UNCLASSIFIED

NRL Report 6024

Parametric Amplifiers and Upconverters

ALAN C. MACPHERSON

Solid State Electronics Branch
Electronics Division

May 5, 1964

U.S. NAVAL RESEARCH LABORATORY
Washington, D.C.

CONTENTS

Copies available from Office of Technical Services
Department of Commerce – $1.25

FOREWORD

This report is an expanded version of material prepared by the author for the Capitol Radio Engineering Institute of Washington, D.C. Some of the material on negative resistance amplifiers and some of the remarks on making a rational selection of a receiver for a given system have not appeared in the literature, but the rest of the report is tutorial. It is assumed that the reader has some familiarity with the most elementary principles of linear circuit theory.

ABSTRACT

The purpose of this report is to place in proper perspective the role that the modern semiconductor diode low-noise amplifier and converter have in the amplification of extremely small signals. There is a discussion of the fundamental principles of parametric amplification, and a quantitive treatment which is valid in the limit of very high Q diodes. Design formulas from the literature are given and discussed for the more practical case of finite Q diodes. An engineering discussion of noise in amplifying systems is presented, because the most important characteristic of parametric amplifiers and converters is low noise. Since the parametric amplifier is built up from a two-terminal negative resistance, it is qualitatively different from the conventional amplifier (such as the vacuum tube triode and traveling wave tube). For this reason some basic material on the properties of negative resistance amplifiers is included. Several typical parametric amplifiers are presented and discussed. Information is included which may be of assistance to a systems designer who is weighing the merits of the parametric amplifier against some other amplifier in a system design. Several examples of system design are given. Finally, some recent experimental data are discussed which suggest that in the future the parametric amplifier will completely prevail over the maser.

PROBLEM STATUS

This is a final report on one phase of the problem; work on other phases continues.

AUTHORIZATION

NRL Problem R08-33
Project RR 008-03-46-5657

Manuscript submitted October 7, 1963.

PARAMETRIC AMPLIFIERS AND UPCONVERTERS

INTRODUCTION

Since the early days of radar there has been a strong demand for more sensitive microwave receivers. Because microwave receiver sensitivity was, at that time, limited by noise generated inside the receiver, it is equally correct to say that there was a strong demand for low noise figure receivers.

The most popular receiver in the microwave region was the combination of a crystal mixer and an i-f amplifier, in which most of the noise was due to the crystal mixer. The input frequency of such a receiver might be 10,000 Mc. In the first stage, the crystal mixer stage, this low level signal would be mixed with a high level local oscillator signal at say 10,030 Mc, and the difference frequency of 30 Mc would be passed on to a vacuum tube i-f amplifier with a bandwidth of a few Mc. Since World War II there has been some improvement in these receivers, mostly through crystal improvement, but internal receiver noise is still dominant in limiting sensitivity. That is, the receiver noise is still large compared with the noise brought in from the antenna.

The invention of the maser completely reversed the situation in the sense that the internal noise of a maser amplifier is so low as to be completely negligible in almost any system. Thus, it is quite true that the noise problem in microwave receivers has been completely solved if one is willing to put up with the disadvantages of the maser. From an engineering point of view, these disadvantages are rather serious. The maser is cumbersome and expensive and requires both a high quality magnet and extremely low temperatures.

The various forms of the parametric amplifier (PA) employing semiconductor diodes offer the engineer a relatively simple means of obtaining a low noise figure. The disadvantage is a somewhat higher noise figure than that obtained with the maser.* Commercially available parametric amplifiers have noise figures considerably lower than the crystal mixer and i-f amplifier combination.

Nomenclature

In this discussion, the term parametric amplifier will be used as a generic name for a class of amplifying and frequency-converting devices which utilize the properties of nonlinear or time-varying capacitances obtained through the use of semiconductor diodes. Thus, vacuum tube parametric devices and magnetic parametric devices will be excluded. Other names which have been used instead of parametric amplifier are variable parameter amplifiers, reactance amplifiers, and mavar (microwave amplification by variable reactance).

The two types of parametric amplifiers which will receive the most attention will be called the upconverter, which is an amplifier and frequency upshifter, and the negative resistance parametric amplifier (NRPA), which is a one-port amplifier (no frequency shift) employing a negative conductance or resistance.

*Some very recent work using liquid nitrogen cooling has yielded PA noise figures comparable with maser noise figures.

The term negative resistance amplifier will be used to refer to any amplifier which operates as a one-port negative resistance or conductance. Examples of negative resistance amplifiers are the maser, the tunnel diode amplifier, and the negative resistance parametric amplifier.

History

Parametric amplification has been understood, in principle at least, since the time of Lord Rayleigh (1883). Indeed, a child pumping a swing by shifting his body twice during each cycle of the swing is employing the principle.

During World War II, H. Q. North observed conversion gain in what appeared to be a conventional microwave crystal mixer (special diodes were used), but the noise was high and the work was dropped, partially because it was believed that large noise was an inherent property of such negative resistance devices.

What was needed was some evidence that the parametric principle was not inherently noisy and a practical device to supply the nonlinear capacitance at microwave frequencies.

Van der Ziel (1) was the first to point out the low noise possibilities in parametric amplification, but it was nearly ten years before practical microwave parametric amplifiers were built using modern semiconductor diodes.

In the early days of designing parametric amplifiers most of the emphasis was placed on low noise and simplicity. As a result, only simple tuned circuits were employed, and the bandwidths were quite low. As will be explained later, some very recent work proves conclusively that bandwidths of the order of 20 percent are feasible without increasing the noise figure significantly.

NOISE

Introduction

If it were not for noise, arbitrarily low level echoes could be detected by any radar receiver with sufficient gain. However, in practice very small signals are obscured by noise arising from inside the receiver or brought in by the antenna along with the signal. The primary concern here is with noise originating in the receiver. Such things as hum and microphonics are not considered because they can, in principle, be reduced to any desired value. These are not intrinsic sources of noise. If the noise brought in on the antenna is large, whatever its origin, then there is little incentive to develop low noise receivers. There would be little point, for example, in developing a super low-noise amplifier for a home phonograph system because the limiting factor here is needle scratch rather than amplifier noise.

In the range above a few hundred megacycles, the antenna or source noise is low in many systems, and there is therefore good motivation for developing low noise receivers. The first strong demand for low noise receivers arose during World War II in connection with radar. The receiver noise was approximately ten times greater than the antenna noise, so that a twofold reduction in receiver noise meant nearly a twofold increase in sensitivity.

It is a fact that in the microwave region the PA is the least complicated low noise receiver. The maser is the only amplifier with lower noise capabilities.

The reader may have noticed the almost interchangeable use of the terms receiver and amplifier. The noise properties of an ordinary receiver, in which there is considerable gain before detection, are determined by the noise properties of the amplifier. Therefore, as far as noise is concerned, the two terms can be used interchangeably. It is fortunate that this is so, because amplifiers are linear devices and can be handled much easier than nonlinear ones. The receiver (including the detector), on the other hand, is basically nonlinear because the detection process is nonlinear. It should be emphasized that a mixer is a linear device, and therefore is included in the general category of amplifier. The term first detector is often used to describe the mixer in a superheterodyne receiver. This is misleading because the signal is not detected. The so-called first detector simply shifts the carrier and sidebands by the same amount in frequency and preserves their relative amplitudes. The gain may be either greater or less than one.

To appreciate the virtues of the PA it is necessary to understand something about noise in linear systems. In this section the following subjects will be covered: The engineering characterization and measurement of noise in two- and four-terminal networks, circuit calculations involving noise currents and voltages, combinations of noisy networks, the relationship of noise figure to system sensitivity, and physical sources of noise.

How Noise Is Described

Consider any linear two-terminal conductance G. Connect this conductance through a filter of center frequency f (let $f >> 1$) and bandwidth of 1 cycle to a zero impedance ammeter which is noiseless and will produce a continuous record of current. This record might look something like Fig. 1. It seems fair to say that our conductance is shunted by some kind of a random current generator. We cannot characterize this current by its average value, which is apparently zero, nor can we use its amplitude, which is undefined because it is of a random nature. As a matter of fact, a complete characterization of such a record is very difficult, and far too complicated for engineering calculations. However, a description which is adequate for our purposes can be made as follows: Call the 1-cycle-bandwidth time-dependent short-circuit current $i(t)$. Consider a very long record. Square it, giving $[i(t)]^2$. Next average this value over the record. This is called the mean square value of the noise currents per cycle, $\overline{i^2}$. The product of $\overline{i^2}$ and some bandwidth is analogous to the mean square value of a sine wave of current and is directly proportional to the heating effect. This expression can be written more formally as the limit of

$$\frac{1}{T} \int_0^T [i(t)]^2 \, dt$$

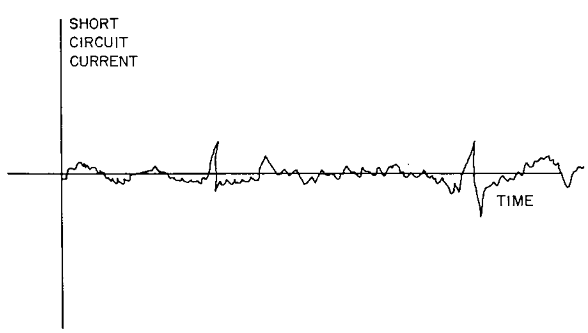


Fig. 1 - Typical short circuit current from a noisy one-port

as $\tau \to \infty$. For our purposes, we can assume that the random noise source is described by the value of $\overline{i^2}$. In general, we must expect $\overline{i^2}$ to be frequency dependent; that is, if the experiment is repeated with the filter centered at a different frequency we will obtain a different result for $\overline{i^2}$. However, over a narrow band $\overline{i^2}$ is constant; thus, for our purposes we can assume that $\overline{i^2}$ is frequency independent. It should be noted that $\overline{i^2}$ has been defined for a bandwidth of 1 cycle. It has been shown that when the bandwidth is doubled the mean square current is also doubled. Thus, in general, the mean square current is given by $\overline{i^2}\Delta f$, where Δf is the bandwidth. Note that the units for $\overline{i^2}$ are current squared per cycle.

There is a very important theorem concerning the value of $\overline{i^2}$ for any conductance which is in thermal equilibrium with its surroundings at temperature T. An exact definition of thermal equilibrium is rather difficult to state, but perhaps it would be helpful to point out that the following two-terminal networks are not at thermal equilibrium: a semiconductor diode drawing current, output terminals of a triode amplifier or mixer, a carbon or metal resistor carrying dc current, and input or output terminals of a fired TR tube. The following could be assumed to be at thermal equilibrium: a carbon or metallic resistor carrying no current which has been held at a relatively constant temperature for a long time, a semiconductor diode at zero bias, and a waveguide matched load. Nyquist has shown that for these latter devices $\overline{i^2} = 4kTG$, where k is Boltzmann's constant. This relationship is completely independent of the physical nature of the device.

Multiply both sides of the above equation by $1/(4G)$. Then

$$\frac{\overline{i^2}}{4G} = kT . \tag{1}$$

The left side of Eq. (1) is analogous to the available power from a sine wave generator, and thus the available noise power per unit bandwidth from a one-port or two-terminal conductance at thermal equilibrium of any G is equal to kT. By multiplying both sides of Eq. (1) by Δf we see that the available noise power from a two-terminal conductance at thermal equilibrium of any G is equal to $kT\Delta f$.

For two-terminal conductances which are not at thermal equilibrium, there is still some value for $\overline{i^2}$ which describes its noisiness. This can be written as $\overline{i^2} = 4kT_nG$ and we can consider this a definition of an equivalent noise temperature T_n. Normally T_n is larger than the ambient temperature of the device.

Circuit Calculations

Consider the following simple problem. Two conductances G_1 and G_2, both at room temperature T_0, are connected in parallel and then fed through a lossless filter of bandwidth Δf. What is the short-circuit mean-square noise current at the output of the filter? This problem involves combining the noise currents of the two conductances to get the resultant mean square current at the output. The proper way to do this is to calculate separately the effect due to each noise source and add the result. Thus, the mean square current contribution from noise source one is $kT_0 G_1 \Delta f$ and that from noise source two is $kT_0 G_2 \Delta f$. The sum of these is $kT_0 (G_1 + G_2) \Delta f$. Of course, we could have obtained this result immediately by observing that the shunt combination of the two resistors (both at thermal equilibrium) is simply a resistor at thermal equilibrium whose conductance is $G_1 + G_2$.

As a second problem we will calculate the available noise power (sometimes called simply the noise) of the network of Fig. 2. There is an easy way to do this problem. If the reactance is short-circuited, the noise becomes $\overline{i^2}(R/4)\Delta f$ where $R = 1/G$. Since we are interested in the available noise power, the reactance, being lossless will not affect

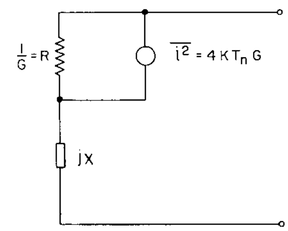


Fig. 2 - Linear noisy network. The resistor is at noise temperature T_n.

the result. By substituting $\overline{i^2} = 4kT_nG$ into the above equation the noise becomes $kT_n\Delta f$. This is a useful result because it indicates that the noise from a two-terminal network of complex impedance depends only on the noise temperature of the resistive component. Thus the noise from any linear two-terminal network in thermal equilibrium at temperature T is given by $kT\Delta f$. Note that this theorem refers to the available noise power. The short-circuit mean-square noise current is not independent of the reactive component.

Let us do the above problem in a different way, so that we can learn something about handling noise currents in circuit problems. To solve the problem we need to know the short-circuit current and the equivalent conductance at the output, G_o. The output admittance is $(R - jX)/(R^2 + X^2)$. Therefore $G_o = R/(R^2 + X^2)$. Now we calculate the short-circuit current i_{ss} just as if it were sinusoidal. Thus

$$i_{ss} = \frac{i/(jX)}{1/R + 1/(jX)} .$$

We wish to calculate $\overline{i_{ss}^2}$, that is, we wish to take the mean square of both sides of the above equation. In so doing we make use of the rule that the mean square value of a complex number is equal to the absolute value squared of that number. Thus

$$\overline{i_{ss}^2} = \frac{\overline{i^2}R^2}{X^2 + R^2} .$$

It is left to the reader to prove that the desired result, $\overline{i_{ss}^2}\,\Delta f/4G_o$, is equal to $\overline{i^2}\,(R/4)\,\Delta f$.

Of course, we can solve the above problem using noise voltages as well as noise currents. Obviously, a conductance G with a noise current $\overline{i^2}$ can also be described as having an open-circuit mean-square noise voltage $\overline{v^2}$, which is given by $\overline{i^2}/G^2$. Thus the available noise power $kT_n\Delta f$ is equal to $\Delta f\,\overline{v^2}/(4R)$ where $\overline{v^2} = 4kTR$.

Depending on the problem, either the noise current or the noise voltage method may be more convenient. For example, the noise voltage method is usually simpler in solving problems involving series connections.

Noise Figure

The next problem to be considered is noise in four-terminal linear networks. Examples of such networks are amplifiers, attenuators, sections of waveguide, T and Π networks, mixers—anything that is linear and has an input and an output. First we wish to define noise figure F (refer to Fig. 3). The impedance Z at temperature T acts as the source impedance for the noisy four-pole being considered. The output of the four-pole network is matched into a lossless filter with bandwidth Δf centered at frequency f. This filter is necessary because the noise properties of the four-pole are in general frequency dependent, and our characterization will be valid only at one frequency. The power

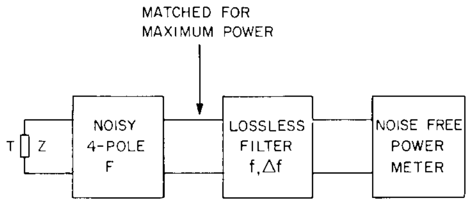


Fig. 3 - Circuit used in the definition of noise figure. The power meter records noise from Z and from the noisy four-pole or two-port.

meter, being noise-free, records only the noise power delivered from the four-pole. Because of the matching, this is really the available noise power from the four-pole, or simply the noise from the output of the four-pole.

For example, if the four-pole is an amplifier, it must be adjusted to some operating point (B^+, gain, etc.) and be kept at that same point during the experiment.

In general, the noise fed to the power meter will be the sum of the noise arising within the amplifier and the amplified noise from Z. In defining F we do not want to associate the latter noise with the four-pole. Theoretically, we can eliminate this noise by cooling Z to absolute zero. Assume that this has been done and call the resultant noise AN_n, where A is the available power gain of the four-pole. Evidently, N_n is the noise caused by the four-pole when it is referred to the input, and it is a good measure of the noisiness of the four-pole except for the fact that it depends on the bandwidth. An increase in the passband of the filter will increase N_n. Since N_n is a noise power, we can set it equal to $kT_E\Delta f$, and consider this to be a definition of an effective noise temperature T_E for the four-pole.

Noise figure F is defined by

$$F - 1 = N_n/(kT_0\,\Delta f)\,. \tag{2}$$

From the definitions of T_E and F we can derive that $T_E = (F - 1)\,T_0$. Thus F and T_E are equivalent ways of specifying the noisiness of a four-pole. If one is known, the other can always be calculated. It can be stated that $F - 1$ is the ratio of network noise to source noise. It should be kept in mind that for purposes of definition the source must be at T_0. This does not mean that F is inapplicable when the source temperature is different from T_0. Some authors have defined an effective noise figure which is a function of the source temperature, but this is unnecessary.

As a simple example, calculate the noise figure of the four-pole formed from a single resistor at temperature T (Fig. 4). Since this is a series circuit, noise voltages will be easier to handle than noise currents. Noise voltages are introduced in Fig. 5. Because F has been defined in terms of available noise power, it is convenient to terminate the four-pole so that a power match will exist. This is also indicated in Fig. 5. First calculate the delivered noise AN_n and assume that R_g is noiseless. The noise voltage across

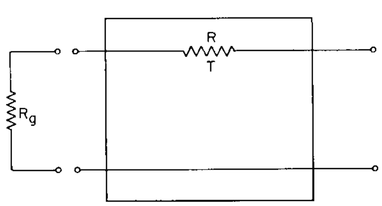


Fig. 4 - Calculation of the noise figure of a series resistance at temperature T

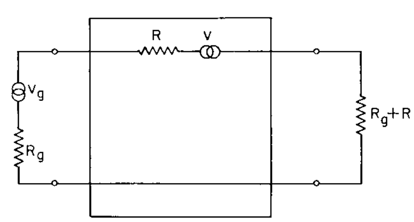


Fig. 5 - Same as Fig. 4, but noise voltages and a matched termination have been introduced

the load is $v/2$; therefore the delivered noise power is $\Delta f\overline{v^2}/4(R+R_g)$. Clearly, the noise caused by the source alone (i.e., the delivered noise if the four-pole were noiseless), is $\Delta f\overline{v_g^2}/4(R+R_g)$. Thus, $F-1$ equals $\overline{v^2}/\overline{v_g^2}$. At this point, we should note two facts. First, as long as we are consistent, it does not matter whether we evaluate the source and network noise at the input or at the output; and second, the ratio of the network noise to the source noise does not depend on the terminating impedance since the terminating impedance cancels out. This means that we can modify our definition of F and speak of noise power delivered to any load instead of available noise power. In fact, since the terminating impedance cancels out, we can calculate (at the output) the ratio of the open-circuit mean-square noise voltage to the open-circuit mean-square noise voltage caused by the source. This ratio equals $\overline{v^2}/\overline{v_g^2}$. Since $\overline{v^2} = 4kTR\Delta f$ and $\overline{v_g^2} = 4kT_0 R_g \Delta f$ it follows that $F-1 = RT/R_g T_0$ and $F = (R_g T_0 + RT)/R_g T_0 = 1 + (T/T_0)(R/R_g)$. If $T = T_0$, that is, if R is at room temperature, $F = (R_g + R)/R_g$. Note that the noise figure of a four-pole depends on the source impedance but not on the load impedance. The dependence of F on R_g in this simple network is indicated by the solid line of Fig. 6. The noise figure approaches infinity for very small R_g and approaches one for very large R_g. The latter result is not very exciting, because the four-pole has a gain less than unity. If we define F^0 as the lowest noise figure and R_g^0 as the particular value of R_g which corresponds to F^0, then $F^0 = 1$ and $R_g^0 \to \infty$. It can be shown that for the most general four-pole the dependence of F on R_g is of the form given by the dotted line; note that R_g^0 is finite and F^0 is greater than unity. This result is extremely important since it is often possible (particularly in narrow-band circuits) to tune at the input of an amplifier so that $R_g = R_g^0$ and so that the minimum noise figure F^0 is obtained. In the technical and commercial literature values of noise figure are often quoted without stating the source impedance. In such cases it is safe to assume that reference is being made to F^0.

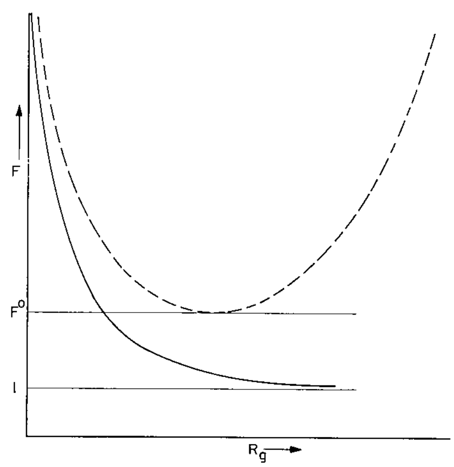


Fig. 6 - Dependence of the noise figure on the source impedance. The solid line is for the circuit of Fig. 4; the dotted line is for a general two-port

System Sensitivity

It should be clear that the noise figure plays an important role in determining the sensitivity of receiving systems that utilize an amplifier in the first stage. This does not imply, however, that a universal definition of sensitivity can be stated. The exact role that F plays in the sensitivity of a system will depend on the details of that system. In some systems, for example, it is not advantageous to reduce F below a certain value. To illustrate these ideas it will be useful to construct meaningful definitions of sensitivity for two important idealized situations in which low noise receivers are employed:

1. <u>The Point Source</u> — Consider a receiver used to detect a signal from a distant communications transmitter or radar target. When compared with the dimensions of the

transmitting antenna or the radar target the receiving antenna pattern is very broad. The receiving antenna, therefore, must be looking not only at the signal source but also at a background which almost completely determines the impedance of the antenna as well as the noise temperature associated with the antenna. The effective noise temperature of the antenna or the effective temperature of the radiation resistance of the antenna will be an average of the space at which the antenna is looking. For example, if an antenna is aimed at the sea, it is reasonable to assume that its noise temperature would be equal to the temperature of the sea.

This antenna noise must be considered a disturbing factor in the attempt to detect the signal from a distant transmitter or radar target. However, the important point is that if the distant transmitter or radar target were suddenly removed, the effect on the antenna impedance and temperature would be negligible. This example then is that of a point source imbedded in a thermal noise background.

The total noise power referred to the input of the receiver is $N_n + kT\Delta f$ where as before N_n is the network or amplifier noise and T is the noise temperature of the antenna. Also N_n is given by Eq. (2); thus, the total noise equals $k\Delta f\,[(F - 1)T_0 + T]$. It follows that if the total noise is doubled, the signal to noise ratio will be halved; or if the total noise is halved, the signal to noise ratio will be doubled. It seems reasonable to define the sensitivity S of the system as the inverse of this quantity. Boltzmann's constant k and the bandwidth Δf can be dropped, because they are not relevant. We then have

$$S = \frac{1}{(F - 1)T_0 + T} . \tag{3}$$

For the case in which the antenna temperature T is equal to the room temperature T_0, then

$$S = \frac{1}{FT_0} . \tag{4}$$

This is one case, at least, in which the sensitivity is dependent upon F alone. The sensitivity of a radar or communications receiver whose antenna is at temperature T_0, is inversely proportional to the noise figure.

As a second case assume T is well below room temperature. This is possible because at certain microwave regions parts of the sky are cold. But if we exaggerate the situation and assume that $T = 0$ we have the extremely interesting result that S is inversely proportional to $F - 1$. In other words a receiver of noise figure 1.05 is twice as sensitive as a receiver of noise figure 1.1.

As a final case, assume that T is very large. As $T \to \infty$, $S = 1/T$ and the sensitivity is independent of the noise figure of the receiver.

2. The Extended White Noise Source — Assume that one points a radio astronomy antenna at a broadband radio source of noise temperature T_s (for example, the sun) and wishes to measure T_s. If the antenna beam is too wide, noise sources around the object of interest will be picked up and the situation will become complicated. If, on the other hand we assume that the antenna beam is narrow and includes only the source of interest, a reasonable definition of sensitivity can be given: Sensitivity is defined to be the ratio of the available signal power to the available amplifier noise power at the amplifier output. In this definition atmospheric absorption is neglected, and T_s is assumed to be independent of frequency over the bandwidth of the receiver. The noise from the source is now the signal (note that the signal is what we choose it to be) and it is degraded as it passes through the amplifier. Again using Eq. (2) we will have $S = \Delta f AkT_s/[\Delta f AkT_0(F - 1)]$, where A (which cancels out) is the available gain of the receiver and

$$S = T_s/[T_0(F-1)] .$$

Thus

$$S \propto \frac{1}{F-1} .$$

This simple relationship indicates why radio astronomers are so intensely interested in super-low noise receivers. For example a receiver of noise figure 1.01 is ten times better than one of noise figure 1.1.

But what does the discussion of system sensitivity have to do with parametric amplifiers? This question will be answered more fully later but a partial answer will be given at this point. When a receiver problem for some particular system is formulated, the engineer may be asked to select the most appropriate type of receiver. It would be foolish to select the maser for an application which did not really require a very low noise figure. Thus the engineer must know something about the relationship between noise figure and system sensitivity if he is to choose properly.

Tandem Connection of Two-Port Networks

The following problem frequently arises. Assume that noise figures, F_1 and F_2, of two-port number 1 and of two-port number 2 are known. Suppose that they are connected in tandem. What is the resultant overall noise figure F? The situation is indicated in Fig. 7. It can be shown that the correct answer is

$$F = F_1 + (F_2 - 1)/A_1 \tag{5}$$

where A_1 is the available power gain of the first two-port. The three quantities F_1, A_1, and F_2 depend on source admittance; this means that the proper values for F_1 and A_1 in Eq. (5) are those obtained when the source admittance for two-port number 1 is Y_s. Likewise F_2 must correspond to the source admittance which is the output admittance of the first two-port. The first term in Eq. (5) is the noise contribution from the first two-port; the second term (always positive) is the noise contribution from the second two-port. The effect of the second two-port can only be a degrading one. Note that the noise contribution from the second amplifier depends on the available gain from the first amplifier. Evidently if F_1 and F_2 are approximately equal, and if A_1 is larger than say 10, the second amplifier will contribute less than 1/10 of the noise. Of course, if A_1 approaches infinity and F_2 is finite, then $F \approx F_1$; that is, the second amplifier will make no contribution to the overall noise. This conclusion can be intuitively understood. Assume for purposes of simplicity that the first and second amplifiers are identical. Then, referred to the input of each, the noise caused by each amplifier is the same. But referred to the input of the second amplifier, the noise caused by the first amplifier has been amplified by the gain of the first amplifier. If the gain is large, the second amplifier noise is swamped out. In the practical receivers, the gain of the first stage is usually considerable. This explains why so much attention is given to the first stage design of low-noise amplifiers. The old fashioned microwave radar receiver which consists of a mixer followed by an i-f amplifier at say 30 Mc is a notable exception because F_1 is large (order of 10), and the

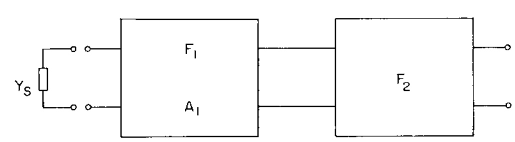


Fig. 7 - Noise figure of a pair of two-ports connected in tandem

mixer is actually lossy (A_1 is of the order of 1/4). Thus, the noise figure of the second stage, which in this case is the i-f amplifier, must be low for a low overall F.

It has already been pointed out that noise figure is a good figure of merit for the noise behavior of four-terminal networks. This statement should be amended to state that noise figure is a useful figure of merit provided that the network has high gain. It is, after all, the overall noise figure of an entire receiver that is important, and it is of small value to have a low noise figure in the first stage if the gain is so low that the noise from the second stage predominates. For example, suppose that an engineer is intrigued by the fact that a network of pure inductances has a noise figure of one, and thus decides to use such a network as a first stage in a receiver. Since power is not lost in such a network the matched power gain or available gain A_1 is also equal to one. Applying Eq. (5), we see that the overall noise figure is equal to the noise figure of the second stage. For a stage of only moderate gain it seems that there should be some sort of noise figure of merit which would properly take into account the gain as well as the noise figure. Such a figure of merit exists; it is called the noise measure and will be discussed later. However, as far as complete amplifiers, composed of many stages, are concerned the gain is always very high; thus the noise figure is a perfectly valid figure of merit.

Some remarks are in order here concerning the laboratory adjustment of receivers for overall minimum noise figure. Refer again to Eq. (5) and Fig. 7, but assume, as is often the case, that there are transformers between Y_s and the input to the first stage, and between the first and second stage.

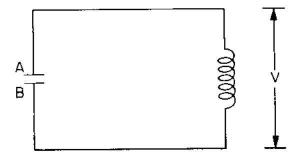
With reference to Eq. (5) it is pertinent to ask how the above system can be adjusted to produce a minimum overall noise figure. The interstage transformer is easily adjusted. We adjust this transformer until the source conductance seen by the second amplifier is the optimum source conductance G^0. Then by definition, F_2 will assume its smallest value. If we assume that this transformer is lossless, this adjustment will have no effect on F_1 or A_1. However, adjustment of the input transformer will affect both A_1 and F_1. We would like to adjust A_1 to be very large (to its maximum value if it exists), and, at the same time, we would like F_1 to equal its minimum value F_1^0. Unfortunately, the maximum value of A_1 is not obtained for that source admittance which minimizes F_1. That is, "tuning for maximum power gain is different from tuning for minimum noise figure." In practice, this problem can be solved in the laboratory by tuning the input transformer for minimum overall noise figure. Since even one measurement of noise figure is time consuming, this can become a laborious task.

QUALITATIVE DESCRIPTION OF PARAMETRIC AMPLIFICATION

The heart of the parametric amplifier (PA) is the capacitance C which varies with time at the pump frequency f_3. The reason for using the symbol f_3 for pump frequency will become clear in a later paragraph. There is, of course, no such thing as a capacitance without losses, but it is a useful idealization and we will first assume that such a capacitor is available. We will see later that the losses in real capacitors constitute the main limitation in PA operation.

We can gain insight into the PA process by considering the circuit of Fig. 8. Assume that this circuit is oscillating at its resonance frequency $f_0 = 2\pi\omega_0$. Since there are no losses, it will continue to oscillate at the same amplitude and frequency for an indefinite period, as shown in Fig. 9. We now wish to describe the pumping process. Suppose that at time t_1 of Fig. 9 we suddenly pull the plates of the condenser apart and hold them at this new separation. This certainly will take mechanical work, and since energy must be conserved, it would appear that we have transferred some mechanical energy into electrical energy in the form of the electric field of the condenser. The voltage across the condenser is given by $V = Q/C$ where Q is the charge on the condenser. If we move the

Fig. 8 - Shunt tuned circuit. The condenser plate spacing can be varied to produce parametric pumping.



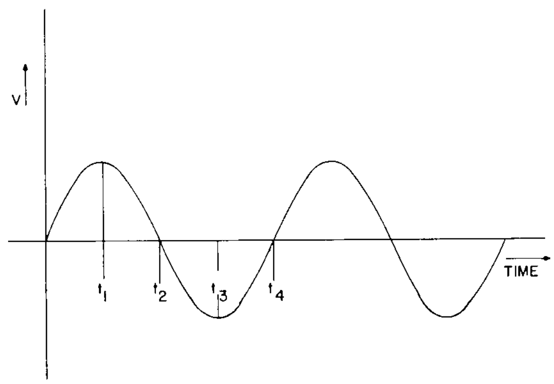


Fig. 9 - Behavior of the circuit of Fig. 8 when there is no pumping

plates fast enough, Q will stay constant but C will decrease (since 1/C is proportional to the plate spacing) and therefore the voltage will make a little jump. If we then wait until time t_2 and quickly restore the plates to their original spacing, work will not be performed since no field exists at this time. At time t_3 we again pull the plates apart causing another voltage jump and at t_4 we again restore the plates to their original position. It is evident that we have arrived at a cyclic process which transfers a net energy to the tuned circuit. Ordinary tuned circuits, which contain losses in the form of positive resistances lose energy in the form of heat to these resistances and thus the amplitude of the oscillation decreases. Here is a case in which a tuned circuit is gaining energy, and it is natural to suggest that this is caused by the presence of negative resistance.

The voltage vs time would now look like Fig. 10. If we were to smooth out Fig. 10 a little, it would look like a sine wave of increasing amplitude, which confirms the above argument. We can make this argument more quantitative by noting that the damped sine

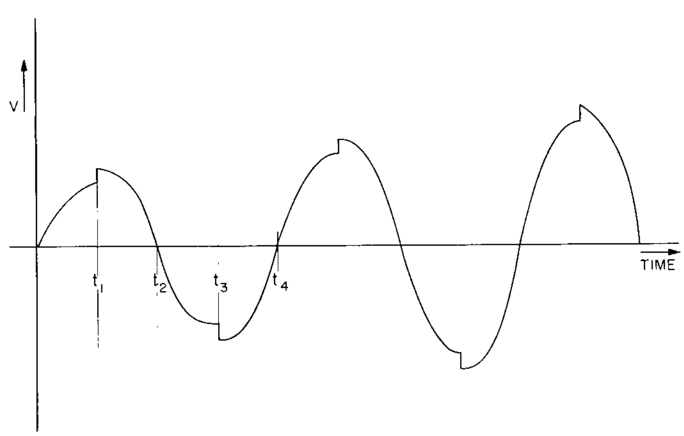


Fig. 10 - Pumped response of the circuit of Fig. 8

wave of a normal tuned shunt circuit is of the form $(\sin \omega_0 t) \exp(-Gt/2C)$. Ordinarily G is positive, but if we let it be negative then we have an exponentially increasing sine wave. In other words, when a shunt tuned circuit is shunted with a negative conductance we obtain a sine wave of increasing amplitude. Thus, the effect of pumping the circuit of Fig. 8 is as if we had placed a negative conductance across the tuned circuit.

It is important to note that the plates must be separated at special times during the resonance cycle of the tuned circuit. That is, the pump frequency must have a definite relationship to the resonance frequency. In the example given, the pump frequency is just twice the resonance frequency. The observant reader will notice, however, that although pumping at twice the resonance or signal frequency is most efficient, it is not really necessary to pump every cycle in order to obtain some increase in amplitude as long as the plates are separated at the right times. Pumping at frequencies lower than the signal frequency is called harmonic pumping.

Once we have established the fact that pumping can produce an equivalent negative conductance, it follows immediately that either an amplifier or an oscillator can be built at 1/2 the pump frequency, depending on the external circuit loading. Even from this simple model, we can see what some of the characteristics of PA's must be. They must be of the negative resistance type (and thus have stability problems), they must be narrow band (at least in single-tuned versions), and we can perhaps even guess that they will be low noise since ordinary pure capacitances do not produce noise.

There is, however, a subtle flaw in the above reasoning. It is absolutely required that there exist an exact relationship between the pump frequency and the signal frequency; for example the signal frequency must be exactly 1/2 the pump frequency. This implies complete knowledge of the signal phase at all times. But ordinarily one does not have such knowledge of real signals. The flaw may be explained in another way. Suppose that the signal is amplitude modulated. The requirement that the pump frequency be twice the signal frequency cannot be met when there are sidebands to deal with. It is therefore clear that what we have been describing is not really an amplifier in the usual sense of the word. It may be described as a lock-in amplifier. The lock-in amplifier does have certain uses, but we are mainly interested in low-noise amplifiers, in the usual sense, so we will not pursue this matter further.

It should be emphasized that the difficulty is not with the circuit of Fig. 8 but rather with the above primitive argument leading to the conclusion that the signal frequency has to be exactly 1/2 the pump frequency. In a more correct treatment, it will be convenient to refer to Fig. 11. Assume that C is pumped at that frequency $f_1 + f_2 = f_3$, which is the sum of the resonant frequencies of the two tanks. (Note that Fig. 8 is a special case of Fig. 11 in which $f_1 = f_2$.) It can be shown that under these conditions, then, looking to the right at AA' at a frequency f_1 we see a pure negative conductance and looking to the left at BB' at frequency f_2 we see a different pure negative conductance. Thus, both tanks oscillate simultaneously at their respective resonant frequencies.

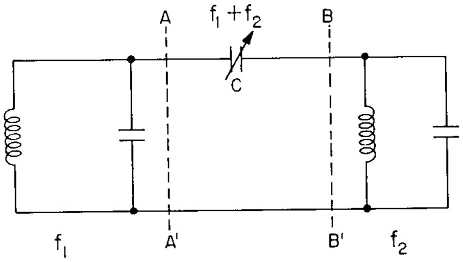


Fig. 11 - Two-tank parametric circuit

To make an amplifier both tanks must be properly loaded. An amplifier can be constructed to operate at either f_1 or f_2. The unused frequency is called the idler frequency. It is important to realize that currents and voltages always exist in the idler tank at the idler frequency. Suppose we choose to construct a negative conductance at f_1. (We will defer a discussion of how two-terminal negative conductances or resistances can be used as amplifiers.) The circuit is as shown in Fig. 12. Note that the idler tank is loaded by the conductance G_I. This is absolutely necessary for parametric behavior. Suppose the pump frequency is adjusted to equal $f_1 + f_2$. We want the signal frequency f_s nearly equal to f_1, but we certainly cannot guarantee that this will be the case; in fact, we must assume that they will differ slightly. A proper solution of this problem shows that a negative conductance will appear across tank 1 provided only that $f_s + f_I$ equals the pump frequency f_3, where f_I is the frequency of currents and voltages appearing in tank 2. Note that since $f_1 \neq f_s$ we have $f_2 \neq f_I$. Note also that we are assuming the circuit is in the nonoscillating condition. The frequency f_I is actually generated by a mixing action between f_s and f_3; thus the condition $f_s + f_I = f_3$ is automatically satisfied. Of course, for significant gain, it is necessary for f_s and f_I to remain within the passband of their respective tanks, so that the device is still narrow-banded. We are now dealing with a negative conductance which can be built into an amplifier.

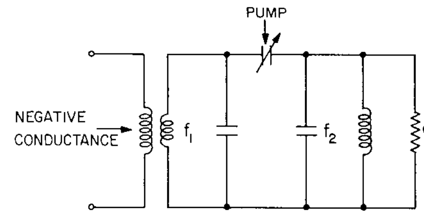


Fig. 12 - The generation of a negative conductance

The above was a physical explanation of the negative resistance parametric amplifier. Unfortunately, a simple physical explanation of the stable upconverter is not possible, and we will need to carry out a quantitative analysis of this device before we can proceed.

PARAMETRIC OR VARACTOR DIODES

In practice one cannot vary the plates of a capacitor fast enough for amplifier operation. An electronic method of varying a capacitance is required. Until recently such a device was not available. The principle of parametric amplification has in fact been known for some time, but the development of the modern low-noise parametric amplifier had to await the invention of the pn junction parametric diode.

The property of the pn junction that interests us is as follows: When a pn junction is biased in the back direction, nearly zero current is drawn. The diode is capacitative, however, and the value of the capacitance depends on the instantaneous value of the applied voltage in the manner indicated by the dashed curve in Fig. 13. The solid curve is current vs voltage at low frequency. The back voltage at which the current rises sharply is called breakdown voltage V_B. It is an important parameter because we must operate between zero voltage and V_B in order to avoid drawing current and to keep the capacitance essentially lossless.

If we bias the diode approximately halfway between zero and V_B and apply a sinusoidal voltage of amplitude $V_B/2$ and frequency f_3 we will get the largest possible variation of C vs time. Filters may be necessary to keep f_3 separate from the signal frequency, but, except for this slight difficulty, this is a perfectly acceptable way of producing a condenser whose value varies at f_3.

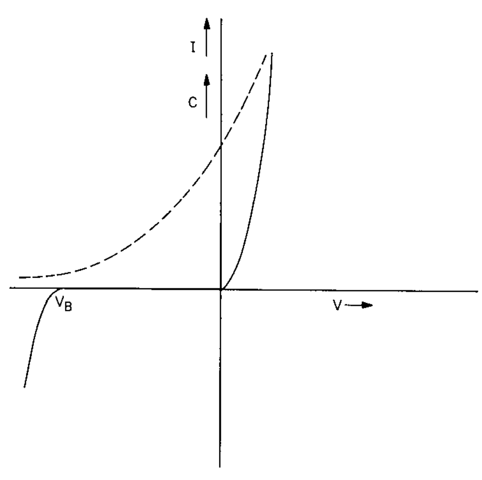


Fig. 13 - Current (solid curve) and capacitance (dotted curve) of typical pn junction as a function of voltage

All real parametric diodes have a spreading resistance R_s in series with the nonlinear

Fig. 14 - Equivalent circuit of back-biased varactor or parametric diode

capacitance. This resistance is, to a good approximation, frequency and bias independent. The equivalent circuit is shown in Fig. 14, where the arrow through the capacitance indicates that C is voltage dependent. The effect of R_s is to degrade the signal performance of the diode. This degradation becomes worse as the frequency rises, and for very high frequencies C behaves as a short-circuit; when this happens only R_s is left which cannot produce parametric effects. In addition, R_s makes a noise contribution and degrades the noise figure. It is useful to define a cutoff frequency f_c as that frequency at which the magnitude of the capacitative reactance is equal to R_s. Thus, $R_s = 1/(2\pi f_c C')$ gives $f_c = 1/(2\pi R_s C')$. It will be shown later that a special capacitance C_0 should appear in the above formula. In practice, however, manufacturers define f_c in terms of the value of the diode capacitance for some particular back bias, and we denote this by C'. The value of f_c so defined is close to the correct value, which is obtained when C_0 is used. Unfortunately, there is no universal agreement as to what bias to use in defining C'. Various authors use zero bias, V_B, $V_B/2$ or some arbitrary value. Still f_c is a useful figure of merit for a PA diode, or, as it is sometimes termed, a varactor diode. Cutoff frequencies for commercially available diodes go up to the order of 200 kMc. As a rough approximation we can say that an acceptable amplifier can be built at 1/10 the cutoff frequency.

GENERAL THEORY OF SMALL SIGNAL NONLINEAR CAPACITANCE

Introduction

We now begin a quantitative study of the properties of the nonlinear capacitor. The reader may wonder why the term nonlinear capacitor is used instead of time-varying capacitor, which was previously employed. This is because the only practical method we have of producing a time-varying capacitor at high frequency is by using the pn junction diode. The pn junction exhibits a time-varying capacitance only because its capacitance is a function of the applied voltage; that is, it is a nonlinear capacitance.

It has been mentioned that a nonlinear capacitor may be used for amplification in several different ways or modes. We decided to use the term parametric amplifier (PA) to include these different ways. However, we are going to make a special study of two circuits: the upper sideband upconverter, hereafter called the upconverter, and the negative resistance parametric amplifier, hereafter called the NRPA. As one might guess, the upconverter is really a mixer with gain, but unlike the usual mixer the output frequency is higher than the input frequency. The upconverter is perfectly stable; that is, there is no combination of source and load impedance and no change in diode characteristics which will make it oscillate. The properties of the NRPA are different. The input and output frequencies are the same, in fact, the input and output terminals are identical, and it will oscillate if loaded improperly. What is the point, then, in discussing both of them in the same section? It is because their properties as well as the properties of several modes, which we are not going to discuss in detail, arise fundamentally from the nonlinear capacitor. Thus the first steps are the same regardless of the mode of operation.

The key idea here is that any form of PA is intrinsically a multifrequency device. This is true even if the same frequency is found at both the input and output terminals. In the case of the NRPA, for example, which has a common input and output frequency, it can be shown that both current and voltage at the idler frequency must be present for the device to operate; in fact, power must be dissipated at the idler frequency. Either currents or voltages must be present at many other frequencies, but normally these channels may be either open or short-circuited so that power is not dissipated at these frequencies.

The easiest way to short-circuit or open these unused frequencies is to use tuned circuits for the frequencies at which power must be dissipated, but this results in a narrow-band device. Those readers who have some familiarity with the conventional microwave mixer will find that it is similar in many ways to the PA.

Mixing in the Nonlinear Capacitance

Let us examine the consequences of the mixing action of a nonlinear capacitance. The problem is greatly simplified if we assume that the pump voltage is large when compared to the signal voltage. This is tantamount to the assumption that as far as the signal is concerned the capacitor is linear; that is, the value of the capacitor is not changed by the signal voltage. This is similar to the assumption made in treating vacuum-tube or transistor amplifiers. If the signal voltage is small enough the amplifier may be treated as a linear four-pole. No distortion and no harmonics of the input frequency appear at the output. But when the signal becomes large, as it may in a power amplifier, we expect distortion and new frequency components in the output. Thus, the amplification factor μ of a vacuum tube depends on the value of the supply voltage but does not depend on the signal amplitude if the signal amplitude is below a certain level. In the same way we can say that the properties of the capacitative mixer (i.e., the way that the capacitance changes with time) depend on the pump but not the signal.

Assume then that the capacitor C is driven by a high level pump at frequency f_3 and also by a low level signal at frequency f_1. We do not wish to say at this point whether the f_1 "signal" is the input, the output, or neither. Assume that f_1 is the lowest frequency low-level signal. Consider only normal pumping, then $f_1 < f_3$. Because of the mixing process, a set of low level signals will be generated whose frequencies are given by $nf_3 \pm f_1$ where n is any integer. In addition there will be a set of high level frequencies generated which are simply the harmonics of the pump frequency. This is given by mf_3, where m is any integer. However, we are not interested in these frequencies since they are not associated with any signal but instead with pump or B^+ supply.

Note that we must deal with an infinite set of low level frequencies since n can assume any value. This is not a very desirable situation either from the theoretical or practical point of view. If the energy of the input signal were to be converted into a large number of new frequencies and then dissipated at these frequencies we could hardly expect low noise and high gain to result. A way out of this difficulty is to short-circuit or open all those frequencies which are not needed. This will prevent the dissipation of power at the unused frequencies. If we assume that this is done it can be shown that, in the theoretical treatment, we can completely ignore these unused frequencies.

It is one thing to assume that the above can be done and quite another thing to do it in a practical amplifier. The fact that amplifiers can be built with properties closely approximating those predicted by the simplified theory indicates that the unused frequencies are being effectively suppressed. It turns out that to account for the parametric phenomena of interest we need to consider only three frequencies, namely f_1, $f_2 = f_3 - f_1$, and $f_4 = f_3 + f_1$.

The above argument exactly parallels the treatment of the conventional radar microwave mixer, in which typical values might be $f_1 = 30$ Mc, $f_3 = 10{,}000$ Mc, and thus the other "signal" frequencies would be $f_4 = 10{,}030$ Mc and $f_2 = 9970$ Mc. In this case f_1 is the output i-f frequency and either f_2 or f_4 is the input microwave frequency. In this particular case, it makes little difference whether the upper sideband f_4 or the lower sideband f_2 is chosen for the input, because this mixer is a resistive mixer rather than a capacitative mixer. However, something must be done about the sideband which is not chosen for the input. It may be short-circuited thus giving the lowest noise figure. Unfortunately, this reduces the bandwidth (since a filter must be employed). This is

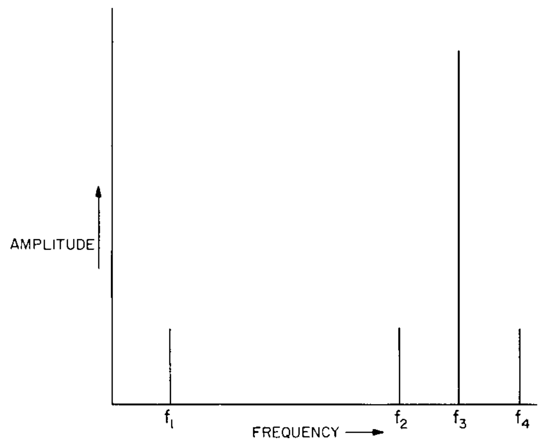


Fig. 15 - Frequency spectrum of a parametric amplifier. In practice, either f_2 or f_4 is reactively terminated.

called narrow-band operation, and it is equivalent to considering this sideband as an unused frequency. If no filter is used the unused sideband is terminated in the same way that the input frequency is terminated, namely with the source impedance.* Since no signal enters this unused channel, but noise associated with the source resistance does enter, this noise figure is higher than the narrow-band noise figure.

Let us return to the capacitative mixer. Figure 15 shows the frequency spectrum we are now treating. The reader may convince himself that since f_1 is by definition the lowest signal frequency, it can at most equal $1/2\ f_3$. The problem now consists of determining the relationships which the pumped capacitor imposes between the voltages and currents at the three frequencies of interest. Since we assumed these currents and voltages are very small, it is natural to assume that these relationships will be linear, just as the relationships between the various currents and voltages in vacuum-tube or transistor amplifiers are linear. Let I_1, V_1, I_2, V_2, I_4, and V_4 denote the complex amplitudes of the currents and voltages at the angular frequencies ω_1, ω_2, and ω_4. Then, assuming that the unused frequencies are short-circuited it can be shown (2a) that the relationships we seek are of the form

$$\begin{aligned} I_4 &= j\omega_4 C_0 V_4 + j\omega_4 C_1 V_1 + j\omega_4 C_2 V_2^* \\ I_1 &= j\omega_1 C_1 V_4 + j\omega_1 C_0 V_1 + j\omega_1 C_1 V_2^* \\ I_2^* &= -j\omega_2 C_2 V_4 - j\omega_2 C_1 V_1 - j\omega_2 C_0 V_2^* \end{aligned} \tag{6}$$

where the asterisks indicate the complex conjugate. The real numbers C_0, C_1, and C_2 will be defined presently.

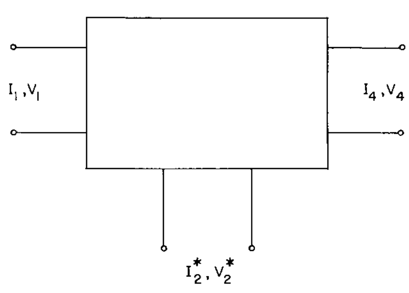


Fig. 16 - Three-port which represents the capacitative mixer. Each port is at a different frequency.

These equations contain all the information relating to the signal properties of any of the more frequently used modes of parametric amplification, and in a sense they even include the noise properties. It must be remembered, however, that these equations refer to the pure nonlinear capacitor; that is, the effect of the spreading resistance is not included.

At this point, we are dealing with a linear six-pole or three-port device, and the properties of such networks have been extensively treated in the literature. The three-port device is indicated in Fig. 16. This sketch should emphasize the fact that this three-port is similar to other three-ports, except that the same frequencies do not appear at each port.

*This is because there are no microwave components left in the circuit which can distinguish between the two sidebands.

Various Operating Modes

Various amplifiers and converters can be formed from the three-port of Fig. 16 by choosing a port for the input, a port for the output (the input and output may be the same port) and by connecting passive loads to the unused ports. For example, there are six possible choices for converters if we require that the input and the output be different ports. They are:

1. Input at port 1, output at port 4.
2. Input at port 4, output at port 1.
3. Input at port 2 and output at port 4.
4. Input at port 4 and output at port 2.
5. Input at port 1 and output at port 2.
6. Input at port 2 and output at port 1.

Note that all these are converters (they shift frequency) since no two ports are at the same frequency. But this still does not indicate all the variations possible in converters. Consider, for example, 1 above. The properties of this converter will depend on what type of passive load is connected to the unused port 2. If port 2 were loaded resistively, for example, the converter would, in general, behave differently than if port 2 were reactively loaded.

Even though there are in theory many modes of operation, most of these modes are not practical for one reason or another. For example some modes do not result in gain but rather in loss.

Let us now discuss these various modes in some detail and indicate which ones are practical. Later we will treat in still further detail the upconverter and the NRPA. For now, statements will be made without proof; later some of these statements will be proved.

Converters will be discussed first. The unused frequency or the unused channel is always terminated by a short-circuit or an open circuit. It really does not make much difference which choice is made, but we will assume that the unused channel is short-circuited. Thus, we need to consider only the six choices listed above. Neither choice 4 or 2 are used because they result in loss rather than gain. Thus downconversion from the upper sideband is never used. Furthermore 3 is never used because it results in low gain. There is only one scheme left which uses the upper sideband, and this is 1, which is called the upper sideband upconverter or often just the upconverter. The maximum possible power gain is given by f_4/f_1; thus, the input frequency must be small compared to the pump frequency if significant gain is to result. The upconverter is the most broadband mode and it is completely stable; that is, it will not oscillate for any combination of load and source impedance. Although infinite gain is possible in 6, it is not used because it is unstable. This leaves 5, which is a lower sideband upconverter (compare it with 1). In principle it is an attractive mode although it has not been used often. Infinite gain and low noise are possible in this mode. It is inherently a narrowband converter compared with the upper sideband upconverter, and it is not stable for some combinations of source and load impedance. It also has the following peculiarity: Since the sum of the input and output frequencies equal a constant f_3, modulation sidebands which appear above the carrier at the input will appear below the carrier at the output. This is called sideband inversion, and it may not be acceptable in some systems. The only converter which seems worthy of further discussion then is the upconverter.

We now turn to single-frequency amplification. How can we make an amplifier, as distinguished from an amplifying converter? Clearly the same port must serve as both the output and the input, but the only way that a one-port can be used for amplification is for it to have negative conductance or resistance. It turns out that no matter how ports 1 and 2 are terminated, only positive resistance appears at port 4. Thus negative resistance cannot appear at a frequency higher than the pump frequency. There are two modes of operation which will yield a one-port or two-terminal negative resistance.

1. Short-circuit port 4, terminate port 2 resistively and a negative conductance appears at port 1.

2. Short-circuit port 4, terminate port 1 resistively and a negative conductance appears at port 2.

Note that in both cases f_4 simply becomes another unused frequency and can be ignored. Furthermore, it can be shown that 2 results in a higher noise figure, and in fact, is never used. We call mode 1 the negative resistance parametric amplifier or NRPA. It is the most popular form of parametric amplification.

It might appear to be better to terminate port 2 by a short-circuit or open-circuit rather than resistively. It turns out, however, that the negative resistance will appear at port 1 only if port 2, often called the idler, is terminated resistively. This can perhaps be understood intuitively from the following: If f_2 is short-circuited (and remembering that the unused frequencies are also short-circuited), then from the point of view of an observer at port 1, the whole device becomes completely reactive, that is, it could not have a resistive component because there is no loss mechanism. Thus the input admittance at port 1 is purely reactive. The nonlinear capacitor when employed in this manner can be thought of as a transformer which transforms the positive conductance or resistance of the idler circuit into a negative conductance or resistance at port 1.

In summary, there are only two commonly used modes of parametric amplification:

1. The upconverter which has input at f_1 and output at f_4. The output frequency f_4 must be large with respect to f_1 since the maximum possible gain is f_4/f_1. This means that the pump frequency f_3 must also be large since it is given by $f_4 - f_1 \approx f_4$. In practice, the gain is usually somewhat less than f_4/f_1, which forces the pump and output frequencies up still further to realize a given gain. The upconverter is completely stable and it is the most broadband mode. There are many situations in which the frequency shift is not acceptable. Suppose, for example, that we want to put a low-noise preamplifier on an existent radar receiver in order to improve the radar range. A converter could not be used because the input and output frequencies of the preamplifier must be the same. On the other hand if we design a low noise receiver from scratch with only the input frequency specified, there would be no objection to the use of a low-noise upconverter first stage followed by a conventional amplifier at f_4.

2. The NRPA. The input and output frequencies are the same. The gain can be very high, but the result is gain instability and the possibility of oscillations. Power must be dissipated at an idler frequency, which is equal to the pump frequency minus the signal frequency. The NRPA is relatively narrow-band.

Properties of the Three-Port Nonlinear Capacitance Mixer

In order to quantitatively treat the upconverter and the NRPA we must reconsider Eq. (6), which describes the three-port from which these devices are constructed. The properties of the three-port (which in turn determine the properties of the upconverter and the NRPA) are contained wholly in the coefficients of the voltages on the right side of Eq. (6). It is convenient to arrange these coefficients as shown below and to call this array the admittance matrix Y of the three-port. Thus,

$$Y = \begin{pmatrix} j\omega_4 C_0 & j\omega_4 C_1 & j\omega_4 C_2 \\ j\omega_1 C_1 & j\omega_1 C_0 & j\omega_1 C_1 \\ -j\omega_2 C_2 & -j\omega_2 C_1 & -j\omega_2 C_0 \end{pmatrix} . \tag{7}$$

We need to discuss this admittance matrix, but first we should note one peculiar feature about Eq. (6). The current and voltage corresponding to the lower sideband appear as complex conjugates. This is a result which is necessary if negative frequencies are to be avoided, and will cause no difficulty.

First note that all the elements of Y are pure imaginary numbers. This seems reasonable because Y was constructed from a lossless nonlinear capacitor. In fact each term of Y has the form of the admittance of a capacitor. Note also that each row of Y contains only one frequency; for example, the second row (which is really the f_1 row) contains only the frequency f_1.

Finally, we see that the element of the first row, second column is not equal to the element of the first column second row and so forth. Networks in which this is the case are called nonreciprocal. There is a very fundamental difference between nonreciprocal and reciprocal networks. Nonreciprocal networks have a "one way" feature in that the maximum power gain differs when the input and the output are interchanged. Examples of reciprocal networks are (a) networks consisting only of resistances, capacitances, and inductances, and (b) passive microwave circuits without static magnetic fields, etc. Examples of nonreciprocal networks are vacuum-tube and transistor amplifiers, microwave isolators, and microwave circulators.

Consider the real parameters C_0, C_1, C_2. These arise as follows: The local oscillator voltage causes a periodic variation of capacitance. This depends on the C vs V curve of the diode (see Fig. 13) and on the local oscillator voltage waveform. Thus, C is a periodic but, in general, a nonsinusoidal function of time (i.e., $C = C(t)$). Therefore, $C(t)$ may be expanded in a Fourier series. The problem is greatly simplified if it is arbitrarily assumed that $C(t)$ is an even function, which is the same thing as saying that we can leave out the sine terms in the expansion. Thus,

$$C(t) = C_0 + 2C_1 \cos \omega_3 t + 2C_2 \cos 2\omega_3 t + \cdots .$$

In words, the C's are one-half the Fourier coefficients of the expansion of the capacitance as a function of time. It is important to note that the C's depend on both the nonlinearity law of the capacitance and on the pump waveform. If the capacitance is linear, that is if C does not depend on V, then $C(t)$ will simply be a constant and C_1, C_2, ... will equal zero. What does Eq. (6) have to say about this situation? Suppose a voltage V_1 is applied at port 1. The first row gives $I_4 = 0$, the second row gives $I_1 = j\omega_1 C_0 V_1$, and the third row gives $I_2 = 0$. In other words, there is neither coupling between ports nor mixing action. This is just what is expected of an ordinary linear capacitor. Evidently the C's are a measure of the amount of coupling between ports and of the amount of mixing which occurs. Also, they are a measure of the degree of nonlinearity of the varactor capacitance under dynamic (i.e., pumped) conditions.

THE UPCONVERTER

The Circuit Equations

In the previous section, we laid the foundation for a study of the upconverter and the NRPA. In this section we will consider the upconverter.

For upconverter action, port 2 of the three-port discussed in the previous section is short-circuited. Letting $V_2^* = 0$ in Eq. (6) we have

$$\begin{aligned} I_1 &= j\omega_1 C_0 V_1 + j\omega_1 C_1 V_4 \\ I_4 &= j\omega_4 C_1 V_1 + j\omega_4 C_0 V_4 . \end{aligned} \tag{8}$$

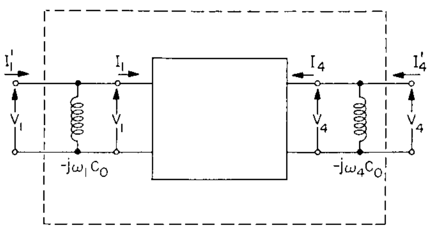


Fig. 17 - The upconverter is tuned by shunt inductances at the input and output

It is permissible to tune the input and the output with reactive elements, since these elements cannot contribute any noise. We then shunt the input of the two-port described by Eq. (8) with a susceptance $-\omega_1 C_0$ and shunt the output with a susceptance $-\omega_4 C_0$. The reason for doing this will appear presently. Once this is done, we next calculate the equations of the new two-port, which was made by simply shunting the input and the output of the original two-port described by Eq. (8). This situation is depicted in Fig. 17. The solid line box represents the original two-port described by Eq. (8), and the terminal variables V_1, I_1 and V_4, I_4 are indicated. The input and output are shunted by inductances, and the new two-port which results is indicated by the dashed line box. Note that the terminal currents for the new two-port are given by $I_1' = I_1 - j\omega_1 C_0 V_1$ and $I_4' = I_4 - j\omega_4 C_0 V_4$. Thus, $I_1 = I_1' + j\omega_1 C_0 V_1$ and $I_4 = I_4' + j\omega_4 C_0 V_4$. Substituting these equations into (8) we get the very simple circuit equations for the new two-port:

$$I_1' = j\omega_1 C_1 V_4 \tag{9a}$$

$$I_4' = j\omega_4 C_1 V_1 . \tag{9b}$$

Input and Output Conductance

If a real source admittance G_s and real load admittance G_ℓ are connected to this two-port, it is easily shown (see Appendix A) that the input admittance G_i equals $\omega_1 \omega_4 C_1^2/G_\ell$ and the output admittance G_0 equals $\omega_1 \omega_4 C_1^2/G_s$. These are both positive pure conductances. The condition for maximum power gain is that $G_s = G_i$ and $G_\ell = G_0$. Thus

$$G_s = \frac{\omega_1 \omega_4 C_1^2}{G_\ell} ; \quad G_\ell = \frac{\omega_1 \omega_4 C_1^2}{G_s} . \tag{10}$$

One of these equations is evidently redundant, and the matching condition for maximum power gain comes out

$$G_s G_\ell = \omega_1 \omega_4 C_1^2 . \tag{11}$$

This result is a little strange. Ordinarily we expect to have a unique answer for the values of G_ℓ and G_s which are required for maximum power gain; here only the product of G_ℓ and G_s is fixed. This result comes about because the coefficients in Eq. (9) are pure imaginary.

Gain

We next want to calculate the maximum available gain A_m, which is defined as the ratio of the power delivered to the load to the power delivered to the input when matched conditions exist. Assume that matched conditions do exist. Then V_1 and I_1' are in phase, and V_4 and I_4' are in phase (remember both the input and output admittance are pure real); therefore, the input power is $I_1'V_1$ and the power delivered to the load is $I_4'V_4$. Multiply Eq. (9a) by I_4' and (9b) by I_1' and equate the two. We then have

$$j I_1' V_1 \omega_4 C_1 = j I_4' V_4 \omega_1 C_1$$

$$\frac{I_4' V_4}{I_1' V_1} = A_m = \frac{\omega_4}{\omega_1} = \frac{f_4}{f_1} .$$

That is, the maximum available gain obtainable from an upconverter is the ratio of the output frequency to the input frequency.

Equivalent Circuit

It is now possible to see the reason for the shunt tuning which was carried out at the output and input of the original two-port described by Eq. (8). In so doing, we made the circuit resonant; that is, we made the input admittance pure real when the load admittance is pure real and we made the output admittance pure real when the source admittance is pure real. We also set up the condition for power match by matching out the susceptance components at the input and the output. Viewed differently, Eq. (9) could be said to describe a nonlinear capacitor whose average value, C_0, is zero. This is possible only if the capacitor has negative values (if any function is to have a zero average over some range it must have negative values); thus it does not correspond to any real capacitor. However, such a capacitor is a useful idealization which is correctly described by Eq. (9).

We cannot pretend that C_0 does not exist, and if we wish to draw an equivalent circuit using this idealized capacitor we can represent the effect of C_0 correctly by shunting both the output and the input with C_0. This is done in Fig. 18. Furthermore, any amount of additional capacitance could be added at either the input or the output as long as the capacitance is compensated by additional inductance. Thus, the shunt capacitance appearing at the output and input in the final equivalent circuit can be any value that is larger than C_0. Of course, the input must be resonant at ω_1 and the output resonant at ω_4. The final equivalent circuit is given in Fig. 19. The above argument was undertaken

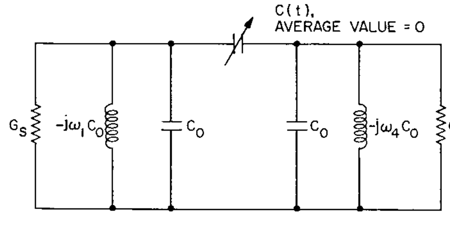


Fig. 18 - Equivalent circuit of the upconverter

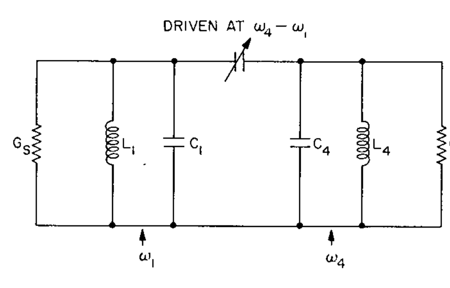


Fig. 19 - Final equivalent circuit for the upconverter

because the circuit of Fig. 19 is often seen in the literature and is in fact sometimes the starting point for the treatment of parametric amplification.

Noise Figure

From an intuitive point of view, it appears that all pure capacitors should be noiseless, whether they are linear or nonlinear. Experimental evidence indicates that this is the case, although some theoretical work indicates that a very small amount of noise can be expected from a pn junction capacitance. We will assume that the pn junction capacitance is noiseless. Then the calculation of the noise figure of the upconverter becomes trivial. There is no source of noise in the upconverter, and therefore, it must have a noise figure equal to one for any source admittance. Incidentally, the same argument holds for any network made up of capacitances and inductances—the noise figure always equals one.

When we include the effect of the spreading resistance (which is a source of noise) the noise figure of the upconverter will be greater than one.

Bandwidth

In the section on the NRPA an expression will be derived for the gain-bandwidth. The gain-bandwidth calculation for the upconverter is similar, but since the algebra is more complicated it will be omitted, and only the result will be given (2b). The maximum possible bandwidth under high gain conditions is $2/Q_1$, where Q_1 is the Q of the input tank circuit as loaded by G_s and any losses present. In other words, the bandwidth of the upconverter is twice the bandwidth of the input tank. This may seem strange, but it is caused by the fact that the input conductance is a pure real positive number and tends to reduce the overall Q. We might expect Q_4, the Q of the output tank circuit, to appear in the expression for bandwidth. In fact, in an exact calculation Q_4 does appear, but we are using an approximation which permits Q_4 to be omitted.

The alert reader may ask what the practical value of the above formula for bandwidth may be. No criterion seems to exist for determining how low a value Q_1 can have. The Q of the circuit, after all, can be made as low as is desired. A complete answer to this question cannot be given until we consider the effects of the spreading resistance, but a partial answer can be given now. We are looking for some logical reason which would prevent one from making Q_1 indefinitely small. The main loss in the input tank which controls Q_1 is the source conductance G_s. Equation (11) tells us what values of G_s are needed for high gain. If the value of G_s is large then G_ℓ will be small. But this will force the value of Q_4 to be large, and if Q_4 is too large the output tank is going to start to limit the bandwidth.

This argument is not a quantitative one, but at least it indicates there is some limit on how small we can make Q_1. Nevertheless, in practice low values of Q_1 are allowed, and the result is that the single-tuned upconverter, particularly when compared with the single-tuned NRPA is a broadband device.

PROPERTIES OF NEGATIVE RESISTANCE AMPLIFIERS

Introduction

The maser, the tunnel diode amplifier, and the negative resistance parametric amplifier are (in their simplest forms) basically negative resistance amplifiers. At resonance, the equivalent circuit of these three devices is simply a noisy negative resistance (or conductance). The particular value of the negative resistance and the noise temperature depends on the detailed physics of the device. Therefore, it seems reasonable to treat the general negative resistance amplifier (at resonance) so that the results may be applied to any of the above three negative resistance amplifiers or to any new negative resistance amplifiers which may be developed. It should be emphasized that this treatment applies only at resonance or at the center frequency of the device. This treatment, then, can yield no information concerning bandwidth.

Transducer Gain and Stability

The circuit for the simplest negative resistance amplifier at resonance is shown in Fig. 20. The source on the left has a conductance G_s and a short circuit current i_s. On the right is shown a load of conductance G_ℓ and in the middle is the negative conductance of value $-G$. (Note that G is a positive number.)

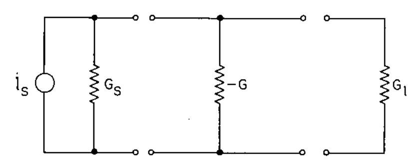


Fig. 20 - Circuit for analyzing a negative conductance amplifier

We will first consider the possibility of oscillations in this circuit, assuming that it is shunt resonant. It can be shown that a shunt resonant circuit will oscillate if the total shunt conductance is less than or equal to zero. For this circuit to be stable we can demand that $G_\ell + G_s - G \geq 0$. This suggests defining a stability factor $\mathbf{S} = G_\ell/G + G_s/G - 1$. The system is then stable if $\mathbf{S} > 0$; and the larger $\mathbf{S}$ is, the more stable is the circuit. If $\mathbf{S}$ is nearly equal to zero, a small decrease in either G_ℓ, G_s or a small increase in G will cause the circuit to oscillate.

The transducer gain A_T is defined as the ratio of the power delivered to the load to the power available from the generator. It is one of the more important parameters of any power amplifier. For the circuit of Fig. 20 it can be shown that

$$A_T = \frac{4G_\ell G_s}{(G_\ell + G_s - G)^2} = \frac{4(G_\ell/G)(G_s/G)}{(G_\ell/G + G_s/G - 1)^2} = \frac{4(G_\ell/G)(G_s/G)}{\mathbf{S}^2}. \tag{12}$$

The expression on the far right indicates that the actual values of conductance are not important but that the conductive ratios are decisive.

The point of view taken here is that G is a parameter over which the engineer has little or no control. However, by means of transformers the engineer can control both G_s and G_ℓ. No matter what the value of G, he can control the ratios G_ℓ/G and G_s/G and thus can control $\mathbf{S}$ and A_T. As a matter of fact, all the quantities of interest at the resonant frequency depend only on these or similar ratios, never on actual values.

We notice that according to Eq. (12), the condition $\mathbf{S} \to 0$ (from the positive side), which is the condition for oscillations, is also the condition for large transducer gain. In fact, oscillations will start when the transducer gain approaches infinity. The above facts are the basis for the often seen statement that NRPA's, and in fact any NRA, are unstable. This behavior is sharply different from the behavior of ordinary (e.g., neutralized triode) amplifiers. If we consider the transducer gain of ordinary amplifiers as a function of G_s and G_ℓ we find that there are special values of each yielding a maximum gain (the matched condition) which is not infinity. Of course, there are no terminal conditions which cause oscillations.

Available Gain and Noise Figure

We should be careful, however, not to overemphasize the importance of the transducer gain for the following reason: In practice, negative resistance amplifiers are nearly always used as low-noise preamplifiers which feed conventional amplifiers, and these, not the NRA, feed the final load. Furthermore, the parameter which is most important is the overall amplifier noise figure. If the overall noise figure is satisfactory, then the transducer gain of the preamplifier (i.e., the NRA) is of little interest. If the overall transducer gain is not high enough, it can always be increased by adding a few final stages; this will not affect the noise figure. It is convenient to think of the overall amplifier as a tandem connection of the preamplifier and a very high gain conventional amplifier. Then the overall noise figure F is given by the tandem formula (Eq. (5)), where F_1 and A_1 refer to the preamplifier and F_2 refers to the conventional amplifier. Naturally, there is no point in using the preamplifier unless F is less than F_2.

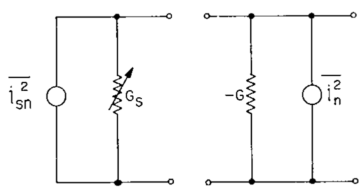


Fig. 21 - Noisy negative conductance with noisy source conductance

We next consider A_1 and F_1, and to investigate the latter quantity we must make G noisy. Since G is in reality a two-terminal network, we can do this by placing a short-circuit noise generator $\overline{i_n^2}$, of Fig. 21, across G. Let us calculate F_1 with aid of Fig. 21. From the definition of noise figure we see that $\overline{i_{sn}^2}$ is given by $4kT_0G_s$. We might expect some difficulty because $-G$ is negative, but it is useful to simply define the noise temperature of the negative conductance from the equation $\overline{i_n^2} = 4kTG$. The network of Fig. 21 is the dual of the network of Fig. 4; thus, we arrive at the proper expression for noise figure by replacing the resistances by the conductances. (The negative sign just does not enter.) It is convenient to call the quantity T/T_0 the noise temperature ratio of G. Let $t \equiv T/T_0$. The number t tells how many times hotter a two-terminal network is than room temperature. The expression for F_1 then becomes

$$F_1 = 1 + t\frac{G}{G_s}. \tag{13}$$

To calculate A_1, which is the ratio of the available output power to the available source power, refer to Fig. 20. Assume that G_s is always greater than G. We get

$$A_1 = \frac{\dfrac{i_s^2}{4(G_s - G)}}{\dfrac{i_s^2}{4G_s}} = \frac{G_s}{G_s - G}. \tag{14}$$

We see that if G becomes larger than G_s, the available gain becomes negative. There is, in fact, a way of handling this possibility, but for the sake of simplicity we will only consider the case $G < G_s$. Note that $G_s - G$ is really the output conductance, and we would expect difficulty when this expression is negative.

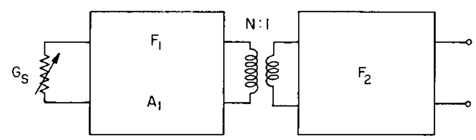
The observant reader may note that according to Eq. (13) F_1 approaches to 1 as G_s approaches infinity. This is really a trap, because Eq. (14) shows that, for the same condition, A_1 approaches a very small value. This illustrates the fact that noise figure alone is a valid figure of merit for noise only if the gain is high.

Overall Noise Figure

We are now ready to consider the overall noise figure of a receiver consisting of a low-noise NRA followed by a conventional amplifier. This system is illustrated in Fig. 22. It is assumed that both G_s and the interstage transformer are adjustable. If Eqs. (13) and (14) are substituted into Eq. (5), then the overall noise figure is

$$F = 1 + t\frac{G}{G_s} + \frac{(G_s - G)(F_2 - 1)}{G_s} . \tag{15}$$

Fig. 22 - Equivalent circuit for the calculation of the overall noise figure of a negative conductance preamplifier and a conventional postamplifier



Of course F_2 is dependent upon the source admittance seen by the conventional amplifier (amplifier number 2 in Fig. 22), and thus the source admittance is dependent upon the turns ratio N and upon $G_s - G$. Thus the overall noise figure is a complicated function of G_s and N. This problem can be treated exactly, but we will come to some useful conclusions by using intuition. The reason for using the preamplifier is to obtain a lower overall noise figure than the one obtained with the conventional amplifier alone. To do this the noise contribution from the second amplifier must be minimized. This is represented by the last term in Eq. (15). This term can be minimized by making G_s nearly equal to G and by adjusting the transformer so that F_2 assumes its minimum value. If this is done then

$$F = 1 + t . \tag{16}$$

This is a rather important result, because for all practical purposes this is the lowest overall noise figure that can be obtained from any receiver which uses the NRA with t as a preamplifier. It should be mentioned that typical values of t at 3 kMc are (a) too-low-to-be-measured (masers), (b) 0.3 for typical NRPA's, and (c) 1.0 for tunnel diodes.

There is still one important point to be made about stability. In making $G_s - G$ very small we are also making the source conductance for the conventional amplifier, $(G_s - G)/N^2$, very small (unless N is also made very small). Making $(G_s - G)/N^2$ very small is undesirable because F_2 assumes its minimum value for some finite value of source admittance. Thus, to keep $(G_s - G)/N^2$ finite we must make N arbitrarily small. But G_ℓ, the load seen by the preamplifier is given by $N^2 G_i$, where G_i (assumed finite) is the input conductance of the conventional amplifier. Thus, $G_\ell \approx 0$. Since it has been assumed that $G_s \approx G$, we have $G_\ell + G_s - G \approx 0$. This is just barely the condition for the circuit to oscillate. Since we must back off slightly the noise figure is somewhat greater than $1 + t$. This is also an important result and bears repeating. The lowest possible noise figure is obtained only at the expense of operating on the verge of the oscillating condition. It can be shown that there must always be an exchange of stability for noise figure, so that great stability is obtained for relatively large noise figures. It turns out, however, that this exchange can be made much more favorable by using nonreciprocal elements such as isolators and circulators. This will be discussed later.

THE NEGATIVE RESISTANCE PARAMETRIC AMPLIFIER

Circuit Equations

A better name for the negative resistance parametric amplifier would be the negative conductance parametric amplifier, but the former term is so well established in the literature that we will use it.

To get NRPA action, port 4 of Fig. 16 is short-circuited. In Eq. (6) let $V_4 = 0$. We have

$$\begin{aligned} I_1 &= j\omega_1 C_0 V_1 + j\omega_1 C_1 V_2^* \\ I_2^* &= -j\omega_2 C_1 V_1 - j\omega_2 C_0 V_2^* . \end{aligned} \tag{17}$$

We shunt-tune port 1 and port 2 exactly the same way and for exactly the same reasons as we did for the upconverter. The identical arguments apply, and the equations for the new two-port (i.e., the two-port with tuning) are

$$\begin{aligned} I_1' &= +j\omega_1 C_1 V_2^* \\ (I_2^*)' &= -j\omega_2 C_1 V_1 , \end{aligned} \tag{18}$$

where the primes indicate that the terminal currents of the new two-port are different from the terminal currents of the original two-port.

Conductance Calculations

Proceeding as we did with the upconverter, the next task is to calculate the admittance looking into port 1 when port 2 is terminated with a conductance and vice versa. We want to be careful not to call one port the input port and the other port the output port, because the NRPA is not used in this manner. Let us connect a conductance G_2 on port 2 and calculate the admittance looking into port 1. Call this admittance $Y_{1i} = G_{1i} + jB_{1i}$.

This calculation is done just as it was for the upconverter (Eq. (10)) and the result is

$$Y_{1i} = G_{1i} = -\omega_1 \omega_2 C_1^2/G_2 . \tag{19}$$

This is an important equation. It says that at resonance a pure negative conductance appears at port 1. Similarly, if we connect a conductance G_1 on port 1 and calculate the admittance looking into port 2, Y_{2i}, we will get

$$Y_{2i} = G_{2i} = -\omega_1 \omega_2 C_1^2/G_1 . \tag{20}$$

This can, in fact, be written down immediately from symmetry consideration.

Note that both Eqs. (19) and (20) demand that a finite conductance be connected to one port if a finite negative conductance is to be observed at the other port.

Noise in the Negative Resistance Parametric Amplifier

According to Eqs. (19) and (20), a negative conductance can be produced at port 1 of the two-port described by Eq. (18) as long as a pure conductance is placed across port 2, and vice versa. It is customary to call the port at which the negative conductance is produced the signal port and to call the other port the idler. It is not clear at this point which port should be chosen as the signal port, but we shall start by picking port 1, which corresponds to the lower frequency.

An equivalent circuit for this amplifier, including source and load is shown in Fig. 23. Here G_I, which was formerly called G_2, is the idler load. This figure should be compared with Fig. 20. Note that there are three distinct conductances which we must consider: the source conductance G_s, the load conductance G_ℓ, and the idler conductance G_I. In addition, of course, the negative conductance of port 1 is shunted across the source and load just as in Fig. 20. This negative conductance (G_{1i} in Eq. (19)) will now simply be $-G$ (just as in Fig. 20).

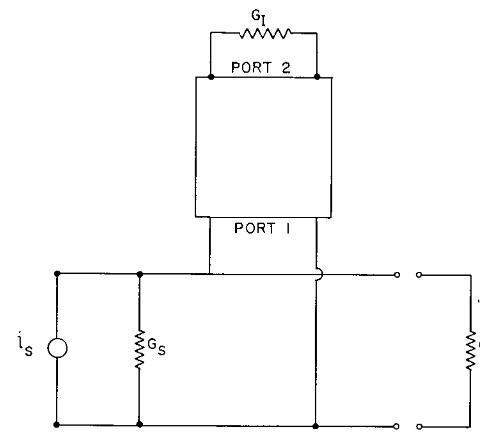


Fig. 23 - Equivalent circuit of the NRPA

The noise figure and available gain of this amplifier have already been worked out and are given in Eqs. (13) and (14). In addition, the minimum possible overall noise figure obtainable using this amplifier as a preamplifier has been given in Eq. (16). One problem remains, however: What is the noise temperature ratio looking into port 1?

Port 1 must be noisy. We know that G_I must have a finite value; thus if G_I is at a finite temperature T_I there will be thermal noise at the terminals of G_I. This noise is at the idler frequency, but because of the mixing process, this noise will appear at the signal port and produce a noise temperature ratio, t. We can calculate t as follows. The thermal noise of G_I can be represented by a short-circuit noise generator i_I, where $\overline{i_I^2} = 4kT_I G_I$. Then we ask the following question: If port 2 is driven by a generator of source conductance G_I and short-circuit current i_I, what short-circuit noise current i_n is produced at port 1? This is a straightforward circuit problem. Once i_n is determined, the noise temperature at port 1 is given from $\overline{i_n^2} = 4kTG$, and thus $t = \overline{i_n^2}/(4kGT_0)$. But we have already calculated G (Eq. (19)); thus

$$t = \frac{T_I}{T_0}\,\frac{\omega_1}{\omega_2}\,. \tag{20}$$

Let $\omega_s \equiv \omega_1$, $\omega_I \equiv \omega_2$, and $T_I/T_0 \equiv t_I$. Then

$$t = t_I\,\frac{\omega_s}{\omega_I}$$

where t_I is the noise temperature ratio of G_I. It is customary to call ω_s and ω_I the signal and idler frequencies respectively. It will be recalled that ω_2 is greater than ω_1, and therefore t is less than t_I. In fact, by making the ratio of the idler frequency to the signal frequency become arbitrarily large, t can be made arbitrarily close to zero. Another way of saying this is that the downconversion process from ω_2 to ω_1 is an inefficient process; thus noise originating at frequency ω_2 is reduced when it appears at frequency ω_1. Had we picked port 2 as the signal port and port 1 as the idler port, the ratio would have gone the other way; that is, t would be given by $t_I(\omega_2/\omega_1)$ and the idler noise would be magnified in the conversion process. For this reason, the idler frequency is never made less than the signal frequency, and if possible, the idler frequency is made very large compared with the signal frequency.

According to (Eq. 20), cooling G_I will reduce t_I. This is done in some practical amplifiers.

If Eq. (20) is substituted into Eq. (16) we have for the lowest possible overall noise figure using a NRPA as a preamplifier

$$F = 1 + t_I(f_s/f_I) . \tag{22}$$

As an example, consider a NRPA operating at 200 Mc with the idler at room temperature at 1000 Mc. Then the minimum noise figure would be 1.2. In practice, the idler circuit is often a resonant cavity and G_I may simply consist of the cavity losses. Then the idler noise temperature ratio is just the noise temperature ratio of the cavity, or T_c/T_0, where T_c is the cavity temperature.

Bandwidth

Up to now, we have only been considering the center or resonant frequency properties of the NRPA. The algebra was relatively simple because everything was pure real at the center frequency. It is important to know (for any amplifier) how the properties of the amplifier change as the input frequency deviates from the center frequency. In particular, we are usually interested in the frequency dependence of the transducer gain A_T. The center frequency value is given by Eq. (12). Normally, one is not interested in A_T at frequencies far removed from the center frequency because A_T would simply be very small and not useful. Therefore, we seek an expression for the transducer gain for frequencies close to the center frequency. Naturally, expressions of the form $G + j(\omega C - 1/\omega L)$ will frequently appear in such calculations. Viewed as a function of ω, the above is a rather unwieldy expression. It is well known that for frequencies close to resonance this expression can be written as $G(1 + j2\delta Q)$, where δ is the fractional deviation of the signal frequency from the resonant frequency ω_0. Thus, $\delta = (\omega - \omega_0)/\omega_0$, and Q is given by $G/\omega_0 C$. This expression simplifies things, because it is linearly dependent upon frequency in contrast to the original expression.

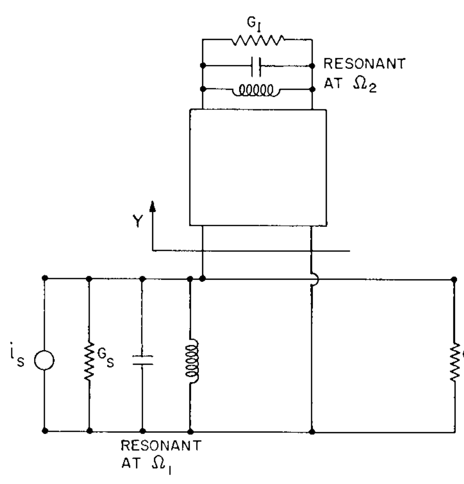


Fig. 24 - Equivalent circuit for calculation of the bandwidth of the NRPA

The equivalent circuit for the bandwidth calculation is shown in Fig. 24. This looks like Fig. 23, except that signal and idler tanks tuned to frequencies Ω_1 and Ω_2 are indicated. Note that the admittance looking into the signal port not including the signal tank is called Y. The actual signal and idler frequencies, which are now assumed to differ from Ω_1 and Ω_2, are designated by ω_1 and ω_2. The load current is given by

$$i_\ell = i_s \frac{G_\ell}{G_\ell + Y_1 + Y}$$

where Y_1 is the sum of G_s and the signal tank circuit. Then the power delivered to the load is

$$P_\ell = |i_s|^2 \frac{G_\ell}{|G_\ell + Y_1 + Y|^2}$$

and the power available from the source is

$$P_s = \frac{|i_s|^2}{4G_s}.$$

The frequency dependent transducer gain $A_{T\omega}$ is by definition P_ℓ/P_s, and thus we have

$$A_{T\omega} = \frac{4G_\ell G_s}{|G_\ell + Y_1 + Y|^2}. \tag{23}$$

Clearly, the frequency dependence of $A_{T\omega}$ is to be found in Y and Y_1.

The details (3) of the calculation will not be given here, but the bandwidth is obtained by setting the denominator (since the numerator is frequency independent) equal to twice its center frequency value and solving for frequency. The result is

$$2\delta' = \frac{G_s + G_\ell - G}{Q_1\left[G_s + G_\ell + (G\Omega_1 Q_2/\Omega_2 Q_1)\right]} \tag{24}$$

where δ' is that particular value of $(\omega - \Omega_1)/\Omega_1 = \delta$ at which the transducer gain has dropped to half its center frequency value, Q_1 is the total Q of the signal tank (i.e., it includes G_s as well as any losses in the resonant circuit), and Q_2 is the Q of the idler tank including all conductance shunting that tank except that seen looking into port 2.

Note that the numerator of this expression is the square root of the denominator of the expression for center frequency transducer gain which was given in Eq. (12). If Eq. (12) is solved for the quantity $G_\ell + G_s - G$ and the result substituted into Eq. (24) we get

$$2\delta\sqrt{A_T} = \frac{2\Omega_2\sqrt{G_s G_\ell}}{\Omega_2 Q_1 G_{T1} + \Omega_1 Q_2 G} \tag{25}$$

where G_{T1} is the total conductance shunting the signal tank (but not including G); that is, it is $G_s + G_\ell$, plus the losses associated with the signal tank, which are usually small when compared with $G_s + G_\ell$.

Since Eq. (25) involves the square root of the transducer power gain and since voltage is proportional to the square root of power, Eq. (25) is often called the gain-bandwidth in analogy to low frequency definition. Usually the Q of the idler tank is large when compared with the Q of the signal tank because the signal tank must of necessity be loaded with G_s and G_ℓ. Thus we can write

$$\text{gain} \times \text{bandwidth} = 2\frac{\Omega_2}{\Omega_1}\frac{1}{Q_2}\sqrt{\frac{G_s G_\ell}{GG}}. \tag{26}$$

For large transducer gain, $G_s + G_\ell \approx G$; thus

$$\text{gain} \times \text{bandwidth} = 2\frac{\Omega_2}{\Omega_1}\frac{1}{Q_2}\sqrt{\left(\frac{G_s}{G_s + G_\ell}\right)\left(\frac{G_\ell}{G_s + G_\ell}\right)}$$

$$= 2\frac{\Omega_2}{\Omega_1}\frac{1}{Q_2}\left(\frac{\sqrt{G_\ell/G_s}}{1 + G_\ell/G_s}\right).$$

As a function of G_ℓ/G_s, this expression is maximum when $G_\ell/G_s = 1$, and for this case we have the final answer for the maximum gain-bandwidth product:

$$\text{gain} \times \text{bandwidth} = \frac{1}{Q_2}\frac{\Omega_2}{\Omega_1}. \tag{27}$$

This is an important formula. It must be remembered that this formula refers to the square root of the power gain and that the bandwidth is the fractional bandwidth.

Let us consider a simple problem using this formula: Is it feasible to build a low-noise NRPA at 400 Mc with an idler frequency at 4000 Mc, a gain of 20 db, and an absolute bandwidth of 0.5 Mc? The transducer power gain equals 100, and the square root of this is 10. A typical Q for a resonant circuit at 4000 Mc would be, say 1000. The fractional bandwidth then equals 10^{-3} which means that the absolute signal bandwidth would be about 0.5 Mc. Thus, it would appear feasible to build such an amplifier.

Circulators, Isolators, and Negative Resistance Amplifiers

At microwave frequencies nonreciprocal circuit elements such as circulators and isolators are available and may be usefully employed with negative resistance amplifiers.

An isolator is a two-port device which passes microwaves with nearly zero loss in one direction and nearly infinite loss in the other direction. At port 1 a waveguide match is seen regardless of what is connected to port 2, and looking in port 2 a match is seen regardless of what is connected to port 1. Isolators can be used with a negative resistance as is shown in Fig. 25.

Note that the apparent source and load conductances which are loading G, namely G_s' and G_ℓ' are nearly constant regardless of changes in G_s and G_ℓ. Thus, this amplifier is more stable against changes in G_s and G_ℓ than the amplifier of Fig. 20. Of course, changes in G are just as important as ever; thus these changes must be minimized by holding the pump power constant and perhaps by stabilizing the diode temperature. There are two other points which should be mentioned: First, we have now a separate input port and output port and second, neither the output conductance nor the input conductance will be negative regardless of the values of G_s and G_ℓ (assuming that the isolator is perfect). The other properties of the negative resistance amplifier such as bandwidth and noise figure are unchanged.

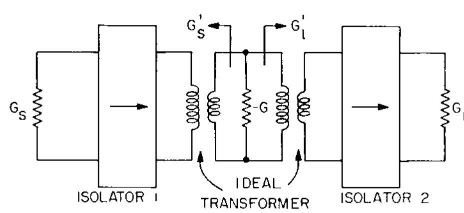


Fig. 25 - A combination of isolator and a negative conductance to produce a more stable low-noise amplifier

A circulator is a lossless three-port device. It is best described in terms of waves. A wave entering port 1 is transmitted without loss to port 2 only, a wave entering port 2 is transmitted without loss to port 3 only, and a wave entering port 3 is transmitted without loss to port 1 only. Since the circulator is lossless it is intuitively plausible that it will not affect the minimum noise figure.

A circulator is used with a negative resistance amplifier as shown in Fig. 26. Ordinarily, the source and load are tuned to the characteristic conductance of the waveguide G_c. The negative conductance G is shown transformer-coupled to the circulator so that the total conductance loading G, denoted by G_T, may be adjusted as desired. An input wave enters the circulator from G_s and is directed by the circulator to G. It is reflected from G with amplification. The reflected, amplified wave is then directed by the circulator into the load G_ℓ. The circulator, like the isolators, improves the stability against changes in G_s and G_ℓ and we can see this in the following way: For simplicity, we will consider only a change in G_s. This is a practical case, because in most systems the antenna impedance is more likely to change than is the impedance of the succeeding stage. For example the impedance of a rotating radar antenna changes as it looks at different parts of the landscape.

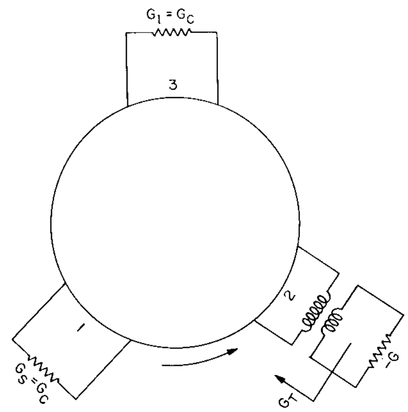


Fig. 26 - A combination of a circulator and a negative conductance to produce a more stable low-noise pre-amplifier and to increase the gain-bandwidth product

This sytem will oscillate if $G \geq G_T$, and therefore the proper question to ask is: How does G_T change for a given change in G_s? We can answer this by replacing (in thought) G by a signal generator. This generator sends a wave into arm 2 and the question now becomes: In what way is the wave that returns to arm 2 dependent upon G_s? It is apparent that the wave that enters arm 2 is directed into G_ℓ. Since G_ℓ is matched to the guide there is no reflection from G_ℓ and therefore no way for energy to return to arm 2, regardless of the value of G_s. Thus the system is insensitive to changes in G_s.

A second advantage of circulator coupling is an increase in the gain-bandwidth product, which is not achieved when isolators are used. It can be shown that the gain-bandwidth product is just doubled (2c). Since the circulator offers these two advantages, it is almost always used in frequency ranges in which circulators are readily available.

It should be noted, however, that to the extent the circulator or isolators are not ideal, the noise figure will be increased, and we must take this effect into account.

Quasidegenerative Negative Resistance Parametric Amplifier

Suppose that the signal and idler frequencies are nearly equal. It appears that it would not be necessary to have a separate tank circuit for the signal and the idler, and that a single tank circuit wide enough to accept both the signal and the idler frequency could be used. This device may be called the quasidegenerative NRPA. The prefix quasi is employed to indicate that the signal and idler frequencies are not exactly equal. For brevity, the term degenerate amplifier (DA) is often used. In Fig. 27 a simplified circuit for a DA is given for the case in which a circulator is employed for bandwidth and stability. It is important to note that both signal and idler power will be delivered to the load (which normally will be the input to the next amplifier). It is possible to speculate on the possibility of sending both the signal and image through the rest of the receiver. However, for reasons which will not be discussed, any such receiver would not be a linear receiver and we wish to limit the discussion to linear receivers. Thus, it is necessary

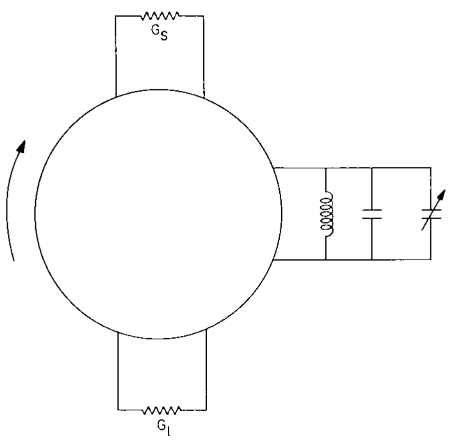


Fig. 27 - Equivalent circuit for a degenerate amplifier used with a circulator

that the postamplifier, which is fed by the degenerate amplifier, have a passband which accepts only the signal frequency.

A gain vs frequency sketch for this type of amplifier is shown in Fig. 28. The pump frequency is approximately twice the center frequency. The useful input bandwidth is indicated by the shaded portion. The center portion of the response curve cannot be used; that is, the amplifier cannot be used at the resonant frequency. This curious fact comes about as follows: Suppose that the signal consists of a carrier plus an upper sideband. The image will consist of a signal plus a lower sideband because of the sideband inversion effect. If the signal carrier were too close to the resonant frequency, the signal sideband and the image sideband would overlap, and the signal and image information could not be separated by succeeding amplifiers. Thus, it is necessary to maintain a guard band around the center frequency. Obviously, the width of this guard band will depend on the bandwidth of the incoming information, that is on the width of the signal spectrum.

In the DA we no longer use separate circuits to distinguish between the signal and idler frequencies. The conductance loading the idler, which we formerly called G_I, is now simply the total conductance loading the signal circuit. That is, the signal and idler loading conductances are now physically identical. This means that G_s will not only furnish input noise to the amplifier (which we do not blame on the amplifier) but will also introduce noise into the idler channel; in other words, it will act as G_I. This idler noise is to be blamed on the amplifier; that is, it contributes to increasing the noise figure.

The noise figure of the DA is thus the same as the noise figure of the regular NRA, which has an equal signal and the idler frequency. Thus, from Eq. (20) we have

$$F = 1 + 2t_I ;$$

but it is important to remember that here the idler cannot be separately cooled; the idler noise temperature ratio is simply the antenna noise temperature ratio.

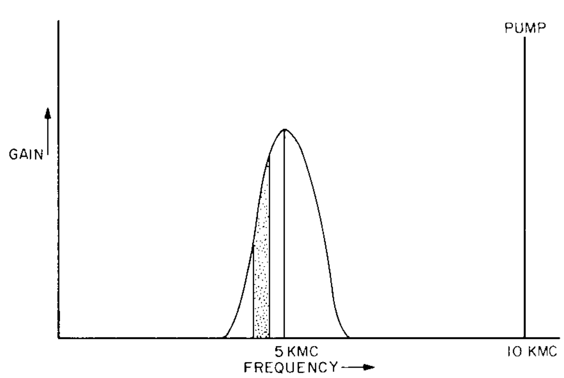


Fig. 28 - Gain characteristics for a 5-kMc degenerate amplifier

The degenerate amplifier, however, has another noise figure called the double channel noise figure (in contrast to the usual noise figure which is called the single channel noise figure). Suppose it were somehow possible to introduce the signal simultaneously into the signal channel and the idler channel. Then the effect of the idler channel would not be a total loss, since it would contribute signal information into the system. It seems reasonable that the double channel noise figure would be smaller than the single channel noise figure. In fact, if the signal and idler channels are nearly equivalent (in the electrical sense), the signal-to-noise ratio would be doubled. Thus, the noise figure would be halved; that is, the double channel noise figure would be 3 db less than the single channel noise figure.

There is one very important application in which information is easily fed in both channels so that the double channel noise figure is appropriate. This was discussed in the section on noise under the heading "The Extended White Noise Source" (p. 15). Here the "signal" consists of a wideband noise source which is relatively frequency independent. The amplitude of this noise represents the noise temperature of the astronomical entity being studied. Since the source is so wideband, it is a simple matter to feed this energy into both channels—in fact, in the DA it would be difficult to avoid feeding it into both channels. It seems reasonable that the double channel noise figure should apply in this case. If, for example, the spreading resistance were nearly zero; that is, if the parametric diode were nearly a pure capacitance, then there would be no noise from the diode and of course no idler noise—therefore a noise figure of zero db would be expected.

It does seem, however, that in the case of a radar type signal the single channel noise figure would apply. It is difficult to conceive of a conventional radar signal which could be fed into both channels without the overlap of signal and idler sidebands mentioned previously.

The concepts of single and double channel noise figure are not new. They arise in fact in any mixer. Consider for example, the conventional radar mixer. Suppose the i-f is 30 Mc and the local oscillator is operating at 10,030 Mc. Then there will actually be two possible input channels, one at 10,000 Mc and one at 10,060 Mc. These frequencies are close together as far as microwave circuitry is concerned, and unless microwave filters are used (and they usually are not) impedances will be the same for both channels. That is, the two channels will be electrically equivalent. Normally, one of the channels is not used but, of course, it contributes noise exactly analogous to idler noise. Such a receiver will exhibit a 3-db improvement in noise figure if it is used for the type of astronomical observation we have been discussing.

Here is a quick review of the degenerate amplifier: The DA is a particular form of the NRPA in which the signal and idler frequency are both nearly equal to one-half the pump frequency and are both contained within the passband of one tuned circuit. Two kinds of noise figure are defined for this amplifier—the single-channel noise figure, which applies to most cw type signals such as radar, communications, etc., and the double channel noise figure which is used when the "signal" consists of broadband noise from an object whose temperature is sought. The single channel noise figure is 3 db larger than the double channel noise figure.

Broadbanding the Negative Resistance Parametric Amplifier

The theoretical bandwidths of the degenerate and nondegenerate NRPA are, of course, quite small and the experimental bandwidths are usually somewhat smaller even than the theory predicts. A natural question arises as to whether these narrow bandwidths are an intrinsic property of NRPA's or whether they are caused by the fact that we have employed simple single-tuned resonant circuits for sorting out the various frequencies which are involved in PA operation. In the final analysis a NRA will give gain at any frequency at which a negative resistance is present. Equation (19) indicates that in the case of the NRPA the frequency dependence of the negative conductance is certainly not great enough to account for bandwidths of only a few percent. The frequency dependence of the transducer gain is contained in the denominator of Eq. (23) in Y and Y_1, where Y is the admittance of the input tank circuit and is strongly frequency dependent, as in any single tuned circuit. We have not given an expression for Y_1, which is the admittance coupled into the signal circuit from the idler circuit, but it can be shown that Y_1 is strongly frequency dependent, mostly because the idler circuit is also a single tuned circuit.

Thus it would appear that if bandpass filters, with their broader, more flat-topped response, were used instead of the single-tuned circuits an increased bandwidth would result. This reasoning is correct, and although the broadbanding of NRPA's by this means is still in the early development stage, it has been shown both experimentally and theoretically (4) that bandwidths of the order of 20 percent are feasible without significant sacrifice of the low noise properties.

Naturally, this procedure is simplest in the case of the DA because only one filter is involved, and most of the early experimental broadband work was done on the DA. The design procedure is at present rather complex, and a detailed analysis will not be given, but some experimental results will be given in a later section.

THE EFFECT OF SPREADING RESISTANCE

Introduction

In the section on parametric diodes, it was pointed out that all practical parametric diodes have a spreading resistance R_s in series with a nonlinear capacitance, and some of the consequences of this fact were noted. It is to be expected that all the properties of any type of PA will be affected by the spreading resistance for high enough frequencies. The gain, bandwidth, and noise figure will be different for finite R_s, and in addition, there will be some new design considerations. For example, the pure nonlinear capacitor ($R_s = 0$) does not absorb pump power; when R_s is added, however, it will absorb pump power proportional to the square of the pump frequency. This can be an important design consideration.

In principle, it would be relatively easy to go over the calculations we made earlier and consider in detail the effect of a finite R_s. It is a relatively straightforward problem in linear circuit theory. The calculations are rather involved, however, and only the results are presented here.

Two parameters are important for describing the parametric diode when the effects of spreading resistance are included. The first, γ, is nearly equal to C_1/C_0. The quantity γ is essentially a measure of the degree of nonlinearity of the diode when driven by the pump. If the pump drive is weak, γ will be small, and as the pump drive is increased, γ will increase. Of course, the pump drive cannot be increased indefinitely or the diode will be driven into conduction, so the maximum value of γ is limited. A reasonable value of γ for an exceptionally good diode driven hard is 0.3; a typical value is 0.25.

The second parameter which is important in describing the parametric diode is the cutoff frequency ω_c. This is the frequency at which the spreading resistance is equal to $1/(\omega C_0)$; in other words, at ω_c, the impedance of the spreading resistance is equal to the impedance of the fictitious capacitor C_0, where C_0 is just the average value of the diode capacitance through the pump cycle. In calculations it is customary to define an effective Q at the lowest frequency ω_1 as follows:

$$Q = 1/(\omega_1 C_0 R_s) . \tag{28}$$

Since by definition $C_0 R_s = 1/\omega_c$, we have

$$Q = \omega_c/\omega_1 . \tag{29}$$

Thus, if we know the cutoff frequency of a diode, we can always calculate the proper Q corresponding to the frequency ω_1. Note that if ω_c is large, which it will be if either R_s is small or C_0 is small, Q will be large. We now discuss the detailed effect of these two parameters on the properties of the upconverter and the NRPA.

The Upconverter

Transducer gain (2d) is given by

$$A_T = \frac{f_4}{f_1} \frac{X}{\left[1 + \sqrt{1 + X}\right]^2} \tag{30}$$

where $X = (f_1/f_4)(\gamma Q)^2$ and $f_1 = \omega_1/2\pi$.

From Eq. (30), we can obtain an upper bound for the gain by letting Q approach infinity. It is easily seen that for this case, X approaches infinity and thus the gain approaches f_4/f_1. This agrees with our treatment for R_s equal to zero.

But we can derive a still more interesting result from Eq. (30). Suppose we have a given diode being driven hard; that is, γ and f_c are given. Suppose that the input frequency f_1 is also given. Assume that we are willing to make the pump frequency (and the output frequency) as high as is necessary to reach the maximum gain. What is the maximum possible gain if we let f_4/f_1 approach infinity? Intuition tells us that the gain will not approach infinity but will approach some finite value because the diode must, after all, operate at f_4, and as f_4 becomes larger and larger, the diode becomes less efficient because of the spreading resistance.

From Eq. (30) the gain $(A_T)_\infty$ becomes

$$(A_T)_\infty = (1/4)(\gamma Q)^2 . \tag{31}$$

Remember that Q is defined at f_1. Insofar as transducer gain is concerned, γQ is considered to be a very important figure of merit for the diode because for a given input frequency it determines what the maximum gain can be.

Consider a practical problem: A diode with a cutoff frequency of 20 kMc and $\gamma = 0.25$ is available for use in an upconverter of input frequency of 0.5 kMc. What is the maximum possible gain, and what is a reasonable pump frequency?

The Q is f_c/f_1, which in this case equals 40; $(\gamma Q)^2$ equals 100; thus the maximum gain is 25. The gain equals 25 when f_4 is infinite, but Eq. (30) gives the gain for any finite value of f_4; therefore we can substitute tentative values of f_4 into Eq. (30) until the gain is as close to 25 as desired.

Let us now consider bandwidth (2e). The fractional bandwidth for a high quality (high Q) diode is given by

$$2\delta' = 2\gamma\sqrt{f_4/f_1} . \tag{32}$$

This equation predicts very large bandwidths for the upconverter; in fact, for large f_4/f_1, the predicted bandwidth is absurdly large. This is due to the failure of some of the assumptions of the derivation. We will simply conclude that very large bandwidths are possible for the upconverter and that the bandwidth increases with an increase in γ and f_4/f_1. At present no simple method of predicting bandwidth exists.

Finally, we consider noise figure. Assuming that $\gamma Q >> 1$, the noise figure for optimum source impedance is given by (2f)

$$F^0 = 1 + \frac{2t_D}{\gamma Q} \tag{33}$$

where t_D is the noise temperature ratio of the spreading resistance. Good agreement with experimental data has been obtained by assuming that t_D is the ambient temperature of the diode. As an example, suppose the diode is operated at room temperature and $\gamma Q = 10$. Then the noise figure is approximately 1.2. Note that the quantity γQ again appears as a sort of figure of merit for a given input frequency.

The Negative Resistance Parametric Amplifier

Gain is not of great interest because it can always be made infinite for any negative resistance amplifier. However, gain bandwidth (2g) is of great interest and it is given by

$$2\delta'(A_T)^{1/2} = \frac{\gamma}{2}\left(\frac{f_2}{f_1}\right)^{1/2} . \tag{34}$$

This formula applies to the case in which a circulator is not used. As an example, suppose that the gain is 20 db, γ is 0.25, and the idler frequency is four times the signal frequency. The maximum possible bandwidth will then be about 2.5 percent.

We next shall consider the noise temperature ratio t of the NRPA when R_s is finite. The general solution is too complicated to be of much value to us, but there is one rather practical special case which gives simple results. If we assume that the losses in the idler circuit are small compared with R_s, then the expression for t becomes*

$$t = t_D \frac{K^2 \Omega^2 + 1}{K^2 \Omega - 1} \tag{35}$$

where t_D is the noise temperature of R_s; $K = \gamma Q$ and $\Omega = \omega_s/\omega_I$. It is natural to suppose that t_D is simply the ambient temperature of the diode. This assumption does seem to agree with the experimental evidence; that is to say, cooling the diode does lower t provided that the diode will continue to operate properly at the reduced temperature.

We note that γQ appears again as a figure of merit for the diode. For a given diode operated at a given temperature and signal frequency, t_D and K are fixed. The designer still has freedom however, in choosing the idler frequency (i.e., in choosing Ω).

It is clear, that unlike the case for $R_s = 0$, the optimum idler frequency is not infinite. In fact, it appears that there is a finite value of the idler frequency which will give the minimum value for t. This can be found by the usual method; that is, by taking the derivative of Eq. (35) with respect to Ω and setting it equal to zero. Assuming that K is much greater than one, we have for the optimum value of Ω

$$\Omega_{op} = \frac{1}{\gamma Q} . \tag{36}$$

The optimum value of Ω can in turn be substituted back into Eq. (35), thus giving us the optimum value of t

$$t_{op} = \frac{2t_D}{\gamma Q} . \tag{37}$$

*Equation (35) can be deduced from Knechtli's Eq. (40) with $\mu = 0$ (small idler circuit losses) and $\alpha = 1$ (high gain) along with Eq. (41), which permits elimination of μ_s. This procedure is permissible because my t is identical with Knechtli's t_{ni}, even though the latter describes a four-terminal amplifier. See Ref. 5.

Finally, when the optimum value of t is substituted into Eq. (16) it will yield an expression for the optimum or lowest noise figure:

$$F_{op} = 1 + \frac{2t_D}{\gamma Q} . \tag{38}$$

In practice, it may not always be feasible to use the optimum idler frequency, because the pump frequency required may be in a difficult frequency range. In this case, one can tentatively choose an idler frequency which does give a convenient pump frequency. The noise temperature ratio for this idler frequency can be computed from Eq. (35) and compared with the lowest possible value given by Eq. (37), and thus the penalty for not using the optimum idler frequency can be adjudged to be either acceptable or not acceptable. This procedure can be repeated until an acceptable solution is obtained.

Equation (38) is one of the most important equations of NRPA theory. We will use it to estimate the noise figures expected from the best NRPA's. At present commercially available varactor diodes have cutoff frequencies up to about 200 kMc. If we assume that t_D equals 0.25 (liquid nitrogen temperature) and γ equals 0.25, then at X-band a 200-kMc-diode will have an F_{op} equal to 1.1. There are very few systems that require a lower noise figure.

Pump Power

Since pump current must flow through the spreading resistance, pump power P_p is dissipated in the spreading resistance. By making two simplifying assumptions, we can estimate what the pump power must be. First, assume that the pump voltage which appears across the diode terminals is sinusoidal, and second, assume that the diode is biased halfway between zero and breakdown voltage V_B. A large γ is always desirable, and clearly γ has its largest value if the pump voltage amplitude is equal to $V_B/2$. Then, the voltage across the spreading resistance V_s is

$$V_s = \frac{V_B}{2}\left[\frac{R_s}{R_s - j/(\omega_3 C_0)}\right] ,$$

where ω_3 is the pump frequency. Since the capacitance varies throughout the pump cycle, it is not clear what value of capacitance should be substituted into the above formula, but C_0 appears to be a good approximation. At any rate we expect only a rough estimate. The power dissipated in R_s is $|V_s|^2/(2R_s)$; thus

$$P_p = \frac{V_B^2}{4} \frac{R_s^2}{(R_s^2 + 1/\omega_3^2 C_0^2)\, 2R_s}$$

$$= \frac{V_B^2}{8} \frac{R_s}{R_s^2 + 1/\omega_3^2 C_0^2} .$$

Divide the numerator and denominator by R_s^2, and recall that the cutoff angular frequency ω_c is given by $1/C_0 R_s$. We have

$$P_p = \frac{V_B^2}{8} \frac{1/R_s}{1 + f_c^2/f_3^2} .$$

For frequencies well below the cutoff frequency,

$$P_p = \frac{V_B^2}{8R_s} \frac{f_3^2}{f_c^2}. \tag{39}$$

The important point is that pump power varies with the square of the pump frequency, and this may be a reason for avoiding high pump frequencies.

APPLICATIONS

Is a Parametric Amplifier Suitable?

Before we can intelligently discuss practical applications of parametric amplifiers, we need to consider in what sort of situation parametric amplifiers are a good choice, and in what sort of situations other amplifiers are a better choice.

To begin with, the frequency range in which parametric amplifiers appear to be applicable extends from about 100 Mc to 10,000 Mc. We must keep in mind that parametric amplifiers have serious drawbacks in comparison with conventional amplifiers. The NRPA has stability problems, and the upconverter employs radically different frequencies at the input and output. Below 100 Mc, vacuum tubes and transistors may be preferable. Furthermore, below 100 Mc there is not much demand for very low noise figures. The reason for this is that the sky temperature, which largely determines the antenna temperature, is rather high. According to the discussion in the section on noise, a low noise receiver is of value only when the antenna temperature is low.

In this connection, Fig. 29 is a rough plot of the sky temperature as a function of frequency (6). The lower curve refers to the temperature looking vertically and the upper curve refers to a temperature 60° to the vertical. Notice that a broad minimum is centered at 3 kMc at which frequency the sky temperature (for vertical incidence) has the very low value of approximately 2°K. We will return to this graph later when we consider several examples of practical receiver problems.

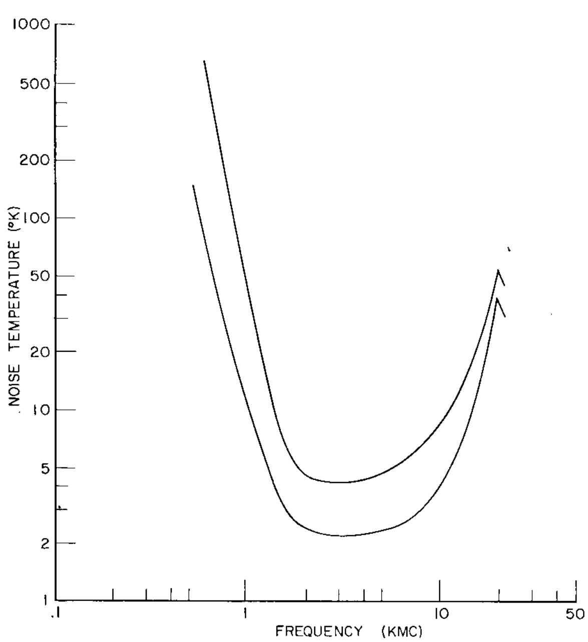


Fig. 29 - Noise temperatures of the sky

Above 10 kMc there are very few practical systems of any kind. Furthermore, the cutoff frequencies of presently available diodes do not promise very low noise figures above 10 kMc.

Let us now consider the competition in the range from 100 Mc to 10,000 Mc. Most of the competition comes from various kinds of vacuum tubes such as triodes, backward wave amplifiers, traveling wave tubes; and the maser. Eventually the tunnel diode amplifier will also be considered, but it is a little too early to be able to say where this amplifier will fit in. The tunnel diode amplifier has stability problems just like the NRPA. It can be extremely simple and lightweight, and its noise figure is similar to that of the vacuum tube.

Generalizations are risky, but it appears that in the frequency range in which we are interested, the maser is the quietest, the vacuum tube is the noisiest and the parametric amplifier falls between the two. At 3 kMc, typical values for the three amplifiers are: the maser, almost too low to measure; the PA, 1.5 db; the vacuum tube, 3 db.

However, with respect to bandwidth, the vacuum tube (traveling wave tube) rates the highest, the PA next, and the maser last.

With respect to simplicity, the tube would seem to rate highest (no pump, no extraneous frequencies to deal with), the PA next, and the maser a very poor last, since both very low temperatures and a magnet are required.

At this time we are not attempting to consider which type of PA would be optimum but only to choose between tubes, masers, and PA's. The three following examples might be helpful:

1. A receiver is to be designed for a scatter communications system in a large land based installation. The narrow beam antenna is to be at an angle of about 30° to the horizontal. The center frequency is to be 3000 Mc and the bandwidth is to be 50 kc. The stability requirements are modest. What type of amplifier should be used? Referring to Fig. 29 we see that we are going to have a very low antenna temperature. Equation (3) is applicable here and indicates that when the antenna temperature T is very low, a very low noise figure receiver should be used. Thus, the maser is suggested, and in a large land-based installation we could handle the weight, cost, and complexity of the maser.

2. A receiver is required for an airborne search radar. The antenna temperature is about 200°K, and the frequency is 1000 Mc. What type of amplifier should be used? The antenna temperature is considerably higher here—in fact, too high to make the maser, with its weight and complexity, at all attractive. At the same time, modest bandwidths are usually acceptable in search radars, and Eq. (3) indicates that a lower noise figure than that usually obtainable from tubes would appreciably increase the sensitivity. Some type of PA is indicated.

3. A short range X-band tracking radar receiver is needed for a system in which very accurate range information is desired. The antenna temperature is near room temperature. What type of amplifier should be used? For good range resolution an ordinary radar system needs fast rise time pulses. This, in turn, means large rf bandwidth. Although the outlook for broadband PA's is good, they are still in the experimental stage. Furthermore, since the antenna temperature is rather high, we can accept the larger noise figure of tubes. It appears that a vacuum traveling wave tube is called for here.

What Kind of Parametric Amplifier?

Let us now assume that a PA will be used. How do we choose between an upconverter and a NRPA, and if we choose the latter, how do we choose between a straight NRPA, with separate idler circuit, and a degenerate amplifier? We must consider the system requirements before making a final choice, but some rather general remarks can be made even though they are largely in the nature of a review.

The first step is to choose between the upconverter and some kind of NRPA. With respect to noise figure, Eqs. (33) and (38) indicate that the two are just about equal. The NRPA is inherently unstable and will require some kind of isolation if stability is important. It has high gain, but the bandwidth (at least for the simpler forms) is rather modest.

The upconverter on the other hand is inherently stable, but the gain is limited. For the most part, the choice between the upconverter and NRPA can be made on the basis on the operating frequency. Assume that at least 10 db of gain is needed to override the noise contribution of the second stage. Equation (31) shows that it may not be possible to acquire this much gain if γQ is small. Anyway, if the output frequency is too high, it becomes increasingly difficult to design a second stage with an acceptable noise level. In addition, the required pump power may be too large. For these reasons, the practical limit for the input frequency of an upconverter is about 1000 Mc.

Above 1000 Mc, then, a NRPA should be used, and if stability is important a circulator or isolators will also be needed.

One rather strong statement can be made concerning the choice between the degenerate and the nondegenerate amplifier. If the system is one in which the double-channel noise figure is appropriate, then the degenerate amplifier is the correct choice. This is because the double channel noise figure of the degenerate amplifier is less than or equal to the noise figure of the nondegenerate amplifier. Even in cases in which the single-channel noise figure is appropriate, the degenerate amplifier may be the better choice because of its greater simplicity (only one tuned circuit) and because it is easier to broadband. Furthermore, if the antenna temperature is low, then the noise fed into the idler channel will be correspondingly low; in other words, in the degenerate amplifier, a cool antenna is equivalent to a cool idler. The degenerate amplifier does have the disadvantage that somewhat more than half the bandwidth must be reserved for the idler in order to prevent both the idler and signal from being fed into the second stage.

The main advantage that the nondegenerative amplifier has over the degenerative amplifier in single channel operation is the possibility of minimizing idler noise and thus minimizing the single channel noise figure. This can be done by making the idler frequency much greater than the signal frequency. As Eq. (38) indicates this procedure has a point of no return when the idler frequency becomes high enough so that we run into diode high frequency effects caused by the spreading resistance (i.e., when γQ is small). At present, for most single channel applications the nondegenerative PA is superior, and up to X-band it is used in most commercial equipment.

For bandwidths of the order of 20 percent, the simple tuned circuits will not be sufficient. The filter technique is the most promising broadband technique.

Examples

Since PA's are used mostly in the microwave region, we should not expect real amplifiers to bear a physical resemblance to the lumped equivalent circuits that are so useful in theoretical analysis. Instead of inductances and capacitances we should expect resonant cavities, and instead of a physical lumped conductance G_s (source conductance) we should expect a transmission line leading to an antenna, and so forth. In microwave circuits the distinction between shunt and series connections becomes blurred—an impedance which is small at one point in a transmission line appears large a quarter wavelength away. In particular, there is nothing special about the parallel-tuned equivalent circuit of Fig. 18; the circuit of Fig. 30 could have also been analyzed and the results would have been substantially the same.

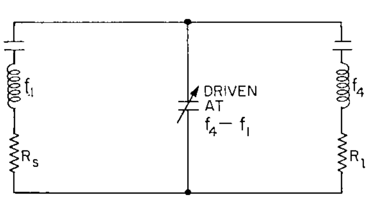


Fig. 30 - Equivalent circuit for series tuned parametric upconverter

As a first example, consider the parametric upconverter developed by Greene and Lombardo (7). This has a 400-Mc input and 9400-Mc output with a gain of 11 db, a noise figure of 0.7 db, and a bandwidth of 22 Mc. The pump (Fig. 31) is fed into the E-arm of the hybrid, with the result that the pump energy reaches the two parametric diodes out of phase. This is easily taken care of by using reversed polarities, so that in essence the diodes are pumped in phase. The two diodes are also fed in phase with the 400-Mc input signal. The 9400 Mc generated in each diode is fed out the H-arm and into the receiver. As the theory requires, a filter is provided in this arm to short-circuit the lower sideband.

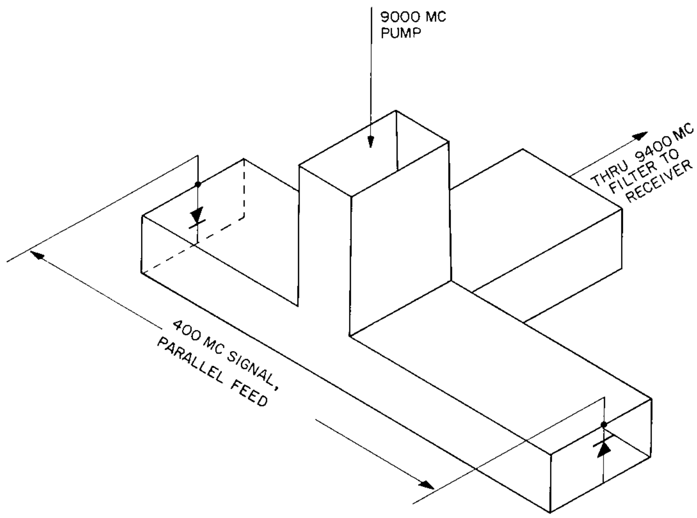


Fig. 31 - Example of a practical upconverter

As a second example, the 5000-Mc degenerate amplifier of Uenohara (8) is pictured in Fig. 32. The figure shows the narrow side of a piece of 10,000-Mc waveguide which is mounted on top of a piece of 5000-Mc waveguide. The diode is mounted so as to be partially in both guides. An adjustable iris in the 5000-Mc guide is used for signal tuning and the diode is really in a resonant cavity which is wide enough for both the signal and the idler. Not shown is a circulator which is connected to the output. The usable bandwidth is 15 Mc, the gain is 20 db, and the single channel noise figure is 5 db. Of course, these figures all refer to one particular set of tuning conditions. The double-channel noise figure must be 2 db.

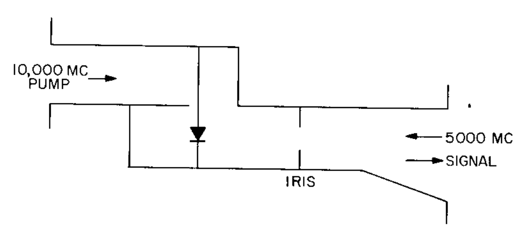


Fig. 32 - Example of a practical degenerate amplifier

A nondegenerative amplifier (2h) at about 2.8 kMc is illustrated in Fig. 33. The pump and idler circuits are in waveguide, and in Fig. 33 we are looking at the broad face of the guide. The signal is introduced coaxially through a low pass filter. The idler circuit is essentially a resonant cavity terminated at one end by the metallic plunger and at the other end by the K_{μ}-band guide and the coaxial cable. However, the idler energy cannot enter the signal circuit because of the low pass filter, and it cannot enter the pump circuit because the pump guide is below cutoff at the idler. Thus, insofar as the idler is concerned, a pure reactance is seen looking to the left and to the right, and this is descriptive of a

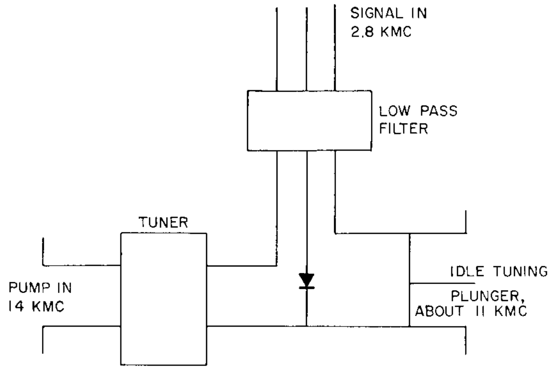


Fig. 33 - Example of a practical nondegenerate NRPA

resonant cavity. The pump energy can undoubtedly enter the idler cavity, but this causes no difficulty. Of course, the signal energy cannot enter either the pump or idler circuits because both the X-band and K_μ-band guide are below cut-off at 2.8 kMc. As is customary, G_I consists of normal losses in the idler circuit.

The fact that the signal circuit is not single tuned indicates that this amplifier constitutes, at least in a rudimentary way, an attempt at broadbanding. It follows that the bandwidth expressions we have derived do not apply.

There is really "no such animal" as an exact equivalent circuit for a microwave device, but Fig. 34 is an attempt to indicate a rough equivalent circuit for Fig. 33. At this signal frequency (2.8 kMc) commercial circulators and isolators are available, and for good stability a circulator was used with this amplifier. It may also be noted that this amplifier has a rather wide tunable bandwidth, since the signal circuit need not be retuned and the idler is easily tuned with the plunger. The figures on this amplifier are as follows: The single channel noise figure equals 2.2 db; the bandwidth equals 70 Mc at a gain of 15 db; and the tunable bandwidth equals 200 mc.

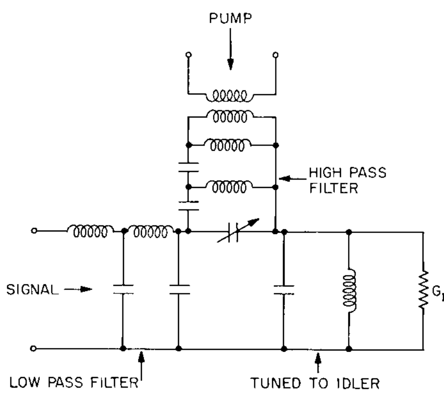


Fig. 34 - Approximate equivalent circuit for Fig. 34

It seems fitting to conclude with some very recent experimental data on the lowest noise parametric amplifier ever reported. Hanson (9) has reported an overall single channel effective noise temperature at 4 kMc of 20°K. This includes noise contributions from a lossy input line and from a circulator. Making a reasonable estimate of these latter contributions he arrived at 2°K as the effective noise temperature of the parametric amplifier alone. This amplifier employs a point contact gallium arsenide diode cooled to liquid helium temperature. The author emphasizes the fact that this amplifier is much simpler than a maser amplifier because the cryogenic system is simpler and because there is no magnet.

This very important work strongly indicates that in the near future the maser will become obsolete, even in extremely low noise applications, and its functions will be taken over by the NRPA.

REFERENCES

1. Van der Ziel, A., "On the Mixing Properties of Non-linear Condensers," J. Appl. Phys. 19:999-1006 (Nov. 1948)

2. Blackwell, L. A. and Kotzebue, K. L., "Semiconductor-Diode Parametric Amplifiers," Prentice-Hall: Englewood Cliffs, N.J., 1961

a. p. 9	e. p. 46
b. p. 43	f. p. 48
c. p. 57	g. p. 57
d. p. 41	h. p. 149

3. Heffner, H. and Wade, G., "Gain, Band Width, and Noise Characteristics of the Variable-Parameter Amplifier," J. Appl. Phys. 29:1321 (Sept. 1958)

4. Mattaei, G.L., "A Study of the Optimum Design of Wide-Band Parametric Amplifiers and Up-Converters," IRE Trans. MTT-9:23-38 (Jan. 1961)

5. Knechtli, R.C. and Weglein, R.D., "Low-Noise Parametric Amplifier," Proc. IRE, 48:1226 (July 1960). Equation (35) can be deduced from Knechtli's Eq. (40) with $\mu_i = 0$ (small idler circuit losses) and $\alpha = 1$ (high gain) along with Eq. (41), which permits elimination of μ_s. This procedure is permissible because my "t" is identical with Knechtli's "t_{ni}" even though the latter describes a four-terminal amplifier.

6. Reed, E.D., "Diode Parametric Amplifiers," Semiconductor Products 4:41 (Feb. 1961)

7. Greene, J.C. and Lombardo, P.P., "Low Noise 400-Mc Reactance Amplifiers," The Microwave Journal 2:28 (May 1959)

8. Herrmann, G.F., Uenohara, M., and Uhlir, A., "Noise Figure Measurements on Two Types of Variable Reactance Amplifiers Using Semiconductor Diodes," Proc. IRE. 46:1301-1303 (June 1958)

9. Hanson, D.C., Fink, H.J., and Uenohara, M., "Varactor Diode Amplifier at Liquid Helium Temperature," Digest of Technical Papers, 1963 International Solid-State Circuits Conference, Winner: New York, 1963

Appendix A

INPUT AND OUTPUT ADMITTANCE OF AN UPCONVERTER

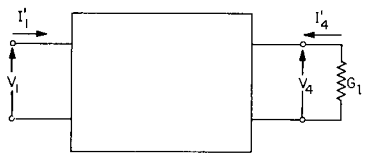


Fig. A1 - Circuit for the calculation of the output and input admittance of an upconverter

Figure A1 shows the circuit for the calculation of input admittance Y_i. Positive polarities of current and voltage are conventional. The load G_ℓ imposes the condition $I_4' = -G_\ell V_4$. From Eq. (9a) we have $Y_i = I_1'/V_1 = j\omega_1 C_1 V_4/V_1$. From Eq. (9b) we have $V_1 = I_4'/(j\omega_4 C_1) = -G_\ell V_4/(j\omega_4 C_1)$.

Substituting this expression for V_1 into the expression for Y_1 yields

$$Y_i = \frac{\omega_1 \omega_4 C_1^2}{G_\ell} .$$

Clearly, to obtain the output admittance Y_o, we need merely substitute ω_1 for ω_4 and vice versa (which has no effect) and substitute G_s for G_ℓ. Thus

$$Y_o = \frac{\omega_1 \omega_4 C_1^2}{G_s} .$$

The significant point is that these are both pure real.

* * *