The Effects of BJT Self-Heating on Circuit Behavior

Robert M. Fox, Senior Member, IEEE, Sang-Gug Lee, and David T. Zweidinger

Abstract—This study demonstrates the circuit and device conditions under which self-heating can significantly affect bipolar junction transistor (BJT) circuit behavior. Simple quantitative measures are supplied that allow estimation of thermally induced errors in BJT small-signal parameters, based on knowledge of the transistor geometry and its Early voltage. It is shown that errors in output admittance and reverse transadmittance can be significant without much power dissipation, especially when the base and emitter driving impedances are small. Other small-signal parameters are less affected unless the power dissipation becomes significant. Thermal effects in large-signal dc analysis can be significant in precision analog circuits that depend on close transistor matching; such circuits can also exhibit long settling-time tails due to long thermal time constants. ECL delay is shown to be insensitive to self-heating. These effects are demonstrated through simulations of a variety of circuits using versions of SPICE modified to include physics-based models for thermal impedance.

I. INTRODUCTION

FEW electronic devices are more sensitive to temperature variations than bipolar junction transistors (BJT's). This fact has been routinely considered in the design and application of power BJT's. There are many papers on the characterization and measurement of thermal impedances for power devices, especially on the effects of packaging and heatsinking [1], [2]. There have also been studies of the effects of cross-chip thermal gradients on circuits [3] and of the effects of heat flow within the transistor itself on BJT behavior [4], [5].

However, relatively few papers have examined the effects of self-heating on BJT operation in relatively low-power circuits [6]-[9]. Several current trends have led to an increase in interest in BJT self-heating. These include a continuing rise in current densities, which tends to raise device operating temperatures, and increasing reliance on simulation in the circuit design process, which makes accurate modeling increasingly important. The trend toward use of SiO$_2$ for lateral or vertical isolation also tends to make self-heating more important, since SiO$_2$'s thermal conductivity is much less than that of silicon.

This paper quantifies the conditions under which circuits and devices are sensitive to modeling errors if self-heating is neglected. Section II presents a classification of thermal effects into three categories, based on different thermal impedance mechanisms. Primary emphasis in this paper is on those thermal effects that are intrinsic to the operation of the transistor, as opposed to those that are controlled by the chip-to-package-to-ambient thermal impedance. Section III presents expressions for the temperature coefficients of the collector and base currents and the base-emitter voltage, based on the Gummel-Poon model. The expressions are applicable in the forward-active region at all current levels. Section IV concerns self-heating effects on small-signal operation, showing that thermal effects in large-signal dc analysis may show errors with moderate power. In Sections III-V, example circuits are simulated using a version of SPICE modified to allow each transistor to be simulated at its own temperature, calculated based on its power dissipation and a thermal impedance computed from the transistor's geometry. The thermal impedance models in the program are summarized in Section II, and the temperature dependences of the relevant Gummel-Poon parameters are presented in Section III. Other details of the implementation were presented in [10]. The software at present works only in ac and dc analyses. In Section VI, a SPICE behavioral thermal subcircuit is used to demonstrate thermally induced transient effects in analog circuits and to study thermal effects on the delay of BJT digital circuits. SPICE and other parameters for the transistor used in most of these simulations are given in Table I. This is a fairly big transistor, with a modest thermal resistance. The predictions are thus conservative; smaller transistors would show even larger thermal effects.

II. THERMAL EFFECTS IN BJT'S

As noted above, several different mechanisms can contribute to heating effects in BJT circuits [8], [10]. In most cases they can be considered separately. Two of these mechanisms can cause a coupling from the power of one transistor to the temperature of another on the same chip. Of these mechanisms the more important is usually the chip-to-package-to-ambient thermal impedance $Z_{TH}^{pkg}$. The total power dissipated on the chip can be multiplied by this value to obtain an overall chip temperature. Thermal transients mediated by $Z_{TH}^{pkg}$ can start less than 1 ms after a power change and last for many minutes. This wide range of time scales is typical of distributed effects such as heat flow. Typical values of package thermal resistance range from 5 to 200 K/W, with differences due to different
package types. This thermal resistance can also be affected by the quality of the die attachment and by heat sinking.

The other effect that can cause cross-chip coupling is the temperature gradient that can arise when power is generated nonuniformly across a chip. In [3], it was shown that thermal feedback from the output stage of an op amp to its input stage can profoundly affect its voltage gain if the output stage dissipates high power. As a result of this work, heuristic guidelines were developed to help avoid such problems through careful layout, especially through use of cross-coupled differential pairs. Heat-flow analysis [4] shows that except for rather high-power transistors, temperature gradients are mostly confined within the heat-generating transistor itself.

The third effect, the primary mechanism considered in this paper, is the rise in a transistor’s emitter junction temperature caused by the transistor’s own power dissipation. A time-domain model for the thermal spreading impedance that controls this effect was derived in [4] and a simplified frequency-domain equivalent was presented in [11]. The models assume that heat is generated uniformly in the collector space-charge region (SCR) below the emitter in a uniform semi-infinite medium with an adiabatic surface. The models show that the temperature drops below 10% of its peak value for all distances greater than about \( \sqrt{WL} \) from the transistor, where \( W \) and \( L \) are the emitter width and length. In most technologies this distance is greater than the minimum interdevice spacing, except for transistors with very large emitters. Thus, even when nominally matched transistors are laid out next to each other, their temperatures can differ if they dissipate different powers.

The simplified model in [11] also provides an easy way to estimate the effective thermal spreading resistance \( R_{TH} \) for a given transistor geometry and operating point. Let \( d = D/\sqrt{WL} \), where \( D \) is the base-collector junction depth, and let \( h = H/\sqrt{WL} \), where \( H \) is the SCR thickness, which can be estimated using the depletion approximation. Let \( a = W/L \), with \( W \geq L \). Then the thermal resistance can be calculated from

\[
R_{TH} \approx 1/(4\pi K \sqrt{WL} \cdot f_1 \cdot f_2)
\]

(1)

where \( K \) is the thermal conductivity (1.45 W/cm-K for Si at 300 K), and

\[
f_1(d,h) = (0.058d + 0.14)h + 0.34d + 0.28
\]

\[
f_2(a) = 0.98 + 0.043a - 6.9 \cdot 10^{-4}a^2 + 3.9 \cdot 10^{-6}a^3.
\]

As an example, consider a transistor with a 10 \( \times \) 7-\( \mu \)m\(^2\) emitter, with \( D = 0.4 \) \( \mu \)m and epitaxial collector doping \( N_{EPI} = 1 \times 10^{16} \) cm\(^{-3}\), operated at \( V_{CB} = 6 \) V. Equation (1) gives \( R_{TH} \approx 200 \) K/W.

It is also shown in [11] that the frequency variation of thermal impedance can be approximated by modeling the collector as a point heat source a distance \( r_{eff} \) away from a point emitter, where \( r_{eff} = 1/(2\pi \cdot K \cdot R_{TH}) = 2\sqrt{WL} \cdot f_1 \cdot f_2 \). Then

\[
Z_{TH}(\omega) = R_{TH} \cdot \exp(-r_{eff} \sqrt{j\omega/\kappa})
\]

(2)

where \( \kappa \) is the diffusivity of silicon, about 0.89 cm/\( \sqrt{\text{s}} \) for silicon at 300 K. This function rolls off rather slowly with frequency. For the transistor in the example above, \( r_{eff} \approx 5.4 \) \( \mu \)m, and \( |Z_{TH}| \) drops below 0.18 \( R_{TH} \) for frequencies above 4.7 MHz. As noted in [11], these approximations have been verified experimentally for a variety of junction-isolated BJT's. They do not apply to dielectric- or trench-isolated transistors. Measured thermal resistances for dielectrically isolated BJT's are typically three or more times those predicted by (1) [12].

III. TEMPERATURE COEFFICIENTS

As will be demonstrated, self-heating effects can be observed at all current levels. For modeling and for parameter extraction, it is important to know how the temperature sensitivities of the transistor currents and voltages vary with operating point. The fractional temperature coefficient of a parameter \( X \) is given by \( TCF(X) = (\partial X/\partial T)/X \). In the forward-active region, the Gummel–Poon model for the BJT [13] gives

\[
I_C = I_{BE1}/K_{ph}, \quad I_{BE1} = I_S \exp(V_{BE}/V_1), \quad V_I = kT/q \text{ is the thermal voltage and } K_{ph} \text{ is the base-charge factor defined by}
\]

\[
K_{ph} = \frac{1 + \sqrt{1 + 4(I_{BE1}/I_K)}}{2(1 - V_{BC}/V_A - V_{BE}/V_{AR})}
\]

(3)

where \( I_K \) is the forward knee current and \( V_A \) and \( V_{AR} \) are the forward and reverse Early voltages. The saturation current can be expressed as

\[
I_S = W \cdot L \cdot I_{SO} \exp \left[ \frac{(V_{GO}/V_I)}{(T/T_0 - 1)} \right]^{\gamma}
\]

(4)

where \( I_{SO} \) is the value of \( I_S \) at some nominal temperature \( T_0 \) and \( V_{GO} \) is bandgap potential. The temperature exponent \( \gamma \) is given by \( 4 + n \), where \( n \) is the temperature coefficient of the base-region minority-carrier mobility, usually near -1.0, so \( \gamma \) is generally about 3.0. The temperature coefficient of the ideal current \( I_{BE1} \) is

\[
D_{BE1} = TCF(I_{BE1})|_{V_{ex} = \text{fixed}} = \frac{1}{T} \cdot [\gamma + (V_{GO} - V_{BE})/V_I].
\]

(5)

The TC\( F \) of the collector current can then be written as

\[
D_C = TCF(I_C)|_{V_{ex} = \text{fixed}} = D_{BE1}(1 - f)
\]

(6)

where

\[
f = \frac{I_C/I_K}{(1 - V_{BC}/V_A)} \cdot \frac{1}{\sqrt{1 + 4(I_{BE1}/I_K)}}.
\]

(7)

The high-current correction factor \( f \) only becomes significant as the current approaches \( I_K \).

At moderate to high currents, base current \( I_B \) is proportional to \( I_{BE1} \), but at low currents nonideal base current components become significant, so \( I_B \) is modeled as \( I_B = I_{BE1}/\beta_F + I_{BE2} \), where \( \beta_F \) is the large-signal current gain and the nonideal component is \( I_{BE2} = I_{SE} \exp(V_{BE}/n_E V_I) \), where \( n_E \) is a constant, typically around 1.5. \( \beta_F \) can be assumed to vary as \( \beta_F = \beta_{FO} e^{-\lambda} \), where \( \beta_{FO} \) is the \( T = T_0 \) value and \( \lambda \approx \Delta V_G/V_I \), where \( \lambda \) is the emitter–base bandgap difference [14], which is controlled by bandgap narrowing for
Table I

<table>
<thead>
<tr>
<th>Parameter Value</th>
<th>Parameter Value</th>
<th>Parameter Value</th>
<th>Parameter Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i_E$</td>
<td>0.015 A</td>
<td>$T$</td>
<td>3.0</td>
</tr>
<tr>
<td>$i_B$</td>
<td>220 mA</td>
<td>$\lambda$</td>
<td>2.1</td>
</tr>
<tr>
<td>$V_A$</td>
<td>150 V</td>
<td>$C_{EE}$</td>
<td>0.4 nF</td>
</tr>
<tr>
<td>$i_K$</td>
<td>20 mA</td>
<td>$C_{JC}$</td>
<td>0.3 $\mu$F</td>
</tr>
<tr>
<td>$i_{SE}$</td>
<td>7.0 mA</td>
<td>$C_{BE}$</td>
<td>0.3 $\mu$F</td>
</tr>
<tr>
<td>$r_E$</td>
<td>2.05 $\Omega$</td>
<td>$r_f$</td>
<td>25 ps</td>
</tr>
</tbody>
</table>

Fig. 1. Fractional temperature coefficients of collector current $D_C$ and $D_B$ assuming fixed $V_{BE}$, computed using (6) and (10), for the transistor whose parameters are given in Table I. Also plotted is $D_3$, the temperature coefficient of the current gain for fixed $I_C$.

homojunction transistors, a typical value for $\lambda$ being around 2.0. In heterojunction BJTs, the bandgap difference can take on a wider range of values, and $X$ is usually negative. The temperature dependence of $I_{SE}$ is given by

$$I_{SE} = W \cdot L \cdot I_{SE0}(T/T_0)^{(\lambda - \gamma/n_e)} \cdot \exp[(T/T_0 - 1)(V_{GO}/n_E)]$$

(8)

so that for fixed $V_{BE}$,

$$D_{BE2} = \Delta C_f(I_{BE2})$$

$$= \frac{1}{n_E} \left( \frac{V_{GO} - V_{BE}}{V_t} + \gamma - n_E \cdot \lambda \right)$$

(9)

Then

$$\Delta C_f(I_{BE})|_{fixed V_{BE}} = D_B = (I_{BE1}/\beta_f)[D_{BE1} - (\lambda/T)] + D_{BE2}I_{BE2}/I_B.$$  

(10)

Also

$$\Delta C_f(V_{BE})|_{fixed I_C} = -\frac{D_B}{1 - f}$$

$$\Delta C_f(I_B)|_{fixed I_C} = -\frac{D_B}{1 - f}$$

where $D_B = D_C - D_B$.

The temperature coefficients depend on loading conditions; for example, if $V_{BE}$ is held fixed, $\Delta C_f(I_C) = D_C$, as given by (6), but if $I_B$ is fixed instead, $\Delta C_f(I_C) = D_B$. At moderate current levels, $D_B$ is generally much less than $D_C$ and is dominated by $\lambda/T$. Fig. 1 shows plots of $D_C, D_B$, and $D_3$ for the typical transistor whose parameters are listed in Table I.

IV. SELF-HEATING EFFECTS ON BJT SMALL-SIGNAL BEHAVIOR

An easy way to understand self-heating effects in circuits is through the effects on BJT small-signal parameters [7], [15]. Assume that with self-heating neglected the common-emitter $g$ parameters are given by

$$g_{mE} = 1/T_e \cdot g_{11E} = g_{12E} = \frac{1}{r_e}$$

and

$$g_{11E} = \frac{g_{11E}}{1 + D_C R_{TH} I_C}$$

$$g_{12E} = \frac{g_{12E}}{1 + D_C R_{TH} I_C}$$

(11)

assuming that $i_C / i_B \cong I_C / I_B$. Similarly, the base current can be written as

$$i_b = \frac{g_{11E} V_{be} + g_{12E} V_{ce} + D_C R_{TH} I_C (V_{be} + I_c V_{ce})}{1 - D_C R_{TH}}$$

(12)

Thus the $g$ parameters, corrected for self-heating, can be expressed as

$$g_{11} = \frac{g_{11E} + D_C R_{TH} I_C}{1 - D_C R_{TH}}$$

$$g_{12} = \frac{g_{12E} + D_C R_{TH} I_C}{1 - D_C R_{TH}}$$

$$g_{21} = \frac{g_{21E} + D_C R_{TH} I_C}{1 - D_C R_{TH}}$$

$$g_{22} = \frac{g_{22E} + D_C R_{TH} I_C}{1 - D_C R_{TH}}$$

(13a)-(13d)

The denominators in (13a)-(13d) become significant as the power approaches the critical values

$$P_C = \frac{1}{D_C R_{TH}}$$

and

$$P_B = \frac{1}{D_B R_{TH}}$$

where plots of $I_C$ and $I_B$ versus the port voltages have infinite slopes. These limits are related to the onset of second breakdown. The typical transistor whose parameters are given in Table I, operated at a $V_{CE}$ of 3 V, requires a current of about 4.5 mA to get a 10% error in the denominators in (13). This is a fairly high current. In most analog circuits the power dissipation is much less than $P_C$ or $P_B$, so that the effects of the denominator are modest except in circuits requiring great precision. Note, however, that as $\lambda$ values for $R_{TH}$ rise, either because of technology scaling or because of the use of dielectric or trench isolation, the power required to cause errors will shrink.

Now consider $g_{21}$. The electrical-only transconductance $g_{21E}$ is approximately $g_m = I_C / V_t$. If the denominator in (13c) is about unity, then the value of $g_{21}$ corrected for self-heating is approximately $g_{21E}(1 + D_C R_{TH} I_C V_t)$. Similarly, the electrical-only input conductance is $g_{11E} = 1/T_e = I_B / V_t$, so that $g_{11} = 1/T_e \cong g_{11E}(1 + D_C R_{TH} I_B V_t)$. For any reasonable conditions, the term $D_C R_{TH} I_B V_t$ is very small, so
these two parameters are only affected by the corrections in the denominators, which require substantial power dissipation.

The situation is different for the other two g parameters. A simple model for the electrical output conductance is \( g_{22E} = 1/r_{OE} = I_C/V_A \), where \( V_A \) is the Early voltage. Thus, from (13d), for \( P \ll 1/(D_C R_{TH}) \), \( g_{22} \approx g_{22E}(1 + K_C) \), where \( K_C = D_C R_{TH} I_C V_A \) can be thought of as a figure of merit giving the fractional error in the output conductance caused by self-heating for low power. \( K_C \) can be substantial even for very modest currents. For example, for the transistor in Table 1, \( K_C \) reaches 1 for a collector current of only about 800 \( \mu A \).

Considering only electrical effects, the effects of changes in \( V_{CE} \) on \( I_B \) and in \( g_{22E} \), are very small in forward-active operation; typically, \( g_{22E} = -g_{22E}/(\eta \beta_0) \) where \( \eta \geq 10 \) for narrow-base BJTs [16]; \( g_{12} \) is neglected in most circuit models. However, including self-heating and assuming \( P \ll 1/(D_B R_{TH}) \), this parameter is much larger in magnitude and positive:

\[
g_{12} \cong D_B R_{TH} I_B I_C = +K_B g_{22E}/\beta_0 \quad K_B = D_B R_{TH} I_C V_A.
\]

Using values of \( D_B \) from (10), \( g_{12} \) measurements can be used with (14) to extract \( R_{TH} \) [17].

Consider a circuit configuration in which \( R_B, R_E, \) and \( R_C \) represent the resistances seen looking out from each of the transistor terminals. The resistance looking into the collector can be shown to be

\[
R_{OUT} = \frac{1}{g_{22}} \left[ 1 + \beta_0 + \frac{R_E + (g_{12}/g_{22})(R_E + R_B)}{r_\sigma + (R_E + R_B)(1 - \beta_0 g_{12}/g_{22})} \right]
\]

where \( \beta_0 = g_m r_\sigma \) is the small-signal current gain. Now, from (14), \( 1 - \beta_0 g_{12}/g_{22} = (1 + K_\beta)/(1 + K_C) \), where \( K_\beta = K_C - K_B \). Therefore

\[
R_{OUT} \cong \frac{\tau_{OE}}{1 + K_C} \left[ 1 + \beta_0 + \frac{R_E(1 + K_C) + R_B K_B/\beta_0}{(R_E + R_B)(1 + K_\beta) + r_\sigma(1 + K_C)} \right]
\]

assuming \( \beta_0 \gg 1 \) and \( g_m r_\sigma \gg 1 \). Now, in the limit \( (R_E + R_B) \ll r_\sigma(1 + K_C)/(1 + K_\beta) \), the error in \( R_{OUT} \) due to neglecting self-heating is approximately \( 1/(1 + K_C) \), which can easily be quite significant. However, in the opposite limit of large \( R_E + R_B \), corresponding to fixed base or emitter current, the error approaches \( 1/(1 + K_\beta) \), which is generally not as significant, since usually \( K_\beta \ll K_C \). (The error due to \( K_\beta \) could be much larger in HBT's, where \( K_\beta \) is usually negative.)

By a similar analysis, it can be shown that the resistance looking into the base is given by

\[
R_{IN} = \left( \frac{R_E R_C}{R_E + R_C} \right) + r_\sigma \left( 1 + \frac{g_{12} r_\sigma R_E + K_B R_C}{r_\sigma + (R_E + R_C)(1 + K_\beta)} \right)
\]

For very small values of \( R_E \), such that \( R_E \ll K_B R_C/\left(g_m r_\sigma \right) \),

\[
R_{IN} \cong r_\sigma \left[ 1 + K_B \cdot \frac{R_C}{r_\sigma + (R_C)(1 + K_\beta)} \right]
\]

which can be significantly larger than \( r_\sigma \) if \( R_C \) is comparable to \( r_\sigma \). However, for larger values of \( R_E \) or smaller \( R_C \) the error is likely to be small.

Note that there are significant thermal effects for the important case of a common-emitter amplifier driven from a low-resistance source with a high-resistance load. For source resistance \( R_S \) and load resistance \( R_L \), the dc voltage gain is

\[
A_v = \frac{g_m r_\sigma R_L}{r_\sigma + R_S + R_L r_\sigma(1 + K_C) + R_S(1 + K_\beta)}
\]

which can be much less than predicted without considering self-heating.

To summarize, when the base–emitter junction is driven with a fixed voltage source, thermal feedback can significantly lower the collector output resistance and the voltage gain. Similarly, for fixed collector current, the relatively large positive value of \( g_{12} \) caused by self-heating can raise the resistance looking into the base if \( R_E \) is very small. If either the base or collector current is fixed, the errors caused by self-heating are smaller, and are dominated by the temperature dependence of \( \beta \). Of course, if there is large power dissipation, all the parameters can be affected.

The dependence of the common-emitter output resistance on the base driving impedance can have a significant effect on dc parameter extraction [17]. Fig. 2 shows two sets of output characteristics, one simulated with \( I_B \) as a parameter and the other with \( V_{BE} \). The parameter \( V_A \), usually extracted from the slopes of the output curves, clearly cannot be found without ambiguity unless thermal effects are accounted for. In fact, because of the extremely long thermal time constants associated with chip-to-ambient thermal impedances, differences in extracted \( V_A \) values have been seen to depend on the integration speed of the parameter analyzer, on the package type, and even on the air-flow rate around the package [10]. On the other hand, the fastest thermal time constants are comparable to the electrical ones, so it is impossible to completely avoid self-heating in most measurements [18].

By treating the variables as phasors, (13) can be converted into the frequency domain if the \( g \) parameters are replaced by the \( y \) parameters and \( R_{TH} \) is replaced by \( Z_{TH} \). This gives

\[
y_{11} = \frac{y_{11E} + D_B Z_{TH} I_B I_C^2}{1 - D_B Z_{TH} P}
\]

\[
y_{12} = \frac{y_{12E} + D_B Z_{TH} I_B I_C}{1 - D_B Z_{TH} P}
\]

\[
y_{21} = \frac{y_{21E} + D_C Z_{TH} I_B I_C}{1 - D_C Z_{TH} P}
\]

\[
y_{22} = \frac{y_{22E} + D_C Z_{TH} I_C^2}{1 - D_C Z_{TH} P}
\]
self-heating would thus significantly affect the compensation of such circuits. Although this is possible, in the most common cases there is surprisingly little effect on the phase margin.

In a typical op amp, for example, the first stage is operated at such a low current that \( K_C \ll 1 \). The second stage is operated at higher currents, and is driven from the low output impedance of an emitter follower, so there is an error in the dc gain due to self-heating. Time-constant analysis shows that if Miller compensation is applied around the second stage, the dominant pole location is also shifted by self-heating. However, above the dominant pole frequency the negative feedback through the compensation capacitor swamps out the positive thermal feedback so there is no thermal effect on the phase margin. Similarly, if a shunt compensation capacitor is placed at the output of the stage, its admittance swamps out the effect of self-heating on \( y_{22} \), and again the phase margin is not affected by self-heating.

For self-heating to affect the phase margin, the stage whose gain is distorted by thermal effects must set the key nondominant pole, which determines the required unity-gain frequency. The compensation must then be applied somewhere else in the circuit without raising the driving impedance or lowering the load impedance of the thermally affected stage. These conditions rarely arise in practice, but they can be simulated using ideal controlled sources [19].

V. LARGE-SIGNAL THERMAL EFFECTS

Most of the thermally induced errors in large-signal modeling can be understood by extrapolating from the small-signal behavior. Errors are mostly restricted to precision circuits which depend on close matching of BJT characteristics. For example, consider a simple current mirror [9]. The base of the output transistor is driven from a low-impedance source, so its output resistance is reduced by a factor \( 1 + K_C \) compared to predictions neglecting self-heating. Thus, any variation in its \( V_{CE} \) from that of the input transistor leads to a larger than expected current mismatch. A typical circuit whose performance is degraded by this effect is the self-biased
Fig. 5. (a) Self-biased $V_{BE}$-based current source. (b) Self-biased current source with current mirror replaced with Wilson mirror, which reduces self-heating effects. Both the n-p-n and p-n-p transistors were modeled using the parameters in Table I.

$V_{BE}$-based current generator [16] shown in Fig. 5(a). The current $I_{out}$ is nominally independent of $V_{CC}$. However, the finite $r_e$ values of $Q_1$ and $Q_4$ cause a sensitivity to $V_{CC}$. With $V_{CC}$ set to 10 V, the fractional sensitivity of $I_{out}$ to $V_{CC}$ is $\frac{dI_{out}}{dV_{CC}}$, which is $0.59\%$ when simulated neglecting self-heating, whereas with self-heating, the value rises to $1.40\%$.

From the discussion of small-signal effects, it is clear that techniques that raise the output resistance of current mirrors can reduce the errors due to self-heating as well. These include cascoding and negative series feedback as with Wilson current mirrors or emitter degeneration. For example, if the $Q_1, Q_2$ current mirror in Fig. 5(a) is replaced with a Wilson mirror, as shown in Fig. 5(b), $I_{out}$'s $V_{CC}$ sensitivity as simulated neglecting self-heating is $0.14\%$, whereas even with self-heating, the value only rises to $0.17\%$.

Another class of circuits subject to errors due to self-heating are the translinear circuits, so-called because they depend on the linearity of the transconductance with collector current. The corresponding exponential dependence of $I_C$ on $V_{BE}$ holds over many orders of magnitude, but it can be distorted by self-heating. The transconductance is affected by self-heating as the denominator of (13c) differs from unity, so errors in translinear circuits are related to the ratio $P/P_C$.

For example, consider the Gilbert multiplier shown in Fig. 6. The current-to-voltage converters were simulated using ideal controlled sources in SPICE. In the simulations, the gains $K_1$ and $K_2$ of the controlled sources were set so that ideally the output is given by $I_{OUT} = I_{O1} - I_{O2} = 2 \times 10^{-5}V_1 V_2$ amperes. $V_{CC}$ was set to 10 V, causing substantial power dissipation in core transistors $Q_3-Q_6$. $V_1$ was fixed at 8 V and $V_2$ was swept from 0 to 8 V. The worst-case error between the ideal and simulated values of $I_{OUT}$ was 5.6% of the full-scale output when self-heating was included, compared to 1.5% when self-heating was neglected. It is common practice to trim the gain constants of multipliers to provide a first-order correction to reduce scale-factor, feedthrough, and gain errors. In the simulations $K_1$ was varied to minimize worst-case errors. After trimming, the simulated error neglecting self-heating was below 0.1%. However, simulations including self-heating showed that trimming could only reduce the error to about 3.2%. The error can be reduced but not eliminated if the core transistors are operated with lower $V_{CE}$. With $V_{CC}$ set to zero, there was still a predicted post-trim error including self-heating of about 1.3%. These results prove that self-heating can be a dominant source of errors in such circuits.

Another class of circuits whose performance is complicated by self-heating are bandgap voltage references, such as that shown as an inset in Fig. 7. The circuit was designed so that the current density is $Q_2$ is 100 times that of $Q_1$. For $V_{CC} = 10$ V, $Q_2$'s power dissipation is about 10% of $P_C$. Assuming $Q_1$ and $Q_2$ operate at the same temperature, the output voltage is

$$V_O = V_{BE1} + \frac{R_2}{R_1} \cdot (V_{BE2} - V_{BE3})$$

$$= V_{BE1} + \frac{R_2}{R_1} \cdot V_1 \cdot \ln(100).$$

(21)
With care, the temperature coefficients of the first and second terms can be made to cancel at a chosen temperature $T_0$, giving zero $TC_F(V_0)$ at that temperature. This circuit was designed neglecting self-heating to give zero $TC_F(V_0)$ at $T_0 = 25^\circ C$, which requires $V_0 = 1.1552$ V. Fig. 7 shows $V_0$ versus $\Delta T$ as predicted by unmodified SPICE and by the version modified to include self-heating, which causes about a 10-mV change in $V_0$ at $25^\circ C$, and a shift in $V_0$ to give $V_0 = 1.1552$ V still leaves an error in $TC_F(V_0)$ caused by self-heating.

VI. SELF-HEATING IN TRANSIENT OPERATION

A thermal subcircuit was used to simulate the transient responses of a variety of circuits using PSpice [20]. PSpice's behavioral modeling allows the inverse Fourier transform of the thermal impedance (2) to be used in transient simulation. The instantaneous power, temperature rise above ambient, and thermal corrections to $V_{BE}$ and $I_B$ were computed using controlled sources. High- and low-current effects were neglected for simplicity.

The results show that translinear circuits can have long tails in their settling times, due to the slowness of thermal step responses. For example, when a fast 8- to 0-V step was applied to $V_1$ in the multiplier of Fig. 6, with $V_0$ held at 8 V and $V_{CC}$ at 5 V, the circuit simulated without self-heating settled to $I_O = 0$ in less than 1 ns. With self-heating, 95% of the drop in $I_O$ occurred within 10 ns, but settling to within 2% took almost 10 $\mu$s. The reason can be seen from Fig. 8, which shows a typical thermal unit-step response, as predicted by the inverse transform of (2) and by the more detailed model of [4]. Note the slow settling response of the temperature and the wide range of time constants.

VII. CONCLUSION

Self-heating can significantly affect BJT circuit behavior, especially in analog circuits. The effects are often subtle. In low-power circuits, self-heating effects are reduced by resistances in the base or emitter or through negative feedback, and they drop at very high frequencies. Thermally induced errors in large-signal modeling generally require some power dissipation and mostly show up in precision circuits. However, as thermal impedances rise due to scaling and the use of thermally insulating dielectrics, it will become more and more important for designers to understand the effects of self-heating on circuit behavior. This paper has presented the basis for such understanding.

ACKNOWLEDGMENT

The authors thank Dr. T. Jung of Harris Corporation for pointing out examples of circuits exhibiting thermally induced settling tails.
REFERENCES


Robert M. Fox (S’78-M’80-SM’92) was born in Birmingham, AL, on November 12, 1950. He received the B.S. degree in physics from the University of Notre Dame, Notre Dame, IN, in 1972 and the M.S. and Ph.D. degrees in electrical engineering from Auburn University, Auburn, AL, in 1981 and 1986, respectively.

He has been on the Electrical Engineering faculty of the University of Florida, Gainesville, since 1986. His current rank is Associate Professor. His research interests center on circuit design for advanced silicon technologies. Dr. Fox is a member of the Audio Engineering Society, ASEE, Phi Kappa Phi, and Eta Kappa Nu.

Sang-Gug Lee was born in KyungNam, Korea, in 1958. He received the B.S. degree in electronic engineering from KyungPook National University, Taegu, Korea, in 1981, and the M.E. and Ph.D. degrees in electrical engineering, both from the University of Florida, Gainesville, in 1989 and 1992, respectively.

From 1983 to 1986 he was an Instructor in the Department of Electronic Engineering at the Naval Academy, JinHae, Korea. Since 1992 he has been with Harris Semiconductor Company, Melbourne, FL, working on analog IC design. His current research interests are focused on device modeling, characterization, and RF circuit design.

David T. Zweidinger was born in Gainesville, FL, on June 22, 1967. He received the B.S.E.E. degree in 1990 and is currently working toward the M.S.E.E. and Ph.D. degrees at the University of Florida, Gainesville. His interests include electronic circuits, semiconductor device modeling, and computer-aided design.