

# Adaptive Inter-router Links for Low-Power, Area-Efficient and Reliable Network-on-Chip (NoC) Architectures

Avinash Karanth Kodi<sup>†</sup>, Ashwini Sarathy<sup>‡</sup>, Ahmed Louri<sup>‡</sup>, Janet Wang<sup>‡</sup>

<sup>†</sup>*Dept. of EECS, Ohio University, Athens, OH 45701, USA*

<sup>‡</sup>*Dept. of ECE, University of Arizona, Tucson, AZ 85719, USA*

E-mail: kodi@ohio.edu, {sarathya,louri,wml}@ece.arizona.edu

**Abstract**— The increasing wire delay constraints in deep sub-micron VLSI designs have led to the emergence of scalable and modular Network-on-Chip (NoC) architectures. As the power consumption, area overhead and performance of the entire NoC is influenced by the router buffers, research efforts have targeted optimized router buffer design. In this paper, we propose **iDEAL** - inter-router, dual-function energy and area-efficient links capable of data transmission as well as data storage when required. **iDEAL** enables a reduction in the router buffer size by controlling the repeaters along the links to adaptively function as link buffers during congestion, thereby achieving nearly 30% savings in overall network power and 35% reduction in area with only a marginal 1 – 3% drop in performance. In addition, aggressive speculative flow control further improves the performance of **iDEAL**. Moreover, the significant reduction in power consumption and area provides sufficient headroom for monitoring Negative Bias Temperature Instability (NBTI) effects in order to improve circuit reliability at reduced feature sizes.

## I. INTRODUCTION

Technology scaling into the deep sub-micron regime has led to the development of multi-core architectures with increased transistor density on a single chip. However, the increasing global wire delays [1] limit the potential improvement in performance, leading to the modular and scalable packet-switched NoC paradigm [2, 3, 4]. As concluded in the recent NSF sponsored workshop on on-chip interconnects [5], one of the major research challenges currently faced by NoC designers is that of power dissipation with the router buffers consuming about 46% of the router power. The increased power density and temperature accelerate device degradation, and hence reduce the lifespan of the circuit. On the other hand, simply reducing the number of router buffers to reduce the power and area overhead degrades the network performance as the performance and flow control of the network are primarily characterized by the router buffers [6].

Current high speed VLSI designs require repeater insertion along the wires to overcome the quadratic increase in delay with the wire length [1]. It has been shown in [7] that the repeaters can be designed to sample and hold a data bit, in addition to their conventional functionality. Therefore, with repeaters as potential buffer elements, we can employ them to

adaptively function as buffers along the links under high network loads, when there are no more buffers in the router. Data storage along the links has been explored by [8, 9] to achieve a latency-insensitive design and by [10] in order to resolve timing errors in overclocked NoCs.

In this paper, we propose **iDEAL** - inter-router dual-function, energy and area-efficient links for NoCs by employing circuit and architectural techniques at the inter-router links and the router buffers respectively. At the links, we deploy circuit level enhancements to the existing repeaters so that they adaptively function as buffers during congestion. As the correct operation of the adaptive link buffers depends on the error-free detection of the congestion signal, we employ a double-sampling technique within the control block to overcome arbitrary timing errors along the congestion signal. The double-sampling technique [11] has been employed in correcting timing errors due to voltage scaling [12] and increased operating frequency [10]. In our proposed architecture we double-sample the congestion signal to overcome potential timing errors due to increased coupling noise and process variations at the deep sub-micron technologies. Compared to a conventional repeater-inserted control line, this control technique operates reliably and consumes significantly less power as it can be disabled in the absence of congestion.

At the router buffer, we deploy architectural techniques such as static and dynamic buffer allocation [13, 14] to prevent performance degradation, while sustaining or improving the performance of a generic router. While static buffer management leads to insufficient or unused buffers [13], dynamic buffer management allocates the incoming flit (the basic flow control unit, a packet consists of several flits) to any free buffer, leading to higher throughput, although at the cost of higher control management. In [13] the number of virtual channels (VCs) and the depth of buffers/VC are dynamically adjusted based on the traffic load, complicating the control logic and increasing the packet latency due to higher packet interleaving [6]. Therefore, in our proposed work, we adopt a dynamic VC table based approach with fixed number of VCs, thereby achieving the flexibility of storing flits dynamically without excess control overhead. Moreover, the congestion control from the downstream to the upstream router enables aggressive flit transmission without waiting for credits to return, thereby further improving the throughput.

The combination of circuit and architectural techniques in iDEAL using the dual-function links and the dynamic router buffer allocation, allows us the flexibility of reducing the router buffer size to decrease the power and area overhead without significant performance degradation. In addition, the reduction in power consumption and area provides sufficient design headroom for improving the reliability of the circuits. Reliability, especially the Negative Bias Temperature Instability (NBTI) effects has emerged as a major threat to the current CMOS technology. Prior research has focused on modeling NBTI at the device level and gate level [15, 16, 17]. [18] has reviewed current NBTI research and provides guidelines towards reliability in general purpose logic and memory circuits, by proposing guard-banding and adaptive body biasing techniques. [19] is one of first papers to propose an NBTI-aware processor design. Our proposed architecture with the iDEAL methodology offers a possible NBTI-aware solution for NoCs. The additional circuits required to monitor NBTI effects increase the overall power consumption and area. However, our proposed low-power, area-efficient architecture provides design headroom for the Thermal Design Power (TDP) of the auxiliary circuits and thereby improves the circuit reliability without additional overhead.

Unlike other NoC designs where performance is improved at the cost of major changes to the router design, the changes in the proposed architecture pertain only to the input buffer and the allocation of the input buffer space to the incoming flits. Synthesized designs in the 90 nm technology at 500 Mhz and 1.0 V, show a reduction of 30% in the overall network power and a reduction of 35% in area, when half the router buffers are removed. Moreover, the proposed architecture enables NBTI monitoring without additional overhead and improves device lifetime by providing both power and area design headroom. Cycle accurate network simulation on  $8 \times 8$  mesh and folded torus network topologies shows only a marginal 1-3% loss in throughput. In addition, the throughput can be further improved by 10% with aggressive transmission of flits without credit turn-around.

## II. DESIGN OF INTER-ROUTER DUAL-FUNCTION LINKS

### A. Proposed Link Implementation

Figure 1 shows the proposed repeater-inserted interconnect, with the conventional repeaters replaced by three-state repeaters. When the control input to a repeater stage is low, the three-state repeaters in that stage function like the conventional repeaters transmitting data. When the control input to the repeater stage is high, the repeaters in that stage are tri-stated and hold the data bit in position. Once congestion is alleviated, the control logic is disabled and the three-state repeaters return to the conventional mode of operation. The adaptive dual-function links hence enable a decrease in the number of buffers within the router and save appreciable power and area.

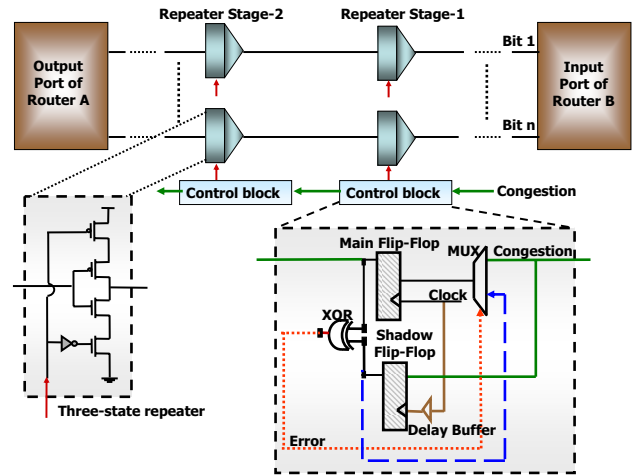


Fig. 1. Proposed inter-router dual-function links with the control blocks.

### B. Control Block Implementation

The control block shown in Figure 1 enables the three-state repeaters to adaptively function as link buffers during congestion. A single control block is sufficient to control the functionality of all the repeaters in one stage. The incoming congestion signal is delayed by one clock cycle at each control block. In the next clock cycle, the repeaters in that stage are tri-stated and the congestion signal travels to the next control block. Hence each repeater stage is successively tri-stated to hold the data in position, until the congestion-release signal arrives. The control block in Figure 1 is more efficient than a conventional repeater-inserted control line [7], as it provides the following advantages: (1) Unlike conventional repeaters, the control circuit operates accurately at variable clock speeds. The congestion signal is double-sampled and the shadow flip-flop makes the circuit tolerant to timing errors. (2) The control circuit can be disabled when there is no congestion, thus reducing the power consumption along the congestion control line.

## III. DESIGN OF ROUTER BUFFERS

In packet-switched NoCs, every processing element (PE) is connected to an NoC router as shown in Figure 2(a), with most NoCs commonly adopting network topologies such as mesh, or folded torus for regularity and modularity [6, 20, 21, 22]. In wormhole switching, each packet that arrives on the input port progresses through router pipeline stages (routing computation(RC), virtual channel allocation (VA), switch allocation (SA), switch traversal (ST)) before it is delivered to the appropriate output port [6]. At each intermediate router, only the header flit of every packet is responsible for the first two pipeline stages of RC and VA.

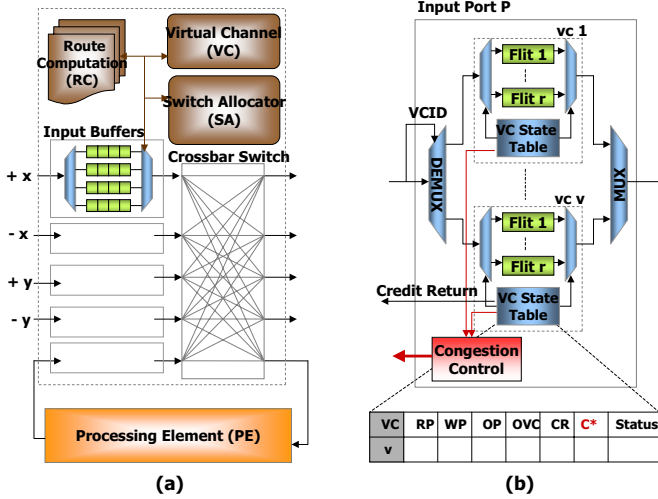


Fig. 2. (a) A generic  $5 \times 5$  NoC router architecture (b) The proposed static buffer allocation with congestion control.

### A. Statically Allocated Router Buffers

The proposed statically allocated router buffer design with congestion control is shown in Figure 2(b). For a router architecture with  $P$  ports,  $v$  VCs/port and  $r$  flit buffers/VC the total number of buffers/port is  $z = vr$ . Each input VC is associated with a VC state table [6, 13]. It maintains the state for each incoming packet and ensures that the body flits are routed to the correct output port. The VC identifier (VCID) of the incoming flit allows the DEMUX to switch to the correct input VC. The Read Pointer (RP) and the Write Pointer (WP) are used to read the flit into the buffer and write the flit out to the crossbar. The Output Port (OP) is provided by the RC stage, Output VC (OVC) is provided by the VA stage. Credits (CR) indicate the total storage available at the downstream router. Given that each VC has  $r$  credits, a credit is consumed when a flit is transmitted to the downstream router. A credit is returned to the upstream router when a flit is read out of the buffer. The Status field indicates the current status of the VC - idle, waiting, RC, VA, SA, ST, and others.

In the generic NoC design, the total number of input buffers is  $vr$  per input port. With the additional  $c$  buffers in the link, the total storage now becomes  $vr + c$ . The number of credits available at each VC is  $(vr + c)/z$ . This allows routers to send additional flits into the network, even if the storage is in the link, instead of the router buffer. Every VC state table maintains another field  $C^*$  which indicates congestion. When  $C^*$  is set, the congestion control is activated which in turn holds the data in the network link itself. When a flit is read from the buffer, the  $C^*$  field is cleared, which in turn allows data flits to enter into the router.

This nominal change does not impact the design of the network router and leads to significant power savings and area gain. However, at high network load, this design leads to head-of-line (HoL) blocking in the link buffers. When the congestion field  $C^*$  is set for a particular VC, the corresponding flits

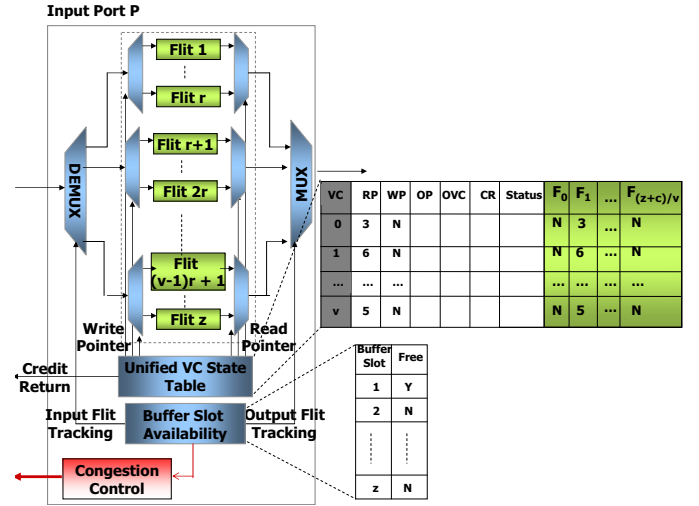


Fig. 3. Proposed dynamic buffer allocation with congestion control.

are held in the network link. These flits block the flits headed towards other VCs, although the other VCs may have their  $C^*$  field cleared. A more attractive alternative is dynamically allocated router buffers as explained in the next section.

### B. Dynamically Allocated Router Buffers

As ViChar's [13] table based approach had solved issues pertaining to latency and scalability, we have adopted a similar idea but limited the number of VCs to prevent excess control overhead. We adopt the unified buffer architecture and augment the architecture with a 'Unified VC State Table' (UVST). The maximum size of the UVST is  $O(v)$  as compared to the ViChar which is  $O(vr)$ . For an incoming flit, the 'Buffer Slot Availability' (BSA) tracking system keeps track of all buffer slots and allocates the first buffer slot found to be free. Similarly, for a departing flit, BSA deallocates the buffer slot and adds it to the list of free slots maintained in the table, as shown in the inset of Figure 2(b). The table contains buffer slots  $F_0, F_1, \dots, F_{(z+c)/v}$  in addition to the regular fields of RP, WP, OP, OVC, CR and Status fields. For fairness purposes, the number of credits is equally divided between all the VCs as  $(z + c)/v$  per VC. When BSA finds only a single non-null pointer in its base table, it triggers the congestion signal and when a free buffer slot is created by a departing flit, the BSA releases the congestion signal.

In iDEAL, the link buffers can be viewed as serial FIFO buffers as opposed to the parallel FIFO buffers used within the routers. Therefore, eliminating the HoL blocking is critical in iDEAL. Dynamic allocation of buffer slots using the table based approach significantly reduces the HoL blocking.

#### IV. PERFORMANCE EVALUATION OF THE PROPOSED ARCHITECTURE

In this section, we evaluate the proposed dual-function links and router buffers in terms of power dissipation and area overhead. The notation followed for the different cases is of  $vn_V - rn_R - cn_C$ , where  $n_V$  is the number of VCs per input port,  $n_R$  is the number of router buffers per VC and  $n_C$  is the number of link buffers. For example, the baseline is denoted as  $v4 - r4 - c0$ , implying 4 VCs per input port, 4 router buffers per VC and 0 link buffers. In each case, the design is implemented in Verilog and synthesized using the Synopsys Design Compiler tool and the TSMC 90 nm technology library, at an operating frequency of 500 MHz and a supply voltage of 1 V.

##### A. Power and Area Estimation for the Links

The links are assumed to be 2 mm long and the baseline design has 8 optimally spaced conventional repeaters along each wire of the 128-bit wide links. For the baseline design, the total power consumed by the link per flit traversal is 2.45 mW. When all the conventional repeaters are replaced by the three-state repeaters, the total power consumed by the link for each flit traversal is found to be 3.94 mW. In the presence of congestion, the control block consumes a power of only about 6  $\mu$ W.

The area consumed by the repeater stages is found to be 8,960  $\mu$ m<sup>2</sup> in case of the baseline and 10,240  $\mu$ m<sup>2</sup> when all the conventional repeaters along the link are replaced by the three-state repeaters.

##### B. Power and Area Estimation for the Router

The router buffers are implemented as FIFO registers with the associated control logic. Considering both the write and read operations in the buffer, the total power consumed for a 128-bit flit in the buffer is estimated to be 19.54 mW, for the baseline design with 16 buffers. Decreasing the buffer size by 50% reduces the power per flit traversal to 11.57 mW, resulting in a 40.77% savings in buffer power alone. For the configurations with 4 VCs per input port, the arbiter in the router consumes a power of 0.15 mW, for a single arbitration task and the switch in the router consumes 0.31 mW per flit traversal. When the number of VCs is decreased to 3 per input port, the power consumed by the arbiter reduces to 0.09 mW per arbitration and the power consumed by the switch decreases to 0.27 mW per flit traversal.

The buffer area in case of the baseline is 81,407  $\mu$ m<sup>2</sup>. A 50% decrease in the buffer size reduces the area to 48,066  $\mu$ m<sup>2</sup>, leading to a 35% savings in the overall area including the links and the router buffers.

##### C. Design Headroom for NBTI-related Reliability Issues

NBTI affects PMOS transistors whose gate input is at logic '0', introducing the need for a higher threshold voltage and constraining the transistor speed. Among all the circuit parameters affecting the influence of NBTI, temperature plays a

decisive role. Higher temperatures accelerate the NBTI-related degradation [15, 19]. In order to effectively manage the influence of NBTI, designers monitor the temperature in cores and functional blocks. Additional circuitry is included to change the input of the PMOS gates to logic '1' while the PMOS gates are in the 'idle' state [19]. This technique utilizes the self-healing ability of the PMOS [19], as the PMOS is not stressed when it is turned 'OFF' by the logic '1' input. However, the additional circuits lead to an increase in power consumption and area overhead.

iDEAL achieves a significant reduction in power density and temperature of the router by 'moving some of the buffers from the router to the link'. As is well known, each router in the NoC is closely placed to the core or the PE in the multi-core architecture. Therefore reducing the router power and area has a significant impact on the temperature of the PEs. Moreover, iDEAL provides additional design headroom for NBTI-related reliability issues. In order to quantify the additional design headroom achieved by the iDEAL methodology we consider the Thermal Design Power (TDP), which is the maximum power that is required to be dissipated by the cooling system [19]. In the current context, it is the additional power that is consumed by the auxiliary circuits enabling the self-healing of the PMOS gates. The TDP for an NoC with  $N$  cores,  $TDP_{NoC}$  is given by

$$TDP_{NoC} = \sum_{i=1}^N TDP_i \quad (1)$$

The complexity of the auxiliary circuits in turn depends on the functional block under consideration [19]. For a power-efficient architecture, it is essential to minimize the overhead due to the auxiliary circuits. The significant reduction in power consumption achieved by iDEAL provides the additional headroom to overcome this constraint, as shown by Equation (2).

$$TDP_{NoC} \leq P_{saved} \quad (2)$$

where  $P_{saved}$  is the reduction in power consumption achieved by iDEAL.

As the area linearly impacts TDP [19], it is required to limit the area in order to decrease TDP. iDEAL achieves an area-efficient architecture, thereby reducing the impact on the TDP. The additional headroom provided by the area-efficient iDEAL methodology is given by

$$TDP_{NoC} = K \times S_r \quad (3)$$

where  $S_r$  is the area reduction achieved by iDEAL and  $K$  is the factor determining the impact of the area on the TDP.

Therefore, the iDEAL methodology offers a good opportunity to add auxiliary circuits such as inverting input circuits for memory-like structures, mitigating mechanisms in latches and self-healing circuits for the combinational logic in the design [19].

## V. SIMULATION RESULTS AND DISCUSSION

A cycle-accurate on-chip network simulator was used to conduct a detailed evaluation of the proposed architecture in  $8 \times 8$  mesh and folded torus networks under several synthetic traffic patterns (Uniform Random(UN), Bit-Reversal(BR), Butterfly(BU), Matrix Transpose(MT), Complement(CO), Perfect Shuffle(PS), Tornado(TO), Neighbor(NE)) as well as the SPLASH-2 suite benchmarks (FFT with input data set 64K points; LU with  $256 \times 256$ ,  $16 \times 16$  block; MP3D with 48000 molecules; RADIX with 1M integers, 1024 radix and WATER with 512 molecules). For simplicity, the test configurations are referred to as  $n_V - n_R - n_C$  in the following discussion.

**Total Network Power and Throughput:** Figures 4(a) and 4(b) show the total power consumed for a network load of 0.5, in the mesh and folded torus networks respectively. The power dissipated in the control blocks is negligible and is not visible at the scale considered. By reducing the router buffer size, all the configurations achieve a reduction in power compared to the baseline, with the 4-2-8 case showing about 30% savings in the total network power. From Figures 4(c) and 4(d), the saturation throughput shows almost similar performance for 4-4-0, 3-4-4 and 4-3-4. The 4-2-8 shows only about 3% drop in performance. This result is significant as we can save about 35% of the area and yet achieve similar performance as the baseline by dynamically allocating the router buffers and using the additional link buffers at high network loads.

**Router Buffer Power and Throughput for All Traffic Patterns:** Figure 4(e) shows the power consumed at the router buffers and Figure 4(f) shows the throughput achieved at a network load of 0.5 for the  $8 \times 8$  mesh, under all the synthetic traffic patterns considered, for the 4-4-0, 4-3-4 and 4-2-8 configurations. Power savings is obtained for both the 4-3-4 and the 4-2-8 cases under all the traffic patterns. From Figure 4(f), there is no significant decrease in throughput under any of the traffic patterns considered.

**Throughput and Power for SPLASH-2 suite benchmarks:** Figures 4(g) and 4(h) show the normalized execution time and normalized total power consumed for the selected SPLASH-2 suite benchmarks for 4-4-0, 4-3-4 and 4-2-8 configurations. From Figure 4(g), the 4-3-4 and 4-2-8 configurations do not show significant drop in performance, in fact the drop is less than 1%. From Figure 4(h), the power savings from the 4-3-4 and 4-2-8 configurations are 20% and 30% respectively.

**Throughput using Aggressive Speculation:** Figure 4(i) shows the saturation throughput using an aggressive speculation technique, for the  $8 \times 8$  folded torus network under uniform traffic. The number of credits available to the upstream router is speculatively increased to 8 as the congestion control circuit enables an aggressive flit transmission without waiting for the credits from the downstream router. This technique improves the performance of iDEAL by about 10% (as seen in the 4-2-8 case), without additional power or area overhead.

## VI. CONCLUSION

As recent research has shown, the major issue in NoC design is the increasing power consumption. iDEAL proposes to reduce the number of router buffers, thereby achieving a significant savings in power and area. As this impacts performance, we provide adaptive dual-function links for data storage when required. Simulation results show that by reducing the router buffer size in half, iDEAL achieves nearly 40% reduction in buffer power alone, more than 30% savings in the overall network power and 35% savings in the total area. In addition, the dynamically assigned buffers with aggressive speculative flow control show up to 10% improvement in performance. The savings in power consumption and area provides significant design headroom to overcome circuit reliability issues due to NBTI effects. This paper shows that eliminating some of the buffers in the router and using adaptive link buffers saves an appreciable amount of power and area, without significant degradation in the throughput or latency.

## ACKNOWLEDGMENT

This research was partially supported by NSF grants CCR-0538945 and ECCS-0725765.

## REFERENCES

- [1] R. Ho, K. W. Mai, and M. A. Horowitz, "The future of wires," *Proceedings of the IEEE*, vol. 89, pp. 490–504, April 2001.
- [2] W. J. Dally and B. Towles, "Route packets, not wires: On-chip interconnection networks," in *Proceedings of the 38th Design Automation Conference (DAC)*, Las Vegas, NV, USA, June 18–22 2001, pp. 684–689.
- [3] L. Benini and G. D. Micheli, *Networks on Chips: Technology and Tools*. Morgan Kaufmann, 2006.
- [4] H. S. Wang, L. S. Peh, and S. Malik, "Power-driven design of router microarchitectures in on-chip networks," in *Proceedings of the 36th Annual ACM/IEEE International Symposium on Microarchitecture (MICRO)*, Washington DC, USA, December 03–05 2003, pp. 105–116.
- [5] 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," *IEEE Micro*, vol. 27, no. 5, pp. 96–108, September–October 2007.
- [6] W. J. Dally and B. Towles, *Principles and Practices of Interconnection Networks*. San Francisco, USA: Morgan Kaufmann, 2004.
- [7] M. Mizuno, W. J. Dally, and H. Onishi, "Elastic interconnects: Repeater-inserted long wiring capable of compressing and decompressing data," in *Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, February 5–7 2001, pp. 346–347.
- [8] L. P. Carloni, K. L. McMillan, and A. L. Sangiovanni-Vincentelli, "Theory of latency-insensitive design," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 9, pp. 1059–1076, September 2001.
- [9] L. P. Carloni and A. L. Sangiovanni-Vincentelli, "Coping with latency in SoC design," *IEEE Micro*, vol. 22, no. 5, pp. 24–35, September 2002.
- [10] R. Tamhankar, S. Murali, S. Stergiou, A. Pullini, F. Angiolini, L. Benini, and G. D. Micheli, "Timing-error-tolerant network-on-chip design methodology," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 7, pp. 1297–1310, July 2007.
- [11] M. Nicolaïdis, "Time redundancy based soft-error tolerance to rescue nanometer technologies," in *Proceedings of the 17th IEEE VLSI Test Symposium*, San Diego, CA, USA, April 25–30 1999, pp. 86–94.
- [12] T. Austin, D. Blaauw, T. Mudge, and K. Flautner, "Making typical silicon matter with Razor," *Computer*, vol. 37, no. 3, pp. 57–65, March 2004.
- [13] C. A. Nicopoulos, D. Park, J. Kim, N. Vijaykrishnan, M. S. Yousif, and C. R. Das, "ViChar: A dynamic virtual channel regulator for network-on-chip routers," in *Proceedings of the 39th Annual ACM/IEEE International Symposium on Microarchitecture (MICRO)*, Orlando, FL, USA, December 9–13 2006, pp. 333–344.

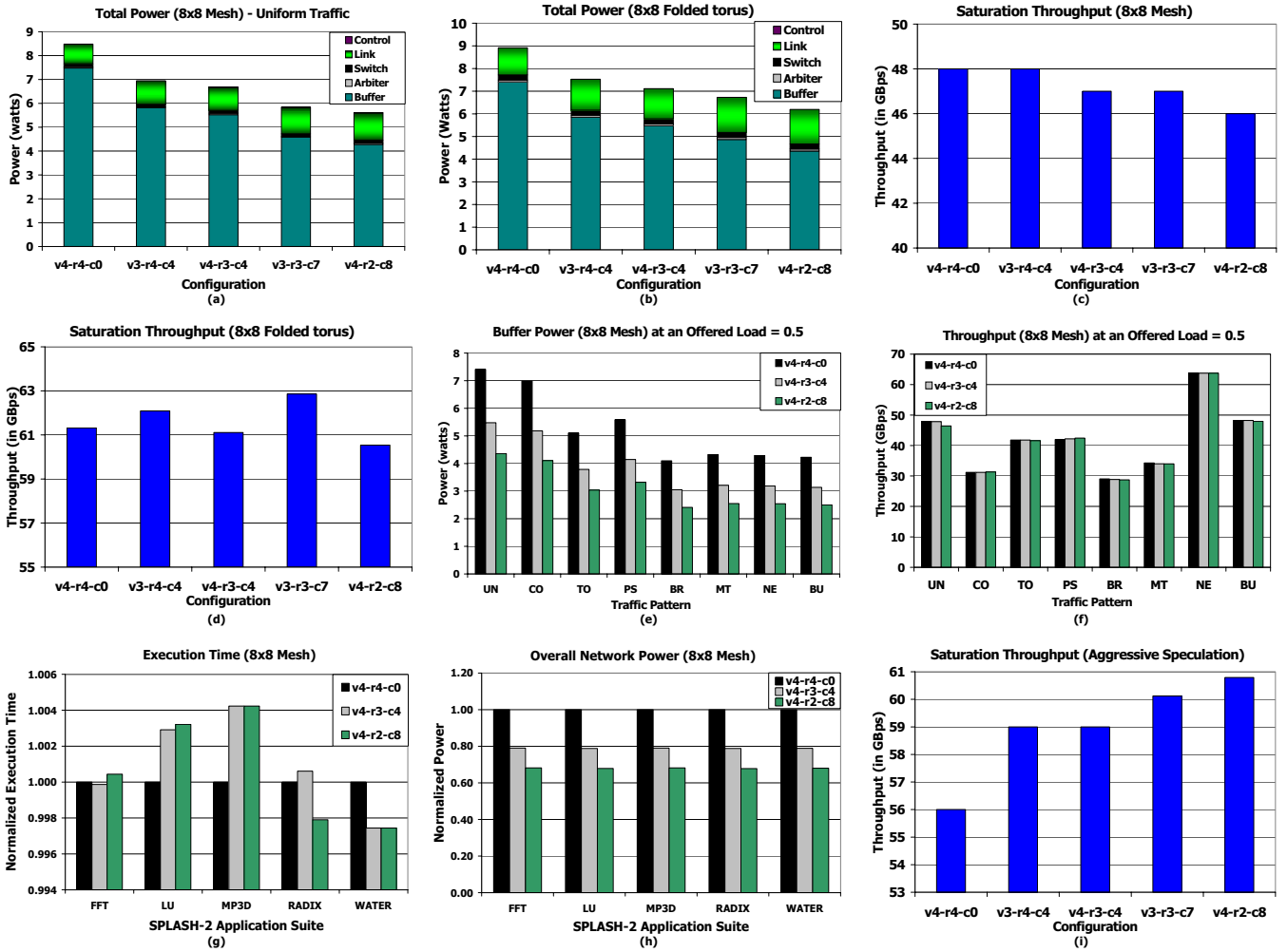


Fig. 4. (a) - (d) Total network power and Saturation throughput under Uniform traffic for  $8 \times 8$  mesh and folded torus networks (e) Buffer power and (f) Throughput at 0.5 network load in the  $8 \times 8$  mesh, using dynamic buffer allocation, for all the synthetic traffic patterns considered (Uniform Random (UN), Complement (CO), Tornado (TO), Perfect Shuffle (PS), Bit-Reversal (BR), Matrix Transpose (MT), Neighbor (NE), Butterfly (BU)) (g) Normalized execution time and (h) Normalized total network power under the SPLASH-2 application suite, for the  $8 \times 8$  mesh (i) Throughput for Aggressive Speculation using 8 credits, under Uniform traffic for the  $8 \times 8$  folded torus network.

- [14] Y. Tamir and G. L. Frazier, "High-performance multiqueue buffers for VLSI communication switches," in *Proceedings of the 15th Annual International Symposium on Computer Architecture (ISCA)*, Honolulu, Hawaii, USA, May-June 1988, pp. 343-354.
- [15] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," in *Microelectronics Reliability*, 2005, pp. 71-81.
- [16] S. Bhardwaj, W. Wang, R. Vattikonda, Y. Cao, and S. Vrudhula, "Predictive modeling of the NBTI effect for reliable design," in *IEEE Custom Integrated Circuits Conference*, San Jose, CA, USA, September 10-13 2006, pp. 189-192.
- [17] W. Wang, S. Yang, S. Bhardwaj, R. Vattikonda, S. Vrudhula, F. Liu, and Y. Cao, "The impact of NBTI on the performance of combinational and sequential circuits," in *Proceedings of the 44th Design Automation Conference (DAC)*, San Diego, CA, USA, June 4-8 2007, pp. 364-369.
- [18] K. Kang, S. Gangwal, S. Park, and K. Roy, "NBTI induced performance degradation in logic and memory circuits: How effectively can we approach a reliability solution?" in *Proceedings of the 13th Asia and South Pacific Design Automation Conference*, Seoul, Korea, January 21-24 2008, pp. 726-731.
- [19] J. Abella, X. Vera, and A. Gonzalez, "Penelope: The NBTI-aware processor," in *Proceedings of the 40th Annual ACM/IEEE International Symposium on Microarchitecture (MICRO)*, Chicago, IL, USA, December 1-5 2007, pp. 85-96.
- [20] J. Hu and R. Marculescu, "DyAD smart routing for networks-on-chip," in *Proceedings of the 41st IEEE/ACM Design Automation Conference*, San Diego, CA, USA, June 7-11 2004.
- [21] Y. M. Boura and C. R. Das, "Performance analysis of buffering schemes in wormhole routers," *IEEE Transactions on Computers*, vol. 46, pp. 687-694, 1997.
- [22] N. Ni, M. Pirvu, and L. Bhuyan, "Circular buffered switch design with wormhole routing and virtual channels," in *Proceedings of the International Conference on Computer Design (ICCD)*, Austin, TX, USA, October 1998, pp. 466-473.