Architecting Reliable Multi-core Network-on-Chip for Small Scale Processing Technology

Xin Fu
CS Department, University of Illinois at Urbana-Champaign
xfu@illinois.edu

Tao Li, José A. B. Fortes
ECE Department, University of Florida
taoli@ece.ufl.edu, fortes@acis.ufl.edu

Abstract
The trend towards multi-/many-core design has made network-on-chip (NoC) a crucial component of future microprocessors. With CMOS processing technologies continuously scaling down to the nanometer regime, effects such as process variation (PV) and negative bias temperature instability (NBTI) significantly decrease hardware reliability and lifetime. Therefore, it is imperative for multi-core architects to consider and mitigate these effects in NoCs implemented using small-scale processing technology. This paper reports on a first step to optimize NoC architecture reliability in light of both PV and NBTI effects. We propose novel techniques that can hierarchically alleviate PV and NBTI effects on NoC while leveraging their benign interaction. Our low-level design improves PV and NBTI efficiency of key components (e.g., virtual channel allocation logics, virtual channels) of critical paths of the pipelined router microarchitecture. Our high-level mechanisms leverage NBTI degradation and PV information from multiple routers to intelligently route packets, delivering optimized performance-power-reliability across the NoC substrate. Experimental results show that our intra-router level techniques (i.e., VA_M1 and VC_M2) reduce guardband by 47% while improving network throughput by 24%. Our inter-router optimization scheme (i.e., IR_M3) results in 50% guardband reduction and 19% network latency improvement.

1. Introduction
The trend towards multi-/many-core processor design has made a scalable and high-bandwidth on-chip communication fabric that connects these cores critically important. The packet-switched network-on-chip (NoC) [1] is emerging as the pervasive design paradigm for multi-core communication fabrics. With the continuous down-scaling of CMOS processing technologies, reliability is becoming a primary target in NoC design [2]. The physical failure mechanisms, such as Negative Biased Temperature Instability (NBTI), Hot Carrier Injection, Electromigration, and Time Dependent Dielectric Breakdown have been playing a major role in limiting the device lifetime. Among those failures, NBTI becomes a growing threat to deep sub-micrometer CMOS technologies due to the convergence of several factors (e.g., the introduction of nitrided oxides, the increase in gate oxide fields, and operating temperature) caused by technology scaling. NBTI increases the PMOS transistor threshold voltage \( V_{th} \) and reduces the drive current, causing degradation in circuit speed, and requires an increase of the minimal voltage in storage cells to keep their stored state. This eventually leads to failures in logic circuits and storage structures due to timing violations or minimum voltage \( (V_{min}) \) limitations. It has been observed that NBTI can increase \( V_{th} \) by as much as 50mV for devices operating at 1.2V or below [3] and the circuit performance degradation may reach upwards of 20% in 10 years [4]. To combat the effect of NBTI, designers add guardbands to their designs. Guardbanding reduces the circuit frequency or increases the minimal voltage (thereby leading to higher power consumption) to defend against the expected degradation in logic circuits or storage structures during the targeted lifetime. For instance, [5] reported that 20% of the cycle time is reserved to combat NBTI degradation. Meanwhile, process variation (PV) - the divergence of transistor process parameters from their design specifications - caused by the difficulty in controlling sub-wavelength lithography and channel doping as CMOS manufacturing technology scales, results in variability in circuit performance/power and has become a major challenge in the design and fabrication of future microprocessors and NoCs [6, 10, 12, 13, 14, 15, 17]. For instance, [17] observed that the network latency can degrade as much as 37%, and the leakage power in NoC router buffers exhibits a 90% variance due to PV effects.

Although PV and NBTI can be addressed at the device- or circuit-levels, such solutions (e.g., gate sizing) are costly in terms of area and power. Several architecture and system level techniques [6, 10, 16, 18, 19] have been proposed to mitigate the effects of NBTI and PV on processor operations and achieve a lower guardband. As a result, the frequency loss and power consumption reserved for the guardband can be reduced. These techniques focus on processor cores and memory hierarchy and largely ignore the emerging NoC architectures whose design is significantly different from processor architectures. Ignoring the reliability of the NoC can turn it into a potential reliability bottleneck of multi-/many-core architectures. In NoC architectures, ultra-low latency designs are desired since shared-memory workloads are sensitive to the interconnect latency [20]. In addition, power management is also critical in a NoC, [12] reports the interconnect power consumption at approximately 30% to 40% of total chip power consumption. Since NBTI and PV affect both NoC delay and power, it is imperative to address these challenges at the NoC architecture design stage to ensure its efficiency as the underlying CMOS fabrication technologies continue to scale [23].

PMOS transistor wear-out caused by NBTI is aggravated in the presence of process variation. Under the impact of PV, the circuit operating frequency decreases significantly after the chip is fabricated (i.e., frequency is determined by the slowest critical path). The NBTI effect further exacerbates circuit performance degradation during chip operation due to increased threshold voltage. Consequently, the decreasing circuit operating frequency

---

1 This work is performed when Xin Fu is a Ph.D. student at University of Florida.
is a cumulative effect of both PV and NBTI. In other words, the guardband considering both NBTI and PV requires adding the guardbands of the two separate effects. NoC design techniques that can effectively address the combined PV and NBTI effects are needed. In this paper, we propose novel enhancements of circuits, router microarchitecture, and inter-router reliability in the presence of PV and NBTI, while achieving attractive trade-offs among NoC performance, power, reliability, and area overhead. Instead of using a flat/centralized method (as [18]), we leverage NoC microarchitecture characteristics and propose three techniques to hierarchically mitigate PV and NBTI effects in NoC at both inter-router (high) and intra-router (low) levels.

The contributions of this work are:

- We propose intra-router (low) level mechanisms aimed at improving the reliability of key components (e.g., virtual channel allocation logic and virtual channels) of critical paths of the pipelined router microarchitecture. In the virtual channel allocation logic, we observe that the NBTI degradation in a private arbiter is highly related to the utilization of its corresponding virtual channel. We propose mitigation technique VA_M1 at virtual channel allocation (VA) stage. It assigns the packets to virtual channels that are attached to faster arbiters, thereby, maximally protecting the slowest arbiters from NBTI degradation to efficiently mitigate the PV and NBTI effect.

- We propose mitigation technique VC_M2 in the virtual channels which performs at intra-router (low) level as well. VC_M2 introduces the concept of recover flits (i.e., flits with inverted values which can mitigate the NBTI degradation in storage-cell-based structures) and dynamically allocates them to virtual channels to achieve the PV-aware NBTI mitigation. VC_M2 significantly improves the reliability of virtual channels since it considers the combined effect of PV and NBTI.

- We propose inter-router (high) level mechanisms (IR_M3) that leverage NBTI degradation and PV information from multiple routers to intelligently route packets. IR_M3 achieves optimized performance-reliability-power efficiency across the NoC routers. In addition, the proposed techniques at intra-router and inter-router levels can be used together to fully explore reliability enhancement opportunities. Our experimental results show that, when compared with existing NBTI mitigation techniques, our proposed techniques (i.e., VA_M1, VC_M2 and IR_M3) achieve significant guardband and network latency reductions, leading to much lower NBTI and PV mitigation overhead.

The rest of this paper is organized as follows: Section 2 provides background on state-of-the-art NoC router architecture, process variation and NBTI. Section 3 proposes our PV-aware NBTI mitigation techniques for NoC. Section 4 describes our experimental methodologies. Evaluation results are presented in Section 5. Section 6 discusses related work. We conclude the paper in Section 7.

2. Background

2.1. Packet-Switched NoC Router Microarchitecture

Peh and Dally [24] proposed the canonical NoC virtual channel router microarchitecture. The router is input-queued and has P-input ports and P-output ports, where P is usually set as five. Four of them connect to the four cardinal directions (North, East, West, and South) and one connects to the local processing element. The router microarchitecture is composed of the following key elements: virtual channel (VC) FIFOs, routing computation unit, VC allocation logic, crossbar allocation logic, and the crossbar itself. There are five stages in the router pipeline. When a flit enters the router through one of the input ports, it is stored in the VC buffer that has been reserved by the upstream node. If the flit is a header (the first flit of a new packet), it proceeds to the route computation (RC) stage. The routing computation unit determines the output port for this new packet. The following pipeline stage is VC allocation (VA), during which the VC allocation logic attempts to assign a free VC in the next hop to the header flit. In the following cycle, if VC allocation is successful, the flit enters the switch allocation (SA) stage, during which it competes with other flits from the router for the output port. The data and tail flits belonging to the same packet as the header flit can skip the RC and VA stages and proceed to SA. Once the link between the input port and its output port is built up, the flit enters the crossbar traversal (XB) stage and is sent to the next hop via the link traversal (LT) stage. To reduce router latency and improve performance, prior studies have proposed a two-stage router pipeline, which incorporates route look-ahead [25] and path speculation [24]. The first technique performs routing one hop in advance. The second technique speculates that the waiting packet will successfully obtain the output VC from the VC allocation logic and parallelizes the switch allocation and virtual channel allocation stages. If both allocation requests are granted, the latency of switch arbitration is absorbed. The router pipeline can be further reduced to a single stage [26] by performing additional speculation, but will incur more mis-speculations at high loads and cause one-cycle penalties.

Prior studies [27, 28] have also proposed adaptive routing for NoC design. Compared with dimension-order routing (DOR), the adaptive routing schemes can achieve better fault tolerance and congestion avoidance capability. In this study, we consider a two-stage adaptive router microarchitecture similar to that proposed by Kim et al. [28]. We opt to use this design since we consider both reliability and performance in NoC design. Figure 1 shows our baseline adaptive router microarchitecture and pipeline. Adaptive routing requires extra logic to collect congestion statistics, which is used to pre-select the preferred output port (the port with the least congestion) for each packet one cycle in advance [28].

![Figure 1. Two-stage adaptive router microarchitecture](image-url)
variation refers to variations of design parameters within a single die. As technology scales, within-die variation, which is the primary focus of this study, has become more significant and is a growing threat to future microprocessor and NoC design [10, 12, 23]. PV is a combination of random effects (e.g., random dopant fluctuations) and systematic effects (e.g., lithographic lens aberrations) that occur during transistor manufacturing. Random variation refers to random fluctuations in parameters from die-to-die and device-to-device. Systematic variation, on the other hand, refers to the layout-dependent variation through which nearby devices share similar parameters. D2D variation primarily presents as a random variation, whereas WID variation is composed of both random and systematic variations. Among the design parameters, gate length (L) and threshold voltage (Vth) are two key parameters subject to large variation. Body biasing is a voltage applied between the source (or drain) and substrate to adjust the Vth. Forward body biasing (FBB) decreases Vth, reducing the delay of the transistor, but makes it leakier. Contrarily, reverse body biasing (RBB) increases Vth, creating a less leaky, slower transistor. However, the circuit overhead of generating the body-biasing voltage is non-trivial. For this reason, the BB technique is not suitable for PV mitigation at the fine grain level (e.g., intra-router microarchitecture).

NBTI is the critical reliability concern in the nano-scale processing technology era. It is the result of interface trap generation in the silicon/oxide interface of PMOS transistors. When the PMOS transistor is under negative voltage, the silicon-hydrogen bonds at the silicon/oxide interface can easily break and generate interface traps (NIT). NIT captures electrons flowing from the source to the drain and increases the PMOS threshold voltage. As a result, the transistor becomes slower and can cause failures when the delay exceeds timing specifications. NBTI leads to failures in the storage cell as well. Higher Vth requires a higher Vmin to keep the content and Vmin in the cell may not be able to satisfy this requirement due to a limited power budget. NBTI degradation can be recovered when the positive voltage is set at the gate of PMOS transistors. It helps to heal the generated interface traps and partially recovers Vth. Thus, a PMOS experiences periods of stress (gate is set as “0”) or recovery (gate is set as “1”) during its lifetime. The NBTI degradation is partially recovered once the stress is moved. Therefore, minimizing the period during which negative voltage is applied at the gate of PMOS can reduce the NBTI effect.

To mitigate NBTI degradation in combinational logic units, [16] proposed the use of special vectors as inputs to the units when they are idle, avoiding the aggressive stress on a specific PMOS. As a result, PMOS transistors in the units degrade evenly, extending the lifetime of the logic units – determined by the most degraded PMOS. In storage-cell-based structures (e.g., 6T SRAM), there is always one PMOS under stress and another under recovery. Therefore, the best NBTI degradation scenario is to wear-out the two PMOS transistors in the SRAM evenly. Storing “0” 50% of the time and “1” 50% of the time can balance NBTI degradation. To achieve this goal, [16] proposed to invalidate and store the sampled inverted values into 50% of the L1 cache lines during the entire lifetime to statistically degrade the two PMOS transistors in each SRAM bit evenly.

Mitigating NBTI degradation can reduce the required guardbands, leading to improvements in frequency and power savings. However, NBTI mitigation techniques can cause performance penalties and power overhead, making it a poor choice if the overhead outweighs that of guardbanding. NBTI_overhead (shown as Eq.1) is proposed in [16] to evaluate the overhead of NBTI tolerant schemes. It quantifies the trade-off among performance (Delay), reliability (the amount of required NBTI guardband), and power and area overhead (TDP, the Thermal Design Power, is the maximum amount of power that the cooling system is required to dissipate. TDP can be used as an overhead metric for power and area since increasing either of them will increase TDP). The Delay and TDP caused by the technique are normalized to the case without NBTI and PV effects. As can be seen, lower NBTI_overhead implies an improved approach.

\[
N_{BTI\_overhead} = (\text{Delay} \cdot (1 + \text{NBTI guardband})) \cdot TDP \quad (\text{Eq.1})
\]

3. A Hierarchical Design of NoC that Mitigates the Effects of NBTI and PV

In this section, we propose a hierarchical design of NoC that efficiently mitigates the impact of NBTI and PV while leveraging the benign interaction between them. Our hierarchical design performs at both intra-router (low) and inter-router (high) levels. Section 3.1 proposes two intra-router NBTI-and-PV-mitigation schemes. In Section 3.2, we propose a technique at inter-router level.

3.1. Intra-router Level NoC PV and NBTI Optimization

The key components of the pipelined router can be classified into combinational-logic structures (e.g., virtual channel allocation logic) and storage-cell structures (e.g., virtual channels). Our intra-router NBTI-and-PV-optimization schemes target both types of structures.

3.1.1. NBTI and PV Optimization for NoC Combinational-Logic Structures

Prior studies [24] have shown that, when compared to other router pipeline stages (e.g., XB and LT), the delay in the VA stage largely determines the frequency of canonical router microarchitecture. We thus focus our study on virtual channel allocation logic as the representative combinational-logic structure in a pipelined router.

Figure 2(a) illustrates the detailed circuit design of the virtual channels and Figure 2(b) shows a zoom-in view of the VA logic in the two-stage adaptive router [28]. A packet in a minimally routed 2D mesh can only proceed to two of the four quadrants (i.e., NE, NW, SE, and SW). The RC units determine the quadrant to which the packet should travel based on its destination direction, and the pre-selection function selects one output port for the quadrant depending on the congestion information. The VA and SA are then performed for the selected output port. In order to support adaptive routing, the VCs are partitioned into four sets. Each VC set is assigned to one quadrant and is used to collect the flits routed to this quadrant. There are three groups of VCs within each set to accept flits from the local processing element (PE) as well as from the two other directions that are not included in this quadrant. For example, the VC set assigned to NW quadrant can only accept flits from the PE and the E and S directions. Note that a packet whose destination is the local PE will be injected into the PE directly without going through the VA, SA, and XB stages. In the adaptive router, the VA stage operates in two steps: the first step assigns a free output VC (VC at the downstream nodes) to
each request from VCs in the four sets. It requires one arbiter per VC in the sets and a total number of 12v arbiters is needed (i.e. there are four sets and each set has three groups of VC, v stands for the number of virtual channels in a group). The second step produces a winner for each output VC among all the competing VCs in the sets. Correspondingly, one arbiter is required for each output VC. The total number of arbiters required for the second step is 8v. Since VA logic consists of multiple arbiters in parallel, the VA delay is sensitive to the PV effect [17]. Therefore, mitigating the NBTI and PV effects in the VA stage will allow a lower guardband and directly boost the router frequency.

Figure 2. Circuit design of the two-stage adaptive router [28]

As shown in Figure 2 (b), each VC entry has a private arbiter in the first VA step. Since there are multiple parallel arbiters in the first VA step, they exhibit various delays due to the PV effect and the slowest critical path under the impact of PV determines the delay of the first VA step. As an NBTI recovery method, the special input values [16] can be inserted into the arbiter when there is no request signal sending out from the associated VC entry (i.e. the arbiter is idle). However, it is possible that the slowest arbiter is frequently utilized, losing the opportunity for NBTI recovery. This will lead to a longer VA delay since the NBTI degradation on the slowest arbiter is not efficiently recovered even when the NBTI optimization technique is applied. In this paper, we propose NBTI and PV mitigation technique 1 at the VA stage (VA_M1) to assign a higher utilization to the faster arbiters and insert the special input values to idle arbiters. By doing so, the slowest arbiters obtain more idle time to perform the NBTI recovery. As can be seen, the fast arbiters absorb more NBTI degradation while maximally protecting the slowest arbiters from the impact of NBTI. Therefore, the guardband decreases. Due to its overhead, the adaptive body biasing (ABB) technique [19] is not suitable for mitigating PV effect at fine granularity. In our study, we apply the ABB technique in a chip-wide manner to tolerate the PV effect across the entire chip.

Figure 3. Circuit design in the first step of the VA stage (one VC entry is shown)

As mentioned earlier, the first VA step selects a free output VC entry from a specific VC group at the downstream node. Since arbiter utilization depends on the requests from the associated VC, instead of blocking the requests to the arbiter we prefer to avoid using the VC even when it is free. If there is no flit stored in the VC, no request will be sent out. In other words, the arbiter gets the opportunity to recover from NBTI degradation if its associated VC is idle. Figure 3 shows the circuit of the first VA step (only one VC entry is shown). When a flit at a local VC sends an output VC allocation request to its private arbiter, the request will be fanned out to v AND gates (v stands for the number of virtual channels in a group) and ANDed with each VC status (i.e. idle: “1”, or occupied: “0”) at the downstream node. The arbiter only accepts the fan-out request signals for downstream VCs whose status is idle. At the downstream node, when the VC associated with the slowest arbiter is free and there is more than one free VC in the group, VA_M1 marks it as occupied before a credit describing its status is sent back to the upstream node. Consequently, other free VCs in the same group that link to faster arbiters will be used to store the flit. In the case where there is only one free VC, which is exactly the one linked to the slowest arbiter, VA_M1 marks it as free in order to maintain the performance. Since VCs are not fully occupied during most of the router service period, VA_M1 has substantial opportunities to mitigate the NBTI degradation from the slowest arbiter to faster ones. In addition, the arbiters use a least-recently-served priority scheme [26], every faster arbiter receives even NBTI degradation migrated from the slowest one. Therefore, the NBTI degradation will not accumulate on a specific fast arbiter, which can eventually become the new bottleneck for the guardband reduction.

In VA_M1, online detection of the aggregated NBTI and PV effects on the arbiters is required to identify the slowest arbiter dynamically. IDDQ, which describes the standby leakage current in the circuit, can be applied to detect such effects. The IDDQ values can demonstrate the underlying parameter variations [31]. Recently, [32] discovered that IDDQ can also be used to detect NBTI degradation because the leakage current decreases exponentially as Vth increases in transistors. Therefore, IDDQ has the capability to capture both the static and dynamic variations in Vth.

Figure 4 presents the implementation of VA_M1. An IDDQ detector is attached to each arbiter. In our study, we assume the critical paths within each arbiter have similar delay due to the systematic effect of PV. Therefore, one detector per arbiter can well describe the NBTI and PV effect in the arbiter, and the arbiter with the lowest IDDQ value is the slowest one. The
detector operates periodically and sends the testing current to a 2-input NMOS analog voltage comparator [30] that determines which VC should be marked as occupied. The comparison result is written into the VC status table (a hardware component existing in the typical flow-control-based router), which reflects the availability of each VC in the router. Note that the IDDQ detector and comparator perform concurrently with the VA stage and the backward credit representing a VC status does not need to wait for the latest VC status before being sent out. Therefore, no extra delay is introduced into the router pipeline. When inserting the special input values into the arbiter, invalid grant signals will be generated. In order to block them from entering the second VA step, the grant signal is fanned out to the arbiters in the second step. Therefore, the delay variation in the first step cannot be tolerated by the second step – the slow arbiters in the first step will affect the entire VA delay. Moreover, VA_M1 mitigates the NBTI degradation within the first step and it does not introduce any extra NBTI effect to the second step. In order to mitigate the NBTI and PV effects in the second VA step, VA_M1 inserts the special input values when the arbiter is idle.

As can be seen, the three-transistor based IDDQ detectors, the 2-to-1 MUXs, and the comparators are the major extra overhead induced by VA_M1. Our gate-level estimation shows that VA_M1 causes around 3% area overhead to the virtual channel allocation logics.

The switch allocation unit in the SA stage, which is another combinational-logic structure, is sensitive to the effects of PV and NBTI as well. The folding mechanism [17] that splits the SA stage into two phases is applied in our study to mitigate the detrimental PV and NBTI effect. In addition, as shown in Figure 2, the crossbar delay is dominated by the wire delay [24]. Wires are immune to the NBTI degradation and techniques for reducing the wire delay are out of the scope of this paper.

### 3.1.2 NBTI and PV Optimization for NoC Storage-Cell-Based Structure

In this section, we present an optimization technique for VC buffers, which are the representative storage-cell structures in NoC. Recall that an efficient NBTI-mitigation approach for the storage-cell structure is to keep 50% of the bits storing the sampled inverted value. To apply this technique in our work, we perform the inversion at half-flit-size granularity, which is large enough to statistically summarize the NBTI degradation in all of the VCs. To implement this technique, only half of the VC capacity can be used to store the flits and the performance loss is significant on workloads that exhibit heavy traffic. Due to the systematic component of process variation (i.e. the spatial correlation effect of WID), transistors usually share similar behavior with nearby transistors. We synthesized and generated the layout of a prototype router at 45nm processing technology. The router layout is similar to those reported in previous work [20]. The methodology to model the systematic and random variations on $V_{th}$ can be found in Section 4. Figure 5 presents the $V_{th}$ variation map for the router. Note that it is based on the physical layout of the router instead of the conceptual router microarchitecture shown in Figure 2. As shown, the distance between the transistors in the same VC set (shown by the rectangular) is much shorter than that between the transistors from different VC sets. As a result, the $V_{th}$ of transistors within each input port exhibit similar characteristics allowing VCs from the same port to be grouped into one area. Transistors from the same area are characterized by one uniform $V_{th}$, which is determined by the worst-case $V_{th}$ (i.e. the highest $V_{th}$, because it affects the minimal voltage has to be applied) in that area. Therefore, there are several areas in the router (one area per set) with various representative $V_{th}$ values. We observe that areas with low $V_{th}$ can tolerate more NBTI degradation and, as long as the final $V_{th}$ does not exceed that in the area with the highest $V_{th}$, the strict inversion percentage (i.e. 50%) can be appropriately relaxed in those areas.

The second mitigation technique proposed in this paper targets VC buffers (VC_M2) and assigns lower inversion percentage to areas with lower $V_{th}$. By doing so, there are more VC buffers to hold flits and mitigate the performance loss, achieving a good trade-off among reliability, performance, and power. Based on our NBTI modeling, we characterize the relation between inversion percentage and the corresponding increase of $V_{th}$. Using the above information and the statistics of the highest $V_{th}$ and the $V_{th}$ of a given area, one can compute out the inversion percentage for that area during the calibration time. Figure 6 (a) shows one VC set when VC_M2 is applied. In VC_M2, a recover flit is defined as the flit with two half-flit-size inverted values, and it does not contain valid data. To implement VC_M2, two extra buffers are added in the set. One is INV buffer which stores the sampled half-flit-size inverted value. The other one is called the status buffer and records the status (”0”: free, “1”: occupied) of all the flit-size buffers in the VCs: if one flit-size buffer is written by a flit or a recover flit, its corresponding status bit will be set to 1. On the contrary, when the buffer is released, the status bit is reset to 0. Note that at the SA stage, the flit will not send out a traversal request if there is no free buffer in the allocated output.
VC. In general, a credit returns to the upstream node when there is at least one free buffer in the allocated VC at the current node. Note that the credit describes the availability of buffers in a specific VC – it is different from the VC status credit (discussed in VA_M1 techniques) which shows the availability of a VC. In VC_M2, when sending back the credit that represents the buffer availability, the buffer holding the recover flit should also be marked as occupied. However, the VC status is still considered as free if the VC only contains recover flits. Since the recover flit does not belong to any packet, allocating a recover-flit-occupied VC to a new packet will not cause the mix of packets into single VC, which is not allowed in the flit based flow control [24]. VC_M2 mainly focuses on the credit of buffer availability and it does not affect the credit of VC availability. Therefore, it is orthogonal to VA_M1.

![Figure 6. The implementation of VC_M2](image)

As introduced in Section 2.1, in a typical router, each VC is a FIFO-based structure [24, 27] and every flit in the VC will move from the tail to the head and finally enter the crossbar. In VC_M2, the recover flit follows the same policy. When it arrives at the head of the FIFO (as shown in Figure 6 (a)), it will not be read out, but will be overwritten by the following flit. Therefore, there is no request to the arbiters in the first VA step. As can be seen, even though a VC is re-defined as “occupied” in VA_M1 due to its link to the slowest arbiter, the buffers in this VC still can be used to hold recover flits and its private arbiter becomes idle during the period for NBTI recovery. In each set, there is a threshold value to describe the number of inverted half-flit-size buffers corresponding to the required inversion percentage. Meanwhile, a counter (IBCNT) is attached to the set to track the inversion number. Once the IBCNT is below the threshold (e.g. the recover flit arrives the head of the FIFO and is overwritten), VC_M2 will read the status buffer, and pick a free flit-size buffer to store the inverted value and maintain the inversion percentage. When the recover flit is overwritten, a buffer at the tail of the FIFO will be released simultaneously. Therefore, there is always at least one free buffer for inversion (shown in Figure 6 (a) to (b)). VC_M2 performs simultaneously with router pipeline stages and it does not increase the router latency. When implementing VC_M2 at each set of virtual channels, it introduces a status buffer with the bit size equal to the number of buffers in the set, a half-flit-size buffer to keep the inverted value, a 6-bit buffer for inversion threshold, a 5-bit counter for IBCNT, and simple combinational-logics. In summary, the area overhead caused by VC_M2 to the virtual channels is around 3%.

3.2. Inter-router Level NoC PV and NBTI Optimization

In the NoC architecture, routers experience WID variations allowing them to support different frequencies. In this work, we assume NoCs with a single frequency domain, which is determined by the slowest router, and apply the chip-wide adaptive body biasing technique [19] to mitigate the PV effect. The routers under the PV positive effect (e.g. faster routers) can be assigned higher utilization (e.g. handling a larger number of packets) while exhibiting less NBTI degradation. By doing so, the routers under negative PV effect will process less packets, which limits the impact of NBTI. Since the NBTI degradation migrates from the slower routers to faster ones, the guardband for the chip frequency significantly decreases. Note that network congestion status should be taken into consideration when mitigating the NBTI effect at the inter-router level. For instance, it is possible that the packets are all routed to the faster routers while mitigating the NBTI effect on slow routers. Consequently, the faster routers quickly become congested, resulting in longer network latency. Prior work [27] proposes regional-congestion-aware routing to balance the traffic load in the NoC and improve the performance. However, the NBTI and PV effects are not considered.

![Figure 7. The implementation of IR_M3](image)
node which can reduce the VA utilization in the slowest router; (4) selects a preferred output port for the packets based on the computed metrics to achieve a good trade-off between reliability and performance; and (5) propagates its aggregated statistic to its neighbor routers. Figure 7 shows the implementation of IR_M3. The aggregation and propagation modules are added into the adaptive router to perform steps (1), (2) and (5). In addition, the pre-selection unit takes the computed metrics to perform step (4). Note that the above-mentioned steps perform simultaneously with the router pipeline and they do not cause any extra delay to the flit traversal.

As mentioned earlier, IDDQ can quantify both NBTI and PV effects. Since the delay in VA stage mainly determines the router frequency, in each router, we reuse those IDDQ detectors deployed in VA_M1 and the lowest IDDQ current is used as the reliability estimation. A high IDDQ value represents small NBTI and PV effects. As suggested in [27], the combination of free virtual channels and crossbar demands (vc_xb) is used for congestion estimation in our study. A high vc_xb value indicates low congestion. In order to route the packets to an optimal path, the collected estimations need to be integrated as a single statistic.

In our study, we use Garnet [33], which is a detailed cycle-accurate NoC simulator, and extend it to support the two-stage adaptive routing. We use Orion 2.0 power model [34] to track the dynamic and leakage power of the NoC. All simulations are performed for a 25-node (5x5) mesh network. We restrict our evaluation to a 2D mesh NoC, but the general principles presented here could be applied to other NoC topologies as well.

Each VC group has 4 virtual channels. Each VC holds four 128-bit flits. We evaluate our techniques using a set of representative synthetic traffic patterns (i.e., uniform random, transpose, bit-complement and tornado). In addition, we use traffic traces from real-world workloads such as SPLASH-2 [29], SPECComp [21] and Specjbb 2005 [22] in our evaluation. For synthetic traffic simulation, we modify the Garnet simulator to inject packets during a period of 1 million cycles (including 100K warm-up cycles). Both one-flit and five-flit long packets are injected.

At the circuit level, we model variations on L and Vth since they are the two major process variation sources. We focus our study on WID variation, which is an additive effect of systematic and random variations. To model the random effects of WID variation, we generate random variables that follow a normal distribution (random variables are generated through Monte-Carlo simulation) with standard deviation $\sigma = \sqrt{\sigma_{sys}^2 + \sigma_{rand}^2}$, where $\sigma_{sys}$ and $\sigma_{rand}$ are standard deviations for random and systematic variation respectively. In this study, we simulate processors developed using 45nm process technology and assume $\sigma/\mu = 12\%$ and $\sigma_{sys} = \sigma_{rand} = \sigma/\sqrt{2}$ based on variability projections from [11]. We scale down the chip floor plan of a 25-core processor and NoC layout to 45nm and generate 200 chips for statistical analysis. Predictive Technology Models (PTM), are used to provide the basic device parameters for HSPICE simulations to obtain the circuit delay. The dynamic NBTI degradation in $V_{th}$ is modeled by applying the reaction-diffusion (RD) model proposed in [37], the PMOS stress and recovery cycles are obtained via the cycle-accurate simulator, and the signal possibility (the ratio of the PMOS stress time to the total simulation time) is computed and inserted into the model to determine the shift in $V_{th}$ due to NBTI.

Since both NBTI and PV effects are considered in our study, we extend the NBTI_overhead metric to NBTI+PV_overhead (shown in Eq.2), which quantifies the technique efficiency in relation to both NBTI and PV. Correspondingly, the NBTI+PV guardband is named as NBTI+PV_guardband. In Eq.2, network latency is applied to describe the delay. Note that NBTI+PV_overhead quantifies the overhead introduced by NBTI and PV mitigation schemes. Therefore, a lower value means a better technique.

$$NBTI_{\&PV\_overhead} = (\text{Delay} - (1 + NBTI_{\&PV\_guardband})) \cdot TDP$$ (Eq. 2)

5. Evaluation

5.1. Effectiveness of VA_M1

As shown in [16], inserting the special input values (SIV) into a combinational logic unit during its idle periods shows a strong capability in NBTI mitigation. We compare the effectiveness of VA_M1 with that of SIV in virtual channel allocation logics. We incorporate adaptive body biasing in SIV scheme for the purpose of fair comparison. Note that the VA_M1 shares the same PV guardband as SIV. Figure 8 (a)-(d) shows NBTTI_guardband and NBTI+PV_overhead achieved by VA_M1 and SIV when using the four synthetic traffic patterns with different packet injection rates. Since there are multiple routers on the chip and VA_M1 performs at intra-router level, we focus our analysis on routers that suffer the most severe NBTI degradation.
Note that VA_M1 does not block any packet during the VA stage so it does not influence the number of cycles that a packet needs to traverse through the router pipeline. Therefore, the network latency is not affected by our proposed techniques. Furthermore, given the fact that the techniques do not affect processor cores, we expect that the overall impact to CMP performance is negligible. The area and power overhead of VA_M1, which affects the TDP, is incorporated in our NBTI&PV_overhead calculation.

Figure 8 shows that VA_M1 provides strong NBTI mitigation under heavy traffic loads. For example, it achieves 47% NBTI_guardband reduction compared to SIV in uniform random traffic when the injection rate is 0.1. (In our study, we set the lifetime target as 10 years and use the reduction of required guardband as the metric to evaluate the effectiveness of our techniques for reliability enhancement. In the case when the guardband is fixed for various workloads, our techniques will extend the lifetime.) When there are light traffic loads (e.g. 0.02 flits/node/cycle), VA_M1 gains less NBTI_guardband reduction. Due to the lower arbiter utilization, the long idle period already provides a good opportunity for SIV to achieve NBTI recovery. Therefore, there is limited room for VA_M1 to further improve it.

On the other hand, when the network loads are extremely heavy (e.g. 0.42 flits/node/cycle in uniform random traffic), all the arbiters are busy and the possibility for VA_M1 to migrate the utilization is low. Therefore, the guardband reduction becomes smaller. In general, VA_M1 reduces the NBTI&PV_overhead by 10% compared to SIV. One may notice that VA_M1 shows smaller NBTI mitigation improvement for tornado traffic. This is because in tornado traffic, packets are only sent along X direction, the VCs and their local arbiters at the east and west network directions are highly utilized. As a result, VA_M1 does not have enough opportunities to hide the NBTI degradations via the faster arbiters at the X dimension. Note that the NBTI guardband is determined by the maximum $V_{th}$. Even though VA_M1 efficiently mitigates the NBTI effect in the Y dimension, its benefit does not show up in the final results.

5.2. Effectiveness of VC_M2

We compare VC_M2 with the scheme (i.e. 50% inversion) that fixes the inversion percentage as 50% for NBTI mitigation. Note that adaptive body biasing technique is trigged in 50% inversion as well. Figure 9 (a)-(d) presents the network latency and NBTI&PV_overhead of VC_M2 and 50% inversion for the different traffic patterns. Note that both techniques target the same maximum $V_{th}$. Therefore, they achieve the same guardband. The TDP of VC_M2 and 50% inversion is 1.03 and 1.02 respectively. As can be seen, when there is heavy traffic, VC_M2 significantly improves the network latency and reduces the NBTI&PV_overhead compared to 50% inversion. Take uniform random traffic as an example, compared to 50% inversion, VC_M2 can absorb 24% more NoC load before becoming saturated (at the NoC saturated point, the network latency is three times larger than in the zero-load case) and this also results in 164X NBTI&PV_overhead reduction. VC_M2 obtains similar latency as 50% inversion when the traffic is light. As most of the VC buffers are free, inverting 50% of the free buffers will not introduce significant performance penalty.

5.3. Effectiveness of IR_M3

We compare IR_M3 with two other schemes: RCA (Regional Congestion Aware routing) [27], which is an adaptive routing scheme that gives the performance metric 100% weight and therefore does not consider the NBTI and PV effect, and AR_reliability, which only takes the reliability statistics into consideration when routing packets. Adaptive body biasing technique is included in both RCA and AR_reliability. Figure 10 shows the NBTI&PV_guardband and network latency of these three techniques across different traffic patterns. Note that SIV is applied to all the three schemes. As Figure 10 shows, in most cases, IR_M3 achieves a much lower guardband (e.g. 50%) compared to RCA. The benefit becomes smaller when the traffic is heavy due to the high router utilization. On the other hand, IR_M3 significantly improves the network throughput (around 19%) when compared to AR_reliability. Figure 11 shows the NBTI&PV_overhead, which considers trade-offs among performance, reliability, and power. As can be seen, IR_M3 outperforms both RCA and AR_reliability. In general, it reduces the NBTI&PV_overhead by 14% and 30% when compared to the two schemes, respectively. Figure 11 shows that when the packet injection rate increases, the NBTI&PV_overhead in AR_reliability first increases to an extremely high value (due to its higher network latency compared to the baseline case without PV and NBTI effect), and then drops down (because the network latency in the baseline case increases to infinite as well). Different from AR_reliability, the NBTI&PV_overhead in IR_M3 and RCA first drops before the variations, because both of them improve the network throughput compared to the baseline case. For example, under the same injection rate (e.g. 0.32 flits/node/cycle in transpose traffic), the network latency in the baseline case becomes infinite while it still stays low in the two schemes. IR_M3, RCA, and AR_reliability merge to the same NBTI&PV_overhead when the injection rate is around 0.48 flits/node/cycle because the network is saturated in all these cases.
Since our intra-router and inter-router techniques are orthogonal to each other, we propose to combine them together to gain additional benefits in NBTI&PV mitigation. The combined techniques cause total 4% area overhead to each router. We compare our combined scheme with the technique (i.e. SIV+50% inversion+RCA) that uses SIV, 50% inversion and RCA together, adaptive body biasing is incorporated as well. Compared with SIV+50% inversion+RCA, our combined technique can reduce the NBTI&PV guardband and NBTI&PV overhead by as much as 70% and 41% respectively while improving network throughput by 5%. Note that the network latency has already been improved by using SIV+50% inversion+RCA due to the use of congestion aware adaptive routing to balance the buffer utilization in routers.

5.4. Real Workload Results

Figure 12 (a)-(c) shows the NBTI&PV guardband, network latency, and NBTI&PV overhead of the real workloads when the proposed schemes are applied. The results are normalized to the technique of SIV+50% inversion+RCA. Our techniques are able to improve the NBTI recovery on benchmarks with high traffic load (the heavy-traffic real workloads stress NoC similarly to the synthetic-workloads with moderate injection-rate). For example, on water-spatial which exhibits high traffic, when compared with SIV+50% inversion+RCA, the combined scheme can reduce the
NBTI&PV guardband and network latency by 59% and 4% respectively, leading to 33% reduction in NBTI&PV_overhead. Similarly, on benchmarks with relatively high traffic (e.g. water-squared, eqquake, fma3d, and mgrid), the combined technique also improves the NBTI&PV guardband, network latency and NBTI&PV_overhead by 48%, 2% and 20%, respectively. The improvement on medium-load benchmarks (e.g. ocean, radix, barnes and fft) is 30% in NBTI&PV guardband, 1% in network latency, and 12% in NBTI&PV_overhead. The combined technique achieves smaller gains in light-load benchmarks such as raytrace, because there are fewer uses of VA arbiters and less congestion in routers. The NBTI&PV guardband reduction is 5% with 0.5% and 2% improvement in network latency and NBTI&PV_overhead, respectively.

6. Related Work

In the past, various PV and NBTI mitigation techniques have been proposed on processor core architectures. Liang et al. [6] exploited variable latency techniques in register file and function units design to tolerate PV. [10] used linear programming to find the best voltage and frequency levels for each core of the CMP in the presence of PV. The NBTI effect on SRAM arrays is modeled and studied in [7], where it is shown that read stability degrades due to NBTI and that the degradation is exacerbated in the presence of PV. The combined effect of PV and NBTI on processor core has been modeled and analyzed in [18, 8].

There have been several studies on PV effect in NoC design. In [12], Li and Peh observed that NoC design choices are very much influenced by the effects of PV. Ogras and Marculescu [13] studied the use of NoC consisting of multiple voltage-frequency islands to cope with parameter variation problems. To tolerate the effect of process variations on NoC, [14] proposed using self-correcting links that automatically detect delay variations and compensate them. The above studies exclusively focus on PV while our work targets both PV and NBTI, which inherently interplay with each other. Specifically, we focus on leveraging NoC microarchitecture characteristics for the purpose of NBTI and PV mitigation. In addition, [9] proposed a technique called “Razor” to tune the supply voltage by monitoring the error rate caused by PV and NBTI during circuit operation, thereby eliminating the need for voltage margins. “Razor” mainly targets combinational logic. In our study, we target the mitigation of NBTI and PV effect in both combinational circuits and storage-cell structures with desirable trade-offs among performance and reliability. To the best of our knowledge, there has been no prior work on addressing PV&NBTI effect in NoCs.

7. Conclusions

NoC is becoming the imperative communication fabric for emerging multi-/many- core processors. Existing NoC designs largely assume reliable processing technologies and uniform transistor characteristics. As CMOS fabrication technologies approach the nano- and atom- scales, process variation can significantly degrade the performance and reliability of these designs. This problem is further compounded by the NBTI effect, which wears-out transistors and reduces their lifetime. Therefore, it is unwise to ignore the impact of PV and NBTI in NoC architecture design. In this paper, we propose novel techniques at both intra-router and inter-router levels to mitigate the impact of PV and NBTI on NoC. Experimental results show that our intra-router techniques (i.e. VA_M1 and VC_M2) reduce guardband by 47% while improving network throughput by 24%. Our inter-router optimization scheme (i.e. IR_M3) results in 50% guardband reduction and 19% network latency improvement.

Acknowledgements

This work is supported in part by NSF grants CCF-0916384, CCF-0845721 (CAREER), and Microsoft Research Safe and Scalable Multi-core Computing Award. Xin Fu is also supported by the NSF under Grant 0937060 to the Computing Research Association for the CIFellows Project. José Forés is also funded by the BellSouth Foundation.

References