Experimental assessment of self-heating in SOI FinFETs

NXP-TSMC Research Centre, High Tech Campus 37, 5656 AE Eindhoven, The Netherlands
†MESA+ Institute for Nanotechnology, University of Twente, Enschede, The Netherlands
‡IMEC, Leuven, Belgium, *ARCES-DEIS University of Bologna and IUNET, Cesena, Italy

Abstract

In this paper, it is shown that self-heating causes a gigantic effect on the capacitances of MOSFETs/FinFETs. The effect is used to determine the SOI FinFET thermal impedance and to determine the temperature rise during FinFET operation.

Introduction

FinFETs are regarded as prospective replacements for bulk-CMOS devices beyond the 22-nm node. Due to the poor thermal conductivity of some of the materials used in FinFET fabrication, and due to the confined nature of the FinFET geometry, self-heating is expected to be more significant in FinFETs than in their bulk-CMOS counterparts, affecting not only device performance, but also NBTI and oxide reliability [1, 2]. To the best of the authors’ knowledge, all studies on FinFET self-heating so far are based on simulations [1, 3, 4], and experimental assessment is lacking. In this paper, we will (i) show that self-heating causes a gigantic effect on the capacitances $C_{\text{DD}}$ and $C_{\text{DG}}$, (ii) use that effect as an ultra-sensitive method to extract the frequency-dependent complex thermal impedance $Z_{\text{th}}(f)$, and (iii) determine the temperature rise in our SOI FinFETs and its impact on the device performance.

Pulsed measurements

Pulsed $IV$ curves [5], with 100 ns pulse width, were measured on SOI FinFETs (described in [6]) and compared to DC $IV$ curves. For channel lengths ranging from 70 nm to 1 μm, and for a number of parallel fins varying from 1 to 168, no significant difference between the two could be detected. Some examples are shown in Fig. 1. From these measurements we may conclude either that the impact of self-heating on FinFET current is very small, or that the thermal time constant is smaller than 100 ns, or a combination of both.

$S$-parameter measurements

An alternative method [7, 8] to extract the thermal impedance relies on the fact that the device temperature can respond to “slow” (w.r.t. the thermal time constant) AC signals, but is unable to follow “fast” excitations, leading to a difference in the low- and high-frequency values of $g_{\text{DS}}$. Therefore, we conducted $S$-parameter measurements in the 100 kHz – 10 GHz range using an Agilent E5071C network analyzer. The measurements were carried out on RF FinFETs in GSG configuration, de-embedded, converted to $Y$-parameters, and subsequently converted into the output conductance $g_{\text{DS}}(f) = \text{Re}(Y_{\text{DD}})$, transconductance $g_m(f) = \text{Re}(Y_{\text{DG}})$, drain-drain capacitance $C_{\text{DD}}(f) = \text{Im}(Y_{\text{DD}})/(2\pi f)$, and gate-to-drain capacitance $C_{\text{DG}}(f) = -\text{Im}(Y_{\text{DG}})/(2\pi f)$.

In Fig. 2, it is observed that the expected rise of $g_{\text{DS}}$ with frequency is largest for the long channel, and levels off in the GHz range. For the short channel, on the other hand, the effect is smaller and is difficult to distinguish from parasitic-resistance effects, which also cause $g_{\text{DS}}$ to rise in the GHz range (Fig. 2). Both effects make the extraction of the thermal impedance from $g_{\text{DS}}(f)$ less reliable. A similar argument applies to the use of $g_m$, which, on top of that, is less sensitive to self-heating than $g_{\text{DS}}$; see Fig. 3. Strikingly, if we inspect the capacitances $C_{\text{DD}}$ and $C_{\text{DG}}$, a much larger effect is seen; see Figs. 4 and 5. At sub-GHz frequencies, the capacitance $C_{\text{DD}}$ takes values that are several orders of magnitude in excess of the high-frequency and the oxide capacitance. The transcapacitance $C_{\text{DG}}$ shows a similar effect, but here also the sign of the measured capacitance changes. The effects on $C_{\text{DD}}$ and $C_{\text{DG}}$ are due to the delay in the heat transport from device to heat sink, and are similar to the self-heating effect on bipolar-transistor capacitances, which has been predicted theoretically in [9]. A mathematical derivation for the MOSFET/FinFET case leads to

\[ \Delta Y_{\text{DD}} = Z_{\text{th}} \frac{dI_{\text{D}}}{dI_{\text{amb}}} \cdot (I_{\text{D}} + V_{\text{DS}} \cdot Y_{\text{iso}}^{\text{DD}}) \]

(1)

\[ \Delta Y_{\text{DG}} = Z_{\text{th}} \frac{dI_{\text{D}}}{dI_{\text{amb}}} \cdot V_{\text{DS}} \cdot Y_{\text{iso}}^{\text{DG}} \]

(2)

where $\Delta Y_{ij}$ is the difference between the actual $Y$-parameters and the isothermal counterparts $Y_{ij}^{\text{iso}}$ (reached at high frequencies). Moreover, $Z_{\text{th}}$ is the (complex) thermal impedance, and $T_{\text{amb}}$ the ambient (not lattice) temperature.

Extraction of thermal network

The gigantic effect of self-heating on $C_{\text{DD}}$ and $C_{\text{DG}}$ can be employed to extract $Z_{\text{th}}(f)$ from measured data only. This is done by inverting Eqs. (1) and (2), leading to

\[ \frac{\text{Im}(Z_{\text{th}})}{2\pi f} \approx \frac{C_{\text{DD}} - C_{\text{DD}}^{\text{iso}}}{\frac{dI_{\text{D}}}{dI_{\text{amb}}} \cdot (I_{\text{D}} + V_{\text{DS}} \cdot \text{Re}(Y_{\text{DD}}))} \]

(3)

\[ \frac{\text{Im}(Z_{\text{th}})}{2\pi f} \approx \frac{C_{\text{DG}} - C_{\text{DG}}^{\text{iso}}}{\frac{dI_{\text{D}}}{dI_{\text{amb}}} \cdot V_{\text{DS}} \cdot \text{Re}(Y_{\text{DG}})} \]

(4)
The values of the isothermal capacitances $C_{DG}^{iso}$ and $C_{GD}^{iso}$ are determined from the plateau observed at GHz frequencies, but their exact value is actually rather inconsequential. Apart from measuring the capacitances, one only needs to measure the dependence of the drain current on the ambient temperature. In Figs. 6 and 7 it is shown that extraction of $\text{Im}[Z_{th}(f)]$ using either Eqs. (3) or (4), and using different bias conditions, yields essentially the same result, confirming the consistency of the method.

Next, we fit the imaginary part of the impedance of an $n$th-order thermal network to the extracted data. In our case $n = 4$ appeared to be sufficient; see Fig. 8. The result of this fit is shown as the solid lines in Figs. 6 and 7. Having determined $\text{Im}(Z_{th})$ versus frequency, all components of the thermal network, and thus also the thermal resistance $\text{Re}(Z_{th})$, are known. The result is shown in Fig. 9. As expected, $\text{Re}(Z_{th})$ decreases with increasing channel length.

Results

We use the extracted thermal network, connected to a FinFET compact model [10], to simulate the FinFET capacitances and conductances. This yields the solid lines in Figs. 2–5. The agreement is very good, confirming the consistency of our results. Next, we use the model to re-simulate the pulsed experiment at the bias condition where the self-heating is greatest; see Fig. 10. It is seen that at the end of the pulse –where the pulsed measurements are taken– the difference between DC and pulsed measurements is far below the 1% level, which explains why our pulsed experiments (Fig. 1) failed to reveal any self-heating effects. Further analysis of the results shows that this is mainly due to the fact that the FinFET thermal time constants are predominantly in the sub–100-ns regime, i.e., faster than our pulses. The total effect of the self-heating on the FinFET IV curves is shown in Fig. 11, which was obtained by switching self-heating on and off in the model. At the highest bias condition and for the shortest channel, the total effect amounts to a fairly modest 4%. Note that this is also the maximum effect that we would have observed in a –hypothetical– pulsed experiment with $\delta$-pulses. Combining this with the experimental difficulties associated with pulsed experiments, it is clear that our capacitance-based method is much more suitable for studying self-heating in FinFETs.

Although the effect of self-heating on the on-current is small, the temperature rise calculated from the experimental data using $\Delta T = I_D V_{DS} \Sigma R_{th,i}$ is quite significant, and amounts to ~80 K at the highest bias; see Fig. 12. This temperature rise is of the same order of magnitude as that obtained from electrothermal simulations [1, 3, 4], and significantly larger than that of a 40-nm bulk MOSFET (see Fig. 12). Of course, for more aggressive channel lengths and fin spacings than those presented here, $R_{th}$ and thus the temperature rise will be larger, and may pose a concern to, e.g., reliability [2] and thermal noise.

Electro-thermal simulations

In order to gain physical understanding of our experimental results, 3D electro-thermal drift-diffusion device simulations have been performed using Sentaurus [11] and calibrated mobility and thermal conductivity parameters [4]. The simulated SOI FinFET structure includes level-1 interconnects to the drain, source and gate. In Fig. 13, the results of these simulations are compared with the experimental capacitance data for our 70-nm FinFET. The agreement is very encouraging. The simulations point out that the dominant path for the heat-flow includes the contacts and the interconnects which, in turn, exchange heat with the silicon bulk through the underlying buried oxide. Thus, the thermal modeling of the interconnects is essential for a good understanding of the SOI FinFET thermal behavior.

Conclusion

We have demonstrated a gigantic effect of self-heating on capacitances and used it to extract the thermal impedance of SOI FinFETs experimentally for the first time. Using the extracted thermal network, we have shown that, although the effect of self-heating on the FinFET on-current is modest, the associated temperature rise is very significant. First electro-thermal simulations compare favorably with the experimental data. The proposed extraction method can be used directly on measured data, without the need for a compact model or simulation software, and is therefore a valuable tool for FinFET technology development.

Acknowledgment

This work was partly sponsored by the European project Nano-RF (IST-027150). The authors would like to thank J.A. van Steenwijk and G. Curatola (NXP) for valuable discussions and support.

References

Figure 1: Comparison between pulsed measurements (markers) and non-pulsed DC measurements (lines) for several geometries. Left: device consisting of 5 parallel fins with $L_{\text{poly}} = 70$ nm. Middle: same, but now for 30 parallel fins. Right: 5 parallel fins with $L_{\text{poly}} = 190$ nm.

Figure 2: Relative change in output conductance $g_{DS}$ for a 1-$\mu$m and a 70-nm FinFET biased at $V_{GS} = V_{DS} = 1.2$ V. Markers: measurements; solid lines: compact model with thermal network.

Figure 3: Relative change in transconductance $g_m$ for a 1-$\mu$m and a 70-nm FinFET biased at $V_{GS} = V_{DS} = 1.2$ V. Markers: measurements; solid lines: compact model with thermal network.

Figure 4: Drain-drain capacitance $C_{DD}$ for a 1-$\mu$m and a 70-nm FinFET biased at $V_{GS} = V_{DS} = 1.2$ V. Markers: measurements; solid lines: compact model with thermal network; dashed lines: isothermal capacitances used in extraction.

Figure 5: Gate-to-drain capacitance $C_{DG}$ for a 1-$\mu$m and a 70-nm FinFET biased at $V_{GS} = V_{DS} = 1.2$ V. Markers: measurements; solid lines: compact model with thermal network; dashed lines: isothermal capacitances used in extraction.

Figure 6: Extraction of $\text{Im}(Z_{th})/(2\pi f)$ from $C_{DG}$ (o) and from $C_{DD}$ (●), yielding virtually the same result, for two channel lengths. The bias is $V_{DS} = V_{GS} = 1.2$ V. Solid line: fit of $\text{Im}(Z_{th})/(2\pi f)$ of the thermal network of Fig. 8 to the extracted data.
Figure 7: Extraction of $\text{Im}(Z_{th})/(2\pi f)$ for different bias conditions (indicated in the figure), yielding virtually the same result. Left: extraction from $C_{DG}$; right: extraction from $C_{DD}$. Solid lines: same as in Fig. 6.

Figure 9: Extracted $\text{Re}(Z_{th})$ versus frequency for three different channel lengths.

Figure 10: Re-simulation of pulsed experiment of a 70-nm SOI FinFET biased at $V_{GS} = V_{DS} = 1.2$ V, using extracted $Z_{th}(f)$. Note that in the actual pulsed experiments, the first part of the signal is not usable because it is affected by unwanted inductive and capacitive effects.

Figure 11: Assessment of the effect of self-heating on FinFET current ($L_{poly} = 70$ nm) using a compact model and the thermal network depicted in Fig. 8. Solid/dashed lines are with/without self-heating, respectively. Different colors refer to different gate-source voltages, as indicated in the figure.

Figure 12: Temperature rise in the 70-nm FinFET (left), and in a 40-nm multi-fingered bulk MOSFET (right) as calculated from experimental data using $\Delta T = I_D V_{DS} \Sigma R_{th,i}$. Although, at the same drain-source voltage, the on-current of the MOSFET (32 mA) is larger than that of the FinFET (20 mA), the associated temperature rise is significantly smaller.

Figure 13: Comparison of capacitance data (markers) and electro-thermal simulations (lines) for the 70-nm FinFET.