Proactive Aging Management in Heterogeneous NoCs through a Criticality-driven Routing Approach

Dean Michael Ancajas, Koushik Chakraborty, Sanghamitra Roy
BRIDGE Lab, Electrical and Computer Engineering, Utah State University
dbancajas@gmail.com, {koushik.chakraborty, sanghamitra.roy}@usu.edu

Abstract—The emergence of power efficient heterogeneous NoCs presents an intriguing challenge in NoC reliability, particularly due to aging degradation. To effectively tackle this challenge, this work presents a dynamic routing algorithm that exploits the architecture level criticality of network packets while routing. Our proposed framework uses a Wearout Monitoring System (to track NBTI effect) and architecture-level criticality information to create a routing policy that restricts aging degradation with minimal impact on system level performance. Compared to the state-of-the-art BRAR (Buffered-Router Aware Routing), our best scheme achieves 38%, 53% and 29% improvements on network latency, system performance and Energy Delay Product per Flit (EDPPF) overheads, respectively.

I. INTRODUCTION

Emerging many-core systems have raised the importance of communication architectures such as Network-on-Chips (NoCs) [11], [20]. While a vast body of past work address the power-performance optimization in NoCs, reliability concerns are rapidly becoming a fundamental design challenge in them [17]. In particular, the asymmetric utilization seen in NoC components, substantially exacerbate aging degradation. For example, Mishra et al. observe that routers in the central region of the mesh can have more than 2X utilization compared to those near the peripheral region [16]. Such higher utilization in some NoC components can manifest as rapid aging induced power-performance degradation in those components, causing system-wide deterioration.

Unfortunately, balancing the reliability and energy efficiency creates a fundamental tension in NoC design. We demonstrate that *heterogeneous NoCs (hNoCs)*—an increasingly popular power-efficient NoC design alternative [22]— greatly aggravate the reliability challenge in the NoC architecture. A typical *hNoC* consists of strategically placed buffered and bufferless routers, so as to facilitate energy efficient resource proportioning. But this structural and functional diversity aggravates the reliability aspect in the design through: (a) disparate aging effects on buffered and bufferless routers; and (b) the rise in utilization asymmetry among NoC components. Due to the storage structures present in the form of buffers, buffered routers suffer substantially more *Negative Bias Temperature Instability (NBTI)* degradation than the bufferless routers. We demonstrate that aging degradation can affect buffered routers by more than $2\times$ compared to the bufferless routers.

To effectively tackle aging degradation in power-efficient NoCs, this work presents a novel adaptive routing algorithm that dynamically modifies traffic flow and exploits opportunities to avoid the use of buffered routers, while minimiz-

ing system-level power-performance impact. Our proposed approach incorporates a *Wearout Monitoring System (WMS)* (composed of NBTI delay sensors) in NoC components, and combines architecture-level packet criticality during routing to relieve the heavy usage in aged components. Overall, our proposed routing algorithm can substantially reduce the effects of aging degradation in *hNoCs*, thereby assuring a graceful degradation in the communication architecture of emerging systems.

In this paper, we make the following contributions:

- **Reliability Analysis of an hNoC:** We present an extensive study analyzing the impact of routing policies in an *hNoC* (Section II). Our analysis shows an alarming increase in utilization asymmetry in certain *hNoC* components (by more than $1.35\times$), which can cause rapid aging degradation in those components, severely affecting system level performance characteristics. In this context, we also uncover a new opportunity in reliability driven routing in an *hNoC*, by demonstrating that a substantial portion of the data packets routed in the network are non-critical (i.e., system performance is insensitive to their latency). Thus, utilization in centrally placed buffered routers can be reduced by minimizing non-critical packets routed through them.
- **Proactive Routing to Mitigate NoC Aging:** We develop a novel dynamic aging-aware routing algorithm based on architecture-level criticality information and accumulated wearout information using our proposed Wearout Monitoring System (Section III). To the best of our knowledge, this is the first work that uses the criticality information of packets to improve the reliability of NoCs.
- **Holistic Evaluation:** We perform a holistic evaluation of our proposed algorithm by integrating: (a) SPICE level simulation to estimate the combined effect of device aging based on circuit-level utilization and process variation (Section IV), (b) statistical timing analysis of synthesized hardware to accurately estimate the delay distribution under aging and expected usage pattern (Section IV), and (c) cycle accurate architectural simulation using GARNET and GEMS toolsets with multithreaded applications on a 16-core system with an *hNoC* (Section V). Compared to the aging overhead seen in a state-of-the-art *hNoC* routing scheme [22], our best scheme shows 38%, 53% and 29% improvement in network latency, system performance and EDPPF degradation (on average), respectively (Section VI).

II. MOTIVATION

In this section, we motivate our proposed framework for mitigating aging effects in an *hNoC* by demonstrating two
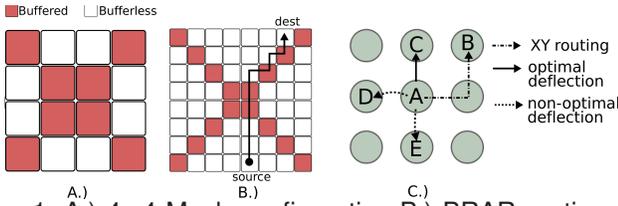
Fig. 1: A.) 4×4 Mesh configuration B.) BRAR routing on an 8×8 mesh C.) Deflection routing on bufferless routers.

important circuit-architectural characteristics. First, we show that existing routing policies in a typical *hNoC* can lead to an alarming rise in the utilization asymmetry, exacerbating the NoC reliability challenge (Section II-A). Second, we analyze the criticality of network packets in an NoC running multi-threaded applications to identify opportunities of improving reliability by exploiting packet non-criticality (Section II-B).

Recently proposed heterogeneous NoC designs exploit the non-uniform traffic across routers to proportion routing resources [16]. Zhao et al. [22] further improve performance by employing a routing algorithm that leverages the non-uniform structure of the network to move data along buffered routers to keep packets on their optimal paths. Routing through the buffered routers prevents these packets from being deflected arbitrarily by bufferless routers, which can increase their network latency. However, this routing approach overburdens the buffered routers as a majority of the flits[1] use buffered routers as a part of their pathways.

A. Utilization Asymmetry in Heterogeneous NoCs

The setup we use in this study is shown in Figure 1A. We employ *Buffered Router Aware Routing (BRAR)* from Zhao et al. [22], and use the identical placement of buffered routers as shown in their work. BRAR routes flits along buffered routers to decrease the chance of packets deflected farther from their destination (Figure 1B). However, if there is output port contention, the flit is stored in the buffer and delayed for another cycle. In contrast, the routing on bufferless routers (Figure 1C) needs to deflect flits towards a different direction because once a packet loses arbitration for the switch, it will be sent out to any free port. Two out of three times, the packet will be sent out towards a non-optimal direction.

Figure 2 shows the increase in utilization asymmetry of the buffered routers when using the BRAR algorithm (distribution of buffered and bufferless routers are shown in Figure 1A). We show all 16 routers in a 4×4 NoC mesh, where the numbers in each router indicate the percentage increase in utilization compared to a homogeneous NoC employing XY routing. Results shown are average across several multithreaded PAR-SEC benchmarks we use in this study. We notice that the centrally placed routers in the *hNoC* show 23–82% increase in utilization, while all of the peripheral routers show a range of reduced utilization. Compared to the homogeneous NoC, this *hNoC* employing BRAR routing shows $1.35\times$ increase in utilization asymmetry[2] which ultimately leads to a more than $2\times$ increase in timing degradation on the buffered routers. Section IV-B discusses in detail the NBTI effect on *hNoCs*.

[1]*flits* are the units of information flow in NoC networks.
[2]Utilization asymmetry is estimated by the range of utilization seen across the NoC components.

Fig. 2: Percentage traffic increase of each router using BRAR (average across PARSEC benchmarks). This utilization difference leads to more than 2× divergence in NBTI induced performance degradation (see Section IV-B).

| Data Messages | | Classification |
source	dest	
L1 Cache	L2 Cache	non-critical
L2 Cache	L1 Cache	critical
Memory	L2 Cache	critical
L2 Cache	Memory	non-critical
Control Messages		**Classification**
source	dest	
L1 Cache	L2 Cache	critical
L2 Cache	L1 Cache	critical
Memory	L2 Cache	critical
L2 Cache	Memory	critical

TABLE I: Packet Criticality Classification.

B. Criticality of different flits in NoCs

The latencies of various packets transmitted through an NoC can have varied effects on performance. Previous works have exploited this criticality to improve system performance [8], [13]. In this work, we exploit this latency criticality of packets in our dynamic aging-aware routing policies. We quantitatively demonstrate this opportunity after briefly outlining our criticality classification.

Criticality Classification: In general, precise estimation of the packet criticality at the NoC router is hard as it merely has information about source-destination and the packet type. A thorough criticality estimation may require information about the relative performance of running program threads [8], [9], detailed cache coherence transitions, and so forth. To mitigate this complexity, we employ a low-complexity approach, requiring no change in existing interfaces. This involves identifying criticality based on packet type and source-destination. Table I shows the summary of our classification. Using this policy, we tag any data packet transmitted from L1 to L2 (destination) as non-critical in a shared two level cache hierarchy. A vast majority of these packets are writebacks because of cache eviction, and thus the system performance is insensitive to their network latency. Some of these packets are also a result of data sharing among on-chip cores, but we expect these to be a much smaller component due to the predominance of private data even in multithreaded programs [5].

Opportunity: Figure 3 shows the percentage of non-critical packets of PARSEC benchmarks averaged across all the buffered routers. An average of 49% of packets traversing through the buffered routers are non-critical and can actually take a different routing path with minimal performance
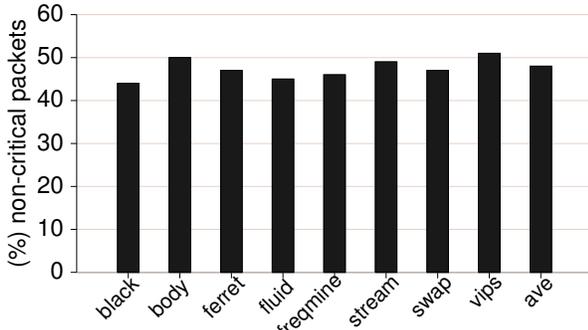
Fig. 3: Percentage of non-critical data packets routed through the buffered routers.

Fig. 4: WMS Circuit. Each path delay is sampled through a buffer sequence and compared with the reference delay to calculate the WF.

degradation. Moreover, all benchmarks show substantial opportunity, ranging from 44% to 51% in these benchmarks. By redirecting non-critical traffic to the bufferless routers, we can minimize utilization of the buffered routers, thereby mitigating the aging effects in the buffered routers.

C. Significance for Reliability Driven Routing in hNoCs

The substantial presence of non-critical packets offers an intriguing opportunity for reliability aware routing in *hNoCs*, while preserving inherent advantages in power efficiency. Instead of always emphasizing the use of buffered routers, the routing algorithm can make a selective choice of routes based on packet criticality. Non-critical packets can be deflected away from the buffered routers, thereby reducing their utilization. We now discuss our proposed approach in exploring such criticality aware routing algorithms.

III. AGING-DRIVEN ROUTING VIA PACKET-CRITICALITY

In this section, we discuss how we use the criticality information of packets to implement a deflection-based aging-driven routing algorithm. We first explain our Wearout Monitoring System (WMS) which keeps track of the health of routers in the network. Then, we discuss how we combine WMS and criticality information to implement aging-aware routing schemes.

A. Wearout Monitoring System for NoC Routers

To be able to guide the aging-aware routing algorithm, the WMS profiles the extent of degradation in each router. The WMS circuit shown in Figure 4 augments all pipeline stages of a router. As the performance degradation of a router is dictated by the worst case delay degradation in any pipeline stage, our proposed monitoring system measures the maximum delay degradation across all paths in different pipeline stages.

Within a stage, the WMS uses a multiplexer to estimate the delay of all *n* paths in a combinational logic. The control unit in Figure 4 alters the multiplexer select signal in each cycle to choose which path to measure. Then, a series of *m* cascaded delay buffers (db_1, db_2, ..., db_m) sample the signal at equal time intervals. The state transition captured at the output of each delay buffer provides an estimate of the delay of the path. Finally, the comparator selects the maximum delay degradation among the *n* paths over a span of *n* cycles. The WMS measures the wearout factor (WF) as follows:
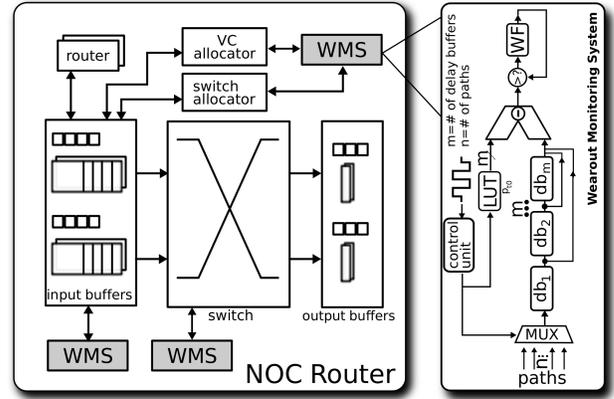
$$WF_{router} = max(wf_1, wf_2, ..., wf_N) \qquad (1)$$

$$wf_i = max(wf_{p1}, wf_{p2}, ..., wf_{pn}) \qquad (2)$$

where, wf_1, wf_2, ..., wf_N are the wearout factors for N stages of the router micro-architecture, and wf_{p1}, wf_{p2}, ..., wf_{pn} are the wearout factors of the n paths in a single stage i.

Implementation Overhead: Since NBTI aging is a slow process, a slow sampling period for WMS (eg. 1 out of 10^{10} times at 1GHz) is satisfactory [12]. Thus, the WMS component is rarely activated and is clock-gated majority of the time. Also, the WMS measurement is accessed in parallel with the pipeline stage avoiding any impact on the performance. Using a well known NoC router model [1] as a baseline, our implementation shows 1.2% and 0.047% area and power overheads using a 45nm TSMC standard cell technology targeted at 1GHz.

The propagation of WF to adjacent routers is done through the flit link network, by triggering a dedicated multiplexer to latch the WF vector. Given the sampling rate, this WF propagation has minimal effect on system performance.

B. Criticality-Driven Path Selection

Our proposed criticality driven routing incorporates two major design considerations: (a) criticality of the incoming packet; and (b) WF that dictates the current aging. We establish a maximum threshold for deflecting non-critical packets, defined as DFL_{Max}. Subsequently, based on the aging degradation in a router, we pro-rate the deflection rate in that router.

Integrating Criticality in Routing: To drive the deflection logic in routing path selection, the source router adds a single-bit to store the criticality in the header flit of every packet. All intermediate routers peek into this criticality bit to select different routing paths based on criticality.

Integrating Wearout Monitoring: Different routers can undergo different aging degradation based on their utilization history. In a given router, the WF provides its current aging degradation. Table II shows the pro-rating scheme used in this work. For example, a router with a WF of 0.8 will deflect 25% of all non-critical packets, assuming DFL_{Max} is 0.5. At every sampling interval of the WMS, the WF will be sent to adjacent routers to communicate the degradation of a particular router and a corresponding link. Each router stores the WF of four

Wearout Factor Range	Scheme
0.0 - 0.50	$\frac{1}{8} \times DFL_{Max}$
0.50 - 0.75	$\frac{1}{4} \times DFL_{Max}$
0.75 - 1.00	$\frac{1}{2} \times DFL_{Max}$
1.00 - $+\infty$	$1 \times DFL_{Max}$

TABLE II: WF based Deflection Estimation.

adjacent routers (North, South, East, West) in dedicated WF registers.

Deflecting Non-Critical Packets: For every incoming flit in a router, the deflection logic uses the WF and packet criticality information to determine whether the packet will be sent in the direction of the pre-established path or deflected away from the buffered router. For a bufferless router, we accomplish this task by a multiplexer and a selection logic. For a buffered router, we add an additional entry in the routing table corresponding to the possible deflection paths for each output port. For instance, an output in the North direction can be deflected to East or West if its coming from the South input. We accomplish this logic by using a 4-bit XOR of the number of ports (N,S,E,W) and the ports used for input and the desired output. Since there can be multiple deflection paths, we use the one that has no pending flits in the output buffer. For ties, we always use the first port using a standard priority encoder.

IV. AGING MODEL FOR AN HNOC

In this section, we derive an aging model for a typical *hNoC*. The aging model we derive in this section is used to calculate the additional delay experienced by flits due to NBTI degradation in heavily utilized routers and links. Unlike homogeneous networks, we have to consider the impact of the presence/absence of buffers in some of the routers in an *hNoC*. We first discuss the impact of NBTI aging on routers and links, and then explain our methodology to estimate their collective aging effect on heterogeneous NoCs.

A. NBTI Impact on Routers and Links

1) Router Effect: The NBTI effect causes components in an NoC to experience stress and recovery phases. The long-term effect on such components is a degradation in its threshold voltage, which in turn worsens the delay of basic gates, leading to system failure. The change in V_{th} in an NoC router for an aging period of t seconds is shown in Equation 3 as reported in [3]. All associated parameters can be referred in [3].

$$\Delta V_{th-router} \approx \left(\frac{n^2 K_v^2 \alpha C t_1 t}{\xi^2 t_{ox}^2 (1-\alpha)} \right)^n \tag{3}$$

To translate this change in V_{th} to timing degradation for NoC *routers*, we use first-order Taylor expansion to estimate timing increase as done in [6]. The delay is modeled as a Gaussian distribution perturbed around its normal value. We then take the $(\mu + 2\sigma)$ value of this delay from our Monte Carlo simulation results and assign it as the router delay in our architectural simulation.

2) Link Effect: For links, NBTI affects the threshold voltages of its repeater circuits, which leads to a higher drive resistance [10]. The additional resistance further degrades the parasitic delay of the NoC interconnects. Our model for ΔV_{th} in links is shown in equation 4 as reported in [19].

Router Type	μ	σ	Worst case delay[1]
Buffered	0.84	0.0304	0.931
Bufferless	0.75	0.0130	0.789

[1] worst case delay is calculated as $\mu + 3\sigma$

TABLE III: Delay Distribution for Buffered (high utilization) and Bufferless (low utilization) Routers in BRAR.

$$\Delta V_{th-link} = b\alpha^n t^n \tag{4}$$

Parameters α and t are the switching probability and aging time in seconds, respectively, while n and b are fitting parameters [14]. The switching probability of each link is taken from the profile of the benchmarks that we used. Using the $\Delta V_{th-link}$, the additional delay for each NoC link is then calculated using an RC model from [10]. Subsequently, the corresponding link delays (in conjunction with the router delays) are also used in the simulator for a system-level evaluation of the NBTI effect.

B. NBTI Effect on Heterogeneous Network-on-Chip

In an *hNoC*, some routers are buffered and some are bufferless. Since the NBTI effects are much more pronounced in sequential circuits, buffered routers are more susceptible to aging degradation as compared to bufferless routers. Moreover, buffered routers are positioned in the network such that a majority of the flits will traverse through them, further increasing the gap between the degradation rate of the two kinds of routers. To prove this phenomenon we conducted an experiment to measure the increase in delay in *hNOC* routers caused by NBTI aging degradation.

Our setup for this experiment uses Synopsys HSPICE, Predictive Technology Models [23], long-term NBTI degradation model [3] and static timing analysis of an *hNoC* derived from an open-source NoC Router [1]. We perform the investigation as follows:

1) We find the effect of NBTI aging and process variation in basic logic gates. We run Monte Carlo HSPICE simulations (10K sample) to get a statistical performance distribution of each gate.
2) We synthesize the buffered and bufferless versions of an open-source NoC Router to obtain a netlist for each router.
3) Using the netlist in 2, we conduct statistical timing analysis to find various critical paths in the router, and their corresponding delay distributions using data from 1 (NBTI effect and process variation) and taking into account the diverse utilization among the routers.

Table III shows the result of this experiment, where the buffered routers experience 2$\times$ more degradation compared to the bufferless ones. Hence, in order to increase NoC lifetime, a routing algorithm should aim to minimize buffered router utilization by deflecting non-critical packets in order to minimize performance impact.

V. METHODOLOGY

In this section, we discuss our simulation infrastructure that combines multiple tools to obtain a holistic analysis and

Parameter	Value
MP Size and Freq	16-core, 2Ghz
Re-order Buffer	64 entries
Pipeline Width	4/cycle
L1 I/D-Cache	16 KB/4-way, private, 2-cycle
L2 cache	128 KB/8-way, shared, 16-cycle
Cache Line	64 Bytes
NoC Network	4×4 Mesh
# of buffered routers	8
# of bufferless routers	8

TABLE IV: Processor and Network Parameters.

Fig. 6: Performance Degradation (*lower is better*).

system-level performance evaluation of NBTI aging in *hNoC* networks.

Our performance evaluation framework uses a full-system architectural simulation of a modern 16-core Chip Multi-processor interconnected via a 4×4 mesh *hNoC*. We have used GARNET [2] built inside GEMS [15] to simulate the NoC network with the many-core system model. Important processor and network parameters are shown in Table IV. We use smaller cache sizes in order to model contention without simulating for longer periods, as has been done by [18]. We use multithreaded PARSEC [4] benchmarks to evaluate our schemes at the system level.

Using our proposed routing algorithms, we evaluate the system performance by simulating 160 million committed instructions in the system (10 million per thread). We use the Fair Efficiency metric to evaluate performance in these multithreaded benchmarks, to avoid distorting the results unfairly due to a large performance improvement in a single thread [7].

To simulate NBTI degradation in each router and link, the additional delay estimated using SPICE and Statistical Timing Analysis (Section IV-B) are added in the simulator to model the system-level impact of NBTI.

VI. RESULTS

In this section, we evaluate the reduction of performance overheads brought about by the use of our proposed criticality-based aging-aware routing algorithm. Our baseline configuration is an *hNoC* architecture that has not experienced any aging-degradation. We use five different schemes for comparison purposes, as described next.

A. Routing Schemes

We show the capabilities of our criticality-based aging-aware routing algorithm by comparing it with the state-of-the-art in *hNoC* routing, the BRAR algorithm. The BRAR algorithm [22] seeks to route flits towards the buffered routers. BRAR does not model any aging-awareness and criticality of the packets. We evaluate four different schemes within our proposed algorithm, varying in their deflection threshold for non-critical packets DFL_{Max} (see Section III-B).

1) **S25:** At most, 25% non-critical packets can be deflected away from the buffered routers ($DFL_{Max} = 0.25$).
2) **S50:** At most, 50% non-critical packets can be deflected away from the buffered routers ($DFL_{Max} = 0.5$).
3) **S75:** Up to 75% non-critical packets can be deflected away from the buffered routers ($DFL_{Max} = 0.75$).
4) **S100:** All non-critical packets can be deflected away from the buffered routers ($DFL_{Max} = 1.0$).

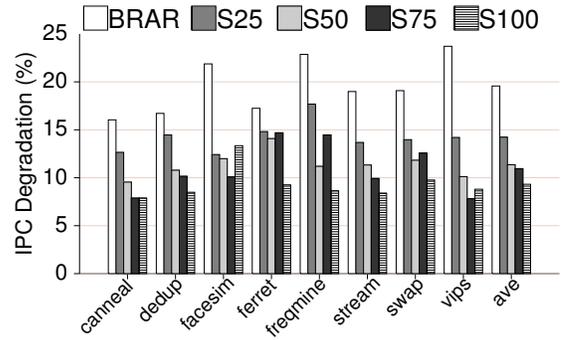### B. Performance Analysis

To show the effectiveness of criticality-based aging-aware routing in mitigating aging degradation effects, we evaluate the impact of aging (45nm, 7 years) on four performance metrics with respect to the baseline system.

- **Overall Network Latency** - Figure 5(a) shows the latency degradation of all schemes. Across several PARSEC benchmarks, our schemes reduce the aging effect on latency by 6–38%. Our scheme achieves lower utilization in buffered routers, thwarting the aging effect substantially.
- **Critical Packet Latency** - As our schemes relieve the burden on buffered routers, the latency degradation for critical packets (Fig. 5(b)) decreases even further due to reduced aging in buffered routers present in their path. For example, our scheme *S100* achieves 75% improvement in critical packet latency compared to BRAR.
- **Non-Critical Packet Latency** - Subsequently, because the non-critical packets are deflected to a non-optimal path, they are showing an opposite trend of increasing latency degradation (Fig. 5(c)): ranging from 4–25% across different schemes. In two cases (facesim and ferret), S75 has more degradation compared to S100, as based on the intrinsic traffic patterns in these benchmarks and sporadic congestion, eagerly bypassing the buffered routers marginally improves the latency in S100.
- **Performance Comparison:** Figure 6 shows the performance degradation of different schemes under aging. Reduced utilization in the central NoC components substantially mitigates their aging induced performance degradation, enabling superior system performance. For example, *S100* achieves an average of 53% reduction in aging impact compared to the BRAR across these benchmarks.
- **Energy Delay-Product per Flit (EDPPF):** Figure 7 shows the EDPPF degradation of all schemes. Except for S25, all of our schemes have lower EDPPF compared to BRAR. S75 and S100 show 16.5% and 29.3% EDPPF improvement, respectively. The reason S25 has worse EDPPF compared to BRAR is that its delay improvement is not enough to compensate for the increase in energy incurred by routing longer paths when packets are deflected.

VII. RELATED WORK

Few works beyond the BRAR [22] investigate routing approaches in an *hNoC*. Yin et al. [21] propose an energy efficient non-minimal path routing algorithm for a heterogeneous NoC running a CPU-GPU system by exploiting the
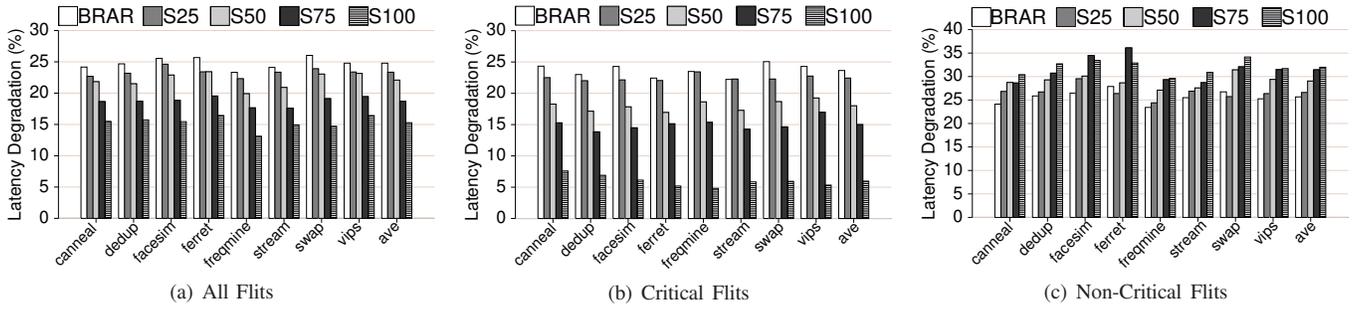
(a) All Flits   (b) Critical Flits   (c) Non-Critical Flits

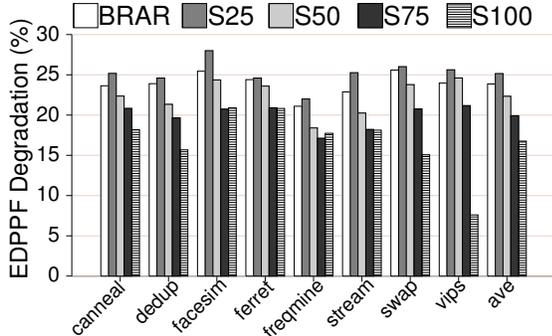Fig. 5: Latency Degradation of Different Types of Flits in *hNoC* (lower is better).

Fig. 7: EDPPF Degradation (*lower is better*).

slack provided by bandwidth-sensitive GPU traffic. In essence, they classify criticality of flits as the source of the traffic (GPU/CPU). Li et al. present runtime techniques to reduce the overall network latency of latency-critical packets by letting them bypass the router pipeline stages, hence improving performance [13].

While previous works focus on using criticality to improve performance and power efficiency in a router-homogeneous NoC, our work is the first to consider criticality in the context of reliability-driven routing in heterogeneous NoCs. Our work mitigates the effect of aging degradation by relieving the burden on routers that are likely to be highly degraded, while minimizing the system level impact of non-optimal flow control. Our WMS and deflection based schemes can also be used in homogeneous NoCs to add aging-aware functionality.

VIII. CONCLUSION

In this paper, we introduce a novel aging-aware routing algorithm for *hNoCs* that minimizes the performance overheads caused due to NBTI using a dynamic routing algorithm. Our algorithm deflects non-critical packets based on the degradation of the routers. As the buffered routers are most affected by NBTI, their minimal utilization through our routing algorithm improves latency, system performance and EDPPF by 38%, 53% and 29%, respectively as compared to BRAR routing.

ACKNOWLEDGMENTS

REFERENCES

[1] *Open Source NoC Router RTL*. https://nocs.stanford.edu/cgi-bin/trac. cgi/wiki/Resources/Router.

[2] AGARWAL, N. AND OTHERS GARNET: A detailed on-chip network model inside a full-system simulator. In *Proc. of ISPASS* (2009), pp. 33–42.

[3] BHARDWAJ, S. AND OTHERS Predictive Modeling of the NBTI Effect for Reliable Design. In *IEEE Custom Integrated Circuits Conference* (sept. 2006), pp. 189 –192.

[4] BIENIA, C. AND OTHERS The PARSEC benchmark suite: characterization and architectural implications. In *PACT* (2008), pp. 72–81.

[5] CANTIN, J. F. AND OTHERS Coarse-Grain Coherence Tracking: RegionScout and Region Coherence Arrays. *IEEE Micro 26*, 1 (2006), 70–79.

[6] CHANG, H., AND SAPATNEKAR, S. S. Statistical Timing Analysis Considering Spatial Correlations using a Single Pert-Like Traversal. In *Proc. of ICCAD* (2003), pp. 621–626.

[7] CHANG, J., AND SOHI, G. S. Cooperative cache partitioning for chip multiprocessors. In *ICS* (2007).

[8] DAS, R. AND OTHERS Application-aware prioritization mechanisms for on-chip networks. In *Proc. of MICRO* (2009), pp. 280–291.

[9] DAS, R. AND OTHERS Aergia: exploiting packet latency slack in on-chip networks. In *Proc. of ISCA* (2010), pp. 106–116.

[10] DATTA, B., AND BURLESON, W. Analysis and mitigation of NBTI-impact on PVT variability in repeated global interconnect performance. In *Proc. of GLSVLSI* (2010), pp. 341–346.

[11] INTEL. *Teraflops Research Chip*, 2006. http://techresearch.intel.com/ ProjectDetails.aspx?Id=151.

[12] KARL, E. AND OTHERS Compact In-Situ Sensors for Monitoring Negative-Bias-Temperature-Instability Effect and Oxide Degradation. In *ISSCC* (2008), pp. 410 –623.

[13] LI, Z. AND OTHERS Latency criticality aware on-chip communication. In *Proc. of DATE* (2009), pp. 1052–1057.

[14] MAHESHWARI, A., AND BURLESON, W. Current sensing techniques for global interconnects in very deep submicron (VDSM) CMOS. pp. 66 –70.

[15] MARTIN, M. M. K. AND OTHERS Multifacets general execution-driven multiprocessor simulator (gems) toolset. *SIGARCH Comput. Archit. News 33* (2005).

[16] MISHRA, A. K. AND OTHERS A case for heterogeneous on-chip interconnects for CMPs. In *Proc. of ISCA* (2011), pp. 389–400.

[17] OWENS, J. D. AND OTHERS Research Challenges for On-Chip Interconnection Networks. *IEEE Micro 27*, 5 (2007), 96–108.

[18] SUDAN, K. AND OTHERS Micro-pages: increasing DRAM efficiency with locality-aware data placement. *SIGPLAN Not. 45* (March 2010), 219–230.

[19] WANG, W. AND OTHERS An efficient method to identify critical gates under circuit aging. In *Proc. of ICCAD* (2007), pp. 735 –740.

[20] WENTZLAFF, D. AND OTHERS On-Chip Interconnection Architecture of the Tile Processor. *Micro, IEEE* (sept.-oct. 2007).

[21] YIN, J. AND OTHERS Energy-efficient non-minimal path on-chip interconnection network for heterogeneous systems. In *ISLPED* (2012), pp. 57–62.

[22] ZHAO, H. AND OTHERS Exploring Heterogeneous NoC Design Space. In *Proc. of ICCAD* (2011), pp. 787–793.

[23] ZHAO, W., AND CAO, Y. *Predictive Technology Model*. http://ptm.asu. edu/.