Mitigating Variability from Technology Scaling

Comparing Timing-Aware, Task Assignment, and Re-synthesis Strategies

Arunkumar Ganesan
Department of Electrical Engineering and Computer Science
University of Central Florida
Orlando, USA
g.arunkumar@knights.ucf.edu

***Abstract*—Technology scaling has resulted in the reduced size of the device components and increased the density of devices that can be implemented on an Integrated Circuit. Equally technology scaling has also increased the variability of transistor characteristics as size is shrinking. Variability issues should be considered seriously in order to design a Reliable system. Mitigating these Variability concerns is a hard time process and each and every part of the system has its own play in variability Mitigation. This paper concentrates on the Mitigation of Various Variability in a device through the techniques of Timing Speculation, Execution-time Task Assignment and Design-time Re-synthesis.**

Keywords—Technology scaling; Variability; Variability Mitigation; Voltage Scaling; Voltage Regulators)

I. INTRODUCTION

Technology Scaling is the most important area that has been concentrated for the decades in the field of VLSI. Technology scaling has increased the VLSI performance by allowing to reduce a size of the transistors and increased the possibility of integrating billions of transistors. Moore's law states the number of Transistors on Integrated circuit doubles for approximately every two years. Power, Energy, Variability, Reliability are the barriers for the future scaling. The size of the Transistor has been so far decreased that billions of Transistor can be used with in an Integrated Circuit. Next the Focus has shifted on Power consumption, Power dissipation and Power delivery of these billions of transistors in an Integrated Circuits. Transistor subthreshold leakage has been increasing as technology scales down and ways for avoiding, tolerating and Controlling these subthreshold leakage becomes necessary in an Integrated Circuit. As Technology scales there will be a problem of Variability, Soft errors, Device performance degradation resulting in the unreliability of the system. It becomes necessary to implement an innovative design to mitigate these issues in a system to make it Reliable.

II. SOURCES OF VARIATION

There are three main sources for variation in the transistor behavior. These are Random dopant Fluctuation, Sub wavelength lithography and Supply voltage variations. These three sources fall into the category of Static and Dynamic sources.

A.Static

Random Dopant Fluctuation

Random Dopant Fluctuation is due to the discreetness of the dopant atoms in the channel of the transistor. The threshold voltage is controlled by doping the transistor channel with the dopant atoms doping can decide the characteristics of the semiconductor. The Figure shows the number of dopant atoms that is doped with the transistor for certain Technology node.

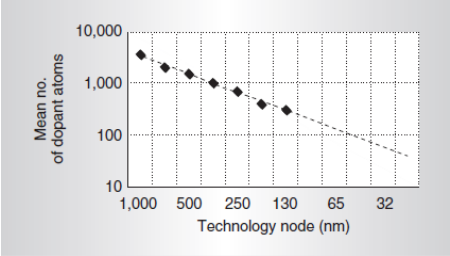

Fig 1 Dopant atom reduction(Borkar 2005)

From the above graph it can inferred that for a micron technology node there are thousands of dopant atoms and as the technology node decreases the number of dopant atom decreases and for the 32 – 16 nm (nanometer) it has decreased to an order of 10s of dopant atoms. With these number of small dopant atoms any small discreetness or changes in the characteristeristics of these dopant atoms can affect the transistors which can cause variability.

Sub Wavelength Lithography

Sub Wavelength Lithography is used for patterning the transistors since 0.25µm technology node. 248nm light wavelength is used for patterning the transistor of size 250nm and 180nm transistors. 193nm wavelength light is used for patterning the 130nm technology node and the same is used for patterning 65nm node. This will widen the difference

between the wavelength of light and the patterning width until 13nm wavelength light concept is introduced for patterning the 65nm and lesser node. Until subwavelength lithography can cause roughness in the edge of designed pattern results in Variation.

B.Dynamic

Dynamic variations occurs during the operations of the transistor and they can change over time i.e. they are time and context variant. For example thermal variations are variations due to the temperature variations which can create hot spots in a circuit degrading the circuit performance. This is due to the fact that an Integrated circuit different unit has different functionality. The Unit that has to perform larger operations need to supply larger current and hence as power increases the temperature also increases resulting in variations

$$P_L \alpha nC_L V_{DD}^2 f \quad (1)$$

From the above equation it is clear that the switching frequency increases when the power increases which also requires in an increase in the supply voltage.

Supply Voltage Scaling will have a worst impact on the future technology scaling which results in a supply voltage variations. Supply voltage scaling are carried out in order to reduce the power consumption of an integrated circuit when the devices are scaled down to smaller size.

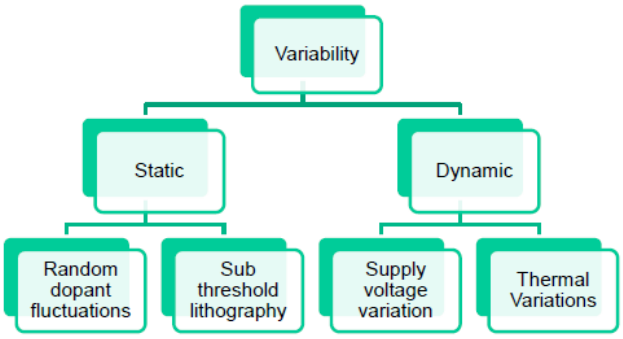

Fig 2 Classification of Variations

III. AFTERMATH OF VARIATIONS

Both Static and Dynamic variations have a greater impact on the performance of an Integrated circuit. Today it has been estimated that due to the above mentioned reasons there is 30 percent variation in operating frequency and 5 to 10 times variation in leakage power. Sub threshold leakage is occurred due to the difference between the threshold voltage and the near threshold voltage which can put the transistor in inversion region. This Sub threshold leakage can contributes about 50 percent variations in total power consumption. This makes the difference between the intended design and the obtained result which could make the system unreliable. So it becomes necessary to find a solution for mitigating these variations to make the system reliable.

Variations also has its impact on test methodology. During Scaling of Transistors the gate dielectric thickness must decrease to improve performance and reduce short channel effects. But the oxide scaling has resulted in the increase in gate leakage current exponentially. This affects the Burn-in test. In burn in test chips are stressed with higher supply voltage at higher temperature for shorter period of time to speed up aging. A fault in chip shows up early during burn in and hence it is caught. Since the gate leakage current increases exponentially with supply voltage, leakage power during burn in could become high which may fail the Burn in test.

IV. VARIABILITY MITIGATION

So far lot of process, methods and design changes have been implemented to mitigate these variations. An example for such a circuit design is that Forward and Reverse body bias is used to control the sub threshold leakage and frequency distributions. i.e. Chips with higher leakage will be faster so using of reverse bias will slow down the operation of the chip and the chips that are slow can be made fast by using Forward bias which requires a slight increase in leakage power. Adaptive supply voltage using body bias technique can control the distribution as well as it can reduce the power dissipation especially in high fan-in logic circuits.

Chip frequency depends on the speed of the critical path. Delays are determined by the designers and the critical paths are designed to satisfy the requirements. But due to static and dynamic variations these delays are changed. After designing of the transistors and integrated circuits designers will downsize the transistors in order to save the active power. When downsizing the transistors, the transistors which are close to the path of the critical path will also get downsized resulting in the increase in critical path. This will reduce the probability of meeting the frequency goal. It is important to find an alternate method of Downsizing because indiscriminate downsizing can result in making non critical path critical. In a conventional microarchitecture design the frequency of operation is increased by increasing the number of critical paths which reduces the probability of meeting the frequency goal. In order to achieve high frequency goals, gate delays are employed in the clock cycle but this is ineffective in averaging and cancelling the impact of variations.

Soft errors are the errors that are occurs due to hitting of alpha particles on silicon chips which can induce some charge on the nodes that flip the memory cell. It is easy to determine and eliminate these errors using parity checker and using some error correcting codes in the memory. But if this soft error occur in logic flip flop then it is difficult to detect and correct this error. Researchers have estimate that about 8 percent increase in soft error rate per logic state bit each technology generation.

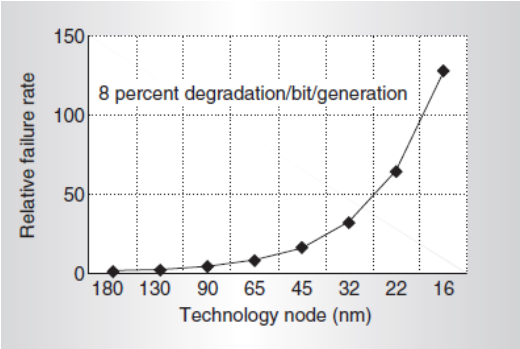

Fig 3 Relative Failure Rate(Borkar 2005)

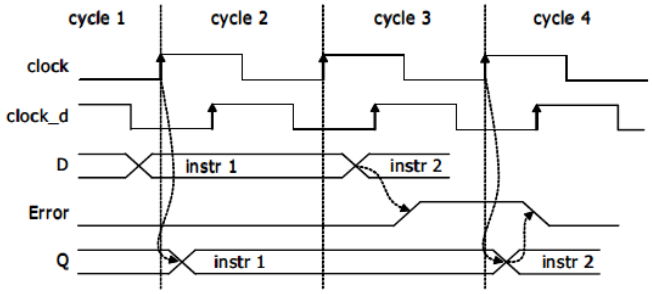

Fig. 5 Timing(Ernst D. 2003)

So it becomes necessary to replace the regular flip flops by soft error tolerant flip flops which can increase the soft error tolerance 10 times. Razor is such a techniques used for determinig and correcting these dynamic errors. Razor is used not only for getecting and correcting errors but also used for its energy efficiency. Because Razor technique does not replicate all the hardware but the flip flop which are critical and reqires checking for correctness.

A.Razor

In Razor technique, pipeline latches augmented with the Shadow latches that hold signals that arrive after the delay. Razor tolerates the timing errors due to Dynamic voltage scaling using Time Redundant shadow latch. The Error detection mechanism is that Razor will find the error if there is a mismatch between the main latch and shadow latch. Razor is based on dynamic detection and correction of speed path failures in digital design. The Razor employs the key idea of tuning the supply voltage by monitoring the error rate during operation. An important feature of Razor is that when Razor operates at sub critical supply voltages it will not result in any ruinous failure but instead there will be some supply voltage penalty.

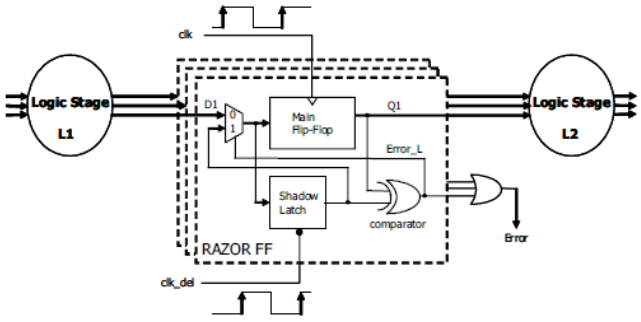

Fig. 4 Flip Flop with Razor(Ernst D. 2003)

B.Error correction and Detection Razor

Each flip flop in the circuit is connected to the shadow latch which is controlled by the delayed clock. The Operation of Razor flip flop can be explained from the fig.5

In the clock cycle 1, the combinational logic L1 meets the setup time by the rising edge of the clock and both the main flip flop and the shadow latch will latch the correct output. Since there is no mismatch between the main flip flop and the shadow latch, the error signal at the output of the XOR gate remains low and hence the operation of the pipe remains unchanged. In the Fig. 5 it is shown that at cycle 2 the combinational logic exceeds the intended delay due to sub critical voltage scaling. So the data is not latched by the main flip flop. But since the shadow latch operates at delayed clock it successfully latches the data at cycle 3. In order to guarantee that the shadow latch will always latch the correct input data, the logic delay is restricted at the design time such that under worst case condition the logic delay does not exceed the set up time of the shadow latch. By comparing the data of the shadow latch and the data in the main flip flop an error signal is produced in cycle 3 and in the subsequent cycle, cycle 4 the data is then restored from the shadow latch into the main flip flop and it is made available to the next pipeline stage L2.

If there is an error in the pipeline stage L1, then the data at the pipeline stage L2 is incorrect and hence it should be flushed from the pipeline. Since the shadow latch consists of the correct output data of the pipeline stage L1, the instruction is need not to be re executed through the failing pipeline stage L1 instead the instruction can be re executed by getting the output data from the shadow latch by the following pipeline stage L3. Hence using Razor technique ensures the forward progressing of the failed instruction. This Error recovery mechanism of Razor is used to correct the soft errors. The Construction of Razor should be in such a way that the power and delay overhead is minimized. The presence of the delayed clock introduces new short path restriction in the design.

In general Razor can be used for the purpose of power conservation if it is applied to all the parts of the microprocessor. But there are three important challenges that should be considered while applying the razor technique to all the parts of the microprocessor.

- The First challenge is the Detection and recovery of timing errors in the combinational logic in the pipeline datapaths.
- The Second challenge is the implementation of Razor in the on-chip SRAM structures because SRAM structures need to be implemented with the Razor compatible sense amplifiers.
- It is necessary to use Razor on pipeline control logic to restore the correct program execution in the presence of incorrect control decisions.

C.Parametric Variations at NTV

As technology scales down the resultant circuit suffers from the problem of growing power density. We can have a large number of devices in an integrated circuit but these devices requires large amount of power supply. On the other hand increase in power increases the temperature limit of the integrated circuit. A possible way to engage large number of cores in a computation is to operate at lower supply voltage V_{DD}. Lowering of V_{DD} to slightly above the threshold voltage V_{th} results in reduced energy per operation. This process is known as Near Threshold Voltage (NTV) computing. Operating at lower supply voltage will ultimately result in power conservation. But at NTV devices are more vulnerable to parametric variations i.e. The Parameters of Devices changes from their normal specifications. These variations slows down the devices and also results in leakage. In addition to that this will alter the core speed and memory structures because memory structures are more sensitive to discrepancies. To depend on the worst case operating margins is also not possible because the frequency here is already low because of the low supply voltage. Conventional variability mitigation technique such as Super Threshold Voltage (STV) can be adopted at NTV but it can certain limitations. These devices depend on the V_{DD} tuning on independent V_{DD} domains on chip. We can mitigate the problem of variation by an architectural model of parametric variation at NTV.

D. Impact of Parametric Variation at NTV

Parametric variations are the mismatch between the actual Values of the device and their intended value. The Parametric variations occurred during the manufacturing process. For example Within Die (WID) variations has a systematic and a random component. The systematic component is caused by the lithographic process and the random component is caused by the random dopant fluctuations. The Operating frequency of the processor or memory block depends on the threshold voltage V_{th} and the effective channel length L_{eff} of the transistor which is vulnerable to the variations. The WID variation in the V_{th} and L_{eff} results in widening the frequency distribution which leads to reduce in the operating frequency. Higher variations in V_{th} and L_{eff} results in variations in the frequency, core speed and memory. Since V_{DD} is close to V_{th} the transistor's switching speed will be very sensitive to variations. The timing guard band required to tolerate certain limit of V_{th} Variation increases as V_{DD} decreases.

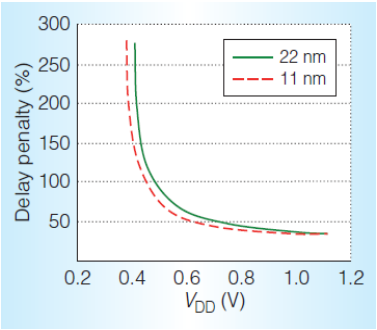

Fig. 6 Device degradation(Karpuzcu 2013)

E.Energysmart technique

Currently variation at STV is mitigated by using Adaptive Body Biasing (ABB) technique and multiple on chip V_{DD} domains but these techniques are not effective at NTV. Multiple on chip V_{DD} has certain limitations which makes it not efficient to use for variability tolerant design at NTV. These limitations are

- Power efficiency of on chip V_{DD} switching regulators are limited 75 to 90 percent range for operating conditions. This requires on chip V_{DD} regulators. These Voltage Regulators (VR) with high frequency are much in demand to meet future microprocessor's requirement. But the high switching loss and body diode loss suffer the efficiency of VR. Two Stage approach for VR reduces the switching loss thereby increasing the VR efficiency.
- Smaller V_{DD} domain are vulnerable to deeper V_{DD} droops and in order to reduce the droop errors it becomes necessary to have larger V_{DD} guardbands in these small V_{DD} domain.
- To have separate on chip V_{DD} regulators and having large number of cores will increase the area of the device

These limitations lead to a design of an architecture for future NTV many core chips. Energysmart technique is proposed that avoids using Multiple on chip V_{DD} domains for energy efficiency. This technique keeps a single V_{DD} in the whole chip and supports dynamic voltage and frequency scaling (DVFS), but applies it globally across the chip. Energy smart support frequency domain. It is organized in cluster cores, where each cluster is potentially a frequency domain. With many frequency domain the chip can easily mitigate the variations.

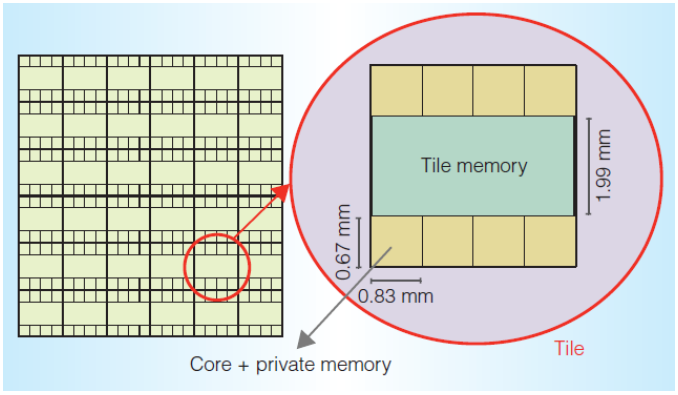

Fig. 7 288 core NTV chip(Karpuzcu 2013)

Energysmart technique uses the core-to-job algorithm in which the frequency of the chosen cluster is selected at the global V_{DD} instead of selecting the V_{DD} and frequency of all clusters

F.Resynthesis technique for delay variation

Conventional methods add timing margin such as guardbands to handle the delay variation problem. But such methods have to trade off performance. Instead of sacrificing the performance Resynthesis technique uses a method of adding a redundant logics to protect the performance. Here we use a concept of slacks and the nodes in the critical paths have zero slack thereby increaing its change of affecting by delay variation. This methods increases the tolerance of delay variation by incresing the number of slacks but there will be additional area penalty of about 2l percent for 10 percent of delay variance tolerance.

Technology scaling has resulted in the low voltage supply, higher frequencies, circuit performance is very sensitive to small variations, noises and delay. These affects the timing behavior of the transistor and hence it becomes necessary to consider the worst case or time margin in order to handle the timing variation. It is also determined that an ASIC may run 40 percent faster than predicted by standard timing analysis. In time critical process adding timing margin will not be possible. So resynthesis method is proposed which trade off area for the delay tolerance.

In some circuits, there are certain gates or wires in the critical path are more vulnerable to gate delay because any delay variations in the gate can result in variations in whole circuit delay. The Vulnerability can be measured in gate slack which means affordable margin without affecting the circuit delay. For example if a circuit has d_t delay tolerance and if each gate delay can be increase d_t delay without affecting the circuit delay i.e. the slack of the circuit is d_t. If a circuit is given and its delay tolerance is d_t then the circuit can be resynthesized in such a way that every gate in the new circuit can tolerate delay variation of about d_t. For this a new architecture was proposed with the voting machine and a original circuit and two auxiliary circuits.

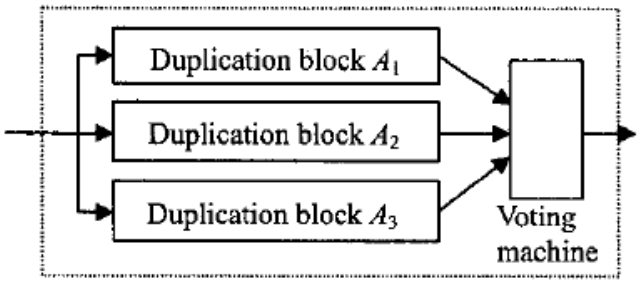

Fig. 8 TMR structure (Shih-Chieh Chang 2004)

The Triple Modular Redundancy (TMR) is used to tolerate delay variation. In a TMR structure the given circuit is made into three copies and its output is connected to the voting machine. The voting machine will produce the correct output if any of the two blocks produces the same output. In a TMR each gate or wire is redundant because removing any of the wires or gates will not affect the function.

It is necessary to increase the number of slacks in the gate or wire inorder to increase the gate delay tolerance. For this we have to add redundant gates or wires to the original circuit so that the slack will be increased. In TMR each component like wire, a gate, and a path has three replications. These three replications of a component are isomorphic components. In Fig. 9 p1, p2, p3 are isomorphic paths and wires w1, w2, w3 are isomorphic wires.

All three isomorphic paths have the same delay. Suppose the delay of isomorphic paths are different due to delay variation. Since the Voting Machine determines the correct output if two of its inputs provides correct result, the final delay is dominated by the second arriving signal. This method of choosing second arriving signal for a voting machine makes all three isomorphic paths not vulnerable to delay variation individually.

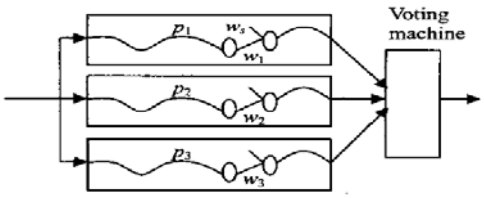

Fig. 9 Isomorphic paths in TMR(Shih-Chieh Chang 2004)

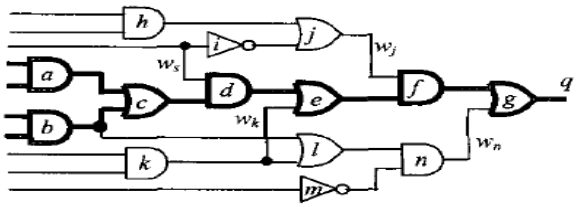

Fig. 10 Original Circuit (Shih-Chieh Chang 2004)

Wires are removed from the redundant logic so that the logic function will not be affected also result in some path change from strictly false paths to true paths. While removing the wires from the logic we have to consider three conditions. They are,

- A wire in d_t critical regions cannot be removed to maintain d_t delay tolerance requirement.
- A side input wire w to a d_t dominator can be removed without violating the requirement of d_t delay tolerance
- One wire among three isomorphic wires (w_{11},w_{12},w_{13}) and one wire three isomorphic wires(w_{21},w_{22},w_{23}) can be removed simultaneously without violating the requirement of d_t delay tolarence.

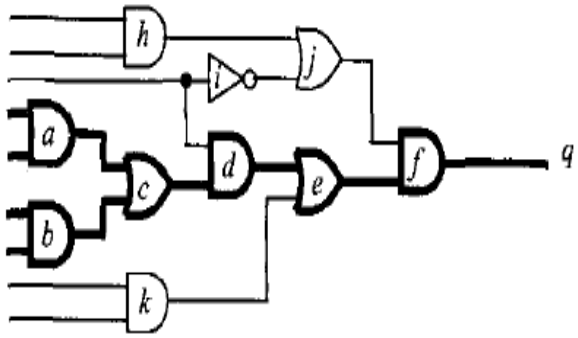

Fig. 11 Removal of side input (Shih-Chieh Chang 2004)

The Following are the necessary and sufficient conditions to remove a wire in a TMR while maintaining the slack of each node to be atleast d_t.

- A node is a critical node if the node's slack is less than d_t.
- A Circuit region is called a d_t critical region which consists of only d_t critical nodes and wires between d_t critical nodes. For example, In Fig.10 Original circuit Consider each gate delay is 1. If the delay tolerance value d_t is 2 Node g is d_t critical node because the slack of g is 0 which is less than d_t = 2.Nodes (a, b, c, d, e, f, g) which are drawn by bold line are all critical nodes. d_t critical region consists of highlighted gates and wires.

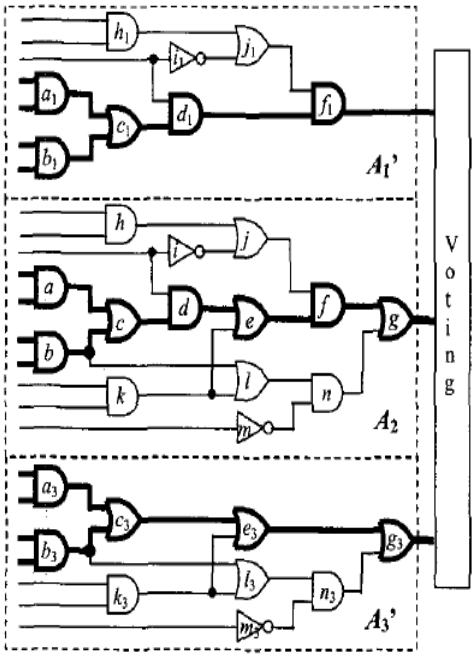

Fig. 12 Removing side input wires(Shih-Chieh Chang 2004)

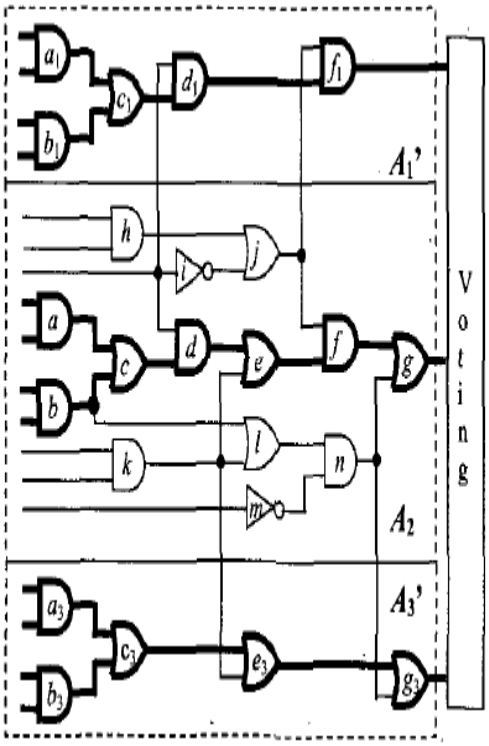

Fig. 13 Final Circuit(Shih-Chieh Chang 2004)

Table.1 Comparision between original and delay tolerance circuit(Shih-Chieh Chang 2004)

Circuit	Original circuit			Delay tolerance structure (d_t = 10%)					Delay tolerance structure (d_t = 15%)					Statistical analysis	
	Area	Circuit delay	Avg. slack	Area	Area overhead (%)	Circuit delay	#nodes with infinite slack	Avg. slack	Area	Area overhead (%)	Circuit delay	#nodes with infinite slack	Avg. slack	Original circuit	Delay tolerance d_t=10%
Alu2	487664	18.71	2.12	531744	9.0	18.66	0	3.62	617584	26.6	19.06	10	3.80	6828	7083
Alu4	1254656	25.71	5.03	1440720	14.8	26.23	16	5.94	1796144	43.2	25.65	38	7.05	5001	8799
Apex6	1165568	11.66	3.55	1357200	16.4	11.29	6	4.43	1442112	23.7	11.47	40	4.38	4062	7929
Apex7	360832	11.41	3.72	411568	12.6	11.24	2	4.01	505296	38.2	11.49	12	3.63	8820	9924
B9	165184	6.54	1.81	202304	22.5	6.64	4	1.94	222256	34.6	6.46	5	2.30	6411	6930
Frg1	186528	11.81	2.26	221792	18.9	10.89	0	3.74	250560	34.3	11.55	4	3.29	8361	9081
Frg2	1425408	11.40	2.56	2065264	44.9	11.82	10	2.28	2269888	59.2	12.02	48	2.53	4179	8385
Pair	2469408	14.46	3.63	2894432	17.2	14.07	12	4.52	3050336	23.5	13.85	24	5.06	6192	8775
Rot	1031936	14.78	6.01	1203152	16.6	14.47	26	6.67	1286672	24.7	15.05	26	6.09	8775	9861
S344	273760	12.34	3.42	307168	12.2	11.86	0	4.13	378624	38.3	11.39	4	4.58	8247	8835
S349	268192	12.80	3.65	327120	22.0	12.27	0	4.33	403216	50.3	12.07	18	4.59	7812	8565
S526	306240	8.98	1.61	364704	19.1	8.83	4	2.06	521536	70.3	8.94	36	2.05	6759	8460
S641	261696	12.94	3.99	301600	15.2	12.40	0	5.31	307168	17.4	11.75	0	5.99	9078	9765
S713	261696	12.92	3.99	294640	12.6	12.38	0	5.36	308560	17.9	11.99	0	5.91	8319	8412
S1196	967904	13.28	2.23	1154432	19.3	13.47	12	2.68	1436544	48.4	14.23	36	2.08	7515	8346
S1238	1026368	14.11	2.11	1116384	8.8	13.93	12	3.10	1402208	36.6	14.69	12	2.52	5607	9714
S1488	900624	13.17	1.79	1253728	39.2	12.74	34	2.91	1530272	69.9	13.27	107	2.52	6933	9990
S1494	857472	12.69	1.77	1314048	53.2	12.78	62	2.08	1494544	74.3	12.86	202	2.23	3351	7842
Avg.					20.8					40.6				6792	8705

Gates not in the d_t critical region will have atleast d_t slacks. If the gate is not in the critical region then we can reduce the area by sharing the equivalent signals that implements the same function. For example in Fig. 12 the Gate j_1 in the block A_1' and the Gate j in the block A_2 has the same functionality and hence it is possible to share the output from the gate j with that of the one in the block A_1'. Thus using the above conditions and removing the wires we obtain the final circuit similar to the one in Figure 13. From the table, the area overhead for 10 percent delay tolerance is 20.8 percent and for 15 percent delay tolerance is 40.6 percent which is far better than 200 percent area overhead in TMR.

G.Resilient Microarchitecture

Scaling has large impact on the device degradation as well as on the Microarchitecture. A new binning strategy is implemented which takes into account individual degradation level of the processing nodes. The test functionality can be distributed as a part of harware to determine, eliminate the faults that can occur and we can implement a multicore design architecture to handle the faults. This type of microarchitecture that can handle the faults is known as resilient microarchitecture. Resilient Microarchitecure can adapt to a temporal and spatial variability of each individual block. Resilient Microarchitecture is built in such a way that it either mitigates the variation by finding its source or uses time speculation i.e. assuming time larger than the normal delay. Time speculation requires a mechanism to detect and correct fault. These mechanisms can also be used to detect and correct faults cause by neutron strikes i.e. soft error.

Besides soft errors and Variations, time dependent degradations makes impossible to design a reliable system. Burn in test also proves that its useless to catch the chip infant mortality. The Hardware test has to be moved into the design cycle in such a way that resilient realization is assured during design and is maintained during product's life cycle. Therefore Resilient Many Core Architecture is proposed which has to provide features like

- Dynamic On Chip testing
- Performance Profiling
- Spare hardware
- Binning Strategy
- Performance and power management
- Coarse-grain redundancy checking
- Dynamic error detection and reconfiguration

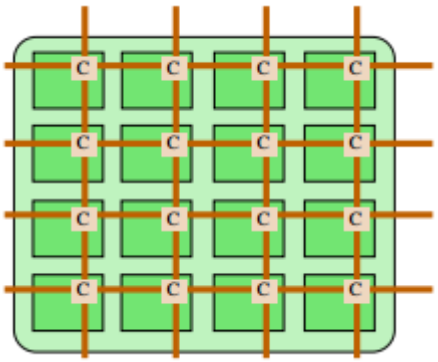

Fig 14 Proposed Many Core Architecture (Glosekotter 2008)

Multiple cores in a system will provide similar performance and redundancy benefits with functional redundancy checking employed at a coarse-grained level. For example, one core could check the results produced by several cores and software and applications will have to support this concept whenever possible.

V. CONCLUSION

Variability becomes a serious issue as Technology scales down and this is the right time to adapt to newer processing, manufacturing and testing technologies to confront variability. Each and every component in a system has to be analyzed and implement with Variability tolerant design. Mitigation of Soft error using Razor requires Time penalty but it results in power conservation. Resynthesis technique for delay variation results in Area penalty of about 21 percent for 10 percent of delay variance tolerance. In Energy smart technique instead of using Multi V_{dd} domain it uses single V_{dd} thereby reducing the Parametric Variation at Near Threshold Voltage (NTV). All the above techniques for Mitigation Variation has a bottleneck of trading off either performance or Area for Mitigating Variation. It becomes necessary to propose a method of Mitigating Variation with Minimum trade off. Resilient Microarchitecture is such a kind of method that detects errors, isolates faults, refreshes processing nodes, and thus adapts the hardware. In Future Resilient Microarchitecture has a strategy to detect and correct variability and it should be concentrated more.

REFERENCES

[1] Borkar, S., "Designing reliable systems from unreliable components: the challenges of transistor variability and degradation," IEEE Micro , vol.25, no.6, pp.10-16, Nov.-Dec. 2005 DOI= 10.1109/MM.2005.110

[2] Shih-Chieh Chang; Cheng-Tao Hsieh; Kai-Chiang Wu, "Re-synthesis for delay variation tolerance," Design Automation Conference, 2004. Proceedings. 41st , vol., no., pp.814,819, 7-11 July 2004

[3] D. Ernst, N. Kim, S. Das, S. Pant, R. Rao, T. Pham, C. Ziesler, D. Blaauw, T. Austin, K. Flautner and T. Mudge, "Razor: a low-power pipeline based on circuit-level timing speculation," MICRO-36, pp. 7 - 18, 2003

[4] Karpuzcu , U.R.; Nam Sung kim; Torrellas, J., "Coping with Parametric Variations at Near-Threshold Voltages," Micro, IEEE, vol.33, no.4, pp.6,14, July-Aug. 2013 DOI=10.1109/MM.2013.71

[5] Iyer, R.K.; Nakka, N.M.; Kalbarczyk, Z.T.; Mitra, S, "Recent advances and new avenues in hardware-level reliability support," Micro, IEEE, Vol.25, no.6, pp.18,29, Nov.-Dec. 2005 DOI= 10.1109/MM.2005.119

[6] J.Abella, A. Gonzalez Heterogenous Way-Size Cache. In ICS 2006

[7] W.Abadeer, W. Ellis. Behavior of NBTI under AC Dynamic Circuit Conditions. In IRPS 2003

[8] R. Ashraf and R.F. DeMara, "Scalability of Modular Redundancy for Near-Threshold Computing," Workshop on Highly-Reliable Power-Efficient Embedded Designs(HARSH 2014), Orlando, Florida, USA, February 16th, 2014.

[9] R. A. Ashraf, A. Alzahrani, and R. F. DeMara, "Exploring Spatial Redundancy to Mitigate Aging-Induced Timing Degradation," ACM/EDAC/IEEE 51st Design Automation Conference (DAC) (poster presentation only), San Francisco, California, USA, June 1 – 5, 2014.

[10] N. Khoshavi, R. A. Ashraf, and R. F. DeMara, "Applicability of Power-Gating Strategies for Aging Mitigation of CMOS Logic Paths," IEEE 57th International Midwest Symposium on Circuits and Systems(MWSCAS-2014)submitted on March 28, 2014.

[11] Jaume Abella; Xavier Vera; Antonio Gonzalez, "Penelope: The NBTI-Aware Processor," MICRO-40,Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture Pages 85-96 2007 DOI= 10.1109/MICRO.2007.32

[12] Glosekotter, P.; Greveler, U.; Wirth, G.I., "Device Degradation and Resilient computing," IEEE International Symposium on Circuits and Systems, 2008. ISCAS 2008, Seattle, WA DOI= 10.1109/ISCAS.2008.4541546

[13] S. Borkar, "Circuit Techniques for Subthreshold Leakage Avoidance, Control, and Tolerance," Proc. Int'l Electron Devices Meeting (IEDM 2004), IEEE Press, pp. 421-424.