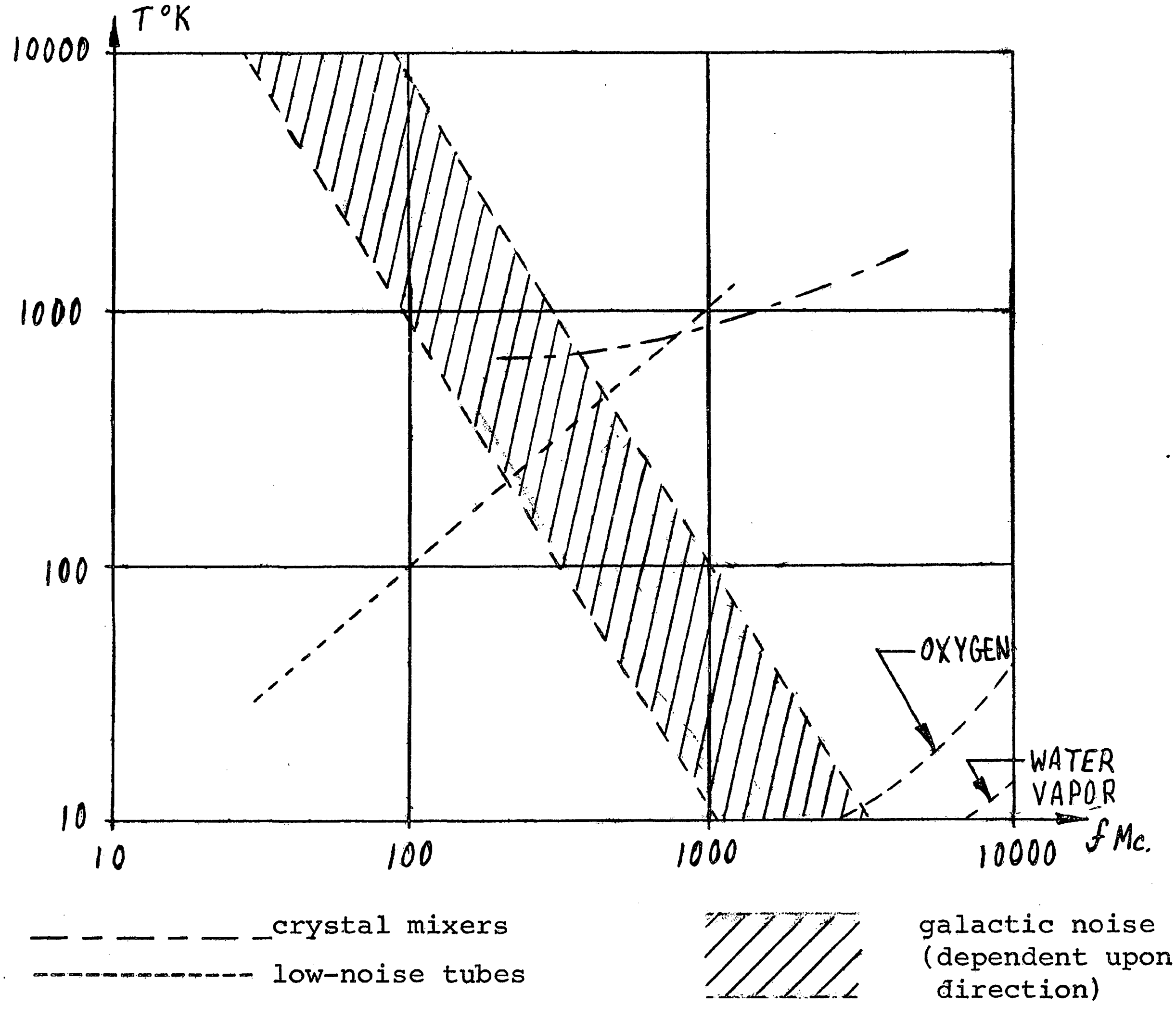


Figure 9. Consideration of system temperature.

In this same figure that galactic background temperature is plotted as a function of frequency. It is obvious from the graph that for frequencies below approximately 100 Mc the system noise temperatures are almost entirely determined by the galactic noise level, and, therefore can not be significantly improved by decreasing the receivers noise temperatures. Above 500 Mc the total system temperature is entirely determined by the receiver noise (neglecting frequencies above 5000 Mc, where atmospheric effects, like water vapor starts to be important) and improvement in receiver noise performances therefore directly pays dividends.

A. Parametric amplifiers

We will now briefly discuss two modern low noise systems, parametric amplifiers and masers, and later discuss their influence on high sensitive radiometry. The parametric amplifiers are based upon the classical principle of amplification at microwave by means of the variation of a parameter in a tuned-rf-circuit. The principle of the gain-mechanism can be most easily understood by considering the following problem: assume that we have a time-varying reactive (lossless) element (condenser, for instance). This reactive element is non-linear (the capacitance might be a function of the applied voltage c(v).) Now, if we apply a strong rf-signal ("pump-power") at a frequency f_p and a weak signal (the signal to be amplified) at a frequency f_s then from the simple theory of mixing we know that the result (due to the presence of the non-linear element) is sidebands on frequencies

$$mf_s + nf_p$$

where m and n are integers. Manley and Rowe (12) showed that the following equations are valid for the average energy ($P_{m,n} = P_{mfs + nfp}$) for a specific combination of the independent integers m and n.

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{m P_{m,n}}{mf_s + nf_p} = 0 \quad \ldots\ldots\ldots \quad (21)$$

$$\sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{m P_{m,n}}{mf_s + nf_p} = 0 \quad \ldots\ldots\ldots \quad (22)$$

Two remarkable results from Equations (21) and (22) are that the energy spectrum produced by the reactive element is independent of the shape of the non-linear device and also independent of the power level introduced by the pump source. A common useful approximation of Equations (21) and (22) results from assuming that only the two lowest sidebands, $f_p + f_s$ and $f_p - f_s$ are present. The energy relation for m = o,1 and n = o,1 are

$$P_s = f_s \left[\frac{P_{p-s}}{f_p - f_s} - \frac{P_{p+s}}{f_p + f_s} \right] \qquad \ldots\ldots\ldots \quad (23)$$

$$P_p \quad - f_p \left[\frac{P_{p+s}}{f_p + f_s} + \frac{P_{p-s}}{f_p - f_s} \right] \qquad \ldots\ldots\ldots \quad (24)$$

the expressions commonly used in the literature. A positive value of P denotes power generated by the amplifier while a negative value of P denotes power absorbed by the amplifier. Equation (24) shows that pump power is always supplied to the reactance ($P_p < o$) if passive loads are coupled to the sideband frequencies. (P_{p+s} and $P_{p-s} > o$). Equation (23), on the other hand, shows that signal power is either supplied or absorbed by the amplifier dependent upon the relative power levels of the sideband frequencies. In other words, the parametric amplifier can either amplify or attenuate the signal depending upon how the sideband frequencies are terminated.

Parametric amplifiers can be divided into two main catagories: (1) circuits which do not introduce negative resistance to the input signal circuit and (2) circuits which do introduce negative resistance. Amplifiers which belong to case (2) are instable and bilateral devices, while amplifiers which belong to case (1) are potentially stable devices. In order to investigate the first mode of operation, we assume that the circuit is tuned to f_s and f_{s+p}. From Equation (23) we obtain

$$P_{p+s} = -P_s \frac{f_p + f_s}{f_s} \qquad \ldots\ldots\ldots \quad (25)$$

showing that if signal power is coupled to the amplifier, then power is generated by the amplifier at the upper sideband. This mode of operation is therfore called an up-converter. In it the signal is applied to the device at frequency f_s while the amplified signal appears at the frequency $f_s + f_p$, with the amplification occuring as a result of the frequency conversion process. Other names for this mode of operation are sum-frequency mode or upper-sideband upconverter. The gain of the up-converter is

$$G = \frac{P_{p+s}}{P_s} = \frac{f_p + f_s}{f_s} = \frac{f_{out}}{f_s} \qquad \ldots\ldots\ldots \quad (26)$$

which favors a low signal frequency and a high pump frequency. If we, on the other hand, use the parametric amplifier as a down-converter, the conversion process is accompanied by a loss of

$$L = 1/G$$

where G is given by Equation (26).

The disadvantages with the up-converter mode of operation is that maximum available gain is limited to the ratio of the output (upper sideband) frequency to the input signal frequency and an undesirable shift of signal frequency occurs.

In order to study the negative resistance mode of amplification, we assume that the power in the upper sideband is minimized by an open-circuit for that signal. Thus $P_{s+p} = 0$, and Equations (23) and (24) then give directly

$$P_s = f_s \frac{P_{p-s}}{f_p - f_s} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (27)$$

$$P_p = -f_p \frac{P_{p-s}}{f_p - f_s} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (28)$$

which shows that the pump source supplies power both at the lower sideband ($f_p - f_s$) and at the signal frequency (f_s). In addition, we see that if gain is to occur on the signal frequency ($P_s > 0$), then P_{p-s} must also be positive. In other words, if signal power is to be generated by the amplifier on the signal frequency the amplifier must also be allowed to generate power on the lower-sideband frequency which, in this case, is called the idler frequency (f_i).

$$f_i = f_p - f_s \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (29)$$

This power must be delivered to a load, an external resistive idler circuit is therefore needed. Because this negative mode of operation is characterized by the necessity of <u>three</u> rf-circuits (signal, pump and idler-circuit), this particular parametric amplifier is also called a three-frequency parametric amplifier. Analysis by Heffner and Wade (13) for instance, has shown that for this mode a <u>negative</u> resistance is presented to the signal circuit, which explains the mechanism for producing gain. The amount of gain can be shown to be limited only by the properties of the circuit, leading to a constant $G^{\frac{1}{2}} \times B$ product. In the negative resistance amplifier, gain can therefore be traded for bandwidth, and when $B \rightarrow 0$, when $G \rightarrow \infty$, and the amplifier goes into oscillations. This points out the big difficulty with the negative resistance amplifier, the stability problem. The negative resistance amplifier physically has the output terminals at the same point as the input terminals. The amplifier is therefore a one-port device, and is consequently called a one-port difference frequency amplifier.

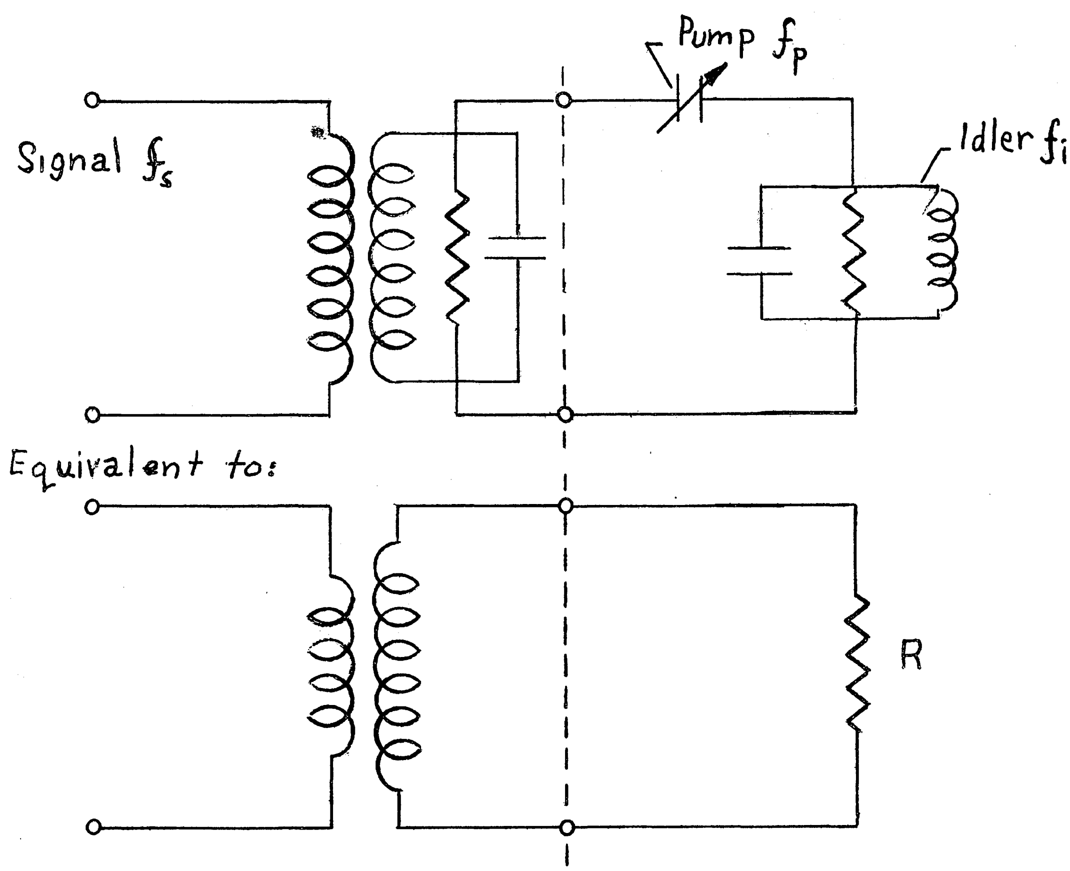
The third main class of operation is, in fact, a mixture of the up-converter and the negative resistance amplifier. Since we have power at the lower sideband according to

$$P_{p-s} = \frac{f_p - f_s}{f_s} \cdot P_s \qquad (30)$$

we can reduce the amount of gain at the signal frequency and extract the useful signal at the lower sideband. In this way one has gain by two processes: first, by the frequency ratio $\frac{f_p - f_s}{f_s}$ and second, by negative resistance. The latter has, however, been reduced, giving less dependence upon matching conditions as compared with the "true" negative resistance amplifier. This mode, sometimes called lower sideband up-converter has some properties (stability) of the up-converter. However, in this amplifier, the output terminals are <u>not</u> at the same point as the input terminals. The amplifier is, therefore, a two-port device, and is sometimes called a two-port difference frequency amplifier. On the other hand, this device is also a three-frequency device.

The last class of amplifiers is the so-called degenerative parametric amplifier which is nothing else than a special case of the negative resistance amplifier obtained by setting $f_p = 2f_s$. This means an idler frequency of $f_i = f_p - f_s = f_s$. Because the idler frequency is equal to the signal frequency, a special idler circuit is not needed; the idler termination can be provided by the signal source. In the case when the signal source is at a low temperature (as an antenna at microwave frequencies in radio astronomy) the noise figure for the degenerate amplifier is lower than for the negative resistance amplifier (this amplifier mode is also called non-degenerate mode of amplification). For the degenerate amplifier there exists an easy model. If the condenser in a tuned circuit is varied at twice the frequency of the signal (for instance by pulling and pushing the condenser plates), then the signal across the condenser increases in amplitude if the pulling apart takes place when the voltage is zero. If the variation of the condenser (the pumping) is shifted 180° in phase, the voltage across the condenser decreases in amplitude, thus making the degenerate amplifier a phase sensitive amplifier.

We have so far introduced four different modes of operation for parametric amplifiers and have illustrated the different types of amplifiers. In Figure 10 are listed the four different modes of operation together with their main properties and their different names. The last list is not supposed to be complete, but it is hoped that it covers the most commonly used names.



MODE	NAMES	R	SPECTRUM	PROTERTY	G
A	1) up-converter 2) sum frequency ampl. 3) upper sideband up-converter	Positive	f_s, f_p, f_{out} — f $f_{out} = f_p + f_s$	stable	$\frac{f_{p+s}}{f_s}$
B	1) negative resistance 2) three-freq. ampl. 3) one-port diff. freq. ampl. 4) non-degenerate ampl.	Negative	f_s, f_i, f_p — f $f_i = f_p - f_s$ $f_{out} = f_s$	unstable	
C	1) lower sideband up-converter 2) two-port diff. freq. ampl. 3) three freq. ampl.	Negative	f_s, f_{out}, f_p — f $f_{out} = f_p - f_s$	somewhat stable	
D	1) degenerate ampl.	Negative	$f_s\, f_i$, $f_p = 2f_s$ — f $f_{out} = f_s$ $f_p = 2 f_s$	unstable	

Figure 10. Different modes of operation for parametric amplifiers.

The time-variable reactance element may either be a condenser (back-biased semiconductor diode), a ferrite material, or an electron beam. The noise figure of a parametric (or reactance) amplifier should be equal to 1 if there were no losses in the circuits, and if the idler was terminated at 0° K. The diodes always have a little loss and so do the tuned circuits, which produces a finite noise temperature. Although diode parametric amplifiers are the most commonly used types of amplifier, except for the up-converter (which in practice can not be used at microwave frequencies) they suffer from lack of stability. The most promising type of amplifier for frequencies below 2 kMc seems to be the electron-beam-parametric amplifier, which is both a unilateral device (because of the properties of the electron beam) and an impedance insensitive device.

B. Masers

The maser amplifier which has obtained its name from the first letters of the phrase "Microwave Amplification (by) Stimulated

Emission (of) Radiation", makes use of paramagnetic properties of certain crystals. The main difference between the existing microwave amplifiers (klystrons and traveling wave tubes) and the maser amplifier are that the latter uses the properties of bound electrons rather than free electrons, in addition to its use of rf-energy as a source of power for amplification rather than dc-energy. (14, 15)

It is known that electromagnetic waves can interact with elementary particles by virtue of changes in the internal energy of the particles. (These particles may be uncharged atoms and molecules, or charged ions.) A molecule consists of an assembly of electrons and atomic nuclei. These can assume only the motion and orientations which yield a discrete set of energies; that is, the energy of the system is quantized. A molecule can interact with electromagnetic radiation by making a transition from the one energy level to another. If the final internal energy exceeds the initial energy, the energy difference must be supplied by the electromagnetic field by the process of <u>absorption</u> of radiation. Conversely, when energy is given to the radiation field, the process is one of <u>emission</u> of radiation and the final energy state of the molecule must have <u>less</u> energy than the initial state. (The radiation energy is also quantized in energy so that it can only exchange descrete amounts of energy with interacting molecules.) If the energy in an upper and lower state is W_m and W_n respectively, then the frequency $f_{m,n}$ of radiation is given by the relation

$$h \cdot f_{m,n} = W_m - W_n$$

where h is Planck's constant.

Not all transitions between energy states of a molecule are possible. Theory predicts and experiment confirms that only certain transitions can occur, thus "selection rules" can be formulated. Paramagnetic ions, for example, present the possibilities of suitable spaced energy levels for making radiation at microwave frequencies possible. The particles possess steady magnetic moments which will attempt to align themselves in the direction of an applied dc-magnetic field. The energy levels of the elementary magnets in the applied field depends upon the magnitude of the field and upon the angle between the vector of the field and the vector of the magnetic moment. Only certain energy values are allowed which correspond to certain angles between the vectors. Since transitions between particular energy levels are permitted, we can choose a suitable paramagnetic substance and a suitable value of dc-magnetic field to have the transition frequencies fall in the microwave range.

Let us now assume that the molecule has two energy states, an upper state and a ground state, between which transitions are permitted. Theory shows that when the molecule is in the upper state there is a definite probability that after a period of time the molecule will return to the ground state with the emission of radiation. This probability has two components, one which is constant, and another which is linearly dependent upon the energy density of radiation at the transition frequency incident on the molecule. Therefore, the presence of radiation at the transition frequency <u>increases</u> the probability for the emission of radiation by the molecule in the upper state. Therefore, if we have an assembly of molecules in the upper of the two energy states, we have a potential source of electro-magnetic energy gain. Molecules in the lower of the two energy states are found to absorb energy at the transition frequency with a probability equal to the varying probability component (giving rise to the so called <u>stimulated</u> emission). Therefore, in order to obtain a net gain for the incoming radiation, we must insure that there is a large excess of molecules in the upper state and maintain that excess by some means. Figure 11 shows an energy-level-digram for a common maser material, ruby, and the relative positions of the energy levels are shown as a function of the applied dc-magnetic field. If a rf-power of frequency f_s (pump frequency) is supplied to the crystal, where f_s corresponds to the transition from level 1 to 3, then many electrons will be pumped up into level 3.

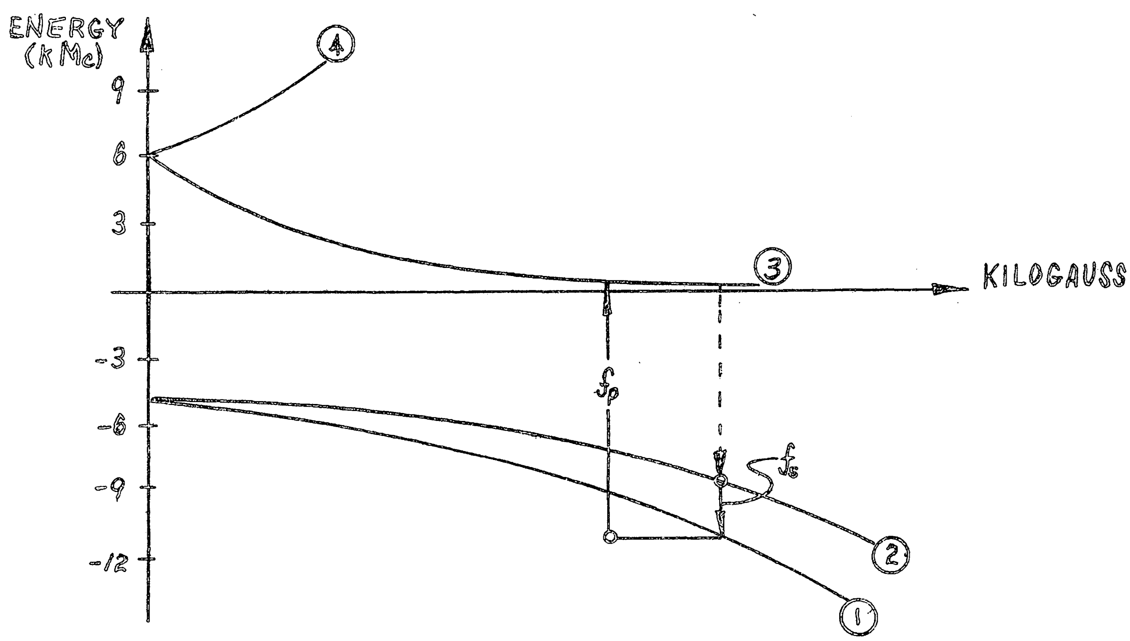


Figure 11. Energy levels for ruby.

Some of the electrons will jump spontaneously from 3 to 2, keeping that level populated. Now, if there are more electrons in level 2 than in the other levels, stimulated emission will exceed absorption and amplification is possible by stimulated emission from 2 to 1. From the figure we also see that by changing the magnetic field and the pump frequency, the idler frequency and the signal frequency are automatically changed. This gives a method of tuning the maser.

Efficient and practical maser operation can only be obtained at temperatures near absolute zero, so the device is usually immersed in liquid helium (4° K). This 4° K temperature produces nearly all of the excess noise temperature of the maser amplifier itself; but unfortunately, losses in electronic component at the input (waveguides, circulators, cavities) increases this temperature to about 10-20°K.

An important parameter for a low noise front-end device is the voltage gain-bandwidth-product, $G^{\frac{1}{2}} \times B$. We have seen earlier that we want a large bandwidth in order to obtain predetection integration, and we want a high gain in order to decrease the back-end receiver's influence on the total system noise. If the receiver following the low noise front end has a noise temperature of T_{FE}, then the total receiver temperature is given by

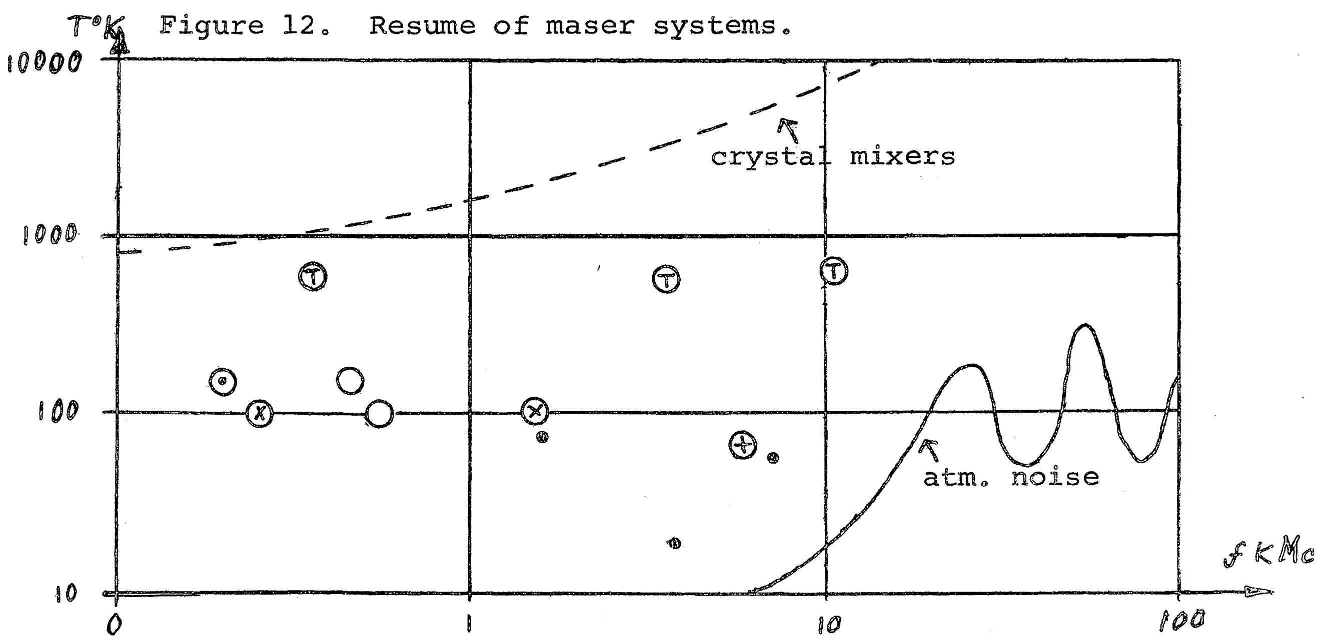
$$T_R = T_{FE} + \frac{T_{BE}}{G_{FE}}$$

If we have a fairly conventional back-end receiver having a temperature of 1000-1500° K, then a front end gain of 20 db (100 times) lowers the back-end temperature to 10-15° K. If we want a bandwidth of 10 Mc, the front end should have a $G^{\frac{1}{2}} \times B$ product of 100 Mc (G measured in times and B measured in Mc). The following list gives important parameters of maser systems which have been in operation in the USA during the last 3 years. System temperatures are only given for masers actually used in connection with an antenna.

The Figure 12 summarizes what has been accomplished as of today with traveling wave tubes, parametric amplifiers, and masers. It should be pointed out that the noise temperatures for the parametric amplifiers listed are amplifier temperatures only; for the masers, the total system temperatures are listed.

NAME	fkMc	BMc	Gdb	$G^{\frac{1}{2}}$ x B	T syst. in °K
NRL	10	5	20	50	87
Ewen-Knight	10	10	20	100	--
Michigan	8.7	30	20	300	75
Airb. I. L.	0.85-2.0	50	23	100	--
Harvard	1.4	--	--	20	84
Hughes	9.3	--	--	100	--
Bell T. Lab	6	25	26	500	20

Figure 12. Resume of maser systems.



(T) traveling wave tube
(•) down-converter
O up-converter
(+) cooled parametric amplifier
(X) electron beam parametric amplifier (Adler-tube)
• maser systems (including antenna noise)

Figure 13. Present status of low noise systems.

In principle, parametric amplifiers and masers have solved the receiver noise problem. The stability, reliability, and ease of operation of these instruments are, however, much poorer than is required for most applications in radio astronomy. For the maser, the system noise consists mostly of input hardware noise and antenna noise. This points out the next obvious step in the development of receiving systems for radio astronomy; that is, the lowering of loss for microwave components like circulators, isolators, and transmission lines while making a new attack on the antenna noise problem.

REFERENCES

(1) J.J. Freeman, Principles of Noise. T. Wiley & Son.

(2) D.A. Bell, Electrical Noise. D. Van Nostr & Co.

(3) Van der Ziel, Noise

(4) H.T. Friis, Noise Figures of Radio Receivers, PIRE 32, (1944), p. 419.

(5) S.O. Rice, Mathematical Analysis of Random Noise, Bell System Technical Journal, 23 and 24, (1944).

(6) R.H. Dicke, The Measurement of Thermal Radiation at Microwave Frequencies, Rev. Science Instruments, 17, July, 1946.

(7) S.J. Goldstein, A Comparison of Two Radiometer Circuits, PIRE, 43, (Nov. 1955). See also PIRE, 45, (March, 1957), pp. 365-66.

(8) J. Galejs, Comparison of Subtractor-type and Multiplier-type Radiometers, PIRE, 45, (Oct. 1957), pp. 1420-22.

(9) W. Selove, A DC-Comparison Radiometer, Rev. Sci. Instr. 25, (Feb. 1954), pp. 120-22.

(10) P.D. Strum, Considerations in High Sensitivity Microwave Radiometry, PIRE, 46, (Jan. 1958), pp. 43-53.

(11) F.D. Drake, H.I. Ewen, A Broad-Band Microwave Source Comparison Radiometer for Advanced Research in Radio Astronomy, PIRE, 46, (Jan. 1958), pp. 53-60.

(12) Manley, Rowe, Some General Properties of Non-linear Elements, PIRE, 44, pp. 904.

(13) H. Heffner, G. Wade, Gain Bandwidth and Noise Characteristics of the Variable Parameter Amplifier, Journal Appl. Physics, 29, pp. 1321.

(14) J.R. Singer, Masers, 1959. Wiley & Son.

(15) G. Troup, Masers, 1959. Meuthen.