

DYNAMIC POWER MANAGEMENT OF HIGH PERFORMANCE
NETWORK ON CHIP

A Dissertation

by

SUMAN KALYAN MANDAL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2011

Major Subject: Computer Engineering

Dynamic Power Management of High Performance Network on Chip

Copyright 2011 Suman Kalyan Mandal

DYNAMIC POWER MANAGEMENT OF HIGH PERFORMANCE
NETWORK ON CHIP

A Dissertation

by

SUMAN KALYAN MANDAL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Rabi Mahapatra
Committee Members,	Duncan M. (Hank) Walker
	Radu Stoleru
	Gwan S. Choi
Head of Department,	Duncan M. (Hank) Walker

December 2011

Major Subject: Computer Engineering

ABSTRACT

Dynamic Power Management of High Performance Network on Chip.

(December 2011)

Suman Kalyan Mandal, B. Tech. (Hons.), IIT Kharagpur, India

Chair of Advisory Committee: Dr. Rabi N. Mahapatra

With increased density of modern System on Chip(SoC) communication between nodes has become a major problem. Network on Chip is a novel on chip communication paradigm to solve this by using highly scalable and efficient packet switched network. The addition of intelligent networking on the chip adds to the chip's power consumption thus making management of communication power an interesting and challenging research problem. While VLSI techniques have evolved over time to enable power reduction in the circuit level, the highly dynamic nature of modern large SoC demand more than that. This dissertation explores some innovative dynamic solutions to manage the ever increasing communication power in the post sub-micron era.

Today's highly integrated SoCs require great level of cross layer optimizations to provide maximum efficiency. This dissertation aims at the dynamic power management problem from top. Starting with a system level distribution and management down to microarchitecture enhancements were found necessary to deliver maximum power efficiency. A distributed power budget sharing technique is proposed. To efficiently satisfy the established power budget, a novel flow control and throttling technique is

proposed. Finally power efficiency of underlying microarchitecture is explored and novel buffer and link management techniques are developed.

All of the proposed techniques yield improvement in power-performance efficiency of the NoC infrastructure.

DEDICATION

To my loving parents

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Mahapatra, and my committee members, Dr. Walker, Dr. Stoleru and Dr. Choi, for their guidance and support throughout the course of this research.

Thanks also go to my wonderful friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. I also want to extend my gratitude to the National Science Foundation, whose generous grants enabled my research. Also, special thanks to Dr. Mohanty from University of North Texas for helping with parts of my research.

Finally, thanks to my wonderful family for their constant encouragement and belief. And, last but most important, thanks to my wife Kasturi, for her invaluable support, patience and love that kept me going.

NOMENCLATURE

CMOS	Complementary Metal Oxide Semiconductor
CNI	Core Network Interface
DRAM	Dynamic Random Access Memory
FIFO	First In First Out Buffer
IP	Intellectual Property
MPEG	Motion Pictures Expert Group
NOC	Network on Chip
PTM	Predictive Technology Model
SOC	System on Chip
SRAM	Static Random Access Memory
TCP	Transport Control Protocol
VLSI	Very Large Scale Integration

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS	viii
LIST OF FIGURES.....	x
LIST OF TABLES	xii
1. INTRODUCTION.....	1
2. POWER BUDGET SHARING.....	7
2.1. Related Research	9
2.2. The Ant System.....	10
2.3. Power Distribution Technique	12
2.3.1. Ant System for Power Management	12
2.3.2. Distributed Power Sharing Scheme.....	14
2.3.3. Power Management Technique	17
2.4. Evaluation Framework	18
2.4.1. Experimental Setup	18
2.4.2. Evaluation Criteria	24
2.4.3. Test Scenarios	25
2.5. Results	26
2.5.1. Performance Comparison	26
2.5.2. Utilization Comparison	29
2.5.3. Overhead Analysis	30
2.5.4. Reactivity Analysis	32
2.5.5. Real Benchmark Evaluation.....	33
2.6. Optimum Power Sharing and PowerAntz	35
2.7. Conclusion and Analysis	36
3. POWER BUDGET SATISFACTION	37
3.1. Related Research	39

	Page
3.2. Flow Control and Adaptive Throttle Mechanism	40
3.2.1. Router Architecture	40
3.2.2. Router Power Model	41
3.2.3. Flow Control with Early Notification	41
3.2.4. Adaptive Throttle Mechanism.....	42
3.3. Experimental Evaluation	44
3.3.1. The Simulation Platform	45
3.3.2. Evaluation Criteria	45
3.3.3. Results	45
3.4. Summary	49
4. POWER EFFICIENT MICROARCHITECTURE	50
4.1. Related Research	51
4.2. Preliminaries.....	52
4.2.1. NoC Flow Analysis	52
4.2.2. Nanoscale CMOS Buffer Design	55
4.3. Router Architecture	59
4.3.1. Virtual Buffer Architecture	61
4.3.2. Central Physical Buffer Design.....	61
4.3.3. Note on Performance of Centralized Buffer	62
4.4. Buffer Power Management	62
4.4.1. Block Level Power Management	63
4.4.2. Flit Level Power Management	65
4.4.3. Dynamic Power Gating	67
4.5. Adaptive Link Control	68
4.5.1. Underutilized Link	68
4.5.2. Link Modes	69
4.5.3. Link Mode Selection	69
4.6. Experimental Results.....	70
4.6.1. Simulation Platform	70
4.6.2. Performance	70
4.6.3. Power.....	73
4.6.4. Overhead	74
4.7. Summary	74
5. CONCLUSION	76
REFERENCES.....	78
VITA	84

LIST OF FIGURES

	Page
Figure 1: NoC in Abstraction	2
Figure 2. Large NoC with Hot and Cold zones	8
Figure 3: The Ant System	11
Figure 4: The Ant Generation Process	15
Figure 5: The Ant Consumption Process	17
Figure 6: A Network Tile	18
Figure 7: Router Architecture	19
Figure 8: Structure of Ant Packet.....	20
Figure 9: NoCSim Architecture	21
Figure 10: A Sample Simulator Configuration	23
Figure 11: Performance Comparison (Latency).....	27
Figure 12: Comparison of Budget Sharing	29
Figure 13: Overhead with Varying FIR	30
Figure 14: Overhead with Varying Network Size and Flit Injection Rate	31
Figure 15: Response Time with Varying FIR for Three Load Scenarios.	33
Figure 16: Media SoC Setup	34
Figure 17: Problem of Router Throttling	38
Figure 18: Pipelined Router Architecture	40
Figure 19: Flow Control State Machine	42
Figure 20: Energy Calculation for Throttle Threshold Determination.....	43

	Page
Figure 21: Latency vs. Flit Injection Rate for Random Traffic.....	46
Figure 22: Throughput and Efficiency vs. Injection	47
Figure 23: Throughput at Different Power Budget using Adaptive Throttling.....	47
Figure 24: Improvement in Throughput using Adaptive Throttling	48
Figure 25: Distribution of Traffic Types for MediaBench Applications	54
Figure 26: Buffer Utilization Distribution in a Mesh Topology	54
Figure 27: Structure and Operation of 7T SRAM.....	56
Figure 28: Total Power Dissipation of 7T SRAM	57
Figure 29: The Centralized Buffer Router Architecture	60
Figure 30: Block Level Feedback System.....	63
Figure 31: The Block Level Power Manager FSM.....	64
Figure 32: Adaptive Control for State Assignment.....	65
Figure 33: FSM for Dynamic Flit Inversion Controller	66
Figure 34: Data Retention Characteristic of the 7T SRAM.....	67
Figure 35: Latency vs. Injection Rate	71
Figure 36: Throughput Comparison	72
Figure 37: Energy Saving Comparison	73

LIST OF TABLES

	Page
Table 1: Energy Consumption of Router Components	20
Table 2: Simulator Feature Set	22
Table 3: Experiment Details	27
Table 4: Power Budget Sharing Experiment Details	28
Table 5: Improvement in Power Budget Utilization	30
Table 6: Real Benchmark comparison of PowerHerd and PowerAntz	34
Table 7: Area and Power Overheads	49
Table 8: Static and Dynamic Power Dissipation of SRAM.	59
Table 9: Link Modes in Proposed Dynamic Link Control	69
Table 10: Controller Overheads	74

1. INTRODUCTION

21st century has seen the pinnacle of digital age. Since Jack St. Clair Kilby invented integrated circuit, digital electronics has come a long way. As amazing it was to have a fully operational circuit on a single piece of silicon, today's chips have an entire computer sitting on a single chip. The era of System-on-Chip (SoC) is here.

During early development of SoCs the communication between different nodes on the chip has been largely ad-hoc. The traditional shared bus communication paradigm has been prevailing as the communication backbone e.g. AMBA, AHI etc. While effective and sufficient on smaller number of nodes, shared bus quickly saturates as more and more nodes compete for the same shared bus for communication. Modern SoCs are already reaching tens of cores with hundreds coming in near future. This explosion in number of cores that a chip can accommodate necessitates a more sophisticated communication infrastructure that is high performance as well as scalable.

Network-on-Chip (NoC) has emerged as a promising interconnect platform for large scale System-on-Chip design [1] [2] [3]. NoC provides the required infrastructure for reliable communication that is based on globally asynchronous and locally synchronous paradigm. In essence, NoC applies the philosophy of the packet switched computer network to enable communication between on chip nodes. Instead of point to point or shared bus connecting multiple nodes on the chip, NoC facilitates

This dissertation follows the style of *IEEE Transactions on Very Large Scale Integration Systems*.

communication by creating a network infrastructure. A typical NoC consists of Core Network Interfaces (CNI) [4], Routers and Links. An abstract view of an SoC with NoC as the communication infrastructure is shown in Figure 1.

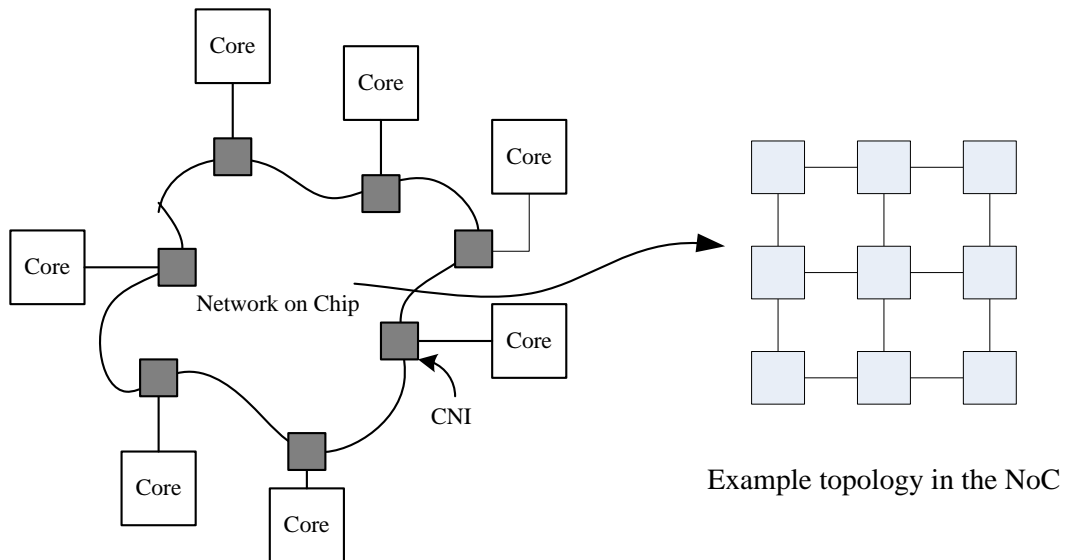


Figure 1: NoC in Abstraction

In this example the NoC infrastructure is a set of routers connected in mesh topology. The key idea to be taken home from here is that NoC allows separation of communication from the rest of the SoC design. Another important aspect of NoC compared to traditional communication paradigm is that, NoC involves introduction of additional active elements on the chip (CNIs, Routers etc).

With increased size and density, power consumption has been increasing. Being an active component on the chip, communication framework accounts for a significant share of chip power consumption [5] [6] [7]. Low power dissipation ability limits the

power consumption that may be allowed in a large scale network on chip. Hence efficient power management technique is necessary to achieve maximum performance while maintaining power consumption within limits. While the active components in a NoC make it more power hungry, they also create opportunity for novel management techniques to make it efficient in terms of power and performance. Consequently power management of NoC has quickly become a popular research topic [5] [8] [1].

In general the power management problem can be classified into two broad categories: Static vs. Dynamic [9]. While variety of circuit level power minimization technique take care of the static power management, the unique nature of NoC and its operation makes dynamic power management more difficult and hence challenging. Power consumption directly reflects in the form of temperature of the chip. Today's dense silicon is severely constrained in terms of the heat it can dissipate. This physically limits the amount of power that can be safely consumed by the chip without burning out. Apart from efficiency, this is a major concern and driving factor for dynamic power management. To ensure safe operation the power management system has to ensure operation within the safe consumption limit. It is necessary to have effective power accounting/estimation and control to provide this kind of safety. A number of techniques have been proposed in literature that addresses this issue [10] [8]. While these techniques manage to keep the power consumption within safe limit, they are not necessarily most efficient in terms of utilization of the available power budget. This creates opportunity for development of more intelligent and efficient solution.

Moreover, any power management algorithm running at the system level needs to and will affect the operation of the underlying network layer and microarchitecture. To build the most efficient network infrastructure, it is necessary to understand the interaction between the system and the network level. Also, to effectively control and manipulate the power consumption, any system or network level management technique requires support from underlying microarchitecture. For example, to make use of varying throughput requirement from the system, a dynamic frequency scaling enabled microarchitecture is necessary. Solutions to power management issues have been addressed at different layers in the literature. This dissertation aims at closing the discontinuity by addressing dynamic power management top down from system to microarchitecture level.

This dissertation makes three novel contributions to solve the dynamic power management problem in high performance NoCs. The first contribution is at the system level power management. Efficiency is the key to get the most of out the watts in a system. And efficiency comes from intelligent distribution of power budget in a power constrained environment. Taking cue from the nature, an ant system based dynamic and distributed power budget sharing scheme is proposed. The proposed design is evaluated against similar techniques found in the literature and shown to be superior. Also, through analysis of overhead and cost the feasibility and practicality of the proposed technique is established.

The second contribution is inspired by closer observation of how system level power management techniques affect the underlying network. For any effective system

level power management technique, it is necessary to manipulate underlying network and hence component functions to estimate and/or control power consumption. This can lead to problems in communication if not designed properly. Packet throttling is a commonly used technique to control power consumption in communication infrastructure [8] [10]. This dissertation proposes an intelligent network level flow control mechanism to minimize performance hit in the network's operation due to actions initiated by higher level power management policies. The proposed technique is evaluated against state of the art and is shown to outperform them in several dimensions.

Finally, to enable efficient dynamic power management by taking advantage of modern extreme nanoscale CMOS devices and their operational characteristics, novel power efficient microarchitecture solutions are proposed. The most significant architectural element responsible for power consumption in a communication infrastructure in today's nanoscale SoCs are the communication buffers and the communication links. To complete the top down approach, novel buffer microarchitecture and link management techniques are proposed. Utilizing adaptive control and dynamic buffer resizing in tandem with dynamic link control, the proposed design enabled very fine grained dynamic power management without sacrificing performance and efficiency.

The dissertation is organized as follows: Section 2 discusses the power budget sharing problem and proposes PowerAntz: the ant system based intelligent and distributed power budget sharing solution. Basics of ant system are described followed by details of how it is adopted in a power sharing problem. Relevant experimental setup

and results are discussed to establish the effectiveness of PowerAntz scheme. Section 3 introduces the implications of a system level power management framework such as PowerAntz on underlying methods of power budget satisfaction. A computer network inspired solution is proposed and evaluated. Section 4 dives deeper into the microarchitecture to explore and exploit the opportunities that modern nanoscale CMOS devices present. A novel dynamic buffer management is proposed along with dynamic link control. The proposed microarchitecture is evaluated against state of the art. Finally Section 5 summarizes the research findings with concluding notes about the future research directions these findings have opened up.

2. POWER BUDGET SHARING

Simple power management offers uniform budget distribution among routers in NoC which may not be adequate for all circumstances. Large scale NoC based systems have non-uniform power consumption due to varying task processing rates and communication requirements [11]. Figure 2 shows a typical power consumption scenario in a NoC. There can be hot zones, i.e. routers need more than allocated maximum power budget to process incoming packets (shown with - in Figure 2) and cold zones, i.e. zones having surplus power budget (shown with + in Figure 2) within the chip separated by a neutral zone (shown with N in Figure 2), consisting of routers consuming power as allocated. Existing power management schemes like PowerHerd [10], PC [12] do not provide enough flexibility to distribute spare power budget from cold zone to hot zones crossing the neutral zone. These schemes restrict power sharing among immediate neighbors. For example, in the scenario illustrated in Figure 2, with traditional power management even though there is surplus power budget in the cold zone, the hot zone can't receive the surplus power budget information from the cold zone. Such inefficiencies in power budget allocation lead to underutilization of available power budget and impacts system performance. Throttling of high activity routers also leads to increased idle period in less active routers and consequently increased idle energy consumption as well.

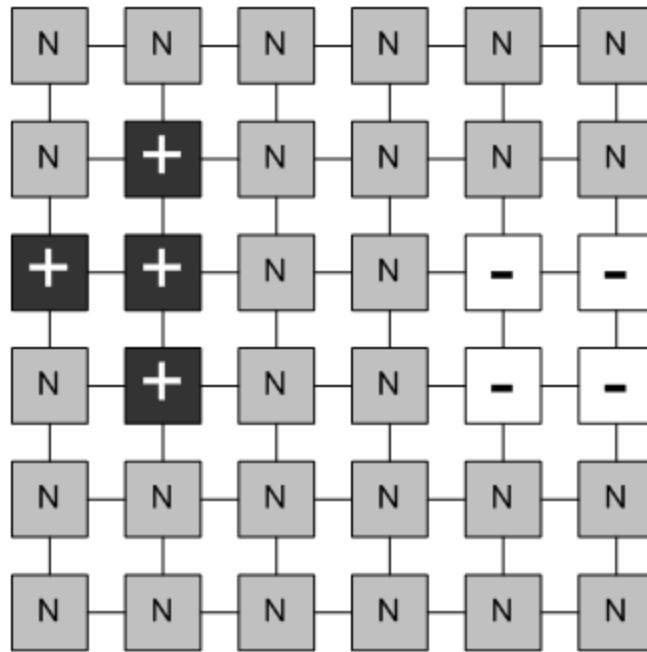


Figure 2. Large NoC with Hot and Cold zones

This dissertation proposes PowerAntz; an ant behavior inspired distributed power sharing scheme in Networks on Chip. PowerAntz attempts to provide improved power budget utilization by allowing power budgets sharing among routers those are beyond neighbors while keeping the distribution overhead to its minimum. This technique utilizes the power sharing history captured in pheromone values to distribute surplus power in the future. The major finding described in this section is a power budget distribution scheme as follows:

1. Efficient: up to 21% improvement in utilization of power sharing in non uniform power consumption scenarios when compared to an existing scheme.
2. Lightweight: Power budget distribution overhead varies from zero to 5% in the best case to the worst case scenario.

3. Scalable: Scheme overhead remains almost constant with varying network size.

2.1. Related Research

The scope of this work is restricted to peak power management, more specifically addressing the issue of power budget distribution. PowerAntz is about methodology to efficiently distribute the power budget. Hence, only power management techniques that have some form of budget sharing technique built into it are discussed here.

Shang et al proposed PowerHerd [10] [13], which handles the power management of a network on chip by sharing power budget among immediate neighbors. PowerHerd shares power budget using explicit control mechanism present among neighboring routers. The design of power herd does not allow power budget sharing between non neighbor nodes.

Kim et al proposed PC [12], which does local power management but no sharing of power budget. They have shown improvement in performance in terms of latency compared to PowerHerd. Bhojwani et al proposed SAPP [8], a non-deterministic peak power management technique that uses immediate neighborhood power consumption information to allocate power budget and periodically adjusts power budget to prevent over allocation. PowerAntz differs from SAPP in the way it shares power budget information. PowerAntz utilizes budget information beyond the immediate neighborhood.

Daneshtalab et al used AntNet [14] approach for power aware routing but they neither addressed power distribution involved among the components on chip nor dealt with explicit power management.

We have restricted the performance comparisons of PowerAntz with PowerHerd and PC due to architecture similarity and hence the results may be considered as more relevant. Since PC does not do dynamic power sharing and budget distribution we will do a network performance comparison with PC. However, we will compare the power sharing performance and budget utilization with that of PowerHerd.

2.2. The Ant System

Ant System was proposed by Dorigo et al and has been used by others as a solution strategy for hard problems [15]. Numerous problems have been identified to be solvable efficiently and easily using this idea. Ant System is inspired by the natural phenomenon of ants finding efficient routes to food sources from their habitats using passive information sharing through modification of the environment. This is called Stigmergy.

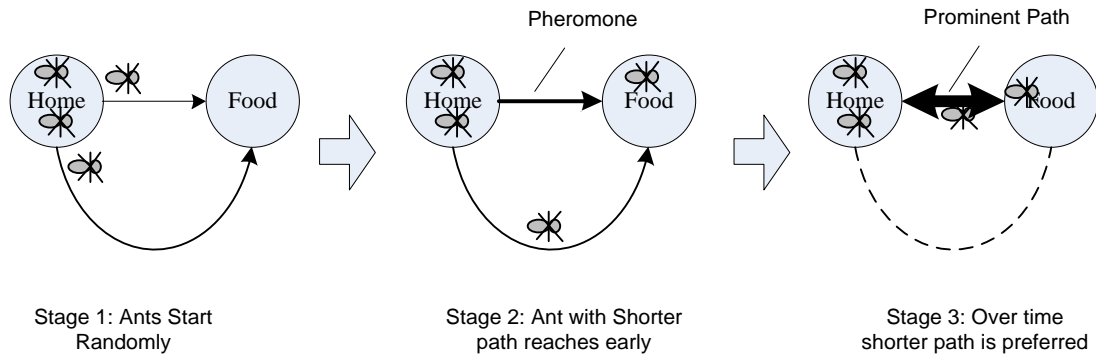


Figure 3: The Ant System

Figure 3 illustrates a simple ant system with a food source and an ant home.

There are two possible paths from Home to Food and back. Initially each ant chooses any of the two paths randomly with equal probability. When an ant takes a path it leaves a trail of pheromone on that path. The strength/concentration of the pheromone trail left on all path decreases gradually. Now as shown in the illustration some paths are shorter in length than others. Consequently the ant taking shorter path will reach **Food** first. While going towards food the ant reinforces the path it has taken by leaving a pheromone trail. Now to come back, the shorter path is already reinforced. So this ant prefers this path and takes it to return and in the process leaves more pheromone. This process repeats and gradually the shorter path is followed by more ants and the longer path is avoided. This is the main idea of Ant System. Mathematically it can be expressed as follows.

Let S_1 and S_2 be two states and there are n links between them. τ_i is the probability of taking link i for going from one state to another. Also let k ants are there in any of the state. The pheromone update operation is defined as,

$$\tau_i(t + \Delta_t) = \tau_i(t) * (1 - \rho) + \sum f_k$$

Where f_k is the reinforcement to τ_i due to k^{th} ant selecting link i between t & $t + \Delta_t$, and ρ is the evaporation of τ_i between t & $t + \Delta_t$. The evaporation rate ρ enables the system adapt to changing situations by making sure once reinforced edges do not remain reinforced forever.

2.3. Power Distribution Technique

We present an approach to power distribution in Network on Chip systems following the Ant System in general. The idea is to make the power distribution system autonomous and adaptive.

2.3.1. Ant System for Power Management

We define two kinds of ants in the system, namely, *Power Ants* and *Beggar Ants*. Power Ants originate from cold zone routers which have surplus power budget to share. Beggar Ants originate from hot zone routers that are consuming their entire budget and require more. The idea behind having two kinds of ants is to facilitate the communication from cold zones to hot zone and vice versa. Both kinds of ants leave pheromone in each router they come across. While making decision regarding movement of the ants these pheromone values are used to determine the best path. Since it's useful for Power Ants to go in a direction from which Beggar Ants are coming and vice versa, in the proposed system the ants update the pheromone level for the other type of ants. That is, on receipt of beggar ants from one link the pheromone for power ants

corresponding to that link is reinforced. Similarly power ants reinforce beggar ants' pheromone values.

We extend the reinforcement function found in general ant system to incorporate the effects of 1. Distance from origin of ant and 2. Share/Demand of power the ant is carrying. We define the power reinforcing function as,

$$f_k^p = C_p * \alpha_p * (1 - h_k/T)^2 * \beta_p * (\delta_k^-/P)$$

Where C_p is the baseline power reinforcement, h_k is hop count and δ_k^- is the demand of power carried by k^{th} beggar ant coming in from a particular link, T is the time to live of all ants and P is the allocated budget for each router/node. Similarly, the reinforcement function for beggar ants is defined as,

$$f_k^b = C_b * \alpha_b * (1 - h_k/T)^2 * \beta_b * (\delta_k^+/P)$$

Where C_b is the baseline beggar reinforcement, h_k is hop count and δ_k^+ is the share of power carried by k^{th} power ant coming in from a particular link. T and P have the same meaning as power reinforcement function. $\alpha_p, \beta_p, \alpha_b$ and β_b are constants controlling the influence of the two factors in the power and beggar reinforcement functions respectively. Notice that, $(1-hk/T)$ decreases as the ant hops from router to router, thereby reducing its contribution to pheromone updates. This is done because, information from longer hop distance away is both spatially and temporally outdated. So in general, power budget information from nearby routers gets higher weight while calculating pheromone values.

To summarize the combined system the pheromone update operations for beggar and power ants are defined as,

$$\tau_i^+(t + \Delta_t) = \tau_i^+(t) * (1 - \rho) + \sum f_k^p$$

$$\tau_i^-(t + \Delta_t) = \tau_i^-(t) * (1 - \rho) + \sum f_k^b$$

Where, τ_i^+ and τ_i^- are pheromone values corresponding to link i for power and beggar ants respectively. Other symbols have meaning as explained before. The Δ_t is the pheromone update interval.

2.3.2. Distributed Power Sharing Scheme

The distributed nature of the algorithm allows operation without any global knowledge. Ants carry out the job of transporting budget information from one region to another. The power budget distribution involves the following phases.

2.3.2.1. System Initialization

On startup the routers are initialized with power and beggar pheromone strength zero for all links. As both types of ants pass through a router the pheromone values are updated according to the pheromone update functions described above in Section 2.3.1.

2.3.2.2. Ant Generation

Ant Generation refers to the task of creating ants to indicate power surplus or need information. Depending on power budget status at a particular router an ant is

generated. Thus a power ant is generated if the router has surplus and it has received beggar ant from somewhere else. Figure 4 illustrates the Ant generation process.

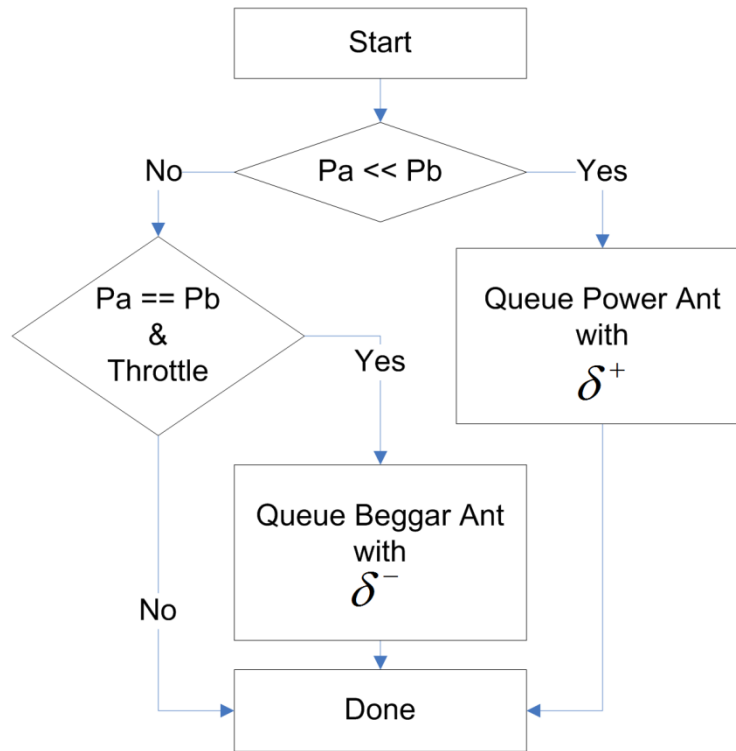


Figure 4: The Ant Generation Process

Let P_a and P_b denote the actual power consumption and the allocated power budget respectively. δ^+ and δ^- are calculated as follows. $\delta_i^+ = (P_b - P_a)/k, i = 1...k$ and $\delta_i^- = (P_b)/\eta k, i = 1...k$. Here, k is a random number between 1 and n where n is the number of links from the current router. η is a constant representing how aggressively a starving router asks for power budget share. A threshold of surplus budget is maintained

so that small fluctuations do not affect the system. Once generated, the Ants are enqueued for forwarding and it follows the normal forwarding process.

2.3.2.3. Ant Propagation

The Ant propagation refers to the routing of ants through the router. It is guided by the corresponding pheromones. An Ant is forwarded to a link based on the relative strength of pheromone on the links. In the routing stage the pheromone values of all the links are looked up. A power ant is forwarded to the link with highest τ_i^+ and similarly, a beggar ant is forwarded to the link with the highest τ_i^- .

2.3.2.4. Ant Consumption

Ant Consumption refers to the end of an ant's existence in the system. When an ant is received in a router, it is handled based on its type and the supply or demand information it is carrying on. When an ant is received, the router updates the appropriate pheromone level for the link the ant came from. Then depending on the routers current power budget level, it consumes the ant if necessary. The process has been illustrated in Figure 5.

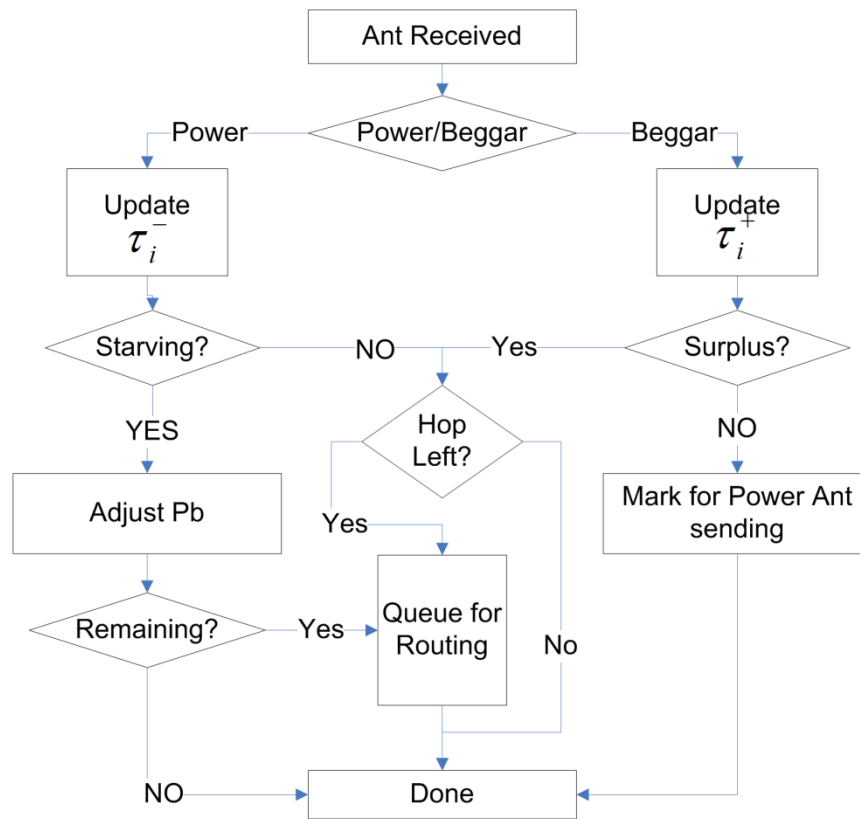


Figure 5: The Ant Consumption Process

2.3.3. Power Management Technique

PowerAntz can be used together with any power management scheme as long as the scheme is able to limit power consumption to a set budget. For example DVFS or packet throttling can be used in the routers to keep power consumption within the allocated budget. In the experiments described in this paper throttling was used to limit power consumption. When the allocated power budget gets exhausted, the routers are put in idle mode and no packet transfer happens until the power consumption goes under the power budget.

2.4. Evaluation Framework

2.4.1. Experimental Setup

To verify the impact of PowerAntz technique, a flit accurate Network-on-Chip simulator was designed for experiments and to compare it with two available power management schemes.

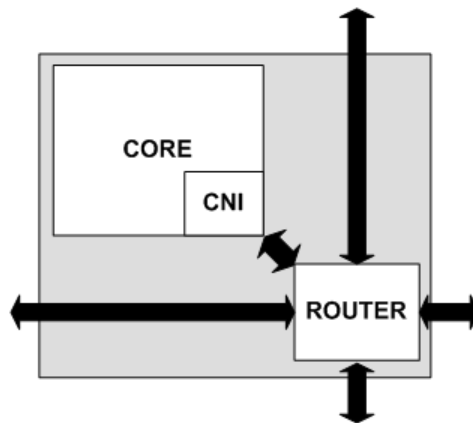


Figure 6: A Network Tile

2.4.1.1. NoC Architecture

PowerAntz scheme was tested using the NoC architecture developed by Bhojwani et al [4]. The network comprises of a collection of connected tiles. Each tile consists of a processing core, router and core network interface. Each router is connected to the core through Core Network Interface (CNI) as illustrated above (Figure 6). The

routers can be arranged and connected in different ways to form different network topologies; e.g. 2D mesh as shown in Figure 2.

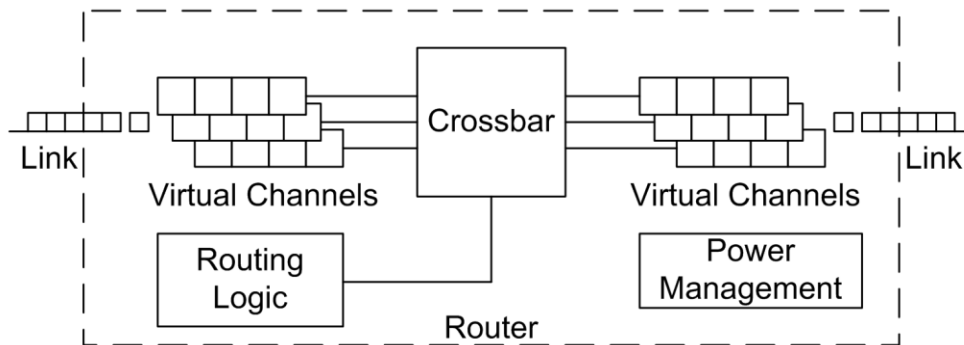


Figure 7: Router Architecture

The router is a standard virtual channel router as shown in Figure 7. These routers can be connected to create topologies such as 2-D Torus, Mesh, Fat-Tree, Ring etc. For comparison purpose, the experiments were restricted to 2-D Torus topology. In our experiments we control the activity (hence power consumption) of routers by limiting flit injection to it. And each router was configured for a given flit injection rate in the simulation setup. We will use this capability to create high or low activity zones in the NoC to evaluate the sharing capability of our scheme. It is important to manage the power consumption in the routers because it accounts for a significant percentage of the total SoC power.

An ant is implemented as a modified control flit in the system. It has the following fields. *Type*, *hop* and *share* in addition to the regular *flit* data structure. Figure 8 illustrates the fields in an ant packet. Here ant type can be either power ant or beggar

ant. Hop count tells how long the ant has been around. An ant is generated with hop count zero and every router increments it when receiving. A router decides to kill an ant when its hop count reaches the set maximum. The third field contains the power share/demand information (δ) as discussed in section 2.3.1.

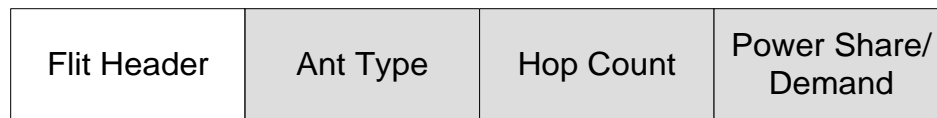


Figure 8: Structure of Ant Packet

2.4.1.2. Power Model

An event based power model was considered. Typical energy consumptions for different events in the routing cycle are adopted from literatures [12] [16]. Table 1 shows the energy consumption of the different steps involved in packet routing through a virtual channel router. To estimate the power consumption we count the number of such events happening in the router and using the energy values we obtained power.

Table 1: Energy Consumption of Router Components

Component	Operation	Energy (pJ)
Buffer	Read	76.41
Buffer	Write	76.62
Routing Logic	Routing	310.00
Crossbar Switch	Traversal	83.00
Link	Write/Read (per bit)	5.52

2.4.1.3. Simulator

A novel, configurable and robust simulation framework, *NoCSim 2* has been developed and was used for evaluation of the proposed scheme. NoCSim 2 is a SystemC based flit-accurate simulator capable of simulating generic NoC architectures. The key features of NoCSim2 are flexibility, speed and accuracy. NoCSim2 allows easy evaluation of different power management schemes' effectiveness on different network architectures. Figure 9 shows that NoCSim 2 consists of the following main components: XML Configuration Parser, Network Component Library and Network Synthesis Engine. This design allows the simulator to support arbitrary network topologies and easy configuration of network components and power management parameters. The modular design is well suited for different levels of abstraction across components according to the need for details. Simulating components of less interest at a higher abstraction level results in significant improvement in simulation time.

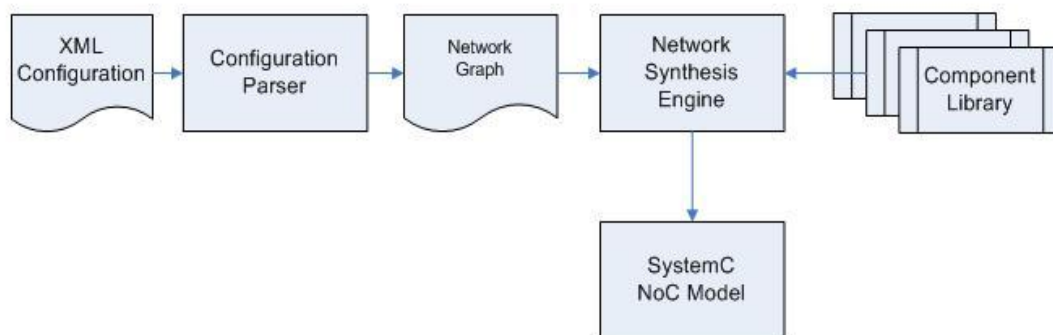


Figure 9: NoCSim Architecture

Currently the Network Component Library consists of a virtual channel router, a Core-Network interface (CNI), and dummy IPCores that generate OCPIP requests. Table 2 describes the feature set of NoCSim2 with respect to measurable metrics and configurable parameters.

Table 2: Simulator Feature Set

Measurable metrics	Configurable parameters
Message round trip times Latencies Virtual channel utilizations Link utilizations Buffer utilizations CNI response times Power dissipation at routers	Topology: Torus, Ring, Mesh Power Model: PowerHerd, PC, PowerAntz Buffer Sizes in routers and CNIs Number of ports in router

As previously mentioned, NocSim 2 features easy configuration of network components through an XML based configuration file. Figure 10 shows a configuration file for a network with two Processing Elements connected through a single router.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Simulator SYSTEM "config.dtd">
<Simulator cycles="20000" maxresponsetime="100000">
  <PowerModel tgpb="10000" pgbp="100" ebuffread="2" ebuffwrite="3" ecrossbar="10" elink="25"/>
  <Network nvcs="8" buflen="8">
    <Node type="Router" name="router1">
      <Router type="powerantz" ports="2" infifolen="8" outfifolen="8"/>
    </Node>
    <Node type="PElement" name="pe1">
      <PElement type="ocp" infifolen="8" outfifolen="8">
        <CNI msgqlen="4" reorder="false" type="ocp"></CNI>
        <CORE></CORE>
      </PElement>
    </Node>
    <Node type="PElement" name="pe2">
      <PElement type="ocp" infifolen="8" outfifolen="8">
        <CNI msgqlen="4" reorder="false" type="ocp"></CNI>
        <CORE></CORE>
      </PElement>
    </Node>
    <Link src="pe1" dst="router1"/>
    <Link src="pe2" dst="router1"/>
  </Network>
</Simulator>

```

Figure 10: A Sample Simulator Configuration

For the particular configuration as shown in Figure 10, the simulation will run for 20000 cycles. Additionally, the parameters for the energy consumption model described in Table 1 can be seen in the Power Model attribute of the NoC simulation. The simulator has also been used for related research with NoC and has produced interesting results [17].

2.4.1.4. Benchmarks

We used both Synthetic and Real benchmarks to evaluate the effectiveness of the PowerAntz approach. Synthetic traffic generators with constant bit rate traffic were used to simulate various utilization scenarios. In addition to the synthetic cases, we also used a multimedia SoC with 16 cores to simulate more realistic test case.

2.4.2. Evaluation Criteria

The power distribution scheme proposed in this paper was evaluated for the following criteria.

- *Performance*: Performance is measured in terms of throughput and end-to-end latency for a given load. Lower latency is better while higher throughput indicated better performance.
- *Utilization*: Utilization is measured by the ratio of power consumption to power budget allocated. Higher utilization is represents a better power distribution scheme.
- *Overhead*: Overhead is measured by additional flits processed in the system due to the power management task. A uniform overhead across network sizes indicates scalability.
- *Reactivity*: How fast the power distribution technique reacts to a change in power management scenario. This is measured by the time taken for a router with budget deficit starting from sending a beggar ant until receiving a power ant. For the scheme to be scalable, it is desirable that this response time does not increase linearly with increase in network size.

2.4.3. Test Scenarios

The proposed scheme is evaluated for the criteria discussed above in the following power consumption scenarios simulated during the experiments.

Non-Uniform Power Consumption: In this scenario power consumption varies from component to component. High and low power consumption zones may be isolated by neutral zones (e.g., Figure 2). This is simulated by limiting the flit injection rates in different routers to different levels.

Low Power Consumption: In this scenario all routers consume power less than their allocated power budget. This is created by low activity system and power does not become a constraint in this scenario. This scenario represents low activity modes like sleep / low power modes in devices.

High Power Consumption: In this scenario every router consumes the power budget allocated to it leaving no surplus. In this case, due to unavailability of surplus power budget improvement due to sharing is not possible. This scenario represents high workload. An example of such system would be a multicore data processor in a web server.

The scenario where the power consumption is random is not considered for simulation/evaluation because of the following reasons. (a) It presents a poor case to evaluate the power redistribution scheme, because in a random power distribution a router with power budget deficit is likely to have a neighbor with surplus hence does not make a compelling case to evaluate the power distribution scheme proposed here and, (b) Any realistic SoC is unlikely to have a random power distribution because the power

intensive components like cache blocks, processor cores etc. tend to co locate and result in clustered power consumption zones.

Media Benchmark: The media benchmark was constructed using a 16 core Mesh setup with MPEG encoding and decoding applications. More detail is provided in the results section.

2.5. Results

2.5.1. Performance Comparison

To compare performance of the three schemes, end to end latency of packet delivery was measured while varying the average load on the system. Load is varied by changing the flit injection rate at each router. The data reported here is average of all the routers in use. Time for a packet to reach the destination core from the source core in cycles, was used to measure the end to end latency. More information about the experimental parameters are given in Table 3.

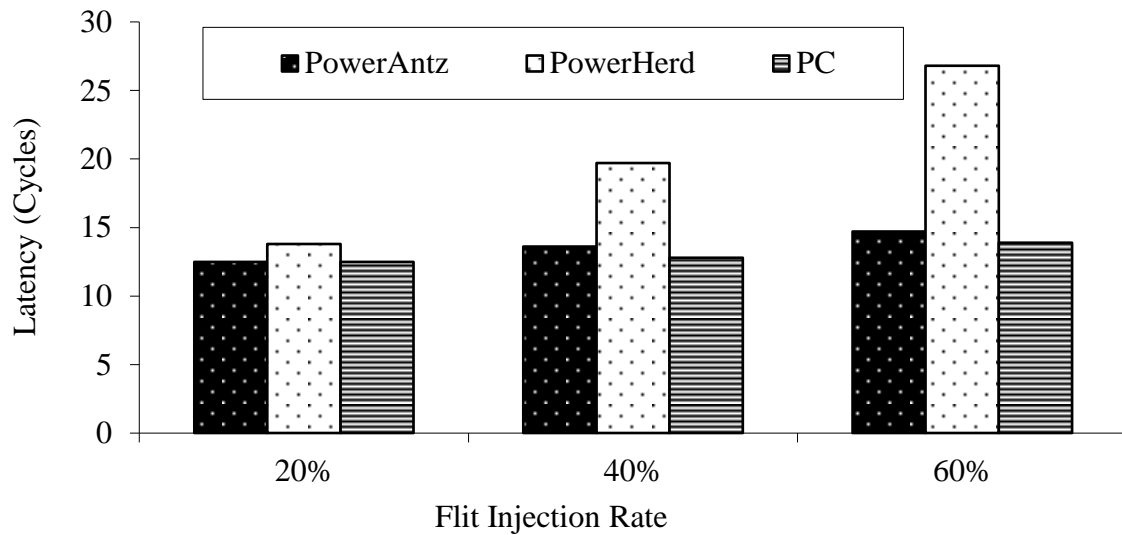


Figure 11: Performance Comparison (Latency)

Table 3: Experiment Details

Experimental Parameter	Value
Network Size	4x4
Topology	2D Torus
Metric Used	End to End Latency
Simulation Duration	1 Million Cycles

Figure 11 shows the comparison of end to end latency using PowerAntz, PowerHerd and PC for varying flit injection rates. The result shows that PowerAntz provides consistent latency across 20% - 60% flit injection rate.

Table 4: Power Budget Sharing Experiment Details

Router Zones	Case1 FIRs	Case 2 FIRs	Case 3 FIRs
Neutral	40%	40%	40%
Cold	10%	20%	40%
Hot	100%	100%	100%

The power budget sharing experiment is designed as follows. The network size of 4x4-2D torus is considered with arrangement of routers as in Figure 2. To create a Non-uniform power consumption scenario (as described Section 2.4.3), two, six and eight routers are configured as Hot, Neutral and Cold zone routers respectively. Three different power consumption cases are considered within this scenario. In all three cases the power budget is set to limit the flit injection rate of all routers to 40%. For Case 1, we set the limit of injection rate in two routers (Hot Zone Routers) to 100% and the neighboring routers are limited to 40%. All other routers (Cold Zone Routers) are limited to 10% flit injection rate. For Case 2 and 3 we increase the flit injection rate in the cold zone routers to 20% and 40% respectively. This is done to reduce the spare budget in the cold zone routers. Flit injection rates corresponding to different routers in the three cases are given in Table 4. The hot zone routers were chosen for comparison because they are the most dependent on the power management scheme's ability to redistribute the available power budget.

Figure 12 shows comparison of throughput of hot zone routers with PowerHerd and PowerAntz in the three cases described above. In Case 1, with PowerAntz, the hot zone routers' throughput increases 30% compared to PowerHerd.

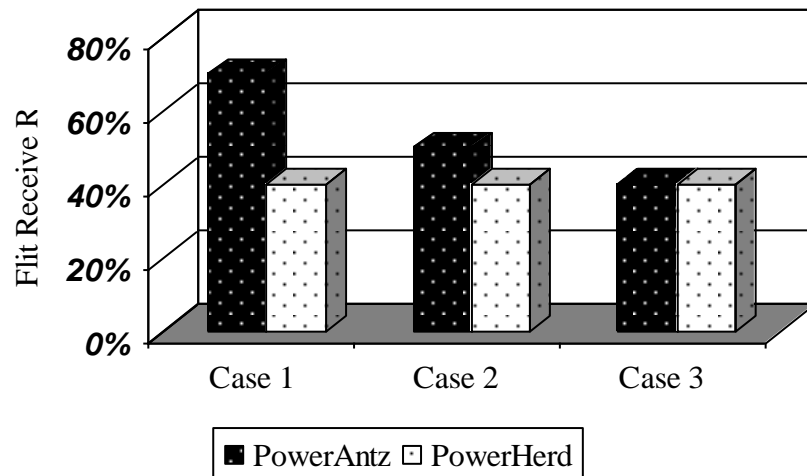


Figure 12: Comparison of Budget Sharing

In case of PowerHerd, the hot zone router is unable to receive surplus power budget from the cold zone routers because of the neutral routers isolating them. Even with PowerAntz, the throughput of the hot zone router has reduced in Case 2 and further decreased in Case 3 to similar level as PowerHerd. This is because in Case 3 the cold zone routers are spending all the allocated budget leaving no surplus.

2.5.2. Utilization Comparison

We used the same experiment as in Section 2.5.1 to measure the power budget utilization using PowerAntz and PowerHerd schemes. Table 5 shows the comparison of

power budget utilization between PowerAntz and PowerHerd. PowerAntz performs better than PowerHerd when there is large variation in the router activities throughout the NoC. In other words, the power budget utilization is high when more surplus power is available in the cold zone routers for sharing. The maximum improvement in power budget utilization is observed when cold zone routers are at 0% FIR, i.e. no activity.

Table 5: Improvement in Power Budget Utilization

Cold Zone FIR	0% FIR	10% FIR	20% FIR	40% FIR
Utilization using PowerHerd	.25	0.3125	0.375	0.5
Utilization using PowerAntz	.303	0.3594	0.39	0.5
Improvement	21.25%	15%	4.17%	0%

2.5.3. Overhead Analysis

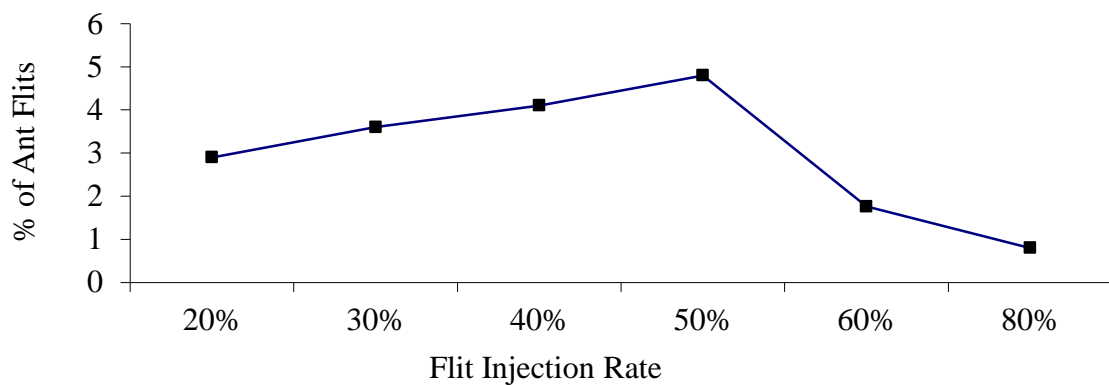


Figure 13: Overhead with Varying FIR

To estimate the ant overhead, a 4x4 2D Torus network was simulated using varying flit injection rate uniform random traffic. Figure 13 shows the overhead of PowerAntz scheme with varying flit injection rate in a 4x4 torus NoC. It is measured in terms of additional ant flits processed due to the PowerAntz scheme. It reveals an interesting characteristic of PowerAntz technique.

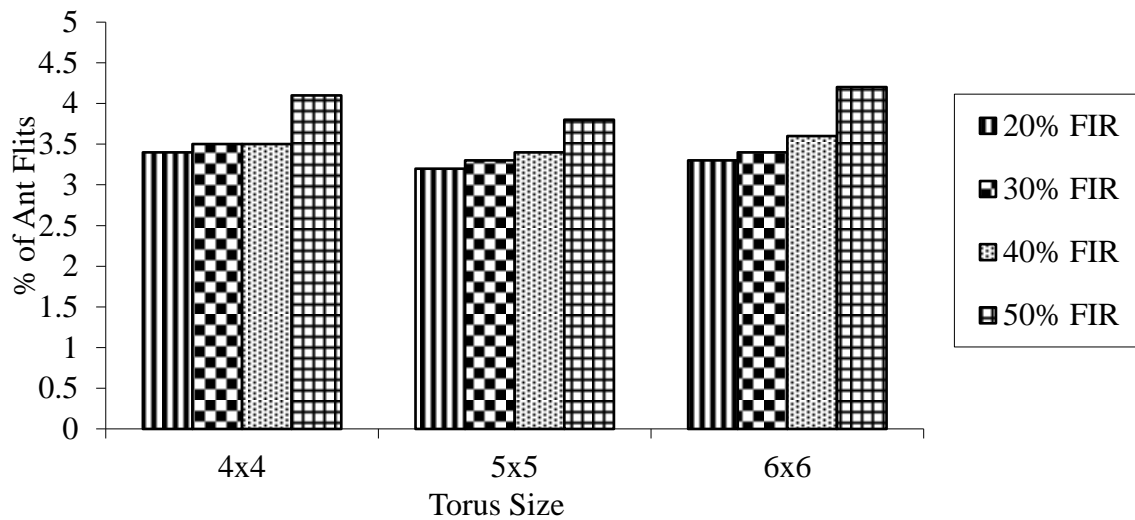


Figure 14: Overhead with Varying Network Size and Flit Injection Rate

The overhead due to PowerAntz scheme reduces towards higher flit injection rates. At high flit injection rate, all routers consume their allocated budget, so less power ants are generated. Figure 14 shows the variation of power management overhead with varied network size and flit injection rate. The overhead remains less than 5% across different network sizes showing the scalability of the technique.

The processing overhead of PowerAntz is small. The link pheromone update function can be computed when the ants are received in the same stage of routing by a simple add/update hardware in one cycle. In pipelined router architecture this cycle can be easily shared with the routing stage and hence will not cause any increase in packet latency.

2.5.4. Reactivity Analysis

The speed of the power re-distribution has been measured by the time taken from sending out a beggar ant to receipt of the first power ant. Four network sizes were simulated using the average injection loads 20, 30 and 40% using random destinations. Figure 15 shows how average response time changed with varying network sizes. The response time slightly increased with increased network. Two important characteristics can be inferred from this experimental result. First is that, the response time is fast enough to allow meaningful budget sharing. The below 20 cycle response time is much smaller than the thermal time constant of a typical chip of that size [10]. Second important observation is that the response time stabilizes and does not increase for networks larger than 6x6. This is because the response time is dependent on the minimum size of the neighborhood that has enough surpluses to meet the power demand. However, with smaller network size the diameter of the network is low hence resulting in a lower response time. In addition to this, with larger network each router is subjected to more traffic and also delays the response.

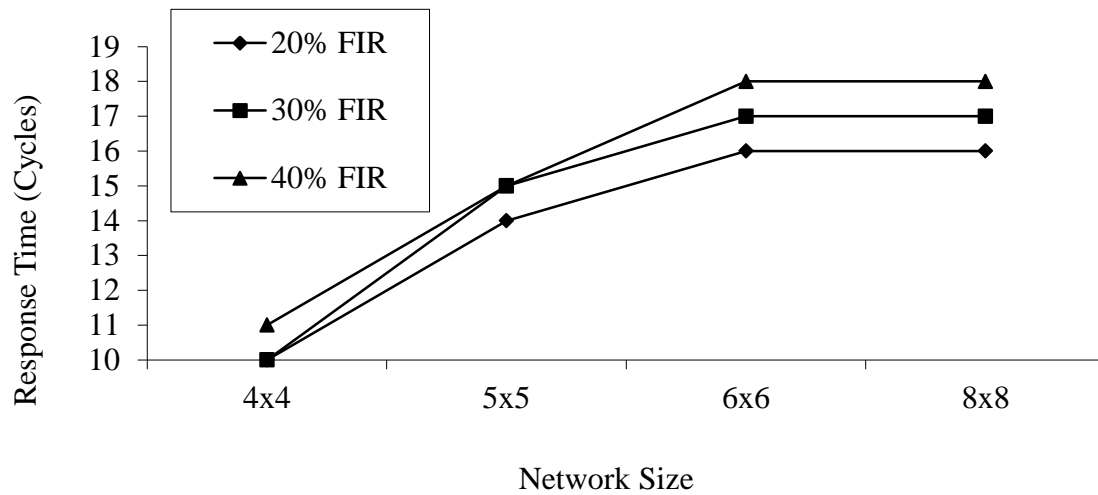


Figure 15: Response Time with Varying FIR for Three Load Scenarios.

2.5.5. Real Benchmark Evaluation

In addition to the synthetic analysis, we evaluated PowerAntz with an experimental SoC configuration (Figure 16) similar to the one used in [11]. The communication volume established in [11] has been used to create a realistic network activity.

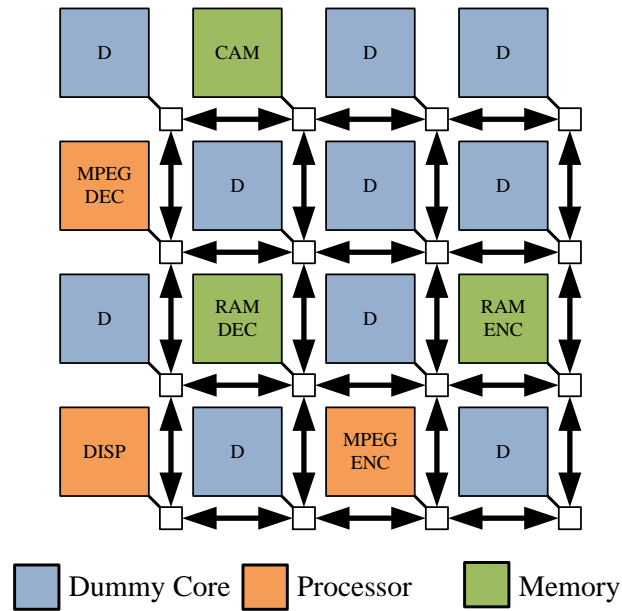


Figure 16: Media SoC Setup

Table 6: Real Benchmark comparison of PowerHerd and PowerAntz

Scheme	Average Packet Latency	Peak Throughput	Budget Violation
PowerAntz	22 Cycles	50%	None
PowerHerd	30 Cycles	45%	None

We used 3 processor cores, 3 memory cores and 10 dummy cores setup to mimic a typical media SoC. The dummy cores are programmed to generate enough traffic to deplete their adjacent routers power budget. The processor cores are programmed to generate the transactions based on the function that is implemented in the corresponding tile. The processor cores will result in most power consumption in the adjacent routers while memories will generally remain idle except for burst transfer to L1 cache within

the processors. This leads to the non-uniform power consumption scenario described in the synthetic cases earlier. We compared PowerAntz scheme with PowerHerd in terms of latency, throughput and power budget satisfaction in Table 6.

2.6. Optimum Power Sharing and PowerAntz

The power sharing problem addressed by PowerAntz technique can also be thought of/formulated as a variant of an optimization problem. Depending on the system's design goals the objective of such a problem could be one of the follows:

1. **Maximizing Utilization:** Here utilization is the fraction of total power budget consumed that produces output after accounting for any efficiency loss introduced by the sharing scheme.

2. **Minimizing Power/Thermal Gradient:** Thermal gradient may be defined as the difference in temperature between neighboring nodes / distance between the nodes.

PowerAntz falls into the first category here. It is designed to enable maximum utilization at the expense of increased variation of power consumption in different parts of the chip. However, with operating system or higher level support PowerAntz can also be used to realize the second power sharing goal. In that case, instead of sharing the budget, workload portions can be shared between busy and idle regions using similar concept. In such a design, the ants can still carry power budget surplus/deficit information. However, upon consumption the receiving node will trigger a task migration to the idle node instead of increasing its own activity level. Interestingly, this approach also has the potential of increasing power budget utilization by engaging idle nodes to pick up excess workload from the busy nodes. This mechanism of course is

dependent on the agility of the operating system in terms of reassigning workload dynamically at a fine granularity.

This alternative approach by itself demand separate detailed investigation and thus has been kept outside the scope of this dissertation.

2.7. Conclusion and Analysis

Efficient power budget distribution and sharing in a complex Network-on-Chip is challenging. PowerAntz provides a robust and scalable technique for power distribution on a NoC system using proven technique of ant system intelligence. Through experiments, PowerAntz has been found to provide up to 30% improvement in throughput of the starving routers while improving the overall budget utilization up to 21.25% compared to PowerHerd. The overhead of PowerAntz was well within 5% across different network sizes from 4x4 to 6x6.

The distributed nature of the PowerAntz technique makes it suitable for large scale budget distribution mechanism. A hierarchical extension of PowerAntz could be utilized to manage datacenter power budget. Our future plan is to evaluate PowerAntz in power distribution starting with the multi core chips extending it upwards to set of chips on a board, boards on a rack and so on. The same will call for extension of the protocol, redefinition of ants and their activities.

3. POWER BUDGET SATISFACTION

As discussed in the previous section, it is important to have effective management technique to make sure the NoC power does not increase beyond what the chip can handle. This problem is referred to as peak power budget satisfaction problem and generally the solutions have relied on communication throttling as a means of limiting power consumption [10].

Closer observation of the throttling mechanism reveals an important drawback while applied in wormhole switched routers. To illustrate a sample situation is provided in Figure 17. In the example the router R1 reaches its power budget limit and is throttled to satisfy the power budget. Due to the nature of wormhole switching, this can result in packets in transmission to get trapped. The trapped packets can potentially more than 1 channel in the routing paths and prevent other packets from being routed efficiently. A clogged packet of n flits can in the worst case span n routers and hence occupy $O(n)$ channels depending on channel allocation policy. This in turn hampers the performance of the neighboring routers and results in overall decreased performance. The performance impact has been shown in the experimental results section.

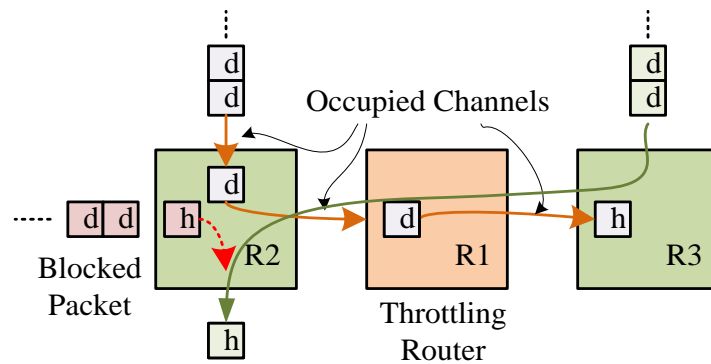


Figure 17: Problem of Router Throttling

To mitigate the performance degradation, this dissertation presents a novel flow control technique which avoids throttle as described before. The flow control mechanism works to flush the congested router before it finally throttles it. The proposed method resembles the early notification scheme of TCP congestion control [18]. In addition, to conserve power further an adaptive throttle mechanism is designed to enable a lower power throttle mode during long throttle periods.

The findings discussed in this section are summarized as follows:

4. An adaptive throttle mechanism for power consumption control without affecting performance
5. A novel flow control mechanism to minimize latency while maintaining power budget
6. Experimental evaluation of the techniques to demonstrate the improvement in latency, throughput and power envelope.

The rest of the section is organized as follows. First related works in the literature are discussed followed by the router architecture along with the adaptive throttle

mechanism and the power aware flow control technique. The experimental setup, evaluation criteria and results are described thereafter.

3.1. Related Research

Since the introduction of NoC, power management has gained significant attention among researchers [19] [16] [11] [20]. For direct relevance, we will discuss the work of Shang et al in detail. Shang et al proposed a power sharing architecture for NoCs [10]. In their approach they proposed the concept of router throttling to limit power consumption of the router when its power budget limit is reached. In this scheme, when the power consumption estimate exceeds the set limit, the router pipeline is halted by stopping the crossbar switch. It is a simple and effective way to limit router power consumption. However, when not used in conjunction with the correct flow control, it can result in packet lockup and significant performance degradation.

Other relevant approach to limit power consumption in the NoC is to control the admission of packets at the entry points, i.e. throttling at CNIs [8]. This method can simplify enforcing algorithms. However, admission control severely limits such schemes' capability to react to local power situations. With large and heterogeneous system compositions this limitation makes them inadequate at best.

3.2. Flow Control and Adaptive Throttle Mechanism

3.2.1. Router Architecture

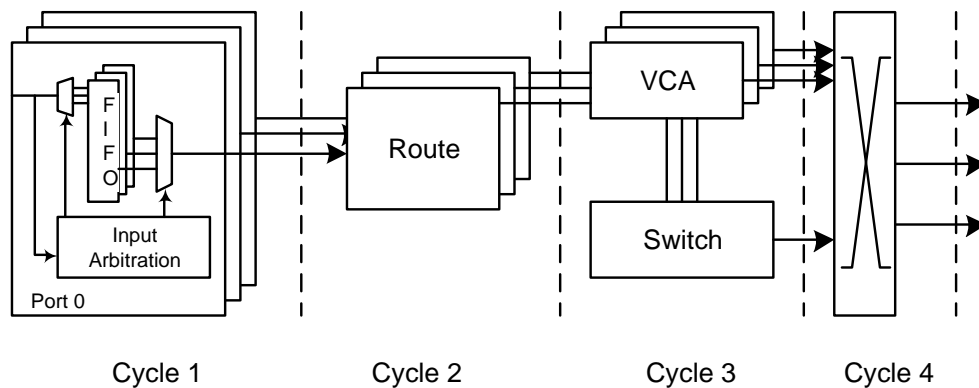


Figure 18: Pipelined Router Architecture

A 4 stage pipelined router architecture was used with input buffering and a hybrid switch allocation-virtual channel allocation policy. The router pipeline and major structures are illustrated in Figure 18. 64 Bit wide flits are used. The virtual channel FIFOs are 8 flits deep and 8 virtual channels per input port are used. The proposed flow control requires that the FIFO length be at least the size of a packet while there can be more FIFOs than virtual channels. To free up the channel after a complete packet is received in the router a the switch allocation is modified to select from a set of arrived packets instead of input virtual channels. This enables freeing channels by letting blocked packet wait in the buffer. Randomized shortest path routing is used using routing table.

3.2.2. Router Power Model

For power estimation the detail router power model Orion 2.0 is used [21]. Power consumption of individual router component is estimated using the Orion model. The dynamic power is estimated by an event based energy model where the energy consumption of each event is estimated using the Orion model and tabulated. The power estimation login in the router uses the tabulated values to estimate the energy consumption in the current power window.

3.2.3. Flow Control with Early Notification

A novel flow control approach inspired by the early notification technique used in TCP congestion control is used to improve performance while throttling traffic to maintain power consumption level. The flow control states are illustrated in Figure 19. During operation each router can be in one of the four states described below.

Begin: In this state all operations proceed normally and no control is enforced.

Notify: When estimated power consumption exceeds the P_{notify} (notify threshold), the router enters this state. In this state, neighbors are given a notify signal and the router stops receiving new packets. However, flits from already accepted packets are received and processed. Similarly, the router stops sending new packets but continues to send flits from already transmitting packet. This is the key state in the performance improvement. The notify state enables the router about to go into throttle mode to clear any occupied channels before going into throttle.

Throttle: The router enters this state when estimated power consumption exceeds $P_{throttle}$ (Throttle Threshold) but it is in Burst Mode. In this state the router stops receiving any flit. However, it is powered on and clocked. This is a temporary suspension state.

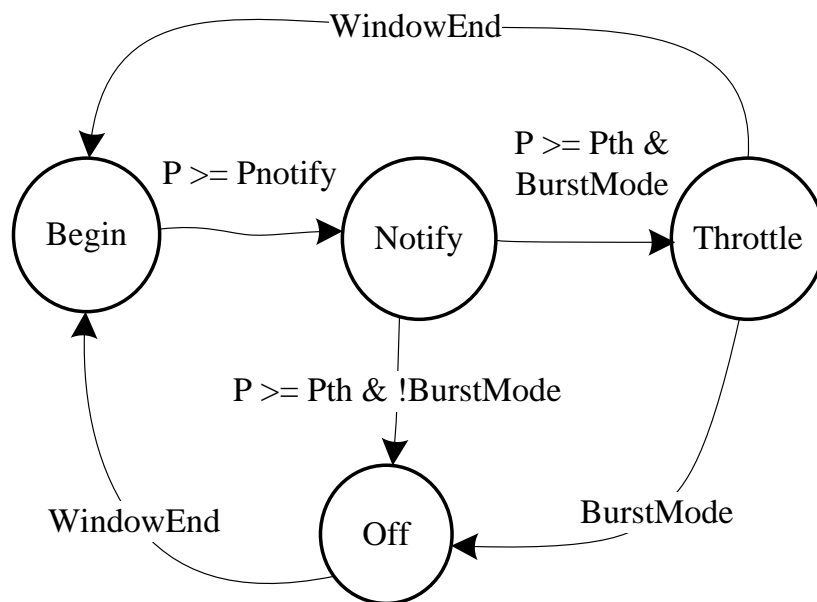


Figure 19: Flow Control State Machine

Off: The router enters this state if estimated power exceeds throttle threshold while in Non Burst Mode. In this state everything except the buffers are switched off using power gating. More details on the Throttle and Off modes are described next.

3.2.4. Adaptive Throttle Mechanism

The throttle mechanism used in PowerHerd and similar power management techniques control power consumption by disabling switch to reduce activity of the

router [10]. Although this technique can bring the overall power consumption to a low level it does not eliminate all activities. As a result, even in throttle mode power is consumed and a rapid burst leading to long throttle can actually result in a violation of power budget.

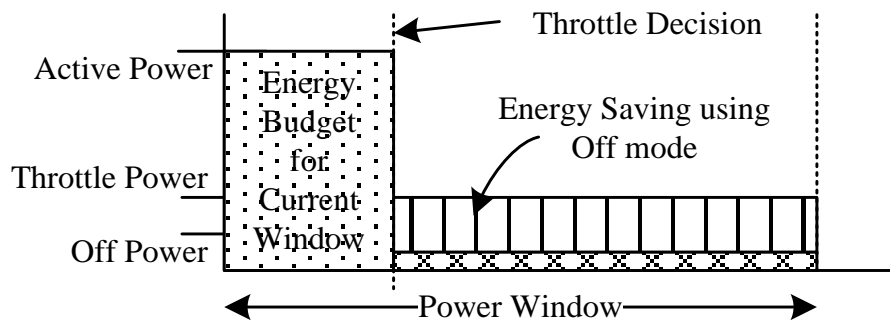


Figure 20: Energy Calculation for Throttle Threshold Determination

One possible solution to this is to turn off the entire router instead of just disabling the switch. This can be done either by clock gating the router device or even more efficient way will be to use power gating techniques to turn of the router completely. However, turning off the router has a downside. It takes some time to bring back up any circuit which is turned off using power gate. A novel design is proposed to achieve the best of both approaches. Figure 19 illustrates the proposed state machine. Here P is the estimated power consumption in the current window. P_{notify} is the threshold for notification P_{th} is threshold for throttling. In this design, two throttle modes are proposed, regular throttle mode (Throttle in Figure 19) where the router is clock gated and a deep throttle mode (Off in Figure 19) where router blocks are power

gated. The decision to move into these states is made based on the flow control state machine and mode determination logic.

The system can be in one of two operating modes: Burst mode and non Burst mode. The general idea is that; if the router will need to get back to normal mode very soon, it is better to be in regular throttle mode. However, if throttle is reached early in the power estimation window, the router will need to be in throttle mode for longer and deep throttle mode is preferred.

The mode selector is implemented as a threshold comparator in the router architecture. Figure 20 illustrates the threshold computation. A timer keeps track of the current position in power estimation window. If the position is beyond a dynamic threshold BurstMode is set. The threshold is updated based on current occupancy of the router according to the following equation.

$$T_{Threshold} = T_{win} - \left(\frac{Power\ Gate\ Energy\ Overhead}{Throttle\ Power - Gated\ Power} \right)$$

Notice that only the gated power (Off mode power) is dynamic and varies based on the state of the router. This is calculated offline by low level device models and is tabulated for dynamic access.

3.3. Experimental Evaluation

We evaluated the proposed scheme with a series of experiments. The proposed flow control technique was compared with the throttle mechanism proposed in PowerHerd [10]. The simulation platform and the experimental setups are discussed in detail below.

3.3.1. The Simulation Platform

We used NoCSim, a flit accurate NoC simulator capable of running SparcV8 processors with integrated cache and memory model to simulate the proposed router design. The simulator is written in SystemC for flexible modeling detail and simulation speed. The proposed router was implemented at RTL level while other components were simulated at transaction and functional level. To simulate real benchmark programs, the ArchC Instruction Set Simulator [22] was integrated with the NoC simulation framework. This simulation system allows of full SoC simulation with a software system kernel. Both synthetic traffic and real application communication can be simulated. For all the experiments performed a 4x4 mesh network was used with 1 core per router tile. CBR and Locally Random synthetic traffic was used while an MPEG ENC/DEC system mapping [11] was used for realistic traffic simulation.

3.3.2. Evaluation Criteria

1. **Latency:** Latency is given by the time it takes for a packet to travel from source to destination.
2. **Throughput:** Throughput was measured by number of injected flits per node per cycle.
3. **Power Budget Violation:** Fraction of thermal cycles where the power consumption of a router exceeded the budget.

3.3.3. Results

We classified the experimental results into the following sections to clearly illustrate the benefits. First we show the performance benefits followed by improvement

in power performance characteristics. All the comparative results are based on PowerHerd scheme discussed in the related research section.

3.3.3.1. Performance

To measure the performance of the proposed flow control end to end latency and throughput was measured using random traffic on a 4x4 mesh network by applying varied injection load. The results are presented in Figure 21 below. The proposed flow control scheme resulted in stable and reduced packet latency across the injection loads.

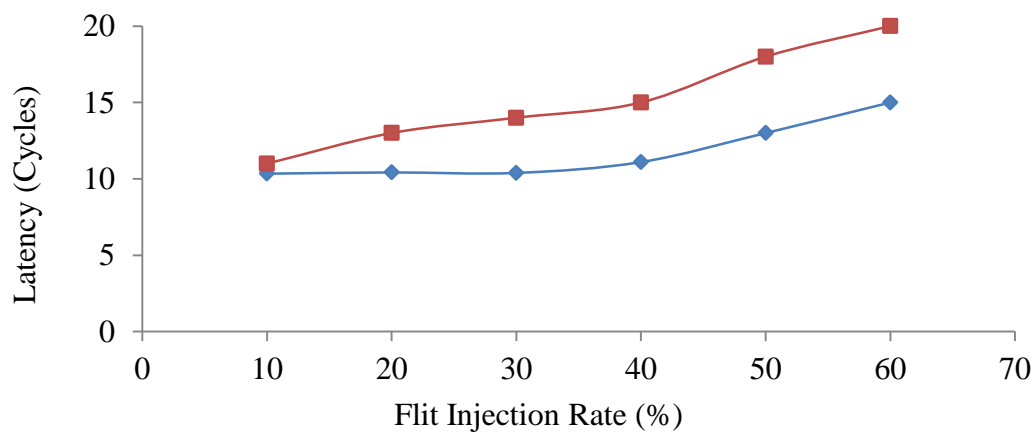


Figure 21: Latency vs. Flit Injection Rate for Random Traffic

Along with latency the effective throughput and efficiency of the network is also measured. Figure 22 shows how efficiency and throughput change with increased load on the network. The efficiency is high at low network load (89% at 10% Injection) and slightly falls (to about 75% at 60% Injection) at higher injections.

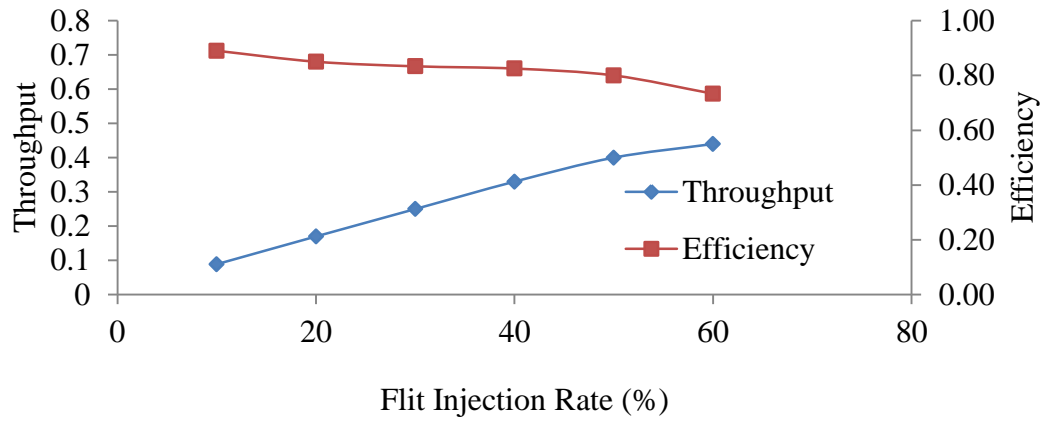


Figure 22: Throughput and Efficiency vs. Injection

3.3.3.2. Effect of Adaptive Throttling

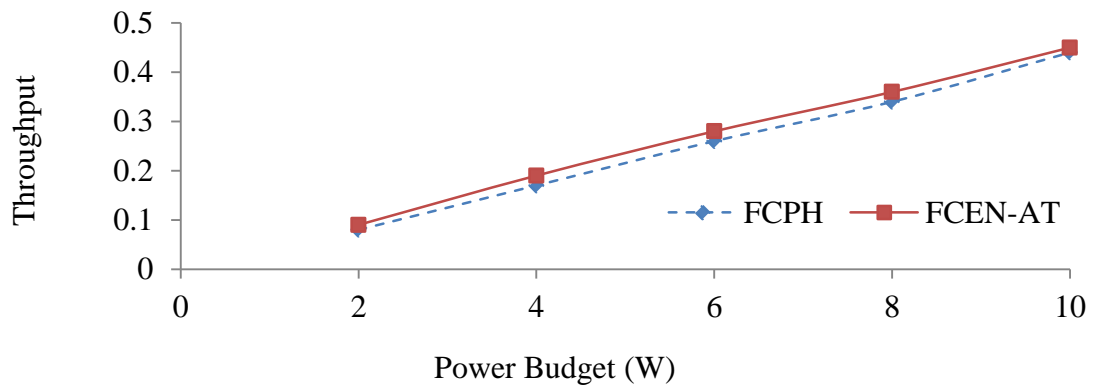


Figure 23: Throughput at Different Power Budget using Adaptive Throttling

Adaptive throttling decreases the throttle mode energy consumption hence enabling more flit transaction with a given power budget. To illustrate this, throughput was measured while changing power budget and keeping packet injection constant.

Figure 23 shows the proposed adaptive throttle mechanism to achieve better throughput at all power budget levels.

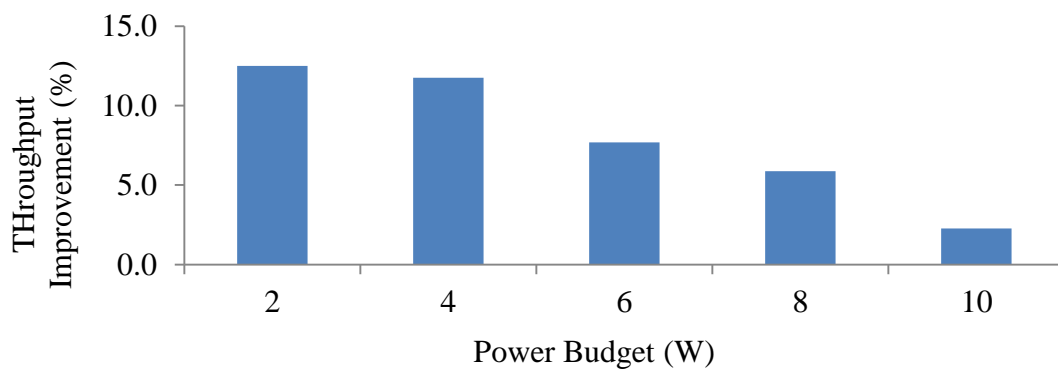


Figure 24: Improvement in Throughput using Adaptive Throttling

The experiments with adaptive throttling's effect on throughput also showed that the improvement in throughput is highest when the system is severely constrained in power budget availability. The result is shown in Figure 24.

3.3.3.3. Design Overhead

The addition of the adaptive flow control logic adds additional processing and hence creates overhead. To evaluate the added cost of this technique the design was synthesized using 90nm TSMC library. The area/power overhead results are summarized in Table 7.

Table 7: Area and Power Overheads

Component	Area (Gate Eq.)	Leakage Power (nW)	Dynamic Power (nW)
State Machine	1243	2.98	345
Threshold	74	0.14	23.73
Router	16087	5830	630×10^3

3.4. Summary

In this section we addressed the problem of packet lockup caused by simplistic router throttle mechanism. A novel power aware flow control technique is proposed to satisfy power envelop while maintaining high performance in terms of throughput and latency. Up to 4X reduction in latency were achieved compared to throttling scheme in PowerHerd. In addition, adaptive throttling technique reduced the tolerance required for power budget satisfaction by 12.5%.

4. POWER EFFICIENT MICROARCHITECTURE

Buffer size and allocation policy play an important role in the performance and efficiency of a NoC router [23] [20]. Furthermore, studies have shown that buffers can consume as much as up to 79% of NoC router power [19]. Thus efficient management is necessary to ensure high performance and low power. Efficient schemes use SRAM arrays for their simplicity and high performance [24].

Nanoscale SRAM buffers are very suitable for NoC router design because of their speed, density and reliability. Power dissipation characteristics of Nanoscale SRAMs are unique. Traditional low power design techniques are not sufficient to ensure minimum power operation. To that end, a dynamic power management technique specifically designed for Nanoscale SRAM buffers is necessary. Such management technique will make use of buffer allocation information and the Nanoscale SRAM power dissipation characteristics to minimize both static and dynamic power consumption while maintaining performance.

It is notable that the buffer utilization in NoC router is dependent on network congestion. Depending on the application communication pattern a given routers buffer utilization will vary over time. To provision for high utilization case it is necessary to provide enough buffer in each router. However, often the buffers are not utilized and remain idle and consume power. To avoid this, we propose dynamic block level buffer power management. To be able to benefit from this scheme it is necessary to use a central buffering strategy in the router. An example design is described in [24]. We propose a buffer design where the new flits are buffered in sets. Each set can hold some

number of flits and is powered by a single source which can be turned on or off using a power gate. Hence depending on usage, the buffers can be turned on/off set by set. The number of active sets required to ensure zero performance hit is determined using a feedback controller.

Another observation about Nanoscale SRAM cells is that, storage of 0 and 1 are significantly different in terms of power consumption. This characteristic is exploited at per flit level to minimize power consumption during storage and read write. This is done by selectively inverting the flits based on their zero density. The decision to invert or not is taken using a simple adaptive controller.

The contributions described in this section are as follows:

1. A Feedback Controlled Block level Buffer Management is proposed for dynamic power management
2. An Adaptive controller for efficient Flit level Power management is proposed
3. Both power management techniques are thoroughly evaluated for performance and energy efficiency and showed to outperform static allocation by 21% increase in throughput and 20% reduction in energy consumption.

4.1. Related Research

Both circuit level and system level techniques have been proposed for NoC power management. There have been significant research works on router buffer power management for low power. Detail discussions on existing designs have been done in [25] [19] [20].

Zhang et al. have proposed a centralized buffer management to achieve enhanced buffer utilization [24]. Their scheme demonstrated a 50% decrease in total buffer requirement in their router. However, they did not provide an active power management strategy which can further reduce dynamic power. The proposed power management technique explores this possibility in central buffer router design to achieve superior power/performance characteristics.

Wang et al. have proposed a zero-efficient design for router buffers that optimizes the circuit level design of router buffer to minimize energy consumption [26]. The basis of their work has been the predominance of zeros in the NoC traffic. This is primarily a circuit level work under the assumption of high zero density and does not necessarily fare well when there is majority of one. Also they do not consider any system level information or active power management technique to adapt to the dynamic nature of the traffic. The proposed scheme differs from this in the way buffers are allocated dynamically and also the way flits are encoded while storage.

4.2. Preliminaries

4.2.1. NoC Flow Analysis

Synthetic traffic has been widely used to evaluate NoC architectures. However, an analysis of actual application communication patterns results in interesting observations that can facilitate advanced management techniques for low static and dynamic power operation. Using a NoC-based, full SoC simulator, a series of application

benchmarks were executed and the traffic flow in the NoC was analyzed. The experiments and observations are discussed below.

4.2.1.1. Experimental Setup

We use a full system simulation environment that models a flit accurate NoC to analyze the traffic flows while running real benchmarks. Further details on the simulation framework can be found in Section 4.6.1. We set up three sets of experiment to analyze the traffic pattern in a typical NoC based SoC.

A 16 node ring network was simulated with 8 processors, 4 memory modules and 4 producer consumers mimicking other devices.

2D Mesh is the most popular NoC topology in literature due to its simplicity and regular nature. The Mesh was also configured with same 8 processors, 4 memories and 4 producer consumer cores.

Similar to 2D mesh is 2D torus. Over 2D mesh, 2D torus has the benefit of a lower network diameter but suffers from increased link count. The core assignment is the same as in 2D mesh case.

The applications were taken from the benchmark Suite Mediabench, Mibench and SPEC2006 [27] [28] . Selected applications from these benchmarks were mapped on to the above mentioned topologies and experiments were performed.

4.2.1.2. Flow Characterization

From the experimental results common NoC traffic flow was characterized as illustrated below. The results showed that majority of NoC utilization was seen in the

burst mode transfers. Figure 25 shows the distribution of the different flow groups in terms of burst lengths. The benchmark traffic was dominated by the longer bursts [4 packets or more].

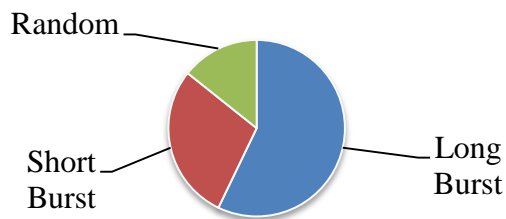


Figure 25: Distribution of Traffic Types for MediaBench Applications

4.2.1.3. Buffer Occupancy Analysis

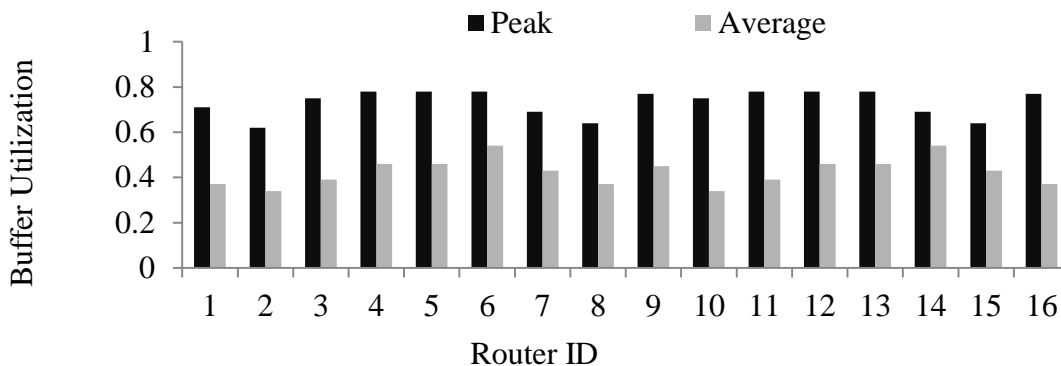


Figure 26: Buffer Utilization Distribution in a Mesh Topology

Figure 26 shows the buffer occupancy analysis for the mesh topology described earlier. This result clearly demonstrates that even though the average utilization of the

buffer space is low, all routers have experienced high peaks of buffer utilization. Due to this nature, a straight forward static buffer allocation based on the average utilization will cause performance degradation [See Section 4.6.2.2].

To mitigate this problem and optimize the buffer utilization a dynamic technique is necessary. This is the key motivation behind the proposed feedback controlled dynamic buffer management.

4.2.2. Nanoscale CMOS Buffer Design

In addition to efficient management technique, inherent efficiency in the buffer design itself plays an important role in the overall energy efficiency of the buffer. Ideally, the design of buffers has to be low power, fast and reliable. Between the alternatives of buffer design, such as SRAM, embedded DRAM [29] and Registers, SRAM is advantageous for speed and power [30]. A novel failure tolerant nanoscale CMOS SRAM based buffer design is adopted. The unique characteristics that make it particularly suitable for NoC router buffer design are analyzed in the rest of the section.

4.2.2.1. Nanoscale CMOS SRAM Buffer

Single-ended SRAMs are known for their tremendous potential of low power dissipation. A seven transistor SRAM cell is shown in Figure 27(A) [31]. The SRAM cell is composed of a read and write access transistor (transistor 1), two inverters (transistors 2, 3, 4 and 5) connected back to back in a closed loop fashion in order to store the 1-bit information and a transmission gate (transistors 6 and 7). The transmission gate opens the feedback connection between inverters during the write

operation. The cell operates on a single bit-line, instead of having two bit-lines as in standard six transistor SRAM cell. Both reading and writing operations are performed over the single bit-line. The word-line (WL) is asserted high prior to write and read operation as similar to standard six transistor SRAM cell. When the cell is in a hold mode, the word-line is low and a strong feedback is provided to the cross-coupled inverters by the transmission.

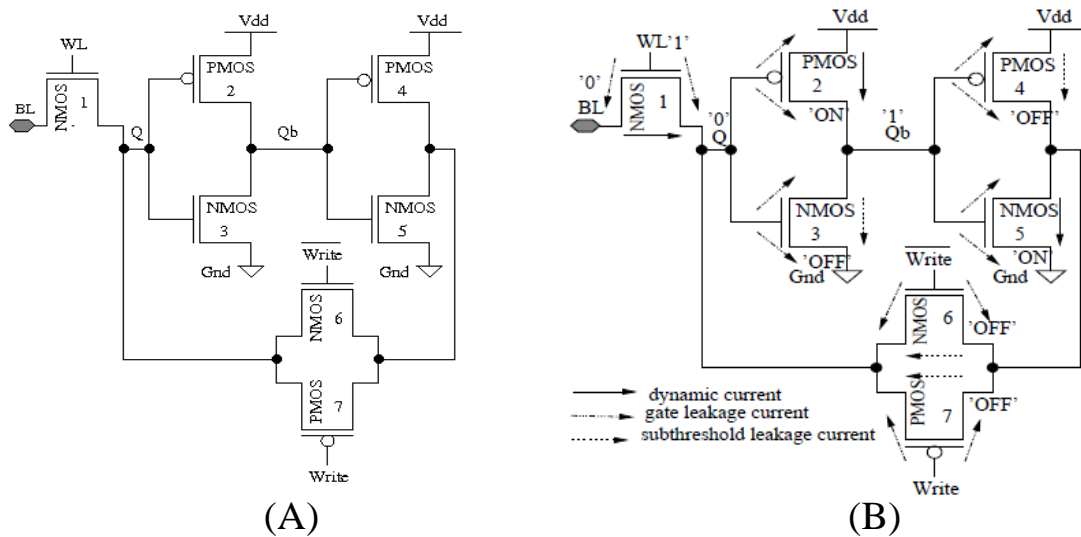


Figure 27: Structure and Operation of 7T SRAM

The total power dissipation of a CMOS based SRAM circuit for sub-65nm technology node is defined as the summation of dynamic power dissipation, subthreshold leakage, and gate-oxide leakage. The SRAM cells have a tendency to retain data for some duration of time as they cannot be shut off. The current flow (or power dissipation) in each device depends of the location the device in the SRAM circuit as

well the operation (e.g. read, write, or hold) being performed. Thus, for accurate measurement of current (power) it is important that the currents are identified. Figure 27(B) shows the current paths for 'write 0' operation. The SRAM cell is simulated at the 45nm CMOS technology node using Predictive Technology Model [32] for nominal sized transistors at a supply voltage of 0.7V. The simulation results are presented in Figure 28 for the above 7-transistor SRAM when designed using dual-threshold voltage technique for low-power dissipation.

4.2.2.2. Statistical Power Model

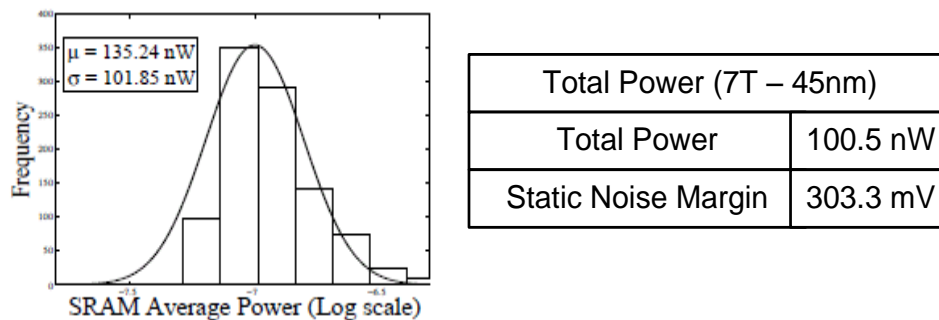


Figure 28: Total Power Dissipation of 7T SRAM

In nanoscale CMOS process variations is a major concern. The process variation has made the designers job much complicated due to loss of circuit yield with reduced time to market. We have selected twelve process parameters for statistical variability study: NMOS/PMOS channel length, NMOS/PMOS channel doping concentration, access-transistor length and width, driver-transistor length and width, load-transistor

length and width. Some of the parameters are independent and some are correlated which is taken into account during simulation for realistic study. Each of the process parameters is assumed to have a Gaussian distribution with mean (μ) taken as the nominal values specified in the PTM for 45nm node and standard deviation (σ) as 5% of the mean. The statistical process variation in parameters is translated to power, leakage, and Static Noise Margin using Monte Carlo simulations. Monte Carlo simulation is an efficient approach because it does not require relating the output to input which otherwise would have been cumbersome for the large number of parameters. For brevity, the statistical distributions of total power dissipation due to nanoscale process variations averaged over different operations are presented in Figure 28.

The static and dynamic power dissipation of the SRAM for different modes of operations is presented in Table 8. The probability density functions of all these are Gaussian in nature.

Table 8: Static and Dynamic Power Dissipation of SRAM.

Power	Operation	Mean (μ)	Standard Deviation (σ)
Gate Leakage	Write 1	21.2nW	9.4nW
	Write 0	21.9nW	9.5nW
	Read 1	12.9nW	5.4nW
	Read 0	7.8nW	3.2nW
	Store 1	2.8nW	1.8nW
	Store 0	1.0nW	0.5nW
Subthreshold Leakage	Write 1	38.2nW	21.1nW
	Write 0	7.8nW	19.0nW
	Read 1	12.3nW	27.0nW
	Read 0	13.5nW	32.1nW
	Store 1	10.8nW	21.0nW
	Store 0	16.2nW	2.3nW
Dynamic Power	Write 1	39.2nW	22.1nW
	Write 0	5.1nW	20.0nW
	Read 1	14.3nW	30.0nW
	Read 0	15.5nW	32.1nW
	Store 1	12.8nW	22.0nW
	Store 0	17.2nW	2.9nW

4.3. Router Architecture

The proposed buffer design is very suitable for routers with centralized buffer management. In this section we discuss the design of a single cycle centralized buffer router design.

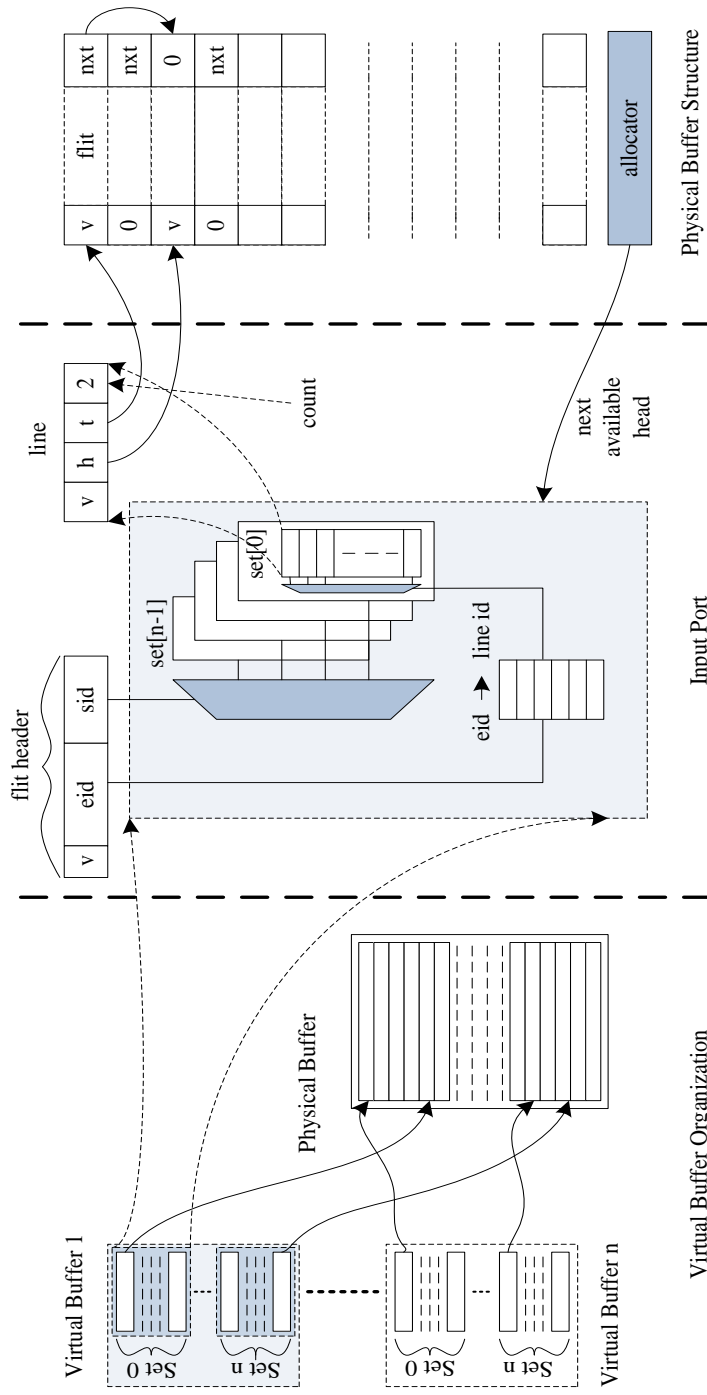


Figure 29: The Centralized Buffer Router Architecture

4.3.1. Virtual Buffer Architecture

Figure 29 illustrates the Router design. To effectively utilize the central buffer design a concept of virtual buffer is introduced. Every input port contains virtual buffer in which each valid entry points to a queue in the central physical buffer. Queue management is performed in the physical buffer design. Instead of allocating the buffer based on virtual channels, a concept of set and line is introduced. A set is a collection of lines allocated to packets going to a given output port of the router. So Set 0 in any input port will contain packets that intend to go to Output Port 0. This requires one step look-ahead routing. Each line queues packets that are going to the same destination. This property avoids head of the line blocking.

4.3.2. Central Physical Buffer Design

The virtual buffers allow independent management of the central buffer structure. The physical buffer is managed centrally and each virtual buffer may or may not be mapped to a physical buffer. To be able to effectively perform power management using power gating the buffer is grouped in blocks. Each block can be turned on and off using a power gating structure. This is desired compared to turning on/off each buffer element because of the power gating structure overhead. The available buffer index logic selects free buffer elements from the fullest buffer block which is not full. This leads to buffer blocks being utilized one by one. Hence allowing blocks to be turned off when no buffer element from that block is used. This can be implemented using a priority encoder based combinational logic block or a PLA.

4.3.3. Note on Performance of Centralized Buffer

The centralized buffering mechanism enables router operation with fewer stages. This effectively reduces the end to end latency that a flit experiences in addition to any buffering delay in general as long as the router can be still operated at the same frequency. While this was true for the proposed design at the 1GHz nominal operating frequency, it is certainly near the higher end of the maximum operating speed of the design. It is however possible to extend the design concept to deeper pipelined router architecture. Note that the centralized buffer based design is not inherently dependent on the router having a shallow pipeline. A deeper pipeline will in fact simplify the centralized buffer operation by reducing the number of ports. But, the added stages will contribute towards higher end to end latency without increasing operating frequency.

4.4. Buffer Power Management

The dynamic power management is motivated by the traffic flow analysis presented in Section 4.2.1.2. It was shown that bursts account for a large proportion of network traffic and also the bursts are in general restricted to a few network paths.

Taking advantage of these characteristics, a mixed mode feedback controller was designed to do buffer power management at block level. To further enhance the control on power consumption an adaptive controller for flit storage encoding management was introduced. We will discuss each level of power management in the following sections.

4.4.1. Block Level Power Management

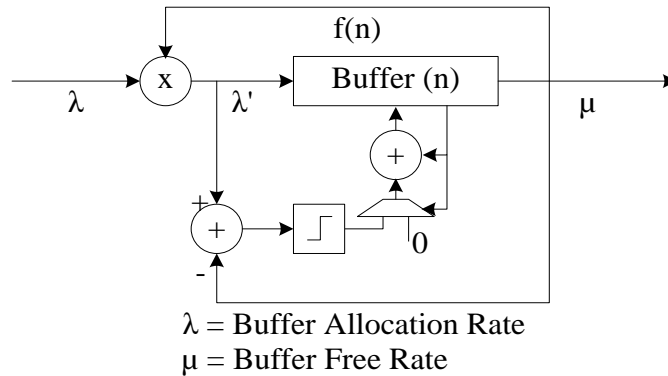


Figure 30: Block Level Feedback System

A non-linear feedback controller was designed for block level power management. Figure 30 shows the feedback system modeling. The observed traffic (λ') is represented by a function of the injection load (λ) and the backpressure (f) of the network which is in turn again a function of the available buffer space. The feedback function $f()$ is estimated from simulation and tabulated. This definition is used to calculate the minimum buffers required to maintain performance.

4.4.1.1. The Feedback Function

To predict the buffer utilization, a flow density metric is used. The flow density of a given set represents the likelihood of that set being occupied. The flow density of a set is given by number of flows * bandwidth of each flow / available buffer for that set.

4.4.1.2. Controller FSM

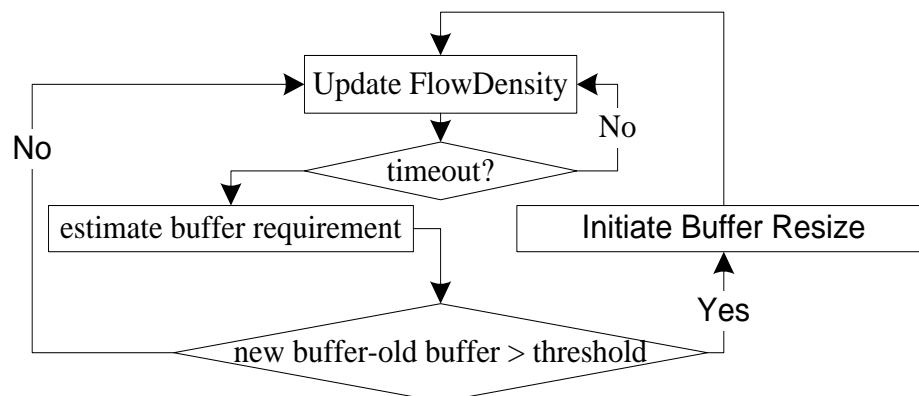


Figure 31: The Block Level Power Manager FSM

The block level power management is done by utilizing flow prediction and buffer power gating. Figure 31 shows the simple FSM used to do this operation. The timeout and the threshold of update are set based on the dynamic nature of the application setup. System level support is used here in predicting flow densities. Every packet is marked as Start Burst, End Burst or Random. Incoming Start/End of Burst packets increase the confidence of prediction based on the flow rate and the length of the burst, while Random packets reduce the prediction confidence. This is utilized to dynamically adjust the buffer switch threshold. The buffer resize is done in a slow mode not to cause power surges. The shown FSM sets a register with the new required buffer number. The power controller shuts down buffers one by one until the remaining buffer matches the required number. Same procedure is followed while increasing the buffers. Buffer blocks are turned on one by one.

4.4.2. Flit Level Power Management

In addition to the block level buffer management, a dynamic encoding technique is applied per flit to further enhance the energy efficiency. This is done by utilizing either positive or negative logic based on the zero density of the data in the flits.

4.4.2.1. Flit Storage States

Any flit can be stored in one of the three states: Active 0, Active 1 or Sleep. In Active 0 – true logic is stored as 0. In Active 1 – true logic is stored as 1. In sleep state data is not stored. Hence sleep is a non-preserving state. A linear adaptive control mechanism is designed to assign the flit storage states dynamically. Wrapper logic is added in the buffer design to make this process transparent to the rest of the system.

4.4.2.2. Adaptive Control

An adaptive control technique was developed for the flit level power management. The overall control operation is shown in Figure 32. The state is a simplified representation of the density of 1's in the flit.

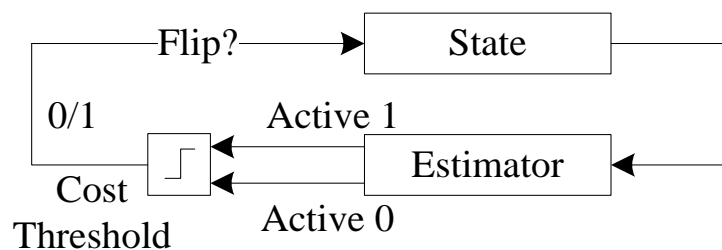


Figure 32: Adaptive Control for State Assignment

4.4.2.3. Estimator Design

For low overhead the estimator needs to be simple. In the proposed design the flits are marked to be '1-dense' by adding a bit to the header. This bit is set when the flit is created. A simple estimate is the frequency of this bit being set in a given time interval T . The corresponding estimator can be easily implemented using a saturation counter.

4.4.2.4. Controller FSM

The estimator described above makes the controller design very simple. The estimate is updated every time a new flit comes in. After every time interval T , this estimate is compared with a threshold. IF the estimate is higher than the threshold the flit is inverted when stored. This decision remains for the next T time. After which the condition is re-evaluated. Figure 33 depicts the operation in a flowchart.

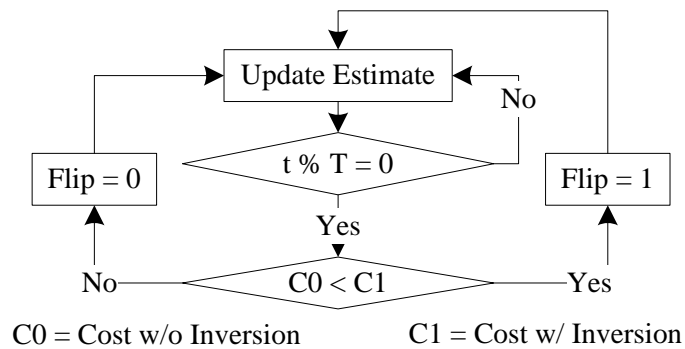


Figure 33: FSM for Dynamic Flit Inversion Controller

4.4.3. Dynamic Power Gating

Power Gating is a popular technique used to reduce leakage power of idle components by switching off the power supply from the component [33] [34] [35]. Power gating is utilized to enable the block level power management discussed in Section 4.4.1. One drawback of power gating is data loss. When the buffer block is power gated it can't retail the data stored in it. Closer experimentation reveals that this is not completely true especially in short timeframe [36]. Depending on the sleep mode bias applied to the block it can retail data for a certain period in time [37]. The data retention characteristics of the 7T SRAM cell used in our design is shown in Figure 34.

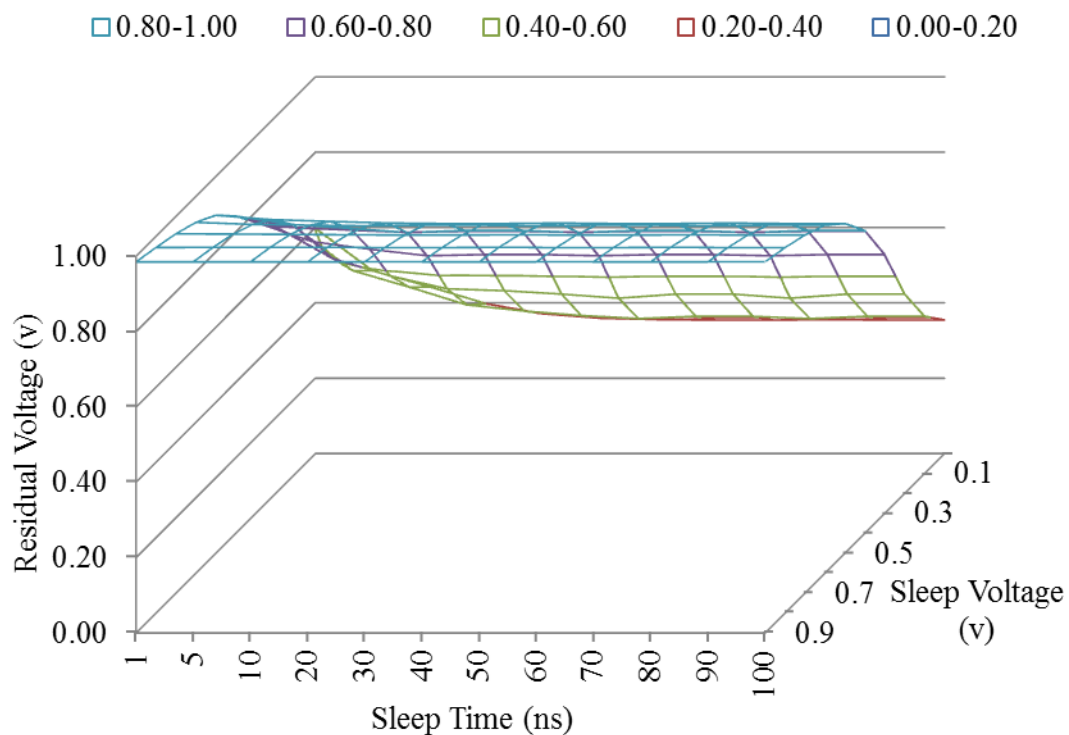


Figure 34: Data Retention Characteristic of the 7T SRAM

This result leads to an interesting idea that if we can predict the duration for which a flit will be sitting in a particular buffer block, it may be possible to drop the operating voltage of that flit buffer and in the process save some energy. We can use the same estimator used for flit encoding to make sleep entry exit decision. To dynamically switch the sleep state of the buffers we use the bias generation circuit proposed by Agarwal et al [33]. Aggressive power gating can lead to failure in buffers. However, techniques have been proposed in literature to address such errors [38].

4.5. Adaptive Link Control

In addition to the flit buffer, the inter router links consume a fair share of power in a typical network on chip. The high width and speed of the links make them even worse in terms of power usage. We proposed a bandwidth demand adaptive link controller to reduce link power consumption without sacrificing performance.

4.5.1. Underutilized Link

Inter routers links are designed to sustain the maximum theoretical capacity of the NoC in question. This often leads to power hungry link design. Although the link is designed to operate at a much higher bandwidth it is not uncommon for incoming links of a busy router to remain idle. This usually results from buffer fill-up or other congestion situation in the network. This presents an interesting opportunity to reduce energy consumption of the links by using an alternative link design while not requiring maximum bandwidth operation. Note that this is slightly different from frequency scaling. The energy savings come from reduced crosstalk between adjacent lines by

disconnecting alternate lines in the link. In addition, we will also use a low swing driver receiver to create an alternative low power link. The operating modes are illustrated below.

4.5.2. Link Modes

To exploit idle link for power reduction we propose four different link modes (Table 9) using a both bandwidth scaling and low swing driver receiver. Each pair of driver receiver can operate at two different widths. Width is selected by a control signal. The driver receiver pair is selected by another control signal. Those two control signal together form the two bit control word to select a link mode.

Table 9: Link Modes in Proposed Dynamic Link Control

Link Mode	Swing	Width
S0	Full	Full
S1	Low	Full
S2	Full	Half
S3	Low	Half

4.5.3. Link Mode Selection

The link mode is selected based on the buffer situation at the current router as well as the flow type being received from the upstream router connected by the link in question. Link up selection (raising performance level at power cost) is always initiated by upstream router when its buffer starts filling up. Similarly link down selection is always initiated by the downstream router when its buffer starts to fill up.

4.6. Experimental Results

The proposed dynamic buffer management technique was compared with a static buffer allocation. The evaluation consists of experiments to demonstrate performance in terms of latency and throughput, power efficiency and design overhead in terms of area and power. The simulation platform and the experimental results are discussed in the following section.

4.6.1. Simulation Platform

We used NoCSim [39], a flit accurate NoC simulator capable of running SparcV8 processors with integrated cache and memory model to simulate the proposed router design. The simulator is written in SystemC for flexible modeling detail and simulation speed. To simulate real benchmark programs, the ArchC Instruction Set Simulator [22] was integrated with the NoC simulation framework. This simulation system allows of full SoC simulation with a software system kernel. A 9 core mesh system with 5 processors at center and 4 memory cores in four corners was used to evaluate the proposed buffer management scheme.

4.6.2. Performance

The proposed buffer and link management schemes are designed to provide operating modes that affect both power and performance. The goal is to utilize the right combination of modes to minimize power without affecting performance. This if effectively done on an individual node basis, will result to better utilization of the total

system power budget. To demonstrate the effect on performance we present the results in terms of throughput and latency.

4.6.2.1. Latency

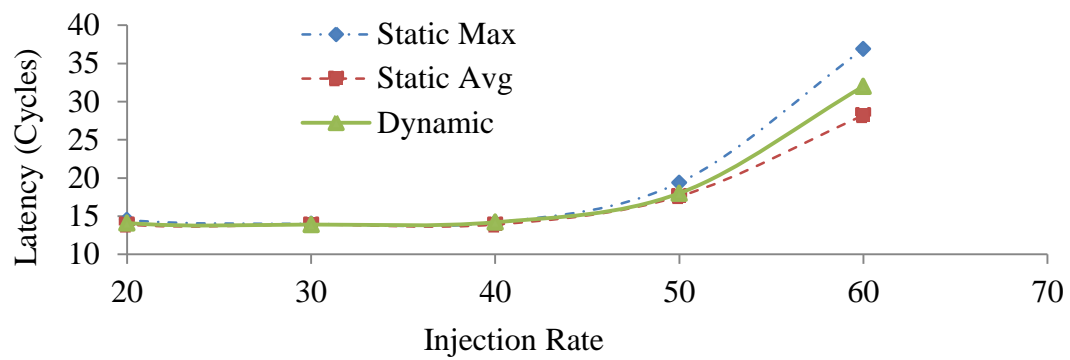


Figure 35: Latency vs. Injection Rate

To measure the effect of the proposed buffer management schemes on end to end packet latency, experiments were performed using statically allocated buffers based on average utilization, statically allocated buffer based on maximum buffer utilization and dynamically managed buffer with the proposed scheme. Figure 35 compares the three schemes based on end-to-end latency. The dynamically managed buffer allocation results in latency comparable to the average or max allocation case. Note that maximum buffer allocation increases the latency toward the higher injection rate. This happens due to increased contention in the router because more flits are buffered.

4.6.2.2. Throughput

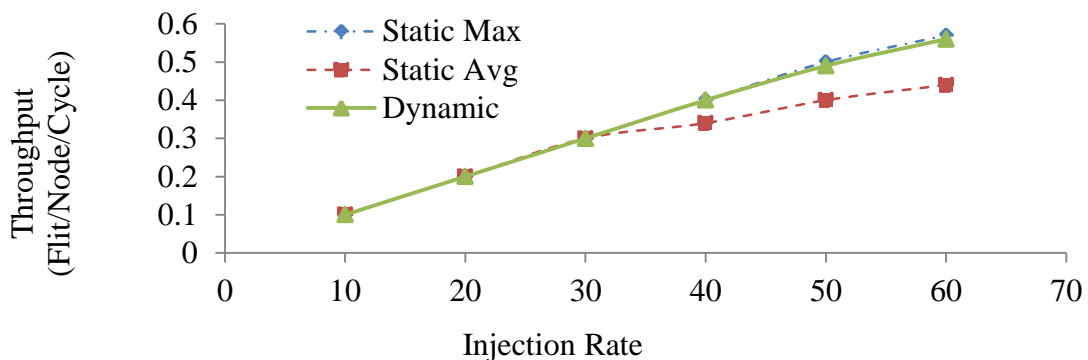


Figure 36: Throughput Comparison

Figure 36 compares the throughput achieved by the three schemes based on varied injection rate. The dynamic buffer allocation achieves virtually the same throughput as the maximum buffer allocation case across the range of injection rate. But the static allocation based on average buffer utilization takes major hit in terms of

throughput beyond 30% flit injection rate. And it deteriorates with higher injection rate.

At saturation (0.56) the throughput reduction is as much as 21%.

4.6.3. Power

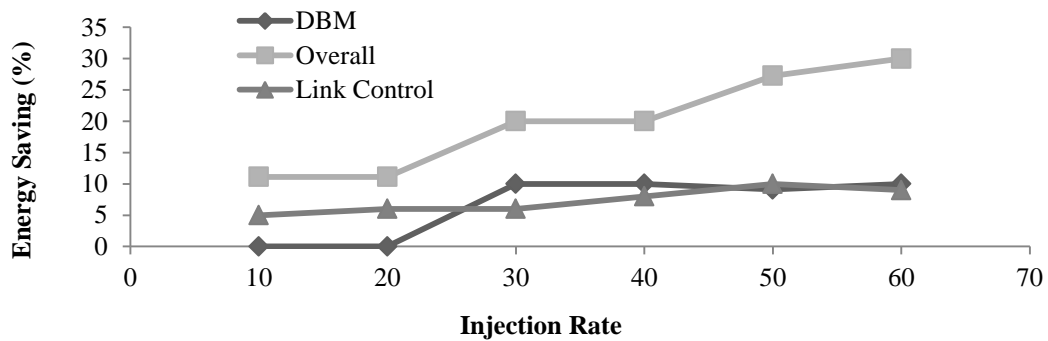


Figure 37: Energy Saving Comparison

Figure 37 illustrates the energy savings achieved by using the dynamic feedback controlled buffer allocation and the adaptive flit storage encoding. The combined technique achieved up to 30% energy saving compared to static buffer allocation technique. At higher flit injection rates congestion causes frequent change in the buffer requirement thus leading to repeated adjustment of buffer allocation resulting in slightly higher energy consumption. Also, frequent on/off of the buffers causes more write overhead in the encoding scheme and hence reduces energy savings little further. In the combined mode the link controller savings make up for the loss in buffer management.

4.6.4. Overhead

The proposed feedback controller for the block level buffer management and the adaptive controller for flit level power management was designed and synthesized using 90nm TSMC Library to calculate the area and power overheads. Table 10 shows that the overheads minimal. 90nm technology was used due to unavailability of sufficient library support for 45nm. However, the results are indicative of the low overall area & power overhead.

Table 10: Controller Overheads

Overhead	Buffer Allocator	Flit Encoding Selector	Total
Area	870 GE	1172 GE	2042 GE
Power	91 μ W	69 μ W	160 μ W

4.7. Summary

In this section, we propose microarchitecture enhancements to complement and support higher level power management techniques. A novel dynamic buffer management technique was presented. The proposed technique utilizes traffic characterization to do predict buffer utilization and perform effective dynamic power management of the router buffer. In addition, a dynamic link control algorithm is presented to optimize link utilization by using multimode links. Experimental evaluation

using the standard NoCSim simulation environment demonstrated up to 30% reduction in energy consumption while improving overall throughput by 21%.

5. CONCLUSION

This dissertation presents a top down approach to dynamic power management in modern nanoscale Network on Chip based System on Chips. A novel ant system inspired distributed dynamic power budget sharing technique is proposed. The proposed scheme automatically establishes sharing paths between budget constrained and over budgeted nodes in the network. This effectively allows more efficient utilization of the given power budget of the system without exceeding it. Experimental evaluation through simulation have shown that proposed sharing scheme improves network throughput by up to 30% and improve utilization by up to 21%.

Careful observation of the working of system level power management schemes revealed their effect in underlying network operation. Use of packet throttling to control power consumption was shown to potentially harm traffic that should not be affected by the power management decision. A novel flow control technique was proposed to avoid power management related congestion and eventual deadlock in the network. The proposed flow control with early notification avoids congestion by preventing packet lockup. Soft power budget limit is used to enter a packet clearing mode before finally throttling the packets completely. The proposed scheme was shown to improve throughput by in a power budget constrained environment by 12.5%. The necessary logic changes were evaluated for overhead. The power overhead was found to be negligible at 0.1% while area overhead was a low 8% compared to a traditional design.

Finally, novel router microarchitecture was proposed to adaptively manage buffer to minimize energy consumption without affecting performance. In addition, a

novel link controller was proposed to reduce energy consumption even further where communication requirement was low. The proposed buffer management was shown to achieve 20% reduction in energy consumption compared to static buffer management. Experimental evaluation also showed that the proposed dynamic management does not affect performance in terms of latency and throughput.

While this dissertation presents solution at different levels of NoC management, there are numerous problems that still need solutions. Next generation microarchitecture will enable even more communication with the higher layer operating system to enable highly integrated power management capability. On the other hand, future device technologies will enable innovative power saving features that will have to be considered at each level of power management system.

To summarize, this dissertation is an attempt at addressing the increasingly complex problem of dynamic power management in high performance nanoscale NoCs from a system level down to microarchitecture level. Solution at each level exposed new problems and inspired new techniques and solutions at the next. The research findings improve the state of the art in dynamic power management and open new research possibilities towards an ultimate goal of maximum energy efficiency. With the limited supply of energy in an ever increasing world of computation this is a crucial step towards sustainable technology.

REFERENCES

- [1] L. Benini and G. De Micheli, "Networks on chips: A new SoC paradigm," *Computer*, vol. 35, no. 1, pp. 70-78, 2002.
- [2] W.J. Dally and B. Towles, "Route packets, not wires: on-chip interconnection networks," in *Proceedings of Design Automation Conference*, 2001, pp. 684-689.
- [3] J. Howard, S. Dighe, Y. Hoskote, S. Vangal, D. Finan et al., "A 48-core IA-32 message-passing processor with DVFS in 45nm CMOS," in *Proceedings of International Solid-State Circuits Conference*, 2010, pp. 108-109.
- [4] P. Bhojwani and R.N. Mahapatra, "Core network interface architecture and latency constrained on-chip communication," in *7th International Symposium on Quality Electronic Design*, 2006, pp. 363-368.
- [5] R. Mullins, "Minimising dynamic power consumption in on-chip networks," in *International Symposium on System-on-Chip*, 2006, pp. 1-4.
- [6] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," in *Proceedings of the 2004 International Workshop on System Level Interconnect Prediction*, 2004, pp. 7-13.
- [7] J. Warnock, Y. Chan, W. Huott, S. Carey, M. Fee et al., "A 5.2GHz microprocessor chip for the IBM zEnterprise™ system," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, San Francisco, 2011, pp. 70-72.
- [8] P. S. Bhojwani, J. D. Lee, and R. N. Mahapatra, "SAPP: scalable and adaptable

- peak power management in NoCs," in *Proceedings of International Symposium on Low Power Electronics and Design*, 2007, pp. 340-345.
- [9] M. Keating, D. Flynn, R. Aitken, A. Gibbons, and K. Shi, *Low Power Methodology Manual for System-on-Chip Design*. New York, NY, USA: Springer US, 2007.
- [10] L. Shang, L. Peh, and N. K. Jha, "PowerHerd: a distributed scheme for dynamically satisfying peak-power constraints in interconnection networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 1, pp. 92-110, 2006.
- [11] A. Lambrechts, P. Raghavan, A. Leroy, G. Talavera, T.V. Aa et al., "Power breakdown analysis for a heterogeneous NoC platform running a video application," in *16th IEEE International Conference on Application Specific Systems*, 2005, pp. 179-184.
- [12] Y. Jin, E.J. Kim, and K. H. Yum, "Peak power control for a QoS capable on-chip network," in *Proceedings of International Conference on Parallel Processing*, 2005, pp. 585-592.
- [13] L. Shang, L. Peh, and N.K. Jha, "PowerHerd: dynamic satisfaction of peak power constraints in interconnection networks," in *Proceedings of the 17th Annual International Conference on Supercomputing*, 2003, pp. 98-108.
- [14] M. Daneshtalab, A. Sobhani, A. Afzali-Kusha, O. Fatemi, and Z. Navabi, "NoC hot spot minimization using AntNet dynamic routing algorithm," in *Proceedings of International Conference on Application-specific Systems, Architectures and*

Processoros, 2006, pp. 33-38.

- [15] M. Dorigo, V. Maniezzo, and A. Coloni, "Ant System: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 26, no. 1, pp. 29-41, Feb 1996.
- [16] X. Chen and L. Peh, "Leakage power modeling and optimization in interconnection networks," in *International Symposium on Low Power Electronics and Design*, 2003, pp. 90-95.
- [17] J.D. Lee and R.N. Mahapatra, "In-Field NoC based SoC testing with distributed test vector storage," in *Proceedings of International Conference on Computer Design*, 2008, pp. 206-211.
- [18] IETF. (2010, May) RFC 2481. [Online]. <http://tools.ietf.org/html/rfc2481>
- [19] A. Banerjee, R. Mullins, and S. Moore, "A Power and Energy Exploration of Network-on-Chip Architectures," in *Proceedings of the First International Symposium on Network on Chip*, 2007, pp. 163-172.
- [20] U. Y. Ogras, J. Hu, and R. Marculescu, "Key research problems in NoC design: a holistic perspective," in *Proceedings of the 3rd IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*, 2005, pp. 69-74.
- [21] A.B. Kahng, Bin Li, L. Peh, and K. Samadi, "ORION 2.0: A fast and accurate NoC power and area model for early-stage design space exploration," in *Design, Automation & Test in Europe Conference & Exhibition*, 2009, pp. 423-428.

- [22] IC-UNICAMP LSC. (2009, September) ArchC Architecture Description Language. [Online]. <http://archc.sourceforge.net>
- [23] C. A. Nicopoulos, D. Park, J. Kim, N. Vijaykrishnan, M.S. Yousif et al., "ViChaR: A dynamic virtual channel regulator for network-on-chip routers," in *IEEE/ACM International Symposium on Microarchitecture*, 2006, pp. 333-346.
- [24] L. Wang, J. Zhang, X. Yang, and D. Wen, "Router with centralized buffer for network-on-chip," in *Proceedings of the 19th ACM Great Lakes Symposium on VLSI*, 2009, pp. 469-474.
- [25] T. Simunic and S. Boyd, "Managing power consumption in networks on chips," in *Design, Automation and Test in Europe Conference and Exhibition*, 2002, pp. 110-116.
- [26] J. Wang, H. Zeng, K. Huang, G. Zhang, and Y. Tang, "Zero efficient buffer design for reliable network-on-chip," in *Design Automation and Test in Europe*, 2008, pp. 792-795.
- [27] C. Lee, M. Potkonjak, and W.H. Mangione-Smith, "MediaBench: a tool for evaluating and synthesizing multimedia and communication systems," in *Proceedings of the 30th Annual ACM/IEEE International Symposium on Microarchitecture*, 1997, pp. 330-335.
- [28] SPEC2006. (2010, September) The SPEC2006 Benchmark Suite. [Online]. <http://www.spec.org>
- [29] J. Barth, D. Plass, E. Nelson, C. Hwang, G. Fredeman et al., "A 45 nm SOI

- embedded DRAM macro for the POWER™ Processor 32 MByte on-chip L3 cache," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 64-75, January 2011.
- [30] A. Agarwal, S. Hsu, S. Mathew, M. Anders, H. Kaul et al., "A 32nm 8.3GHz 64-entry \times 32b variation tolerant near-threshold voltage register file," in *IEEE Symposium on VLSI Circuits*, 2010, pp. 105-106.
- [31] G. Thakral, S.P. Mohanty, D. Ghai, and D.K. Pradhan, "A combined DOE-ILP based power and read stability optimization in nano-CMOS SRAM," in *Proceedings of the 23rd IEEE International Conference on VLSI Design*, 2010, pp. 45-50.
- [32] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45nm design exploration," in *Proceedings of International Symposium on Quality Electronic Design*, 2006, pp. 585-590.
- [33] K. Agarwal, H. Deogun, D. Sylvester, and K. Nowka, "Power gating with multiple sleep modes," in *7th International Symposium on Quality Electronic Design*, 2006, pp. 633-637.
- [34] A. Lungu, P. Bose, A. Buyuktosunoglu, and D.J. Sorin, "Dynamic power gating with quality guarantees," in *Proceedings of International Symposium on Low Power Electronics and Design*, 2009, pp. 377-382.
- [35] E. Pakbaznia and M. Pedram, "Design and application of multimodal power gating structures," in *Proceedings of 10th International Symposium on Quality of*

Electronic Design, 2009, pp. 120-126.

- [36] H. Jiao and V. Kursun, "Power gated SRAM circuits with data retention capability and high immunity to noise: A comparison for reliability in low leakage sleep mode," in *Proceedings of International SoC Design Conference*, 2010, pp. 5-8.
- [37] H. Xu, R. Vemuri, and W. Jone, "Dynamic characteristics of power gating during mode transition," *IEEE Transactions on VLSI Systems*, vol. 19, no. 2, pp. 237-249, February 2011.
- [38] M. Zhang, S. Mitra, T. M. Mak, N. Seifert, N. J. Wang et al., "Sequential element design with built-in soft error resilience," *IEEE Transaction on VLSI Systems*, vol. 14, no. 12, pp. 1368 - 1378, December 2006.
- [39] S.K. Mandal, N. Gupta, A. Mandal, J. Malave, J.D. Lee et al. (2009, April) UCAS 2009: NoCBench: A Benchmarking Platform for Network on Chip. [Online].
http://ispass.org/ucas5/session1_1_tamu.pdf

VITA

Suman Kalyan Mandal received his Bachelor of Technology (Hons.) degree in computer science and engineering from Indian Institute of Technology at Kharagpur, India in 2006. He entered the Computer Engineering program at Texas A&M University in September 2006. His research interests include System on Chip, Network on Chip, Power Management and VLSI. He will join Intel Corporation to continue his engineering career.

Mr. Mandal may be reached at sumankalyan@gmail.com. His permanent home address is below:

C/O: Mr. Dharanidhar Mandal
College Para,
Beliatore
PO: Beliatore, Dis: Bankura
West Bengal, 722203
India