IMPACT OF OFF-STATE LEAKAGE CURRENT ON ELECTROMIGRATION DESIGN RULES FOR NANOMETER SCALE CMOS TECHNOLOGIES

Sheng-Chih Lin, Anirban Basu, Ali Keshavarzi*, Vivek De*, Amit Mehrotra† and Kaustav Banerjee

Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106

*Circuit Research Lab, Intel Corporation, Hillsboro, OR 97124, †Department of ECE, University of Illinois, Urbana-Champaign, IL 61801

ABSTRACT

This paper introduces the idea of *electrothermally coupled evaluation of junction temperature* to accurately estimate the lifetime of interconnects under electromigration (EM) constraints in leakage dominant technologies. The junction temperature thus evaluated with our proposed methodology is incorporated into the interconnect temperature-rise equation (which includes Joule-heating) and is solved self-consistently with both unipolar and bipolar EM lifetime equations to estimate accurate interconnect metal temperature and to provide comprehensive design guidelines for maximum allowable current density in power/ground and signal lines.

I. INTRODUCTION

While transistor scaling continues to be necessary, interconnects are now becoming the dominant factor determining system performance, power dissipation, and reliability. The requirement to meet the demand for system performance and increasing functionality, has resulted in an increasing number of interconnect levels and reduction in interconnect pitch [1]. These interconnect scaling requirements drive several technology enhancements such as using low-k materials particularly for intra-level dielectric to reduce the interconnect capacitance and introducing Cu as a substitute for Al to lower line resistance and improve current density capability. Unfortunately, these technology trends will exacerbate thermal effects and will have a strong impact on the system thermal management and EM lifetime of interconnects [2].

Failure of chips due to EM effects in interconnects is a reliability concern for sub-100 nm technologies. The rate of EM depends upon the metal temperature, current density, microstructure and robustness of interfaces. By using new metallization systems, such as copper, interconnects have better current density capability and lower resistance that can alleviate EM failures and improve system performance. However, the EM phenomenon has (inverse) exponential dependence on the metal temperature, which in turn, depends on the chip junction temperature and the temperature rise due to thermal effects caused by self-heating. Due to increasing number of interconnect levels and poor thermal conductivity of low-k dielectric materials, thermal effects are known to become worse [2] and may offset any gains from purely increasing current density capability [3]. Additionally, it has been recently shown that strong *electrothermal couplings* between supply voltage, frequency, power dissipation and junction temperature exist in leakage dominant nanometer scale technologies, mainly due to the exponential dependence of subthreshold leakage current on temperature, which can significantly impact various power-performance-reliability-cooling cost optimization schemes [4]. Therefore, an accurate estimation of interconnect metal temperature is extremely important for design and reliability evaluations. This requires a precise quantification of all coupled metrics to project the final reliability (lifetime) of interconnects. The electrothermally coupled (ET-coupled) junction temperature, in addition to the temperature rise due to self-heating of the metal line, can significantly affect the interconnect temperature estimation.

We discuss the EM phenomenon, the major failure mechanism of VLSI interconnects and its dependencies in the next section. Then, in Section III, we present the electrothermal couplings to explain the significance of using electrothermally coupled methodology for metal temperature estimation in leakage dominant technologies. We follow this up, in Section IV, with a quantitative analysis of the impact of off-state leakage currents on the EM design rules for both unipolar and bipolar stress conditions. We also demonstrate the benefit of applying electrothermally coupled methodology from the perspective of maximum allowable current density for interconnects. Finally, we illustrate the effect of technology scaling on allowable current density thereby setting up new guidelines for interconnect design. Conclusions are made in Section V.

II. INTERCONNECT RELIABILITY ESTIMATION

Electromigration is the mass transport in metal lines due to momentum transfer between conducting electrons and metal ions under an applied electric field, and is widely regarded as a major wear-out or failure mechanism of VLSI interconnects. The mean time to failure (MTTF) due to EM is modeled by the well-known Black's equation [5], given by

$$MTTF = A^* j^{-n} \exp(\frac{Q}{K_B T_m}) \tag{1}$$

where MTTF is the mean time to failure in hours, A^* is a constant that is dependent on the geometry and microstructure of the interconnect, j is the current density, and the exponent n is typically 2 under normal use conditions. The parameter Q is the activation energy for EM, K_B is the Boltzmann's constant, and T_m is the metal temperature. The EM lifetime (MTTF) is exponentially dependent on temperature (higher the metal temperature worse the MTTF).

Traditionally, in order to combat EM failures and have superior current density capability, an alternative choice of material is suggested. However, it should be noted that in addition to EM, the maximum allowed current density through interconnects is governed by Joule-heating or self-heating. In fact, it has been shown that Joule-heating effect will be the primary criterion that will dictate the maximum allowed current density in scaled technologies because addition of more levels of metals and introduction of low-k materials with poor thermal conductivity will nullify the gain obtained from higher current density capability of new interconnect materials [3, 6]. Joule-heating also determines the metal temperature (T_m) by the following relation:

$$T_m = T_j + \Delta T_{self-heating} \tag{2}$$

where T_j is average junction temperature or die temperature of the entire chip. The temperature rise of the metal interconnect due to the flow of current, $\Delta T_{self-heating}$ is given by

$$\Delta T_{self-heating} = (I_{rms})^2 \cdot R \cdot R_\theta \tag{3}$$

where R and R_θ are the resistance and thermal impedance of the interconnect respectively, and I_{rms} is the RMS value of the current. Therefore, it is clear that both EM and Joule-heating have to be considered simultaneously in order to evaluate metal temperature to determine the MTTF accurately [6]. However, it should be noted, as indicated by equation (2), in addition to EM and self-heating, the junction temperature will also be directly reflected onto the metal temperature. Therefore, precise estimation of junction temperature is imperative to predict the metal temperature and hence, the lifetime of interconnects. However, the accurate estimation of junction temperature in leakage dominant nanometer scale technologies requires a detailed consideration of different electrothermal couplings. The ET-coupled methodology introduced in [4] for a rigorous treatment of electrothermal couplings in estimating the junction temperature is outlined below.

III. ET-COUPLED METHODOLOGY

As the feature size in modern VLSI technologies scales to sub-100 nm regimes, power dissipation issues and thermal management of integrated systems are becoming key factors in most high-performance ICs including microprocessors [7]. They not only affect the circuit performance but also significantly impact the circuit reliability. Due to threshold voltage (V_t) scaling and process variations [8], leakage power dissipation ($P_{leakage}$), which is dominated by subthreshold leakage for high performance ICs [9], becomes a significant component of total chip power dissipation (P_{chip}) (**Figure 1**). Moreover, subthreshold leakage current is exponentially dependent on temperature and the dependence gets stronger with scaling (**Figure 2**). Hence, as illustrated in **Figure 3**, the junction temperature (T_j) increases superlinearly with thermal impedance (θ_j) of the chip due to the coupling between P_{chip} and T_j instead of the usual linear increase as given by

$$T_j = T_{amb} + P_{chip} \cdot \theta_j \tag{4}$$

where T_{amb} is the ambient temperature. From equation (4), T_j will increase superlinearly with θ_j if there is coupling between P_{chip} and T_j, (primarily due to the strong dependence of $P_{leakage}$ on T_j).

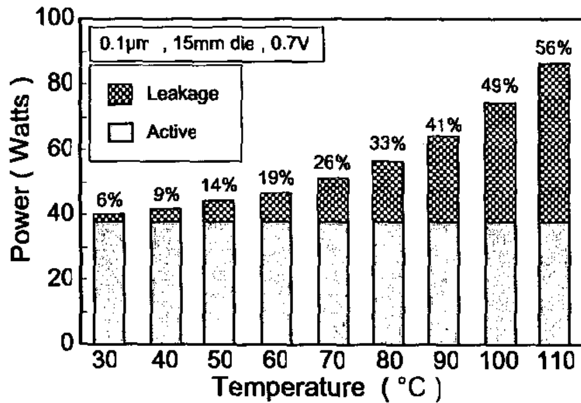

Figure 1. Temperature dependence of the leakage power dissipation for a *100 nm / 0.7 V CMOS* technology. The fraction of leakage power component increases superlinearly with temperature.

It can be observed from figure 3 that in order to maintain a desired value of T_j for leakage dominant nanometer scale technologies, a lower value of θ_j will be required, leading to an increase in packaging/cooling cost. Also, T_j affects the metal temperature (T_m) of interconnects that may adversely impact EM lifetimes and hence, failure of the chip. Furthermore, as the supply voltage (V_{dd}) is increased, chip frequency increases, and so does the total power dissipation and hence T_j. Besides, frequency itself is dependent on temperature due to the dependence of the transistor on-current (I_{on}) on T_j. The details of the various electrothermal couplings are summarized in **Figure 4**.

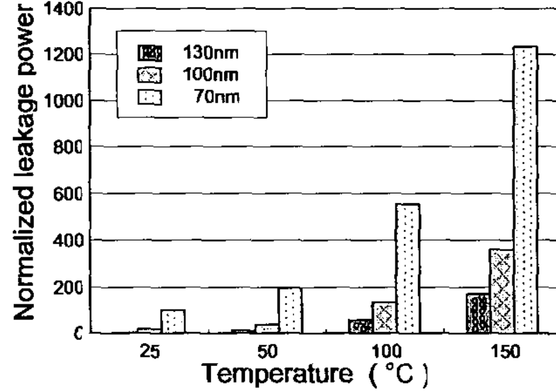

Figure 2. Leakage power of an *NMOS* device for different technology nodes based on SPICE simulations using BSIM3 models shows the effect of temperature. The leakage power is normalized w.r.t I_{off} at *130* nm node at *25 °C*.

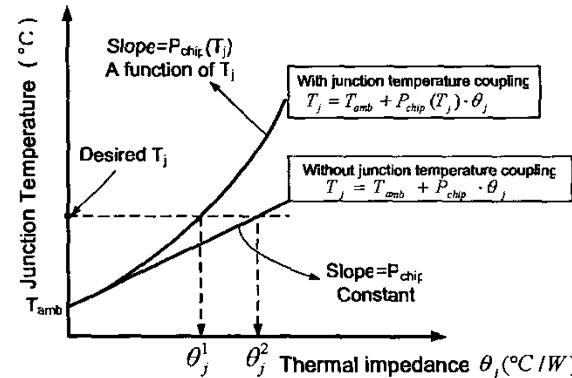

Figure 3. The schematic diagram illustrates the dependence of junction temperature on thermal impedance for the conventional method and the electrothermally coupled power estimation method.

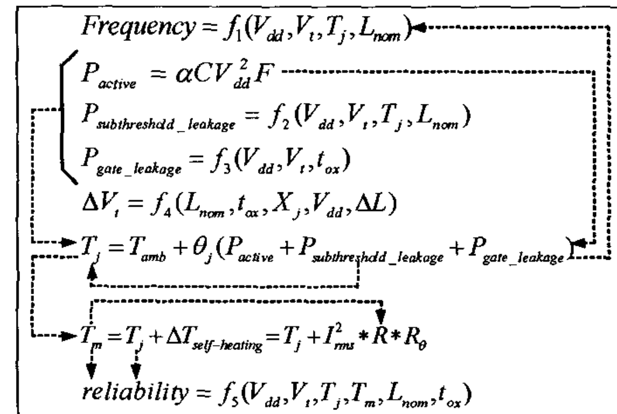

Figure 4. Models used for various metrics are expressed in functional format. Electrothermal couplings are indicated using broken lines.

From figure 4, it is obvious that power, frequency, reliability and temperature are strongly coupled quantities. Hence, an overall self-consistent electrothermal analysis method is necessary for accurately estimating T_j for any value of V_{dd} (or frequency). Once T_j is obtained, T_m can be calculated by equation (2) and (3). **Figure 5** provides an equivalent thermal circuit model and illustrates the relationship between metal, junction, ambient, and external temperatures.

The electrothermally coupled methodology uses both analytical models and results from circuit simulation based on an integrated device, circuit, and system level modeling approach [4] and has been summarized in **Figure 6**.

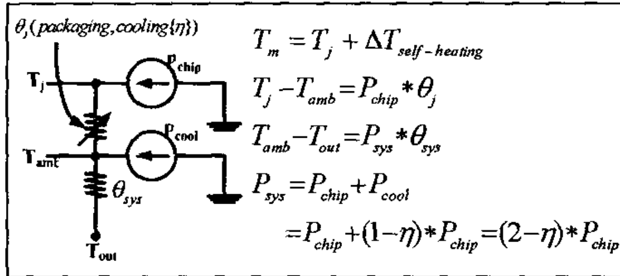

Figure 5. An equivalent thermal circuit illustrating the relationship between metal, junction, ambient and external temperatures in terms of chip and system level power dissipation and thermal impedances.

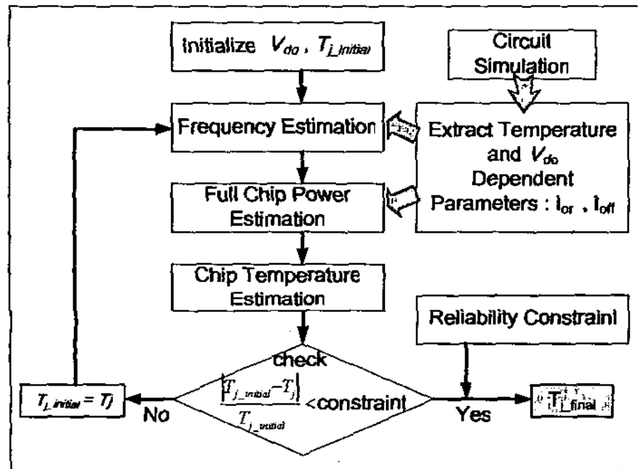

Figure 6. An overview of the electrothermally coupled junction temperature estimation technique. The methodology has been implemented as an automated computer program.

For a given supply voltage (V_{dd}) and initial (average) junction temperature (T_j), the operating frequency of the chip is first estimated using extracted I_{on} value (dependent on temperature and V_{dd}) through a circuit simulator. The estimated frequency is then used to calculate active power dissipation (P_{active}). The leakage power ($P_{leakage}$) estimation uses extracted off-current value, I_{off}, (also V_{dd} and T_j dependent) from the circuit simulator. Effect of process (especially channel length) variation is also taken into account for estimating $P_{leakage}$. The total chip power dissipation is then used to calculate the new junction temperature using compact thermal models for the IC packaging and cooling technology. The estimated junction temperature is then compared with the initial value of T_j to check for convergence and the process continues till a convergence in the value of T_j is achieved. However, if the value of T_j is not acceptable (too high for a chip) or there is no convergence within reasonable iterations, error messages will be generated and invalidate T_j. Besides, a chip-level reliability constraint is also applied to check the validity of the estimated value of T_j for the particular value of V_{dd} in the analysis. This methodology was also calibrated against measured power (both active and leakage), frequency and junction temperature data for a 32-bit microprocessor and the results were found to be in excellent agreement. The high performance IC product used in this case was fabricated using 130 nm technology with an I_{on}/I_{off} ratio of 660 and nominal $V_{dd} = 1.2V$. Once the value of T_j is achieved in an electrothermally coupled way, it is then solved self-consistently with the EM and self-heating equations as discussed in [10].

IV. IMPLICATIONS ON INTERCONNECT RELIABILITY AND DESIGN

Application of ET-coupled Methodology for Interconnect Reliability Estimation

It is important to mention that for *constant system power dissipation* (P_{sys}), employing ET-coupled methodology will always yield a lower estimation of junction temperature (T_j) than that obtained using non-ET-coupled methodology. This is because the ET-coupled methodology takes into account the coupling between P_{chip} and T_j (primarily due to the strong dependence of $P_{leakage}$ on T_j), where the latter will be limited by $P_{leakage}$, which cannot increase unbounded due to the constant system power constraint. The non-ET-coupled methodology will, therefore, clearly overestimate the junction temperature as it neglects the coupling between P_{chip} and T_j.

Since employing ET-coupled methodology will yield a lower junction temperature (T_j) as well as the metal temperature (T_m) of interconnect, more accurate and higher MTTF will be obtained. The difference in estimated MTTF for a unipolar line based on ET-coupled junction temperature and nominal junction temperature has been shown in **Figure 7** and (**Table 1**).

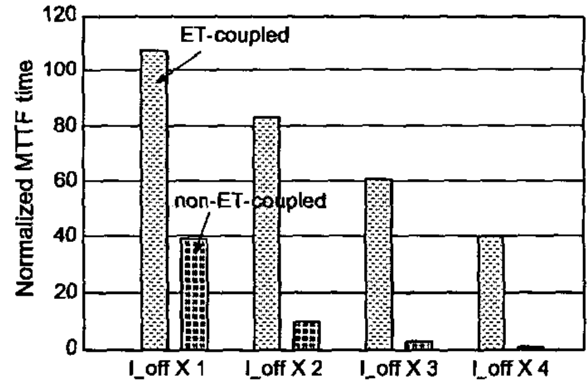

Figure 7. MTTF values of a unipolar line for different I_{off} values. The MTTF is normalized w.r.t I_{off} x4 for the non-ET-coupled methodology.

	I_{off} x 1	I_{off} x 2	I_{off} x 3	I_{off} x 4
MTTF ratio	**2.73**	**8.36**	**20.70**	**40.39**

Table 1. Estimated MTTF ratios between ET-coupled and non-ET-coupled methodologies as shown in figure 7.

It can be observed from figure 7 and table 1 that as technology gets leakier, the MTTF ratio increases. The ET-coupled methodology predicts higher interconnect lifetime as opposed to significantly lower MTTF by non-ET-coupled methodology. This is in agreement with yielding lower metal temperature corresponding to lower

junction temperature under ET-coupled consideration. As discussed earlier, under constant system power dissipation, lower junction temperature will be obtained by the application of the ET-coupled methodology. Also, estimating the junction temperature without electrothermally coupled methodology will lead to increasingly significant errors as the leakage power dissipation rises or the fabrication technology becomes leakier.

Impact of ET-coupled Methodology on Interconnect Current Density Limits in Power/Ground and Signal Lines

We now use the methodology outlined in the previous section for accurately determining the metal temperature and maximum allowed current density of interconnects. There are two major types of interconnects, signal lines and power lines. They differ in that currents in signal lines (clock and data busses are special cases of signal lines) are bidirectional (or bipolar), while those in power/ground lines are usually unidirectional (or unipolar). Therefore, the average current density J_{avg} in signal (bipolar) lines is much less than that of power (unipolar) lines because the current in signal lines has both positive and negative excursions [10]. Also, for signal (bipolar) lines, the EM problem is not as severe as lines with unidirectional current flow. The effect on metal temperature T_m for both power (unipolar) and signal (bipolar) lines at different I_{off} values are shown in **Figure 8**.

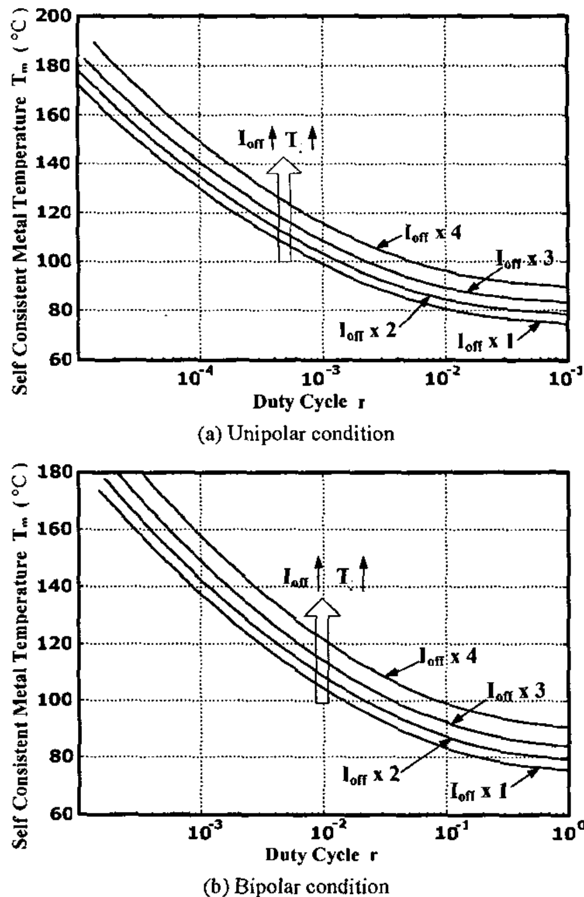

Figure 8. Metal temperature (T_m) is estimated using the ET-coupled methodology and plotted as a function of Duty cycle (r) for increasing values of I_{off}. (a) Unipolar condition (b) Bipolar condition

From figure 8, for a given duty cycle r, self-consistent interconnect metal temperature T_m increases with I_{off} under both unipolar and bipolar conditions. Also, the self-consistent interconnect metal temperature can be higher for the bipolar condition due to the higher EM lifetime under AC (bipolar) stress [10]. In both cases of bipolar and unipolar current conditions the self-consistent temperature increases as leakage increases. This is in tune with the fact that higher leakage will account for increase in total power dissipation, which leads to higher junction temperature.

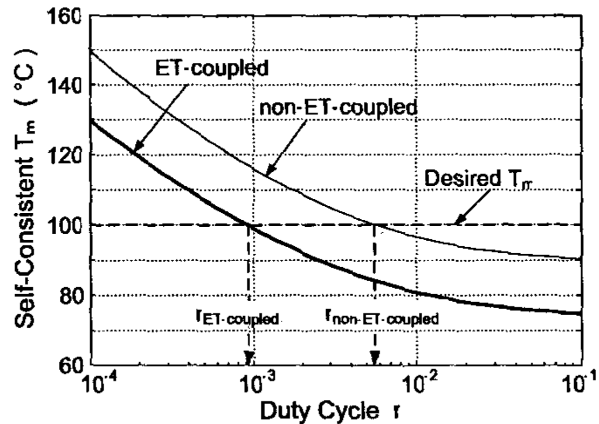

Figure 9a. Metal temperature (T_m) vs. Duty cycle (r) for a unipolar line. Non-ET-coupled metal temperature is also plotted. For any given metal temperature constraint (T_m=100°C condition is shown here), employing ET-coupled method yields a lower limit on the achievable duty cycle.

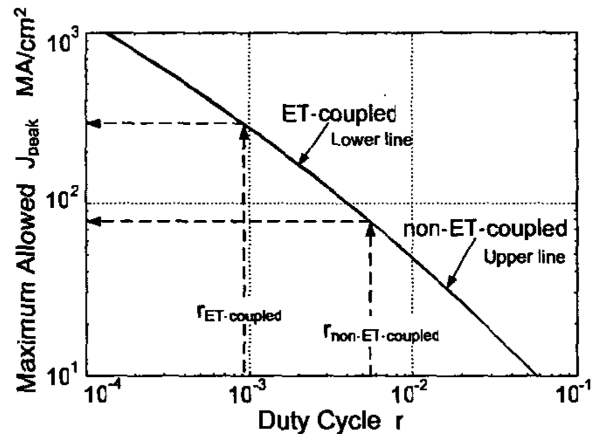

Figure 9b. Maximum allowed current density (J_{peak}) vs. duty cycle (r) for a unipolar line. J_{peak} curves estimated by ET-coupled and non-ET-coupled methods are very close but for a given metal temperature constraint (T_m=*100°C*) as illustrated in figure 9a, the ET-coupled approach yields a higher allowable J_{peak}.

Figure 9 provides an example to illustrate the advantage of employing the electrothermally coupled methodology. As discussed in the previous section, for a given system power dissipation, the electrothermally coupled methodology yields a lower junction temperature (T_j) as well as the metal temperature (T_m) of interconnects. Therefore, the metal temperature (T_m) of interconnect is always lower when applying electrothermally coupled methodology as shown in **Figure 9a**. Also, for a required metal temperature (T_m), the electrothermally coupled approach yields a

lower duty cycle. Hence, a higher peak current density (J_{peak}) can be allowed during the design phase as shown in **Figure 9b**. Similarly, peak current density limits can be obtained for signal lines once the exact shape of the bipolar waveform is known as explained in [10].

As introduced earlier in equation (3), the interconnect thermal impedance (R_θ) is a function of line geometry, surrounding insulator's thermal conductivity and thickness of underlying dielectric. It can be observed from **Figure 10** that for a given duty cycle, interconnects with higher R_θ result in higher metal temperatures (T_m) and thus reduce maximum allowed peak current density (J_{peak}). Also, as the leakage current increases, the effect gets stronger on metal temperature leading to two different limits on the duty cycle. This leads to two different J_{peak} values even though the curves for maximum allowed J_{peak} for different I_{off} values overlap.

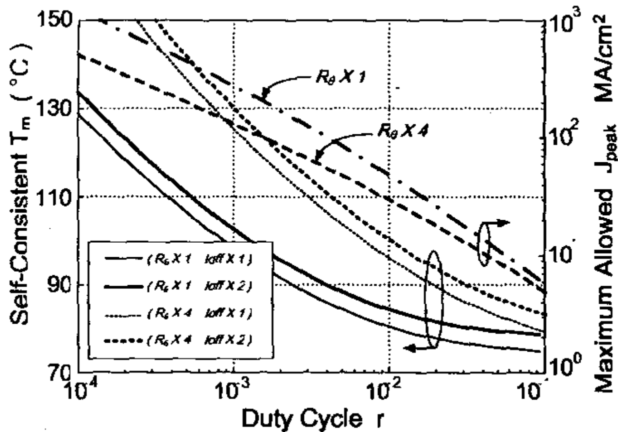

Figure 10. Metal temperature (T_m) as well as maximum allowed peak current density (J_{peak}) vs. Duty cycle (r) at different interconnect thermal impedance (R_θ) and different values of I_{off} (The curves for maximum allowed peak current density (J_{peak}) for different I_{off} values overlap).

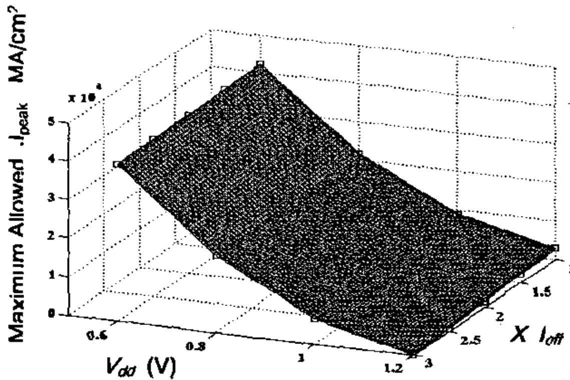

Figure 11. Maximum allowable peak current density (J_{peak}) as a function of off-state leakage current (I_{off}) and supply voltage level (V_{dd}) determined using the ET-coupled methodology for unipolar lines (X I_{off} indicates the increasing factor of I_{off}).

Finally, **Figure 11** shows the overall trend for the maximum allowable interconnect peak current density (J_{peak}) as a function of the chip supply voltage (V_{dd}), (which decreases with technology scaling) and off-state leakage current (I_{off}), (which is progressively increasing with technology scaling). It can be observed that for a given supply voltage V_{dd}, the maximum allowed J_{peak} decreases as I_{off} increases. For example, as shown in figure 11, at the *100* nm node, where nominal V_{dd} is *1.0V*, J_{peak} will decrease by *27.6%* if the technology becomes three times leakier, which is likely under process variations [8]. This result will have significant implications for leakage dominant sub-*100* nm *CMOS* technologies.

V. CONCLUSION

In conclusion, we have demonstrated that for leakage dominant technologies, estimation of the interconnect temperature should take into account various electrothermal couplings between the supply voltage, frequency, power dissipation and junction temperature of the chip. We have also quantified the impact of off-state leakage on EM design rules for both unipolar and bipolar stress conditions. It is shown that for a given system power dissipation, consideration of various electrothermal couplings in subthreshold leakage dominant technologies can lead to lower estimated junction and metal temperatures, which in turn lead to more accurate estimation of electromigration lifetime. This would allow designers to avoid employing overly conservative design rules and thereby improve performance. The results obtained in this work will also have important implications for burn-in testing of leakage dominant ICs.

ACKNOWLEDGEMENT

This work was supported by Intel Corporation, Fujitsu Labs of America and the University of California-MICRO program.

REFERENCES

[1] *International Technology Roadmap for Semiconductors* (ITRS), 2002 edition, http://public.itrs.net/

[2] S. Im and K. Banerjee, "Full Chip Thermal Analysis of Planar (2-D) and Vertically Integrated (3-D) High Performance ICs," *IEEE International Electron Devices Meeting* (IEDM), 2000, pp. 727-730.

[3] P. K. Chatterjee et al., "Trends for Deep Submicron VLSI and their Implications for Reliability," *in Proc. IEEE International Reliability Physics Symposium (IRPS)*, 1995, pp. 1-11.

[4] K. Banerjee et al., "A Self-Consistent Junction Temperature Estimation Methodology for Nanometer Scale ICs with Implications for Performance and Thermal Management," *IEEE International Electron Devices Meeting (IEDM)*, 2003, pp. 887-890.

[5] J. R. Black, "Electromigration—A brief survey and some recent results," *IEEE Trans. Electron Devices*, vol. ED-16, 1969, pp. 338–347.

[6] W.R. Hunter, "Self-consistent solutions for allowed interconnect current density—Part I: Implications for technology evolution," *IEEE Trans. Electron Devices*, vol. ED-44, 1997, pp. 304-309.

[7] P. P. Gelsinger, "Microprocessors for the new millennium: Challenges, opportunities, and new frontiers," *IEEE International Solid State Circuits Conference (ISSCC)*, 2001, pp. 22-25.

[8] V. De and S. Borkar, "Technology and design challenges for low power and high performance," *Proc. International Symposium on Low Power Electronics and Design (ISLPED)*, 1999, pp. 163-168.

[9] Y-S. Lin et al., "Leakage Scaling in Deep Submicron CMOS for SoC," *IEEE Trans. Electron Devices*, Vol. 49, No. 6, 2002, pp. 1034-1041.

[10] K. Banerjee and A. Mehrotra, "Global (Interconnect) Warming," *IEEE Circuits and Devices Magazine*, 2001, pp. 16-32.