Broad-Band Characterization of FET Self-Heating

Anthony Edward Parker, Senior Member, IEEE, and James Grantley Rathmell, Member, IEEE

Abstract—The temperature response of field-effect transistors to instantaneous power dissipation has been shown to be significant at high frequencies, even though the self-heating process has a very low time constant. This affects intermodulation at high frequencies, which is examined with the aid of a signal-flow description of the self-heating process. The impact on broad-band intermodulation is confirmed with measurements over a range of biases. Intermodulation measurements are then used to obtain parameters that describe the heating response in the frequency domain. This description is then implemented in a time-domain model suitable for transient analysis and compared with measured heating and cooling step responses.

Index Terms—Intermodulation, memory effect, microwave field-effect transistor (FET), self-heating, thermal response.

I. INTRODUCTION

MODERN microwave circuit performance is susceptible to the dispersion of microwave transistor characteristics. This is because the complex signals used in communication systems invoke charge trapping and self-heating mechanisms that are sensitive to bias, temperature, and frequency variations. These mechanisms have slow time constants, so have not been considered significant at high-frequencies. Recently, however, the interaction of complex signals with slow time constants has been linked to the generation of intermodulation products and the asymmetry in intermodulation [1], [2]. This has been described as a memory effect.

Electrothermal simulations have alluded to the possible influence of self-heating on intermodulation [3], and a simulation of heating as a first-order memory effect has demonstrated an impact on intermodulation [4]. However, a standard finite-element thermal analysis, computed with finite differences [5], of a typical transistor structure has shown that the frequency response of self-heating is much more significant at high frequencies than a first-order model would suggest [6].

Conductance measurements can be used to determine the frequency response of self-heating for small-signal models [7]. The problem is that conductance is also influenced by trapping mechanisms that may not be separable. However, under specific bias conditions, intermodulation terms can be dependent on the thermal response to difference frequencies present in the signal. This can be used to characterize the frequency response of heating [8].

This paper develops a full characterization and a model of the self-heating process suitable for broad-band applications. Section I examines the impact of self-heating on intermodulation with the aid of a signal-flow description of the process. Section II presents a measurement of intermodulation that confirms the impact suggested by a previous simulation [8]. The impact on broad-band signals is discussed and the transfer response of self-heating is extracted from these measurements. Section III develops a time-domain model of self-heating and fits this to measured step responses of a microwave pseudomorphic high electron-mobility transistor (pHEMT). Section IV discusses the implementation of the self-heating model in circuit simulators.

II. SELF-HEATING AND INTERMODULATION PROCESSES

The self-heating process is a temperature change caused by power dissipation that has a frequency response determined by the physical structure of the transistor. The power dissipation is a function of the drain current. The drain current is influenced by the temperature dependence of carrier mobility, threshold voltage, and carrier saturation velocity [7]. Thus, the heating process is a feedback mechanism. Additional intermodulation products are generated by this feedback mechanism and its interaction with the inherent nonlinearity of a transistor.

A. Self-Heating

This study focuses on microwave field-effect transistors (FETs) operating in the active region where the temperature dependencies of threshold voltage and carrier mobility are dominated by the temperature dependence of carrier saturation velocity. In this case, the temperature dependence of drain current $I_d$ is well described in the time domain by

$$I_d(t) = I_{d0}(T_0) \cdot [1 - \delta \cdot P_d(t) \cdot h(t)]$$

where $I_{d0}(T_0)$ is the isothermal current (A) at ambient temperature, which is a nonlinear function of gate and drain potentials. The term $\delta$ (1/W) is a function of the thermal resistance of the transistor structure and the temperature dependence of the drain current. The power dissipation $P_d$ (W) is convolved with a thermal impulse response $h(t)$. The extraction from intermodulation measurements of a frequency-domain representation of $h(t)$ is given in Section III and the extraction from step-response measurements of $\delta$ is given in Section IV.

Fig. 1 shows a simplified signal-flow schematic of the heating process (1). The input signal passes through an intrinsic non-linearity to produce the isothermal current. This then passes through a negative-feedback process that implements the reduction in current caused by self-heating. The feedback loop passes instantaneous power dissipation through a low-pass filter. The instantaneous power is determined from the output current and voltage. The low-pass filter has a frequency-domain response $H(\omega)$ that is the Fourier transform of the thermal impulse response $h(t)$.
The salient features of the thermal response are captured in the following transfer function, which was chosen because it fits the simulation well:

$$H(\omega) = \frac{1}{(1+j\omega/\omega_c)^n(1+j\omega/\omega_L)^{(1-n)}}$$  \hspace{1cm} (2)

where \(\omega_c\) and \(\omega_L\) are the upper and lower rolloff frequencies and \(n\) is the order, which is less than unity, of the response in the intervening band [8].

The lower rolloff (<100 kHz) is set by the overall thermal path (the heat capacity and thermal resistance of the whole device). The upper rolloff (>1 GHz) is set by the size of the channel region where power is generated. Note that, in many cases, the choice of \(\omega_c\) is not significant because \(H(\omega_c)\) is small and other reactive components will dominate.

In the intervening band, the response reduces by less than 20 dB/decade. This implies that self-heating has a measurable effect at microwave frequencies. However, the main reason that self-heating has a significant impact on intermodulation is the interaction of a multitone signal with the low-frequency thermal response.

### B. Third-Order Intermodulation

There are three sources of nonlinearity within the system of Fig. 1 that contribute to intermodulation. The first is the calculation of instantaneous power, which involves an independent multiplication of drain current and drain voltage. This produces components at the fundamental frequencies (being the signal current multiplied by bias potential and signal voltage multiplied by bias current) and second-order products (from the multiplication of signal voltage and signal current).

The second source of nonlinearity is the multiplier that combines the isothermal drain current (produced by the intrinsic nonlinearity) with the output of the thermal response.

The third source of nonlinearity is the intrinsic nonlinearity of the device that generates fundamental, second-order, third-order, and higher order products.

For a two-tone input with frequencies \(\omega_a\) and \(\omega_b\), the third-order intermodulation product at a frequency of \((2\omega_a - \omega_b)\) is given by

$$\text{IMD} \propto \left\{ [H(\omega_a) + H(\omega_b)] \cdot \delta \cdot g_m \cdot (I_D R_L - V_{DS}) ight.$$  
$$+ [H(\omega_a - \omega_b) + H(2\omega_a)] \cdot 2\delta \cdot g_m^2 \cdot R_L$$
$$+ 2g_m^3 \right\} \cdot V_{DS}^2 R_L$$  \hspace{1cm} (3)

where \(g_m, g_m^2\), and \(g_m^3\) are first, second, and third derivatives of the output of the intrinsic nonlinearity with respect to its input [6]. The amplitude of the input tones are \(V_i\) (V), and the output bias current and voltage are \(I_D\) (A) and \(V_{DS}\) (V). The ratio of output voltage and current is \(R_L\) (\(\Omega\)), which is generally complex and frequency dependent. For clarity of the explanation, the transistor’s output conductance is not isolated in this simple expression. To understand the effect of \(H(\omega)\) alone, \(R_L\) is chosen to be constant over all frequencies and is set to 50 \(\Omega\) in the measurements.

Simulations and analyses [6] confirm that the three terms in (3) are produced by the following mechanisms.

- The term involving \([H(\omega_a) + H(\omega_b)]\) is generated by the multiplication of the intrinsic nonlinearity and thermal response outputs. Second-order products of the intrinsic nonlinearity at \((\omega_a - \omega_b)\) and \(2\omega_a\) are mixed with the fundamental products of the thermal response. The latter are proportional to both \(H(\omega_a)\) and \(H(\omega_a)\), and include the bias dependence associated with the first-order product of the power dissipation calculation.
- The term involving \([H(\omega_a - \omega_b) + H(2\omega_a)]\) is also generated by the same multiplier. Fundamental products of the intrinsic nonlinearity at \(\omega_a\) and \(\omega_b\) are mixed with the second-order products of the thermal response. The latter are proportional to \(H(\omega_a - \omega_b)\) and \(H(2\omega_a)\), respectively.
- The remaining term in (3) is the third-order product of the intrinsic nonlinearity. This is dominant at most biases, but is zero at a point near pinchoff in all FETs.

It is possible to control these contributions by the choice of signal frequencies and bias. The contribution involving \(H(\omega_a - \omega_b)\) can be made dominant in a two-tone measurement by using high-frequency tones and selecting a bias such that the other contributions cancel each other. The latter occurs at a bias near pinchoff. This leads to an extraction procedure for \(H(\omega)\) from a plot of intermodulation distortion (IMD) versus \((\omega_a - \omega_b)\). It is also possible to eliminate the first contributor by setting \(V_{DS} = I_D R_L\), though, in practice, this may be more difficult.

### III. INTERMODULATION MEASUREMENTS

The nature of the dependence of intermodulation on frequency spacing for a pHEMT can be seen in the measurements of Fig. 2, which is one of a set of many that were performed over varying drain bias and frequency conditions. For the measurements, two tones were used near 500 MHz, with a difference frequency less than 10% of the center frequency. Typically, these frequencies are orders of magnitude greater than \(\omega_c\) in (2) and orders of magnitude less than \(\omega_r\). Thus, \(H(\omega_a), H(2\omega_a), \text{and } H(\omega_b)\) change at a rate of 20\(\alpha\) dB/decade. Since there is only a 10% variation in frequency, these changes are less than 0.8 dB over the entire measurement for \(n < 1\). The value of \(H(\omega_a - \omega_b)\) varies considerably because \(\omega_a - \omega_b\) ranges from 0 to 50 MHz (>25 dB in Fig. 3).

The signal frequencies were set significantly lower than the \(f_1\) of the device so that access and capacitance elements are negligible. Drain bias was delivered via an RLC network that presented 50 \(\Omega\) to the drain at all frequencies. This was done so that
A. Broad-Band Variation

Self-heating coexists with, and in most cases is dominated by, other intermodulation mechanisms. However, for linear applications, operating points are selected to minimize intermodulation. At these points, self-heating is significant and gives intermodulation a dependence on frequency spacing.

Consider the case of low difference frequency, at the front edge of Fig. 2, where \( H(\omega_0 - \omega_T) \approx 1 \). There are two nulls that occur at gate biases such that the terms in (3) cancel. One is at the point near pinchoff where \( g_m'' = 0 \). The other, near \( V_{GS} = -0.8 \text{ V} \), is a point where \( g_m'' \) is sufficiently negative to cancel the other terms.

Consider the case of high difference frequency, at the rear edge of Fig. 2, where \( H(\omega_0 - \omega_T) \ll 1 \). The second term in (3) is substantially reduced so the \( g_m'' \) term near \( V_{GS} = -0.8 \text{ V} \) is no longer cancelled. The null near pinchoff still exists, but at a slightly different gate bias.

A significant problem for broad-band applications is the variation with difference frequency of the null near \( V_{GS} = -0.8 \text{ V} \) (for which the current is \( \approx 40\% \) of \( I_{DSS} \)). This variation is due to self-heating and persists in the active region over a very wide range of drain biases. The null can be exploited in applications with bandwidths less than the self-heating rolloff frequency \( \omega_T \).

However, components of a wider bandwidth signal would be presented with substantially larger intermodulation.

In the region of the intermodulation null, the thermal transfer function leads to a significant asymmetry in the intermodulation products. This is due to the conjugate relationship between the thermal transfer function \( H(\omega_0 - \omega_T) \) for the upper \( (\omega_0 > \omega_T) \) and lower \( (\omega_0 < \omega_T) \) intermodulation products.

Self-heating is not the only frequency-dependent process evident in the pHEMT of Fig. 2. Impact ionization and electron trapping are also present to varying degrees depending on bias. In Fig. 2, there is a marked reduction in intermodulation in the high gate-bias high-difference frequency region. The frequency at which this reduction occurs varies exponentially with drain bias, which suggests that it is related to charge trapping (possibly impact ionization [9]). Trapping-related dependencies are beyond the scope of this paper, but will be the subject of further investigations.

B. Transfer Function

The local minimum in intermodulation near pinchoff is visible on the right-hand-side edge of the surface in Fig. 2. Although not clear in Fig. 2, the \( H(\omega_0 - \omega_T) \) response is visible in this null, which can be used to extract \( H(\omega) \). This occurs at a bias at which all terms not involving \( H(\omega_0 - \omega_T) \) in (3) cancel each other. There is also a contribution from threshold-voltage temperature dependence that is not considered in (3). However, this is dependent on the same thermal transfer function so the net effect is an offset in the magnitude of the intermodulation, but not a change in the rolloff frequency or slope.

The third-order intermodulation for the pHEMT of Fig. 2 is shown in Fig. 3 for a region near pinchoff (5\% \( I_{DSS} \)). To observe a full view of the thermal transfer function \( H(\omega_0 - \omega_T) \), a
gate bias that minimizes the other contributors to intermodulation can be used. This is the bias that gives minimum intermodulation at large difference frequencies.

The $H(\omega)$ curve in Fig. 3 is an estimate of the thermal response of the device with $\omega_c$ and $n$ fitted to the data. This line is seen to approximate the actual response, albeit with a sharper rolloff. Note that $H(\omega)$ in (2) was chosen to simply capture the salient features of the thermal response.

The difference frequencies used are too small to observe $\omega_c$. This term could be estimated from knowledge of the channel dimensions and its thermal properties. However, any value for $\omega_c > 1$ GHz would be suitable for most applications because the thermal response is relatively small at this frequency. The magnitude of the $H(\omega)$ curve in Fig. 3 was arbitrarily set to fit the graph, but cannot be used to determine $\delta$ because the intrinsic nonlinearity terms are not known. A transient measurement can be used to determine $\delta$.

IV. TRANSIENT RESPONSE

Step response measurements for the pHEMT are shown in Fig. 4. Each measurement is for a step change from a cool or warm bias point. After stepping from the cool bias to a higher power dissipation point, there is a significant droop in current over time as the device heats up. After stepping from the warm bias to a lower power point, there is an increase in current as the device cools down. For a sufficiently long time, the measurements settle to the same dc characteristics.

In the time domain, it is possible to approximate the step response corresponding to the thermal transfer function $H(\omega)$ by a stretched exponential $u(t)$, which was chosen because it fitted the thermal analysis simulation well, as follows:

\[ u(t) = 1 - \exp \left[ - \left( \frac{\omega_c t}{n^2} \right)^n \right]. \tag{4} \]

From a consideration of (4), with $\omega_c = 2\pi \times 8000 \text{ rad/s}$ and $n = 0.35$ (taken from Fig. 3), the thermal response to a step input has settled to 0.2% at $t = 50 \mu s$. Thus, the isothermal current after the step can be approximated by

\[ I_{do} \approx I_d(t=50 \mu s)/(1 - \delta P_d), \tag{5} \]

For a step change from a point dissipating power $P_q$ to another point dissipating power $P_d$, the drain current is

\[ I_d \approx I_{do} \left[ 1 - \delta P_q [1 - u(t)] - \delta P_d u(t) \right]. \tag{6} \]

This assumes an ideal step change that ignores the variation in power dissipation associated with the current droop. This avoids the calculation of the convolution that would be better done in a circuit simulation. The simplification provides an effective process for determining the parameter $\delta$ and a means to extrapolate isothermal characteristics from step response measurements. For example, the parameter $\delta$ in Fig. 4 was selected for a good fit over all lines in the data.

The lines of Fig. 4 are consistent with the heating and cooling of the step changes. There are, however, deviations at extremes of step changes and long times. There is evidence of a 100-ns impact ionization process that is dependent on drain potential and is considerably slower at lower drain biases. At the 5.0-V bias selected here, this trapping process was faster than the thermal process being investigated. Since the thermal response derived from intermodulation measurements indicates that the process has settled by 50 $\mu s$, the long-time droop in drain current must be due to other mechanisms such as electron trapping.

V. MODEL IMPLEMENTATION

The thermal impulse response $h(t)$ in (1) is represented above in either the frequency domain by (2) or as a step response by (4). These representations, i.e., $H(\omega)$ and $u(t)$, are approximations that well describe the salient features of the thermal response in their respective domains. It should be noted that the inverse Fourier transform of $H(\omega)$ and the differentiation of $u(t)$ yield different impulse responses. However, simulations have been carried out to confirm that these impulse responses are approximately the same. Figs. 3 and 4 demonstrate that $H(\omega)$ and $u(t)$ are self-consistent enough to fit data from the same device. There is scope for further study to devise self-consistent and yet tractable $H(\omega)$, $u(t)$, and $h(t)$ functions.

In frequency-domain simulators, the implementation of $H(\omega)$ can be straightforward. Some simulators (such as ADS) provide a circuit element that is explicitly described by a user-defined function of $\omega$.

For other simulators, including a time-domain simulator such as SPICE, the array network in Fig. 5 can be used. The network consists of a current source proportional to the instantaneous
power dissipation and a ladder of resistor and capacitor elements. The values of the resistors and capacitors follow a geometric series with growth rates $a$ and $c$, respectively. For upper and lower rolloff frequencies of $\omega_c$ and $\omega_0$, the resistors and capacitors of Fig. 5 can be estimated with the aid of Table I. The values of $a$, $c$, and $C_n$ need to be tuned iteratively to optimize the simulated values of $n$, $\omega_0$, and $\omega_c$. A spread sheet is useful for determining initial values before optimizing the response with simulation. For a smooth response, the number of elements should be set so that there is approximately one node per decade of frequency between the lower and upper rolloff points. Fewer points will improve simulation speed at the expense of accuracy.

Fig. 6 shows the frequency- and time-domain responses of two ladder networks compared with $H(\omega)$ and $u(t)$, respectively. The eight-element network provides a reasonable simulation of both phase and magnitude. The four-element network is also reasonable in magnitude and may be adequate in most applications.

Note that $\omega_c$ need be no larger than the highest harmonic frequency of interest in the simulation. Reducing $\omega_c$ on a case-by-case basis can speed up simulations that would otherwise need to resolve the very-fast time constants involved.

VI. CONCLUSION

Measurements presented here show that the self-heating process in FETs has a significant impact on intermodulation of broad-band signals. This is due to a feedback mechanism that is sensitive to difference frequencies present in the signal.

For broad-band applications, simulators can account for self-heating by including it in the device models with the architecture of Fig. 1. The implementation of these models requires a description of the thermal response. This response is dependent on the physical thermal properties of the device, but has features that can be captured by suitable approximations. A frequency-domain approximation $H(\omega)$ and a time-domain sub-circuit (Fig. 5) have been presented here. These approximations compare well with measured intermodulation and transient responses. They should be suitable for most applications.

Note that electron trapping and impact ionization, which will each exhibit a frequency dependence, still need to be investigated. A similar analysis and sub-first-order frequency response may prove appropriate. The isolation of the thermal response with intermodulation measurements will be helpful in an investigation of this.

The approach to thermal characterization and modeling presented here is based on simple physical concepts. Therefore, it has broad application to other devices and material systems.

TABLE I
GUIDE FOR SETTING UP THE LADDER NETWORK IN Fig. 5

<table>
<thead>
<tr>
<th>Variable</th>
<th>Initial Guess</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>$&gt; \log_{10}(\omega_c/\omega_0)$</td>
<td>Adjust for smoothness</td>
</tr>
<tr>
<td>$a$</td>
<td>$&gt; 1$</td>
<td>Adjust for correct $n$</td>
</tr>
<tr>
<td>$c$</td>
<td>$\frac{n-1}{\omega_c} \sqrt{\frac{\omega_c}{\omega_0}}$</td>
<td>Adjust for correct $\omega_0$</td>
</tr>
<tr>
<td>$R_k$</td>
<td>$a^{k-1} \frac{1}{n-1}$</td>
<td>Sets $\sum R_k = 1$</td>
</tr>
<tr>
<td>$C_k$</td>
<td>$c^{k-1} \frac{1}{\omega_c R_0}$</td>
<td>Adjust for correct $\omega_c$</td>
</tr>
<tr>
<td>$n$</td>
<td>$1 - \frac{c}{c+a} \frac{N}{N+1}$</td>
<td>Guide only</td>
</tr>
</tbody>
</table>

![Fig. 6. Responses of the ladder network in Fig. 5 with four (–) and eight (—) nodes compared with ( - - -): (a) the frequency-domain response $H(\omega)$ and (b) the time-domain response $u(t)$. Parameters are $n = 0.5$, $\omega_0 = 2 \times 10^3$ rad/s, and $\omega_c = 2 \times 10^9$ rad/s.](image)

REFERENCES


Anthony Edward Parker (S’84–M’90–SM’95) received the B.Sc., B.E., and Ph.D. degrees from The University of Sydney, Sydney, Australia, in 1983, 1985, and 1992, respectively.

In 1990, he joined Macquarie University, Sydney, Australia, where he is currently Head of the Electronics Department. He is involved with a continuing project on pulsed characterization of microwave devices and design of low-distortion communications circuits. He has consulted with several companies including M/A-COM, Lowell, MA, and Agilent Technologies, Santa Rosa, CA. He has developed accurate circuit simulation techniques, such as used in field-effect transistor (FET) and high electron-mobility transistor (HEMT) models. He has authored or coauthored over 120 publications. His recent research has been in the area of intermodulation in broad-band circuits and systems, including a major project with Mimix Broadband Inc.

Prof. Parker is a member of the Institution of Telecommunications and Electronic Engineers, Australia, and a committee member of the IEEE Antennas and Propagation (AP)/Microwave Theory and Techniques (MTT) New South Wales (N.S.W.) Local Chapter.

James Grantley Rathmell (M’89) received the B.Sc., B.E., and Ph.D. degrees from The University of Sydney, Sydney, Australia, in 1977, 1979, and 1988, respectively.

From 1979 to 1981, he was involved with radio astronomy with both the Molonglo and Fleurs aperture synthesis telescopes. From 1982 to 1986, he was involved with gate array and very large scale integration (VLSI) design of digital integrated circuits. During this time, he was a member of the teaching staff with The University of Sydney. In 1986, he joined The Nucleus Group, where he was involved with the design of biomedical equipment. He managed the research and development of advanced ultrasound imaging equipment. In 1989, he returned to The University of Sydney. In 1995, he began a collaborative with Macquarie University, Sydney, Australia, where he is involved with collaborative work on the development of pulsed-bias and pulsed 5–parameter characterization of microwave devices.