

# Investigation on the thermal distribution of $n$ MOSFETs under different operation modes by scanning thermal microscopy

E. Hendarto<sup>1</sup>, A. Altes<sup>2</sup>, R. Heiderhoff<sup>2</sup>, J.C.H. Phang<sup>1</sup>, L.J. Balk<sup>2</sup>

<sup>1</sup>Centre for Integrated Circuit Failure Analysis and Reliability (CICFAR)  
Faculty of Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576  
Tel: 65-6874-2244; Fax: 65-6874-7912; E-mail: eng10627@nus.edu.sg

<sup>2</sup>Bergische Universität Wuppertal  
Fachbereich Elektrotechnik, Informationstechnik, Medientechnik, Lehrstuhl für Elektronik  
Rainer-Gruenter-Str. 21, D-42119 Wuppertal, Germany

## ABSTRACT

A Wollaston resistive probe based scanning thermal microscope (S<sub>Th</sub>M) is used to investigate the thermal distribution within the gate channel of an  $n$ MOSFET with width to length ratio ( $W/L$ ) of  $2.0\mu\text{m}/3.0\mu\text{m}$ . The dynamic mode of S<sub>Th</sub>M is used to detect and characterize thermal wave propagation in  $n$ MOSFET biased at different operation modes, *i.e.* linear, near saturation, far saturation, and near breakdown. An ultra compliant polyimide thermal probe is later employed and the resulting 2D thermal distribution is compared with the results obtained using the Wollaston resistive probe. Measurements using the two probes reveal that the highest thermal wave amplitude region in the gate is closest to the drain at pinch-off, and slowly spreads out and extends towards the source as the MOSFET approaches the near breakdown operation mode.

## INTRODUCTION

As electronic devices become smaller and smaller into the micrometer and nanometer range, it results in increased performance and decreased cost per transistor [1]. However, as transistor gets smaller, more power and heat dissipation issues also develop. High power densities dissipated in smaller and faster devices cause major thermal problems in semiconductor devices. The local heat dissipation in such devices induces harmful effects such as malfunctions and faster degradation. Thermal study and characterization of devices with small feature sizes are therefore crucial to investigate and understand device degradation and breakdown mechanism. The scanning thermal microscope (S<sub>Th</sub>M) has demonstrated to be a suitable and promising tool to investigate devices in the micrometer and nanometer regime. By incorporating a Wollaston resistive thermal probe in S<sub>Th</sub>M, a spatial resolution of  $30\text{nm}$  [2] and temperature sensitivity of  $5\text{mK}$  have been achieved [3].

Today, MOS integrated circuits (ICs) are the dominant technology in the semiconductor industry, and it will remain as the most important technology for the coming decade. Knowledge of heat generation in  $n$ MOSFETs thus becomes essential in understanding the physical phenomena of device failures. During the last three decades, the basic MOSFET size has shrunk from a feature size of  $20\mu\text{m}$  to less than a micron [1]. With decreasing gate channel length to much less than a micron, higher vertical electric fields  $\bar{E}$  across the gate oxide as shown in figure 1 are experienced by the carriers, electrons in the case of  $n$ MOSFET and holes in  $p$ MOSFET, in the channel region near the drain end due to constant voltage scaling. This results in increased channel hot-carrier effects, which is deleterious to the transistor, leading to faster device failure. Many studies have been carried out to understand the mechanism of transistor failure in the past several years. Hot carrier induced degradation, effects of punchthrough, short channel effects and the

effects of gate-drain overlap and the resulting parasitic capacitances have been extensively studied, and several models have been developed to explain the failure mechanisms [4].

The purpose of this paper is to measure the thermal distribution in an  $n$ MOSFET using S<sub>Th</sub>M, and examine how the behaviour of the high thermal wave amplitude region changes with increasing drain voltages at a fixed gate voltage higher than the threshold voltage. The shift in the high thermal wave amplitude region in MOSFET is expected due to the shift of the channel pinch-off region  $X$  of the inversion layer under the gate contact as shown in figure 1. When a MOSFET is turned on by applying a gate voltage higher than its threshold voltage, increasing the drain voltage causes the inversion layer channel below the gate to first pinch off at the drain and then shift slowly towards the source as the drain voltage is increased further. A depletion region exists between the pinch-off point and the drain. As the lateral electric field continues to accelerate the electrons and injects them into the depletion region, these accelerated electrons cause the channel closer to the drain to be heated up as the electrons travel through a region of high resistance through the depletion region. Hot-carrier effects also play a significant role in heating up the device. With shorter channel length, the punchthrough effect can play a part in heating up the gate channel as well. Thermal analysis by S<sub>Th</sub>M is performed on the  $n$ MOSFET to correlate such electronic behaviour in  $n$ MOSFET.

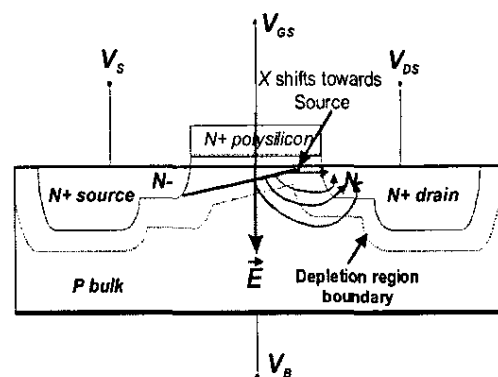


Figure 1. State of the inversion channel and pinch-off at saturation with directions of electron flow from inversion channel to the drain in the bulk, and the vertical electric field

## EXPERIMENTAL SET-UP

The dynamic temperature measurement mode of the S<sub>Th</sub>M is employed in the investigation of the  $n$ MOSFET. This mode has been described by Altes *et al* [5], and the schematic set-up is shown in figure 2.

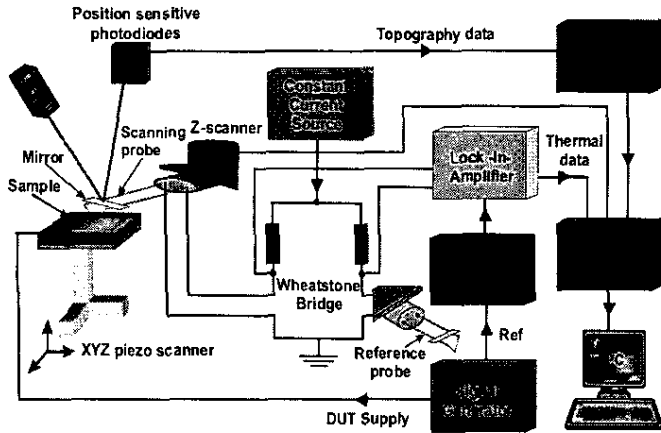


Figure 2. Set-up of SThM for topography and dynamic temperature measurement

For probe-sample distance control, a common scheme consisting of a laser and a four-cell position sensitive photodetector is used. A mirror fixed to the thermal probe reflects the laser beam that falls on it to the photodetector which gives topography information of the device under test (DUT). In the dynamic temperature measurement technique, the probe is supplied with a small current from a constant current source to avoid self-heating of the tip, while the DUT is biased with a periodic *ac* voltage from a signal generator. This bias arrangement generates thermal waves within the DUT which is detected by the tip. Since the current  $I$  passing through the MOSFET has a frequency of  $\omega$ , the resulting thermal waves generated in the DUT have a frequency of  $2\omega$ , because the relationship between temperature  $T$ , power  $P$  and current  $I$  is  $T \propto P \propto I^2$ . The Wheatstone bridge measures the resulting harmonic resistance variation of the thermal probe at  $2\omega$ . A frequency multiplier is used to multiply the modulating frequency  $\omega$  to  $2\omega$  to be used as the reference signal of the lock-in-amplifier (LIA).

Two kinds of thermal probes have been used. The first thermal probe is the commercial Wollaston resistive probe shown in figure 3. In the dynamic SThM, the resistive probe acts like a sample thermometer. The probe is made of a wire with a thin platinum core with a diameter of  $5\mu\text{m}$  surrounded by a thick silver cladding of  $75\mu\text{m}$  diameter. The end of the loop is etched, uncovering a small length of platinum of approximately  $200\mu\text{m}$ . This is the thermal element which determines the resistance of the whole sensor. Information about the local temperature is obtained by monitoring the probe resistance as it changes due to variations of the sample surface temperature. The change in temperature of the sample surface causes a corresponding change in resistance of the probe and the output voltage of the Wheatstone bridge. The contact area between sample and probe has been determined to be  $30\text{nm}$  [2].

The second probe is the ultra compliant polyimide thermal probe [6-8]. The probe has dimensions of  $350\mu\text{m}$  in length,  $50\mu\text{m}$  in width and  $3\mu\text{m}$  in thickness. It has a spring constant of only  $0.082\text{N/m}$  and this characteristic allows the SThM to operate without the need of force control [8]. The probe is made of polyimide which is mechanically stable and has a low thermal conductivity of  $0.147\text{W/mK}$ , offering a high degree of thermal isolation [7]. With the flip-over manufacturing technique, a tip diameter of  $50\text{nm}$  can be achieved. Figure 4 shows the probe cross-section and the resulting ultra compliant polyimide thermal probe. A lateral spatial resolution of  $<50\text{nm}$ , topography resolution of  $<1\text{nm}$  and thermal resolution of  $1.2\text{mK}$  have been achieved [8]. Unless otherwise stated, the thermal distribution images are obtained using the Wollaston resistive probe.

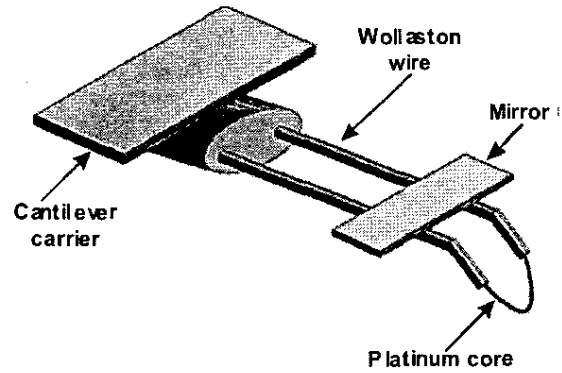


Figure 3. Schematic diagram of the Wollaston resistive probe

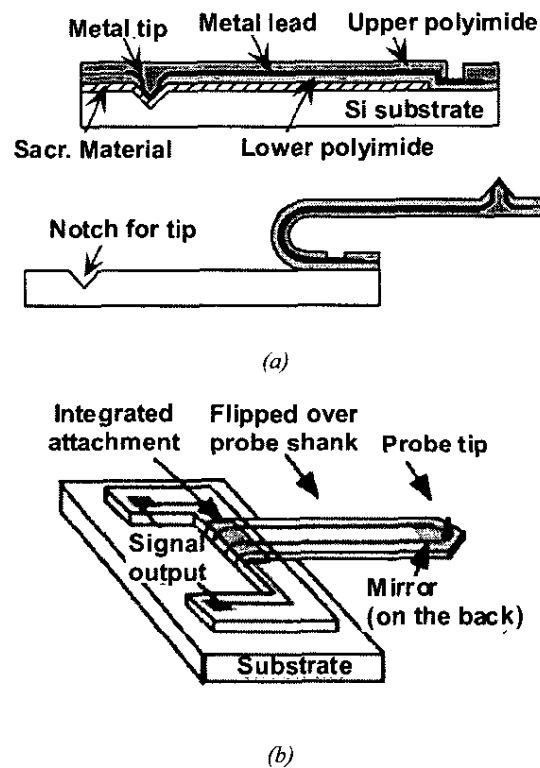


Figure 4. (a) Schematic diagram of a device cross-section illustrating the six mask fabrication sequence of the micromachined resistive thermal probe [7]; (b) Schematic of the resulting ultra compliant polyimide thermal probe [8]

## RESULTS AND DISCUSSIONS

The test device is an *n*-channel enhancement LDD MOSFET with  $W/L=2.0\mu\text{m}/3.0\mu\text{m}$  built on a *p*-substrate, sharing a common source, gate and substrate with many other transistors in two columns. The optical micrograph of the device in figure 5 shows the source, drain and gate regions of the *n*MOSFET. It can be seen that a  $10\mu\text{m} \times 10\mu\text{m}$  scan area would include the gate, source and drain outline structures, although it is also possible to observe the high thermal wave amplitude in a  $5\mu\text{m} \times 5\mu\text{m}$  scan area.

An outline of the gate contact, source, drain and gate channel has been drawn as white lines. This white outline is overlaid in the dynamic thermal distribution images in the following sections for

registration purposes. This is done with the help of the topography images which are obtained simultaneously during dynamic temperature measurement. This outline will assist in the analysis of the location and shift of the high thermal wave amplitude region within the channel in the thermal distribution images. The  $I_{DS}-V_{DS}$  characteristic in figure 6 is obtained at two different gate voltages,  $V_{GS}=3V$  and  $4V$ . The operating points used in the experiments are also indicated.

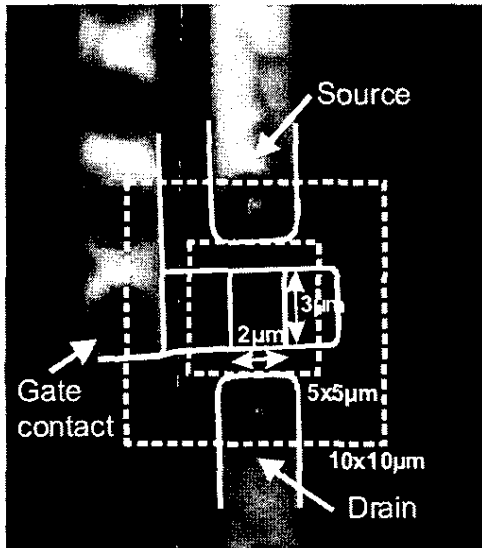


Figure 5. Optical micrograph of nMOSFET device with  $W/L=2.0\mu\text{m}/3.0\mu\text{m}$

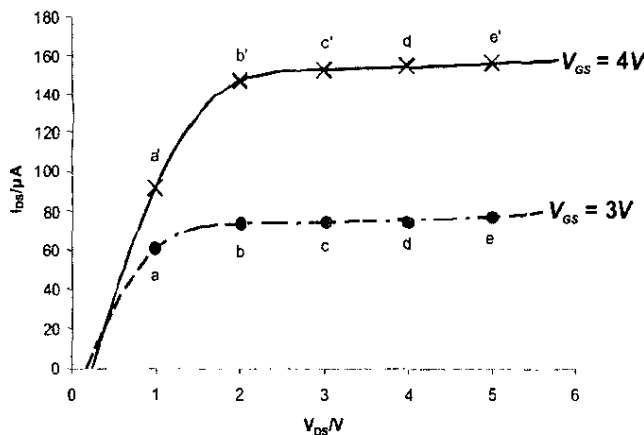


Figure 6.  $I_{DS}-V_{DS}$  characteristic of nMOSFET with  $W/L=2.0\mu\text{m}/3.0\mu\text{m}$  at  $V_{GS}=3V$  and  $4V$ , showing the operating points a-e and a'-e' used in the experiments

In the dynamic SThM mode, the nMOSFET is modulated by passing an alternating voltage  $v_{gs}$  to the gate, i.e. the transistor gate is pulsed to turn the transistor on and off rapidly, and a dc voltage to the drain while the source and substrate are grounded. The signal of the ac voltage supplied to the gate of the nMOSFET is sinusoidal with a frequency of 570Hz, because this frequency produces thermal distribution images with the least signal-to-noise ratio (SNR) problems. A small current of 3mA passes through the scanning thermal probe. Figure 7 shows the thermal distribution of the nMOSFET with  $V_{DS}$  increasing in 1V steps and  $v_{gs}$  amplitude of 3V at

570Hz, while figure 8 shows the corresponding thermal distribution with  $v_{gs}$  amplitude of 4V at 570Hz.

The grey circles in figure 6 indicate the operating points at which the nMOSFET is biased with  $v_{gs}$  having amplitude of 3V. The respective dynamic thermal distribution images, set to the same scale, are shown in figures 7(i)(a)-(i)(e). The steady increase in thermal wave amplitude can be attributed to the fact that as the drain voltage is increased, the electrons now have higher energies. The source-drain lateral electric field also increases with higher drain voltages, accelerating the electrons towards the drain even more. More current is also pumped into the gate channel, resulting in higher heat dissipation. The grey crosses in figure 6 indicate the corresponding operating points when  $v_{gs}$  amplitude is 4V at 570Hz, while figures 8(i)(a')-(i)(e') show the increase in thermal wave amplitude of the nMOSFET with increasing  $V_{DS}$ . They also show that thermal penetration depth increases with higher  $V_{DS}$ . It can be seen that the highest thermal wave amplitudes are always generated in the region between the source and the drain closer towards the drain. 50nm resolution is exhibited by the images despite the high thermal conductivity of silicon. This resolution is achievable due to the high temperature gradient and strong localized heating within the gate of the device when the nMOSFET is biased.

Figures 7(ii)(a)-(ii)(e) and 8(ii)(a')-(ii)(e') show how the behaviour of high thermal wave amplitude generated within the DUT shifts with different gate and drain biasing. Figures 7(ii)(a)-(ii)(e) clearly show that the high thermal wave amplitude region always develops in the area closer to the drain, and this result corresponds with earlier investigations [9-11]. The effects of increasing  $V_{DS}$  will now be analyzed using figures 8(ii)(a')-(ii)(e') when the device is biased with  $v_{gs}$  having amplitude of 4V, because the change in the amplitude of the thermal wave generated is more obvious and drastic. In figure 8(ii)(a') with  $V_{DS}=+1V$ , the nMOSFET is operating in the linear region, and the thermal distribution is still homogeneous throughout the device. A significantly high amplitude thermal wave region is observed when  $V_{DS}=+2V$  in figure 8(ii)(b'). This region develops first closer to the drain area, as the pinch-off begins to develop near to the drain and the device is operating in the near-saturation mode. Figure 8(ii)(c') shows the thermal image of the nMOSFET operating beyond the saturation mode at  $V_{DS}=+3V$ . The high thermal wave amplitude, still located closer to the drain, begins to expand as the pinch-off departs from the drain, so that electrons now have to travel longer distance through the depletion region to reach the drain. Figure 8(ii)(d') shows the penetration of the thermal wave more and more into the drain area, and also the source as the pinch-off moves closer and closer to the source as the device is operated in the far saturation mode when  $V_{DS}=+4V$ . Figure 8(ii)(e') in which  $V_{DS}=+5V$  and the device is operating in the near-breakdown region, shows that high amplitude thermal wave penetrates further into the source. The punchthrough effect plays a significant role in heating up the device as well.

To obtain a further understanding on the thermal distribution in the nMOSFET, vertical line analysis along  $PP'$  as shown in figure 9(a) is performed on each thermal image. Line measurements are performed on each of the thermal image in figures 7(ii)(a)-(ii)(e) and 8(ii)(a')-(ii)(e'). The thermal line profiles for the nMOSFET biased at  $v_{gs}$  having amplitude of 3V are shown in figure 9(b) for different drain voltages. For the line profile when  $V_{DS}$  is 5V, two peaks are observed as shown in figure 10 denoted by 'M' and 'N'. At  $y=2\mu\text{m}$ , it is the end of the source structure and beginning of the gate channel. It is observed that the temperature gradually increases from  $y=2\mu\text{m}$  to  $6\mu\text{m}$ . This is because the electrons are traveling along an inversion channel which is gradually narrowing, like that shown in figure 1, which means that the electrons experience resistance  $R$  which is gradually increasing as well. The power dissipated, which is given by  $I^2R$ , gradually increases too, and so does the temperature of the test device.

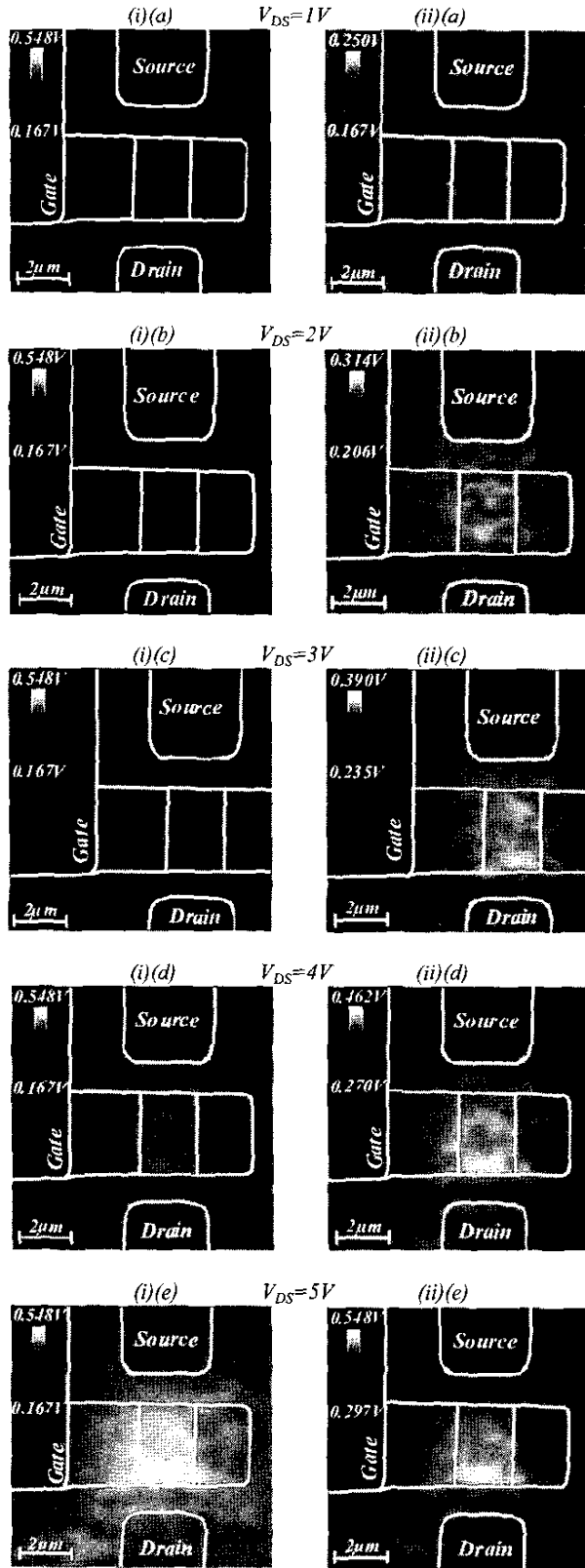


Figure 7. Thermal distributions with  $v_{gs}$  having amplitude of 3V at  $f=570\text{Hz}$

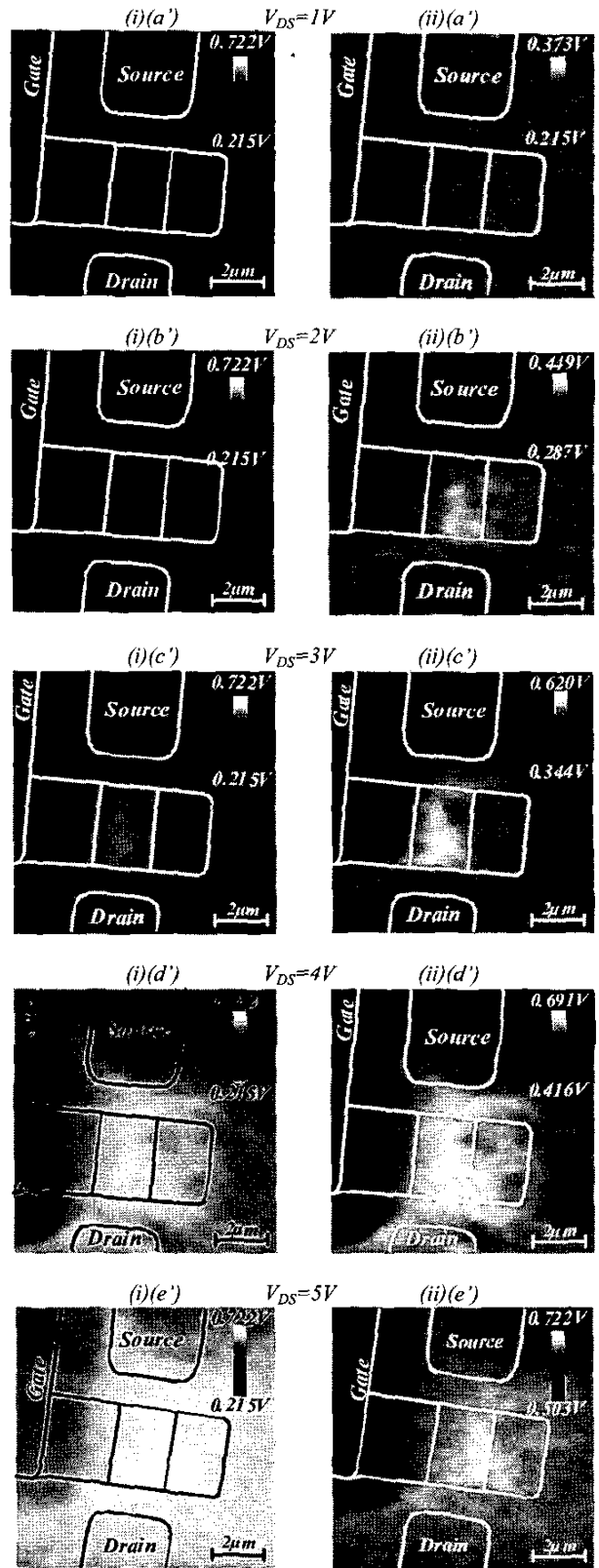
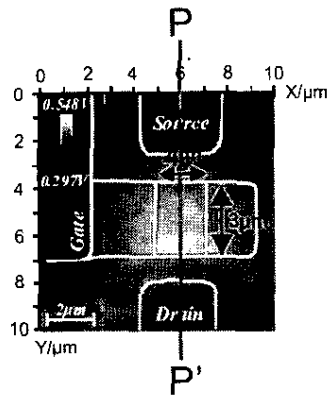
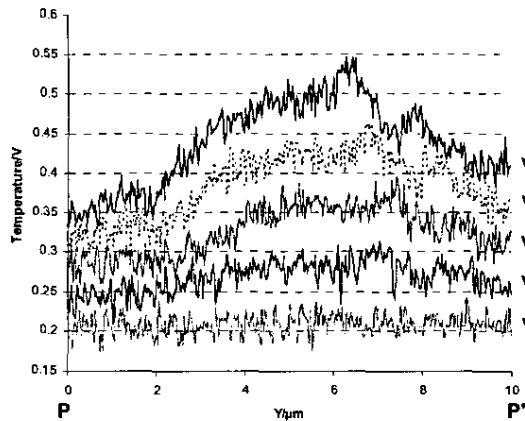


Figure 8. Thermal distributions with  $v_{gs}$  having amplitude of 4V at  $f=570\text{Hz}$



(a)



(b)

Figure 9. (a) Vertical line measurement on thermal images across  $PP'$ ; (b) Thermal line profiles across  $PP'$  at different drain voltages when the amplitude of  $v_{gs}$  is  $3V$

A possible explanation to the first peak at ' $M$ ' is due to the pinch-off. There are high electric fields at the inversion channel-depletion region interface that can cause high acceleration of electrons. Also, from [11], the power dissipated at the pinch-off  $P_{po}$  is given by:

$$P_{po} = I_D(V_D - V_{Dsat}) \quad (1)$$

while the power dissipated in the inversion channel  $P_{ch}$  is given by:

$$P_{ch} = I_D V_{Dsat} \quad (2)$$

where  $I_D$  is the drain current,  $V_D$  is the drain voltage, and  $V_{Dsat}$  is the saturation drain voltage. With our bias conditions, the power dissipated in the pinch-off should be higher than that dissipated in the inversion channel, and our line measurement in figure 10 supports this.

Since ' $N$ ' corresponds to the drain edge where there is transition from  $p$  to  $n$ - region of the LDD structure, high vertical electric field is present there, which may be the reason for the second peak ' $N$ '. Higher silicon lattice vibrations and high velocity electrons may have caused the heating up. There is a fall in temperature in the region between ' $M$ ' and ' $N$ ' despite the fact that electrons are traveling in the depletion region, and experiencing higher resistance than in the inversion channel. One possible explanation to this could be that, as the electrons are accelerated towards the drain from the inversion channel, the electrons flow deeper in the volume of the MOSFET and not just on the surface, like that shown in figure 1. Therefore, heat dissipation occurs deeper in the substrate and not just at the surface and the thermal probe is not able to pick up this information. The second explanation is that, since electrons now travel through the bulk, the effective resistance experienced by the electrons drops

as it now travels across a bigger surface area and away from the surface, and therefore the power dissipation drops.

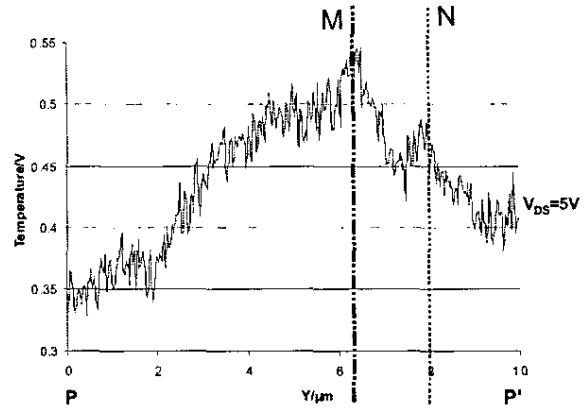
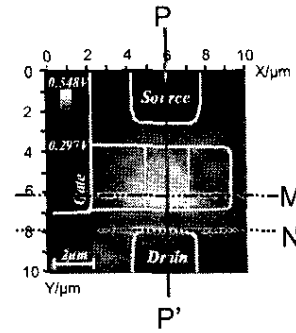


Figure 10. Thermal line profile across  $PP'$  when the amplitude of  $v_{gs}$  is  $3V$ , and the drain voltage is  $5V$

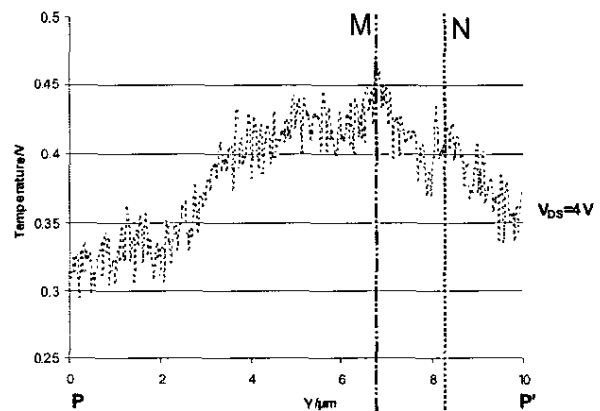
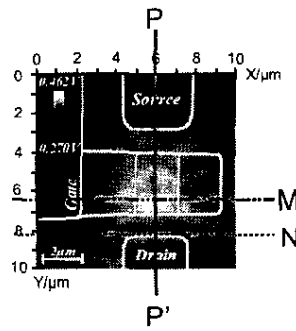


Figure 11. Thermal line profile across  $PP'$  when the amplitude of  $v_{gs}$  is  $3V$ , and the drain voltage is  $4V$

Figure 11 shows the line measurement at  $v_{gs}$  having amplitude of  $3V$  and drain voltage of  $4V$ . Comparing the line measurements in figures 10 and 11, it is observed that there is a slight movement of peak 'M' away from the drain and thus increasing the distance of  $M-N$  as the drain voltage increases, indicating that the pinch-off moves away from the drain with higher drain voltage, and so does the highest thermal wave region observed along the channel.

The measurements are repeated for bias conditions similar to that in figures 7(i)(e) and 7(ii)(e), i.e. the amplitude of  $v_{gs}$  is  $3V$  at  $570Hz$ , and  $V_{DS}=+5V$ , using the ultra compliant polyimide thermal probe. Figure 12 shows the topography images as obtained by the two different probes. With the ultra compliant probe, it is clear that we are able to obtain a topography image that reveals more details with respect to the contour of the structure. This can be attributed to the higher flexibility of this probe compared with the resistive probe, allowing the probe to follow the topography structures very easily. The ability of the two probes to measure thermal wave amplitudes is also compared. The results are obtained in figure 13. While the ultra compliant probe is able to reveal the area generating high amplitude thermal wave at a large scan size of  $100\mu m \times 100\mu m$  better compared with the resistive probe, the thermal distribution obtained by the ultra compliant probe at a scan size of  $10\mu m \times 10\mu m$  tends to have greater SNR problems, and the resulting image is noisier. The thermal wave distribution images obtained using both probes reveal that the highest amplitude of thermal waves generated by an operating nMOSFET is closest to the drain at pinch-off and slowly spreads out and extends towards the source as the nMOSFET approaches the near breakdown operating point.

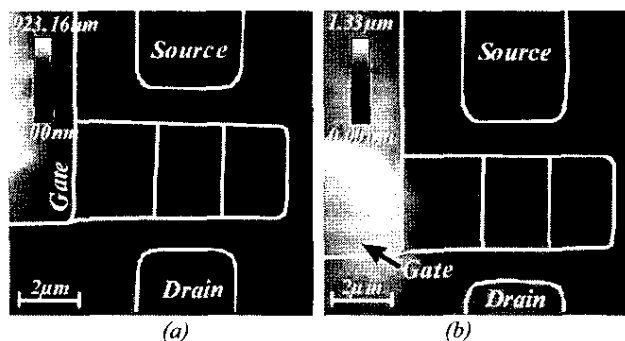


Figure 12. Topography obtained by the use of (a) Wollaston resistive probe, and (b) ultra compliant probe

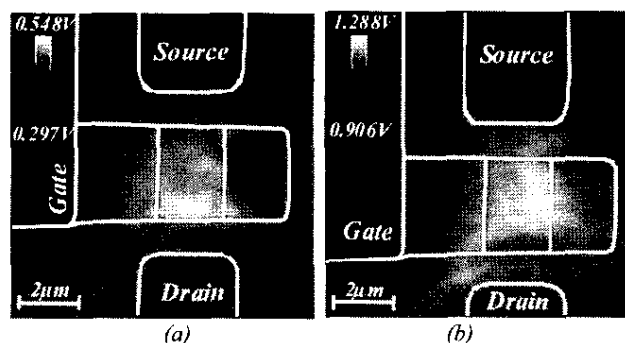


Figure 13. Thermal distributions obtained by the use of (a) Wollaston resistive probe, and (b) ultra compliant probe

## CONCLUSION

The thermal distribution of the nMOSFET under different operation modes has been mapped with a resolution of  $50nm$  by

dynamic SThM. All images reveal that the highest thermal wave amplitude generated in the MOSFET is at the region closer to the drain. This thermal wave amplitude slowly increases in magnitude, and also spreads out towards the source as the drain voltage is increased, and the MOSFET functions in different operating conditions. Line analysis of the thermal images gives deeper insight into our understanding of heat generation in the MOSFET, as two peaks of high thermal wave amplitudes are observed, one due to high electric field at the pinch-off and the other high electric field at the drain edge. With a greater understanding of heat generation and dissipation in MOSFET, new development and design methods can be adopted by future designers of transistors to minimize the problems of transistor failures due to thermal production, and thus increasing the lifetime of future transistors.

## ACKNOWLEDGEMENT

The authors acknowledge Prof Y.B. Gianchandani of University of Michigan, Ann Arbor, for the supply of ultra compliant polyimide thermal probes for the measurements.

## REFERENCES

- [1] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors," 2004 Edition, <http://public.itrs.net/>
- [2] L.J. Balk, M. Maywald, R.J. Pytkki, "Nanoscope Detection of the Thermal Conductivity of Compound Semiconductor Materials by Enhanced Scanning Thermal Microscopy," in *Inst Phys Conf Ser*, No 146, 1995, pp. 655-658.
- [3] G.B.M. Fiege, V. Feige, J.C.H. Phang, M. Maywald, S. Goerlich, L.J. Balk, "Failure Analysis of Integrated Devices by Scanning Thermal Microscopy (SThM)," *Microelectronics Reliability*, Vol 38, No 6-8, 1998, pp. 957-961.
- [4] N. Arora, "MOSFET Models for VLSI Circuit Simulation: Theory and Practice," Springer-Verlag, 1993.
- [5] A. Altes, K. Mutamba, R. Heiderhoff, H.L. Hartnagel, L.J. Balk, "Scanning Near Field Thermal Microscopy on a Micromachined Thin Membrane," *Superlattices and Microstructures*, Vol 35, 2004, pp. 465-476.
- [6] M. Li, J.J. Wu, Y.B. Gianchandani, "Surface Micromachined Polyimide Scanning Thermocouple Probes," *Journal of Microelectromechanical Systems*, Vol 10, No 1, 2001, pp. 3-9.
- [7] M. Li, Y.B. Gianchandani, "Microcalorimetry Applications of a Surface Micromachined Bolometer-type Thermal Probe," *Journal of Vacuum Science & Technology B*, Vol 18, No 6, 2000, pp. 3600-3603.
- [8] M. Li, J.H. Lee, F. Cerrina, A.K. Menon, Y.B. Gianchandani, "Chemical and Biological Diagnostics using a Micromachined Polyimide-shank Scanning Bolometer Probe," in *Solid State Sensors, Actuators and Microsystems Workshop, Hilton Head Island, South Carolina, USA, 2002*, pp. 235-238.
- [9] M. Palaniappan, G.B.M. Fiege, V. Ng, R. Heiderhoff, J.C.H. Phang, L.J. Balk, "Correlation of Thermal and Electrical Properties of a Short Channel nMOSFET," in *Proc Int Symp Testing & Failure Analysis (ISTFA 1999), 14-18 Nov 99, Santa Clara, California, USA, 1999*, pp. 465-470.
- [10] T.H. Lee, G.B.M. Fiege, A. Altes, G. Zimmermann, V. Ng, R. Heiderhoff, J.C.H. Phang, L.J. Balk, "Characterization of MOS Devices by Scanning Thermal Microscopy (SThM)," in *Proc Int Symp Testing & Failure Analysis (ISTFA 2001), 11-15 Nov 2001, Santa Clara, California, USA, 2001*, pp. 191-197.
- [11] R. Ostermeir, K. Brunner, G. Abstreiter, W. Weber, "Temperature Distribution in Si-MOSFET's Studied by Micro Raman Spectroscopy," *IEEE Trans Electron Devices*, Vol 39, No 4, 1992, pp. 858-863.