Effect of Static Power Dissipation in Burn-in Environment on Yield of VLSI

Arman Vassighi
ECE Department
Univ.of Waterloo
Waterloo,Canada

Oleg Semenov
ECE Department
Univ.of Waterloo
Waterloo,Canada

Manoj Sachdev
ECE Department
Univ.of Waterloo
Waterloo,Canada

Ali Keshavarzi
Research Labs
Intel Corp.
Hillsboro, OR

Abstract

The leakage power is expected to increase with scaling of CMOS technology. The increased leakage is a strong function of the elevated temperature and voltage stress. As a consequence, under the burn-in (BI) conditions the elevated leakage power may cause increased post Burn-in fallout. In this paper the impact of elevated leakage and technology scaling in burn-in environment on post BI yield is analyzed. We have also shown that to maintain a constant post-BI yield loss, the burn-in temperature should go down by 10°C for each technology generation. We also show that at static burn-in conditions, the die temperature is increased exponentially and range of optimal stressed voltage and temperature for fixed post burn-in yield loss is reduced significantly, when CMOS technology is aggressively scaled down.

1: Introduction

Burn-in (BI) is an important test technique to weed out infant mortality from the main population. As a result, quality and reliability of outgoing Integrated Circuits (ICs) are improved. During the BI, ICs are subjected to elevated voltage and temperature stress. As a consequence, the field and temperature enhanced failure mechanisms are accelerated [1]. Optimization of effectiveness of BI test has been the constant focus of quality and reliability engineers.

As we scale the transistor to deep sub-micron regime, it's off state leakage increases significantly. A linear reduction in transistor threshold voltage with technology scaling, results in exponential increase in its leakage. This leakage is further increased under voltage and temperature stress conditions. High leakage causes further elevation in chip temperature and eventually leads to a positive feedback. Therefore, handling the power dissipation, becomes an important consideration during the burn-in. In this article, we investigate static power dissipation as a function of temperature and voltage. The objective is to minimize the burn-in temperature and voltage, while maintaining constant yield during the burn-in test for future technologies.

1063-6722/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY

2: Impact of CMOS technology scaling on junction temperature increase in burn-in conditions

The die temperature, commonly referred as the junction temperature or T_j, is defined as [2]:

$$T_j = T_a + P \times \theta_{ja} \tag{1}$$

$$P = P_{Dynamic} + P_{Static} \tag{2}$$

$$P_{Static} = I_{Leak} \times V_{dd} \tag{3}$$

where T_a is the ambient or set point temperature, P is the device total power, and θ_{ja} is the junction-to-ambient thermal resistance. In equation 2, the dynamic power is $C.V_{dd}^2.f$ where C is the total circuit capacitance and f is clock frequency. Since in BI environment, the device operates at lower frequency (often $\frac{f_{nominal}}{1000}$), the dynamic portion of power dissipation can be ignored. On the other hand, the leakage power component is elevated. Combining all leakage components, and using curve fitting techniques, it can be shown that transistor leakage is an exponential function of the die temperature:

$$I_{off} = f(T_j) = Ae^{\frac{(T_j+B)(T_j+C)}{D}} \tag{4}$$

where A is a prefactor and B, C, and D are constants which are used to fit the function to the experimental data. The power of the high performance circuits is increasing with respect to coming technology generations and it can be safely concluded that the elevated temperature and voltage results into excessive power dissipation causing further increase in die temperature. In other words, resulting in a positive feedback mechanism that may eventually lead to a thermal runaway. Therefore experts have been trying to limit the leakage current using techniques such as back body biasing [3], dual-threshold transistors [4], and dynamic threshold transistors [5][6], to reduce the static power dissipation.

Recently, a relationship between thermal resistance of a MOSFET and its geometrical parameters was derived using 3-D heat flow equation [7].

$$\theta_{ja} = \frac{1}{2\pi K}[\frac{1}{L}Ln(\frac{L+\sqrt{W^2+L^2}}{-L+\sqrt{W^2+L^2}}) + \frac{1}{W}Ln(\frac{W+\sqrt{W^2+L^2}}{-W+\sqrt{W^2+L^2}})] \tag{5}$$

Where K is the thermal conductivity of silicon ($K = 1.5 \times 10^{-4} W/\mu m^{\circ}C$ [7]), L and W are channel geometry parameters. The thermal conductivity of the silicon exhibits weak temperature dependence described as [8]:

$$K = 154.86 \times (300/T)^{4/3} \qquad (W/m/K \tag{6}$$

However, in our research we assume that the thermal resistance of silicon is temperature independent [7,9]. Considering temperature dependence of silicon thermal conductivity is more important in silicon on insulator (SOI) technologies where self-heating contributes to rise in junction temperature. We will use this model (5) for thermal resistance calculation for MOSFETs in different CMOS technologies.

1063-6722/02 $17.00 © 2002 IEEE

IEEE COMPUTER SOCIETY

In this work we designed an inverter where HSPICE simulations were carried out with BSIM model level 49. The load of inverters was the standard load element (N-MOSFET), which is used in Taiwan Semiconductor Manufacturing Corporation (TSMC) for inverter ring-oscillators simulations. The sizes of load elements were adopted from TSMC SPICE models file specified for each of analyzed CMOS technology. From Eq. 5, thermal resistance of an average transistor is computed. The average size of a transistor is achieved by averaging the NMOS and PMOS transistor widths. Since, the transistor dimensions are reduced with scaling, the thermal resistance is increased. Fig. 1 illustrates inverter power dissipation and thermal resistance of an average transistor as functions of technology.

The $0.35\mu m$ CMOS technology was used as the reference technology in this investigation. Furthermore, from Eq. 1 ΔT is defined as the temperature difference between junction and the ambient. Considering ΔT as unity for $0.35\mu m$ technology, we may calculate the normalized change in ΔT with respect to the reference technology. Using Eq. 1 and data presented in Fig. 1, we estimated the normalized average temperature increase of average size transistor for different technologies. The following principle was used for the calculations:

$$\frac{\Delta T_{0.25-CMOS}}{\Delta T_{0.35-CMOS}} = \frac{(T_j - T_a)_{0.25-CMOS}}{(T_j - T_a)_{0.35-CMOS}} = \frac{(P \times \theta_{ja})_{0.25-CMOS}}{(P \times \theta_{ja})_{0.35-CMOS}} \quad (7)$$

In order to estimate the average normalized temperature increase of CMOS chip, we must also consider increase in transistor density with scaling. The transistors density numbers for microprocessors were adopted from International Technology Road map for Semiconductors (ITRS) [10,11]. The normalized temperature increase of CMOS chip with technology scaling was calculated by multiplication of temperature increase per transistor and transistor density divided by two (in fully static burn-in we assume that half of transistors are off and half of them are on). To estimate the average power consumption of inverters for different operating conditions and technologies, we simulated these inverters at different temperatures and V_{DD}. For the static burn-in testing, we assume that the stressed temperature is changed from $25°C$ to $125°C$, the stressed voltage is changed from nominal V_{DD} for the given technology to $V_{DD} + 30\%$, and inverter inputs are grounded. The simulation (I_{av}) and calculation results (P, ΔT) are given in Table 1. In this table, I_{av} and P are the average current and power dissipation of inverter and ΔT is the (Tj - Ta) per 1 mm^2 of chip area. The obtained results are presented in Fig. 2. In this calculation, we assumed that the ambient temperature is the same for all analyzed technologies.

For static burn-in conditions, from Fig. 2 we can conclude that the estimated average junction temperature increase in CMOS chip should be ~ 70 times higher for $0.13\mu m$ CMOS technology in comparison with $0.35\mu m$ CMOS technology. This junction temperature increase with technology scaling is the result of drastic standby leakage power increase, the higher transistors density in advanced CMOS dies and the thermal resistance increase of scaled MOSFETs.

3: Time to breakdown and yield under thermal stress

The yield during the BI process is of critical importance. Therefore, BI parameters should be evaluated to have a constant predictable yield.

1063-6722/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY

The logarithm of the time to breakdown t_{bd}, for gate oxide of a transistor, based on constant voltage test, has been shown to be [12]:

$$t_{bd} = \tau_o exp[\frac{G}{E_{ox}}] = \tau_o exp[\frac{GT_{ox}}{V_{ox}}] \quad (8)$$

where T_{ox} is the oxide thickness, V_{ox} is the voltage across the oxide, and G(350 MV/cm) and $\tau_o(1 \times 10^{-11}s)$ are the slope and intercept of the $\ln(t_{bd})$ versus $1/E_{ox}$ plot, respectively [12]. In this equation, τ_o and G are the only temperature dependent factors. These parameters, using Arrhenius equation form with activation energy, E_b, can be expressed as:

$$\tau_o(T) = \tau_o exp(\frac{-E_b}{k}[\frac{1}{T} - \frac{1}{300}]) \qquad G(T) = G(1 + \frac{\delta}{k}[\frac{1}{T} - \frac{1}{300}]) \qquad \delta = \frac{k}{G}\frac{dG(T)}{d(1/T)} \quad (9)$$

where k is Boltzmann's constant, T is the absolute temperature, and τ_o and G are the room temperature constants given earlier. In the temperature range between $25°C$ and $150°C$, δ and E_b have been determined to be 0.0167 and 0.28 eV, respectively [13].

Although the burn-in is related to the removal of infant mortality, it may affect the yield of semiconductor devices. This is due to the fact that defects will grow during burn-in and some of them will cause yield loss. The amount of the defect growth and yield loss depends on burn-in environment, such as stressed voltage, stressed temperature and burn-in time.

In order to choose burn-in conditions which maximize the burn-in yield while improving the projected failure rate, the defect distribution models and their growth model should be studied. For the gate oxide defects in a transistor with a width of W, the growth model for defects are given as follows [14]:

$$W_b = W \times \nu, \quad where \quad 0 \le \nu \le 1 \quad and \quad \nu = \frac{t_b}{\tau_o(T_b)} exp[-\frac{G(T_b)T_{ox}}{V_b}] \quad (10)$$

$$W_g = W \times u, \quad where \quad 0 \le u \le 1 \quad and \quad u = \frac{t}{\tau_o(T_o)} exp[-\frac{G(T_o)T_{ox}}{V_o}] \quad (11)$$

W_b and W_g are defect growths in gate oxide during burn-in and operation, V_b and T_b are the burn-in voltage and temperature, and V_o and T_o are the operation voltage and temperature respectively. Temperature-dependent expressions for $\tau_O(T)$ and $G(T)$ are given in (9). It has been shown in [14] that the yield loss during burn-in will be:

$$Y_{loss} = Y - Y_l = Y(1 - Y^{\nu/(1-\nu)}) \quad (12)$$

where Y_l is the yield after burn-in and equals to $Y^{1/(1-\nu)}$.

In Fig. 3 we have shown the yield loss percentage for two given yield (0.9 and 0.99) before burn-in in $0.18\mu m$ technology for maximum voltage stress ($1.4 \times V_{dd}$). It can be seen that for a given pre-BI yield, the post-BI yield loss will increase exponentially with increasing temperature. Furthermore, the pre-BI yield will play an important role in post-BI yield loss. The lower the pre-BI yield is, the higher will be the post-BI yield loss.

In Fig. 4 we illustrate the post-BI yield loss as a function of temperature, using Eq. 12, for three technology generations ($0.18\mu m$, $0.13\mu m$, and $0.10\mu m$). In this simulation, voltage stress of $1.4 \times V_{dd}$ and 12 hours BI time are used. In Table 2, the electric field across the gate oxide for above mentioned technologies is calculated. It can be seen that with scaling, electric field across the gate oxide is increasing. Hence the yield loss is expected

1063-6722/02 $17.00 © 2002 IEEE
IEEE COMPUTER SOCIETY

to increase. In order to compensate this, we need to reduce the burn-in temperature with each technology generation. In Fig. 4, the horizontal line shows that to maintain a constant yield loss during the burn-in for three technology generation we need to reduce the burn-in temperature by $10°C$.

Table 1. DC simulation (I_{av}) and calculation results ($P, \Delta T$) of CMOS inverter for different technologies.

		25°C	25°C	25°C	85°C	85°C	85°C	125°C	125°C	125°C
CMOS Technology	V_{DD} (V)	I_{av} (pA)	P (pW)	ΔT ($°C/mm^2$)	I_{av} (pA)	P (pW)	ΔT ($°C/mm^2$)	I_{av} (pA)	P (pW)	ΔT ($°C/mm^2$)
$0.35\mu m$	3.3	7.7	25	0.00071	0.07	0.23	0.0066	2.05	6.77	0.2
	3.8	9.2	35	0.00099	0.084	0.32	0.0091	2.15	8.17	0.23
	4.3	11.1	47.7	0.0014	0.11	0.47	0.014	2.27	9.76	0.28
$0.25\mu m$	2.5	19.3	48.3	0.0023	0.418	1.04	0.05	3.96	9.9	0.29
	2.9	22	63.8	0.0031	0.47	1.36	0.065	4.41	12.80	0.35
	3.3	25	81.3	0.0039	0.531	1.75	0.08	4.81	15.87	0.45
$0.18\mu m$	1.8	90.5	163	0.02	1.33	2.39	0.24	8.96	16.13	0.97
	2.1	101	210	0.022	1.48	3.08	0.31	9.75	20.48	1.23
	2.35	112	264	0.027	1.62	3.81	0.39	10.9	25.6	1.51
$0.13\mu m$	1.2	766	920	0.2	8.45	10	2.32	28	34	7.79
	1.4	1200	1680	0.38	12.3	17	3.94	34	47	10.97
	1.56	1860	2900	0.67	17.7	27.6	6.4	55	85	19.81

Table 2. $1/E_{ox}$ for different technologies.

Technology	V_{dd} (V)	T_{ox} ($A°$)	$1/E_{ox}$ (cm/V)
$0.18\mu m$	1.8 V	$30A°$	$11.90 \times 10^{-8} cm/V$
$0.13\mu m$	1.3 V	$20A°$	$11.80 \times 10^{-8} cm/V$
$0.10\mu m$	1.0 V	$15A°$	$11.70 \times 10^{-8} cm/V$

Now let us assume that post burn-in yield loss for advanced CMOS technologies should not be worse than post burn-in yield loss for $0.35\mu m$ CMOS technology. It means that the junction temperature increase over ambient temperature during burn-in testing for advanced technologies should not be higher than the burn-in junction temperature increase for $0.35\mu m$ CMOS technology. From Table 1 for $0.35\mu m$ CMOS technology, the junction temperature increase (ΔT) over ambient stressed temperature per 1 mm^2 of chip is $0.28°C$ at $V_{DD} = 4.3V$ and $T = 125°C$. If we plot ΔT versus stressed temperature for different stressed voltages, we can find the optimal burn-in temperature and voltage when $\Delta T = 0.28°C/mm^2$ for other CMOS technologies. For example, Fig. 5 presents this technique for $0.25\mu m$ CMOS technology. Data for this graph was used from Table 1. Similarly, we can find the optimal burn-in temperature and voltage for other technologies, using data of Table 1. The obtained results are shown in Fig. 6. In this figure, the optimal burn-in temperature and voltage are presented for different technologies, at which the average junction temperature increase of die for these technologies is the same as the average

Proceedings of the 17th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'02)
1063-6722/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY

junction temperature increase for 0.35μm CMOS technology. In this case, we can expect that the post burn-in yield loss for scaled CMOS technologies has the same value as the post burn-in yield loss for 0.35μm CMOS technology.

4: Conclusion

In this work, post-BI fallout was considered for the setting of optimal BI conditions. Under the burn-in conditions, the elevated leakage power may cause increased post Burn-in fallout. We have shown that to maintain a constant post-BI yield loss, the burn-in temperature should be reduced at least by $10°C$ for each technology generation. We have also calculated the thermal resistance of MOSFET with the typical average size for the given technology and high performance design. Using the obtained result and typical transistors density data for microprocessors, we calculated the average junction temperature increase over ambient temperature under BI conditions for hypothetical high performance CMOS die, implemented in different CMOS technologies. Finally we showed that the range of optimal stressed temperature and voltage in burn-in environment is significantly reduced with technology scaling. The effectiveness of voltage acceleration factor is increased faster with technology scaling in comparison with the increase of effectiveness of thermal acceleration factor.

References

[1] R.-P. Vollertsen, "Burn-In", IEEE International Integrated Reliability Workshop Final Report, 1999, pp. 167 -173.

[2] P. Tadayon, "Thermal Challenges During Microprocessor Testing", Intel Tech. Journal Q3, 2000.

[3] A. Keshavarzi, S. Narendra, C. Hawkins, K. Roys and V. De, "Technology Scaling Behaivior of Optimum Body Bias for Standby Leakage Power Reduction in CMOS ICs", 1999, ISLPED, pp. 252-254.

[4] L. Wei, Z. Chen, K. Roy, M. C. Johnson, Y. Ye, and V. K. De, "Design and Optimization of Dual-Threshold Circuits for Low-Voltage Low-Power Applications", IEEE Transaction (VLSI) Systems, Vol. 7, No. 1, March 1999, pp. 16-24.

[5] F. Assaderaghi, "Dynamic Threshold-Voltage Mosfet (DTMOS) for Ultra-Low Voltage VLSI", IEEE Transactions On Electron Devices, Vol.44, No.3, March 1997, pp. 414-422.

[6] F. Assaderaghi, "DTMOS: Its Derivatives and Variations, and Their Potential Applications", The 12th International Conference on Microelectronics, 2000, pp. 9-10.

[7] N. Rinaldi, "Thermal Analysis of Solid-State Devices and Circuits: an Analytical Approach," Solid-State Electronics, vol. 44, No. 10, pp. 1789-1798, 2000.

[8] D.L. Blackburn and A.R. Hefner, "Thermal Components Models for Electro-Thermal Network Simulation," Proc. of 9th IEEE SEMI-THERM Symposium, 1993, pp. 88-98.

[9] N. Rinaldi, "On the Modeling of the Transient Thermal Behavior of Semiconductor Devices," IEEE Trans. on Electron Devices, vol. 48, No. 12, pp. 2796-2802, 2001.

[10] D.P. Valett and J.M. Soden, "Finding Fault With Deep-submicron ICs," IEEE Spectrum, pp. 39-50, October 1997.

[11] International Technology Roadmap for Semiconductors, http://public.itrs.net/

[12] R. Moazzami and C. Hu, "Projecting Gate Oxide Reliability and Optimizing Reliability Screens" IEEE Transactions on Electron Devices, vol. 37, NO 7, July 1990, pp. 1643-1650.

[13] R. Moazzami, J. Lee, and C. Hu, "Temperature Acceleration of Time Dependent Dielectric Breakdown", IEEE Trans. Electron Device, Vol. 36, No. 11, Nov. 1989, pp. 2462-2465.

[14] T. Kim, W. Kuo, and W. K. Chien, "Burn-in Effect on Yield", IEEE Transactions on Electronics Packaging Manufacturing, Vol. 23, No. 4, Oct 2000, pp. 293-299.

1063-6722/02 $17.00 © 2002 IEEE

IEEE COMPUTER SOCIETY

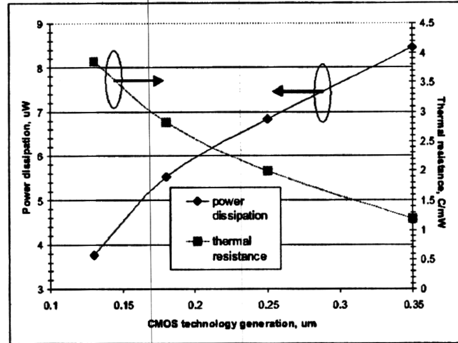

Figure 1. Inverter power dissipation (P) and transistor thermal resistance (θ_{ja}) versus CMOS technology scaling.

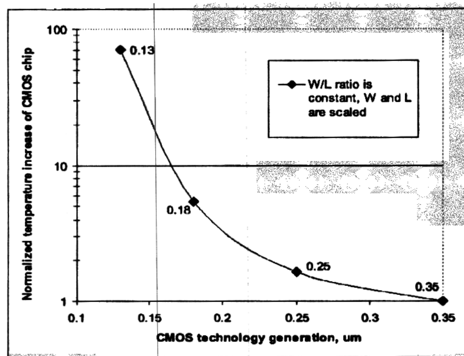

Figure 2. Impact of technology scaling on normalized temperature increase of CMOS logic chip at burn-in conditions ($V_{DD} + 30\%, T = 125°C$).

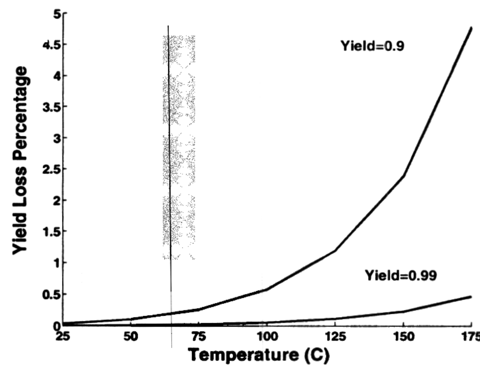

Figure 3. Post BI yield loss percentage vs. BI temperature for two pre-BI yields.

Proceedings of the 17th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'02)
1063-6722/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY

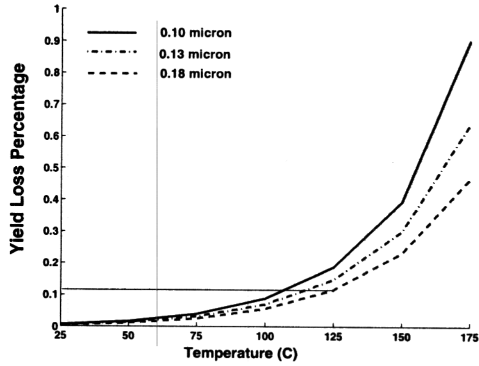

Figure 4. Hypothetical post BI yield loss percentage for different technology and pre-BI yield of 0.99.

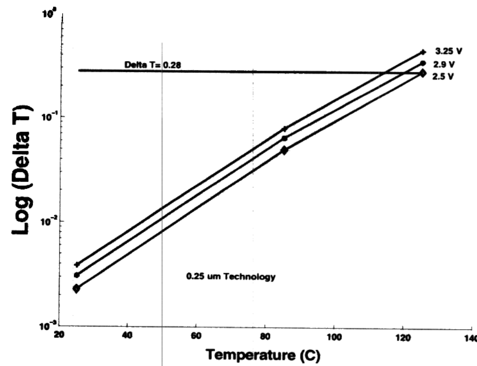

Figure 5. Junction temperature increase over ambient stressed temperature per 1 mm^2 chip area versus stressed temperature and different V_{DD}.

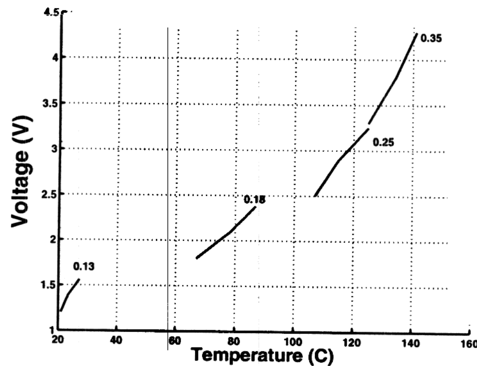

Figure 6. Optimized burn-in voltage and temperature for constant burn-in loss, which is equaled of burn-in loss of $0.35\mu m$ CMOS technology.

1063-6722/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY