Analysis of Power Consumption on Switch Fabrics in Network Routers

Terry Tao Ye
Computer Systems Lab
Stanford University
taoye@stanford.edu

Luca Benini
DEIS
University of Bologna
lbenini@deis.unibo.it

Giovanni De Micheli
Computer Systems Lab
Stanford University
nanni@stanford.edu

ABSTRACT

In this paper, we introduce a framework to estimate the power consumption on switch fabrics in network routers. We propose different modeling methodologies for node switches, internal buffers and interconnect wires inside switch fabric architectures. A simulation platform is also implemented to trace the dynamic power consumption with bit-level accuracy. Using this framework, four switch fabric architectures are analyzed under different traffic throughput and different numbers of ingress/egress ports. This framework and analysis can be applied to the architectural exploration for low power high performance network router designs.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Design studies, Modeling techniques

General Terms

Design, Experimentation, Performance

Keywords

Networks on Chip, Interconnect Networks, Systems on Chip, Power Consumption

1. INTRODUCTION

With the fast development of Internet applications, the network bandwidth and capacity are moving quickly to the Giga-bit and Tera-bit domains. The switch fabric circuit is the fundamental building block inside a network router, it distributes all network traffic from ingress ports to egress ports. The performance of switch fabrics is very critical in network applications.

While most attention is focused on speed and capacity issues of switch fabrics, power consumption is becoming a more and more serious problem, especially for single chip network routers, where switch fabric circuit contributes a significant part of the total power consumption.

There are many different switch fabric architectures used in network routers. They have different characteristics in terms of bandwidth, throughput and delay [1]. In this paper, we will focus on the power consumption analysis of these different architectures, and estimate how the power consumption scales with the network capacity. Particularly, the following questions are to be addressed and answered:

1. What is the power consumption of different switch fabric architectures under different network traffic loads?

2. How does the power consumption scale with different numbers of ingress and egress ports?

In a switch fabric circuit, the power is dissipated on three components: 1) the internal node switches, located on the intermediate nodes between ingress and egress ports; 2) the internal buffers, used to temporarily store the packets when contention between packets occurs; and 3) the interconnect wires that connect node switches. The power consumption on these three components changes differently under different traffic loads and configurations. Therefore, they need to be analyzed with different modeling methodologies.

In this paper, we first propose different power consumption models for each of the above three components. The total power consumption is then analyzed and simulated on four widely-used switch fabric architectures; 1)Crossbar, 2)Fully Connected 3)Banyan and 4)Batcher-Banyan. A Simulink based multi-threading simulation platform is created for this analysis. It performs a time domain simulation with bit level accuracy. The power consumption of every bit in the traffic flow is traced as the bit moves among the components inside switch fabrics.

Previous approaches for switch network power estimation are either based on statistical traffic models [2], or analytical models [3] [4]. The simulation is performed on gate or circuit levels, it is time consuming and not practical for large switch fabric designs. Furthermore, these approaches are not suitable for tracing the power consumption dynamically in real-time network traffic conditions. For example, the power consumption on internal switch buffers depends on the dynamic contention between packets. Compared with previous approaches, our modeling is an architectural-level estimation with bit level accuracy. It traces the power consumption based on dynamic packet dataflows. This approach is ideal for architectural design exploration as well as application specific power analysis.

The paper is organized as follows. Section 2 describes the network router architectures and functions of switch fabrics. Section 3 proposes the power modeling techniques for switch fabrics. Using these techniques, four switch fabric architectures are analyzed in Section 4. Experiment benchmarks are described in Section 5 with results in Section 6.

2. ROUTERS AND SWITCH FABRICS

A network router consists of four parts: 1) the ingress packet process unit, 2) the egress packet process unit, 3) the arbitration unit (arbiter), and 4) the switch fabrics. Fig. 1 shows the block diagram of a network router architecture.

The ingress packet process unit parallelizes the serial dataflow on the transmission line into bus dataflow, and inspects the header and the content (e.g. in content-aware routing algorithms) of the incoming packets. The egress process unit re-assembles the processed

DAC 2002, June 10-14, 2002, New Orleans, Louisiana, USA.

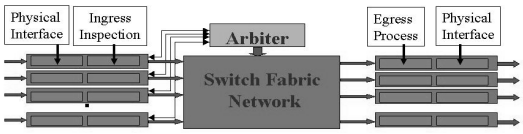

Figure 1: Block Diagram of a Network Router

packets and delivers the packets to their destination ports. The arbitration unit (the arbiter) determines when and where a packet should be routed from the ingress ports to the egress ports. It uses the packet information from the ingress inspection and makes the decision based on arbitration algorithms.

A switch fabric circuit is an interconnect network that connects the ingress ports to the egress ports. Different switch fabric architectures have different impacts on the network router performance, e.g. throughput, delay, power, etc. In the following sections, we will only analyze the power consumption issues on switch fabrics and estimate the power consumption of different switch fabric architectures with different numbers of ingress and egress ports.

3. POWER MODELING WITH BIT ENERGY

A switch fabric circuit is an on-chip interconnect network. The power consumption on switch fabrics comes from three different sources: 1)the internal node switches; 2) the internal buffer queues; and 3) the interconnect wires. Inside the switch fabrics, different packets travel on different data paths concurrently, and the traffic load on each data path may change dramatically from time to time. To estimate the dynamic power consumption in this multi-process interconnect network, we propose a new modeling approach: the *Bit Energy* E_{bit}.

The bit energy E_{bit} is defined as the energy consumed for each bit when the bit is transported inside the switch fabrics from ingress ports to egress ports. The bit energy E_{bit} is the summation of the bit energy consumed on node switches, $E_{S_{bit}}$, on internal buffers, $E_{B_{bit}}$ and on interconnect wires, $E_{W_{bit}}$. We will analyze these three bit energies in details in the following sections.

3.1 Node Switch Power Consumption

Node switches are located on the intermediate nodes inside switch fabrics. They direct the packets from one intermediate stage to the next stage until reaching the destination ports. In different switch fabric topologies, node switches may have different functions and node degrees. For example, in Banyan switch fabrics, the node switch is a 2 $\times$ 2 switch and has degree of 4.

When a data bit travels through a node switch, The logic gates on the data path inside the node switch consume power as they toggle between power rails. We will analyze the 2 $\times$ 2 switch used in Banyan switch fabrics as an example (Fig. 2).

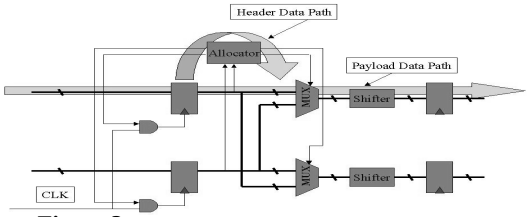

Figure 2: A 2 $\times$ 2 Node Switch in Banyan Switch

The 2 $\times$ 2 switch directs the packets from its two inputs to the outputs, according to the destination addresses of the packets. The ingress process unit had already parallelized the serial dataflow on the transmission line into a parallel bus dataflow (16-bit or 32-bit wide), so the destination address can be read out at one clock cycle. The destination bits are first detected by the allocator. If the destination port is available, the allocator allocates the output port to the packet and preserves the allocation throughout the packet transmission. The allocation process and packet transmission process are denoted in Fig. 2 as "header data path" and "payload data path" respectively. Both processes consume energy when packet bits travel through the data paths. The energy consumed by the header bits is actually different from the payload bits. However, the header normally occupies a small portion of the entire packet. Without loss of generality, we will use the payload bit energy as the node switch bit energy $E_{S_{bit}}$ in our analysis.

In reality, the bit energy $E_{S_{bit}}$ also depends on the presence or absence of packets on other ports of the node switch. For example, the switch will consume more power to process two packets at the same time, but the power consumption is not necessarily twice as much as that of processing a single packet. Therefore, the bit energy $E_{S_{bit}}$ is an input state-dependent value and should be expressed in an input vector indexed look-up table. For a switch with n input ports, there will be 2^n different input vectors with different $E_{S_{bit}}$ values.

In our analysis, the look-up table is pre-calculated from Synopsys Power Compiler simulation. The node switch circuit is first simulated based on input vectors. Then, the switching activities on every gate are also traced. Last, the power consumption of the entire circuit is estimated. We simulate different combinations of the input vectors and the results are shown in details in Section 5. Using the look-up table, the power consumption on the switch node can be estimated under different traffic conditions.

3.2 Internal Buffer Power Consumption

When contention occurs between packets, internal buffers are needed to temporarily store the packets with lower priorities (Fig. 3). There are two different types of contention between ingress packets, namely, the *destination contention* and *interconnect contention*.

- *Destination contention* – When there are two or more packets in the ingress ports requesting the same destination port at the same time, *destination contention* will occur. This type of contention is application dependent. In our analysis, we assume the arbiter had already solved this type of contention before the ingress process unit delivers the packets to the switch fabrics.

- *Interconnect contention (Internal Blocking)* – Inside switch fabric circuits, the same interconnect link may be shared by packets with different destinations. The contention on the shared interconnects is called *interconnect contention* or *internal blocking*. It happens inside switch fabric interconnect networks and it is architectural dependent.

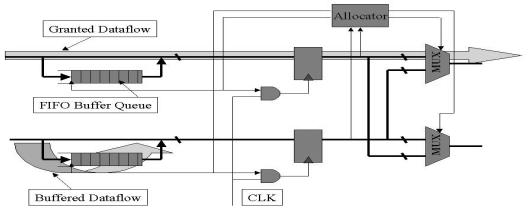

Figure 3: Buffers in a 2 $\times$ 2 Node Switch

In switch fabric circuits, the buffers are normally implemented with shared SRAM or DRAM memories. The energy consumption in buffers comes from two sources: 1) the data access energy, consumed by each READ or WRITE memory access operation, and 2) the refreshing energy, consumed by the memory refreshing operation (in the case of DRAM). The bit energy on the internal buffers $E_{B_{bit}}$ can be expressed by the following equation(Eq. 1):

$$E_{B_{bit}} = E_{access} + E_{ref} \quad (1)$$

where E_{access} is the energy consumed by each access operation and E_{ref} is the energy consumed by each memory refreshing operation. In reality, memory is accessed on word or byte basis instead of a single bit, the E_{access} is actually the average energy consumed for one bit.

The energy consumed by memory access is determined by the contentions between the ingress packets. As discussed earlier, *destination contention* is application dependent, regardless of what switch fabric circuits are used. In comparison, *interconnect contention* is switch fabric architecture dependent. Different architecture topologies result in different contention occurrence. In this paper, we are interested in comparing the power consumption on different switch fabric architectures under the same network traffic, therefore, we assume the *destination contention* has already been resolved by the arbiter before the ingress packets are delivered to the switch fabrics. We only compare the internal buffer energy consumption occurred from *interconnect contention*.

3.3 Interconnect Wires Power Consumption

When the node switch delivers one bit with flipped polarity to the interconnect wires, the signal on the wire will toggle between logic "0" and logic "1". Energy is dissipated in this charging or discharging process. Only bits with flipped polarity consume energy, namely, $E_{W_{bit0\to1}}$ or $E_{W_{bit1\to0}}$ have bit energy values, $E_{W_{bit0\to0}} = 0$, $E_{W_{bit1\to1}} = 0$.

Assuming a rail-to-rail toggling, the bit energy on the interconnect wires $E_{W_{bit}}$ for bit 1→0 and 0→1 can be described by the following equation (Eq. 2).

$$E_{W_{bit}} = \frac{1}{2}C_{wire}V^2 + \frac{1}{2}C_{input}V^2 = \frac{1}{2}C_W V^2 \qquad (2)$$

Here C_{wire} is the wire capacitance on the interconnect, C_{input} is the total capacitance of the input gates connected to the interconnect. $C_W = C_{wire} + C_{input}$ is the total load capacitance propagated by that bit. The rail-to-rail voltage is denoted by V, assuming the CMOS gates are switching from Vdd to GND.

The bit transmitted on interconnect wires consumes energy only when its polarity is flipped from previous transmitted bit. The switching activities on interconnects can be traced by our simulation approach proposed in Section 5.

3.4 Interconnect Wire Length Estimation

The wire capacitance C_{wire} is a function of wire-length and coupling capacitance between adjacent wires [5]. Estimation of interconnect wire-length of the switch fabric network is essential for the bit energy calculation on wires.

Here we adopt the Thompson model [6] for wire-length estimation. The estimation is based on a graph embedding process as described below and also illustrated in Fig. 4.

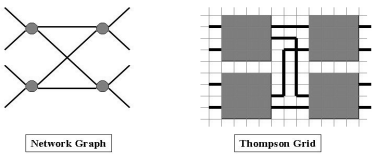

Figure 4: Thompson Wirelength Estimation Model

We are given a source graph $G(V_G, E_G)$, where $V_{G_i}(0 < i \leq n)$ are the vertices of graph *G* and $E_{G_i}(0 < i \leq m)$ are the edges. The source graph represents the network topology. The target graph is denoted by $H(V_H, E_H)$, where V_H and E_H are the vertices and edges of *H*. Graph *H* is a 2-dimensional grid mesh consisting of *p* columns and *q* rows. An embedding of graph *G* into graph *H* is performed as follows. Each vertex in *G* is mapped into a $d \times d$ square of vertices in *H*, where d is the degree of vertex $v_i \in V_G$ and no more than one vertex in V_G occupies the same vertex in V_H. Each edge in *G* is mapped into one or more edges of graph *H*, and no more than one edge in E_G occupies the same edge in graph *H*. The optimal Thompson embedding of graph *G* into graph *H* is to find the minimum number of columns p_{min} and rows q_{min} in *H* that are needed for this embedding. The interconnect wire-length is defined as the number of grids that an edge E_G covers.

In our approach, we manually map the switch fabric topologies into Thompson grids and estimate the interconnect wire-length by counting the number of grids the interconnect covers. The detailed mapping of each particular switch fabric topology will be shown in Section 4.

The Thompson model is only a global wire-length estimation. It is not as accurate as detailed wire-routing, but it is an effective way of architectural planning for interconnect networks, especially when the network topology is regular. In the case of switch fabrics, it is straightforward to map the regular switch nodes and interconnects into a 2-dimensional mesh of regular rows and columns. Therefore, the Thompson model is an ideal model for switch fabric wire-length estimation.

For an interconnect wire of length equal to one Thompson grid, we define the wire bit energy consumption as $E_{T_{bit}}$. If an interconnect wire has its length equal to m Thompson grids, its wire bit energy is $E_{W_{bit}} = m \times E_{T_{bit}}$.

4. SWITCH FABRIC ARCHITECTURES

With the above proposed power consumption model, we will analyze four widely-used switch fabric architectures in this section.

4.1 Crossbar Switch Fabrics

An $N \times N$ crossbar network connects N input ports with N output ports. Any of the N input ports can be connected to any of the N output ports by a node switch on the corresponding crosspoint (Fig. 5).

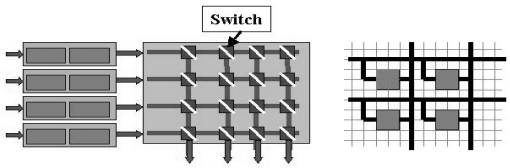

Figure 5: Crossbar Switch Fabrics

Crossbar topology uses space division multiplexing for input-output connection. Every input-output connection has its own dedicated data path, therefore, crossbar is *interconnect contention* free.

The node switch on the crosspoint of crossbar network can be a simple CMOS pass gate, or a tri-state CMOS buffer. Both are relatively simple compared to the node switches used in other network topologies.

Every bit will propagate throughout the long interconnect wires that connect to the input port (the row interconnect, in Fig. 5) and output port (the column interconnect). It also toggles the input gates of all the node switches connected to the same row. The load on the input port is the total of the wire capacitance and the sum of all input capacitances of N switches.

Some crossbar switch networks use buffers at every crosspoint to solve the *destination contention* problems. As discussed earlier in this paper, we assume the *destination contention* is already resolved by the arbiter, i.e., there are no buffers needed in the power modeling of crossbar network.

A Thompson embedding of crossbar switch network is also shown in Fig. 5. Under the Thompson model, the mapping is straightforward and the total bit energy for the crossbar switch fabrics is described in Eq. 3.

$$E_{bit_{crossbar}} = N \times E_{S_{bit}} + 8N \times E_{T_{bit}} \qquad (3)$$

where $E_{S_{bit}}$ is the bit energy for the switch and $E_{T_{bit}}$ is the bit energy of a Thompson grid wire. Each crossbar node switch has degree of 4, however, two of the ports are used as feed-through ports, so we assume it occupies 2×2 Thompson grids. Two extra grids are also needed for horizontal and vertical interconnects for each node switch. Each bit traveling from input i to output j will propagate both the interconnect wires connected to the input port i and output port j, each of the interconnect has length of $4N$ of a Thompson grid.

Crossbar has the benefit of being free of *interconnect contention*. However, the bit energy will increase linearly with the number of

input and output ports N. The power consumption will be very high for switch fabrics with large port numbers.

4.2 Fully-Connected Network

An $N \times N$ fully-connected network uses MUXes to aggregate every input to the output (Fig. 6). Each MUX is controlled by the arbiter that determines which input should be directed to the output.

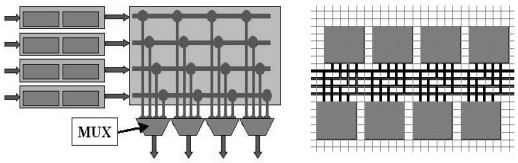

Figure 6: Fully Connected Switch Fabrics

Similar to the crossbar network (fully connected switch network is also often referred as crossbar), in fully connected switch network, each source-destination connection has its dedicated data path. The network is also free of *interconnect contention*. There are no internal buffers needed in its power modeling.

The bit energy for a fully connected switch network is consumed on the interconnect wires and the MUXes. A Thompson embedding is shown in Fig. 6, where the MUXes are placed in a double-row fashion. The bit energy can be estimated with the following equation (Eq. 4).

$$E_{bit_{fullyconn}} = E_{S_{bit}} + \frac{1}{2}N \times N \times E_{T_{bit}} \quad (4)$$

where $E_{S_{bit}}$ is the bit energy on the MUX. Compared with crossbar switch, each bit only consumes energy on one of the MUXes, instead of N switches as in the case of crossbar. However, the N-input MUX has more complicated logic gates, and its power consumption and complexity scale up with the number of inputs N.

4.3 Banyan Network

Banyan network (Fig. 7) is an isomorphic variation of Butterfly topology. It has $N = 2^n$ inputs and $N = 2^n$ outputs, where n is called the dimension of Banyan. It has total of $\frac{1}{2}N \log_2 N$ switches in n stages, each stage is referred as stage i where $0 \leq i < n$ [1].

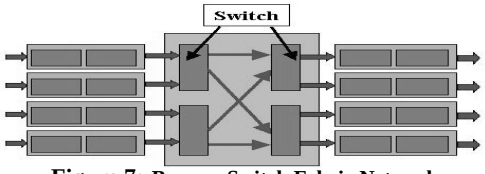

Figure 7: Banyan Switch Fabric Network

The switch used in a Banyan network is a binary switch, as described in Section 3. It has two inputs and two outputs. The input packet with destination bit *"0"* or *"1"* goes to output *"0"* and *"1"* respectively. Stage i in the Banyan network checks the i^{th} bit of the destination address of the packet, therefore, the packet will be routed automatically from stage 0 to stage $n-1$. This routing scheme is also called self-routing switch fabrics. If two input packets have the same destination bit coming at the same time, one of the input packets will be buffered.

Banyan network has *interconnect contention* problems [1]. The same interconnect might be shared by different data paths. A buffer is needed at each internal node switch.

The binary switches in Banyan network have more complex logic as compared to the crosspoint node switches in crossbar network. A binary switch also consumes more power when bit is switched from the input port to the output port.

A simple Thompson embedding of Banyan network is shown in Fig. 4. Detailed analysis of Thompson embedding of Banyan isomorphic networks can be found in [7]. Using the Thompson model, the longest interconnect wire-length for stage i of Banyan network can be estimated as 4×2^i Thompson grid. Different input/output connections have different interconnect data paths. The worst case bit energy of Banyan network (the longest interconnects a bit may travel) can be estimated using Eq. 5 below.

$$E_{bit_{Banyan}} = \sum_{i=0}^{n-1} q_i E_{B_{bit}} + 4\sum_{i=0}^{n-1} 2^i E_{T_{bit}} + nE_{S_{bit}} \quad (5)$$

where $N \geq 2$ and $n \geq 1$. q_i has the value of 0 or 1. When there is a contention at stage i, q_i is 1, otherwise it is 0. The value of q_i is determined by the contentions between packets on the interconnect.

4.4 Batcher-Banyan Network

To solve the *interconnect contention* problem, the Batcher-Banyan network architecture is introduced, as shown in Fig. 8. It consists of a Batcher sorting network, followed by the Banyan network. After sorting network, each input-output connection will have its own dedicated path, therefore there is no *interconnect contention* [1].

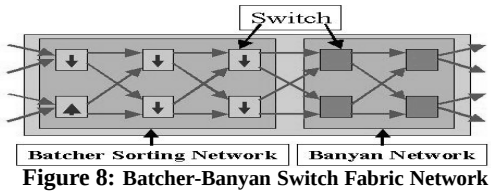

Figure 8: Batcher-Banyan Switch Fabric Network

Although Batcher-Banyan network solves the *interconnect contention* problem, it pays the price by increasing the number of stages between the inputs and outputs. It has total of $\frac{1}{2}(log_2 N)(log_2 N + 1)$stages, which will in turn increase the bit energy consumed on switches and interconnect wires.

The Thompson embedding of Batcher-Banyan is very similar to Banyan network, except it has more stages. The worst case bit energy for Batcher-Banyan network can be expressed in the following equation (Eq. 6).

$$E_{bit_{Batcher}} = 4\sum_{j=0}^{n-1}\sum_{i=0}^{j} 2^i E_{T_{bit}} + 4\sum_{i=0}^{n-1} 2^i E_{T_{bit}} \quad (6)$$

$$+\frac{1}{2}n(n+1)E_{SS_{bit}} + nE_{SB_{bit}} \quad (7)$$

where $N \geq 4$ and $n \geq 2$. Because the sorting switch used in Batcher sorting network is different from the binary switch used in Banyan network. we denote $E_{SS_{bit}}$ for the bit energy of sorting switches and $E_{SB_{bit}}$ for the bit energy of binary switches.

5. BIT-ENERGY CALCULATION AND EXPERIMENTS

In this paper, we present a framework to estimate switch fabric power consumption. The calculation and results are based on case studies. The accurate values of each parameter for a specific switch fabric implementation depend on particular circuit design techniques and technologies. However, the methodology introduced here can be applied in other cases.

5.1 Bit Energy Calculation

1. Bit energy of Node Switches

As discussed in Section 3, the bit energy of node switches is state-dependent, it depends on the presence or absence of the packets on other input ports. For a node switch with n-input ports, there are 2^n different input vectors with different bit energy values. Normally switch fabrics with a large input/output ports are constructed from node switches with smaller number of degrees (2×2 or 4×4), therefore, the vector number 2^n is not prohibitively large. In this paper, the bit energy is pre-calculated from Synopsys Power Compiler simulation. We build each of the node switches with 0.18μm

libraries, apply different input vectors and calculate the average energy consumption on each bit. The circuits for each node switch range from a few hundred gates to 10K gates, therefore, the simulating time is very quick (tens of seconds, in most cases). The bit energy results for crossbar switches, the N-input MUXes, the Banyan binary switches, the Batcher sorting switches are listed in Table 1.

Table 1: Bit Energy Under Different Input Vectors

Switch Fabric Architectures	Input Vector	Bit Energy	Input Vector	Bit Energy
		$10^{-15} joule$		$10^{-15} joule$
Crossbar 1×1	[0]	0	[1]	220
Banyan 2×2	[0,0]	0	[0,1]	1080
	[1,0]	1080	[1,1]	1821
Batcher 2×2	[0,0]	0	[0,1]	1253
	[1,0]	1253	[1,1]	2025
N-input MUX	N = 4	431		
	N = 8	782		
	N = 16	1350		
	N = 32	2515		

The input vectors for the 2×2 switches in the above table indicate presence or absence of the packets on the corresponding input ports, e.g. [1,0] means only input port 0 has packet coming in. For the N-input MUXes, bit energy values are very close among different input vectors, but they increase with the number of inputs N, as shown in the table.

2. Bit Energy of Buffer Queues

We use SRAM as the shared buffers inside Banyan switch fabrics. We adopt techniques similar to those proposed by [8] and [9] to estimate the memory access power consumption. SRAMs with different sizes have different access time and current, therefore they have different memory access power dissipation. The buffer size at each node switch greatly affects the performance of Banyan switch. The trade-off analysis between buffer size and switch throughput is beyond the scope of this paper. Researches in [10] and [11] show that buffer size of a few packets will actually achieve ideal throughput under most network traffic conditions. In our experiments, we use 4K bit buffer queue for each Banyan node switch. Based on the buffer size of each switch, we calculate the size of the shared memory. The memory access energy consumption can then be estimated based on selected memory size.

An off-the-shelf $0.18\mu m$ 3.3V SRAM is used as a reference for bit energy estimation. The calculation of access energy is based on 133MHz operation with access time specified in the data sheet. The results are shown in Table 2.

Table 2: Buffer Bit Energy of $N \times N$ Banyan Network

In/Out Size	Number of Switches	Shared SRAM Size	Bit Energy $(10^{-12} joule)$
4×4	4	16K	140
8×8	12	48K	140
16×16	32	128K	154
32×32	80	320K	222

3. Bit Energy of Interconnect Wires

The length of the Thompson grid can be estimated from the bus pitch distance of the interconnect. In Thompson model, each interconnect is a signal bus and occupies one grid square. Assuming the bus width is 32 bit, in 0.18μm technology, the wire pitch of global buses is around 1μm, therefore, the Thompson grid is around 32μm. The interconnect wire capacitance can be calculated from the method presented in [5]. For a global wire in 0.18μm technology, the wire capacitance is around 0.50fF/μm. Using these estimations, under 3.3V, the bit energy on interconnect wire of a Thompson grid length $E_{T_{bit}} = 87 \times 10^{-15}$ joule.

Comparing the buffer bit energy $E_{B_{bit}}$ values in Table 2 with the interconnect wire bit energy $E_{T_{bit}}$ value calculated above, we can see that storing a packet in buffer consumes far more energy than transmitting the packet on interconnect wires. This "buffer penalty" indicates that energy consumed in buffers is a significant part of total energy consumption of switch fabrics, and the buffer energy will increase very fast as the packet flow throughput increases. Our experiments in Section 6 will show this result.

5.2 Simulation Platform

The bit energy introduced in Sections 3 and 4 is the energy consumption for one bit. In switch fabric interconnect networks, different packets will travel on different data paths. To calculate the total power consumption of the entire switch fabrics, we need to trace the dataflow of every packet and summarize the bit energy of every bit on nodes switches, internal buffers and interconnect wires. In this paper, we introduce a Simulink based bit-level multi-threading simulation platform.

The complete network router system is implemented in Simulink. The ingress process units, the egress process units, the global arbiter and different switch fabric architectures are all written in C++ and then compiled into Simulink S-functions. The switch fabric architecture is constructed hierarchically. The activities of every bit in the packet are traced at every node switches, every buffers as well as interconnect wires (all implemented in S-functions).

A TCP/IP packet traffic flow is generated as inputs for the network router. Because we only need to simulate the switching activities inside the switch fabrics, the packet payloads are random binary bits. The IP address of each packet has been translated into destination port address by the ingress process unit. The arbiter uses the first-come-first-serve arbitration with round robin policy. The destinations of the TCP/IP packets are random. We use input buffer scheme to store the packets when there is *destination contention*. The input buffers are located at each ingress process unit. Because the input buffers are outside the switch fabric network, they are not counted for switch fabric power consumption. The line data rate is assumed to be 100BaseT.

Power consumption is measured under three metrics: 1) different traffic throughput, 2) different switch fabric architectures, and 3) different numbers of ingress and egress ports.

We implement four switch fabric architectures with different numbers of input/output ports, namely, 4×4, 8×8, 16×16 and 32×32. Packet dataflow is generated at each input port. The throughput of the packet dataflow can be adjusted by controlling the packet generation intervals. The throughput is measured at the egress process units. The throughput indicates the traffic loads that flow through the switch fabric networks.

6. RESULTS AND ANALYSIS

Fig. 9 shows the power consumption of different switch fabric architectures under traffic throughput from 10% to 50%. The figure also shows the power consumption/throughput relationship under different numbers of ingress/egress ports. Because we use input buffering scheme to store the packets with *destination contention*, the theoretical maximum throughput is 58.6% (measured at egress ports). In reality, the 58.6% throughput is not achievable [1].

From these results, we have the following observations:

1) *Interconnect contention* has a dramatic impact on the power consumption of Banyan switch. Banyan switch has the lowest power consumption under low traffic throughput, as the throughput increases, the power consumption increases exponentially. This is caused by the "buffer penalty" as discussed in Section 5. However, as the number of ingress and egress ports increases, the interconnect wire-length and bit energy also increase, the "buffer penalty" domination will have less impact. This can be seen by comparing the "Banyan curve" in the figures with the curves of other architectures. Actually, in the 32×32 configuration, Banyan had the lowest power consumption when the traffic throughput is less than 35%.

2) *Fully connected switch* has the lowest power consumption among all four architectures with different numbers of ports. But the difference with Batcher-Banyan narrows down as the number of ports increases. This is because in switch fabrics with a small number of ports, the power consumption on the internal node switches dominates, as the switch fabrics is getting bigger (more ports), interconnect power consumption will dominate.

The impact of number of ports on power can be better seen from Fig. 10. In this figure, power consumption of each architecture is compared with different number of ports. The traffic throughput is 50%. The power consumption difference between fully con-

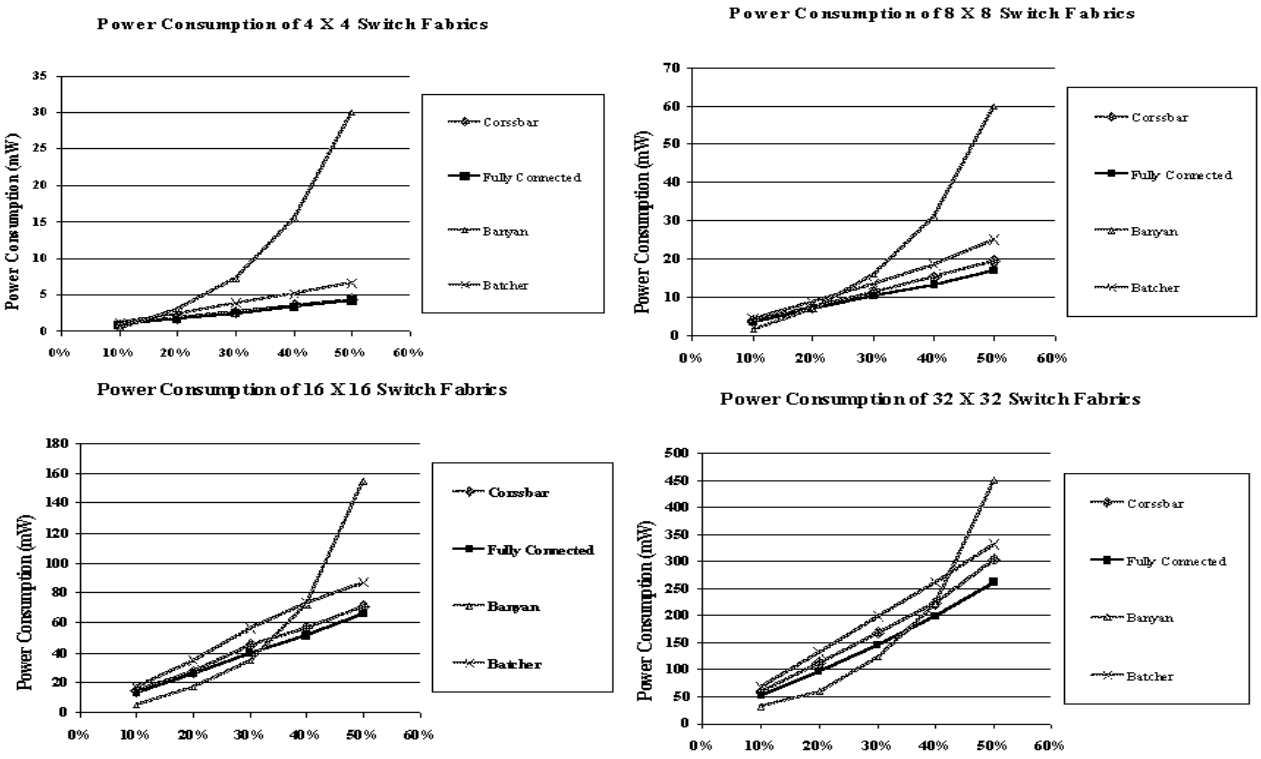

Figure 9: Power Consumption Under Different Traffic Throughput

nected switch and Batcher-Banyan switch decreases from 37% in 4×4 switches to 20% in 32×32 switches.

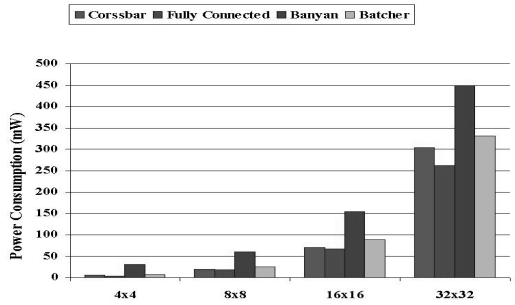

Figure 10: Power Consumption Under Different Number of Ports

3) The power consumption of crossbar, fully connected and Batcher-Banyan networks increases almost linearly with the increase of the traffic throughput, except the Banyan network, which is dominated by the power consumption on internal buffers. This observation is not surprising because data flow with higher throughput needs more power to process the bits along the data path.

7. CONCLUSION

A power consumption estimation framework is proposed in this paper. Using the framework, we analyze the power consumption of different switch fabric architectures used in network routers. From our analysis, we have the following conclusions:

1) *Interconnect contention* (internal blocking) induces significant power consumption on internal buffers, and the power consumption on buffers will increase sharply as throughput increases.

2) For switch fabrics with a small number of ports, internal node switches dominate the power consumption, for switch fabrics with a larger number of ports (e.g. beyond 32×32), interconnect wires will gradually dominate the power consumption.

The power consumption estimation in the results only represents the power dissipated on switch fabrics, the total power consumption for the entire network router needs to include other components as well, e.g. the ingress units, egress units, arbitration logics, peripheral logics, and IOs. The analysis and comparison above are only based on case studies with specific technology parameters. Different implementations of switch fabrics will have different comparison results. However, the methodology presented in this paper is general, it can be applied to different switch fabric designs.

8. ACKNOWLEDGMENTS

This research was supported by GSRC/MARCO.

9. REFERENCES

[1] H. Jonathan Chao, Cheuk H. Lam, Eiji Oki "Broadband Packet Switching Technologies: A Practical Guide to ATM Switches and IP Routers" *Wiley-Interscience 2001*

[2] Wassal, A.G.; Hasan, M.A. "Low-power system-level design of VLSI packet switching fabrics" *CAD of Integrated Circuits and Systems, IEEE Transactions on* , June 2001

[3] C. Patel, S. Chai, S. Yalamanchili, D. Shimmel, "Power constrained design of multiprocessor interconnection networks," *Int'l Conference on Computer Design*, 1997.

[4] Langen, D.; Brinkmann, A.; Ruckert, U. "High level estimation of the area and power consumption of on-chip interconnects," *IEEE Int'l ASIC/SOC Conference, 2000.*

[5] R. Ho, K. Mai, M. Horowitz, "The Future of wires," *Proceedings of the IEEE*, April 2001.

[6] C. D. Thompson, "A Complexity Theory for VLSI, PhD thesis", *Carnegie-Mellon University*, August 1980

[7] Andre DeHon, "Compact, Multilayer Layout for Butterfly Fat-free" 12th ACM Sym. on Parallel Algorithms and Architectures, 2000

[8] E. Geethanjali, N. Vijaykrishnan, M. Kandemir, M. J. Irwin, "Memory System Energy: Influence of Hardware-Software Optimizations", Int'l Sym on Low Power Design and Electronics, July 2000.

[9] Wen-Tsong Shiue; Chakrabarti, C. "Memory exploration for low power, embedded systems", 36th Design Automation Conference, 1999. Proceedings.

[10] Moustafa A. Youssef, M. N. El-Derini and H. H. Aly "Structure and Performance Evaluation of a Replicated Banyan Network Based ATM Switch" IEEE Sym. on Computers and Communications 1999

[11] Oktug, S.F.; Caglayan, M.U. "Design and performance evaluation of a banyan network based interconnection structure for ATM switches", Selected Areas in Communications, IEEE Journal on , June 1997