

WEAROUT RESILIENCE IN NOCS THROUGH AN AGING AWARE ADAPTIVE ROUTING ALGORITHM

¹R.Suganya, ²N.L.Thangadurai,

¹PG Scholar, Dept of VLSI Design, Roever Engineering College, Perambalur,

²Asst prof, Roever Engineering College, Perambalur.

Abstract:

With the proliferation of on-chip cores allowed through rapid technology scaling, Network on-Chips (NoCs) are becoming a critical determinant of overall system power-performance characteristics. Consequently, the growing reliability challenges, which are continuously reshaping the system design considerations, must now be thoroughly analyzed in the context of NoC designs. Two primary mechanisms studied in this work that are responsible for circuit wear-outs in an NoC design are: Negative Bias Temperature Instability (NBTI) and Electro migration. it may be necessary to use an alternate route. However, employing such an alternate route can increase the network latency, thereby degrading the system level power-performance.

Keywords – NOC, NBTI, Latency, Wear outs.

1. INTRODUCTION

An NoC architecture comprises two major components: NoC router and link. A pipelined NoC router consists of both combinational logic structures (e.g., virtual channel allocation logic) and storage-cell structures (e.g., virtual channels). Due to the presence of these structures, NBTI is the major aging mechanism associated with NoC routers [2]. NoC links, on the other hand, are implemented using repeated copper interconnects [3]. Therefore, NBTI (repeaters) and electromigration (copper interconnects) are the two primary aging problems associated with NoC links. Unfortunately, previous works on NoCs have completely ignored the role of links in their reliability analysis, focusing solely on the routers [2,4]. Experiments in this work demonstrate that such limitations can grossly under-estimate the NoC lifetime by nearly a factor of 2.

In the context of reliability in NoCs, another critical design challenge stems from the asymmetric usage of NoC components. Mishra et al. have shown this non-uniform pattern of router buffer and link utilization [5]. They observed that the routers in the center of the mesh are highly (75%) utilized, while the peripheral routers have low (35%) utilization. Similarly, the experiments with multithreaded workloads on a 4×4 mesh indicate a wide disparity in buffer utilization of different routers. Due to such asymmetric utilization, each router and link will also suffer from different amounts of aging degradation. Therefore, there is a need for an aging-aware routing algorithm for NoCs that considers the aspect of asymmetric aging while routing packets, so as to improve the system reliability.

2. LITERATURE SURVEY

In this paper, we investigate the characteristics of SRAM cells with high- k metal-gate Si/Si_{1-x}Ge_x dual channel structures. The characteristics are compared with those of the unstrained structures. The results show that the strain degrades read SNM slightly while increases read current considerably. In addition, it increases write ability while decreases standby power. Moreover, NBTI

and PBTI effect for two cases of symmetrical and asymmetrical stresses is investigated. In the symmetrical case, read and write stability don't reduce while read current decreases. For the case of the asymmetrical stress, both read and write stabilities degrade. In addition, read current decreases more than that of the symmetrical case. The results demonstrate while NBTI and PBTI cause less read current reduction in the strained cells, the degradations of other metrics are comparable to those of the unstrained cell, Negative-bias temperature instability (NBTI) and positive-bias temperature instability (PBTI) weaken PFET and NFET over the lifetime of usage, leading to performance and reliability degradation of nanoscale CMOS SRAM. In addition, most of the state-of-the-art SRAM designs employ replica timing control circuit to mitigate the effects of leakage and process variation, optimize the performance, and reduce power consumption. In this paper, we provide comprehensive analyses on the impacts of NBTI and PBTI on a two-port 8T SRAM design, including the stability and Write margin of the cell, Read/Write access paths, and replica timing control circuits. We also discuss degradation tolerant design techniques to mitigate the performance and reliability degradation induced by NBTI/PBTI.

3. RELATED WORK

NoCs are the most scalable and power-efficient solutions towards developing interconnects that can connect many number of processors on chip. NoCs comprise of routers and links as the basic building blocks. simple NoC architecture connecting two on-chip processing elements (PE1, PE2). The processing elements in the form of processors are connected to the routers (R1, R2) through Network Interface (NI). These routers are then connected to each other via links (L), which are repeated copper interconnects.

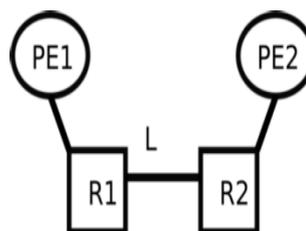


Fig.1.A simple NOC

Routing algorithms are used to decide the best path that a flit will take to reach its destination router. As the performance and power consumption of a network is dependent on the paths that flits take (shortest or longest), routing algorithms play an important role in controlling the power-performance characteristics. Routing algorithms can be divided into the following categories.

Oblivious Routing: In oblivious routing, the path is completely determined by the source and destination address. This type of routing enables simple and faster router designs. However, oblivious routing algorithms are not able to address various runtime issues such as congestion.

Adaptive Routing: In adaptive routing, the path is decided dynamically based on the runtime conditions such as network congestion. This way, they are able to achieve better performance as compared to the oblivious routing algorithms. However, the downside is a more complex and costly

adaptive router. Moreover with manufacturing defects and hardware malfunctions, contemporary research is also directed towards fault-tolerant routing algorithms for NoCs. Shi et al. recently proposed a scalable and distributed fault tolerant routing algorithm for NoCs that divides the system into regions and each region guarantees fault-tolerance of its own area. Fick et al. developed a highly resilient routing algorithm that reconfigures around the fault components to improve robustness of the system. Chaix et al. proposed a fault-tolerant adaptive routing algorithm that is able to route packets in the presence of multiple nodes and link failures without using routing tables. Akbari et al. address the issue of poor vertical links yield in case of 3D mesh-based many core ICs by introducing AFRA, a low cost high performance deadlock-free routing algorithm that tolerates faults on vertical links.

4. AGING AWARE ADAPTIVE ROUTING ALGORITHM

For each flow at runtime, the routing algorithm selects the best shortest path from the routing table that (a) suffers from least aging degradation, i.e., the path that suffers from least delay variation due to aging (1-a); and (b) is least congested (1-b). A higher priority is given to a path that is least degraded as compared to a path with the least congestion. For example, in case of an arbitrary flow F , if the available number of deadlock-free shortest paths is four (path0....path3) then the algorithm maintains congestion and aging scores (sc_{cong} and sc_{age}) for each of these paths. These scores are calculated based on both local and global aging and congestion information obtained from different routers and links present in the paths. Now if the sc_{age} is least for path0 but sc_{cong} is least for path3 then the algorithm selects path0 for routing. In a different situation, if sc_{age} is same for all paths but sc_{cong} is least for path3 then path3 is only selected for routing. congestion score (sc_{cong}) of the paths. Similarly, based on the number of stressed links in SL_{set} for each epoch, an aging score (sc_{age}) is calculated for each path. The route computation unit then selects the output link corresponding to the shortest path which has the least sc_{age} and sc_{cong} . Note that the additional logic-based circuitry is introduced in parallel paths rather than in sequence, for example sc_{age} and sc_{cong} are calculated in parallel.

5. RESULT ANALYSIS

In order to show the robustness degradation in a 4×4 NoC mesh that does not consider aging aware-routing (NO-AGE), the reliability of this scheme is compared with the MIPROUT scheme that models the aging-aware routing. the router and link utilization for both the schemes for canonical benchmark run. As evident from the table, under NO-AGE, several routers and links exceed their TAC limit (TAC-LIM), which will render them faulty after some period of stress. MIP-ROUT adjusts these utilizations, and therefore improves NoC robustness. In the case of TAC-AGE, since stressed routers and links meet their TAC limits, its robustness is better than that of NO-AGE but lower than that of MIP-ROUT. the mesh reliabilities of the schemes for an aging period of 7 years using the reliability's exponential dependence on failure rate. As expected, NO-AGE shows higher failure rate compared to MIP-ROUT, as its design does not adapt to the wear-out degradation of NoC components. As stressed routers and links are well below their TAC-LIM in MIP-ROUT, its reliability is better than TAC-AGE also.

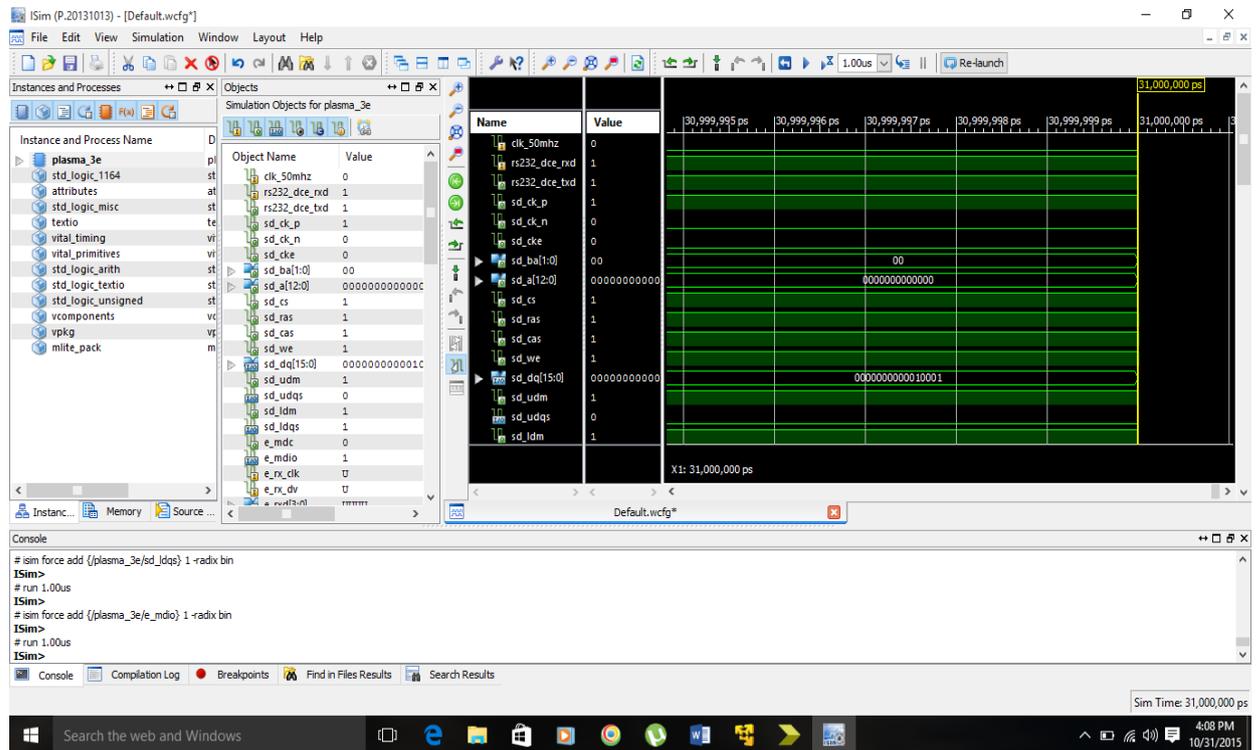


Fig.2.Output result

Extensive experimental analysis incorporating power-performance impact of aging demonstrate 13% and 12.17% improvements in network latencies and EDPPF for our algorithm, respectively. At the system level, our algorithm show 10.4% performance improvement for real workloads.

CONCLUSION

A critical need is observed to optimize NoC power-performance metrics while considering wear-out degradation resulting from asymmetric utilization of NoC components. To efficiently tackle this multi-objective design challenge, two routing algorithms are proposed: (a) Aging and congestion aware adaptive routing algorithm, and (b) Aging-aware oblivious routing algorithm. Extensive experimental analysis using real workloads demonstrates improvements in network latencies and EDPPF for both the algorithms. At the system level also, the proposed algorithms show reduction in performance overhead, mitigating the effects of aging mechanisms such as NBTI and electro migration. Extensive experimental analysis incorporating power-performance impact of aging demonstrate 13% and 12.17% improvements in network latencies and EDPPF for our algorithm, respectively. At the system level, our algorithm show 10.4% performance improvement for real workloads.

REFERENCES

- [1] J. D. Owens, W. J. Dally, R. Ho, D. N. Jayasimha, S. W. Keckler, and L.-S. Peh, "Research challenges for on-chip interconnection networks," in *IEEE Micro*, vol. 27, no. 5, pp. 96–108, 2007.

- [2] X. Fu, T. Li, and J. A. B. Fortes, "Architecting reliable multi-core network-on-chip for small scale processing technology," in *IEEE Dependable Systems and Networks (DSN)*, pp. 111–120, 2010.
- [3] C. Hernandez, F. Silla, and J. Duato, "A methodology for the characterization of process variation in noc links," in *Design Automation and Test in Europe (DATE)*, pp. 685–690, 2010.
- [4] A. K. Kodi, A. Sarathy, A. Louri, and J. M. Wang, "Adaptive inter-router links for lowpower, area-efficient and reliable network-on-chip (noc) architectures," in *Asia-Pacific Design Automation Conference (ASP-DAC)*, pp. 1–6, 2009.
- [5] A. K. Mishra, N. Vijaykrishnan, and C. R. Das, "A case for heterogeneous on-chip interconnects for cmps," in *International Symposium on Computer Architecture (ISCA)*, pp. 389–400, 2011.
- [6] K. Bhardwaj, K. Chakraborty, and S. Roy, "Towards graceful aging degradation in nocs through an adaptive routing algorithm," in *Design Automation Conference (DAC)*, pp. 382–391, 2012.