

Thermal Instabilities in High Current Power MOS Devices: Experimental Evidence, Electro-thermal Simulations and Analytical Modeling

P. Spirito, G. Breglio, V. d'Alessandro, N. Rinaldi

Abstract - The phenomenon of the thermal instability presented by some high current power MOS has been intensively investigated, both by experimental means and by numerical simulations. An analytical expression for the positive temperature coefficient of the Drain current has been developed and a model for the thermal instability in transient operation has been proposed. The results explain the main causes of the thermal instability and give some rules to evaluate the possible failure occurrence for a given device.

I. INTRODUCTION

The limitation of the Forward Bias Safe Operating Area (FBSOA) in power bipolar transistors due to thermal instability is a well known effect. The positive temperature coefficient of the collector current due to both a reduction of the base voltage and the increase of current gain as temperature increases, gives rise to the so called "hot spot", namely a large current and temperature increase in a localized region of the device area, that eventually leads to the destruction of the device [1]. Power MOS devices on the contrary are assumed to have their S.O.A. limited only by constant power dissipation, and they are granted to be thermally stable, due to the negative temperature coefficient of the Drain current, caused by the mobility drop as the temperature increases.

However, in recent papers [2,3] thermal instability effects in MOS devices have been presented, that limit their SOA capability. The transient thermal investigation in low voltage high current MOS reported in [2], shows a hot spot growth on the device area when the Drain voltage of the power MOS is increased above a given value, that eventually leads to the device failure.

In this paper we present an extensive analysis of this phenomenon, to show that thermal instability can be a significant phenomenon for power MOS of the new generation with high transconductance and VLSI cell design.

The paper is organized as follows. In Sect. II an experimental evidence of the thermal instability in Power MOS is given, by means of an experimental set-up that allows to obtain the 2D temperature distributions

onto the chip area in transient operation. In Sect. III a parametric analysis for the elementary transistor cell is performed, using a numerical 2D device simulator, to better understand the physical reasons for this quite large positive temperature dependence of the drain current. In Sect. IV the dynamics of the temperature instability are studied by means of an electro-thermal simulator that takes into account the layout and the geometry of actual power devices, composed of a large number of elementary cells connected in parallel. In Sect. V a new analytical model will be presented to explain this type of instability, based on the results of the analysis. The model is capable to predict the onset of thermal instability for a given MOS device, and can be used both to define the allowed SOA for the device and as a design guide to make more rugged devices.

II. EXPERIMENTAL OBSERVATION OF THERMAL INSTABILITIES

The detection of the temperature distribution across the area of the device in transient conditions has been made possible by the use of our radiometric 2D measurement system, that allows to acquire the transient temperature maps of the device surface [4]. The measurements are done by scanning the surface with a x-y motor stage, and by transforming the 2D radiance distributions recorded in the transient mode into the 2D temperature distributions by using the real emissivity map of the surface previously acquired with the same set-up. In this way the true temperature values can be evaluated point by point, using a fast infrared radiometer with a very small spot size; our system has a time resolution of 2 μ s and a spatial resolution of 8 μ m, in a range of 5x5 cm^2 .

The measurements have been performed on a power MOS rated 60A 60V (indicated as #1), with a chip area of 5x5 mm^2 ; the chip is mounted on an open TO3 package to allow the radiometric temperature measurements with our set up. The device was biased at different Drain voltages and currents, and operated in pulse bias to analyze the transient response but preventing device failure. The temperature maps, taken at different time intervals with a constant pulse power of 180 W, show different temperature distributions on the chip passing from low drain voltages (less than 20 V) to higher ones (above 20 V). In the former case, the temperature distribution is quite uniform, (considering the heat sink effect given by the two source terminals connected to the top active area of the chip) and the distribution remains the same during

The authors are with the Dept. of Electronic Engineering, Università di Napoli "Federico II", Via Claudio 21, 80125 Napoli, Italy, E-mail: pspirito@unina.it

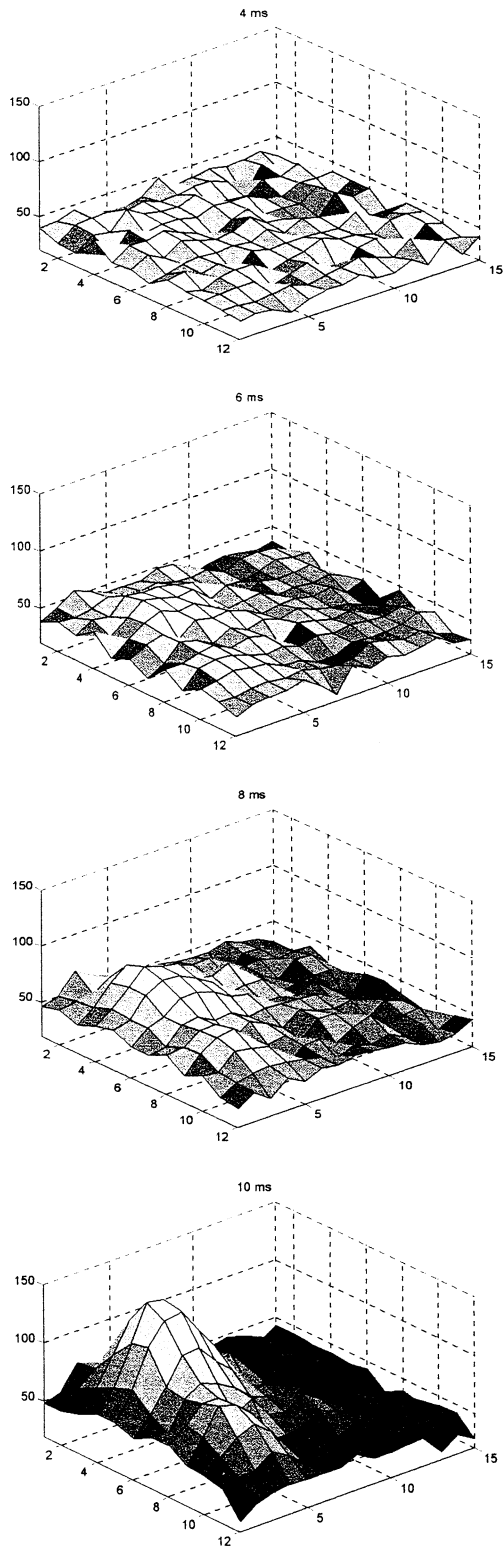


Fig. 1 Temperature distribution of the Power MOS at four time intervals after the application of the pulse power (90W): 4 μ s, 6 μ s, 8 μ s and 10 μ s.

the time, with an even increase of the temperature due to the thermal chip time constant. Instead, when biased at

the same power but at larger drain voltage, the temperature plots clearly present an uneven increase of the temperature after some milliseconds, with a large increase of the temperature in a localized region of the device, as it can be seen by the 2D temperature plots of fig. 1. This is a clear evidence of a formation of a hot spot, generated after a relatively short time after the power application in pulse operation, with a hot spot area less than 10% of the active chip area, carrying an average temperature more than twice the average temperature in the remaining chip area.

A general proof of a temperature instability phenomenon for this case is given by the failure presented by the device if the time of the pulse bias is increased, while the same device would not fail when biased at drain voltages below 20 V, at the same power.

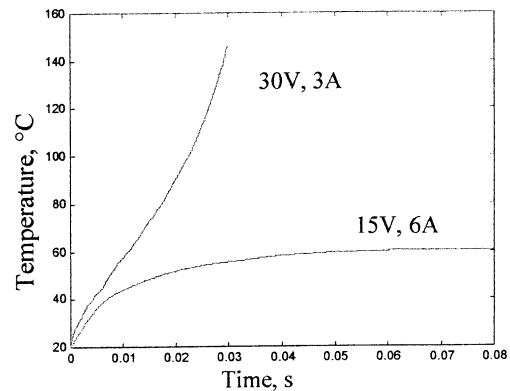


Fig. 2. Temperature transient detected for the Power MOS of Fig. 1 at the hot spot location, for the same pulse power (90 W) but for two different bias values.

A more convincing experimental proof of this instability can be given by the plots of the temperature transients in a single point of the chip, inside the region where the hot spot is likely to develop, reported in fig. 2. Both plots refer to a pulse power of 125 W, well below the maximum ratings. The first plot, at $V_{DS} = 15$ V shows an asymptotic increase due to the thermal time constant up to a temperature of 80 °C. The second one, for the same power but now at $V_{DS} = 30$ V presents a change in curvature after some ms, with temperature increasing at a much faster rate than the previous case, and leading to unacceptable values if not limited by the pulse duration of 10 ms used in this case.

As in all semiconductor devices, a necessary condition for the thermal runaway is that the current of the device in the operating range increases with temperature. For high voltage and low current Power MOS the positive temperature coefficient is usually restricted to a limited range of low currents near the threshold voltage V_{TH} , far enough from the operating range to be neglected for the majority of applications. However this is not the case for Power MOS of new generation with high transconductance K , where the temperature coefficient can be positive for a significant range of currents, even above the maximum current rating of the device, as it will be shown.

III. DEVICE PHYSICS

For a better understanding of the role played by the technology used to realize the elementary transistor on the Drain current temperature coefficient, an exhaustive evaluation of the influence of technological parameters is needed. This task has been done by means of the two-dimensional device simulator MEDICI.

We have assumed as reference values for the elementary cell a structure with peak doping in channel region $N_{AP} = 3.5 \times 10^{17} \text{ cm}^{-3}$, oxide thickness $t_{OX} = 350 \text{ \AA}$ and $L_{CH} = 0.76 \text{ \mu m}$. From the I_D - V_{GS} characteristics of fig.3 obtained for different substrate temperatures, it can be seen that for $I_D < 3.2 \times 10^{-4} \text{ A/\mu m}$ the Drain current shows a positive temperature coefficient, while for higher currents the coefficient becomes negative.

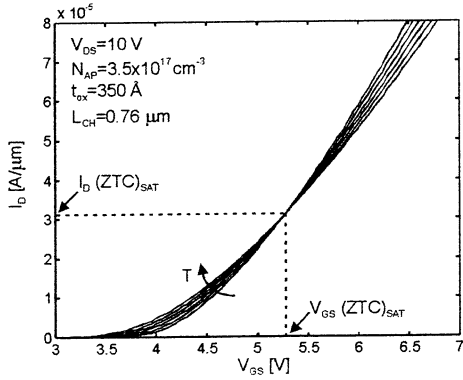


Fig.3. I_D - V_{GS} characteristics for the cell parameters indicated, at different temperatures.

This effect has been already analyzed in literature and can be explained as follows: at low Drain currents, the temperature behavior is dominated by the decrease in the threshold voltage, which leads to a current increase, while at high Drain currents the decrease in channel mobility prevails, leading to a negative temperature coefficient.

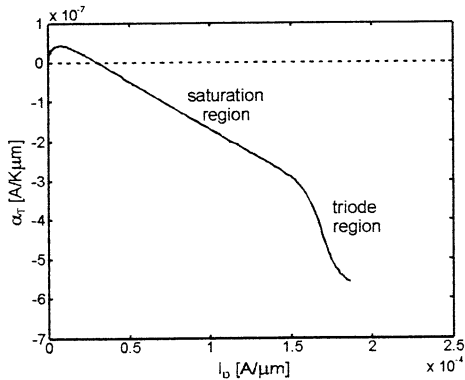


Fig.4. Temperature coefficient as a function of Drain current

In fig.4 the temperature coefficient α_T as a function of Drain current is shown; as can be seen, the positive temperature coefficient occurs in saturation operating mode, where the current is independent of the drift resistance.

In the following, we have considered the effect of a variation in body (and channel) doping, assuming different maximum values of the doping itself:

- 1) $N_{AP} = 4.8 \times 10^{17} \text{ cm}^{-3}$ (device #1);
- 2) $N_{AP} = 3.5 \times 10^{17} \text{ cm}^{-3}$ (device #2);
- 3) $N_{AP} = 2.5 \times 10^{17} \text{ cm}^{-3}$ (device #3);
- 4) $N_{AP} = 1.6 \times 10^{17} \text{ cm}^{-3}$ (device #4);
- 5) $N_{AP} = 9.0 \times 10^{16} \text{ cm}^{-3}$ (device #5).

First, let us consider the threshold voltage variation with temperature. In fig. 5 there are shown both the effects of the gate oxide thickness and of body doping. The dV_{Th}/dT coefficient increases from $-2 \text{ mV/}^\circ\text{C}$ of MOS for integrated circuits, to more than $-5 \text{ mV/}^\circ\text{C}$ for Power MOS structures.

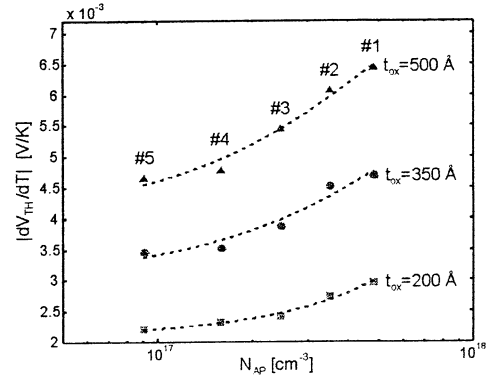


Fig.5. Threshold voltage temperature coefficient as a function of peak doping in channel region for various values of gate oxide

Then we consider the effect of body doping in the whole temperature coefficient α_T , shown in fig. 6. As it can be noted, the body doping has a large effect on the α_T . From the above figure, it is clear that a low value of peak doping in channel region should be needed to reduce thermal effects.

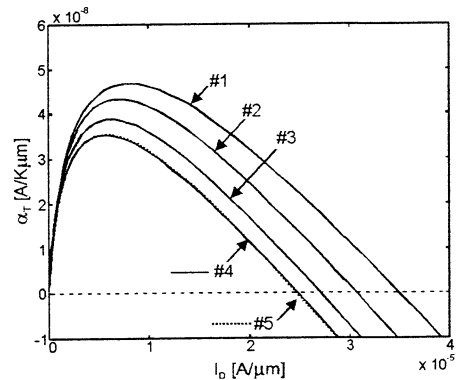


Fig.6. Temperature coefficient α_T as a function of Drain current for various values of peak doping in channel region.

Unfortunately, modern power MOSFET's for low-voltage applications are typically realized with relatively high doping values, due to various factors:

- The trend toward short channel lengths using VLSI techniques requires a high doping level in order to prevent punch-through effects.
- the body resistivity should be made relatively high to reduce the parasitic BJT operation, that is detrimental for SOA.
- The gate reduction in oxide thickness in order to obtain more efficient devices causes the threshold voltage to decrease, so that a high channel doping is used to avoid very low values of the threshold voltage itself.

The over-all effect of gate oxide thickness on α_T is shown in fig. 7, and it is not so effective as for the threshold voltage.

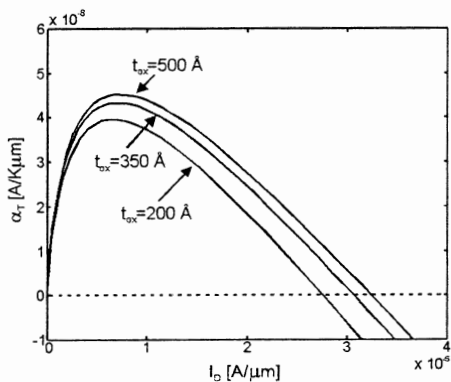


Fig.7. Temperature coefficient as a function of Drain current for various values of gate oxide thickness

This is substantially due to the fact that a t_{ox} increase causes an increase of the threshold voltage temperature dependence, but on the contrary it causes a transconductance lowering. By increasing the oxide thickness, the first factor prevails, leading to less thermally robust devices.

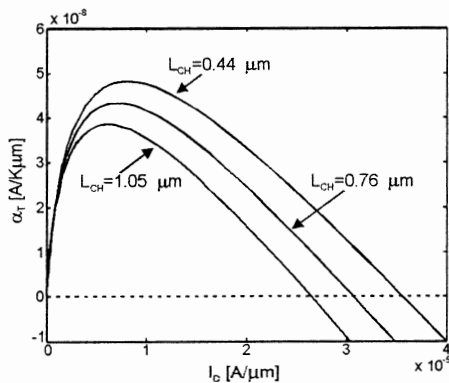


Fig.8. Temperature coefficient as a function of Drain current for various values of channel length

Lastly an analysis of the effect of the channel length has been performed. Simulation results, depicted in fig.8, show a noticeable increase of α_T if channel length L is decreased.

IV. ELECTROTHERMAL SIMULATOR

We have developed a numerical simulator to describe the electrical-thermal behavior of power electronic devices. The electro-thermal interaction is taken into account by the simultaneous solution of the two nonlinear systems and by the interchange of variable and/or parameters [5].

The electrical network takes into account a compact model of the active elementary device cell (in this case the NMOS LEVEL3 SPICE model) connected by means of resistances which are dimensioned according to the lay-out of the metallization pattern. The thermal network is a 3D electrical equivalent network which is composed by resistances, capacitances and current generators. The die-case-heat-sink contact is simulated by means of a double RC network. Realistic air-silicon thermal boundary conditions at the lateral and upper surfaces of the die have been taken into account.

As inputs are considered the biasing condition (V_{GS} and V_{DS}), the time of simulation and the time step (time resolution). With these starting condition, the simulator core works on the recursive resolution of the two network, and because they can be analytically described by means of two non linear equation systems (corresponding to the Kirchoff's law to the nodes) that are characterized by a large number of rows (equations) with very few non zeros coefficients; sparse matrix solving routine have been used to reduce the computational effort.

By means of the above described simulator we have analyzed the electro-thermal behavior of the characterized Power MOS #1.

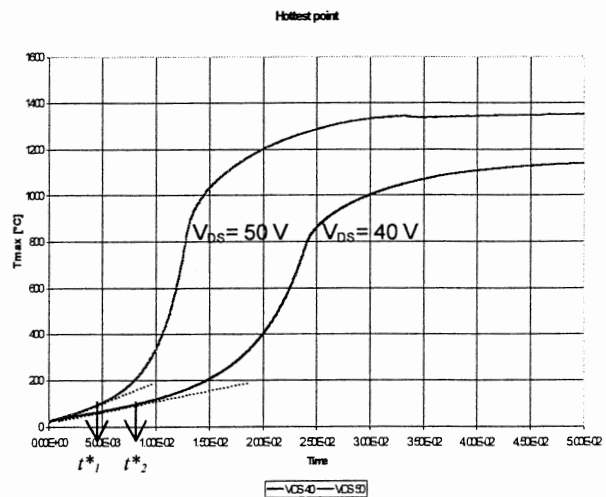


Fig. 9. Dynamic temperature behavior of the hottest point over the simulated PowerMOS, for two different biasing conditions: $V_{DS} 50V$, $I_D 2.3A$, and $V_{DS} 40V$, $I_D 2.3A$.

In fig. 9 the plot of the temperature transient on the hottest point for two different bias voltages are reported. In both cases the temperature instability is clearly noticed, as for the experimental plot of fig. 2: it can be noted a dramatic temperature increase after a time t^* , indicating the reaching of the thermal instability condition. The time t^* for the triggering of the instability is smaller if the bias voltage is increased.

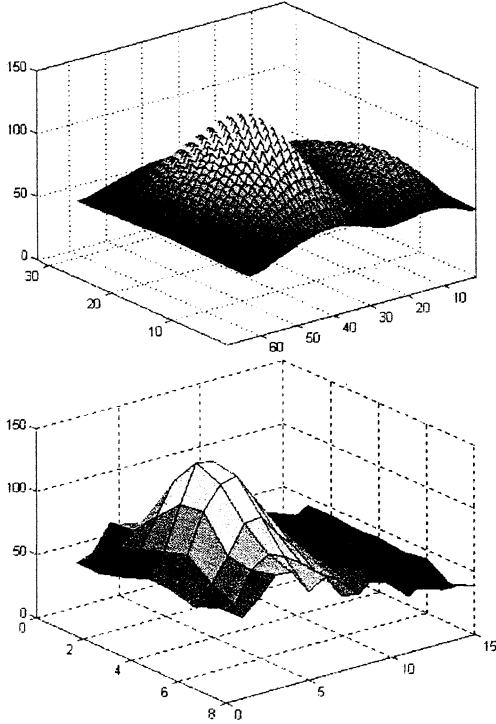


Fig. 10. Comparison of thermal maps obtained by simulation data (upper plot) and experimental data (bottom plot), for half of the active device. The maps refer to a time of about 10ms from the power pulse application.

The comparison of the simulation results with experimental data is quite good, as it can be seen from fig. 10, both on the spatial localization and in the time evolution of the thermal instability, demonstrating the good accuracy of the simulator itself.

V. ANALYTICAL MODELING

The usual condition for thermal instability of power devices is assumed from [6], indicating P_G the electrical power and P_D the thermally dissipated one, as:

$$\frac{\partial P_G}{\partial T} \geq \frac{\partial P_D}{\partial T}$$

For Power MOS this leads to:

$$V_{DS} \frac{\partial I_D}{\partial T} \geq \frac{1}{R_T(t)} \quad (1)$$

where $R_T(t)$ is the time dependent thermal resistance, and $\partial I_D / \partial T = \alpha_T(I_D)$ is the temperature coefficient of the drain current. From (1) the necessary condition for instability to occur is a positive temperature coefficient $\alpha_T(I_D)$. Once defined adequate analytical expressions for both the temperature coefficient $\alpha_T(I_D)$ and the time dependent thermal resistance $R_T(t)$, one can introduce an instability factor S from (1), and the onset of the instability can be defined as:

$$S \equiv V_{DS} \alpha_T(I_D) R_T(t) = 1 \quad (2)$$

From (2), to model the time dependence of the onset of the thermal instability, the only time dependent quantity is the term $R_T(t)$, while the temperature coefficient of current can be obtained by the steady state MOS model (by assuming the time scale for developing thermal instability much larger than that related to the electrical transients for the device itself).

A. Device modeling

The temperature dependence of the Drain current of a MOS can be obtained, as a first order approximation, from the well known expression in saturation:

$$I_D(T) = K(T) [V_{GS} - V_{TH}(T)]^2 \quad (3)$$

where $K(T)$ is the transconductance, given by:

$$K(T) = \frac{\mu_0(T) C_{ox} W}{2L} \quad (4)$$

It can be demonstrated that the threshold voltage temperature dependence can be described by a linear law:

$$V_{TH}(T) = V_{TH}(T_0) - \phi(T - T_0) \quad (5)$$

while the low-field channel mobility can be modeled by a simple power law:

$$\mu_0(T) = \mu_0(T_0) \left(\frac{T}{T_0} \right)^{-m} \quad (6)$$

From the above expressions, an analytical expression is derived for the temperature coefficient $\alpha_T(I_D)$ of the Drain current:

$$\alpha_T(I_D) = -\frac{m}{T} I_D + 2\phi \sqrt{KI_D} \quad (7)$$

where ϕ is the absolute value of temperature coefficient of the threshold voltage and m is the mobility power factor.

From (7) the peak value of α_{Tmax} and the Drain current I_D^* corresponding to a zero-temperature-coefficient can be easily achieved:

$$\alpha_{T_{max}} = \frac{\phi^2 K(T) T}{m} \quad (8)$$

$$I_D^* = \frac{4\phi^2 K(T) T^2}{m^2} \quad (9)$$

Recalling that the threshold voltage is given by:

$$V_{TH} = \Phi_{MS} - \frac{Q_{OX}}{C_{OX}} + 2\psi_B + \frac{\sqrt{4qN_{AP}\epsilon_S\psi_B}}{C_{OX}} \quad (10)$$

where ψ_B is the bulk potential, N_{AP} is the peak doping in channel region, n_i is the intrinsic carrier concentration, Φ_{MS} is the work-function difference, one can compute the temperature coefficient of threshold voltage as:

$$\frac{dV_{TH}}{dT} = \frac{d\Phi_{MS}}{dT} + \frac{d}{d\psi_B} \left(2\psi_B(T) + \frac{\sqrt{4qN_{AP}\epsilon_S\psi_B(T)}}{C_{OX}} \right) \frac{d\psi_B}{dT} \quad (11)$$

From (11) with some algebra we can obtain the following expression for the absolute value of the threshold voltage temperature coefficient:

$$\phi = \frac{1}{T} \left(-\frac{E_G(T)}{2q} + \psi_B(T) - 3\frac{kT}{q} \right) - \left(2 + \frac{1}{C_{OX}} \sqrt{\frac{qN_{AP}\epsilon_S}{\psi_B(T)}} \right) \left(\frac{\psi_B(T)}{T} - \frac{1}{q} \left(\frac{3k}{2} + \frac{E_G(T)}{2T} \right) \right) \quad (12)$$

As it can be seen from (8) and (9), both the maximum of temperature coefficient, and the current I_D^* where the temperature coefficient is zero, can be relatively high for large transconductance K and short channel devices. In fact both terms depend on the product $\phi^2 K$, that can be considered as a *figure of merit* for thermal instability.

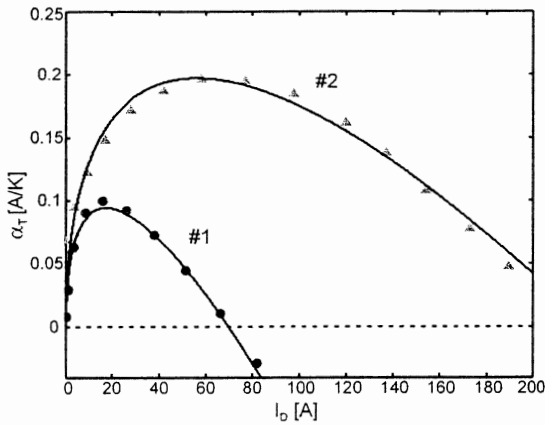


Fig. 11. $\alpha_T(I_D)$ for two MOS devices with different channel length (#1 $L=0.76 \mu\text{m}$; #2 $L=0.4 \mu\text{m}$). Solid line - analytical model; dots - experimental values.

In fig. 11 we can see that the agreement between the analytical model for $\alpha_T(I_D)$ and the experimental data is quite satisfactory; in both cases a substantial positive temperature coefficient has been found; both the peak value $\alpha_{T_{max}}$ and the current I_D^* corresponding to a zero-temperature-coefficient are increasing with channel length decrease.

B. Thermal modeling

A new analytical model for $R_T(t)$ has been derived in [7], for a device area of dimensions L, W with $W \leq L$:

$$R_T(t) = \frac{\sqrt{D}}{WLk\sqrt{\pi}} \cdot \begin{cases} 2\sqrt{t} & [t < t_1] \\ \frac{W}{\sqrt{\pi D}} \left(1 + \ln \sqrt{\frac{t}{t_1}} \right) & [t_1 \leq t \leq t_2] \\ \frac{W}{\sqrt{\pi D}} \left(2 + \ln \sqrt{\frac{t_2}{t_1}} \right) - 2\sqrt{\frac{t_2}{t}} & [t \geq t_2] \end{cases} \quad (13)$$

where D is the thermal diffusivity, k the thermal conductivity, while $t_1 = W^2/4\pi D$, and $t_2 = L^2/4\pi D$ are characteristic times which define different phases of the thermal transient. This model allows a deeper insight on the effect of geometric parameters on the thermal response, and has been found in good agreement with experimental data for the Power MOS examined. From (13) it can be seen that the thermal resistance is a decreasing function of the aspect ratio L/W . It should be noted that (13) refers to a semi infinite substrate; this approximation is accurate if the substrate thickness h is much higher than the device dimensions. If this is not the case, (13) overestimates the thermal resistance. In order to take into account the effect of a finite substrate thickness we use the image method [8].

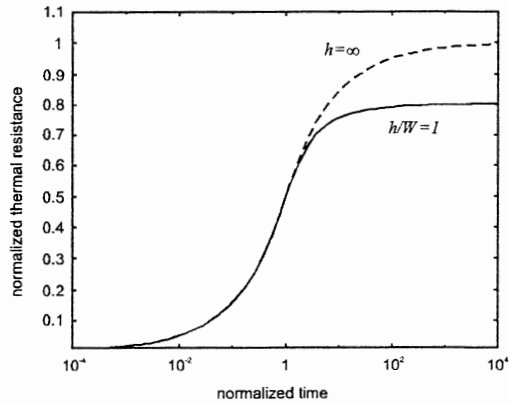


Fig. 12. Normalized thermal resistance as a function of normalized time t/t_1 for a square device with a substrate thickness $h=W$ (solid line) and for an infinite substrate (dashed line).

Using this approach it is possible to show that the thermal resistance of a substrate of thickness h placed on top of a heat

sink can be calculated by adding a correction term to eq. (13), given by:

$$\Delta R_T = -\frac{\ln 2}{2\pi K} \left[1 - \operatorname{erf} \left(\frac{h}{\sqrt{4Dt}} \right) \right] \quad (14)$$

From the plots of fig. 12 one can notice that the heat sink does not affect the small time scale of the thermal transient. Only when the heat flow reaches the heat sink, the cooling effect becomes noticeable.

C. Modeling of the instability onset

Relation (1) can be rewritten for the case of instability as:

$$\alpha_T(I_D) \geq 1/V_{DS} R_T(t) \quad (15)$$

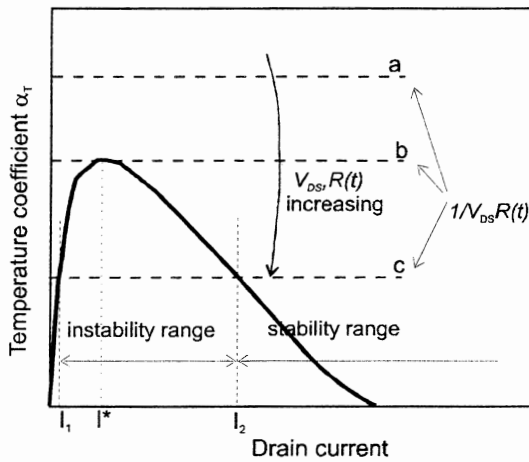


Fig. 13. Graphical analysis to define the thermal instability range of currents. Dashed lines correspond to the right hand side of (15), while the solid line represents the temperature coefficient.

One can evaluate this instability condition by means of a graphical method based on the plot of the temperature coefficient $\alpha_T(I_D)$, as indicated in fig. 11. In this plot, case (a) refers to an unconditionally stable operation, because for the given V_{DS} and $R_T(t)$ values the dashed line does not intercept the positive $\alpha_T(I_D)$ curve. Case (b) corresponds to the onset of instability at the current I^* corresponding to the maximum of the $\alpha_T(I_D)$ curve (α_{Tmax} given by (8)). Case (c) corresponds to an unstable range of currents, where I_1 is the current that triggers the instability, and I_2 is the current the device should reach in the hot spot after the thermal transient (provided that the device does not fail before reaching that value). The thermal instability starts at lower currents if either the drain voltage or the thermal resistance increase.

To evaluate analytically the time for the onset of thermal instability we must consider the time dependence of the thermal resistance $R_T(t)$. Equation (2) allows to define the time t^* at which the instability factor becomes unity:

$$S \equiv V_{DS} \alpha_T(I_D) R_T(t^*) = 1 \quad (16)$$

by expressing the relevant quantities with the expressions (7) and (13) given. A graphical solution for t^* from eq. (4) can be obtained by using the plot of the transient thermal resistance $R_T(t)$ and determining the time t^* from the intercept of this curve with the horizontal line of ordinate $1/[V_{DS} \alpha_T(I_D)]$ as indicated in fig. 14. The $R_T(t)$ plot can be either obtained from eq. (13) of our model, or from the datasheet of the device under study.

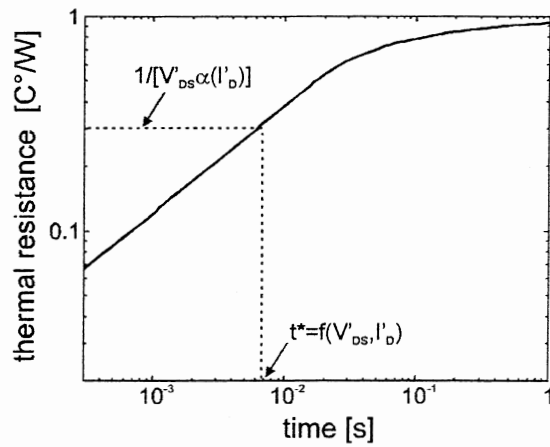


Fig. 14. Graphical analysis showing the definition of the time t^* where instability begins to occur for a given set of bias values V_{DS}' and I_D' . The solid line represents the time-dependent thermal resistance $R_T(t)$ for a given device.

From the example of fig. 14 we can see that the time t^* for the onset of the thermal instability is shorter if either the voltage is increased or the positive temperature coefficient is larger. In any case the time for the onset is defined by the maximum value α_{TMAX} of the temperature coefficient.

We define the ideal SOA in pulse operation by assuming, for a given pulse duration T_p , a constant value $P_{GMAX} = \Delta T_{MAX}/R_T(T_p)$. The thermal instability can reduce this SOA for pulse duration T_p if, for some of the values of V_{DS} $I_D = P_{GMAX}$, the time t^* for the onset of instability, obtained from (4), is lower than T_p . From (16) one can extract the relevant t^* values for each pulsed bias current and voltage; moreover one can obtain the current and voltage bias values corresponding to a t^* value larger than the pulse duration T_p , to define analytically the SOA of the device free from thermal instability.

As an example in fig. 15 there are plotted the S.O.A curves for different T_p pulse values for the two MOS devices #1 and #2 analyzed, by assuming the same data of fig. 11 for $\alpha_T(I_D)$, together with the appropriate values of $R_T(t)$ obtained

from (13). One can see that the thermal instability puts a strong limitation on the SOA curves for $T_p = 10$ ms and 100 ms; the instability limits the maximum voltage for which a constant power rule can be assumed for the SOA. As seen from the plots of fig. 15, the maximum voltage for a constant power law decreases as the peak value α_{TMAX} of the temperature coefficient increases (passing from MOS #1 to MOS #2).

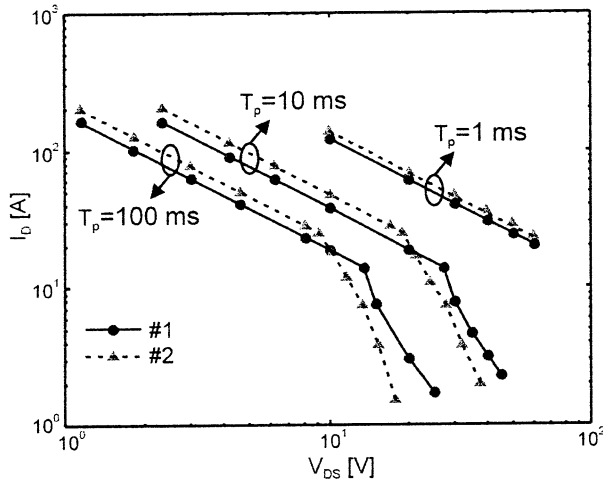


Fig. 15. SOA in pulse operation for the two devices #1 and #2 analyzed, obtained from the model of Sect. V

VI. CONCLUSIONS

An extensive analysis of the dependence of the temperature coefficient of the drain current has been presented for different MOS structures, both by experimental means and by numerical simulations. An analytical expression for the positive temperature coefficient of the Drain current has been developed and a model for the thermal instability in transient operation has been proposed.

It has been shown that the thermal instability can pose serious problems for the new generation of short channel, high current MOS structures using VLSI technology. The model proposed can explain the main causes of the thermal instability and give a some rules to evaluate the possible failure modes of operation for a given device.

REFERENCES

- [1] P.L.Hower, D.L.Blackburn, S. Rubin, "Thermal characterization of power transistors" IEEE Trans. on Electron Devices, vol. ED-23, pp. 831-838, 1976
- [2] G. Breglio, F. Frisina, A. Magri, p. Spirito "Electro-thermal instability in low voltage power MOS: experimental characterization" Proc. ISPSD '99, pp. 233-236, Toronto, 1999
- [3] P. Hower, C. Tsai, S. Merchant, T. Efland, S. Pendharkar, R. Steinhoff, J. Brodsky, "Avalanche induced thermal instability in LDMOS transistors", ISPSD 2001, pp. 153-156
- [4] G. Breglio, S.Pica, A. Tartaglia, P. Spirito, "Experimental detection of time dependent temperature maps in power bipolar transistors" Proc. of '97 Terminic Workshop, pp., Cannes, France, 1997
- [5] G. Breglio, N. Rinaldi, P. Spirito, "Thermal mapping and 3D numerical simulation of new cellular power MOS affected by electro-thermal instability" Microelectronics J., Vol. 31/9-10, pp. 741-746, 2000
- [6] P. Hower, G. Govil, IEEE Trans. ED, ED-21, pp. 617- 623, 1974
- [7] N. Rinaldi, "On the modeling of the transient thermal behavior of semiconductor devices," IEEE Trans. on ED (in print)
- [8] N. Rinaldi, "Thermal analysis of solid-state devices and circuits: an analytical approach", *Solid-State Electronics*, vol. 44, pp. 1789-1798, 2000.