Measurements of Inter-and-Intra Device Transient Thermal Transport on SOI FETs

P. M. Solomon, M. Shamsa*, K. A. Jenkins, C. P. D'Emic, A. A. Balandin* and W. Haensch
IBM, SRDC, T.J. Watson Research. Center, Yorktown Heights, NY 10598, e-mail: solomonp@us.ibm.com
*Nano-Device Lab., Dept Elect. Eng., University of California. Riverside, CA 92521.
Tel: (914) 945-2841, Fax: (914) 945-2141, Email: solomonp@us.ibm.com

Introduction:

With the continued downscaling of CMOS with ever increasing power densities especially in the local on-chip environment, and with the use of silicon-on-insulator (SOI) technology, the measurements related to local heating of single devices and the transfer of heat between devices are becoming increasingly important[1]. Furthermore it is important to be able to characterize the full thermal transient, since the chip thermal activity fluctuations tend to be on the same time scales as typical thermal time constants resulting in peak transient temperatures which may well exceed steady-state average temperatures. There have been few attempts at transient thermal characterization of CMOS devices with recent notable advances[2] but the work we report below is *the first resolving detailed thermal transients for CMOS devices*. Furthermore we investigate different heat paths between and inside devices to reveal the importance of the thermal conductivity of the gate. This work is extended to studying thermal transport within a sub-micrometer CMOS FET where we are able to detect the delayed heat pulse at the source due to heat generation in the drain.

Table 1: Thermal transport media

Structure		L_M (nm)
I	(shallow trench isolation)	200, 300, 500, 700
RX[a]	(doped and silicided SOI)	300, 500, 1100
OP[a]	(dope d, unsilicided SOI)	0, 300, 500, 700, 1100
PC[a]	(floating polysilicon gate)	200, 500, 1100

[a] These designations refer to names of mask levels.

Experimental Approach

This work describes our experimental approach with the simulation technique reported elsewhere[3]. The layout of our SOI CMOS devices is shown in Fig. 1.Two 10 µm wide FETs are placed back-to-back with a medium separating them. The different media and their lengths (L_M) are given in Table 1. The FETs are connected to microwave probes with the gates grounded. This minimizes electrical crosstalk.

The DC temperature sensitivity of the 250 nm FET is shown in Fig. 2. Note that the sensitivities are opposite for below threshold, injection-limited current compared to the above threshold, mobility limited, regime. (The sub-threshold sensitivity is much smaller for the longer, 600nm, gate length[3] transistor.) The sub-threshold sensing has the advantage in that it is sensitive to the local temperature at the source junction rather than an average over the whole device length.

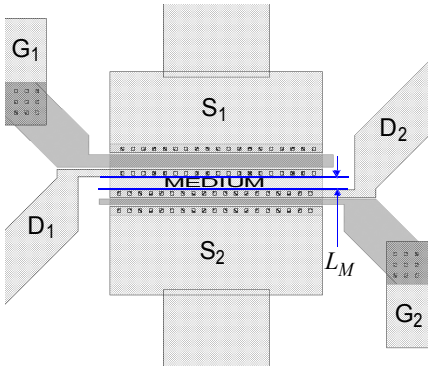

Fig. 1. Device layout where subscripts 1 & 2 indicate the preferred terminals of the 600nm and 250nm gate length FETs respectively. The medium, and L_M are given in Table I.

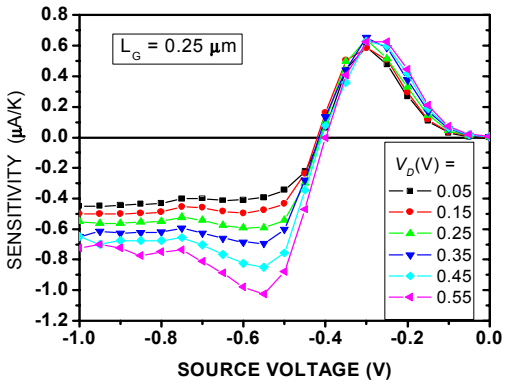

Fig. 2. DC temperature sensitivity of a 250 nm x10 µm n-FET in common-gate mode.

The pulse-probe measurement circuit is shown in Fig. 3. Negative going power and sense pulses are applied to the source of power and sense transistors respectively, with a varying delay between them. The amplitude of the pulses is adjusted to give a DC current corresponding to the positive peak of the sensitivity curve. Typical transients are shown in Fig. 4a & b. Note that in Fig. 4b a large transient is caused by the power pulse being diverted to the sense FET through conducting bridge between the FETs when silicide or n+ silicon media are used. We have implemented a lock-in type sampling scheme by alternately applying and

suppressing the power pulses at a 1Hz rate and averaging the difference of the sense pulses with and without the power pulses. The technique was verified by observing positive and negative-going current transients which were observed when the source pulse amplitude was varied from below to above threshold.

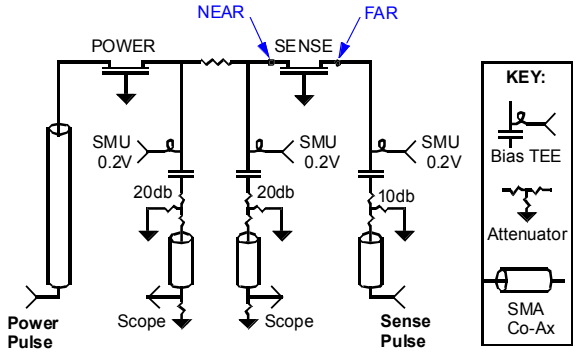

Fig. 3. High-speed Pulse-Probe circuit showing the case where the sense pulse is applied to the FAR FET S/D contact.

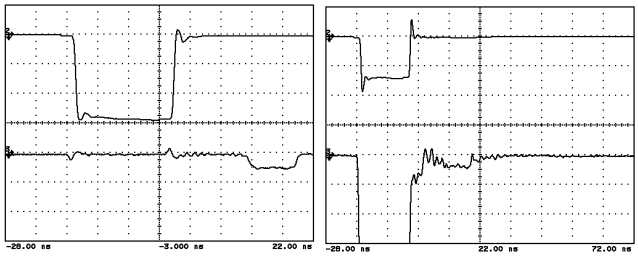

Fig. 4. Transient current in power transistor (top) and sense transistor (bottom), showing cases for (a) an insulating and (b) a conducting connection between the two transistors. Horizontal scales are 5 ns/div and 10 ns/div respectively. The power pulse was 16ns and the sense pulse 8ns.

Table 2: Thermal conductivities

Material	Thermal Cond. (W/m-K)
copper	400
bulk silicon	160
polysilicon	125
45 nm thick SOI	45
silicide	40
silicon dioxide	1.024

Simulations:

Simulation of the thermal transients in the SOI device structures were carried out by solving the 2D heat-diffusion equation using the finite element method (FEM) implemented in FEMAB software package, as described in Ref. 3. Spreading of heat into the contacts (see Fig. 1) is not considered. Thermal conductivities used are given in Table 2.

Results: Two-Transistor Measurements:

Thermal transients obtained after 10 hrs of averaging in four separate sweeps, are shown in Fig. 5a for a power pulse of 16 ns and a sense pulse of 8 ns duration. This figure illustrates that the sub-threshold sensing mode detects heat arriving at the source. In (b) the source is placed at the FAR end of the sense transistor so that heat has to transverse an extra 600 nm of transistor channel length, so the signal is attenuated and delayed with respect to NEAR end sensing (a). This is shown (b) to be equivalent to an external 0.6 μm floating gate (PC). Results for different PC lengths are shown in Fig. 9 where we also compare measurements to simulations. The agreement is remarkably good considering the simplifying assumptions in the simulations. Results for the different media (see Table 2) are shown in Fig. 7 where the peak temperature vs. length for the medium is plotted. In the PC case an extra 100 nm was added to account for mask differences. Note that the PC case has the largest thermal transmission, as expected, but note also that there appears to be some thermal interface resistance

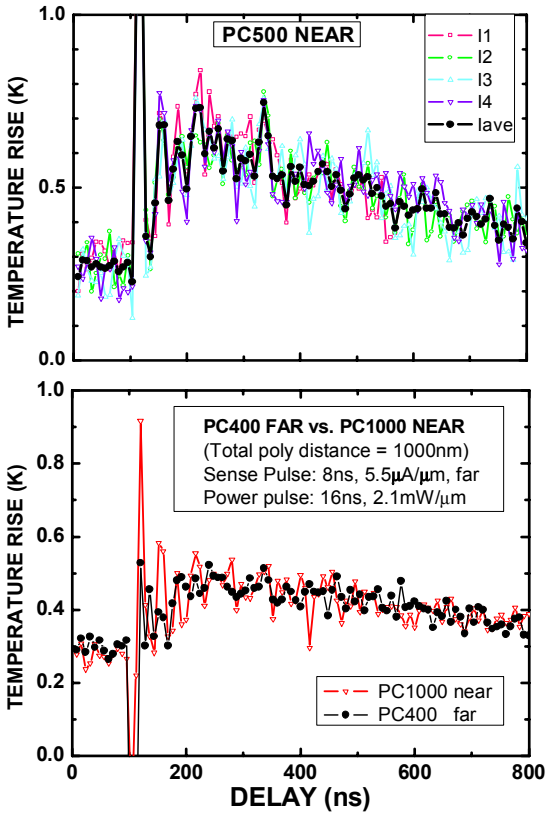

Fig. 5. Transient sense pulse signal (after overnight averaging), converted to temperature rise, for FETs separated by a floating gates (PC). (a) The sense pulse is applied to the NEAR S/D contact with a 500 nm PC medium and (b) to the FAR contact with 500 nm PC compared to the NEAR contact with an 1100 nm floating gate.

for transfer of heat to the floating gate, which is not captured by the simulations, since the temperature falls off so rapidly from the zero-gap case.

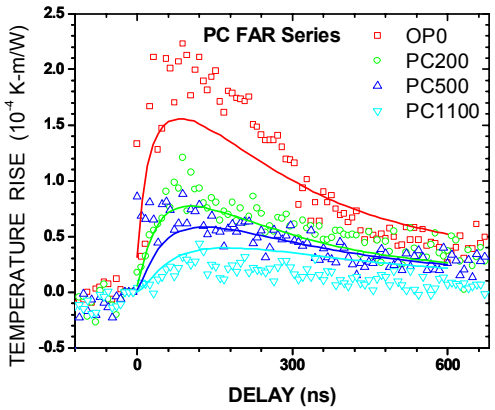

Fig. 6. Temperature transients for no floating gate (OP0) and for PC lengths of 200, 500 and 1100 nm. Solid lines (in order of decreasing amplitude) are simulated results for the same cases.

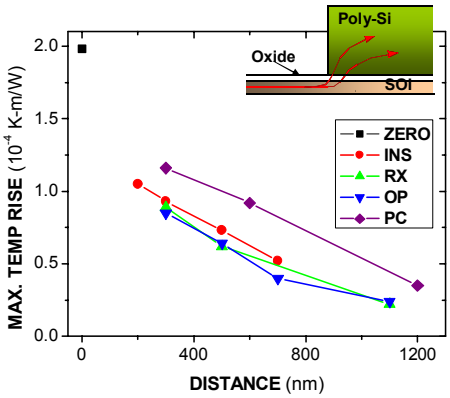

Fig. 7. Max. Temp. vs. distance for the I, RX, OP and PC media. Insert illustrates thermal transfer resistance of the PC case.

Single-Transistor Measurements:

Measurements were also done on a single transistor using the circuit shown in Fig. 8. A fast Schottky diode suppresses the positive going overshoot of the negative going power pulse allowing the application of a sense pulse, shortly afterward, to the same terminal, with minimal interference. Transients are shown in Fig. 9 for the two configurations of the circuit. Note that for sensing at the power drain (Fig. 9b) the sense pulse has to be large enough to overcome the DC drain bias and momentarily reverse the bias on that terminal. Both power and sense currents are averaged, but the sense pulse is switched on-and-off and the two average currents subtracted, allowing an accurate measure of the sense current even though the power current is orders of magnitude larger. To prevent the heat pulse from traversing the transistor during the power pulse a short (2 ns) pulse is required Results in Fig. 11 clearly show that the thermal transient is obtained only after a delay when measuring at the source of the power transistor (a) but is seen decaying immediately after the power pulse when measuring temperature on the drain (b). This interpretation is supported by measurements on the shorter, 250 nm transistor (c) where the maximum of the temperature response is much closer to the end of the power pulse. (The spike at ~20 ns delay seen in all of these figures is an artifact).

The results were supported by simulations. Fig. 11 shows that heat generated at the drain does not have time to arrive at the source during a 2ns heat pulse. The magnitude of the temperature rise predicted by simulations agrees with the experimental values.

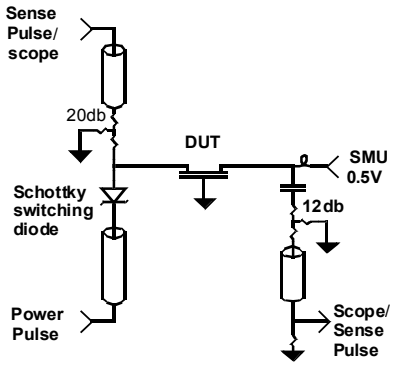

Fig. 8. Measurement circuit for a single transistor showing use of Schottky diode. The sense pulse may be applied at the same or opposite sides of the transistor to the (negative going) power pulse.

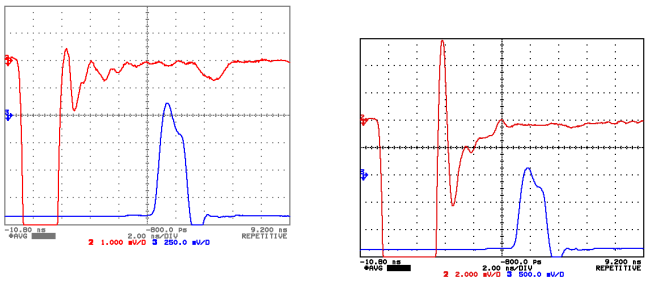

Figure 9. (a) Single transistor transients at 2 ns/div, show sense pulse (bottom) measured on the complement channel of the pulse generator, and source current (top). The current scales are magnified, clipping the power pulse, to show the sense pulse. (a) Sense pulse (250 mV/div) applied at the power source, current scale is 80μA/div. (b) Sense pulse (500 mV/div) applied at the power drain, current scale is 400 μA/div.

Conclusions

Our results clearly demonstrate capability of measuring thermal transients in separated devices on the ns time-scale and the ability to sense local source temperature, both in separated and in a single FET. We show both by measurements and simulations that oxide does not afford good isolation and that the main cooling mechanism of SOI devices is to the gate, with transfer resistance playing an important role.

Acknowledgements

M.S. acknowledges IBM PhD Fellowship Award, which allowed him to carry out measurements at the IBM T.J. Watson Research Center. The work at UCR was supported through the SRC/FCRP center on Functional Engineered Nano Architectonics (FENA), contract no. 2003-NT-1107. We acknowledge the support of the MRS process facility and to Xiaoyan Shao for electroless plating.

References

1. B. M. Tenbroek, M. S. L. Lee, W. R-White, R. J. Bunyan and M. J. Uren, IEEE J. Sol. State Circuits, vol. 33, p.1037, 1998.
2. M. B. Ketchen, K. Xiu* and M. Bhushan submitted to 2007 SOI conference.
3. M. Shamsa, P.M. Solomon, K. A. Jenkins, C. A. D'emic, A. A. Balandin* and W. Haensch, manuscript submitted for publication.

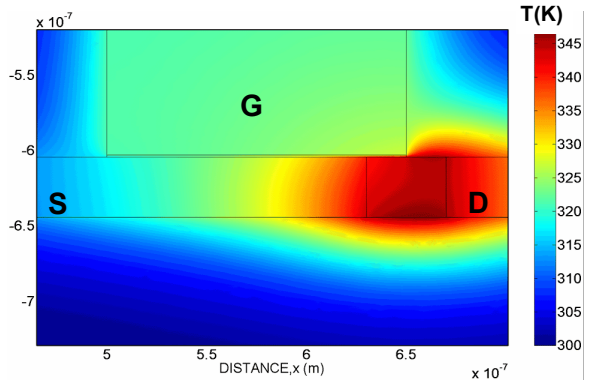

Fig. 11. Simulation of temperature distribution in a 600nm gate length FET at the end of a power pulse of 2nS duration..

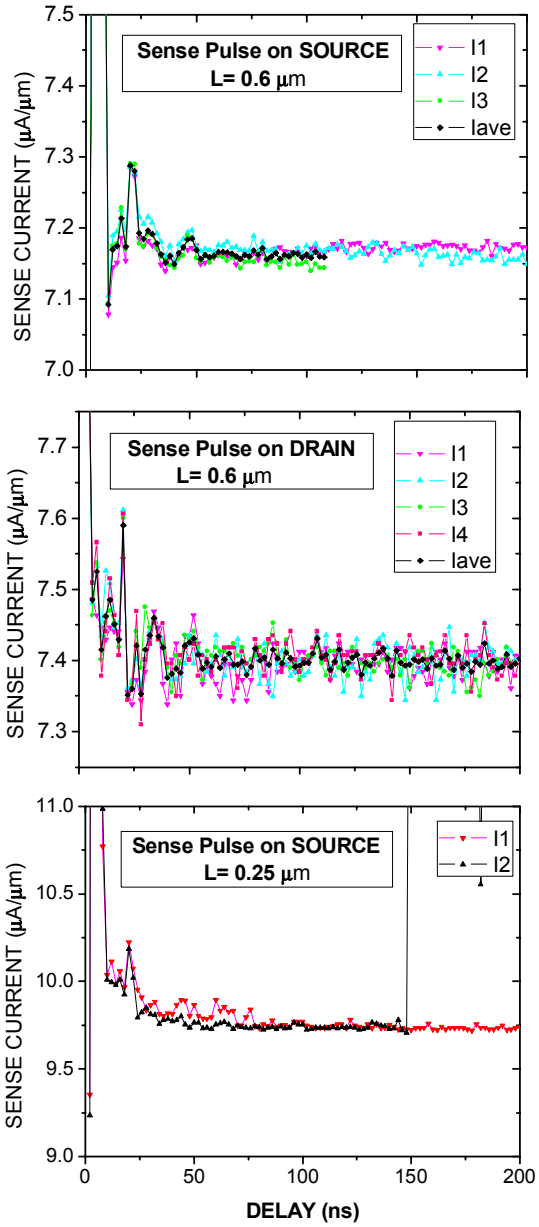

Fig. 10. Current transients after application of a 2 ns power pulse (a) and (b) for a 0.6µm gate length transistor with sense pulse applied to the power source and drain respectively and (c) for a 0.25 µm gate length transistor with sense pulse applied to the power source. Note that the spike at ~20 ns seen in (a) and (c) is spurious, caused by a reflection.