Chapter 2

POWER, JUNCTION TEMPERATURE, AND RELIABILITY

Abstract: In this chapter, we initially discuss the power consumption trends and its impact on junction temperature elevation. Junction temperature plays a pivotal role in determining long term integrated circuit reliability. Therefore, we discuss the impact of junction temperature on the reliability of *CMOS* integrated circuits.

Key words: *CMOS*, Junction Temperature, Reliability, Leakage Power, Failure Mechanisms.

1. POWER IN NANOMETER REGIME

The power of an integrated circuit (*IC*), for a fixed operating voltage and temperature, increases linearly with the clock frequency (the frequency of a master signal with which all operations must be synchronized) driving the *IC*. Extrapolation of the power vs. frequency response down to a frequency of zero (which may be realized in a sleep mode) yields a non-zero power, which is referred to as the static power, P_{static}. Increasing amount of attention

is paid on to the techniques to reduce the overall power consumption, and an interested reader is referred to several outstanding work in this domain [1]. However, we briefly discuss various dynamic and leakage current components in subsequent sub-sections. This discussion will be useful in building intuitive understanding for following sections in this chapter.

1.1 Dynamic Power

The component of power which is proportional to the frequency is referred to as the dynamic power, $P_{dynamic}$. The dynamic power is due primarily to the charging and discharging of capacitances in the *IC*, and can be represented by an effective switching capacitance, *C*, via the well known relationship,

$$P_{dynamic} = C \cdot V_{DD}^2 \cdot f \tag{2.1}$$

In this equation *C* does not necessarily represent the actual total capacitance being switched by the chip since many of the circuits may be switching at some fraction of *f* (or, for that matter, at some multiple of *f*). Furthermore, another source of active power, sometimes referred to as short-circuit; shoot-through, or crossover power, is also lumped into *C*. This short-circuit power is due to current which completes a path from the power supply node to ground directly through a network of *n*-type and *p*-type *FETs* during the short but finite time interval when the gates are close to $V_{DD}/2$, and hence both *n*- and *p*-type *FETs* are in a conducting state. Typically this component represents several percent of the active power.

1.2 Static Power

The standby current density increases exponentially as the channel length is decreased. This follows from the demand that V_{TH} decreases with V_{DD}, to maintain a high drive current and achieve the performance improvement. This current causes an additional power demand in the operation of *CMOS* which is often referred to as static power, since, unlike switching, or active power, static power is dissipated by all *CMOS* circuits all of the time, whether or not they are actively switching.

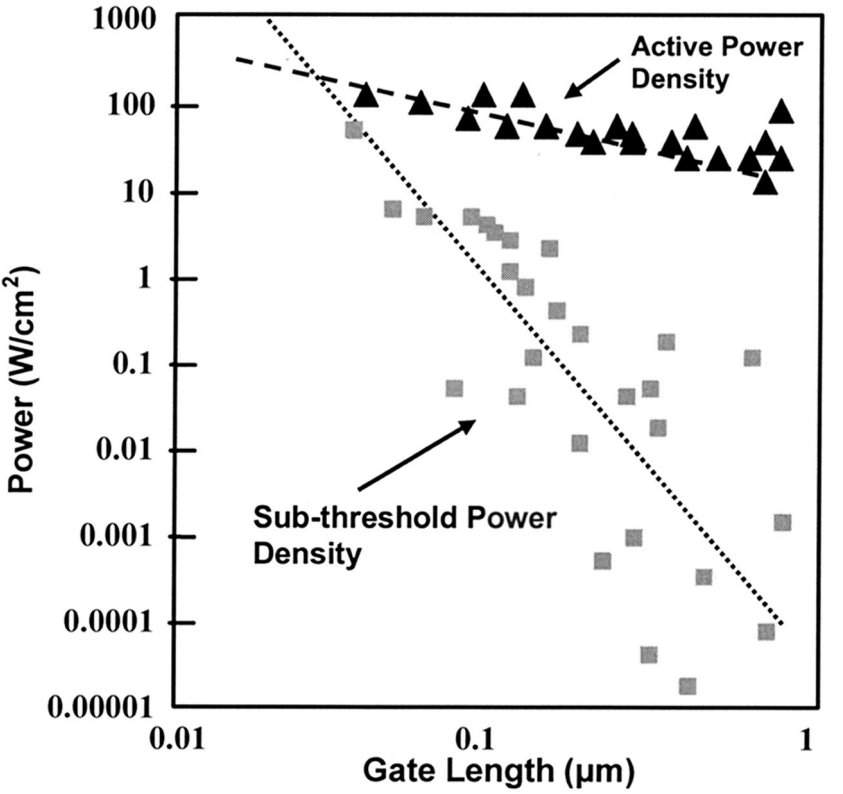

Figure 2-1. Static and dynamic power trends, vs. L_{gate} for a junction temperature of 25°C [2].

Figure 2.1 illustrates the static power trend based on sub-threshold currents calculated from the industry trends of V_{TH} , all for a junction temperature of T_j = *25°C*. More practical values of T_j only serve to exacerbate this situation, with the off current of *MOSFETs* rising nearly two times for each *10°C* increase in T_j. For reference, the active power density is shown in Figure 2.1 in the same scale to illustrate that the sub-threshold component of power dissipation is emerging to compete with the long battled active power component for even the most power-tolerant, high speed *CMOS* applications. Empirical extrapolation (dashed curves) suggest that sub-threshold power will equal active power at L_{gate} of 20nm and this point will be encountered closer to L_{gate} of 50nm at elevated temperatures [2].

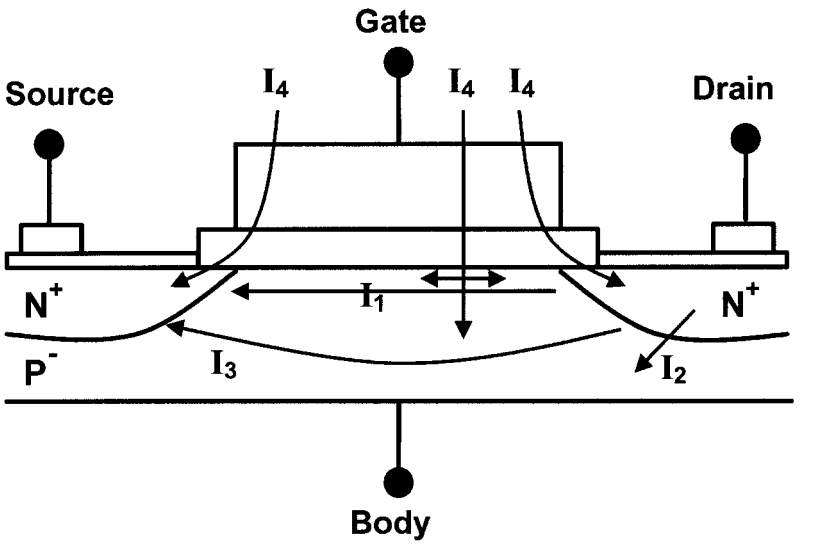

Figure 2-2. Figure 2.2: Leakage current mechanisms for *CMOS* transistors [3].

The total leakage current I_{OFF} is influenced by threshold voltage, channel physical dimensions, channel/surface doping profile, drain/source junction depth, gate oxide thickness, and V_{DD}. Understanding the different components of leakage current is a necessity for developing techniques to effectively reduce the off-state leakage. Figure 2.2 shows these leakage current mechanisms [3].

1.2.1 Subthreshold Leakage (I_1)

Subthreshold leakage is the weak inversion conduction current that flows between the source and the drain of a *MOS* transistor when gate voltage is below the threshold voltage, V_{TH}, [4]. Unlike the strong inversion region in which drift current dominates, the subthreshold conduction is dominated by the diffusion current. In a similar manner to charge transport across the base of a bipolar transistor, carriers move by diffusion along the surface. Weak inversion or subthreshold current typically dominates modern device off-state leakage due to low V_{TH}.

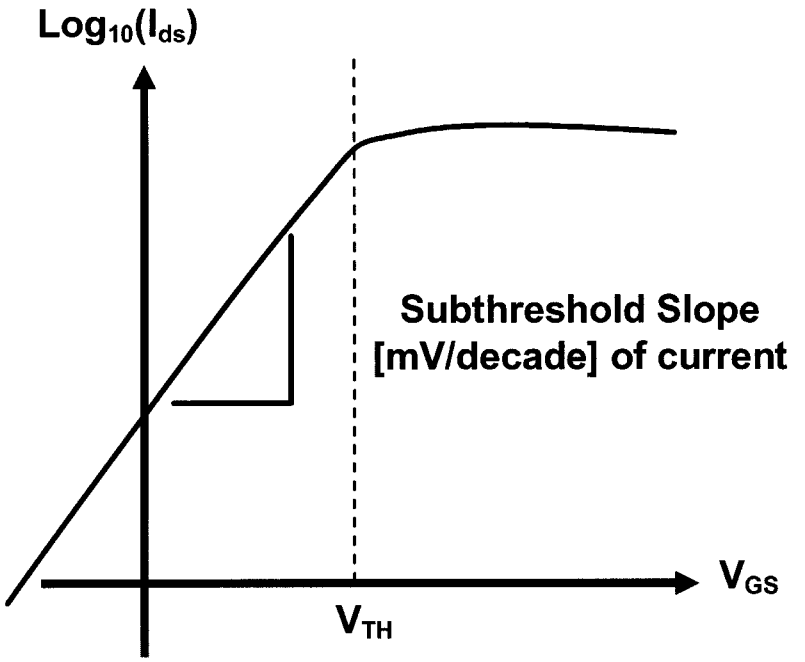

Figure 2-3. Subthreshold leakage in *NMOS* transistor.

Subthreshold current is exponentially related to the gate voltage as illustrated in Figure 2.3. The inverse of the slope of the $log_{10}(I_{ds})$ versus V_{gs} characteristic is called the subthreshold slope (S_t) [4]. Subthreshold slope indicates how effectively the transistor can be turned off when V_{gs} is reduced below V_{TH} thus it is desirable to reduce S_t. Typical values of S_t for bulk *CMOS* technology range from 70mV/decade to 120mV/decade (70 mV/decade means that 70 mV reduction in gate voltage reduces the subthreshold current by one order of magnitude).

Subthreshold current is an exponential function of the drain voltage. However, in long channel devices this dependency is very weak for very low V_{DS} (in the order of few $V_t = kT/q = 26mV$ at 25^0C). The subthreshold leakage increases with V_{DS} increase due to drain induced barrier lowering (*DIBL*) effect. When in short channel devices a high V_{DS} is applied to the drain, it lowers the potential barriers height. As a consequence more carriers can flow and give rise to higher subthreshold current. *DIBL* does not change the subthreshold slope (S_t), but it lowers the threshold voltage. Figure 2.4 shows that *DIBL* effect shifts the I_{DS}-V_{GS} curve to the left and to the top [3]. Devices with shorter channel length suffer more from *DIBL* effect. The

effect of *DIBL* on subthreshold current can be reduced by increased surface and channel doping and shallower source/drain junction depths.

1.2.2 Band-to-Band Tunneling Current (I2)

If a high electric field is applied to a reverse biased *pn* junction, electrons tunnel across the junction giving rise to tunneling current. The electrons tunnel from the valance band of the *p*-region to the conduction band of the *n*-region as it is shown in Figure 2.5. Tunneling occurs when the voltage drop across the reverse biased *pn* junction is greater than the semiconductor band-gap.

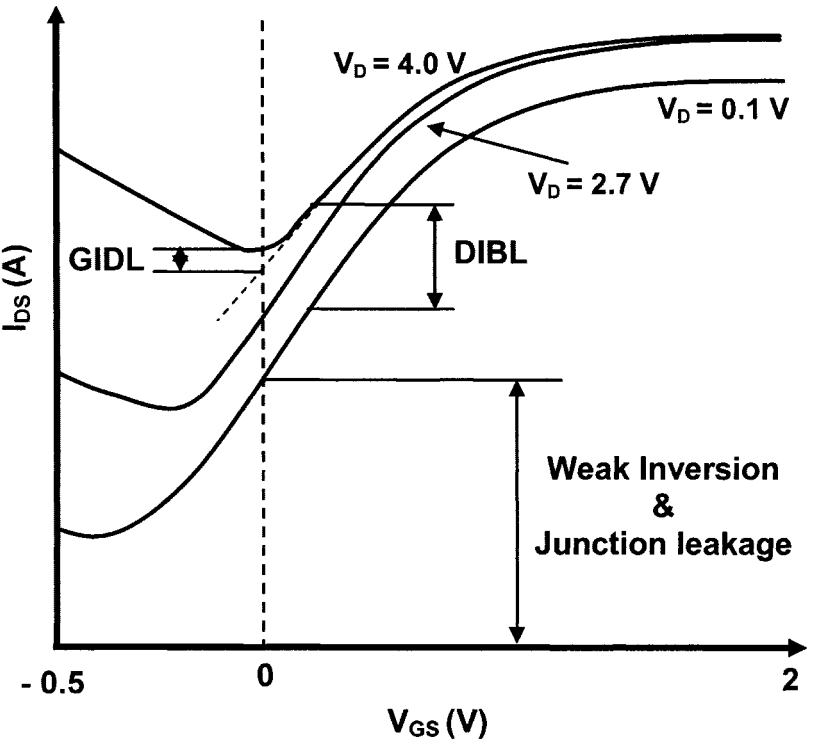

Figure 2-4. Figure 2.4: *n*-channel I_{DS} vs. V_{GS} showing *DIBL, GIDL*, weak inversion and *pn* junction leakage components [3].

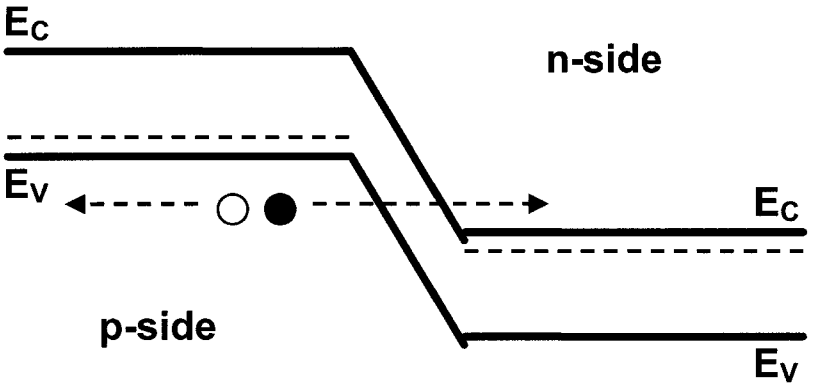

Figure 2-5. Band-to-Band tunneling in reversed bias *pn* junction[5].

When the drain (or source) of an *NMOS* device is biased at a higher voltage with respect to the substrate, *BTBT* current flows from the drain (or source) to substrate through the reverse biased *pn* junction [5]. In scaled *MOSFET* technologies, *p* and *n* regions are heavily doped, and halo implant is utilized to reduce the short channel effects (*SCE*). Owing to the heavily doped nature of junctions, the *BTBT* current increases significantly. The *BTBT* current can be a major component of the transistor off current. Reducing substrate doping near the substrate-drain/source junction is an effective method to reduce *BTBT* current, but it increases the short channel effects and leads to higher subthreshold leakage. An application of small forward body bias reduces the electric field across the junction, and hence reduces the *BTBT* current [6].

1.2.3 2.1.2.3 Punch through (I3)

As the channel becomes shorter in scaled *CMOS* technologies, the depletion layer widths of source and drain junction become comparable to the channel length. Using an abrupt one dimensional approximation, the width of the source junction W_S and that of the drain junction W_D are given as [7].

$$W_S = \sqrt{\frac{2\varepsilon_S}{qN_A}(V_{bi} + V_{BS})} \tag{2.2}$$

$$W_D = \sqrt{\frac{2\varepsilon_S}{qN_A}(V_{DS} + V_{bi} + V_{BS})} \tag{2.3}$$

Where ε_S is the permittivity of silicon, q is the electronic charge, N_A is doping of the drain and source region, V_{bi} is the built in potential and V_{DS} is drain to source voltage, and V_{BS} is the body to source voltage. If the transistor is off and if the V_{DS} is increased, at certain voltage the drain and source depletion layers will merge (W_S+W_D = L) resulting in the punch through. Under these conditions, the gate loses the control over the channel. Therefore, punch through is a major limitation of device operation in short channel *MOSFETs* and often retrograde implant is used to prevent the occurrence of the punch through.

1.2.4 Gate Oxide Tunneling (I4)

There are mainly two gate oxide leakage mechanisms between gate and the substrate of a *MOSFET*. One is the Fowler-Nordheim (*FN*) and another is the direct tunneling. Gate oxide leakage in scaled technologies is mainly due to the direct tunneling. In thin oxide layers (less than 3-4 nm), electrons tunnel directly from the silicon surface to the gate through the forbidden energy gap of the SiO_2 layer [5]. There are five different mechanisms that contribute to the direct tunneling current in *MOSFETs*. Two of these mechanisms cause the leakage between gate and source; and gate and drain extension overlap region, respectively. Next two mechanisms contribute leakage between the gate; and source and drain, respectively through the channel. Finally, the last component is the gate to substrate leakage current. The modeling of each of these components can be found in [8, 9].

2. LEAKAGE REDUCTION TECHNIQUES

As illustrated in Figure 2.1, the leakage power is becoming significant component of the total power and may contribute to majority of the power dissipation in future *CMOS* technologies [10]. Therefore, in last decade and half a lot of attention has been paid on low power circuit and process techniques to reduce the elevated power consumption. Figure 2.6 shows the static or leakage power and dynamic power of the Intel microprocessors in different *CMOS* technologies. It is clear that the leakage power is increasing exponentially with the scaling.

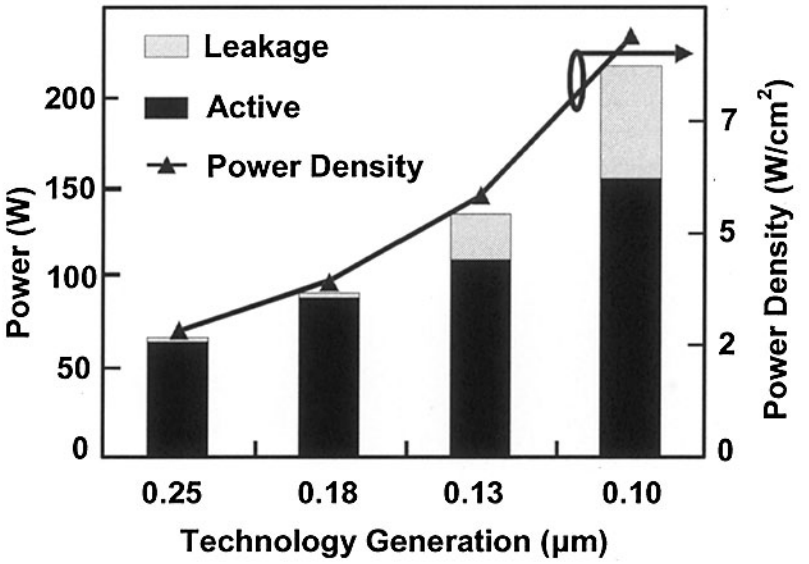

Figure 2-6. Static and dynamic power trends for Intel microprocessors [10].

Leakage current reduction can be achieved by utilizing either process techniques or circuit techniques. For low power applications combination of these two techniques are applied to reduce the leakage current. While at process level optimizing the device physical dimensions (length, oxide thickness, junction depth, etc) and the device doping profile lead to leakage current reduction, at circuit level, optimum biasing of the transistor terminals and other circuit techniques reduces the leakage current under different operating conditions. It must be noted that some of the leakage mechanisms like punch through can be optimized at process level while others can be optimized either at process level or at circuit level.

2.1 Process Level Leakage Reduction Techniques

Channel engineering is the process level technique which is used to reduce the leakage current while maximizing the linear and saturation drive currents. Changing the device doping profile in the channel region affects the electric field and potential contours and consequently changes the different current components. Steep retrograde wells and halo implants have been used to increase the drive current without affecting the off-state leakage current. Figure 2.7 shows the device with retrograde well and halo implants [6]. In a retrograde well structure, the doping concentration near the surface

is low and it increases in subsurface region. This results in higher channel mobility on the surface and prevents punch through in the subsurface region.

Halo doping is a laterally non-uniform channel profile which controls the dependence of the threshold voltage on the channel length. In halo doping, the doping near the two edges of the channel is increased by the injection of point defects during sidewall oxidation, which gathers doping impurities from the substrate. As the channel length decreases, these highly doped regions consume a larger fraction of the total channel width, therefore reduce the width of the depletion region in the drain-substrate and source-substrate regions and guard against the threshold voltage degradation and leakage current increase. Although the halo implants reduce the subthreshold leakage current, they increase the band-to-band tunneling current in reverse bias *pn* junction due to increased doping.

2.2 Circuit Level Leakage Reduction Techniques

Often process level leakage reduction techniques are not enough to achieve low leakage objectives in contemporary VLSIs. Hence, circuit level leakage reduction techniques are utilized. In general, circuit level techniques offer greater flexibility and can be optimized for specific applications. Moreover, these techniques can be utilized to reduce the leakage current of the circuit either under nominal or stress conditions. Stress conditions refer to reliability screening test environment, where *ICs* are subjected to higher junction temperature and higher supply voltage.

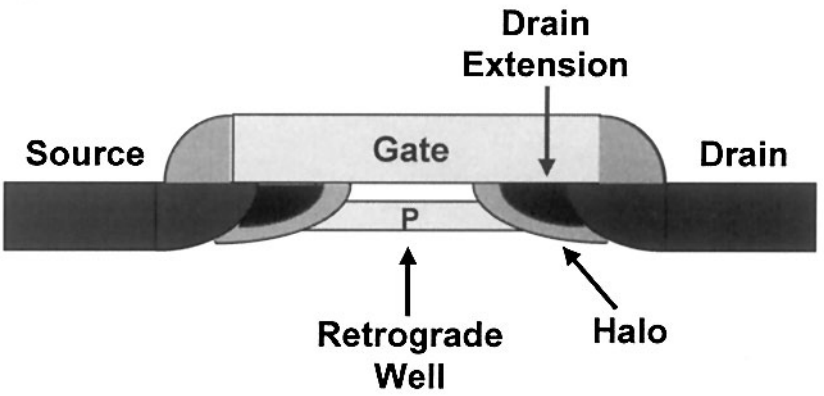

Figure 2-7. Different aspect of well engineering [6].

All the leakage reduction techniques under nominal conditions reduce the leakage current under nominal and stress conditions. There are some techniques that are designed specially for stress conditions, so under nominal conditions these techniques are disabled through proper signaling.

Major research has been carried out on low power and leakage current reduction [11-18]. The transistor I_{OFF} comprises of several different components [17, 3] of which the subthreshold current is the most dominant in scaled technologies. Subthreshold current of a *CMOS* transistor is modeled as follow:

$$I_{SUB} = A \cdot \exp\left(\frac{V_{GS} - V_{THo} - \gamma V_{SB} + \eta V_{DS}}{nV_t}\right)\left(1 - \exp\left(\frac{-V_{DS}}{V_t}\right)\right) \quad (2.4)$$

In this equation $A=\mu_o C_{ox} W/L_{eff} V_t^2 e^{1.8}$, $\mu_{o\ is}$ the zero bias carrier mobility, C_{ox} the gate oxide capacitance, L_{eff} is the transistor effective channel length, W is the transistor width, η is the drain induced barrier lowering coefficient, γ is the linearized body effect coefficient, n is the transistor subthreshold swing coefficient and V_t is the thermal voltage given by KT/q (~33mV at 110°C). In addition, V_{TH0}, V_{GS}, V_{SB} and V_{DS} denote the transistor zero-bias threshold voltage, gate-source, source-body and drain-source voltages respectively. In the following sub-sections different techniques to reduce subthreshold current are described.

2.2.1 Non-Minimum L_{eff}

There is an exponential relationship between threshold voltage and effective channel length. Increasing the channel length will increase the threshold voltage. Figure 2.8 shows this exponential dependency. It must be noted that in very long channel lengths the dependency of threshold voltage on channel length decreases and extent of leakage reduction due to increased threshold voltage is limited by V_{TH} roll-off [14, 19].

In the region of interest, the threshold voltage increases almost linearly for small increases in the drawn channel length (L_{eff}). As a result, the increase in the transistor zero body bias threshold voltage is approximated as in Eq. (2.5) [19].

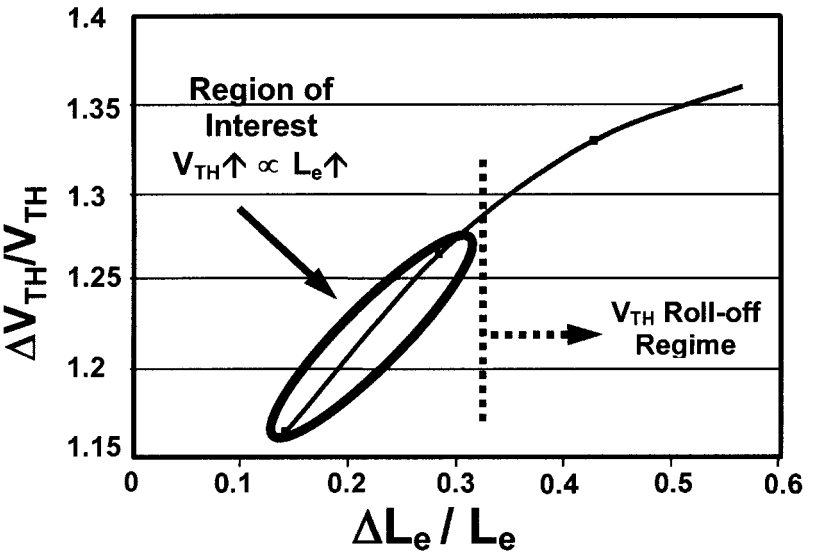

Figure 2-8. Normalized effective channel length vs. normalized threshold voltage.

$$\Delta V_{THo} = V_{THo}\left(\frac{\Delta L_{eff}}{L_{eff}}\right) \tag{2.5}$$

It must be noted that in nano-scale technologies, the channel mobility remains constant and independent of channel length due to velocity saturation. Therefore, the reduction in leakage current using non-minimum channel length transistors can be modeled as [19]:

$$\frac{\Delta I_{OFF}}{I_{OFF}} = 1 - \left[\frac{L_{eff}}{L_{eff} + \Delta L_{eff}} . \exp\left(\frac{-\Delta V_{THo}}{nV_t}\right)\right] \tag{2.6}$$

Where, ΔL_{eff} is the change in effective channel length.

2.2.2 Stack Effect

Another solution to the increasing leakage is placing a non-stack transistor on a stack of two transistors without affecting the input load [20]. It has been shown that stacking two off-transistors significantly reduces the sub-threshold leakage compared to a single off-transistor. The drawback of

this technique is the increased delay. This delay increase is comparable to high V_{TH} logic implementation in a dual V_{TH} technology. A significantly large fraction of the non-critical path implemented with this technique shows minimal performance degradation while reducing the sub-threshold leakage. The stack forcing technique can be either used in conjunction with dual V_{TH} or with a single V_{TH} technology [20].

Figure 2.9 shows a configuration of two stack transistors. The drain of the transistor *N1* is biased at V_{DD} and the source of the transistor *N2* is connected to the ground. The intermediate node voltage (V_N) reaches a steady state *DC* value for the two stack *N*-MOS pull down when both transistors are *OFF*. This value is within an order of magnitude of the thermal voltage (V_t). In 130 nm technology V_N is approximately 100mV and is due to $V_N = I_{OFF}R_{OFF}$ voltage drop across the bottom transistor (*N2*). As a result, a negative gate drive voltage (V_{GS}) appears across the top *N-MOS* transistor (*N1*) of the stack. Furthermore, there is a reduction in V_{DS} which suppresses the leakage current due to *DIBL* effect and also there is a negative body to source bias, V_{BS} (reverse body bias) which reduces the subthreshold current. Thus, the leakage current reduction using stack effect is as follow [19]:

$$\frac{\Delta I_{OFF}}{I_{OFF}} = 1 - \exp\left(\frac{-I_{OFF}.R_{OFF}.(1+\gamma+\eta)}{n.V_t}\right) \tag{2.7}$$

It must be noted that by using the stack transistor technique the drive current of the transistor *N1* is reduced accordingly. So it is important to place the stack transistors in the paths that charging the load is initial design target and discharging through the stack transistors with reduced drive current does not degrade the overall performance.

2.2.3 Reverse Body Bias (*RBB*)

The subthreshold leakage is reduced when the body of the transistor is biased to a negative voltage with respect to the source of the transistor ($V_{SB}<0$). The reduction in the leakage current is proportional to the extent of the applied reverse body bias. However, beyond a certain optimal reverse body bias voltage, the transistor off-state leakage current starts to increase due to increased band-to-band tunneling current as shown in Figure 2.10 [15]. The optimum *RBB* for sub-130 nm technologies is approximately 30% of the V_{DD}.

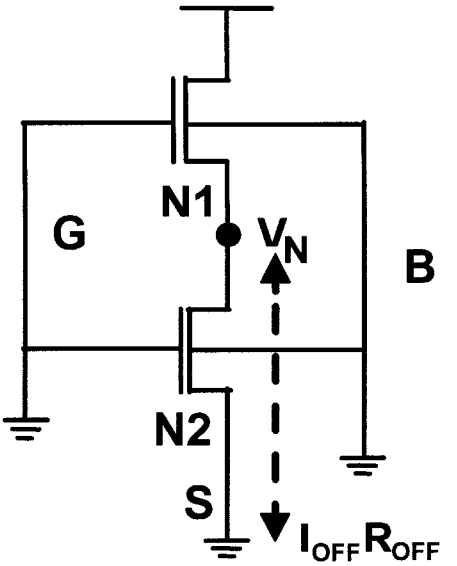

Figure 2-9. Leakage reduction using the stack effect.

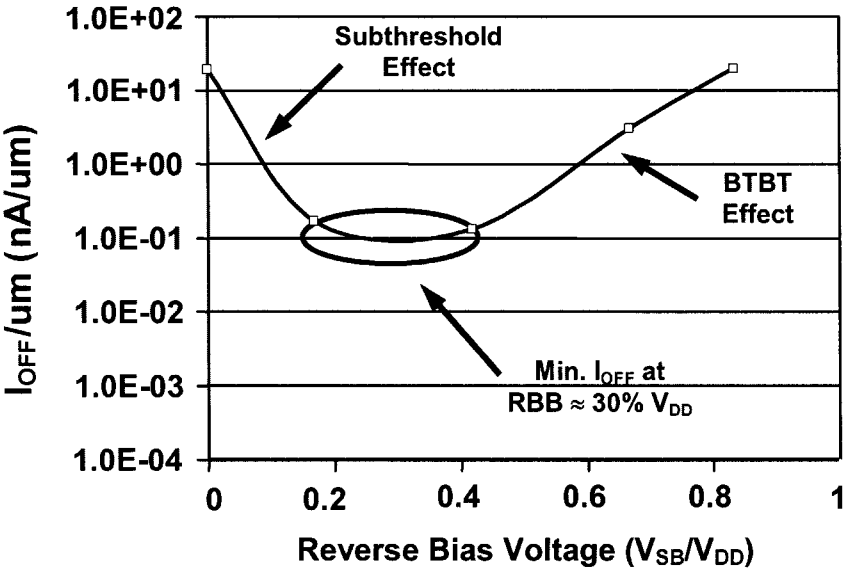

Figure 2-10. An optimum Reverse Body Bias (*RBB*) reduces the leakage current to its minimum [19].

For the region around the optimum reverse body bias, the leakage reduction can be modeled as [19]:

$$\frac{\Delta I_{OFF}}{I_{OFF}} = 1 - \exp\left(\frac{-\gamma . V_{SB}}{n . V_t}\right) \tag{2.8}$$

Reverse body bias technique is used to reduce leakage current during active operation, burn-in, as well as in standby mode. During active operation, *RBB* is applied to the idle portion of the chip to reduce overall chip leakage power without impacting the performance. Since the chip operational frequency is very low during burn-in, *RBB* can be applied to the whole chip simultaneously.

2.2.4 Multi-threshold Logic

This technique adjusts high performance critical path transistors with low V_{TH} while non-critical paths are implemented with high V_{TH} transistors. Hence, performance and power objectives are achieved at the cost of additional process complexity. Wei, et. al., reported a reduction of more than 80% in leakage power while meeting the performance objectives by using a dual V_{TH} technology [21].

Alternatively, a high V_{TH} transistor can be placed between power supply/ground and the high performance circuit or block (Figure 2.11). In the active mode, the high V_{TH} transistors are on and since their on-resistance is low, the performance impact is minimal. In the standby mode, the high V_{TH} transistor is off, and hence the leakage is limited to the leakage of a high V_{TH} transistor [22]. Traditionally, multi-threshold transistors are realized through different doses of threshold adjust ion implantations. Adjusting the threshold voltages can also be done by depositing two different oxide thicknesses or by different channel lengths [21].

2.2.5 Comparison of Leakage Reduction Techniques

Table 2.1 shows the leakage reduction achieved by using different techniques. In this table the theoretical models are compared with simulation results. Although theoretical models track simulation results for all techniques, they under estimate the reduction in leakage current. This is due to simplification in leakage reduction models. These results indicate that the stack effect technique reduces the leakage current by up to 12x while, the non-minimum L_{eff} technique (L_{eff} increased by 30%) reduces the leakage current by 9.3 x. Reverse body bias equal to 30% of V_{DD} reduces the leakage current by 2.2-2.3x [19].

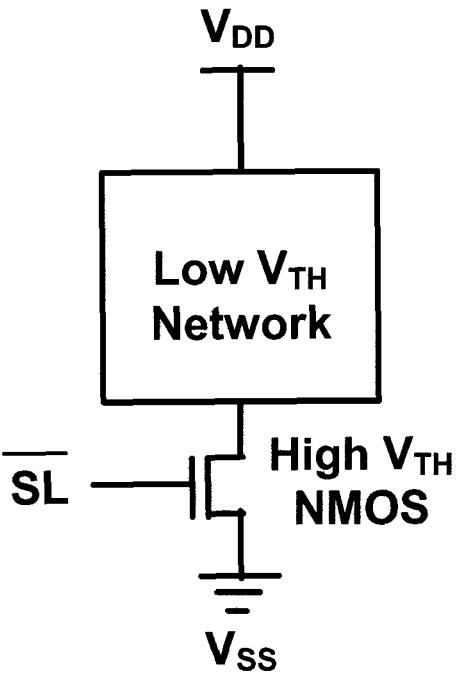

Figure 2-11. MTCMOS used to reduce the leakage of the low V_{TH} network.

Table 2-1. Leakage current reductions for 130nm technology [19].

Technique	Simulation Results	Theoretical Model
Non-minimum L_{eff} (L_{eff} +30%)	9.3x	8.7 x
Stack Effect	12.0 x	11.5 x
RBB (30% Reverse Bias)	2.3 x	2.1 x

2.2.6 Leakage Reduction and Impact on Performance

The leakage reduction is always desirable since it is wasted power. However, one must also look at the cost of reducing the power consumption with respect to its impact on performance. One of the most important trade-offs is leakage reduction and performance degradation trade-off. Design engineers often look at the I_{ON}/I_{OFF} ratio rather the absolute value of the I_{OFF} reduction when they examine any leakage reduction technique. The amount of I_{ON} determines the performance of the circuit and any reduction in I_{ON} leads to performance degradation.

Figure 2.12 shows the I_{ON}/I_{OFF} trends with technology scaling (solid lines) and also the threshold voltage trends with technology scaling (dash lines) for two cases of low and high threshold voltage. It can be seen that transistor $I_{OFF}/\mu m$ is increasing by 3-5x per generation resulting in the degradation of I_{ON}/I_{OFF} ratio with technology scaling. This results in excessive leakage currents for the 70 nm generation and offsets the reduction in switching energy obtained from scaling.

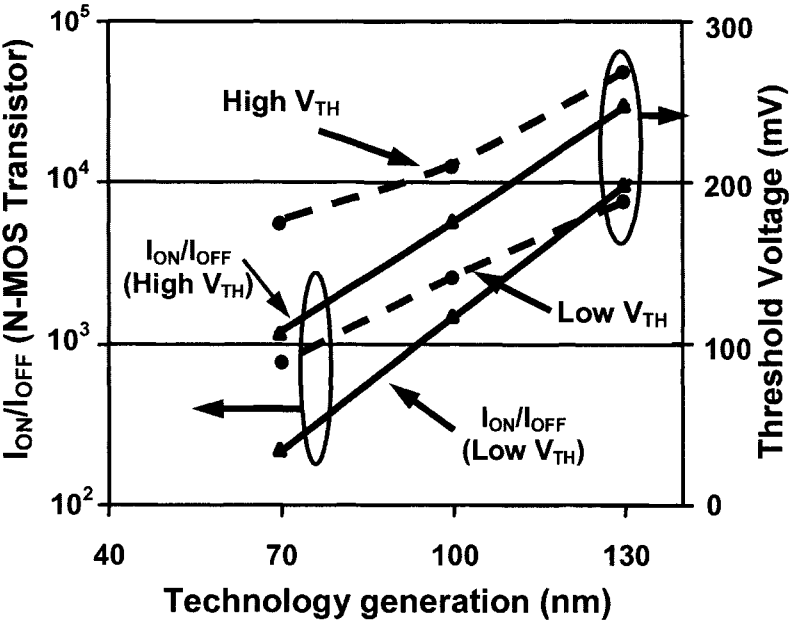

Figure 2-12. I_{ON}/I_{OFF} and V_{TH} scaling trends for typical corner device in 110°C and nominal V_{DD} [19].

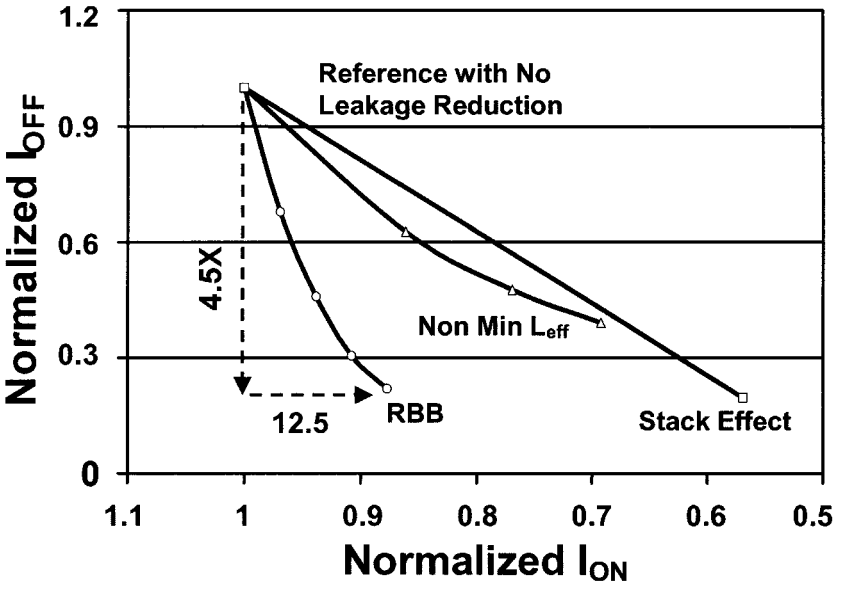

Figure 2-13. Comparison of leakage reduction techniques in 70 nm technology in 110°C [19].

Figure 2.13 quantify the I_{OFF} vs. I_{ON} tradeoffs for each leakage reduction technique for the 130nm and 70nm technologies. This figure indicates that both reverse body bias and non-minimum L_{eff} techniques result in lesser degradation of transistor than stack effect. Consequently, both reverse body bias and non-minimum L_{eff} techniques have steeper gradients in the I_{OFF}-I_{ON} plane making them more efficient in reducing leakage current for 130-70nm technologies [19].

2.2.7 Leakage Reduction Techniques and Scaling Trends

Figure 2.14 shows the effectiveness of leakage reduction techniques with respect to scaling. In this figure ΔI_{OFF} is the reduction in leakage and ΔI_{ON} is reduction in drive current where certain leakage reduction technique is incorporated.

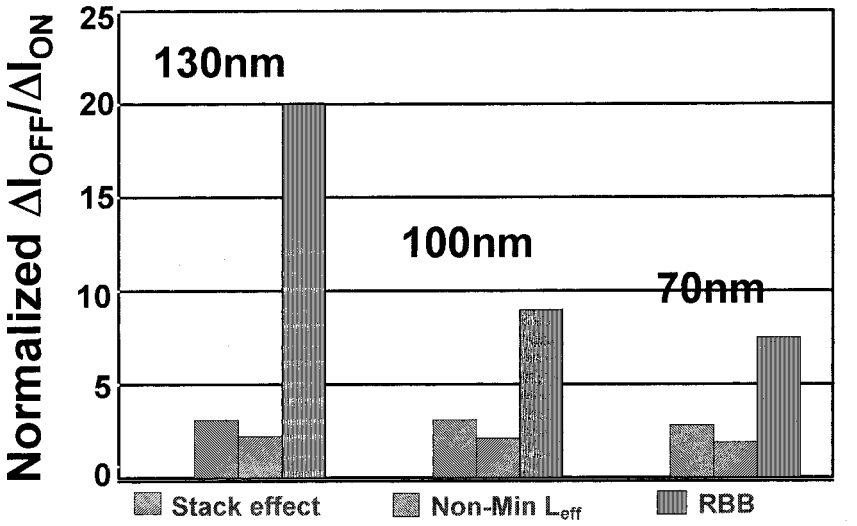

Figure 2-14. Trend for effectiveness of leakage reduction techniques with technology scaling.

It can be seen that for reverse body bias the ratio of $\Delta I_{OFF}/\Delta I_{ON}$ is the highest where, stack effect has the worse ratio. Another important observation from this figure is that the effectiveness of all these leakage reduction techniques is reducing with technology scaling. For reverse body bias technique, the $\Delta I_{OFF}/\Delta I_{ON}$ has decreased from 20x to 7.5x and for non-minimum L_{eff} it is reduced from 3.1x to 2.8x with scaling from 130 nm technology to 70 nm technology. This ratio for stack effect is reduced from 2.2x to 1.9x with scaling from 130 nm technology to 70 nm technology [19].

3. JUNCTION TEMPERATURE PROJECTIONS FOR DEEP SUB-MICRON TECHNOLOGIES

Several techniques can estimate junction temperature. One method directly measures junction temperature with thermal sensors at several on-chip locations during normal and burn-in conditions [23, 24]. Another method uses chip level 3D electro-thermal simulators that can find the steady-state *CMOS* VLSI chip temperature profile at the corresponding circuit performance [25, 26]. However, thermal sensors are relatively large devices, and accurate prediction requires a number of them placed on the *IC*. Sensors require calibration. Gerosa, et. al., reported a 0.2mm^2 thermal sensor with a sensing range of 0-128°C and a 5-bit resolution (4°C) [27]. Thermal sensors can only be used for verification, and one may have to use other

techniques for prediction and estimation. 3D electro-thermal simulators cannot be used for large-scale integrated circuits such as microprocessors because of long simulation time. The simulation time of a 2D Discrete Cosine Transformation (*DCT*) chip (107,832 transistors, 8 MHz) was reported at 12 hours [26].

In this section, a method for average junction temperature (T_j) estimation that can be used for normal and burn-in operating conditions is proposed. The method can predict the impact of technology scaling on junction temperature. The packaging issues, such as the thermal impedance of the package and other such factors were not considered. In this work the focus was on the intrinsic die behavior under the burn-in and normal conditions since package thermal properties tend to be user-specific.

3.1 Semiconductor Thermal Resistance Models

The Arrhenius model predicts that the failure rate of integrated circuits is an inverse exponential function of the junction temperature. A small increase of 10-15°C in junction temperature may result in ~2x reduction in the life span of the device [28]. While T_a represents the ambient temperature for an *IC*, the relationship between ambient and average junction temperature for a VLSI is often described as in [29]:

$$T_j = T_a + P_{chip} \times R_{ja} \tag{2.9}$$

Where T_a is the ambient temperature, P_{chip} is the total power dissipation of the chip, and R_{ja} is the junction-to-ambient thermal resistance. The impact of technology scaling on Eq. (2.9) must be analyzed to estimate the average junction temperature for several technologies. In this work the power dissipation and thermal resistance change with technology scaling were investigated in order to predict how these parameters will change.

The initial investigations on technology scaling and thermal resistance were carried out on bipolar transistors. For these devices, the thermal resistance was estimated as in [30]:

$$R_{ja} = \frac{1}{4K(L \times W)^{0.5}} \tag{2.10}$$

Where K is the thermal conductivity of silicon, (LxW) is the emitter size, and R_{ja} is the thermal resistance (°C/mW). It was shown that the thermal

resistance increased as the emitter size was reduced. Recently, a relationship between the thermal resistance of a *MOSFET* and its geometrical parameters was derived using a 3-D heat flow equation [31].

$$R_{ja} = \frac{1}{2\pi k}\left[\frac{1}{L}\ln\left(\frac{L+\left(W^2+L^2\right)^{0.5}}{-L+\left(W^2+L^2\right)^{0.5}}\right)+\frac{1}{W}\ln\left(\frac{W+\left(W^2+L^2\right)^{0.5}}{-W+\left(W^2+L^2\right)^{0.5}}\right)\right] \tag{2.11}$$

Where k is the thermal conductivity of silicon (k = *1.5x10-4W/mm°C* [32]), W and L are channel geometry parameters. The thermal conductivity of silicon has a temperature dependence described in [33].

The temperature dependence of silicon thermal conductivity is more important in silicon on insulator (*SOI*) technologies where self-heating contributes to a rise in junction temperature. So, our calculations assumed that the thermal resistance of silicon was temperature independent [31, 32]. Eq. (2.11) was used for the thermal resistance calculations for *MOSFETs* in different *CMOS* technologies.

3.2 Estimation of Junction Temperature Increase with Technology Scaling at Normal Conditions

F_{max} is defined as the maximum toggle frequency of an inverter in a given technology. The dynamic power consumption calculation under normal operating conditions was done at 70% of F_{max}. *HSPICE* simulations were carried out with BSIM model level 49. Transistor models for a 0.13μm *CMOS* technology were taken from United Microelectronics Corporation (*UMC*). Transistor models for other *CMOS* technologies were adapted from the Taiwan Semiconductor Manufacturing Corporation (*TSMC*). The simulation results and transistor sizes are given in Table 2.2.

The inverter load was the standard load element (*N-MOSFET*) used by *TSMC* for inverter ring-oscillator simulations. The load element sizes were taken from the *TSMC* and *UMC SPICE* models file specified for each of analyzed *CMOS* technologies. The International Technology Road map for Semiconductors (*ITRS*) 2002 [34] indicates that scaling down of device sizes is still in progress. Planar type transistors with 15-30 nm gate lengths have already been demonstrated [35]. However, 90-100 nm *CMOS* technology is currently the state-of-the-art for production of microprocessors and *SRAM* chips [36-38]. Therefore, we included the 90 nm *CMOS* technology node in

our study of burn-in testing. The effective channel length of transistors for this technology was assumed to be 55-65 nm.

Table 2-2. Simulated *CMOS* inverter parameters and F_{max}.

CMOS Technology $\mu m/V_{DD}$, *V*	*N-MOSFET W/L* $\mu m / \mu m$	*P-MOSFET W/L, Load* $\mu m / \mu m$	*N-MOSFET W/L, Load* $\mu m / \mu m$	F_{max} *MHz*	$F_{operating} = 0.7F_{max}$ *MHz*
0.35/3.3	4/0.35	10.0/0.35	3.0/0.35	1450	1015
0.25/2.5	2.86/0.25	7.140.25	2.15/0.25	1950	1365
0.18/1.8	2.06/0.18	5.14/0.18	1.55/0.18	2300	1610
0.13/1.3	1.49/0.13	3.71/0.13	1.12/0.13	4000	2800

The total power consumption of an inverter toggling at $0.7F_{max}$ in four different technologies is simulated, with results given in Table 2.2. The thermal resistance of an average transistor was computed from Eq. (2.11). The average size of a transistor was estimated by averaging the *NMOS* and *PMOS* transistor widths.

As the transistor dimensions are reduced, the thermal resistance increases. Figure 2.15 illustrates inverter power dissipation at an operating frequency of $0.7F_{max}$ and the thermal resistance of an average transistor as functions of technology. Owing to lack of access to 90 nm *CMOS* technology, an alternative method was utilized to obtain the inverter power and thermal resistance estimates in Figure 2.15. For the 1.0 V, 90 nm *CMOS* technology, the *ITRS* predicts the transistor density in a microprocessor chip to be about 0.27 millions/mm^2. It is assumed that the transistor density is doubled with technology scaling for each new process generation. An industrial estimate of the power density of a microprocessor chip, implemented in 90 nm technology, is approximately 0.5W/mm^2 [37-39]. Power density is defined as the power dissipated by the chip per unit area under nominal frequency and normal operating conditions. Using these assumptions we can estimate the inverter power dissipation at normal operating conditions (V_{DD} = 1V, T = 25°C) and speed (Figure 2.15).

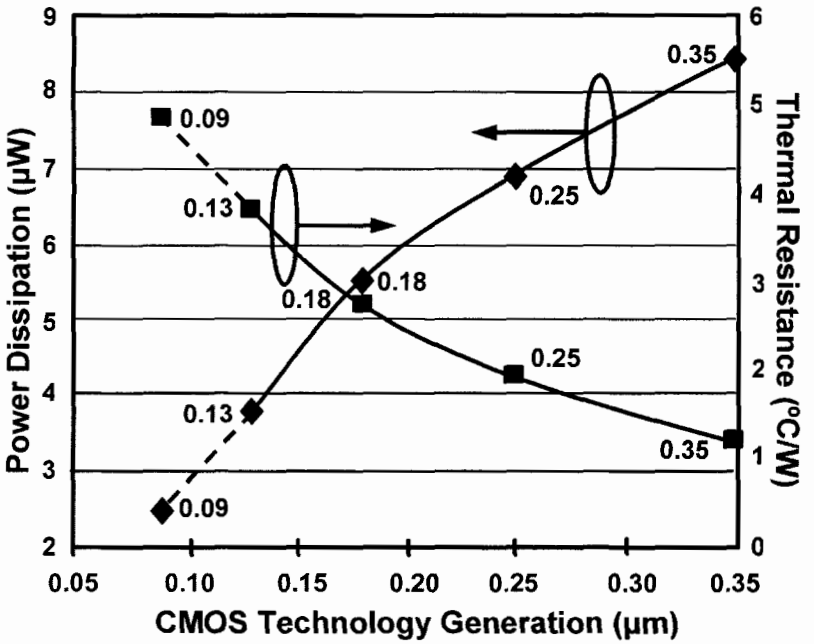

Figure 2-15. Inverter power dissipation and transistor thermal resistance for different *CMOS* technologies.

The scaling scenario of transistor sizes in a *CMOS* inverter was extended to 90 nm *CMOS* technology to calculate the thermal resistance. Transistor sizes of *PMOSFET (W/L)=3.0/0.1* and *NMOSFET (W/L)=1.0/0.1* were used. The calculated transistor thermal resistance for 90 nm technology using Eq. (2.5) is shown in Figure 2.15.

The 0.35μm *CMOS* technology was used as the reference technology. Using Eq. (2.9), ΔT can be defined as the temperature difference between junction and the ambient. If ΔT is set to unity for a 0.35μm technology, then the normalized change in ΔT with respect to the reference technology can be calculated. Using Eq. (2.9) and data presented in Figure 2.15, the normalized average temperature increase for different technologies was estimated. For example, Eq. (2.12) is used for calculation of $\Delta T_{0.25}=\Delta T_{0.35}$ ratio:

$$\frac{\Delta_{0.25}}{\Delta_{0.35}} = \frac{\left(T_j - T_a\right)_{0.25}}{\left(T_j - T_a\right)_{0.25}} = \frac{\left(P \times R_{ja}\right)_{0.25}}{\left(P \times R_{ja}\right)_{0.35}} \tag{2.12}$$

Figure 2.16 shows the normalized *MOSFET* junction temperature change with respect to the 0.35μm technology using Eq. (2.12). As the technology shrinks from 0.35μm to 0.18μm, the normalized temperature increased primarily from the increase in thermal resistance with scaling. However, scaling from 0.18μm to 0.09μm results in lower normalized *MOSFET* junction temperature with respect to 0.18μm technology. The reduction in normalized transistor temperature is due to the drastic reduction in power dissipation. The reduced parasitic capacitance is the primary reason for the reduced power dissipation. As a result of scaling from 0.18μm technology, P reduces faster than R_{ja} increases.

The increase in transistor density with scaling when estimating the average normalized temperature increase must also be considered. The density numbers were adopted from the International Technology Road map for Semiconductors (*ITRS*) [34, 40]. Figure 2.17 shows the increased number of transistors and chip size with scaling. These graphs allow us to calculate the transistor density in the chip for the given technology.

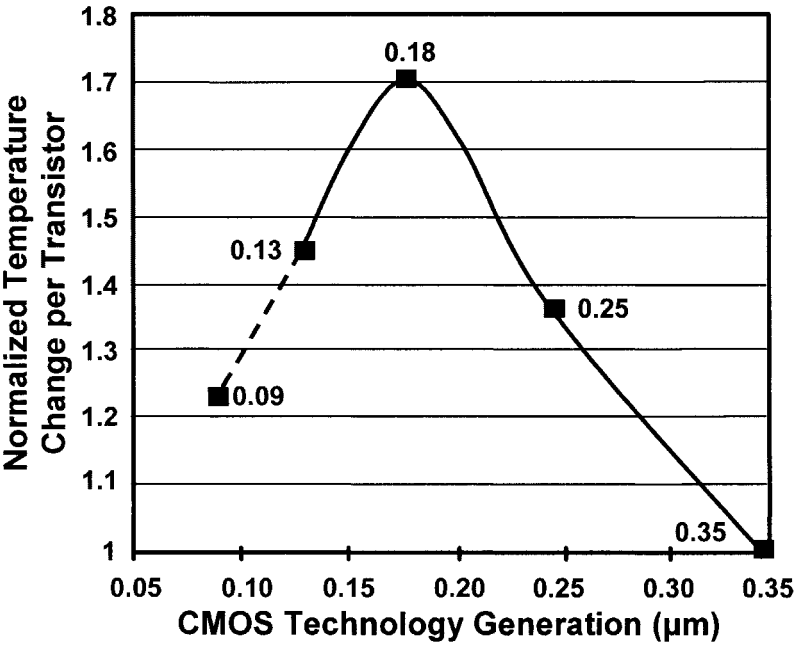

Figure 2-16. MOSFET junction temperature vs. technology.

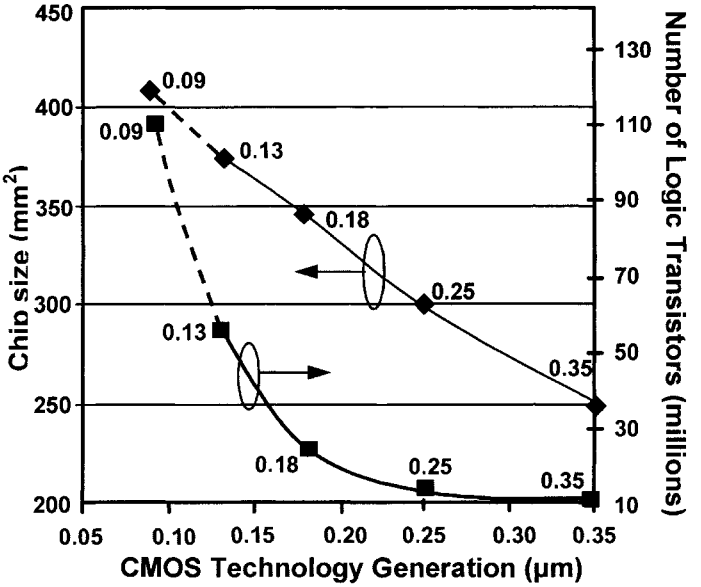

Figure 2-17. The trends of *CMOS* logic chips (data for graphs were adopted from [17, 20]).

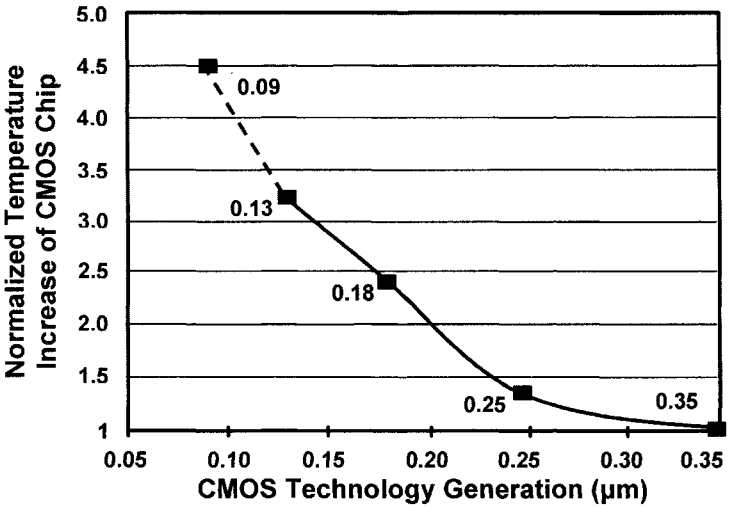

Figure 2-18. Normalized chip junction temperature increase with technology scaling for normal operating conditions.

The normalized temperature increase of a *CMOS* chip with technology scaling was calculated by multiplying the temperature increase per transistor in Figure 2.16 times the transistor density calculated from Figure 2.17. The results are shown in Figure 2.18. It can be concluded from Figure 2.18 that the normalized temperature increase of the chip is significantly elevated with *CMOS* technology scaling from 350 nm to 90 nm under normal operating conditions. The estimated junction temperature of a 90 nm *CMOS* chip is ~4.5 times higher than the junction temperature of a 0.35μm *CMOS* chip. This calculation assumed that the ambient temperature was the same for all analyzed technologies.

4. RELIABILITY ISSUES IN SCALED TECHNOLOGIES

The effects of temperature and V_{DD} on microelectronic devices are often assessed by accelerated tests carried out at high temperature and voltage to generate reliability failures in a reasonable time period. Burn-in is often used as a reliability screen to weed out infant mortalities. Weak gate oxides are one of the major components of such failures. These failures are accelerated due to elevated electric field and temperature. Several dielectric breakdown models exist in the literature that can describe intrinsic as well as the defect-related breakdown. In the next subsections, we consider some widely used models. It is apparent that electric field and junction temperature influence the time to breakdown of a gate oxide. Metal failures are another typical reliability failure mechanism activated by burn-in. Most metal failures are due to electro-migration [42, 43] or stress voiding [43]. The increase in chip junction temperature results in an exponential increase in cooling cost [24].

4.1 Time-Dependent Dielectric Breakdown (TDDB) and Gate Oxide Breakdown Models

The fundamental physical mechanisms of gate oxide breakdown are divided into two groups: intrinsic and extrinsic oxide breakdown mechanisms. The intrinsic oxide breakdown and wear out refers to defect-free oxide. The failure mechanism can be defined at the critical density of accumulated charge traps in the gate oxide through which a conductive path is formed from one interface to the other. The extrinsic breakdown refers to defects in the oxide whose failure mechanisms are the result of plasma damage, mechanical stress inside of oxide film, contamination, hot carrier damage, or oxide damage by ion implantation. The extrinsic damages in gate oxide typically appear during relatively short time burn-in testing (~12

hours). Both breakdown mechanisms appear during burn-in as well as life testing [44, 45].

The *E* and *1/E* models are widely used in intrinsic gate oxide reliability predictions for oxide thickness greater than 50Å. Both models have a physical basis. The E-model is expressed as:

$$t = A\exp(-\gamma E)\exp\left(\frac{E_a}{kT_j}\right) \tag{2.13}$$

Where t is the time to breakdown, A is a constant for a given technology, γ is the field acceleration parameter, E is the oxide field, E_a is the thermal activation energy, k is Boltzman's constant, and T_j is the junction temperature (°K). The *E*-model is based on thermo-chemical foundation and it indicates that increasing electric field across the gate oxide will decrease the time to break down. On the other hand, if we assume that the breakdown process is a current driven process, then the *1/E* model predicts:

$$t = \tau_0 \exp\left(\frac{G}{E}\right)\exp\left(\frac{E_a}{kT_j}\right) \tag{2.14}$$

Where τ_0 and G are constants, E is the oxide electric field, E_a is the activation energy, and T_j is the junction temperature. The *1/E* model implies that the dielectric will not degrade in the absence of electric field. The *1/E* model ignores important thermal/diffusion processes that are known to degrade all materials over time, even in the absence of an electric field. Figure 2.19 shows the comparison of *E*-model and *1/E*-model failure in time (*FIT*) to time dependent dielectric breakdown (*TDDB*) in *T=175°C*.

To increase the drive current and to control the short channel effects, the oxide thickness should decrease at each technology node. The experimental measurements of time to breakdown of ultra thin gate oxides with thickness less than 40 Å show that the conventional *E* and *1/E TDDB* models cannot provide the necessary accuracy for calculation and prediction [46]. Hence, starting from about the 180 nm *CMOS* technology (T_{OX} range is about 26-31 Å) a new *TDDB* model is proposed [46, 47]. Experiments show that the generation rate of stress-induced leakage current (*SILC*) and charge to breakdown (Q_{BD}) in ultra- thin oxides is controlled by gate voltage rather than the electric field. This model (Eq. 2.15) includes the gate oxide thickness (T_{OX}) and the gate voltage (V_G) [48].

$$T_{bd} = T_0 \cdot \exp\left[\gamma\left(\alpha \cdot T_{OX} + \frac{E_a}{kT_j} - V_G\right)\right] \tag{2.15}$$

Where γ is the acceleration factor, E_a is the activation energy, α is the oxide thickness acceleration factor, T_0 is a constant for a given technology, and T_j is the average junction temperature. Time to breakdown physical parameter values were extracted from experiments as follows: $(\gamma \cdot \alpha) = 2.0$ *1/Å*, $\gamma = 12.5$ *1/V* and $(\gamma \cdot E_a) = 575$ *meV* [48].

Historically, the activation energy has been an independent parameter in gate oxide breakdown models. However, starting from 130 nm technology, it becomes a function of accelerating electric field, as shown in Eq. (2.16) [49].

$$E_a \approx 1.15 - 0.07 \cdot E_{ox} \quad [\text{eV}] \tag{2.16}$$

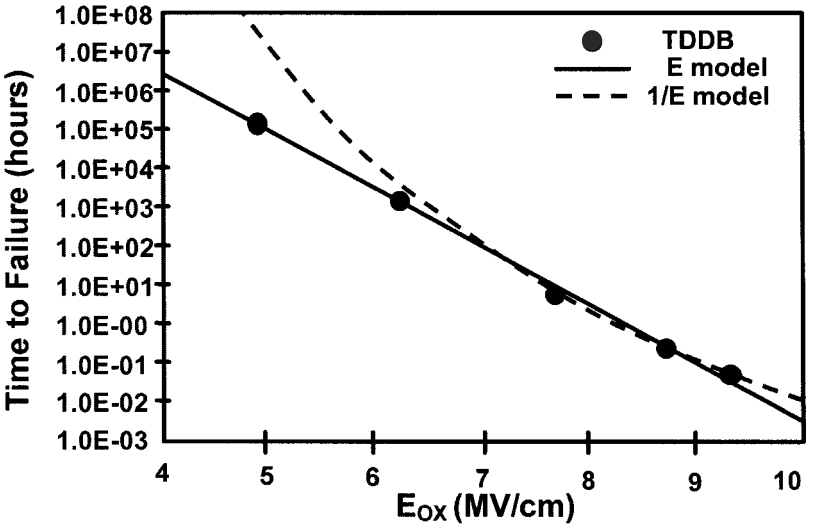

Figure 2-19. Comparison of *E*-model and *1/E*-model fit to T=175°C *TDDB*.

All the above methods describe the behavior of the intrinsic, good quality gate oxide. However, these models can also predict the time to breakdown

under extrinsic oxide breakdown conditions, which include oxide damage by ion implantation, plasma damage, mechanical stresses, and contamination from technology processes. Under these conditions, the E_a is reduced. Since, the time to breakdown is a strong function of T_j and E_a, above mentioned oxide breakdown models can be used to predict the defect related breakdown.

To explain the time-dependent dielectric breakdown (*TDDB*) mechanism of extremely thin oxide films (~20-30 Å), researchers proposed two different approaches: (1) the anode hole injection model [50], (2) the electron trap generation model [51]. According to the first model, injected electrons generate holes at the anode that can tunnel back into the oxide. Intrinsic breakdown occurs when a critical hole concentration (Q_{BD}) is reached. The second model claims that a critical density of electron traps generated during stress is required to trigger oxide breakdown. Based on this model the breakdown event is presented as the formation of a conductive path of traps connecting the anode to the cathode interface. Recently, it was shown that the anode hole injection model and the electron trap generation model can be directly linked. A new model based on a percolation concept and statistical properties of oxide breakdown was developed [52]. Accordingly, breakdown can occur only when a connecting path of traps is formed across the gate oxide from the substrate to the gate due to the random defect generation throughout the insulating film. The physics-based analytical model [53], which is the extension and simplification of the common percolation concept, allows us to calculate (Eq. 2.17) the critical density (N_{crit}) of defects per unit of area at breakdown conditions as a function of gate oxide thickness (t_{ox}).

$$N_{crit}^{BD} = \frac{t_{ox}}{\alpha_0^3} \exp\left(-\frac{\alpha_0}{t_{ox}} \ln\left(\frac{A_{ox}}{\alpha_0^2} \right) \right) \tag{2.17}$$

Where, α_0 is the lattice constant of a cubic structure in the oxide bulk ($\alpha_0 \approx 2.34$ nm), and A_0 is the oxide area.

The relationship between the charge-to-breakdown (Q_{BD}), the critical defect density (N_{crit}), and the injected electron density (P_g) is [54]:

$$Q_{BD} = \frac{qN_{crit}^{BD}}{P_g} \tag{2.18}$$

The time-to-breakdown of thin oxides is determined by:

$$T_{BD} = \frac{Q_{BD}}{J_g} = \frac{qN_{crit}^{BD}}{P_g J_g} \tag{2.19}$$

Where, J_g is the tunneling current across the gate oxide. The tunneling current (J_g) and the injected electron density (P_g) can be extracted from the experiments using *SILC* and *C-V* measurements [54]. Gate oxide defects have traditionally been a major reason for burn-in. Although other defects are activated during burn-in, it is important to understand the theory of oxide wear out and breakdown.

4.2 Electro-migration (EM)

Interconnect *EM* is the movement of metal atoms in the direction of electron flow due to momentum transfer from electrons to the metal ions under thermal and voltage stresses. *EM* is usually modeled by the empirical Black's formula [55], which relates the Mean-Time-To-Failure (*MTTF*) to the stressing conditions and is given as:

$$MTTF = A \cdot J^{-n} \exp\left(\frac{E_a}{kT_j}\right) \tag{2.20}$$

Where A is the process constant dependent on material and geometry of the metal strip, n is a current exponent factor, T_j is the absolute junction (chip) temperature, k is the Boltzmann's constant, E_a is the activation energy and J is the current density. The activation energy for *Al-Cu* metal is in the range of 0.76-0.86 eV [56], and the activation energy for *Cu* interconnections, can vary widely from 0.7-0.9 eV to 1.0 eV. The lifetime of interconnects is decreased with the reduction of line width [57]. The accuracy of lifetime prediction is strongly dependent on the accuracy of the junction temperature measurement during the acceleration testing.

4.3 Hot Electron Effect

Substrate current has been successfully used as the Hot Carrier Injection (*HCI*) reliability indicator [58]. General substrate current model shows I_{sub} decreases while temperature increases. This is due to carrier mean free path reduction from more lattices scattering at higher temperature [59]. Recent study shows I_{sub} is insensitive to temperature over the range 77°K to 300°K due to the insensitivity of the carrier mean free path to the temperature. The latest study shows I_{sub} of 0.25μm device has reversed temperature

dependence: I_{sub} increases with the temperature when V_{DD} is lower than a specific value, which can be defined as V_{DD} transition point. For *HCI* reliability, the field operation lifetime is projected from accelerated high bias stress data. The fully understanding of this reversed temperature effect and V_{DD} transition point will greatly impact on how to correctly set stress conditions and accurately project *HCI* lifetime to field operation conditions [60].

The widely used I_{sub} model is:

$$I_{sub} = I_d \frac{A_i \lambda}{\varphi_i}(V_{DD} - V_{Dsat}).\exp\left(\frac{-\varphi_i}{\lambda k_m + \beta\frac{3}{2}kT}\right) \tag{2.21}$$

Where I_d is the source drain current, A_i is the coefficient, λ is carriers mean free path, φ_i is the energy required to generate electron-hole pair. $\beta 3kT/2$ is the thermal kinetic energy which has been taken into account.

4.4 Negative Bias Temperature Instability (NBTI)

Negative bias temperature instability (*NBTI*) is a *PMOS* degradation mechanism that can result in threshold voltage shifts up to 100 mV or more in very thin oxide devices [61][62]. Negative bias temperature instability can occur whenever a *PMOS* is biased in inversion. The damage has been shown to consist of both the formation of positive fixed charge and interface state generation [63][61]. The effects of these damage mechanisms are a negative V_{TH} shift (making the device harder to turn on) and transconductance degradation. Unlike hot carrier mechanisms, *NBTI* does not depend on any lateral (channel) field. The degradation is generally worst at $V_{DS} = 0$ [62][64]. Therefore, *NBTI* does not display significant channel length dependence, as hot carrier does.

Analog circuits and *SRAMs* are particularly sensitive to *NBTI* related problems. First of all, analog circuits often require matching among transistors, and utilize larger channel lengths to achieve good V_{TH} and I_D matching. Secondly, many of analog transistors, such as biasing, are constantly on making them prone to *NBTI*. *NBTI* could be expected to be the dominant degradation mechanism for these devices. Even for *PMOS* normally operating at a low gate bias (for which the *NBTI* shift would be small), there may be other circuit operation modes (such as power-down mode) which expose the devices to a relatively high V_{GS} [65].

5. SUMMARY

In this chapter active and leakage power consumption trends were reviewed. Subsequently, we discussed various strategies to reduce the total power consumption. The increased power consumption results in higher on chip, junction, temperature which in turns negatively influences various aging mechanisms. Therefore we discussed the impact of higher junction temperature on long term reliability.

References

1. D. M. Brooks, P. W. Cook, P. Bose, S. E. Schuster, H. Jacobson, P. N. Kudva, A. Buyuktosunoglu, J. Wellman, V. Zyuban and M. Gupta. "Power-aware microarchitecture: design and modeling challenges for next-generation microprocessors". IEEE Microelectronics, Vol. 20, No. 6, 2000.
2. E. J. Nowak. "Maintaining the benefits of CMOS scaling when scaling bogs down". IBM Journal of Research and Development, Vol. 48, No. 2/3, pages 26-44, 2002.
3. A. Keshavarzi, K. Roy and C.F. Hawkins. "Intrinsic leakage in deep submicron CMOS ICs: measurement based test solutions". IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 8, No. 6, pages 717-723, 2000.
4. Y. Taur and T. H. Ning. "Fundamentals of modern VLSI Devices" Cambridge University Press, pages 120-128, 1998.
5. Y. Taur and T. H. Ning. "Fundamentals of modern VLSI Devices" Cambridge University Press, pages 94-95, 1998.
6. S. Tompson, P. Packan and M. Bohr. "MOS scaling: transistor challenges for 21st century". Intel Technology Journal, Q3, 1998.
7. S. M. Sze. "Physics of semiconductor device". John Wiley & Sons, 1936.
8. BSIM 4.2.1 MOSFET model, BSIM group, UC Berkeley, http://www-device.eecs.berkeley.edu/~bsim3/.
9. K. Cao, W. C. Lee, W. Liu, X. Jin, P. Su, S. Fung, J. An, B. Yu and C. Hu. "BSIM4 gate leakage model including source drain partition". Technical Digest of International Electron Devices Meeting, pages 815-818, 2000.

10. S. Rusu. "Trends and challenges in VLSI technology scaling toward 100 nm". Presented at ESSCIRC. [Online]. Available: http://www.esscirc.org/esscirc2001/C01_Presentations/404.pdf.
11. J. P. Halter, and F. N. Najm, "A gate-level leakage power reduction method for ultra-low power CMOS circuits". Proceedings of the IEEE CICC, pages 475-478, 1997.
12. T. Sakurai , H. Kawaguchi , T. Kuroda. "Low-power CMOS design through VTH control and low-swing circuits". Proceedings of the international symposium on Low power electronics and design, pages 1-6, 1997.
13. M. R. Stan. "Optimal Voltages and Sizing for Low Power". Proceedings of the 12th International Conference on VLSI Design - 'VLSI for the Information Appliance', pages 428, 1999.
14. N. Sirisantana, L. Wei and K. Roy. "High-Performance Low-Power CMOS Circuits Using Multiple Channel Length and Multiple Oxide Thickness". Proceedings of the 2000 IEEE International Conference on Computer Design: VLSI in Computers & Processors, pages 227-232, 2000.
15. A. Keshavarzi , S. Ma , S. Narendra , B. Bloechel , K. Mistry , T. Ghani , S. Borkar , V. De. "Effectiveness of reverse body bias for leakage control in scaled dual Vt CMOS ICs". Proceedings of the 2001 International Symposium on Low Power Electronics and Design, pages 207-212, 2001.
16. T. Kuroda, T. Fujita, S. Mita, T. Nagamatsu, S. Yoshioka, K. Suzuki, F. Sano, M. Norishima, M. Murato, M. Kako, M. Kinugawa, M. Kakumu, and T. Sakurai. "A 0.9V, 150-MHz, 10-mW, 4mm2, 2-D discrete cosine transform core processor with variable threshold-voltage (VT) scheme," IEEE Journal of Solid State Circuits, Vol. 31, No. 11, pages 1770-1779, 1996.
17. A. P. Chandrakasan , W. J. Bowhill , F. Fox. "Design of high-performance microprocessor circuits". Wiley-IEEE Press, 2000.
18. A. P. Chandrakasan and R. W. Brodersen. "Low power digital CMOS design". Kluwer Academic Publishers, Norwell, MA, 1995.
19. B. Chatterjee, M. Sachdev, S. Hsu, R. Krishnamurthy, S. Borkar. "Effectiveness and scaling trends of leakage control techniques for sub-130nm CMOS technologies". Proceedings of the International Symposium on Low Power Electronics and Design, pages 122-127, 2003.
20. S. Narendra, S. Borkar, V. De, D. Antoniadis, and A. Chandrakasan. "Scaling of stack effect and its application for leakage reduction". Proceedings of International Symposium of Low Power Electronics and Design (ISLPED), pages 195–200, 2001.

21. 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 Transactions on VLSI Systems, Vol. 7, pages 16–24, Jan. 1999.
22. L. Wei, K. Roy, and V. K. De. "Low voltage low power CMOS design techniques for deep submicron ICs". Proceedings of International Conference of VLSI Design, pages 24–29, 2000.
23. T.J. Goh, A.N. Amir, C.P. Chiu, and J. Torresola. "Novel thermal validation metrology based on non-uniform power distribution for Pentium III Xeon cartridge processor design with integrated level two cache". Proceedings of Electronic Components and Technology Conference, pages 1181-1186, 2001.
24. S.H. Gunter, F. Binns, D.M. Carmean, and J.C. Hall. "Managing the impact of increasing microprocessor power consumption". Intel Tech. Journal, Q1:1-9, 2001.
http://developer.intel.com/technology/itj/archive.htm.
25. Y.K. Cheng, C.C. Teng, A. Dharchoudhury, E. Rosenbaum, and S.M. Kang. "A chip-level electrothermal simulator for temperature profile estimation of CMOS VLSI chips". Proceedings of International Symposium on Circuit and Systems, pages 580-583, 1996.
26. C.C. Teng, Y.K. Cheng, E. Rosenbaum, and S.M. Kang. "ITEM: A temperature dependent electromigration reliability diagnosis tool". IEEE Transaction on Computer Aided Design of Integrated Circuits and Systems, Vol. 16, No. 8, pages 882-893, 1997.
27. P. Reed, M. Alexander, J. Alvarez, M. Brauer, C.C. Chao, C. Croxton, L. Eisen, T. Le, T. Ngo, C. Nicoletta, H. Sanchez, S. Taylor, N. Vanderschaaf, and G. Gerosa. "A 250-MHz 5-W PowerPC microprocessor with on-chip L2 cash controller". IEEE Journal of Solid-State Circuits, Vol. 32, No. 11, pages 1635-1649, 1997.
28. R. Viswanath, V. Wakharkar, A. Watwe, and V. Lebonheur. "Thermal performance challenges from silicon to systems". Intel Technology Journal, Q3, pages 1-16, 2000.
http://developer.intel.com/technology/itj/archive.htm.
29. P. Tadayon. "Thermal challenges during microprocessor testing". Intel Technology Journal, Q3, pages 1-8, 2000.
30. R.C. Joy and E.S. Schlig. "Thermal properties of very fast transistors". IEEE Transaction on Electron Devices, ED-Vol. 17, No. 8, pages 586-594, 1970.
31. N. Rinaldi. "Thermal analysis of solid-state devices and circuits: an analytical approach". Solid-State Electronics, Vol. 44, No. 10, pages 1789-1798, 2000.

32. N. Rinaldi. "On the modeling of the transient thermal behavior of semiconductor devices". IEEE Transaction on Electron Devices, Vol. 48, No. 12, pages 2796-2802, 2001.
33. D.L. Blackburn and A.R. Hefner. "Thermal components models for electro-thermal network simulation". Proceedings of 9th IEEE SEMI-THERM Symposium, pages 88-98, 1993.
34. International Technology Roadmap for Semiconductors (ITRS). http://public.itrs.net/.
35. B. Doyle, R. Arghavani, D. Barlage, S. Datta, M. Doczy, J. Kavalieros, A. Murthy, and R. Chau. "Transistor element for 30 nm physical gate lengths and beyond". Intel Technology Journal, Vol. 6, No. 2, pages 42-54, 2002.
36. S.F. Huang, C.Y. Lin, Y.S. Huang, T. Schafbauer, M. Eller, Y.C. Cheng, S.M. Cheng, S. Sportouch, W. Jin, N. Rovedo, A. Grassmann, Y. Huang, J. Brighten, C.H. Liu, B.V. Ehrenwall, N. Chen, J. Chen, O.S. Park, and M. Common. "High-performance 50 nm CMOS devices for microprocessors and embedded processor core applications". IEDM, pages 237-240, 2001.
37. A. Ono, K. Fukasaku, T. Hirai, S. Koyama, M. Makabe, T. Matsuda, M. Takimoto, Y. Kunimune, N. Ikezawa, Y. Yamada, F. Koba, K. Imai, and N. Nakamura. "A 100 nm node CMOS technology for practical SOC application requirement". IEDM, pages 511-514, 2001.
38. D.J. Frank. "Power-constrained CMOS scaling limits". IBM Journal of Research and Development, Vol. 46, No. 2/3, pages 235-244, 2002.
39. S. Borkar. "Design challenges of technology scaling". IEEE Micro, pages 23-29, July-August 1999.
40. D.P. Valett and J.M. Soden. "Finding fault with deep-submicron ICs". IEEE Spectrum, pages 39-50, October 1997.
41. B. Davari, R.H. Dennard, and G.G. Shahidi. "CMOS scaling for high performance and low power - The next ten years". Proceedings of the IEEE, Vol. 83, No. 4, pages 595-606, 1995.
42. E. T. Ogawa, Ki-Don Lee, V. A. Blaschke, and P. S. Ho. "Electromigration Reliability Issues in Dual-Damascene Cu Interconnections". IEEE Transaction on reliability, Vol. 51, No. 4, pages 403-419, 2002.
43. C.F. Hawkins, A. Keshavarzi, and J.M. Soden. "Reliability, test and IDDQ measurements". IEEE International Workshop on Iddq testing, pages 96-102, 1997.
44. J.W. McPherson, V.K. Reddy, and H.C. Mogul. "Field-enhanced Si-Si bond-breakage mechanism for time-dependent dielectric break-down in thin-film SiO2 dielectrics". Applied Physics Letter, Vol. 71, No. 8, pages 1101-1103, 1997.

45. A.M. Yassine, H.E. Nariman, M. McBride, M. Uzer, and K.R. Olasupo. "Time dependent breakdown of ultra-thin gate oxide". IEEE Transaction on Electron Devices, Vol. 47, No. 7, pages 1416-1420, 2000.
46. J.H. Suehle. "Ultra thin gate oxide reliability: Physical models, statistics, and characterization". IEEE Transaction on Electron Devices, Vol. 49, No. 6, pages 958-971, 2002.
47. P.E. Nicollian, W.R. Hunter, and J.C. Hu. "Experimental evidence for voltage driven breakdown models in ultra thin gate oxides". Proceedings of IEEE International Reliability Physics Symposium, pages 7-15, 2000.
48. F. Monsieur, E. Vincent, D. Roy, S. Bruyere, G. Pananakakis, and G. Ghibaudo. "Time to breakdown and voltage to breakdown modeling for ultra-thin oxides (TOX < 32A°)". Proceedings of IEEE International Reliability Workshop (IRW), pages 20-25, 2001.
49. M. Kimura, "Field and temperature acceleration models for time-dependent dielectric breakdown". IEEE Transaction on Electron Devices, Vol. 46, No. 1, pages.220-229, 1999.
50. I.C. Chen, S. Holland, K.K. Young, C. Chang and C. Hu, "Substrate hole current and oxide breakdown". Applied Physics Letter, Vol. 49, No. 11, pages 669-671, 1986.
51. P.P. Apte and K.C. Saraswat, "Modeling ultra thin dielectric breakdown on correlation of charge trap-generation to charge-to-breakdown". Proceedings of International Reliability Physics Symposium, 1994, pages 136-142.
52. R. Degraeve, G. Groeseneken, R. Bellens, J.L. Ogier, M. Depas, P. J. Roussel, and H.E. Maes, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown". IEEE Transaction on Electron Devices, Vol. 45, No. 4, pages 904-911.
53. J. Sune, "New physics-based analytical approach to the thin-oxide breakdown statistics". IEEE Electron Device Letters, Vol. 22, No. 6, pages 296-298, 2001.
54. J.H. Stathis, "Physical and predictive models of ultra thin oxide reliability in CMOS devices and circuits". Proceedings of International Reliability Symposium, 2001, pages 132-149.
55. J. Black. "Electromigration: A brief survey and some recent results". IEEE Transaction on Electron Devices, ED-Vol. 16, No. 4, pages 338-347, 1969.
56. W.B. Loh, M.S. Tse, L. Chan, and K.F. Lo. "Wafer-level electromigration reliability test for deep submicron interconnect metallization". IEEE Hong Kong Electron Device Meeting, pages 157-160, 1998.
57. C.K. Hu, R. Rosengerg, H.S. Rathore, D.B. Nguyen, and B. Agarwala. "Scaling effect on electromigration in on-chip Cu wiring". IEEE International Interconnect Technology Conference, pages 267-269, 1999.

58. W. Wang, C.F. Hsu, L.P. Chiang, N.K. Zous, T.S. Chao and C.Y. Chang, "Voltage scaling and temperature effects on drain leakage current degradation in hot carrier stressed N-MOSFET" Proceedings of Reliability Physics Symposium, pages 209-212, 1997.
59. H. Sasaki, "Hot-carrier induced drain leakage current in N-channel MOSFET". IEEE Transaction on Applied Superconductivity, pages 726-729, 1987.
60. R. Bouchakour, L. Hardly, I. Limbourg and M. Jourdain, "Modeling and characterization of the MOSFET transistor stressed by hot-carrier injection". Proceedings of the 38th Midwest Symposium on Circuits and Systems, Vol. 1, Pages 61-64, 1995.
61. S. Ogawa, M. Shimaya, and N. Shiono, "Interface-trap generation at ultrathin Si02 (4–6 nm)-Si interfaces during negative-bias temperature aging", Journal of Applied Physics, Vol. 77, No. 3, pages 1137–1148, 1995.
62. P. Chaparala, J. Shibley, and P. Lim, "Threshold voltage drift in PMOSFETs due to NBTI and HCI", IEEE International Integrated Reliability Workshop Final Report, pages 95-97, 2000.
63. C. E. Blat, E. H. Nicollian, and E. H. Poindexter, "Mechanism of negative-bias-temperature instability", Journal of Applied Physics, Vol. 69, No. 3, pages 1712–1720, 1991.
64. A. Acovic, E. Crabbé, F. Guarin, G. La Rosa, J. Lukaitis, and S. Rauch, "NBTI-channel hot carrier effects in pMOSFETs in advanced CMOS technologies", Proceedings of IEEE International Reliability Physics Symposium pages 282–286, 1997.
65. R. Thewes, R. Brederlow, C. Schündler, P. Wieczorek, A. Hesener, B. Ankele, P. Klein, S. Kessel, and W. Weber, "Device reliability in analog CMOS applications," IEEE Transaction on Electron Devices, Vol. 45, pages 2505–2513, 1998.

http://www.springer.com/978-0-387-25762-4

Thermal and Power Management of Integrated Circuits

Vassighi, A.; Sachdev, M.

2006, X, 179 p., Hardcover
ISBN: 978-0-387-25762-4