FLAW: FPGA Lifetime AWareness

Suresh Srinivasan,
Prasanth Mangalagiri, Yuan Xie,
N. Vijaykrishnan
Dept. of Computer Science and Engineering,
Pennsylvania State University, PA 16802
{ssriniva, mangalag, yuanxie,
vijay}@cse.psu.edu

Karthik Sarpatwari
The Engineering Science and Mechanics Dept.,
Pennsylvania State University, PA 16802
kzs127@psu.edu

ABSTRACT

Aggressive scaling of technology has an adverse impact on the reliability of VLSI circuits. Apart from increasing transient error susceptibility, the circuits also become more vulnerable to permanent damage and failures due to different physical phenomenon. Such concerns have been recently demonstrated for regular micro-architectures. In this work we demonstrate the vulnerability of Field Programmable Gate Arrays (FPGA)s to two different types of hard errors, namely, Time Dependent Dielectric Breakdown (TDDB) and Electro-migration. We also analyze the performance degradation of FPGAs over time caused by Hot Carrier Effects (HCE). We also propose three novel techniques to counter such aging based failures and increase the lifetime of the device.

Catagories and Subject Descriptors:
B.8.1 [Hardware]: Performance and Reliability: Reliability, Testing and Fault-Tolerance

General Terms: Reliability, Algorithms, Experimentation

Keywords: FPGA, Time Dependent Dielecric Breakdown, Electromigration, Hot Carrier Effects

1. INTRODUCTION

Recently, Field Programmable Gate Arrays(FPGAs) have been quicker in adopting smaller feature size technologies than the microprocessor industry. While adoption of these newer technologies faster has provided significant performance and power benefits, they have also driven the need to address several reliability concerns that are emerging as serious concerns with smaller feature sizes. Apart from the increased vulnerability of the circuits to transient errors, for which redundancy based techniques have been extensively explored [1], there is an increase in the impact of different aging phenomenon that result in permanent failures of the devices and interconnect. Consequently, many accelerated aging mechanisms such as Hot Carrier Effect (HCE) [2], Time Dependent Dielectric Breakdown (TDDB) [3] [4], Electromigration (EM) [5], Thermal Cycling and Stress Migration [6] that have been extensively studied in the past, are revisited anew due to their increasing impact with device scaling [6].

The rate of degradation due to the accelerated aging phenomena is dependent on different factors such as supply voltage, temperature, switching activity,resulting current in the devices,and leakage currents. For example, the Mean Time To Failure (MTTF) due to TDDB reduces as gate leakage increases. The impact of switching activities on HCE and EM have been presented in recent works [7]. The authors also demonstrated a logical redesign strategy to reduce the impacts of circuit degradation due to HCE and EM. Circuit aging due to switching activity was presented in [8], which demonstrated an aging model, where the age is directly proportional to the signal switching probabilities. In a FPGA, based on the mapped design, different portions of the FPGA exhibit varying switching currents and leakage currents. Consequently, different portions of the FPGA will age differently when implementing the same design for a long period of time.

The first contribution of this paper is to analyze the aging impact of TDDB, EM and HCE on Xilinx style FPGAs using a set of MCNC benchmarks. Our results show that a significant portion of the FPGA resources may fail in the first 3 to 5 years of operation. Consequently, we propose three strategies that can prolong the lifetime of the device by mitigating the impact of the accelerated aging mechanisms. The first technique employs a gate leakage optimization technique to mitigate the impact of TDDB. The second technique employs a region constrained placement technique that periodically changes the placement of a design to age the different components of the FPGA more uniformly. The final technique focuses on mitigating impact of electromigration by periodically changing the routes (and interconnects) for the highly switching interconnects. Our experimental evaluation of the proposed techniques show that the lifetime of the device can be prolonged significantly using the proposed techniques.

The rest of this paper is organized as follows. Section 2 provides an overview on the experimental setup. Section 3 analyzes the impact of the three different accelerated aging mechanisms on the benchmark circuits. Section 4 introduces techniques to increase the lifetimes of the FPGAs, and section 5 provides conclusions.

2. EXPERIMENTAL SETUP

We used Xilinx ISE tools based [9] design flow to implement a set of MCNC benchmarks on the smallest devices possible. Table 1 shows the usage information of different MCNC benchmarks and the device they were implemented on. We obtained the switching activities and the transition probabilities of each of the nets using the placement and routing information provided by the Xilinx De-

DAC 2006, July 24–28, 2006, San Francisco, California, USA.

velopment Language (XDL) file, which is a text file comprising of all the placement and routing information.Different tools used for individual fault observation and solutions have been described over in the respective sections as and when we describe the particular technique.

Design	Slices Used	IO pins	Device
alu4	252	22	xc2v40
apex2	204	42	xc2v40
apex4	549	28	xc2v250
ex1010	567	20	xc2v250
ex5p	286	21	xc2v80
misex3	131	28	xc2v40
seq	605	76	xc2v250
spla	116	62	xc2v40
pdc	308	56	xc2v80

Table 1: Characteristics of MCNC benchmarks used

3. TYPES OF ERRORS AND THEIR IMPACTS ON FPGAS

The permanent aging based hardware failures that we study in this paper include TDDB, EM and HCE. In this section we describe the cause and impact of each of the errors in details.

3.1 Time Dependent Dielectric Breakdown

With reduction in the gate oxide thickness, Time Dependent Dielectric Breakdown (TDDB) is one of the major reasons of permanent failures in recent VLSI technologies. The cause for such failures is primarily due to the trapping of charges in the oxide that create a electric field, followed by charge flow through the oxide, resulting in a breakdown after sometime. The MTTF due to TDDB is a strong factor of temperature and the gate leakage component of a transistor. With increasing gate currents there is an increase in the charges trapped in the oxide which create an electric filed, which keeps increasing until the dielectric breakdown threshold is reached. Such an increase over the dielectric breakdown threshold leads to a runaway problem, since it leads to an increased gate leakage, followed by increased charge trapping. Such a phenomenon is aggravated further due to defects in the oxide while manufacturing that is on an increase with the scaling of technologies.

The MTTF due to TDDB is strongly dependent on the thickness and area of the oxide, the gate voltage, the temperature and the leakage current through the gate. Time to breakdown (t_{bd}), was empirically evaluated [3] for an older technology (t_{ox1}=10nm , A_1=1mm^2, V_1=7V), where t_{ox1} depicts the oxide thickness, A_1 depicts the area of the oxide and V_1 is the applied gate voltage. Using this empirical value, we evaluate t_{bd}, for a newer technology (t_{ox2}, A_2, V_2) using equation(1).

$$\frac{MTTF_{tddb1}}{MTTF_{tddb2}} = 10^{\frac{t_{ox1}-t_{ox2}}{0.22}} (\frac{V_1}{V_2})^{a-bT} e^{\frac{X+\frac{Y}{T}+ZT}{kT}} \quad (1)$$

The values of constants in the equation, X, Y, Z, a, and b were obtained from empirically fitted numbers for scaling from [6]. Note that, the first term in equation (1), $10^{\frac{t_{ox1}-t_{ox2}}{0.22}}$ is actually the factor used to compare the gate leakage current, I_{leak}, for the two different technologies. Hence, this provides an estimate of the actual time to breakdown for a specified amount of gate leakage current. However, gate leakage varies based on the state of the device. In this work, we evaluate the gate leakage number of different components of the FPGA using the static signal probabilities. To validate the scaling model with the existing data available, we first obtained the MTTF due to TDDB under stressed conditions, for 90nm technology node, with t_{ox} of 1.2nm and oxide area of 0.25mm^2, at 125 deg.C and 3.0V. The static signal probability being assumed to be 1 at this point, since the gate is always turned on. We obtained a MTTF of 9842 hrs under stress testing for a gate oxide breakdown, which is quite similar to measured failure numbers provided for FPGAs in 90nm technology [9].

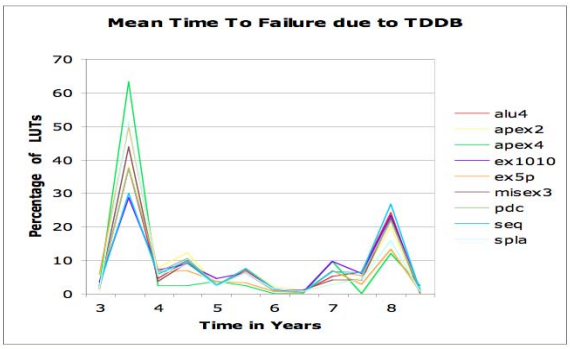

Figure 1: MTTF of LUTs based on their active gate leakage consumption

In this work we have considered the failure in the Look Up Tables (LUTs) in FPGAs. LUTs primarily comprise of 16x1 multiplexer circuit which are designed using pass transistor logic. The gate inputs are driven by different nets, while the actual inputs of the multiplexer are preconfigured and stored in SRAM cells. We custom designed a layout in 70nm CMOS technology and computed the gate leakage power using BSIM4 model [14]. Note that the gate leakage of the LUTs in each of the slices depends on the input vector, thereby the static signal probabilities. Using the static probabilities and the inputs from the SRAM cells to each of the LUT multiplexers for the benchmarks, we obtain the gate leakage power. The gate leakage numbers along with the static probability values were used to determine the MTTF due to TDDB for each of the LUTs. We obtained the MTTF for different components of the FPGA based on their active gate leakage consumption, and predicted their lifetimes using scaling equation(1).

Figure 1 demonstrates the failure rates for different LUTs in a FPGA for nine benchmarks from MCNC suite. It is clear from the figure that there are significant number of LUTs in a low lifetime region (fail less than 4yrs). This can be attributed to the observations represented in [10], which shows that the distribution gate leakage in LUTs is mainly towards low and high extremes. This observation calls for balancing static probabilities or using techniques which may reduce the maximum gate leakage consumption of the LUTs. A similar analysis can be easily performed for the interconnect components as well using the gate leakage characterization of interconnect available from [12].

3.2 Electromigration

Electromigration is one of the causes of permanent failures of interconnect in the form of development of voids in metal lines, due to heavy current densities over a period of time. Such a phenomenon is greatly aggravated by increasing current densities, and smaller feature sizes of the interconnect wires. Based on the most commonly adopted Black's model for Electromigration, the MTTF of a metal line depends primarily on the current density and the

length of the wire. To obtain the MTTF for a Cu wire for 90nm technology, we used a tool called SysRel [5], which takes in as its input the CMOS layout of any circuit and provides the time to failure plots for wires in the design which are prone to failures due to EM. We provided the tool with a simple inverter driven wire layout of length 100um and, obtained its MTTF of the circuit that was observed to be close to 4 years. Such an MTTF was then scaled using a combined equation from [6] and [5] as shown in 2,

$$MTTF_{em} \propto \frac{(J)^{-n} e^{\frac{E_{aEM}}{kT}}}{L} \quad (2)$$

where J is the current density, n is a constant, which is typically equal to 1 [5], L is the length of the wire, E_{aEM} is a constant, k is the Boltzmann's constant, and T is the temperature of the copper interconnect. Since the current density through a wire has a direct impact on the transition probabilities of the net driving the wire, the lifetime of the wire may be characterized as an inverse proportion of the transition probabilities and the length of a wire segment. FPGAs have variable length segments as demonstrated in Figure 2 and significantly varying transition probabilities of different nets.As a consequence, the lifetimes of the wires may vary considerably for a design mapped on an FPGA. Such variations in switching activities are shown in figure 3, which demonstrates the region based distribution of switching activities for an MCNC benchmark alu4 mapped onto a VirtexII xc2v40 device. It is also important to note that, the effective length used for computing such failure periods are the longest runs between two vias through which there is a current flow. It can be observed from the figure 2 that a long wire can have, a current path spanning minimally 1, to as large as the length or width of the FPGA. Similar, a hex may have its current path spanning either 2, 3 or 6 CLBs. We obtained such an information about the effective lengths of the wires by parsing the post-routing information file provided by XilinX, namely the Xilinx Development Language(XDL), for different wires. Using JBits we obtained the transition densities of different wires for MCNC designs mapped onto the smallest possible VirtexII device. Based on their lengths and transition probabilities we estimated the MTTF for different wires. Figure 4, demonstrates the cumulative failure times of the wires in FPGA due to electromigration. Wires beyond 20 year lifetimes are considered immortal and are not depicted in the figure. Figure 5, the cumulative failure times of different wires in FPGA architecture, that provides a clear distinction on the focus of reliability techniques to safe guard the longer wires.

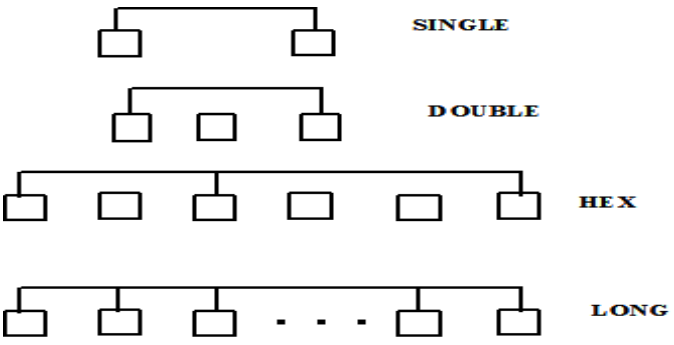

Figure 2: Different wire segments used in an FPGA

3.3 Hot Carrier Effects

Hot carrier effects are attributed to slow creation of traps at the oxide surface. Such interface trap generation may affect the I-V

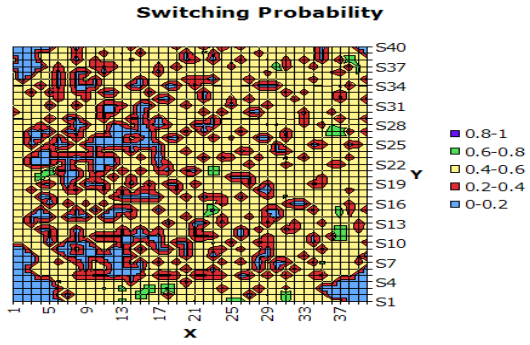

Figure 3: Switching probability distribution in alu4 benchmark

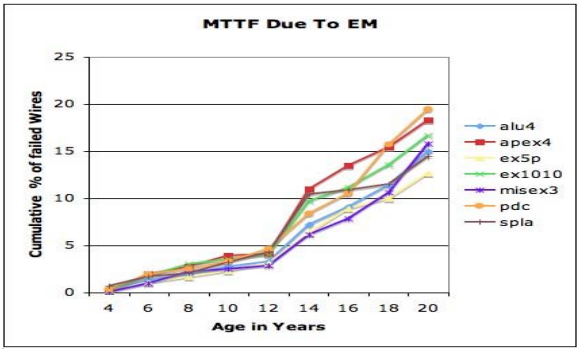

Figure 4: MTTF for different wires in FPGA based on their lengths and switching

characteristics of the transistors and thereby have an impact on the circuit characteristics over a period of time. The number of electrons trapped in the oxide is proportional to the substrate current passing through transistor and is computed using analytical models presented in [16]. The immediate impact of such trap generation is the change in the threshold voltages of the transistors which consequently affects the power and performance of the circuits.

Figure 6 demonstrates the degradation of the threshold voltage over time for a NMOS transistor obtained from our simulations, when there is a continuous current flow through the device for 70nm technology parameters. There is nearly an increase in threshold voltage of 0.3V over a period of 300 hrs, that however saturates since, as the device degrades, V_{th} approaches V_{gs}, due to which device behavior shifts to sub threshold region behavior. The

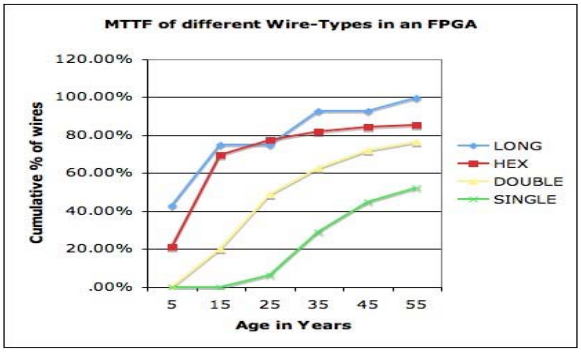

Figure 5: MTTF for different wires types in FPGA

actual age correlation for a transistor is obtained using the substrate current passing through the transistor any time it switches. However, there is only a small period of time during the transition when there is substrate current passing through the device. Consequently, the actual age of the transistors depend upon the switching of the transistor and the rise and fall delay.

Xilinx based FPGA architectures comprise of pass transistor based multiplexers for their LUT and routing fabric. Since each of the transistors in such multiplexers is in the critical path, the degradation of such multiplexers depends on the maximum switching activity of its transistors. The impact on performance of the LUT over time, when its inputs are switching with a probability of 0.5, is demonstrated in figure 7. Consequently, the circuit mapped onto the FPGA may no more be able to operate in the same frequency as before over time, due to degradation in the delay of individual LUTs. To observe such an impact on the frequency we incorporated the age degradation impact on each of the LUTs and obtained a post placement timing estimate using Versatile Place and Route (VPR) tool, for different MCNC benchmarks. There is a degradation of nearly 6% on an average for all the benchmark designs. Figure 8 demonstrates such performance degradation over 3.5 years, of 5 representative MCNC benchmarks implemented on a FPGA.

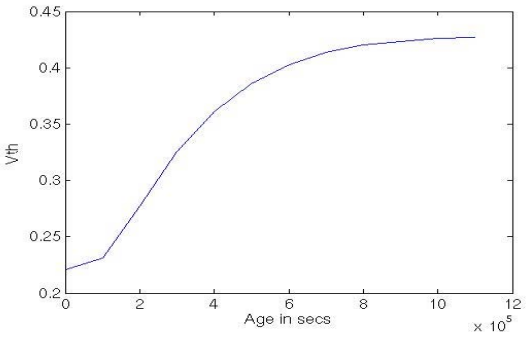

Figure 6: Threshold degradation with time(continuous current flow conditions)

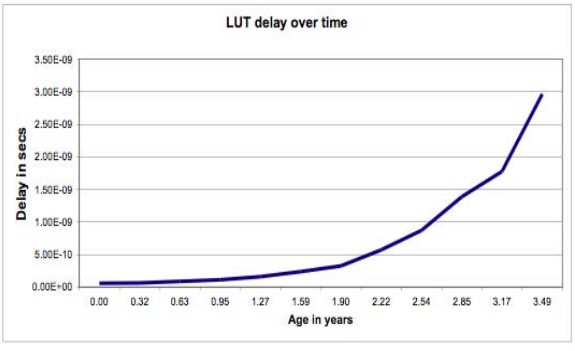

Figure 7: LUT delay degradation over time (actual age)

4. TECHNIQUES FOR UNIFORM AGING

Many simple techniques may be employed to reduce the impact of aging and increase the lifetime of the FPGAs. The techniques applied are based upon the parameters that each of the failures are dependent upon. In this, paper we demonstrate three techniques which may be employed to improve the average lifetime of the FPGAs due to the three main causes of permanent failures and device degradation described in section 3.

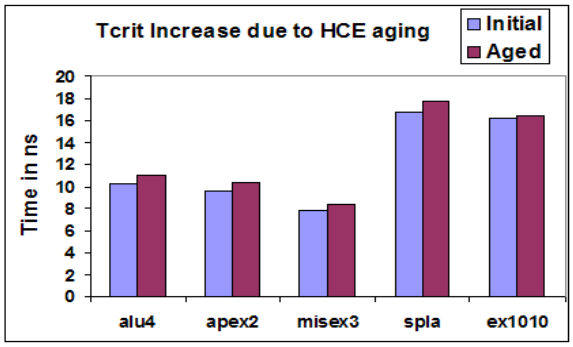

Figure 8: Performance degradation of different MCNC benchmarks over 3.5yrs

4.1 Improving MTTF due to TDDB

As described in section 3, TDDB is a strong function of the gate leakage of any transistor. Consequently, standard gate leakage optimization techniques may be employed to reduce the gate leakage power consumed by the individual components of the FPGA and thereby increase the MTTF due to TDDB.

We employed the input vector technique proposed in [10] to optimize upon the active leakage power consumed by the logic blocks and interconnect routing multiplexers. It is important to note that the lifetime of the circuit is more dependent on the maximum gate leakage of any component instead of the overall gate leakage of the whole device, since the device with maximum active gate leakage is expected to fail the earliest. Figure 9 depicts the average increase in the lifetime of the device on employing one such gate leakage optimization technique proposed in [10]. It is interesting to observe the significant impact of such power optimization scheme on the average lifetime of the device, for different benchmarks. The results depict an average improvement in the MTTF due to TDDB by 24% averaged over different benchmark designs. Moreover, such an approach for increasing the lifetime has no performance impact. Figure 10 demonstrates the percentage of LUTs that may fail at a given period of time before and after the optimization proposed for two representative benchmarks. It is clear from the figure that the percentage of LUTs with low lifetimes significantly drops due to the proposed gate leakage optimization.

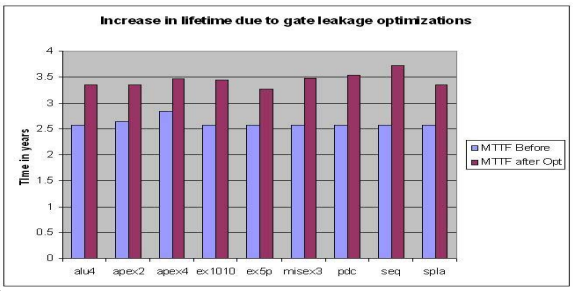

Figure 9: Increase in lifetime due to gate leakage optimizations

4.2 Load balancing to mitigate HCE impact

The degradation of different resources in FPGAs depends on their usage. Since most FPGAs have a utilization of 60% in logic blocks, and still lower utilization of 40% of routing resources, we exploit the ability of re-mapping the design onto the unused por-

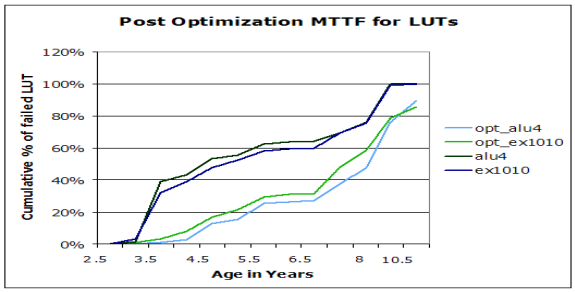

Figure 10: Cumulative increase in lifetime due to gate leakage optimizations

tions of the device to prolong the lifetime of the device and minimize the impact of degradation on performance of the device.

A Region Constrained Placement for Reliability (RCPFR) approach is proposed, that performs periodic re-mapping of the design to less used regions for increasing the lifetimes of the device.As shown in Figure 11 The whole device is treated as a set of rectangular regions and all the designs are mapped using the smallest possible number of rectangles. Each of the regions' average and maximum switching activity is estimated, which provides an estimate on the lifetime of the device. Using this information a new placement is generated, avoiding the regions which have aged most in the initial configuration. The shaded blocks in Figure 11 show the used portion of the FPGA before and after the reconfiguration for reliability. The darker blocks are the most used blocks, with the highest average switching activites and hence are avoided while generating the new placement, whenever possible.

Figure 12, demonstrates the implementation flow of the proposed scheme. Such a placement shift however may not be optimal with respect to the operating frequencies of the device. However, as demonstrated in section 3, the performance of the circuit anyways reduces by nearly 6% over time, and therefore the performance degradation due to RCPFR strategy, may still prove better with respect to both the lifetime and performance benefits. Table 2 demonstrates the increase in the lifetimes of different designs and the corresponding performance impact of such a modification in the placement. We observe an average increase in the lifetimes of the designs by 28%, however the same circuit frequency is not achievable, which is due to over constraining the design. There is an average performance impact of 1.53% . Note that the lifetime of some of the designs double due to the new placement of the design. This behavior is due to the fact that those designs had their slice usage less than half the total number of slices available and hence could be moved to an entire new region in the FPGA.

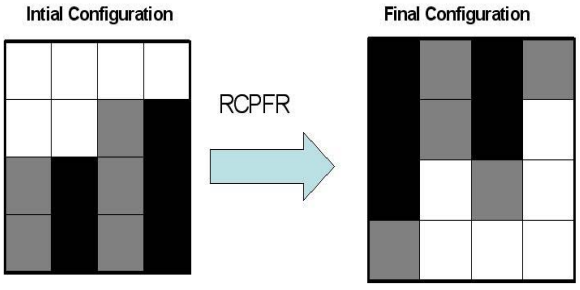

Figure 11: The RCPFR strategy

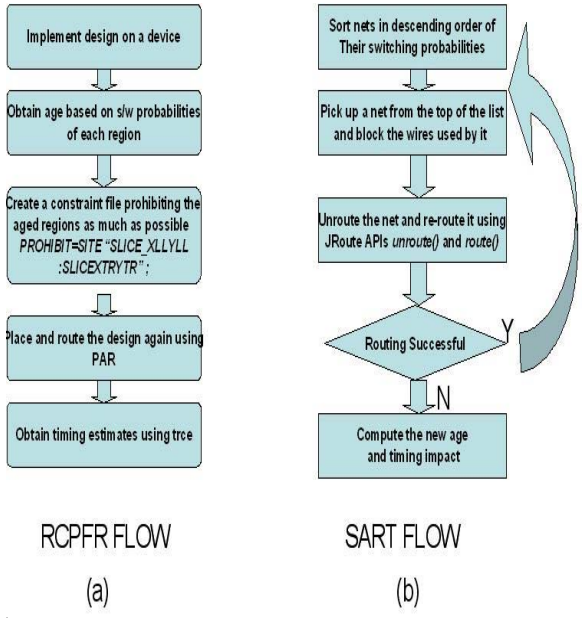

Figure 12: The implementation flow for RCPFR and SART

Design	Increase Age %	freq. impact %
alu4	22.68	3.29
apex2	4.2	1.91
apex4	32	0.74
ex1010	20.48	0
ex5p	12	0.11
misex3	16	2.9
seq	100	0.24
spla	28	3.06

Table 2: Age impact on MCNC benchmarks after RCPFR

4.3 Selective Alternate Routing Technique to increase MTTF due to EM

As demonstrated in section 3, the lifetimes of different wire segments due to Electromigration based failures in an FPGA vary significantly because of the length and the switching characteristics of the nets driving them. Most of the FPGA vendors typically provide large number of routing resources to avoid any type of timing conflicts. Consequently, even a design with as high slice utilization as 99% has close to 60% of unused routing resources [12]. This provides an opportunity to replace the aged wires by unused wire segments to achieve the same pin connections.

We propose a dynamic reconfiguration based Selective Alternate Routing Technique(SART), which may improve the lifetime of the FPGA wires due to EM failures as demonstrated in Figure 13. Note that even with alternate routing a small portion of redundant wire is shared by both the routes. However, the length of the wire may not be sufficiently long to cross the mortality length or even if it does, the MTTF of that wire will be really huge. Since the nets driving the wires determine the aging impact due to Electromigration, the SART approach may be employed in the priority order of the switching probabilities of different nets in the design. The SART technique if applied dynamically over time may, provide benefits. Firstly, in the form of increased operating lifetimes and secondly, by avoiding the performance degradation due to the performance

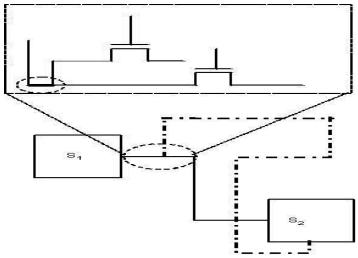

Figure 13: The SART technique: The dotted line indicates the alternate path

impact of Electro Migration on the wires. Such a technique may directly be employed by modifying the bit-stream, using tools that may read back and directly modify the bit-stream. Note that the alternate route proposed may not be the optimal route with respect to the timing.Due to the presence of large number of routing resources, however, there is a good possibility of finding a route almost equal to the existing route with respect to timing. We used a tool JRoute [13], which provides the ability to implement SART directly on the bit-stream and generate a new bit-stream. JRoute is a set of API's incorporated with JBits, which is another java tool that is provided by Xilinx to read back information from bitstreams and provide partial/full dynamic reconfiguration ability. Figure 12 demonstrates the algorithm to selectively reroute the wires. Based on the switching probabilities we continue to reroute the wires until we obtain any unroutable path. This, essentially restricts the age benefits that may be obtained, since the path with highest switching activity that may not be rerouted will really determine the MTTF of the device (for the given design and placement) due to electromigration failures. Table 3 provides an estimate on the percentage improvement in the lifetime of the device due to the proposed optimization along with the average performance degradation due to such rerouting.

Design	Increase Age %	Maximum Freq Decrease
alu4	13.66	1.7
apex2	4.76	0.15
ex1010	15.97	1.82
ex5p	40	1.1
misex3	7.42	0.66
seq	14.51	2.25
spla	2.48	1.63

Table 3: Age impact on MCNC benchmarks after SART technique

5. CONCLUSION

In this paper we demonstrate the impact of aging on lifetime and performance of different components of FPGAs. Modeling of three different failure phenomenon namely, TDDB, HCE and EM, and the lifetime of the FPGA devices due to such failures are charecterized. The FPGA lifetime is greatly affected by the unbalanced signal switching probabilities and the gate leakage of different components of the device. Using such an observation we have proposed three techniques to increase the average lifetime of the devices and have demonstrated their effectivenes on different applications mapped onto FPGAs.

6. ACKNOWLEDGEMENTS

This work was supported in part by grants from DARPA,MARCO, GSRC,NSF 0093085 Career:Compiling for Energy and NSF 0202007.

7. REFERENCES

[1] P. K. Samudrala, J. Ramos, and S. Katkoori "Selective Triple Modular Redundancy for SEU Mitigation in FPGAs" In *Proceedings of Military and Aerospace Applications of Programmable Logic and Devices (MAPLD)*, 2003.

[2] S. Mahapatra, V. R. Rao, B. Cheng, M. Khare, C. D. Parikh, J. C. S. Woo and J. M. Vasi. "Performance and hot-carrier reliability of 100 nm channel lengthjet vapor deposited Si3N4 MNSFETs" *IEEE Transactions on Electron Devices*, vol.48, (no.4), April 2001. p.679-84.

[3] E. Rosenbaum, P. M. Lee, R. Moazzami, P. K. Ko and C. Hu "Circuit reliability simulator - oxide breakdown module" *International Electron Devices Manufacturing Technology Digest*, p. 331, 1989.

[4] J. H. Stathis "Reliability limits for the Gate Insulator in CMOS Technology" In *IBM journal of R&D, Vol. 46*, 2002.

[5] S. M. Alam, C. L. Gan, D. E. Troxel, and C. V. Thompson "Circuit-Level Reliability Analysis of Cu Interconnects" In *Proceedings of International Symposium on Quality Electronics Design(ISQED)* , 2004.

[6] J. Srinivasan, S. V. Adve, P. Bose and J. A. Rivers, "The Impact of Technology Scaling on Lifetime Reliability " In *Proceedings of International Conference on Dependable Systems and Networks(DSN)*, 2004.

[7] X. Xuan, A. Chatterjee, and A. D. Singh "Local Redesign for Reliability of CMOS Digital Circuits Under Device Degradation" In *proceedings of International Reliability Physics Symposium(IRPS)*, 2004.

[8] F. N. Najm "Transition density, a stochastic measure of activity in digital circuits" In *Proceedings of Annual ACM IEEE Design Automation Conference*, 1991.

[9] " www.xilinx.com" Xilinx product datasheet for Virtex-II.

[10] J. H. Anderson, F. Najm, and T. Tuan. "Active leakage power optimization for FPGAs," In *Proceedings of ACM/SIGDA International Symposium on Field-programmable gate arrays*, 2004.

[11] "Critical Reliability Challenges for the International Technology Roadmap for Semiconductors" In *International Sematech Technology transfer 03024377A-TR*, 2003.

[12] S. Srinivasan, A. Gayasen, N. VijayKrishnan and T. Tuan "Leakage control in FPGA routing fabric" In *Proceedings of Asia-Pacific Design Automation Conference (ASPDAC)* , 2005.

[13] E. Keller "JRoute: A Run-Time Routing API for FPGA Hardware. " In *Proceedings of Seventh Reconfigurable Architectures Workshop (RAW)* , 2000.

[14] " BPTM 65nm: Berkeley Predictive Technology Model. "

[15] V. Betz and J. Rose "VPR: A new packing, placement and routing tool for FPGA research" In *Proceedings of the 7th International Workshop on Field-Programmable Logic and Applications*, 1997.

[16] Y. S. Jean and C. Y. Wu "The threshold-voltage model of MOSFET devices with localized interface charge" In *Proceedings of IEEE Transactions on Electron Devices*, 1997.