

How **HOT** Are Interconnection Networks  
— A Power Model for Routers



Hang-Sheng Wang Li-Shiuan Peh Sharad Malik  
{hangshen,peh,sharad}@ee.princeton.edu  
Princeton University



*Outline*



- ◆ Motivation
- ◆ Related work
- ◆ Interconnection router power model
- ◆ Results and insights
- ◆ Conclusion

## *What are interconnection networks?*



### ◆ Fabric connecting subsystems within a digital system

- Multiprocessors, computer clusters, server blades, on-chip networks, etc.
- Performance was primary concern
- Now power efficiency becoming critical
  - ◆ Alpha 21364 integrated router takes about 20% of the total system power (25W out of 125W)
  - ◆ Mellanox server blade designers allocate the same power budget to the router as to the microprocessor
  - ◆ We need to estimate power before we can reduce it

08/21/02

3

## *We need power models and power estimators*



### ◆ Interconnection network power estimation has been long ignored

- Processor power estimation tools abound
  - ◆ Wattch, SimplePower, TEM<sup>2</sup>P<sup>2</sup>EST, SimpleScalar/ARM Power Analyzer, etc.
- But none for interconnection networks

### ◆ So we build one

- Requirements
  - ◆ power models for interconnection networks — routers and links
  - ◆ integration of power estimation and network simulation
- This work addresses the first requirement

08/21/02

4

## Related work



### ◆ Power models for interconnection networks

- Router model based on transistor counts [Chirag S. Patel *et al.* 1997]
- Models for on-chip switch-box networks [Hui Zhang *et al.* 1999]
- Analytical crossbar models [G. Essakimuthu *et al.* 2002]

### ◆ Power models for other networks

- Models for Internet router switch fabrics [Terry Tao Ye *et al.* 2002]

08/21/02

5

## Basics of CMOS circuit power modeling



### ◆ Types of power dissipation

Dynamic power: switch power, short-circuit power

Static power: leakage power

We only model switch power in this work

### ◆ Architectural-level model vs. empirical model

Empirical models — gate counts, energy density, curve fitting, etc.

- ◆ Derived from physical measurements
- ◆ Suitable for product power estimation

Architectural-level models

- ◆ Formulated with architectural and technology parameters
- ◆ Suitable for design evaluation and comparison

08/21/02

6

## Router power modeling methodology



- ◆ Switch power:  $E=0.5\alpha CV_{dd}^2$ ,  $P=f_{clk}E$ 
  - Need to estimate: C and  $\alpha$
- ◆ Divide and conquer to derive C
  - Model different components separately
    - ◆ FIFO buffers, crossbars, arbiters
- ◆ Two ways to estimate  $\alpha$ 
  - Simulation with real traffic
  - Probabilistic estimation
    - ◆ Switch probability  $P_d$
    - ◆  $P_d=1$  for maximum power
    - ◆  $P_d=0.5$  for average power

08/21/02

7

## Some notations

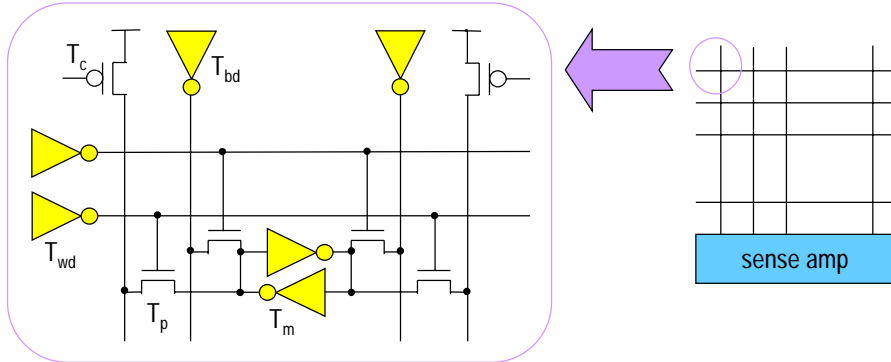


T	a transistor or a gate
$C_g(T)$	gate capacitance of T
$C_d(T)$	diffusion capacitance of T
$C_a(T)$	$C_g(T)+C_d(T)$
$C_w(L)$	capacitance of metal wire of length L

08/21/02

8

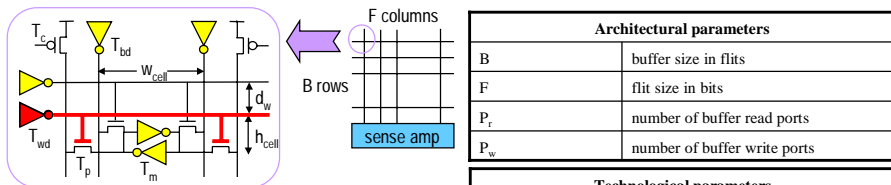
## Modeling details — FIFO buffer model



08/21/02

9

## Modeling details — FIFO buffer model



From  $C_x$  to  $E_x$

$$E_x = 0.5 C_x V_{dd}^2$$

$$E_{write} = E_{wl} + P_d F (E_{bw} + E_{cell})$$

$$E_{read} = E_{wl} + F (E_{br} + 2E_{chg} + E_{amp})$$

Architectural parameters	
B	buffer size in flits
F	flit size in bits
$P_r$	number of buffer read ports
$P_w$	number of buffer write ports

Technological parameters	
$h_{cell}$	memory cell height
$w_{cell}$	memory cell width
$d_w$	wire spacing

Model equations	
wordline length	$L_{wl} = F(w_{cell} + 2(P_r + P_w)d_w)$
bitline length	$L_{bl} = B(h_{cell} + (P_r + P_w)d_w)$
wordline cap.	$C_{wl} = C_w(L_{wl}) + C_a(T_{wd}) + 2FC_g(T_p)$
read bitline cap.	$C_{br} = BC_d(T_p) + C_d(T_c) + C_w(L_{bl})$
write bitline cap.	$C_{bw} = BC_d(T_p) + C_a(T_{bd}) + C_w(L_{bl})$
precharge cap.	$C_{chg} = C_g(T_c)$
memory cell cap.	$C_{cell} = 2(P_r + P_w)C_d(T_p) + 2C_a(T_m)$
sense amp energy	$E_{amp}$ from empirical model

08/21/02

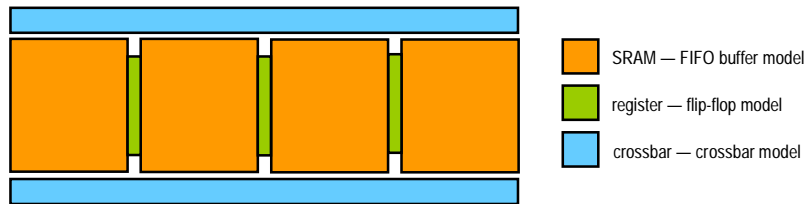
10

## Modeling hierarchy



### ◆ Models have layers

- Separate technology and architectural dependency
- Reusable building blocks
  - ◆ Central buffer — pipelined shared memory



- Fine-grained modeling granularity
- Easy maintenance

08/21/02

11

## Putting it together — from workload to power



### ◆ Activity per flit of a wormhole router

#### FIFO buffer

- ◆ 1 write

#### Arbiter

- ◆  $1/L$  arbitration, assuming  $L$  flits per packet
- ◆ Register clock, independent of workload

#### FIFO buffer

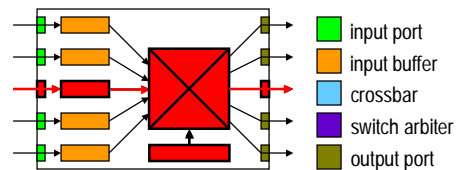
- ◆ 1 read

#### Crossbar

- ◆ 1 flit traversal

### ◆ Scaled by workload

Flit arriving rate: the probability that each input port receives a flit in every cycle



Block diagram of a wormhole router

08/21/02

12

## Case studies — Alpha 21364 integrated router and IBM InfiniBand 8-port 12X router



	Alpha 21364 integrated router	IBM InfiniBand 8-port 12X router
technology	0.18 $\mu$ m	0.11 $\mu$ m
voltage	1.65V	1.2V
frequency	1.2GHz	250MHz
input ports	8	8
output ports	7	8
flit width	32-bit	128-bit
input buffer	319/250/127/190 flits, 2 read ports, 1 write port	256 flits (4KB), 1 read port, 1 write port
switch fabric	two 8x5 crossbars	central buffer: 2560 chunks, 4 banks (160KB), 2 read ports, 2 write ports
local arbiter	19-input matrix arbiter	4-input matrix arbiter
global arbiter	7-input matrix arbiter	7-input matrix arbiter

08/21/02

13

## System design questions

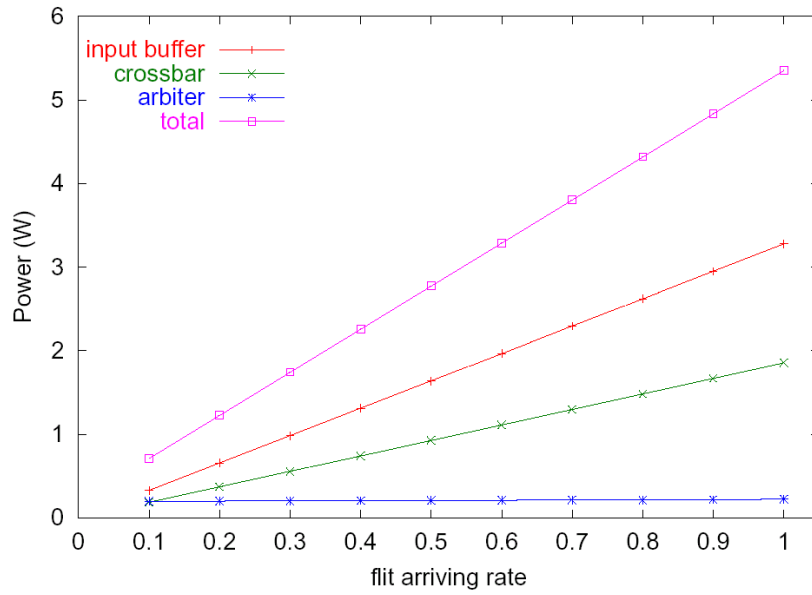


- ◆ How is router power distributed among components?
- ◆ How is router power sensitive to workload?

08/21/02

14

## Results of the Alpha 21364 router

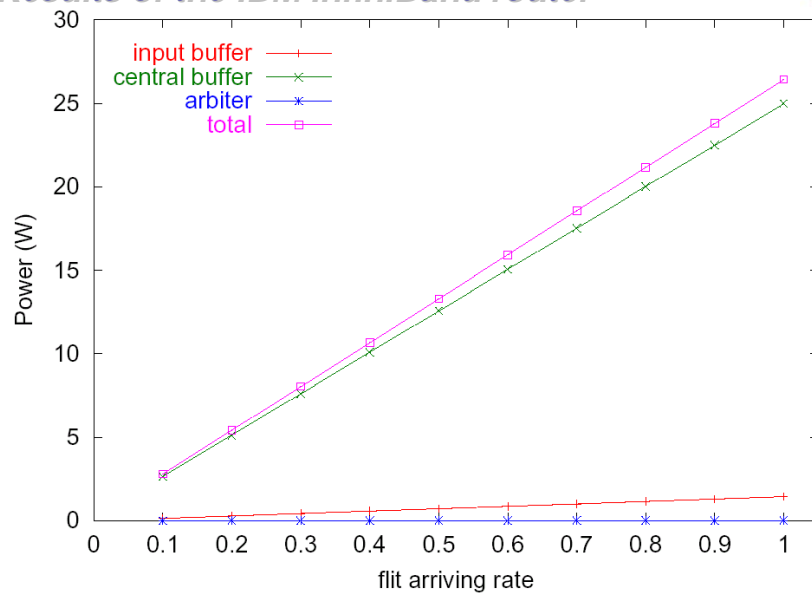


08/21/02

Maximum power of Alpha 21364 router

15

## Results of the IBM InfiniBand router



08/21/02

Maximum power of IBM InfiniBand router

16

## Analysis and insights



### ◆ Router power distribution among components

- ◆ Input buffer and switch fabric dominate
  - ◆ IBM InfiniBand router: nearly 100% of total power
  - ◆ Alpha 21364 router: more than 80% of total power
  - ◆ Memory is the most power consuming component
- ◆ Arbiter power is (almost) negligible

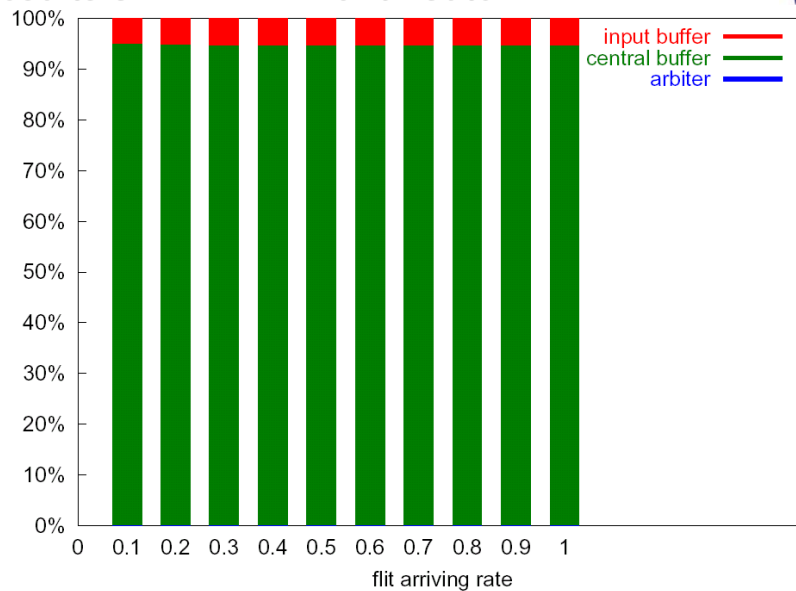
### ◆ Router power sensitivity to workload

- ◆ Approximately linear
  - ◆ Both buffer power and crossbar power are linear in the data traffic
  - ◆ This is good news because it eases macro-modeling
- ◆ IBM InfiniBand router: power breakdown is invariant

08/21/02

17

## Results of IBM InfiniBand router



Maximum power breakdown of IBM InfiniBand router

08/21/02

18

## *Model accuracy*



### ◆ Alpha 21364 router

- Designer data: maximum 7.6W
- Our estimate: maximum 5.36W when flit arrival rate is 1

### ◆ IBM InfiniBand router

- Designer data: average 11W
- Our estimate: average 11W when flit arrival rate is 0.6

### ◆ Hard to tell the accuracy

- Exact measurement conditions unknown
- Actively working on validation

08/21/02

19

## *Conclusions*



### ◆ Propose a “complete” architectural-level power model for interconnection network routers

Enables rapid exploration of power-performance tradeoffs at architecture design time

Easy to reuse and extend

### ◆ Current Status

Power models integrated into Orion — a complete interconnection network simulator (will appear in MICRO 2002)

Actively pursuing opportunities for validations

Working on improving and extending our power models

Public release coming soon

08/21/02

20

## References



- ◆ [Chirag S. Patel *et al.* 1997] Power constrained design of multiprocessor interconnection networks. In *Proc. International Conference on Computer Design*, 1997.
- ◆ [Hui Zhang *et al.* 1999] Interconnect architecture exploration for low-energy reconfigurable single-chip DSPs. In *Proc. IEEE Computer Society Workshop on VLSI*, 1999.
- ◆ [G. Essakimuthu *et al.* 2002] An analytical power estimation model for crossbar interconnects. Technical Report CSE-02-009, Penn State University, Department of Computer Science and Engineering, 2002.
- ◆ [Terry Tao Ye *et al.* 2002] Analysis of power consumption on switch fabrics in network routers. In *Proc. Design Automation Conference*, 2002.