Shot Noise in p-n Junction Frequency Converters*

By A. UHLIR, JR.

(Manuscript received January 22, 1958)

General equations are derived for the noise figure of a frequency converter using a p-n junction diode with arbitrary minority-carrier storage. A p-n junction with a purely capacitive nonlinear admittance permits, theoretically, noiseless amplification. Structures are suggested for approximating this ideal. A diffused silicon junction diode gave 9 db of gain with a 2-db noise figure in converting 460 to 9375 mc and gave a 3-db noise figure as a 6000-mc negative-resistance amplifier. Nonlinear-resistance diodes cannot amplify but can give low-noise frequency conversion if the local-oscillator drive produces a sharply pulsed current waveform in the diode.

I. INTRODUCTION

The shot noise output of any p-n junction device in any circuit can be calculated by enumerating the possible hole and electron motions and summing the noise energies delivered to the load. This method of analysis will be illustrated by applying it to the dc-biased p-n junction (Section IX). A more complicated example, and the principle topic of this paper, is the general p-n junction frequency converter (Section XI).

The technological significance of two kinds of semiconductor diodes used as frequency converters will be discussed briefly. Section II will explain how this theory of shot noise stimulated the recent development of low-noise microwave amplifiers using the nonlinear capacitance of p-n junctions. Section III mentions the nonlinear resistor, for which shot noise appears to be a vital problem.

The "particle-method" formulation of p-n junction theory in terms of the statistics of individual carrier motions (Section VI) is an alternative to the diffusion equations ordinarily used in the analysis of p-n junctions. Either approach leads directly to the impedance of the junction

* Supported in part by the U. S. Army Signal Corps under contract DA 36-039 sc-5589. Preliminary reports of this work were given at the Microwave Crystal Rectifier Symposium, Fort Monmouth, N. J., February 1956, and the I.R.E.-A.I.E.E. Semiconductor Device Research Conference, Boulder, Colo., July, 1957.

(Sections VII and VIII) and to the small-signal transmission properties of frequency converters (Section X). However, the particle method is the most direct way to calculate shot noise.

After the general equations for noise figure are derived in Section XI, the consequences for nonlinear capacitors and nonlinear resistors are discussed in Sections XII and XIII. One of the virtues of the analysis is the possibility of relating impedances and noise without regard to the structure of the p-n junction. However, a few examples of nonlinear-capacitor structures are given in Section XII.

As explained in Sections IV and V, the particle method and the diffusion equation method deal with just one aspect of p-n junction phenomena: minority-carrier effects, which have varying importance, depending upon the device and its use. Minority carrier effects are nowhere more important than in p-n junction transistors. Thus, the particle method is an excellent way to calculate the shot noise in transistors, although such a calculation will not be given in this article.*

II. NONLINEAR-CAPACITANCE AMPLIFIERS

A time-varying capacitor is an active device that can be used as a linear amplifier of small signals, usually, but not necessarily, with a shift in signal frequency. The power required for the amplification is supplied by the motor, of whatever kind, that varies the capacitance. This principle is used in vibrating-reed electrometers for amplifying very low frequencies.[1,2]

To obtain microwave amplification, one needs a capacitance that varies at microwave or higher frequencies. Such a time-varying capacitance can be obtained by applying a microwave voltage to the voltage-dependent capacitance of a p-n junction diode. Remarkably general gain and power relations can be derived for circuits of this type.[3]

Amplifying diode-frequency converters are not new. During World War II, welded-contact germanium diodes were found to give conversion gain when used in superheterodyne receivers of the type then undergoing extensive development for radar use.[4] With the then-existing technology, it was found that these diodes were not as good as point-contact silicon diodes in the over-all system noise figure that could be obtained. As a result of this observation, point-contact silicon diodes took over the entire radar field until recently, and the manufacture of welded-contact diodes was discontinued.

* For such a calculation see van der Ziel, A. and Becking, A. G. J., Theory of Junction Diode and Junction Transistor Noise, Proc. I.R.E. **46**, March 1958, pp. 589–594.

These results, though entirely empirical, may very well have led to a belief that amplification by diodes was a freak effect and necessarily noisy. In the years following World War II, casual experiments on diode amplification were performed in many laboratories, but low noise was not found, nor was it intensively sought. The amplification effect was put to practical use in transmitting modulators, where noise would not be important.[5]

At this point, the shot-noise theory of p-n junction frequency converters was developed. It provided one logically complete example of how a diode could amplify with arbitrarily low noise. It was recognized that the minority-carrier mechanism assumed in the analysis was probably not as important, at high frequencies, as the depletion-layer capacitance; at the same time, the understanding acquired from the shot-noise theory suggested that the depletion-layer capacitance would also be noiseless.

Subsequent efforts to make improved nonlinear capacitance diodes[6] and to give them a fair circuit evaluation have been extremely well rewarded. A diffused-junction silicon diode has given 9 db of gain with a 2-db noise figure in converting 460 mc to 9375 mc.[7] Also, such a diode has given 35 db of negative-resistance gain at 6000 mc with a 3-db noise figure (for double sideband input).[7]

In view of the large variety of possible frequency relations, to say nothing of the circuits involving combinations of diodes, it is fortunate that it can be deduced that an ideal nonlinear capacitor is noiseless in any circuit. By p-n junction techniques it is possible to make nonlinear capacitors that are practically ideal, except for a series resistance. Working theories of nonlinear capacitors will therefore be concerned with the noise from this series resistance.[8]

III. NONLINEAR RESISTORS

Point-contact crystal rectifiers are in widespread practical use in microwave superheterodyne receivers. It has long been believed that the useful action of these devices results from nonlinear resistance and, as yet, there is no reason to doubt that this is true.

As far as noise is concerned, there is no difference between the mathematical results for the p-n junction type of nonlinear resistance and the results that would have been obtained from earlier ideas of metal-semiconductor contacts. But the older theory was not carried very far for the calculation of noise figure. Therefore, it seems worthwhile to point out the importance of the local-oscillator waveform.

In the past, most analyses have been made on the assumption of short-circuited harmonics. This rather arbitrary postulation of a waveform is not realizable in practice, but it does give rise to some simple relations between conversion loss and noise figure. Another case, open-circuited harmonics, will be found to have an entirely different relation between conversion loss and noise figure. Thus, the analysis of nonlinear resistors will show, by example, that no general relation between gain and noise figure is true.

In contrast to the nonlinear resistor, a nonlinear capacitor gives low noise for any local-oscillator waveform.

IV. p-n JUNCTION THEORY

As shown in Fig. 1, there are two important methods of analyzing p-n junctions, each complementing the other. The pseudo-equilibrium method is an accurate means of calculating the capacitance-voltage characteristic of a p-n junction if one knows, from other considerations, that it is a nonlinear capacitor. However, this method gives no information about frequency dependence and noise.

The kinetic method begins with an approximate separation of physical mechanisms, as discussed in Section V. This separation makes it relatively easy to calculate the frequency-dependent admittance and noise. The kinetic method can be applied not only to nonlinear resistors and nonlinear capacitors but also to the host of intermediate types of p-n

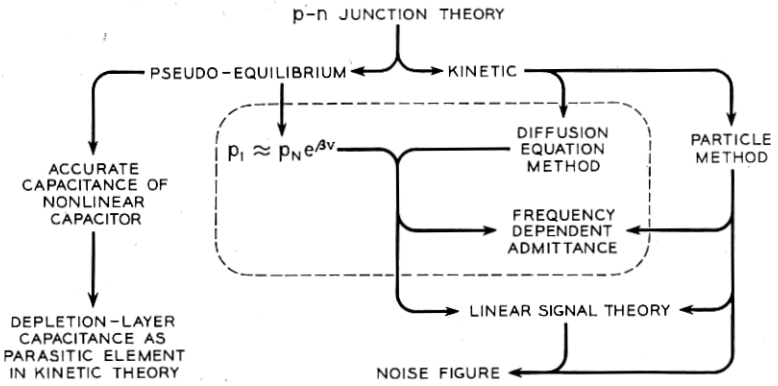

Fig. 1 — Structure of p-n junction theory involved in or related to the calculation of noise figure. The encircled elements constitute the "classical" kinetic theory of p-n junctions.

junctions (including the so-called "simple" p-n junction between homogeneous semi-infinite p and n regions).

The usual mathematical approach in the kinetic method involves solving the diffusion equation, a differential equation for the average motion of carriers. This approach, which will be referred to as "classical" p-n junction theory, is satisfactory for calculating junction admittance. In calculating noise, it seems to be much easier first to calculate the noise due to individual carrier motions and then to sum them. This second procedure will be referred to as "the particle method," since the individuality of the carriers is preserved to a later stage in the calculation.

The particle method gives relations between small-signal junction admittances and the noise figure of frequency converters. The train of reasoning to be used is indicated by the arrows in Fig. 1. However, the connecting lines in Fig. 1 may be followed in any direction. For example, the small-signal admittances may be calculated from the diffusion equation or measured experimentally and used to calculate gain and noise figure.

V. APPROXIMATE SEPARATION OF PHYSICAL MECHANISMS

The equivalent circuit of Fig. 2 denotes the principal physical mechanisms of p-n junction action. The p-n junction is considered to be divided into essentially neutral p and n regions and an intervening space-charge region known as the depletion layer.

The capacitance of the depletion layer may be analyzed by pseudo-equilibrium methods under the approximation that the depletion layer is free of carriers.[9] This capacitance varies with voltage, and amplifying frequency converters can be built that rely on it alone. It is reasonably certain that the depletion-layer capacitance will be very important in all high-frequency nonlinear-capacitance diodes. Nevertheless, we will ignore the depletion-layer capacitance entirely in the following theory of gain and shot noise, except for qualitative remarks. This theory will be developed around the motion of carriers *across* the depletion layer.

The theory of noise thus obtained will illustrate the fact that at least one conceptual type of nonlinear capacitance amplifier is noiseless. It will then perhaps be easy to believe that the depletion-layer capacitance also should not be a source of noise. The analysis should be directly applicable to practical nonlinear resistors, for these must have relatively small values of depletion-layer capacitance for satisfactory operation.

The inescapable importance of the depletion-layer capacitance is con-

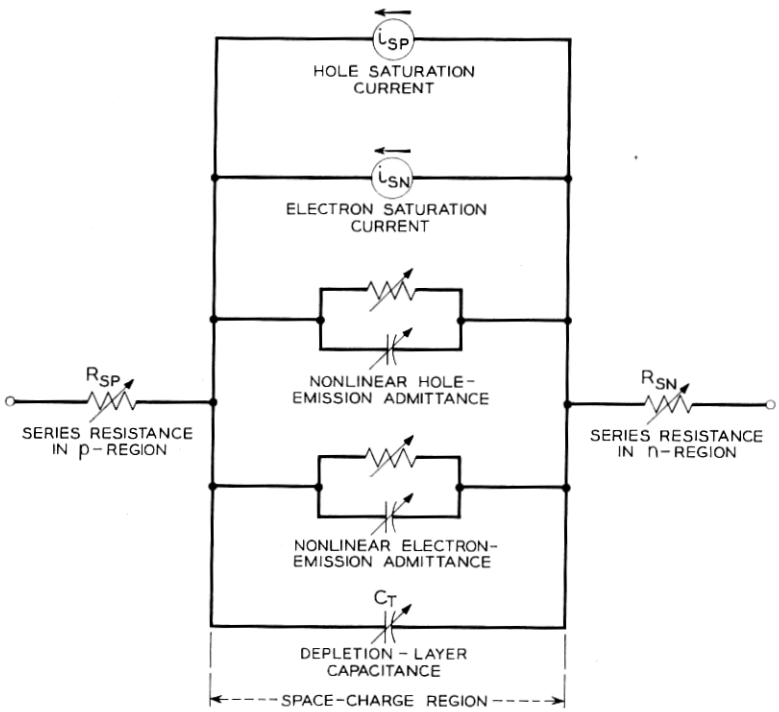

Fig. 2 — Equivalent circuit of p-n junction. In general, the nonlinear resistors and capacitors may be frequency-dependent.

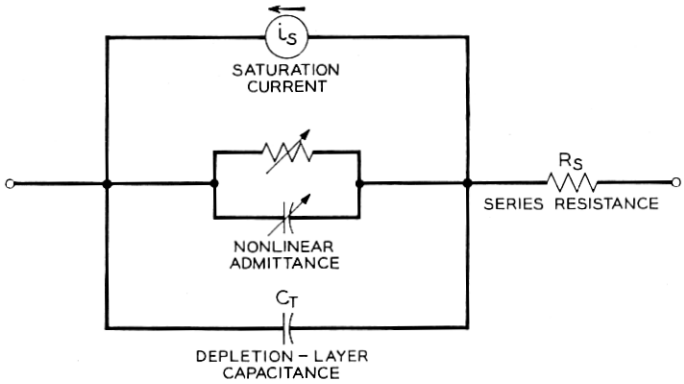

Fig. 3 — Simplified equivalent circuit in which constant parasitic elements are assumed and hole and electron admittances are combined.

nected with the fact that there is series resistance R_s between the junction and the metallic contacts, as indicated in Fig. 2. For many diodes, the series resistance is found to be reasonably constant as bias is varied, giving the equivalent circuit in Fig. 3. This observation is consistent with (but does not prove) the conjecture that the series resistance exhibits thermal noise. If this conjecture is true, the noise figure will depend upon the frequencies involved and upon the product $R_sC_{\min}$, where $C_{\min}$ is the value of the depletion layer capacitance at the maximum reverse voltage. But the immediate problem is to calculate shot noise due to minority carriers (i.e., due to carriers that cross the space-charge region), and R_s and the depletion-layer capacitance will both be neglected. That is, the analysis will be carried out for the equivalent circuit of Fig. 4. It should be noted that this neglect corresponds to the usual first approximation made in analyzing junction transistors.[10]

VI. STATISTICAL DESCRIPTION OF MINORITY-CARRIER MOTION

Each minority carrier — a hole in an n region or an electron in a p region — can be thought of as an individual. A hole may begin its existence as a minority carrier by surmounting the barrier potential of the depletion layer and entering the n region, or by being spontaneously (thermally) generated in the n region. Its existence may end by recombination in the n region or by return to the p region. The situation for electrons is analogous.

One need consider only the holes and electrons that cross the barrier. The life-histories of these carriers, which are the elementary events of the theory, fall into the three classes shown in Fig. 5. Their meandering paths indicate that the motion of a carrier is governed by statistical rules. These rules depend upon the detailed structure of the junction, but it is in keeping with classical p-n junction theory to assume that they do not depend upon the bias applied to the junction.

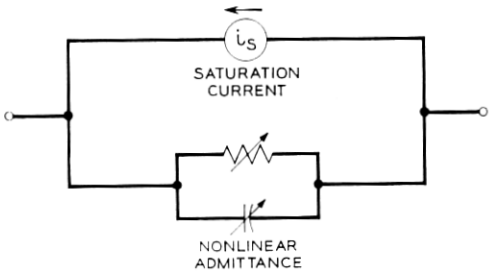

Fig. 4 — Equivalent circuit of p-n junction with negligible parasitics.

Of the three classes of elementary events, the reverse carriers will, throughout the analysis, be the simplest to deal with because they are assumed to occur entirely at random. Holes thermally generated in the neutral region of a p-n junction move about according to statistical rules, and some of them come in contact with the depletion layer. The electric field in the depletion layer is in such a direction as to move these holes across to the p region (forward bias decreases this field but ordinarily does not change its direction). Similarly, some of the electrons generated in the p region cross to the n region. These "saturation currents" are represented by current generators i_{sp} and i_{sn} in Fig. 2. They are direct-current generators and do not enter physically in the ac admittance of the junction, although their values can be introduced, by mathematical manipulation, into expressions for the ac admittance. But the saturation currents do produce shot noise. The sum of i_{sp} and i_{sn} is the saturation current, which may be determined experimentally by measuring the current at a moderate reverse voltage.

Of the many holes in the p region, a few will overcome the potential barrier of the space-charge region and be emitted into the n region. The rate of emission will be assumed to be a nonlinear function of the potential across the space-charge region. This function will be written $\mathfrak{R}(v)$, where v is the deviation (in the forward direction) from the equilibrium potential. "Classical" p-n junction theory, to be considered in Section VIII, involves assumptions equivalent to taking

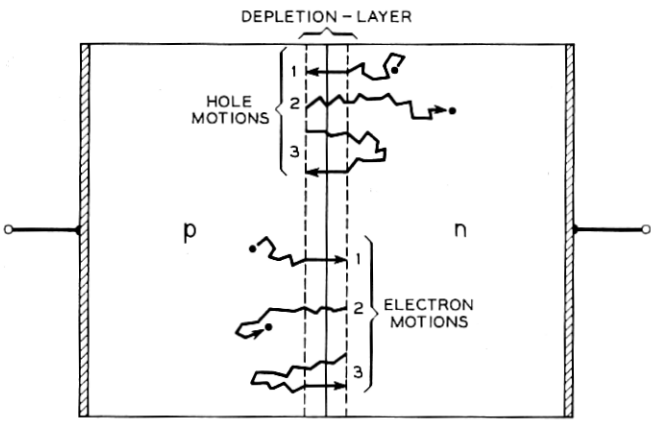

Fig. 5 — The three classes of hole and electron motions ("elementary events"): (1) reverse carrier from spontaneous generation; (2) forward carrier — emission followed by recombinations; (3) forward-and-back carrier — emission and return.

$$\mathfrak{R}(v) = \text{constant} \times e^{\beta v} \qquad \beta \equiv q/kT, \tag{1}$$

where q is the electronic charge, k is Boltzman's constant, and T is absolute temperature. The constant of proportionality is not specified. The conversion and noise analyses will be carried out with an arbitrary function $\mathfrak{R}(v)$, to accommodate refinements in p-n junction theory that may specify an emission relation differing from (1).

The probability of an emitted carrier *not* returning will be denoted P_0. Carriers with such a fate lead to a nonlinear conductance in the equivalent circuit. The probability of an emitted carrier returning at time t' after emission will be $P(t')$. Forward-and-back carriers result, in general, in a nonlinear conductance and a nonlinear capacitance whose values and frequency dependence can be calculated from $P(t')$. Since an emitted carrier must either return or not return,

$$1 = P_0 + \int_0^\infty P(t')\, dt'. \tag{2}$$

It is consistent with the usual approximations of p-n junction analysis to assume that P_0 is a constant and that $P(t')$ depends only upon t'. These assumptions will permit the noise analysis to be carried out rather simply. However, the effect of high current densities upon P_0 and $P(t')$ may be an important feature of point-contact diodes.

All that has been said about the emission of holes from the p side applies to the emission of electrons from the n side. Of course, $\mathfrak{R}(v)$, P_0 and $P(t')$ may be quite different for the electrons. In general, one must perform analyses for the holes and electrons and then add the currents. However, if (1) holds, $\mathfrak{R}$ for electrons is proportional to $\mathfrak{R}$ for holes, for all values of v. Then one can define suitable $\mathfrak{R}$, P_0 and $P(t')$ to apply to holes and electrons taken together. To simplify the presentation, it will be assumed that this has been done.

In assuming a certain probability for recombination of an emitted carrier, we are, in effect, assuming linear recombination. The assumption that the probability of return depends only on the time after emission is equivalent to the assumption of a linear diffusion equation including, possibly, drift in a steady electric field. These are, of course, usual assumptions which are made in analyzing p-n junction devices. It is not necessarily assumed that the recombination occurs uniformly throughout the volume. Surface recombination or recombination at localized centers is consistent with the assumptions as long as the recombination is linear. Also, we are not restricted to simple diffusion; drift in a steady electric field is permitted.

The currents for the three types of carrier events are illustrated in

Fig. 6. The reverse carrier gives a single negative pulse with an area equal to one electronic charge; the forward carrier gives a positive pulse with an area of one electronic charge. The detailed shape of the pulses might differ from one to the next and, furthermore, the average shape of the forward pulses might be somewhat different from the average shape of the reverse pulses. However, it will be assumed in most of the analysis that the entire pulse is of such short duration that the exact shape is of no consequence at the frequencies of interest. The third type of event — forward emission followed by the return of the same carrier — leads to a current whose integral is zero net charge. The forward pulse part of this single event should be similar to the forward pulse of a carrier that is emitted and does not return. Likewise, the reverse pulse part should be similar to the reverse pulses of the thermally generated carriers. However, the fact that the two pulses correspond to a single event is statistically important.

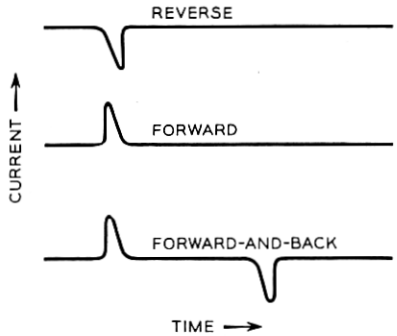

Fig. 6 — Current pulses for the three types of minority carrier events.

VII. SMALL-SIGNAL ADMITTANCE

The small-signal admittance of the p-n junction will be calculated from $\mathfrak{R}$, P_0 and $P(t')$. The calculation introduces some relationships needed in the main analysis. To avoid the introduction of a large number of symbols, a time function and its Fourier coefficients or spectrum will be denoted by the same symbol, for example, $i(t)$ and $i(\omega)$; $P(\omega)$ is an exception with a special definition to be given in (10).

In general, the current is equal to the difference between the current of forward carriers and the current of reverse carriers. For instantaneous forward and reverse transits, this means

$$i(t) = -i_s + q\mathfrak{R}[v(t)] - q\int_0^\infty dt'\mathfrak{R}[v(t - t')]P(t'). \tag{3}$$

The spectrum of i may be found from

$$i(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} i(t)e^{-j\omega t}\,dt. \tag{4}$$

This normalization of the spectrum will be adhered to in the calculation of noise. The small-signal admittance (also, the conversion matrix discussed in Section X) is a ratio of current to voltage, so any consistent normalization may be used. Consider, then a voltage of the form

$$v(t) = v_0 + \text{Re } v(\omega)e^{j\omega t}, \tag{5}$$

where $v(\omega)$ is the complex amplitude of a vanishingly small ac voltage of angular frequency ω. The expected solution is

$$i(t) = i_0 + \text{Re } i(\omega)e^{j\omega t}, \tag{6}$$

$$i(\omega) = Y(\omega)v(\omega). \tag{7}$$

The problem is to find $Y(\omega)$ and i_0.

Because $v(\omega)$ is vanishingly small,

$$\mathfrak{R}[v(t)] = \mathfrak{R}(v_0) + \mathfrak{R}'(v_0)\text{ Re } v(\omega)e^{j\omega t}. \tag{8}$$

Inserting this expression in (3), and performing the integration indicated in (4), one finds that

$$Y(\omega) = q\mathfrak{R}'(v_0)P(\omega), \tag{9}$$

where

$$P(\omega) = 1 - \int_0^{\infty} e^{-j\omega t'}P(t')\,dt'. \tag{10}$$

Also,

$$i_0 = -i_s + q\mathfrak{R}(v_0)P_0 \tag{11}$$

after (2) is used.

VIII. CLASSICAL p-n JUNCTION ANALYSIS

The analysis given thus far will be developed later into ways of relating conversion gain and noise to the small-signal admittances of a junction. Classical p-n junction analysis provides a way of predicting the small-signal admittances of postulated structures, although it does not give unique values for $\mathfrak{R}(v)$, P_0 and $P(t')$. Instead, it gives the products $\mathfrak{R}(v)P_0$ and $\mathfrak{R}(v)P(t')$, which suffice for the desired calculations.

The basic assumption of the classical theory is that the hole concentration p_1 on the n-side of the depletion layer is given by[9]

$$p_1 = p_N e^{\beta v} \tag{12}$$

where p_N is the equilibrium value of p_1. Equation (12) is a boundary condition for the diffusive flow of holes in the neutral n-region. The

generalized diffusion equation (including recombination and drift) is

$$\frac{\partial p}{\partial t} = -\frac{p - p_N}{\tau} + D\nabla^2 p - \mu\nabla\cdot(p\mathbf{E}). \tag{13}$$

The hole concentration p is a function of time and of position within the neutral n region; D, μ and τ are diffusion constant, mobility and lifetime for holes in the n region. The equilibrium hole concentration p_N will vary with position in the n region if the donor concentration is not uniform. For the formal analysis, τ and the electric field $\mathbf{E}$ will be assumed to be independent of p and t, so that (13) is a linear partial differential equation.

The hole current density can be calculated from

$$\mathbf{J} = -qD\nabla p + q\mu p\mathbf{E}. \tag{14}$$

If the voltage, v, is composed of a number of sinusoidal components' the hole concentration will be composed of sinusoidal components. Because (12) is nonlinear, the set of frequencies involved in p_1 will not, in general, be the same as the set of frequencies in v. Even so, the relation between p_1 and J is linear and may conveniently be expressed in the form

$$i(\omega) = \frac{AqD\ p_1(\omega)}{L_\omega}, \tag{15}$$

where A is the junction area and L_ω is the "effective ac diffusion length" for the structure in question. In other words, L_ω is simply a quantity that makes (6) and (15) give the right answer for $i(t)$. This right answer is the current obtained by solving (13), applying (14) and integrating over the junction area; this procedure will be used in Section XII in the calculation of L_ω for a particular nonlinear-capacitor structure.

The small-signal hole admittance can be found by considering in (12) a voltage of the form given in (5). Then

$$Y(\omega) = \frac{\beta AqDp_N e^{\beta v_0}}{L_\omega}. \tag{16}$$

The electron admittance is calculated similarly, usually with a quite different value of L_ω.

A comparison with the statistical description can now be made using (9) and (16). One finds

$$\mathcal{R}'(v_0)P(\omega) = \frac{\beta ADp_N e^{\beta v_0}}{L_\omega}, \tag{17}$$

or

$$\mathfrak{R}(v_0)P(\omega) = \frac{ADp_N e^{\beta v_0}}{L_\omega}, \tag{18}$$

since $\mathfrak{R}'$ equals $\beta\mathfrak{R}$, from (1), which is the form of $\mathfrak{R}$ that must be assumed to effect a comparison of the particle theory with the classical theory. Also, if (1) is assumed, (9) can be written

$$Y(\omega) = q\beta\mathfrak{R}(v_0)P(\omega). \tag{19}$$

In the analysis of conversion and noise, it will be found that all important quantities can be expressed in products of $\mathfrak{R}P(\omega)$ or $\mathfrak{R}'P(\omega)$. Classical p-n junction theory therefore gives a *complete* solution of the problem. Separately defined values of $\mathfrak{R}$ and $P(\omega)$ are not needed because a carrier that goes across and then comes back in an infinitesimal time might just as well not have been emitted in the first place.

The noise analysis will be carried through for an arbitrary emission function, $\mathfrak{R}(v)$, in case the classical theory is not a satisfactory one for some situations.

IX. SHOT NOISE FOR dc BIAS

The application of the above principles to the calculation of shot noise in dc-biased p-n junctions has been published.[11] For this simple case, many of the steps in the calculation combined so obviously that explicit mention of each step was not necessary. Here, a step-by-step derivation of the dc results will be given as a model for more complicated cases.

The calculation of shot noise will be based on the following general principle: Each elementary event delivers a certain amount of noise energy to the load. The noise power is the rate at which noise energy is delivered to the load.

At a given terminal pair of a linear circuit, the noise can be represented by a mean-square current generator $\langle i^2 \rangle$, since the noise power delivered to a load connected to the terminals can readily be calculated from such a generator, the load admittance and the output admittance of the terminals. Representation by a current generator seems particularly appropriate for shot noise, where the elementary events are represented by current generators.

In p-n junctions, the elementary events are the three classes of minority-carrier histories: reverse, forward and forward-and-back. Associated with each carrier event is a small-signal current generator $g(t)$, as illustrated in Fig. 6, with Fourier transform $g(\omega)$.

Suppose that $g(\omega)$ is observed for an infinitesimally small frequency band B. Then the transform back to the time domain of this selected

element of the spectrum is a current generator

$$\hat{g}(t) = \int_B d\omega g(\omega) e^{j\omega t} + \text{complex conjugate}, \tag{20}$$

where the indicated integration is over the values of ω that lie within the selected band.

If one writes

$$\hat{g}(t) = A + A^*, \tag{21}$$

then $\hat{g}^2(t)$, which is proportional to the instantaneous input power, is

$$\hat{g}^2(t) = 2AA^* + A^2 + (A^*)^2. \tag{22}$$

Integration over total time (to obtain a quantity proportional to the energy) can be shown to drop out the latter terms, so that

$$\int \hat{g}^2(t)\, dt = 2 \int AA^*\, dt. \tag{23}$$

Now, AA^* can be written as the double integral

$$AA^* = \int_B d\omega \int_B d\omega' g(\omega) g^*(\omega') e^{j(\omega - \omega')t}. \tag{24}$$

Integrating with respect to time gives

$$\begin{aligned} \int AA^*\, dt &= \int_B d\omega \int_B d\omega' g(\omega) g^*(\omega') [2\pi\delta(\omega - \omega')] \\ &= 2\pi \int_B d\omega \mid g(\omega) \mid^2, \end{aligned} \tag{25}$$

so that

$$\int \hat{g}^2(t)\, dt = 8\pi^2 B \mid g(\omega) \mid^2. \tag{26}$$

If the carrier transits are instantaneous, a forward or reverse transit at time t_0 gives a current generator

$$g(t) = \pm q\delta(t - t_0), \tag{27}$$

where the impulse function $\delta(t)$ has negligible width in time, but its integral over time is unity.

The spectrum of (27) is

$$g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \pm q\delta(t - t_0) e^{-j\omega t}\, dt = \pm \frac{q}{2\pi} e^{-j\omega t_0}. \tag{28}$$

For either a forward or a reverse transit, (26) becomes

$$\int \hat{g}^2(t)\,dt = 2q^2B. \tag{29}$$

The development thus far could be used only to treat the nonlinear-resistor type of p-n junction. For the general p-n junction, forward-and-back events must be considered. A forward pulse at t_0, followed by a reverse pulse at $t_0 + t'$, gives a current generator

$$g(t) = q\delta(t - t_0) - q\delta(t - t_0 - t'), \tag{30}$$

$$g(\omega) = \frac{q}{2\pi}\,[e^{-j\omega t_0} - e^{-j\omega(t_0+t')}]. \tag{31}$$

For this case, (26) gives

$$\int \hat{g}^2(t)\,dt = 2q^2B(2 - 2\cos\omega t'). \tag{32}$$

The shot noise for each class of carrier event is found by multiplying the appropriate integral in (32) by the rate of occurrence of such carrier events. Thus,

$$\langle i^2\rangle_{\text{reverse}} = 2q^2B\,\frac{i_s}{q} = 2qBi_s\,, \tag{33}$$

$$\langle i^2\rangle_{\text{forward}} = 2q^2B\mathcal{G}(v_0)P_0\,, \tag{34}$$

$$\begin{aligned}\langle i^2\rangle_{\text{forward-and-back}} &= 2q^2B\mathcal{G}(v_0)\int_0^\infty (2 - 2\cos\omega t')P(t')\,dt' \\ &= 2q^2B\mathcal{G}(v_0)[P(\omega) + P(\omega)^* - 2P_0].\end{aligned} \tag{35}$$

The total shot noise is the sum of (33), (34) and (36):

$$\langle i^2\rangle_{\text{total}} = 2qB\{i_s + q\mathcal{G}(v_0)[P(\omega) + P(\omega)^* - P_0]\}. \tag{36}$$

For the important special case of classical p-n junction theory, (36) can be transformed through use of (11) and (19) into

$$\langle i^2\rangle_{\text{total}} = 4kTBG(\omega) - 2qi_0B, \tag{37}$$

where $G(\omega)$ is the real part of $Y(\omega)$. This result is easy to remember because it has the form "thermal noise minus shot noise". However, it is evident from the derivation that no subtraction of one noise from another is actually involved.

A previous derivation of (37) was given under less general assumptions,[12] and the equation was compared with experimental results for

p-n junctions.[13,14] The good agreement with experiment suggests that it is worthwhile to analyze the noise of a p-n junction frequency converter.

X. SIGNAL TRANSMISSION IN LINEAR FREQUENCY CONVERTERS

In a linear frequency converter, the output signal voltage and current are linearly related to the input signal voltage and current.

Of course, frequency conversion by a nonlinear device requires a finite local-oscillator voltage. The local-oscillator interaction with the device is quite nonlinear. Since the local-oscillator drive is regarded as a power supply, not an intelligence-bearing signal, a linear frequency converter can be made with a nonlinear device. Linear behavior is obtained when the signal voltages are assumed to be arbitrarily small.

The analysis of signal transmission in the linear frequency converter will be made on the assumption that the local-oscillator waveform is periodic, with a fundamental frequency $b/2\pi$. This assumption is not trivial, for nonlinear capacitors can be tuned to oscillate at frequencies incommensurate with the driving-oscillator frequency, as well as at subharmonics of this frequency. The formal results could be generalized to cover the nonperiodic case, but most practical situations involve periodic waveforms.

The frequency components of the emission rate $\mathcal{R}(v)$ must be related to the frequency components of the voltage, $v(t)$, applied to the junction. The voltage will be given in the form

$$v(t) = v_{\text{LO}}(t) + v_{\text{SIG}}(t), \tag{38}$$

where $v_{\text{LO}}(t)$ is the large, periodic local-oscillator voltage (plus any dc voltage) and $v_{\text{SIG}}(t)$, which will be regarded as arbitrarily small, is the sum of all the signal voltages. Here the word "signal" includes the input and output signals and the image-frequency signals, all of which are linearly related. The method of calculating these linear relationships will next be given. The local-oscillator voltage, whatever its waveform, may be written

$$v_{\text{LO}}(t) = \sum_m v_{\text{LO}}(m) e^{jmbt}, \tag{39}$$

where m takes on all positive and negative integer values including zero.

Most frequency-conversion problems can be treated by proper identification of s in the following representation for v_{SIG} :

$$v_{\text{SIG}}(t) = \text{Re} \sum_m v(mb + s) e^{j(mb+s)t}. \tag{40}$$

For example, $s/2\pi$ could be the intermediate frequency (output) of a superheterodyne receiver (Figs. 7 and 8), or the input frequency of a transmitting modulator (Fig. 9). All of these circuits and many others can be obtained by suitable connections to the multiterminal "black box" indicated in Fig. 10. Negative frequencies are included in the set $(mb + s)/2\pi$; they are the "lower sidebands". For the sake of obtaining noise-figure expressions, it will be assumed that the input and output are at two distinct frequencies $(m_1b + s)/2\pi$ and $(m_2b + s)/2\pi$, as indicated in Fig. 11. In case the input and output are at the same frequency, as in a negative-resistance amplifier, the noise-figure expressions will not be applicable, but (65) will give the shot-noise current at the frequency of interest.

Any information-bearing signal will contain a number of frequencies (in general, a continuous spectrum). The linear nature of the small-signal problem enables one to analyze for each frequency separately.

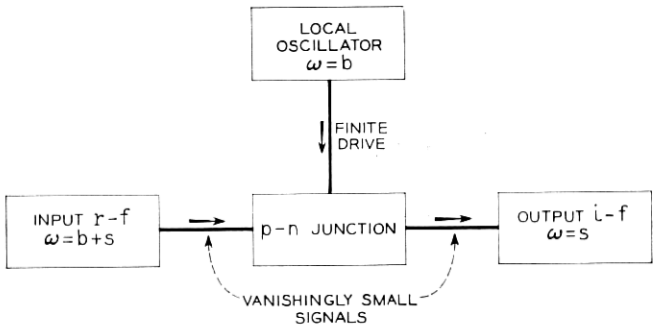

Fig. 7 — Upper-sideband superheterodyne receiver (down-converter).

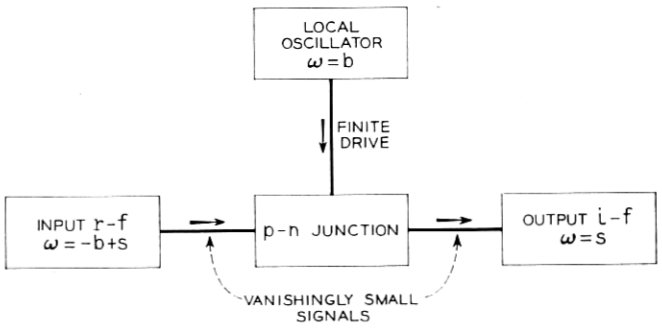

Fig. 8 — Lower-sideband superheterodyne receiver (down-converter).

Since v_{SIG} is arbitrarily small, one can write

$$\mathcal{R}[v(t)] = \mathcal{R}[v_{LO}(t)] + \mathcal{R}'[v_{LO}(t)]v_{SIG}. \tag{41}$$

The dc and large-signal dc components of $\mathcal{R}$ are given by the Fourier expansion of $\mathcal{R}[v_{LO}(t)]$. The Fourier expansion of $\mathcal{R}'[v_{LO}(t)]$ is required for the signal transmission, that is, for calculating the conversion matrix. Both expansions are necessary to calculate the noise figure.

Assume that the required Fourier coefficients have been calculated,

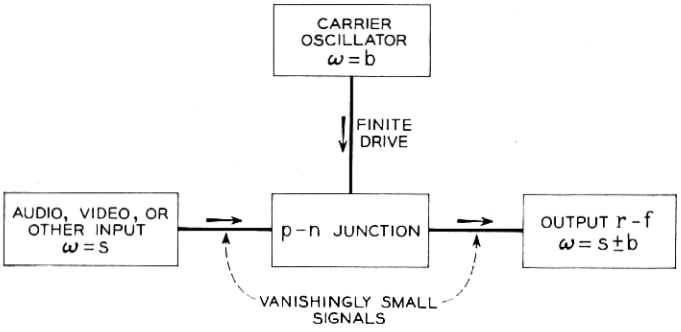

Fig. 9 — Double-sideband transmitting modulator (up-converter).

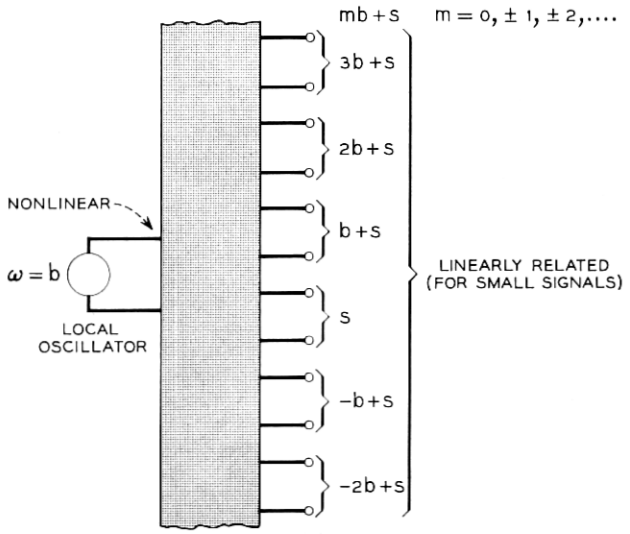

Fig. 10 — "Black box" representation of a linear frequency converter.

so that one can write

$$\mathcal{R}[v_{LO}(t)] = \sum_m \mathcal{R}(mb)e^{jmbt}, \tag{42}$$

and

$$\mathcal{R}'[v_{LO}(t)] = \sum_m \mathcal{R}'(mb)e^{jmbt}. \tag{43}$$

Then, one can say that $\mathcal{R}$ is made up of a large periodic part with Fourier components $\mathcal{R}(mb)$ and a vanishingly-small signal part with Fourier components

$$\mathcal{R}(mb + s) = \sum_n R'[(m - n)b]v(nb + s). \tag{44}$$

The complex amplitudes of the alternating currents, both large and small, can be calculated by multiplying $qP(\omega)$. First, the large-signal current will be called i_{LO}, with Fourier components

$$i_{LO}(mb) = qP(mb)\mathcal{R}(mb). \tag{45}$$

It should be noted that $i_{LO}(0)$ differs from the net direct current, i_0, by

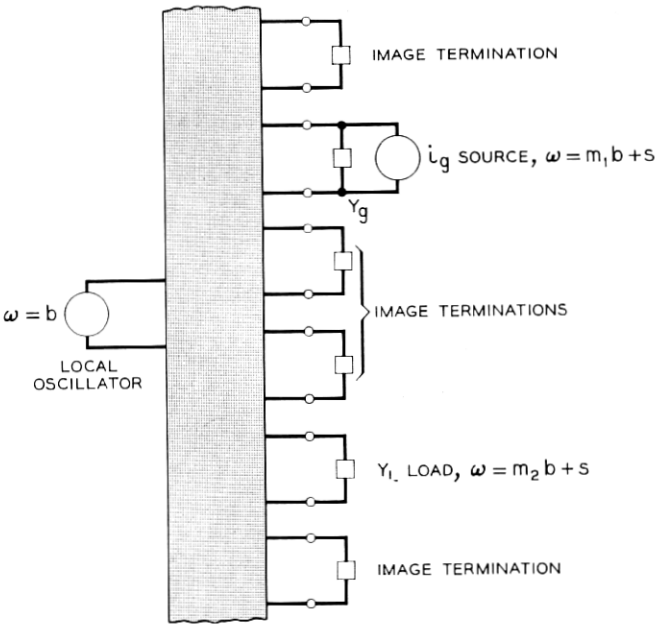

Fig. 11 — Linear frequency converter used as an active four-pole network.

the saturation current, i_s :

$$i_{LO}(0) = i_0 + i_s . \tag{46}$$

Second, a complete specification of the small-signal transmission of the frequency converter is given by

$$i_{SIG}(t) = \text{Re} \sum_m i(mb + s)e^{j(mb+s)t}, \tag{47}$$

with

$$i(mb + s) = qP(mb + s) \sum_n \mathcal{R}'[(m - n)b]v(nb + s). \tag{48}$$

Equation (48) can be written in the form

$$i(mb + s) = \sum_n Y(m, n; s)v(nb + s), \tag{49}$$

where

$$Y(m, n; s) = qP(mb + s)\mathcal{R}'[(m - n)b]. \tag{50}$$

The array $Y(m, n; s)$ is called the conversion *matrix*. For brevity,

$$Y(m, n; s) \equiv Y_{mn} . \tag{51}$$

In operation, the linear frequency converter will be terminated with admittances Y_m at each of the "terminals" $mb + s$. A system of current generators $\mathcal{I}_m$ at each of these terminals will produce currents and voltages that obey the relationship

$$\mathcal{I}_m = i_m + v_m Y_m . \tag{52}$$

Therefore,

$$\mathcal{I}_m = \sum_n (Y_{mn} + Y_m\delta_{mn})v_n , \tag{53}$$

where δ_{mn} is 1 if m equals n and δ_{mn} is 0 if m does not equal n. One can also write

$$v_m = \sum_n \lambda_{mn}\mathcal{I}_n , \tag{54}$$

where

$$\| \lambda_{mn} \| = \| Y_{mn} + Y_m\delta_{mn} \|^{-1} \tag{55}$$

in matrix notation.

The effect at terminal m of a current generator $\mathcal{I}_n$ at terminal n can be duplicated by a fictitious current generator $\eta_{mn}\mathcal{I}_n$ placed at terminal

m, where

$$\eta_{mn} = \frac{\lambda_{mn}}{\lambda_{mm}} \tag{56}$$

does not depend on Y_m (Thévenin-Norton theorem). This property makes η_{mn} a desirable quantity to use in the analysis so that a necessary characteristic of the noise figure — that it be independent of the load admittance — will appear quickly and naturally.

XI. NOISE IN THE GENERAL p-n JUNCTION FREQUENCY CONVERTER

Suppose that the appropriate value of $g(\omega)$ for a carrier event is applied to each terminal of the linear frequency converter. Then the equivalent current generator at the output $m_2b + s$ will be

$$\sum_n \eta_{m_2n} g(nb + s). \tag{57}$$

This sum can be substituted for $g(\omega)$ in (26). For either a forward or a reverse transit, (26) and (57) give

$$\int \hat{g}^2(t)\, dt = 2q^2B \sum_{n,n'} \eta_{m_2n}\eta^*_{m_2n'} e^{j(n'-n)bt_0}. \tag{58}$$

The development thus far could be used to treat the nonlinear-resistor type of p-n junction. For the general p-n junction, forward and back events must also be considered. A forward pulse at t_0 followed by a reverse pulse at $t_0 + t'$ gives

$$\int \hat{g}^2(t, t')\, dt = 2q^2B \sum_{n,n'} \eta_{m_2n}\eta^*_{m_2n'} e^{j(n'-n)t_0}[1 - e^{-j(nb+s)t'} - e^{j(n'b+s)t'} + e^{j(n'-n)bt'}]. \tag{59}$$

In all applications of (59), the weighted sum over all possible values of t' will be required:

$$\int_0^\infty dt' P(t') \int \hat{g}^2(t', t)\, dt = 2q^2B \sum_{n,n'} \eta_{m_2n}\eta^*_{m_2n'} e^{j(n'-n)t_0}[P(nb + s) + P^*(n'b + s) - P(nb - n'b) - P_0]. \tag{60}$$

11.1 *Reverse Carriers*

The thermally generated reverse carriers occur at random but at a steady average rate i_s/q. Multiplying (58) by this rate gives

$$\langle i^2\rangle_{\text{reverse}} = 2qBi_s \sum_n |\ \eta_{m_2 n}\ |^2, \tag{61}$$

since the terms in (58) for which n does not equal n' are oscillatory and give no contribution to the noise power when averaged for a steady rate.

Equation (61) may be interpreted as simple shot noise applied to each terminal and independently converted to the output (except the term for $n = m_2$, which represents the direct shot noise at the output).

11.2 *Forward Carriers*

The rate of emission $[P_0\Re(t_0)]$ of forward carriers varies periodically, so that an average of (58) over the local-oscillator cycle is required:

$$\langle i^2\rangle_{\text{forward}} = \frac{b}{2\pi}\int_{-\pi/b}^{\pi/b} dt_0 P_0\Re(t_0)\left[2q^2B\sum_{n,n'}\eta_{m_2 n}\eta^*_{m_2 n'}e^{j(n'-n)bt_0}\right]. \tag{62}$$

The Fourier expansion, (42), of $\Re(t_0)$ may be inserted in (62). One then finds

$$\langle i^2\rangle_{\text{forward}} = 2q^2BP_0\sum_{n,n'}\Re(nb - n'b)\eta_{m_2 n}\eta^*_{m_2 n'}. \tag{63}$$

11.3 *Forward-and-Back Carriers*

In similar fashion, one can average (60) over the local-oscillator cycle to obtain the mean-square output current generator for all classes of forward-and-return carriers:

$$\begin{aligned}\langle i^2\rangle_{\text{forward-and-back}} = 2q^2B\sum_{n,n'}\Re(nb - n'b)\eta_{m_2 n}\eta^*_{m_2 n'}[P(nb + s)\\ + P^*(n'b + s) - P(nb - n'b) - P_0].\end{aligned} \tag{64}$$

11.4 *Total Shot Noise*

The sum of (61), (63) and (64) is

$$\begin{aligned}\langle i^2\rangle_{\text{total}} = 2qB\Big\{i_s\sum_n |\ \eta_{m_2 n}\ |^2\\ + q\sum_{n,n'}\Re(nb - n'b)\eta_{m_2 n}\eta^*_{m_2 n'}[P(nb + s) + P^*(n'b + s)\\ - P(nb - n'b)]\Big\}.\end{aligned} \tag{65}$$

11.5 *Image Conversion Noise*

The frequency converter is supposed to be terminated at the frequencies $(mb + s)/2\pi$ with specified admittances which may or may not generate noise. What noise is generated will be transmitted to the output according to the coefficients $\eta_{m_2 m}$.

The input and output terminations $m_1 b + s$ and $m_2 b + s$ are in a class apart, because noise generated in these terminations is chargeable to the transducers that precede or follow the frequency converter. The other terminations $mb + s$, with $m_1 \neq m \neq m_2$, will be called images. The frequency converter (but not the diode) must be held responsible for output noise originating in the image terminations.

The properties of the image terminations as noise generators depend upon the physical nature of the terminations and are not uniquely determined by their impedances. For example, the conductance G_m for each of the image terminations $mb + s$ might consist of a resistor at temperature T_m. Then the image termination noise at the output would be

$$\langle i^2 \rangle_{\text{image terminations}} = 4kB \sum_m{}' T_m G_m \mid \eta_{m_2 m} \mid^2, \tag{66}$$

where the symbol $\sum'$ indicates that terms in $m = m_1$ or $m = m_2$ are not included in the summation. The terminations may be devices that are not in thermal equilibrium. If several diodes are used in combination, so that they terminate each other at some image frequencies, the noise generated in the mutual terminations can be much greater or much less than thermal noise.

11.6 *Noise Figure*

The noise output of a linear transducer must be judged in relation to the signal output for a given input signal. As a standard value for the input signal, it is conventional to use thermal noise corresponding to an absolute temperature of 290°K, which will be denoted T_s. The thermal noise current is $\sqrt{4kT_s G_s B}$, where G_s is the source conductance, and an rms current of this value will be the standard input signal.

The noise figure F can be defined by

$$F - 1 = \frac{\text{output noise power due to transducer}}{\text{output signal power from standard input signal}}. \tag{67}$$

The noise figure of the input stage determines receiver sensitivity if the gain of the transducer is sufficiently high. For the frequency converter,

$$F - 1 = \frac{\text{shot noise at output} + \text{image conversion noise at output}}{\text{output signal power from the standard input signal}}. \tag{68}$$

In place of power, we may use in each case the equivalent mean-square current generator at the output, which for the standard signal is

$$\langle i^2 \rangle_{\text{standard}} = 4kT_sBG_s \mid \eta_{m_2m_1} \mid^2. \tag{69}$$

For convenience, one may define

$$\mu_n = \frac{\eta_{m_2n}}{\eta_{m_2m_1}} = \frac{\lambda_{m_2n}}{\lambda_{m_2m_1}}. \tag{70}$$

Then the total noise figure may be written

$$F - 1 = \frac{q}{2kT_sG_s}\left\{q\sum_{n,n'} \Re(nb - n'b)\mu_n\mu_{n'}{}^*[P(nb + s) + P^*(n'b + s) - P(nb - n'b)] + i_s\sum_n \mid \mu_n \mid^2\right\} + \sum_m{}' \frac{G_mT_m}{G_sT_s} \mid \mu_m \mid^2. \tag{71}$$

The last term of this equation represents image conversion noise. It is an important practical problem to design the circuit to minimize image conversion noise. But, in principle, this can always be made to vanish by refrigerating the terminations, so the subsequent theoretical discussion will consider only the shot noise.

XII. NONLINEAR CAPACITORS

A nonlinear-capacitor type of p-n junction is one in which all emitted carriers are collected within a short time after emission. If all emitted carriers are collected, P_0 is zero, since it is the probability of a carrier not being collected. Hence, there are no forward-only carriers. Moreover, the reverse current i_s of the thermally generated carriers must be zero. One may show this by considering the diode in equilibrium, when the net direct current must be zero; if there are no forward-only carriers, there must be no reverse-only carriers.

Thus, the only shot noise is that due to forward-and-back carriers, governed by $P(t')$. The exact form of $P(t')$ is not too important. The simple rectangular function shown in Fig. 12 will serve to illustrate the main result. The maximum storage time, Δ, will be allowed to become very small compared to $1/\omega$, to approach the case of an ideal nonlinear capacitor. The function $P(\omega)$ is, according to (10),

$$P(\omega) = 1 - \int_0^\Delta \frac{1}{\Delta} e^{-j\omega t'}\, dt' = \tfrac{1}{2} j\omega\Delta + \tfrac{1}{6}(\omega\Delta)^2 + \cdots. \tag{72}$$

The shot noise due to forward-and-back carriers is obtained when this $P(\omega)$ is inserted in (71). Each term contains a factor

$$P(nb + s) + P^*(n'b + s)\, P(nb - n'b) =$$
$$\tfrac{1}{2} j(nb + s)\Delta - \tfrac{1}{2} j(n'b + s)\Delta - \tfrac{1}{2} j(nb - n'b)\Delta$$
$$+ \frac{\Delta^2}{6}[(nb + s)^2 + (n'b + s)^2 - (nb - n'b)^2] + \cdots \quad (73)$$
$$= \frac{\Delta^2}{3}(nb + s)(n'b + s) + \cdots .$$

The first-order terms in Δ cancel. As Δ goes to zero, i.e., as the frequency response of the diode is increased, the shot noise goes to zero as Δ^2.

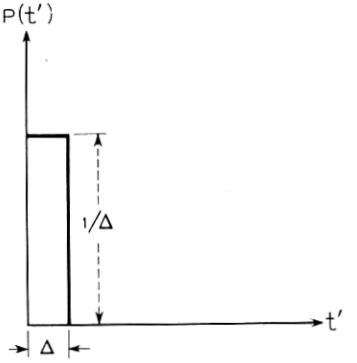

Fig. 12 — Example of a return probability distribution leading to nonlinear capacitor action.

The capacitance, for a given voltage, is proportional to Δ. To approach an ideal nonlinear capacitor, let Δ approach zero. At the same time, the capacitance waveform can be kept essentially unchanged if the emission rate is increased (by appropriate adjustment of the local-oscillator voltage). The shot noise then goes to zero with Δ and we get this result: *A frequency converter using an ideal nonlinear capacitor has no shot noise output regardless of the local-oscillator waveform, the source and load admittances, and the image terminations.*

An interesting qualitative interpretation* of this result is that, as Δ gets smaller, an increasing number of carrier events are required to obtain a given capacitance waveform. A smoother current results from the statistical effect of having a large number of elementary events.

* Suggested by D. Leenov.

If the nonlinear capacitance does not arise from the minority-carrier storage mechanism that has been postulated, but is wholly or partly due to depletion-layer capacitance, there is even less reason to expect noise. The depletion-layer capacitance involves widening and narrowing of the space-charge region. It requires a relatively small motion of majority carriers instead of their emission all the way across the junction, or of any subsequent existence as minority carriers. The frequency response of C_T should therefore be better than that of minority-carrier storage capacitance. Also, the elementary events are likely to be representable approximately as forward-and-back current generators, with the area under each pulse being much less than q, since the carriers do not go all the way across the junction.

The problem of finite frequency response ($\Delta \neq 0$) can be studied by putting the expression given in (73) in (71), to get the leading term in the shot noise:

$$F - 1 = \frac{q^2\Delta}{6kT_sG_s} \sum_{n,n'} \Delta\mathcal{R}(nb - n'b)(nb + s)(n'b + s). \qquad (74)$$

It is evident in (74) that F − 1 goes to zero with Δ if $\Delta\mathcal{R}(nb - n'b)$ is held constant. But, for a finite value of Δ, the shot noise could be very large in the case of significant conversion from high-order harmonic images. Hence, a circuit design consideration is to make μ_n small for all values of n such that nb is greater than $1/\Delta$. Short-circuiting of the harmonic images is one way of accomplishing this objective.

Apart from finite frequency response, actual diodes will fall short of ideal nonlinear-capacitor action because of the finite lifetime of minority carriers. Some minority carriers will recombine before they have a chance to return to the junction. The effect of this recombination is equivalent to that of a nonlinear resistor in parallel with a nonlinear capacitor having infinite minority-carrier lifetime. As will be shown, low-noise frequency conversion is possible with nonlinear resistors when the local-oscillator current is sharply pulsed. Therefore, low-noise frequency conversion is possible with a nonlinear capacitor suffering from recombination, but the local-oscillator waveform can no longer be arbitrary, and the conversion gain will be reduced.

12.1 *Nonlinear Capacitor Structures*

Four semiconductor structures shown in Fig. 13 exhibit nonlinear capacitance, for reasons to be discussed. The first of these can be made from a transistor, and thus is a readily available specimen for relatively low-frequency (about 10 mc) experiments with nonlinear capacitors.

However, the great promise of the nonlinear capacitor is the possibility that a diode structure can be made to work at higher frequencies than transistors. Transistors, as will be explained below, are not one-dimensional structures, a fact that aggravates the series-resistance problem. The p^+-n-n^+ and p-i-n and the graded p-n junction can be truly one-dimensional structures, as long as skin-effect is not important. Of

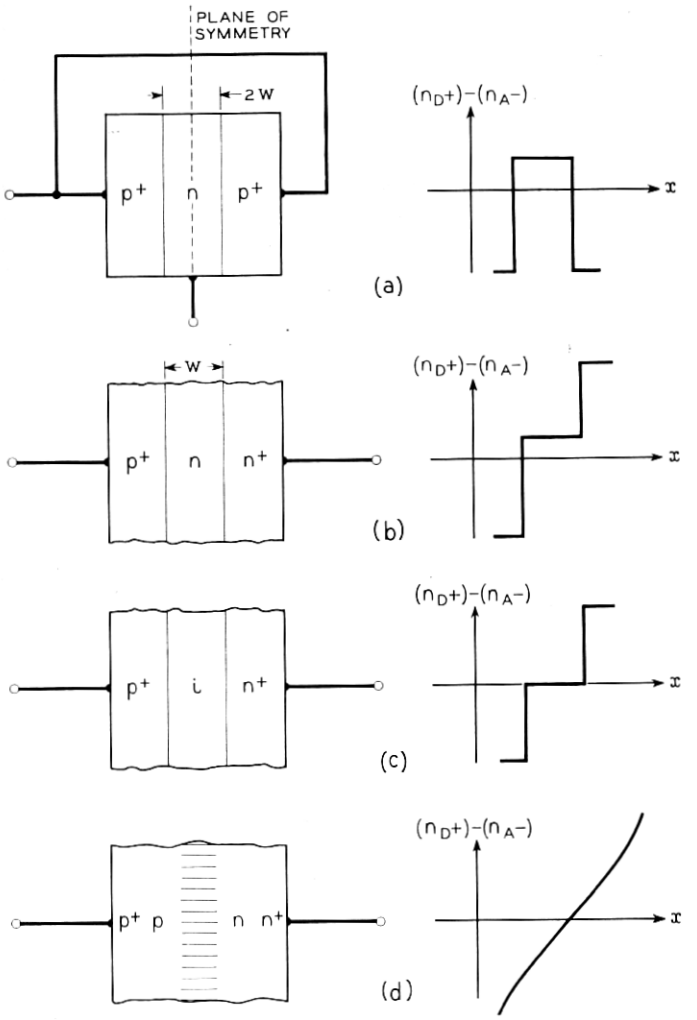

Fig. 13 — Nonlinear capacitor structures: (a) p^+-n-p^+; (b) p^+-n-n^+; (c) p^+-i-n^+; (d) graded p-n junction.

these, the graded p-n junction seems to be the easiest structure to fabricate by solid-state diffusion of impurities into semiconductors.

12.2 p^+-n-p^+ Structure

The p^+-n-p^+ structure in Fig. 13(a) can be made by connecting the emitter and collector leads of an alloy-junction transistor. A symmetrical alloy transistor (one with emitter and collector junctions of equal size) is theoretically preferred for maximum nonlinear-capacitor action. The pains taken by the transistor manufacturer to prevent emitter-collector short circuits are, of course, completely wasted in this application.

The emitter and collector junctions thus become, in effect, one junction which almost completely surrounds the n region. Because of the high concentration of holes in the heavily doped p^+ region, compared to the concentration of electrons in the n region, nearly all of the emitted carriers are holes entering the n region, rather than electrons going into the p^+ regions. Holes going into the n region may recombine and become forward-only events. But if the base width $2w$ is narrow enough, most of the holes will escape recombination, either by returning to the junction from which they were emitted or by being collected at the opposite junction. In either case, such holes become forward-and-return events with respect to the junction, consisting of emitter and collector junctions taken together.

The n region is assumed to be of uniform resistivity and thus to be field-free except for the space-charge regions at the junction. Hence, the motion of the emitted holes is simple diffusion. The diffusion equation (13) then reduces to

$$(j\omega + 1/\tau)p_\omega = D\nabla^2 p_\omega . \tag{75}$$

If one lets x equal zero at the plane of symmetry indicated in Fig. 13(a), the solution of (75) appropriate to this symmetry is

$$p_\omega \propto \cosh x\sqrt{(j\omega + 1/\tau)/D}. \tag{76}$$

At $x = \pm w$, p_ω equals $p_{1\omega}$, so

$$p_\omega = p_{1\omega}\frac{\cosh x\sqrt{(j\omega + 1/\tau)/D}}{\cosh w\sqrt{(j\omega + 1/\tau)/D}}. \tag{77}$$

The current density is given by $qD\nabla p$ evaluated at $x = w$. Then, by comparison with (15), one finds that

$$1/L_\omega = \sqrt{(j\omega + 1/\tau)/D}\tanh w\sqrt{(j\omega + 1/\tau)/D}. \tag{78}$$

According to (18), $1/L_\omega$ is proportional to $P(\omega)$. For comparison with

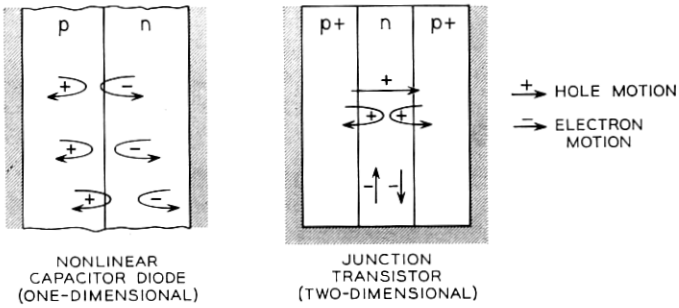

Fig. 14 — One-dimensional versus two-dimensional bipolar (hole and electron) flow patterns in junction devices.

(58), recombination may be neglected by setting $1/\tau \approx 0$, and (64) can then be expanded on the assumption that w is small (i.e., that the frequency response is high relative to ω). One finds

$$P(\omega) \propto \frac{1}{L_\omega} = \frac{1}{w}\left[j\frac{\omega w^2}{D} + \frac{1}{3}\left(\frac{\omega w^2}{D}\right)^2 + \cdots\right] \tag{79}$$

Hence, the leading terms are proportional to those of (72), with w^2/D playing the role of the time Δ in which (nearly) all of the emitted carriers return.

The postulated condition of neutrality of the n region implies that, when holes are emited into the n region, an essentially equal number of electrons must enter the n region from the base contact. A voltage drop is required to move these carriers, and this may be crudely described as a series resistance. The thinner the base layer is made for the sake of high frequency response, the higher the series resistance. This dilemma is well known in transistor design. As shown in Fig. 14, the electrons move in a direction perpendicular to the average motion of the holes; hence, two directions in space are involved. Thus, any one-dimensional analysis of the transistor is inherently an oversimplification, although sometimes a very useful approximation.

12.3 p^+-n-n^+ *Structure*

As an introduction to a structure that is truly one-dimensional, note that the plane of symmetry in Fig. 13(a) could be replaced by a surface that has the property of reflecting holes. A heavily doped n^+ region would have this property, and this suggests the p^+-n-n^+ structure shown in Fig. 13(b). Equation (78) should apply to this structure as well as to that of Fig. 13(a).

For each element of p^+-n junction area, there is a corresponding area of n-n^+ junction that reflects emitted holes and also provides the electrons to neutralize the holes. Thus, the area of this structure can be increased in two dimensions without affecting its performance, except for a proportionate lowering of impedance level. This is a characteristic of a one-dimensional structure.

The depletion-layer etching technique has been used to make very thin germanium p^+-n-x structures, where x is supposed to be a region of high recombination.[15] The unsolved problem in making this structure is that of obtaining a satisfactory x region for nonlinear resistance action at high frequencies. No such problem exists in making a p^+-n-n^+ nonlinear capacitor, since a well-understood n^+ region is required in place of the little-understood x region.

12.4 *p^+-i-n^+ Structure*

The p^+-i-n^+ diode is another one-dimensional structure that has a carrier-storage behavior described qualitatively by (78) because the basic carrier-storage action is quite similar to that of the p^+-n-n^+ structure. Thus, holes emitted from the p^+ region are reflected at the n^+ junction. In addition, electrons emitted from the n^+ junction are reflected at the p^+ junction.

An accurate analysis of the p^+-i-n^+ structure requires consideration of the simultaneous motion of holes and electrons, and would represent an advance over present p-n junction theory. In the meantime, the qualitative considerations mentioned above indicate that the p^+-i-n^+ structure should give nonlinear-capacitance action up to frequencies determined by the width of the i region. At much higher frequencies, the structure will just be a resistance whose value depends upon the density of carriers in the i region, which density is proportional to the forward direct current.

12.5 *Graded p-n Junction*

By drawing a smooth curve through the concentration gradients for either the p^+-n-n^+ or the p^+-i-n^+ structures, one obtains the graded p-n junction shown in Fig. 13(d).

The graded junction has features that recommend it as a nonlinear-capacitor structure.[6] The electric field associated with the impurity gradient tends to reflect emitted carriers — the steeper the gradient, the larger the field and the higher the frequency response of the capacitance. Analysis of the frequency response, based on the approximate kinetic

theory of p-n junctions, can apply (if at all) only to an intermediate region of bias where emission is significant but not so copious that the electric field is altered by the forward current.

The frequency response of the capacitance will depend on bias, being very high for reverse voltages and deteriorating gradually as the voltage approaches and enters the forward region. However, microwave experiments suggest that the series resistance will generally be a much more important source of loss and noise than any frequency limitations of the nonlinear capacitance. A steep impurity gradient makes for a series resistance that is low both in absolute value and in relation to the capacitance. However, the capacitance per unit area increases as the gradient increases.

XIII. NONLINEAR RESISTOR NOISE

The nonlinear resistor type of p-n junction has the property that none of the emitted carriers return, except possibly some carriers that return in a time negligibly short in comparison to the frequencies under consideration. Within the frequency range of nonlinear resistor action, the admittance due to emitted carriers is real and independent of frequency.

In spite of the absence of return pulses to cancel the shot current of emission, the shot noise delivered to the load can be made very small if the local-oscillator current flows in short pulses. The plausibility of this result will be argued by representing the p-n junction as a noiseless admittance in parallel with a generator of shot current pulses. The shot current and the time variation of the admittance are governed by the local-oscillator drive. For the nonlinear resistor, the admittance is a pure conductance and the shot current to be considered is that of emitted carriers that do not return.

In a frequency converter, the nonlinear resistor is connected to a linear passive network which includes the terminations at all the signal frequencies. If practically all of the shot current flows at a time when the junction conductance is very large compared to the network admittance, the shot current will be short-circuited through the junction conductance and little noise will be delivered to the network.

Another way of describing the situation is to consider that shot noise results because one cannot predict with certainty that a carrier will be emitted in any given short time interval. When a local-oscillator voltage is applied to a junction, it is known that the sum of the carrier currents, in spite of their individual unpredictability, is very nearly equal to a definite macroscopic current waveform that can be determined from the junction characteristics. When this macroscopic current consists of short

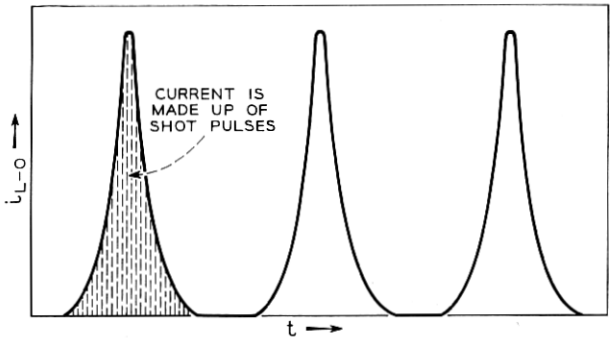

Fig. 15 — Pulsed local-oscillator current waveform giving partial reduction of shot noise.

sharp pulses, it is possible to predict that many carriers will be emitted during the pulses and practically none during other times. Unpredictability, hence shot noise, is greatly reduced for a pulsed current waveform like the one shown in Fig. 15.

The principle of using pulsed local-oscillator current can be applied in the design of nonlinear-resistance frequency converters. In a simple design, for example, it would be preferable to use shunt conductance rather than series resistance to attenuate the power from the local oscillator. More sophisticated mixers may use filters and additional diodes to shape the local-oscillator waveform.

The image terminations chosen below to illustrate the waveform effects should not be regarded as the optimum conditions for practical use. For example, the main image frequency might well be open-circuited to minimize conversion of thermal noise from the series resistance, while short-circuiting of this image will be assumed for mathematical simplicity (in the case of short-circuited harmonics).

13.1 *Short-Circuited Harmonics*

As an example of a circuit permitting low noise and good signal transmission, the mathematically simple case of short-circuited harmonics will be considered. Another way of defining this case is to say that the local-oscillator voltage waveform is purely sinusoidal (plus a dc component, v_0). Thus,

$$v(t) = v_0 + v_b \cos bt. \tag{80}$$

The exponential emission function, (1), will be assumed, so that

$$qP_0\mathcal{R}(v) = i_s e^{\beta v} \tag{81}$$

$$= i_s e^{\beta v_0} e^{\beta v_b \cos t}. \tag{82}$$

The constant i_s is necessarily equal in magnitude to the thermally generated reverse current, so that no net external current flows when v equals zero. From (81),

$$qP_0\mathcal{R}'(v) = \beta i_s e^{\beta v} \tag{83}$$

which, with (80), gives

$$qP_0\mathcal{R}'(v) = \beta i_s e^{\beta v_0} e^{\beta v_b \cos t}. \tag{84}$$

The Fourier expansion of (84) involves modified Bessel functions of the first kind,[16] denoted by I_n. Thus,

$$qP_0\mathcal{R}'(nb) = \beta i_s e^{\beta v_0} I_n(\beta v_b). \tag{85}$$

From (50) and (85), one finds that the conversion matrix is

$$Y(m, n; s) = \beta i_s e^{\beta v_0} I_{m-n}(\beta v_b). \tag{86}$$

To take a specific example, conversion between two frequencies differing by $b/2\pi$ will be considered [such as $(b + s)/2\pi$ and $s/2\pi$]. All other signal frequencies (images) will be assumed to be short-circuited. Then the conversion matrix reduces to

$$\| Y(m, n; s) \| = \beta(i_s + i_0) \begin{Vmatrix} 1 & y \\ y & 1 \end{Vmatrix}, \tag{87}$$

where

$$y = \frac{I_1(\beta v_b)}{I_0(\beta v_b)} \tag{88}$$

and i_0, the direct current, is introduced in view of the relation

$$i_0 = i_s[e^{\beta v_0} I_0(\beta v_b) - 1], \tag{89}$$

which may be obtained from (11).

For minimum conversion loss, the four-terminal network represented by (87) should be used with source and load admittances that simultaneously match the input and output admittances of the network. These admittances are, of course, real for a nonlinear resistor and are found, by standard methods,[17] to be

$$Y_{in} = Y_{out} = G_s = G_L = \beta i_s e^{\beta v_0} I_0(\beta v_b)\sqrt{1 - y^2}. \tag{90}$$

Since the images are short-circuited, all the μ_n's defined in (70) are zero except

$$\mu_{m_1} = 1$$

and

$$\mu_{m_2} = -\frac{1 + \sqrt{1 - y^2}}{y}. \tag{91}$$

For the noise figure, algebraic manipulation that will not be detailed here gives

$$F - 1 = \frac{T}{T_s}\frac{u^2 + \theta}{u + u^2}, \tag{92}$$

where

$$u = \sqrt{1 - y^2} \tag{93}$$

and

$$\theta = \frac{i_s}{i_0 + i_s}. \tag{94}$$

Equation (90) gives the dependence of F_x upon the local-oscillator drive (represented by u) and the direct current, represented by θ. The quantity θ defined in (94) is the ratio of the thermally generated current to the current carried by emitted carriers.

For any value of θ, there is a certain value of u,

$$u' = \sqrt{\theta^2 + \theta} - \theta, \tag{95}$$

that gives a minimum noise figure

$$F_{\min} - 1 = \frac{T}{T_s}\frac{2u'}{1 - 2u'}. \tag{96}$$

As θ approaches zero, $F_{\min}$ approaches unity. At the same time, the optimum value of u approaches zero. The optimum local-oscillator voltage goes to infinity and the conversion loss approaches unity (that is, no loss at all). This result shows that low noise figure and minimum conversion loss are not incompatible.

For most practical microwave diodes θ is quite small, since i_s may be a tenth of a microampere and the direct current i_0 is of the order of a milliampere. An approximate formula is therefore relevant:

$$F_{\min} \approx 1 + 2\sqrt{\theta}; \qquad \theta < 0.01. \tag{97}$$

For $\theta = 10^{-4}$, $F_{\min} = 1.02 = 0.1$ db. Thus, the observed thermally generated currents would not produce much noise and probably are not

responsible for much of the observed noise figure of crystals, except possibly at elevated temperatures.

This analysis has considered only the shot noise of the emitted carriers. In actual diodes, the series resistance and the depletion-layer capacitance impair the noise figure. The depletion-layer capacitance should be essentially noiseless, but, in combination with R_s, it sets bounds on the usable impedance level of nonlinear resistance of the emitted carriers. With the impedance level thus established, R_s causes a definite signal loss and contributes thermal noise. Furthermore, R_s interferes with short-circuiting the local-oscillator harmonics or otherwise obtaining the pulsed local-oscillator current waveform essential for low shot noise.

13.2 *Open-Circuited Harmonics*

When the harmonics of the local oscillator are open-circuited, the local-oscillator current is sinusoidal (in addition to a direct current). It will be shown that this waveform is unfavorable for noise, although it permits a low conversion loss.

The emission rate and the local-oscillator current are simply related:

$$qP_0\mathfrak{R}(t) = i_{LO}(t) + i_s = (i_0 + i_s)\left(1 + \frac{2\xi}{1+\xi^2}\cos bt\right), \tag{98}$$

where

$$0 \leqq \xi \leqq 1. \tag{99}$$

The reason for introducing the quantity ξ will be apparent later. The Fourier analysis of $P_0R(t)$ is obtained by inspection of (98):

$$\begin{aligned} qP_0\mathfrak{R}(nb) &= i_0 + i_s && \text{if } n = 0 \\ &= \frac{(i_0 + i_s)}{1+\xi^2} && \text{if } n = \pm 1 \\ &= 0 && \text{otherwise.} \end{aligned} \tag{100}$$

Since an exponential emission rate is being considered, $\mathfrak{R}'$ equals $\beta\mathfrak{R}$ and, from (50),

$$\begin{aligned} Y(m, n\,;\, s) &= \beta(i_0 + i_s) && \text{if } m = n \\ &= \frac{\beta(i_0 + i_s)\xi}{1+\xi^2} && \text{if } m = n \pm 1 \\ &= 0 && \text{otherwise.} \end{aligned} \tag{101}$$

The admittance representation of a frequency converter lends itself readily to consideration of cases where the image frequencies are short-circuited, since the elements of $Y(m, n; s)$ corresponding to the images may then simply be omitted. One finds, however, that, with open-circuited harmonics and short-circuited images, the conversion loss is at least 3 db, quite apart from the further losses due to the parasitics. As an absolute theoretical limit, this conversion loss appears to be excessive. In seeking an escape from this limit, one naturally wonders if open-circuited images might be appropriate for open-circuited harmonics.

Indeed, open-circuited images permit unity conversion gain. To deal with open-circuited images, we may invert the infinite-dimensional matrix $\| Y(m, n; s) \|$ to obtain the conversion matrix in impedance form:

$$Z(m, n; s) = \frac{(-\xi)^{|l-m|}}{\beta(i_0 + i_s)} \frac{1 + \xi^2}{1 - \xi^2}. \tag{102}$$

The images can now be open-circuited by considering only the 2×2 impedance matrix pertaining to the signals. This matrix may be inverted to give the admittance matrix for open-circuited images:

$$\| Y \| = \beta(i_0 + i_s) \begin{Vmatrix} 1 & \xi \\ \xi & 1 \end{Vmatrix}. \tag{103}$$

For the conversion between two frequencies differing by $b/2\pi$, with all images open-circuited, one has, in analogy with the short-circuit case,

$$Y_{\text{in}} = Y_{\text{out}} = G_L = G_s = \beta(i_0 + i_s)\sqrt{1 - \xi^2}. \tag{104}$$

The noise figure is found to be

$$F - 1 = \frac{T}{T_s} \frac{1 + \theta(2 - z^2)}{z(1 - z)(2 - z^2)}, \tag{105}$$

where

$$z = \sqrt{1 - \xi^2}. \tag{106}$$

For any value of z, the lowest noise figure is obtained when θ equals 0, that is, when the thermally generated current is negligible. For $\theta = 0$, the minimum noise figure is $3.25 = 5.1$ db; this occurs when z equals 0.45 and the conversion loss is 4.2 db. Evidently the noise performance

with open-circuited harmonics is inferior to that obtainable with short-circuited harmonics.

XIV. CONCLUSIONS

The nonlinear capacitance of a p-n junction can be used in amplifiers and amplifying frequency converters. To the extent that the junction exhibits pure capacitance, it will be noiseless in any of these applications. Suitably designed p-n junctions are remarkably good nonlinear capacitors at microwave frequencies, and preliminary experiments indicate that good noise figures can be obtained.

A pulsed local-oscillator current is required for low-noise frequency conversion with nonlinear resistance p-n junctions. This requirement is a logical consequence not only of p-n junction theory, but also of earlier metal-semiconductor rectification theories that assumed emission of carriers across a barrier.

XV. ACKNOWLEDGMENT

The author wishes to thank H. W. Andrews for advice and assistance in setting up the first up-converter experiment to give proof of the low-noise possibilities of diffused nonlinear capacitor diodes.

REFERENCES

1. Palevsky, H., Swank, R. K. and Grenchik, R., Design of Dynamic Condenser Electrometer, Rev. Sci. Instr., **18,** May 1947, pp. 298–314.
2. Scherbatskoy, S. A., Gilmartin, T. H. and Swift, G., The Capacitive Commutator, Rev. Sci. Instr., **18,** June 1947, pp. 415–421.
3. Manley, J. M. and Rowe, H. E., Some General Properties of Nonlinear Elements, Proc. I.R.E., **44,** July 1956, pp. 904–913.
4. Torrey, H. C. and Whitmer, C. A., *Crystal Rectifiers*, McGraw-Hill, New York, 1948, Chap. 13.
5. Semiconductor Diodes Yield Converter Gain, Bell Labs. Record, **35,** October 1957, p. 412.
6. Bakanowski, A. E., Cranna, N. G. and Uhlir, A., Jr., Diffused Silicon and Germanium Nonlinear Capacitance Diodes, I.R.E.-A.I.E.E. Semiconductor Device Research Conference, Boulder, Colo., July 1957.
7. Herrmann, G. F., Uenohara, M. and Uhlir, A., Jr., Proc. I. R. E., **46,** June 1958, pp. 1301–1303.
8. Leenov, D., this issue, pp. 989–1008.
9. Shockley, W., The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors, B.S.T.J., **28,** July 1949, pp. 435–589.
10. Shockley, W., Sparks, M. and Teal, G. K., p-n Junction Transistors, Phys. Rev., **83,** July 1951, pp. 151–162.
11. Uhlir, A., Jr., High-Frequency Shot Noise in p-n Junctions, Proc. I.R.E., **44,** April 1956, pp. 557–558; Nov. 1956, p. 1541.
12. van der Ziel, A., Theory of Shot Noise in Junction Diodes and Junction Transistors, Proc. I.R.E., **43,** November 1955, pp. 1639–1646.

13. Anderson, R. L. and van der Ziel, A., On the Shot Effect of p-n Junctions, Trans. I.R.E., **EO-1,** November 1952, pp. 20–24.
14. Guggenbuehl, W. and Strutt, M. J. O., Theory and Experiments on Semiconductor Diodes and Transistors, Proc. I.R.E., **45,** June 1957, pp. 839–854.
15. Rediker, R. H. and Sawyer, D. E., Very Narrow Base Diode, Proc. I.R.E., **45,** July 1957, pp. 944–953.
16. McLachlan, N. W., *Bessel Functions for Engineers*, Oxford Univ. Press, London, 1955, Chap. 6.
17. Lo, A. W., Endres, R. O., Zawels, U., Waldauer, F. D. and Cheng, C. C., *Transistor Electronics*, Prentice-Hall, Englewood Cliffs, N. J., 1955, pp. 100–102.