Thermal Noise in Field-Effect Transistors*

A. VAN DER ZIEL†, FELLOW, IRE

Summary—The limiting noise mechanism in field-effect transistors is thermal noise of the conducting channel. The noise can be represented by a current generator $\sqrt{\overline{i^2}}$ in parallel to the output. The value of $\overline{i^2}$ is calculated; for zero drain voltage the noise corresponds to thermal noise of the drain conductance, and for other bias conditions the noise at a given gate voltage depends only slightly upon the drain voltage. Because of modulation effects in the channel, $\overline{i^2}$ is somewhat larger than the thermal noise of the dc drain conductance, except for zero drain bias and beyond saturation. The noise resistance of the device is approximately equal to g_{max}/g_m^2, where g_m is the transconductance of the transistor and g_{max} its maximum value. The approximation becomes even closer if feedback due to the series resistances of the channel must be taken into account.

INTRODUCTION

SHOCKLEY[1] has given a theory of the field-effect transistor. This paper aims at applying his model for calculating the noise of the device, caused by thermal noise generated in the conducting channel.

Let the field-effect transistor be a planar transistor made on p-type material and let it be provided with two gate contacts G, a source contact S and a drain contact D (Fig. 1). Let the transistor have unit width, let $2a$ be the distance between the gate contacts and L the length of the conducting channel. Let, for a given bias W between the gate and the channel, the width of the channel be $2b$, and let W_{00} be the bias needed for cutoff ($b = 0$). Then, according to Shockley,

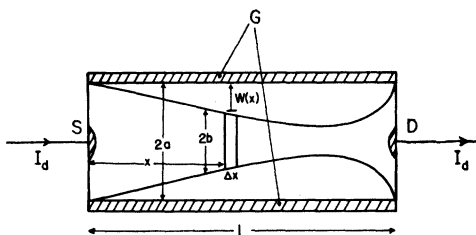$$W = W_{00}(1 - b/a)^2; \quad \text{or} \quad b/a = [1 - (W/W_{00})^{1/2}]. \quad (1)$$

Fig. 1—Cross section of a planar field-effect transistor showing the source S, the gate G, the drain D, the conducting channel of width $2b$ and the space-charge regions of width ($b-a$). $W(x)$ is the bias between the conducting channel and the gate.

Here W and b are slow functions of the distance x to the source. If V_g is the potential of the gate with respect to the source, V_d the potential of the drain and V_{dif} the diffusion potential, then $W = W_s = (V_g + V_{\text{dif}})$ at the source and $W = W_d = (V_g + V_{\text{dif}} - V_d)$ at the drain.

[1] W. Shockley, "A unipolar field-effect transistor," PROC. IRE, vol. 40, pp. 1365–1376; November, 1952.

For a given gate voltage the field strength component E_x in the X direction is

$$E_x = \frac{dW}{dx}. \quad (2)$$

If σ_0 is the conductivity of the p-type channel, then the dc current is

$$I = 2\sigma_0 b \frac{dW}{dx} = g(W) \frac{dW}{dx} \quad (3)$$

where

$$g(W) = 2\sigma_0 b$$
$$= g_0\left[1 - \left(\frac{W}{W_{00}}\right)^{1/2}\right], \quad \text{and} \quad g_0 = 2\sigma_0 a. \quad (3a)$$

Consequently, the current is

$$I = \frac{1}{L}\int_{W_s}^{W_d} g(W)dW$$
$$= \frac{g_0}{L}\left[W_d - W_s - \frac{2}{3}\frac{(W_d^{3/2} - W_s^{3/2})}{W_{00}^{1/2}}\right]. \quad (4)$$

The transconductance g_m of the device is

$$g_m = -\frac{\partial I}{\partial V_g} = \frac{g_0}{L}\left[\left(\frac{W_d}{W_{00}}\right)^{1/2} - \left(\frac{W_s}{W_{00}}\right)^{1/2}\right], \quad (5)$$

and the output conductance g_d of the device is

$$g_d = -\frac{\partial I}{\partial V_d} = \frac{g_0}{L}\left[1 - \left(\frac{W_d}{W_{00}}\right)^{1/2}\right]. \quad (6)$$

At $V_d = 0$, $W_d = W_s = (V_g + V_{\text{dif}})$, which is the smallest value W_d can take for given V_g. In that case $I = 0$ and $g_m = 0$ and

$$g_d = g_{d0} = \frac{g_0}{L}\left[1 - \left(\frac{W_s}{W_{00}}\right)^{1/2}\right]. \quad (6a)$$

For $W_d \rightarrow W_{00}$, $g_d \rightarrow 0$, the (I, V) characteristic becomes saturated and

$$g_m \rightarrow g_{max} = \frac{g_0}{L}\left[1 - \left(\frac{W_s}{W_{00}}\right)^{1/2}\right] = g_{d0} \quad (5a)$$

attains its maximum value. Adding (5) and (6) one obtains the general relationship

$$g_m + g_d = g_{max} = g_{d0}. \quad (5b)$$

If the drain voltage is more negative than is needed for saturation, the current I is practically independent of voltage and Shockley's solution does not hold.

CALCULATION OF THE NOISE IN FIELD-EFFECT TRANSISTORS

Let it now be assumed that the dc current is kept constant at the value I. A thermal noise voltage developed between x and $(x+\Delta x)$ will then modulate the width of the channel between x and the drain and give an amplified noise voltage at the drain. By integrating over all sections Δx one obtains the total output noise voltage.

To take this into account, one puts $W(x) = W_0 + \Delta W$ and $b(x) = b_0 + \Delta b$, where W_0 and b_0 are the dc values and ΔW and Δb are the fluctuations induced by the thermal noise developed between x and $(x+\Delta x)$.

square value is

$$\overline{\Delta W_x^2} = 4kT\Delta f \frac{\Delta x}{g(W_0)} = \frac{4kT\Delta f \Delta W_0}{g(W_0)dW_0/dx}$$

$$= \frac{4kT\Delta f}{I}\Delta W_0. \tag{11}$$

Hence we have

$$\overline{\Delta W_d^2} = \frac{4kT\Delta f}{I}\left[\frac{1 - (W_0/W_{00})^{1/2}}{1 - (W_d/W_{00})^{1/2}}\right]^2 \Delta W_0. \tag{12}$$

The total mean square open-circuit noise voltage is obtained by integrating over the length of the sample, that is, between the limits W_s and W_d. This yields

$$\overline{e^2} = \frac{4kT\Delta f}{I} \frac{\left[(W_d - W_s) - \frac{4}{3}(W_d^{3/2} - W_s^{3/2})/W_{00}^{1/2} + \frac{1}{2}(W_d^2 - W_s^2)/W_{00}\right]}{[1 - (W_d/W_{00})^{1/2}]^2}. \tag{13}$$

Eq. (1) thus gives

$$W_0 = W_{00}\left(1 - \frac{b_0}{a}\right)^2;$$

$$(W_0 + \Delta W) = W_{00}\left[1 - \left(\frac{b_0 + \Delta b}{a}\right)\right]^2. \tag{7}$$

Hence, neglecting the term in Δb^2,

$$\Delta W = -2\frac{W_{00}}{a}\left(1 - \frac{b_0}{a}\right)\Delta b = -\frac{2}{a}(W_{00}W_0)^{1/2}\Delta b. \tag{7a}$$

Eq. (3) gives

$$I = 2\sigma_0 b_0 \frac{dW}{dx} = 2\sigma_0(b_0 + \Delta b)\left(\frac{dW_0}{dx} + \frac{d\Delta W}{dx}\right)$$

$$= 2\sigma_0 b_0 \frac{dW_0}{dx}, \tag{8}$$

since I is kept constant. Neglecting second-order terms yields

$$\Delta b \frac{dW_0}{dx} + b_0 \frac{d\Delta W}{dx} = 0. \tag{8a}$$

Substituting (7a) into (8a) yields

$$\frac{d\Delta W}{\Delta W} = \frac{1}{2} \frac{d(W_0/W_{00})}{(W_0/W_{00})^{1/2}[1 - (W_0/W_{00})^{1/2}]} = \frac{du}{1 - u} \tag{9}$$

where $u = (W_0/W_{00})^{1/2}$. Integrating between the limits x and L, one obtains

$$\frac{\Delta W_d}{\Delta W_x} = \frac{1 - u_x}{1 - u_L} = \frac{1 - [W_0(x)/W_{00}]^{1/2}}{1 - (W_d/W_{00})^{1/2}} \tag{10}$$

where ΔW_x is the fluctuation in the section Δx and ΔW_d is the resulting fluctuation at the drain; $W_0(x)$ is the value of W_0 at the point x where the fluctuation occurs. Since ΔW_x is caused by thermal noise, its mean

The short-circuit noise current has the mean square value

$$\overline{i^2} = \overline{e^2}g_d^2 = 4kT\frac{g_0}{L}\Delta f$$

$$\cdot \frac{\left[(x - y) - \frac{4}{3}(x^{3/2} - y^{3/2}) + \frac{1}{2}(x^2 - y^2)\right]}{\left[(x - y) - \frac{2}{3}(x^{3/2} - y^{3/2})\right]}. \tag{14}$$

as is found by substituting for I and g_d and putting $x = W_d/W_{00}$ and $y = W_s/W_{00}$. The equation holds for $0 \le W_s/W_{00} \le W_d/W_{00} \le 1$, but does not hold in the saturated region of the characteristic.

For zero drain bias ($V_d = 0$), $x = y$ and the expression (14) can be written

$$\overline{i^2} = 4kT\frac{g_0}{L}\Delta f(1 - y^{1/2}) = 4kTg_{d0}\Delta f, \tag{14a}$$

as is found by taking the limit $x \to y$ in (14). Since g_{d0} is the ac output conductance for zero drain bias, (14) indicates that the device gives thermal noise for zero bias, as expected.

Since the dc conductance $g_{dc} = I/(V_s - V_d)$ approaches g_{d0} for $V_d \to V_s$, the device gives also thermal noise of the dc conductance for zero drain bias. If $V_d \ne V_s$ this is no longer the case, but it is still worth while to compare (14) with the thermal noise of the conductance g_{dc}. To that end one writes

$$\overline{i^2} = 4kTg_{dc}\Delta f P(x, y). \tag{15}$$

Substituting for $\overline{i^2}$ and I, putting $(V_s - V_d) = (W_d - W_s)$ and writing again $x = W_d/W_{00}$ and $y = W_s/W_{00}$, one obtains for the factor $P(x, y)$

$$P(x, y) = \frac{(x - y)\left[(x - y) - \frac{4}{3}(x^{3/2} - y^{3/2}) + \frac{1}{2}(x^2 - y^2)\right]}{\left[(x - y) - \frac{2}{3}(x^{3/2} - y^{3/2})\right]^2}.$$ (15a)

For $0 \leqq y \leqq x \leqq 1$, the function $P(x, y)$ is a monotonically *increasing* function of x for fixed values of y and a monotonically *decreasing* function of y for fixed values of x. $P(x, y)$ has the value 1.00 for $x = y$ (zero drain voltage), the value 3/2 for $x = 1$, $y = 0$ and the value 4/3 for $x = 1$, $y = 1$.

It can thus be concluded that the noise is nearly equal to the thermal noise of the dc conductance g_{dc} of the channel for all values of W_s and W_d, as long as the device is not operating on the saturated region of the characteristic.

OTHER EXPRESSIONS FOR THE NOISE

It is convenient to write the expression for $\overline{i^2}$ in a different form by putting

$$\overline{i^2} = 4kTg_{max}\Delta f Q(W_d, W_d)$$ (16)

where

$$Q(W_d, W_s) = Q(x, y)$$

$$= \frac{\left[(x - y) - \frac{4}{3}(x^{3/2} - y^{3/2}) + \frac{1}{2}(x^2 - y^2)\right]}{(1 - y^{1/2})\left[(x - y) - \frac{2}{3}(x^{3/2} - y^{3/2})\right]}$$ (16a)

as is found by substituting (14) and introducing the expression (5a) for g_{max}.

For $0 \leqq y \leqq x \leqq 1$ the function $Q(x, y)$ is a monotonically *decreasing* function of x for fixed values of y and a monotonically *increasing* function of y for fixed values of x. $Q(x, y)$ has the value 1.00 for $x = y$ (zero drain voltage), the value 1/2 for $x = 1$, $y = 0$ and the value 2/3 for $x = 1$, $y = 1$. Since W_s/W_{00} cannot be smaller than V_{dif}/W_{00} for positive gate bias, at least for the type of field-effect transistor under discussion, the smallest value of $Q(x, y)$ is always somewhat larger than 1/2.

In the saturated condition $x = 1$ and

$$\overline{i^2} = 4kTg_{max}\Delta f Q(1, y).$$ (16b)

The theory does not hold beyond saturation, but it is found experimentally that (16b) is nearly correct in the saturated part of the characteristic as long as the field strength in the cutoff part of the channel is not too large. Eq. (16b) may thus be applied under saturated condition.

It is often convenient to introduce the noise resistance, R_n of the device by the equation

$$\overline{i^2} = 4kTR_n\Delta f g_m^2.$$ (17)

Substituting for $\overline{i^2}$, one obtains for R_n

$$R_n = \frac{g_{max}}{g_m^2} Q(x, y).$$ (18)

Since $Q(x, y)$ has a value close to unity, R_n may be approximated as

$$R_n \simeq \frac{g_{max}}{g_m^2}.$$ (18a)

For a more accurate evaluation of R_n one has to determine the value of $Q(x, y)$.

At saturation $g_m = g_{max}$ and $x = 1$, hence

$$R_n = \frac{Q(1, y)}{g_{max}}.$$ (18b)

The theory does not hold beyond saturation, but in view of what was said about $\overline{i^2}$, (18b) remains valid in good approximation in the saturated region of the characteristic, as long as the field strength in the cutoff part of the channel is not too large. The value of $Q(1, y)$ usually lies between 0.60 and 0.67.

INFLUENCE OF THE SERIES RESISTANCES IN THE CHANNEL

If the gate contacts cover only part of the channel, one has to take into account the series resistance r_s at the source side of the channel and the resistance r_d at the drain side of the channel. The corresponding equivalent circuit is now as shown in Fig. 2.
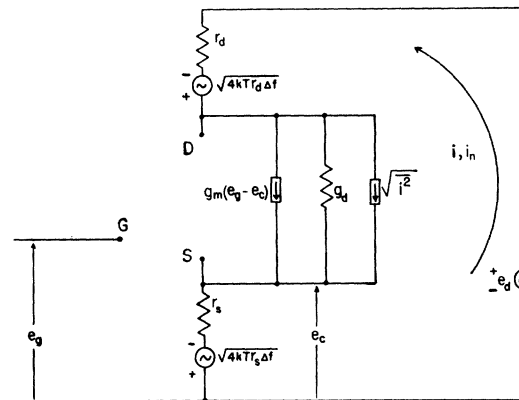

Fig. 2—Equivalent circuit of a field-effect transistor including the thermal noise of the two series resistances r_s and r_d of the channel.

It is easily demonstrated that the "apparent" transconductance g_m' of the device is given by

$$g_m' = \frac{g_m}{1 + r_s g_{max} + r_d g_d},$$ (19)

and that the "apparent" output conductance g_d' of the device is

$$g_d' = \frac{g_d}{1 + r_s g_{\max} + r_d g_d} \cdot \qquad (20)$$

At saturation $g_d = 0$ and the apparent maximum transconductance $g_{\max}'$ is

$$g_{\max}' = \frac{g_{\max}}{1 + r_s g_{\max}} \cdot \qquad (19a)$$

For zero drain bias $g_d = g_{d0} = g_{\max}$ and g_d' has the value g_{d0}'

$$g_{d0}' = \frac{g_{\max}}{1 + r_s g_{\max} + r_d g_{\max}} \cdot \qquad (20a)$$

Hence it is now no longer true that $g_{d0}' = g_{\max}'$; this might be used as a means for demonstrating series resistance effects.

The mean square value $\overline{i_n^2}$ of the noise current in the short-circuited output is found to be

$$\overline{i_n^2} = 4kT\Delta f \frac{[g_{\max}^2 r_s + g_{\max} Q(x, y) + g_d^2 r_d]}{(1 + g_{\max} r_s + r_d g_d)^2}$$

$$= 4kT g_{\max}' \Delta f Q'(x, y) \qquad (21)$$

where

$$Q'(x, y) = \frac{[Q(x, y) + g_{\max} r_s + g_d^2 r_d / g_{\max}](1 + g_{\max} r_s)}{(1 + g_{\max} r_s + r_d g_d)^2} \cdot \qquad (22)$$

Since $Q(y, y) = 1$ for zero drain bias and $g_d = g_{\max}$ in that condition, one finds for $Q'(y, y)$

$$Q'(y, y) = \frac{1 + g_{\max} r_s}{(1 + g_{\max} r_s + g_{\max} r_d)} < 1. \qquad (22a)$$

Under saturated conditions $x = 1$ and $g_d = 0$, so that

$$Q'(1, y) = \frac{Q(1, y) + g_{\max} r_s}{1 + g_{\max} r_s} \cdot \qquad (22b)$$

Since $Q(1, y) < 1$, (22b) indicates that

$$Q(1, y) \leq Q'(1, y) < 1. \qquad (22c)$$

The effect of the series resistances is thus to make the factor $Q'(x, y)$ more nearly independent of the drain voltage. If the device is completely cut off, $Q(1, y) \rightarrow Q(1, 1) = 2/3$ and $g_{\max} \rightarrow 0$. Consequently $Q'(1, 1) = Q(1, 1) = 2/3$.

Introducing the total noise resistance R_n' of the device, by putting

$$\overline{i_n^2} = 4kT R_n' \Delta f g_m'^2 \qquad (23)$$

yields

$$R_n' = \frac{g_{\max}'}{g_m'^2} Q'(x, y) \qquad (24)$$

so that the total noise resistance can be approximated by $(g_{\max}'/g_m'^2)$ in good approximation in many applica-

tions. For a more accurate calculation of R_n' the factor $Q'(x, y)$ must be evaluated.

The resulting noise resistance is quite low. For example, in a field-effect transistor in or near the saturated region with attainable values as $g_m' = g_{\max} = 1000$ μmho and with $Q(1, y) = 0.75$, which is a typical value, gives $R_n = 750$ ohms. This is about a factor four better than the shot noise resistance of a vacuum tube with comparable transconductance and considerably better than the noise resistance of transistor circuits at high input impedance levels.

OTHER NOISE SOURCES

It has been suggested[2] that the observed noise can be interpreted as suppressed shot noise. There is no physical basis for such a suggestion. For evidently the field-effect transistor operates on the principle of true conductance modulation, as Shockley's theory indicates. Generally one associates *thermal* noise with a true conductance and *not* shot noise. It is hard to see how shot noise could ever be generated and, if generated, how it could be partly suppressed. As this paper indicates, the assumption of thermal noise allows a straightforward explanation of the observed noise.

Even the cutoff portion of the channel behaves as a true resistor, though perhaps a nonlinear one, and thus should have thermal noise associated with it. This does not deny, of course, the possibility of an additional noise source. Present experimental evidence does not seem to indicate that this noise is important.[3,4]

Generation and recombination of carriers in the conducting channel would be another possible source of noise. This effect has been observed[3,4] in CdS field-effect phototransistors, where the carriers in the conducting channel are generated by the absorption of light. Noise due to deep lying traps would also belong to this category.

Besides the thermal noise of the conducting channel, field-effect transistors should also show shot noise of the gate current and $1/f$ noise of the gate and the channel currents.[2,3]

The shot noise of the gate current can be described as follows. Let the gate current I_g consist of a part $-I_{g1}$ due to holes arriving at the gate and electrons leaving the gate and a part $+I_{g2}$ due to holes leaving the gate and electrons arriving at the gate, then

$$I_g = -I_{g1} + I_{g2}. \qquad (25)$$

The noise can then be represented by a current generator $\sqrt{\overline{i_g^2}}$ between the gate and the source

$$\overline{i_g^2} = 2e(I_{g1} + I_{g2})\Delta f. \qquad (26)$$

[2] P. O. Lauritzen, "Field effect transistors as low-noise amplifiers," 1962 Internatl. Solid-State Circuits Conf., Philadelphia, Pa., February 14–16, 1962, *Digest of Technical Papers*, pp. 62–63; February, 1962.
[3] E. R. Chenette, to be published.
[4] W. C. Bruncke, to be published.

For low-noise operation the currents I_{g1} and I_{g2} should be kept as small as possible. Because of the location of the current generator $\sqrt{i_g^2}$, its effect will be especially pronounced for large values of the impedance in the gate circuit.

Some $1/f$ noise should be present at low frequencies, both in the gate current and in the channel current. The latter is most pronounced for low impedances in the gate circuit whereas the former will show up for large impedances in the gate circuit. As is well known, the magnitude of these $1/f$ noises can be substantially reduced by appropriate surface treatment, so that it need not be bothersome above a few hundred cycles.

At high frequencies the effect of the thermal noise in the conducting channel will also show up in the gate circuit. This comes about because of the capacitive coupling between the channel and the gate; as a consequence, a capacitive noise current will flow to the gate at high frequencies. This noise should be partly correlated with the channel noise and might thus be used to eliminate part of the latter.[2] It can be represented by a current generator $\sqrt{i_c^2}$ between the gate and the source; because of the capacitive coupling, i_c^2 should be proportional to the square of the frequency over a wide frequency range. Though it resembles induced grid noise in vacuum tubes in this respect, its interpretation is different, as the above discussion shows.

The shot noise of the gate current and the $1/f$ noises can be reduced by the choice of the transistor material, by improved construction of the device and by proper treatment of the finished product. The thermal noise of the conducting channel and the capacitive gate current resulting from it are always present; however, they determine the lowest noise figure that can be obtained with field-effect transistors.

Conclusions

The theoretical discussion shows that the basic limitation of the noise figure of field-effect transistors is the thermal noise of the conducting channel. Expressions are derived for the equivalent output noise current generator and for the equivalent noise resistance of the device that can be easily checked with experiment.

The theoretical predictions of this paper have been verified experimentally by Dr. E. R. Chenette[3] and W. C. Bruncke,[4] both of the Electrical Engineering Department of the University of Minnesota. The author is indebted to these investigators for several stimulating discussions about the interpretation of their experimental results, which led to the development of the theory presented in this paper.

Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz*

R. BECHMANN†, FELLOW, IRE, A. D. BALLATO†, MEMBER, IRE, AND
T. J. LUKASZEK†, MEMBER, IRE

Summary—The first-, second-, and third-order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz have been derived from thickness mode resonances of double-rotated quartz plates employing Christoffel's theory of wave propagation. The temperature dependence of all possible thickness modes can be calculated from the values of the elastic stiffnesses and their temperature coefficients as derived during this investigation. A curve showing the locus of the first-order zero temperature coefficient of frequency of thickness-shear modes has been calculated and compared with experiments. The second- and third-order temperature coefficients of frequency of the first-order zero quartz cuts are given. Applications to AT, BT, CT, and DT cuts are made by comparing the calculated with the experimental values which characterize the temperature behavior of frequencies and new useful piezoelectric cuts of quartz are indicated.

Introduction

THE FREQUENCY-temperature behavior of the various quartz cuts and consequently the behavior of the elastic constants, *i.e.*, the stiffnesses $c_{\lambda\mu}$ determining the thickness modes and the compliances $s_{\lambda\mu}$ determining the contour modes, is in general a nonlinear function of temperature. Therefore higher-order temperature coefficients have to be taken into consideration for these quantities, *e.g.*, by use of a power series.

The Christoffel theory [1] of propagation of plane waves governs the thickness modes of piezoelectric crystals. The six so-called Christoffel moduli $\Gamma_{ik} = \Gamma_{ki}$ ($i, k = 1, 2, 3$) which are combinations of the 21 stiffnesses $c_{\lambda\mu}$ ($\lambda, \mu = 1, 2, \cdots, 6$) and the direction cosines of the wave propagation were thereby introduced.