centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

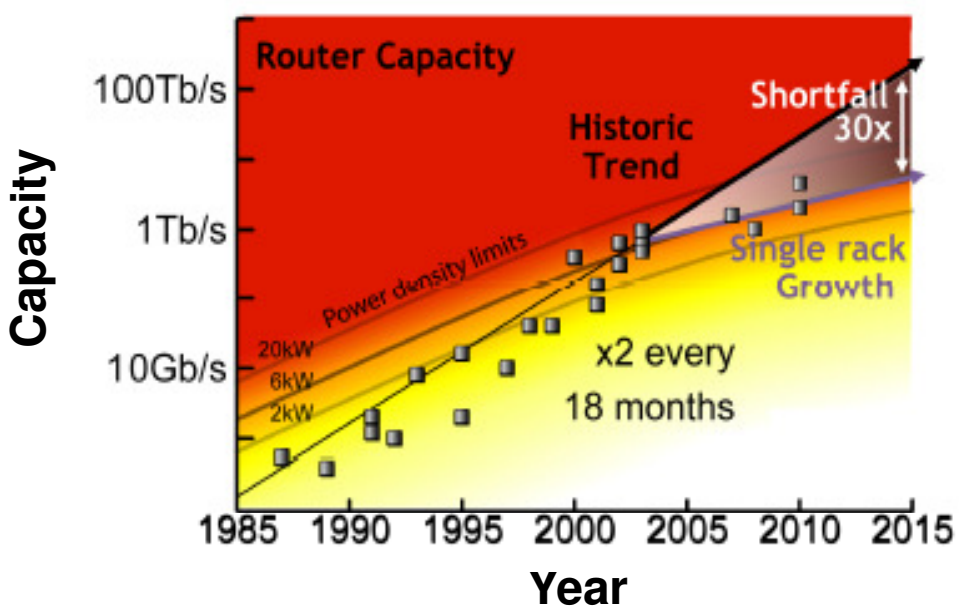
Adapting Router Buffers for Energy Efficiency

Arun Vishwanath
CEET, University of Melbourne

Joint work with Vijay Sivaraman (UNSW), Zhi Zhao (UNSW),
Craig Russell (CSIRO), Marina Thottan (Bell Labs)

Special thanks also to Rod Tucker, Director CEET

Router Capacity Trends



CRS-3 core router

- 446 W/line-card
  - 140 Gb/s
- 12.3 KW/rack
  - 16 cards
  - 4.5 Tb/s (Total)

Courtesy: David Neilson, Bell Labs

- Historically, router capacity x2 every 18 months
  - Future: Unlikely due to power density issues
  - Energy/bit falling 10% per year (3x per decade)
  - Traffic growing 40% per year (30x per decade)

ceet centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

Scaling Problem

- With 10% per year power reduction and 40% per year traffic growth

- 30x increase in traffic from 2010 → 2020
  - 20 Tb/s to 600 Tb/s
  - 12 racks
  - 768 KW just for the line-cards!

ceet centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

Ongoing Work

- At network design stage
  - Optimise number of interfaces and chassis
  - Right combination of optical grooming and IP ports
- Selectively turn-off/underclock line-cards
- Redirect traffic to "greener" areas
- Promise high savings, however ...
  - Clean-slate approaches
  - Major architectural/protocol/design changes
  - Increases the barrier to adoption by ISPs

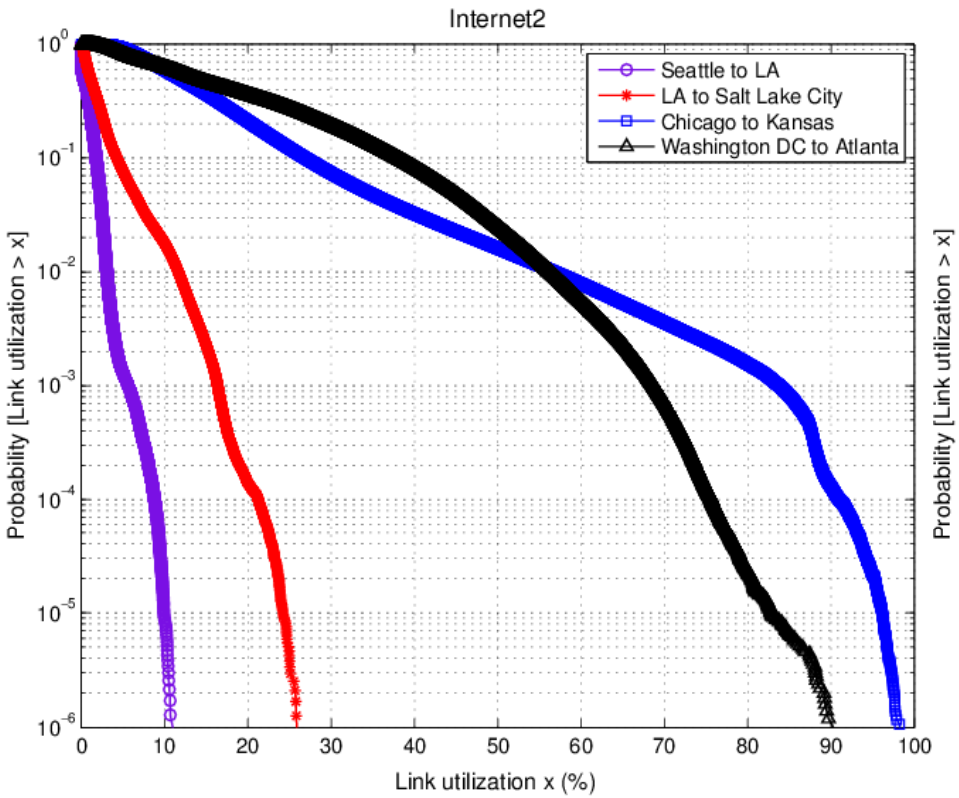
Our Objective

- An evolutionary approach ...
- Focus on packet buffer memory
  - Have large buffers at build time
  - But, adapt its size dynamically
- Incentive – tangible energy savings
  - About 10% of line-card power consumption
  - Memory chips and controllers
  - Useful gains network-wide
  - 30 W × 16 line-cards × 50 routers = 24 KW → **2 core routers**
- ISPs become comfortable operating with small buffers
  - Pave the way for novel 2020 line-card designs

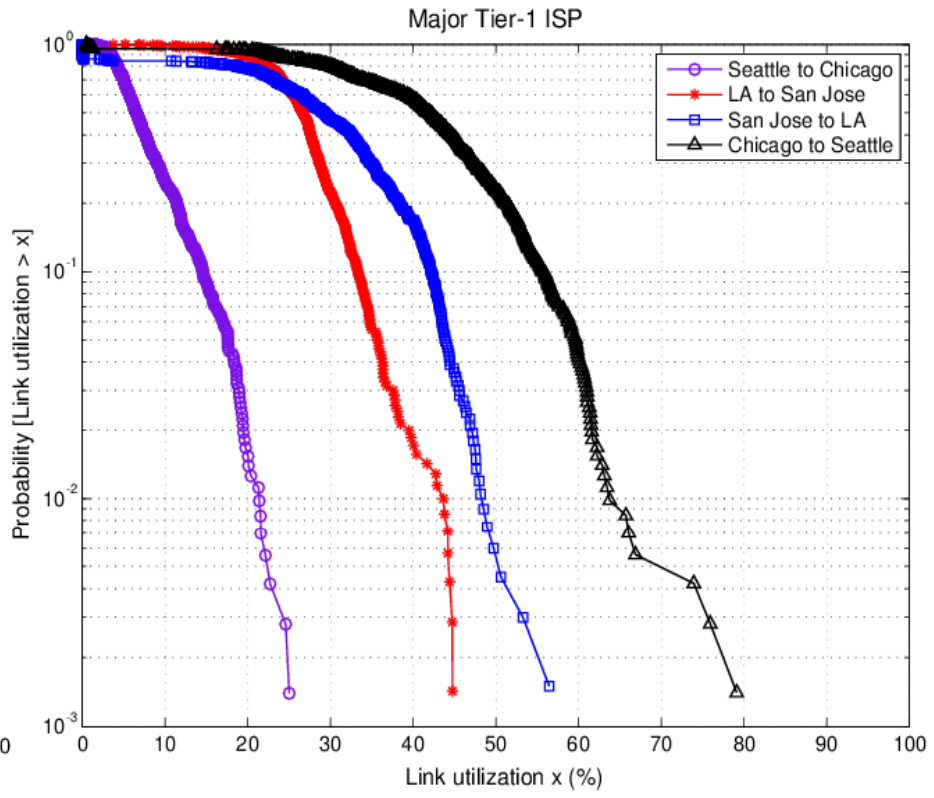
Talk Outline

- Case for reducing buffering energy
  - Link loads
  - Buffer occupancy
- Generic buffering architecture
  - Power saving mechanism
  - Energy model and algorithm
- Performance evaluation
  - Simulations (with TCP)
  - Experimental validation (NetFPGA)
- Conclusions

Link Loads in Operational Networks



CCDF of link load: Internet2

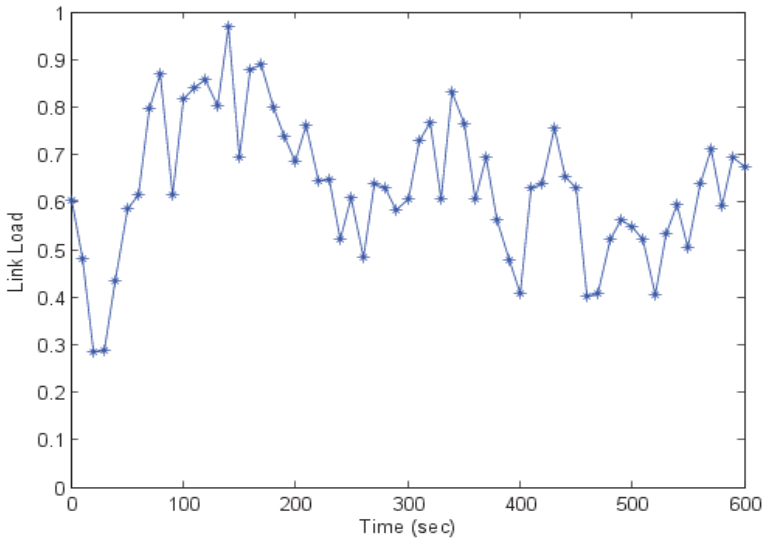


CCDF of link load: Major Tier-1 ISP

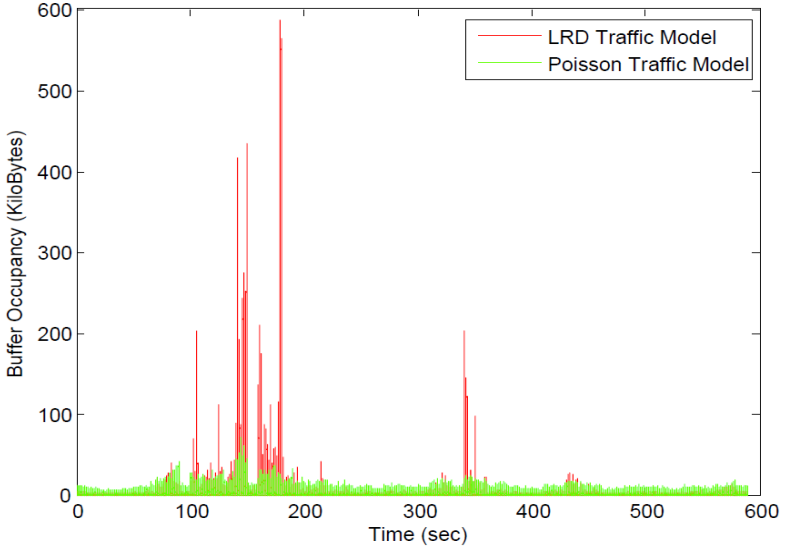
- Link load data spanning nearly 3 years
- Congestion is a relatively infrequent event

ceet centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

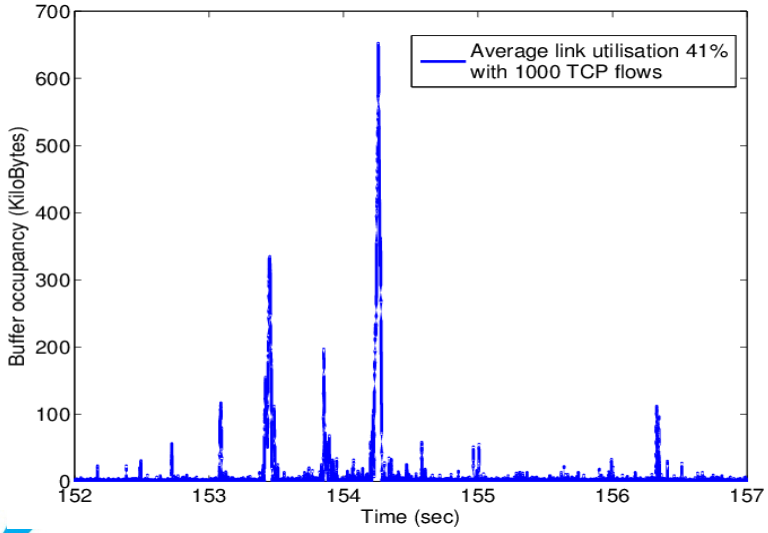
Buffer Occupancy – Opportunity to Save Energy



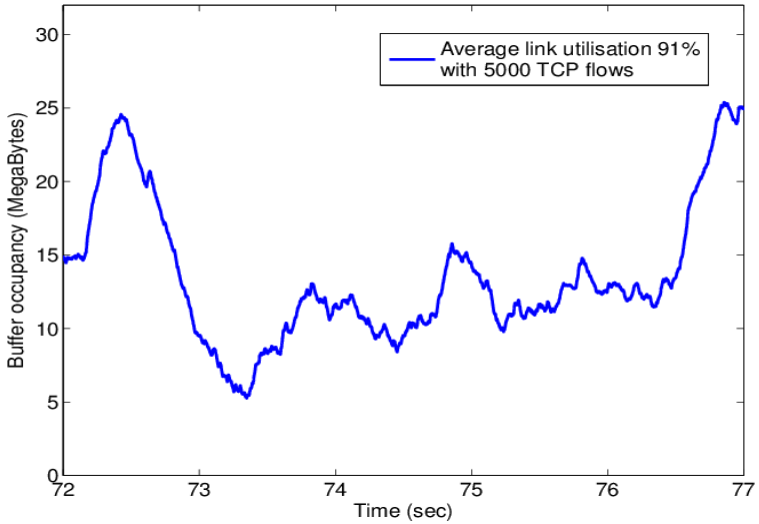
Load from Chicago-Kansas link



Buffer occupancy



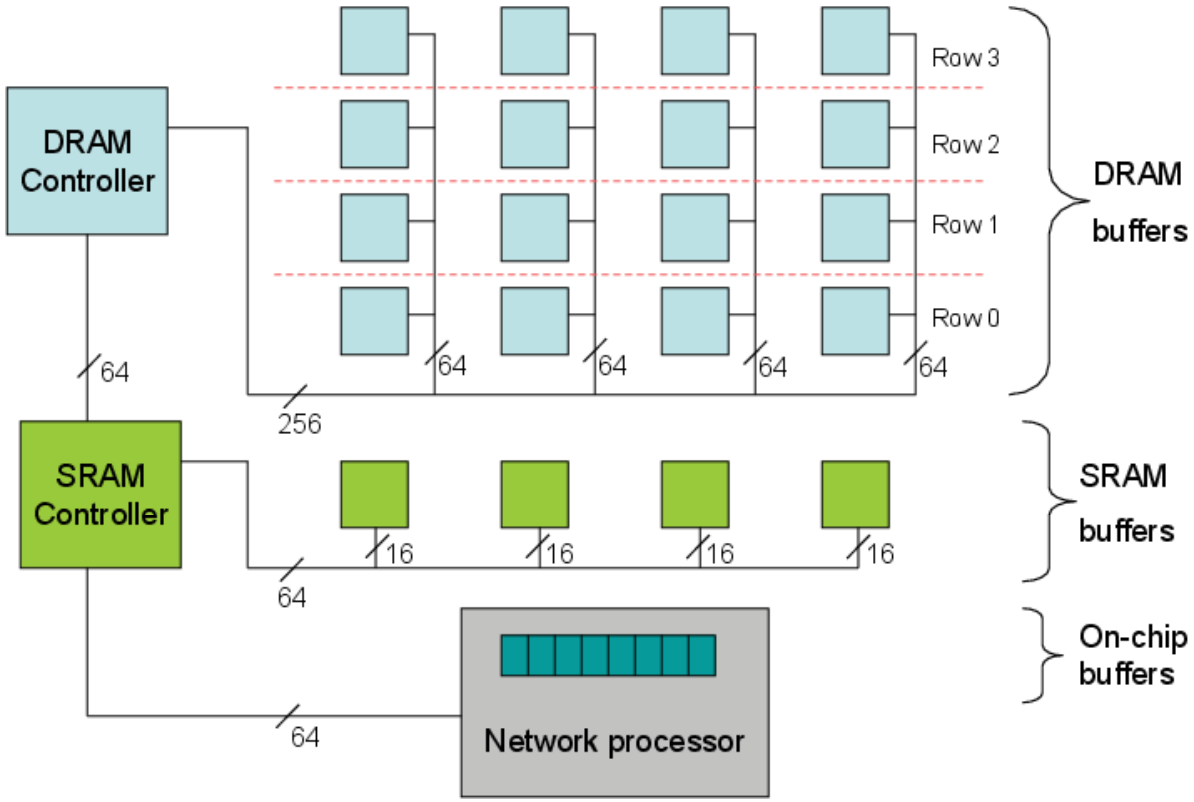
1000 TCP flows



5000 TCP flows

ceet centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

Generic Model of Buffer Architecture



- Memory chips in parallel to meet throughput/latency requirements
- Packet stored in off-chip memory straddles all chips

Power Saving Mechanism

- As buffer occupancy falls
  - Put to sleep DRAM row-3, then row-2, ... (~ 256 MB each)
  - Some point entire DRAM (controller + rows) is asleep
  - Occupancy falls further, put to sleep SRAM (~ 4 MB)
  - On-chip buffers (~ 100 KB) always-on
- Conversely, as buffer occupancy grows
  - Activate SRAM, then DRAM row-0, row-1, ...
- Hysteresis protects against rapid toggling (sleep/awake)

Energy Model

- DRAM – each row 256 MB
  - Power depends on frequency of read/write operations
  - Three states (for simplicity)
    - Active – high frequency of read/write (2W)
    - Idle – little or no read/write (200mW)
    - Sleep – read/write disabled (20mW)
- SRAM – one row 4 MB
  - Static component due to leakage current (dominates)
    - Proportional to the number of transistors
  - Two states – active 4W, sleep 40mW
- Controller power ½ of entire memory

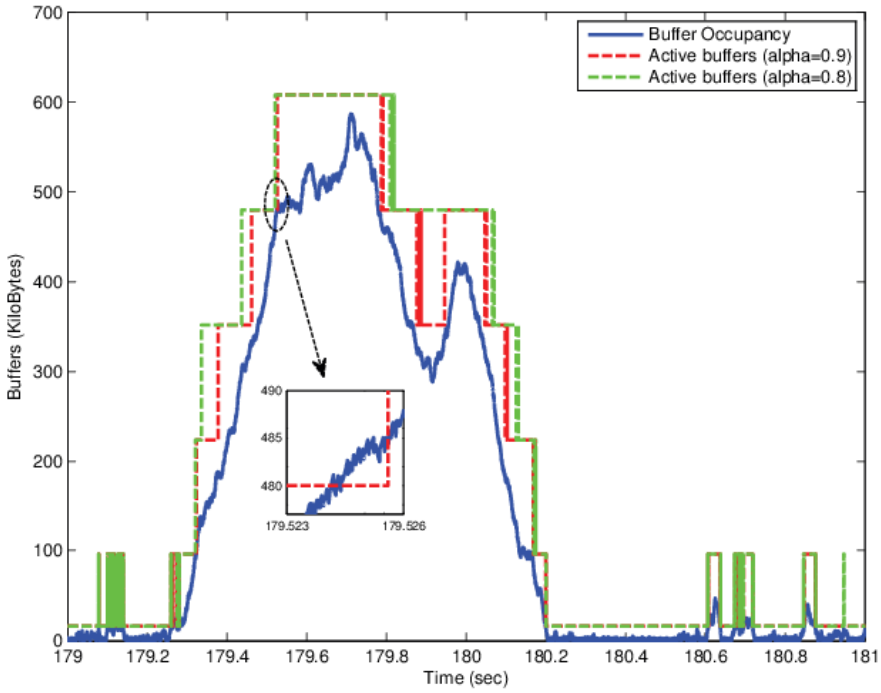
Algorithm

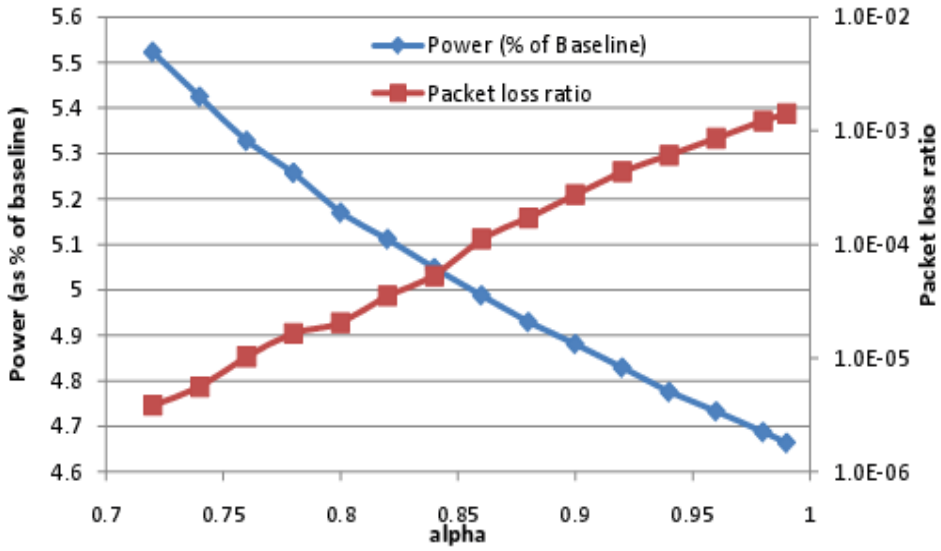
- B_I – amount of on-chip buffers
- B_S and B_D – SRAM and DRAM buffer size
- Q current buffer occupancy
- Control parameter $\alpha \in [0,1)$
- Set buffer capacity B to be between Q and maximum available buffers, i.e.,

$$B = \alpha Q + (1-\alpha)(B_I + B_S + B_D)$$

- $\alpha = 0$ disables algorithm
- Intentionally chosen a very simple scheme and have strived to have **only one** control knob

Evaluation – Internet2 Traffic



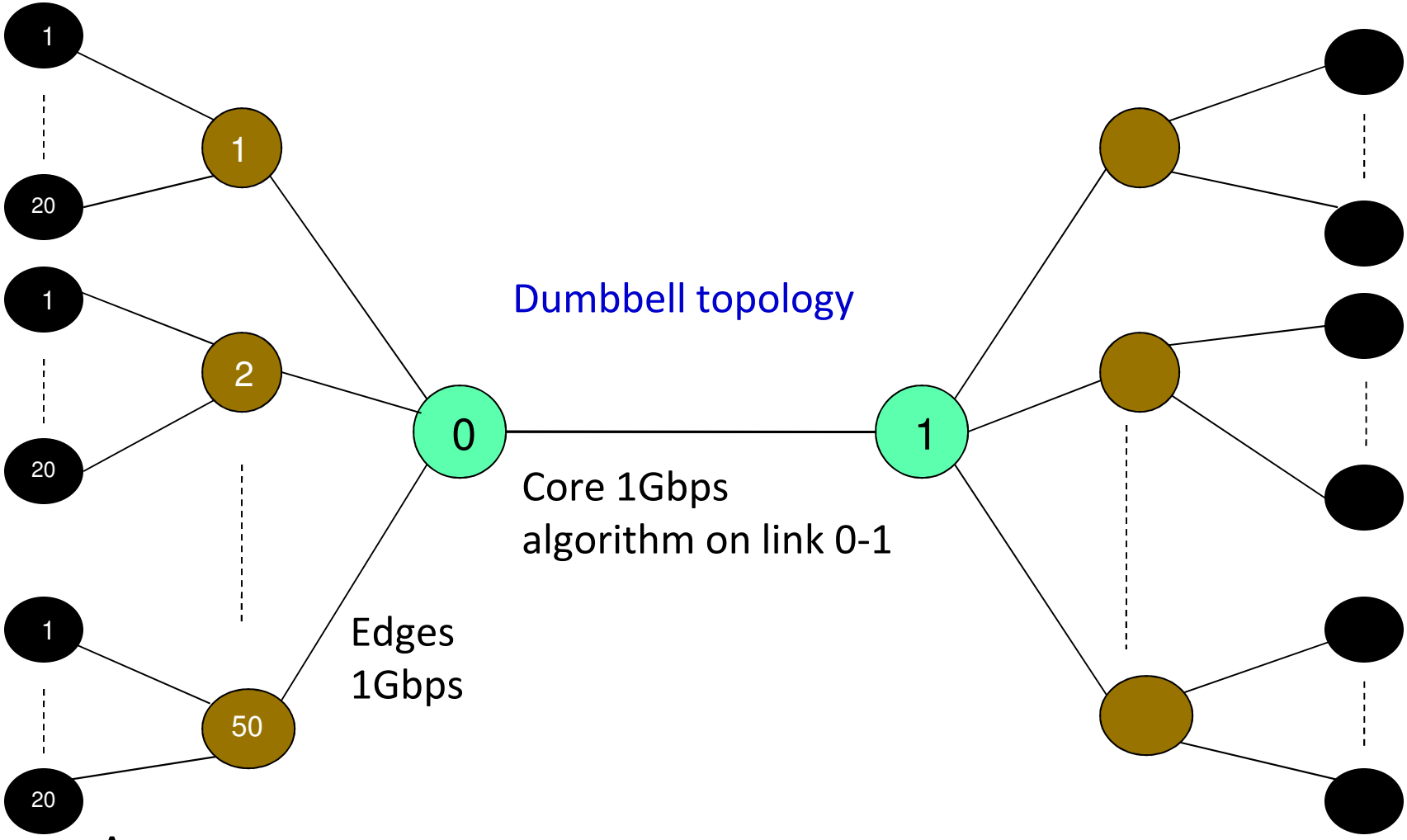


Power versus loss trade-off

- 10 min window of heavy load between Chicago and Kansas
- Assume B_I 16 KB, B_S 80 KB, B_D 512 KB
- For $\alpha = 0.8$, SRAM and DRAM on 2.66% and < 1% of the time
- 95% of off-chip buffering energy saved
- Packet loss 2 in about 100,000

ceet centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

ns2 Simulations with TCP



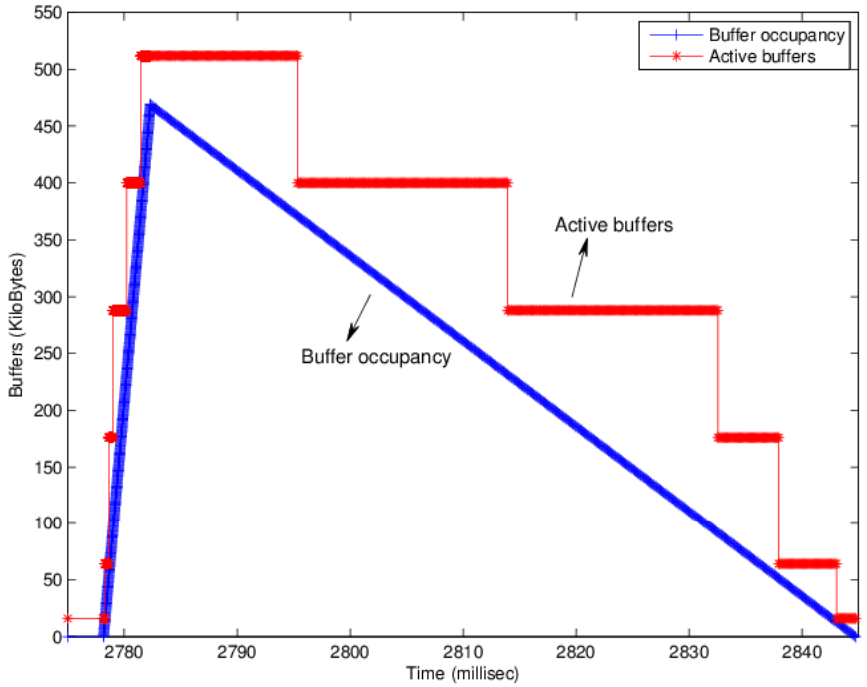
ceet centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

Performance of TCP (Reno) Flows

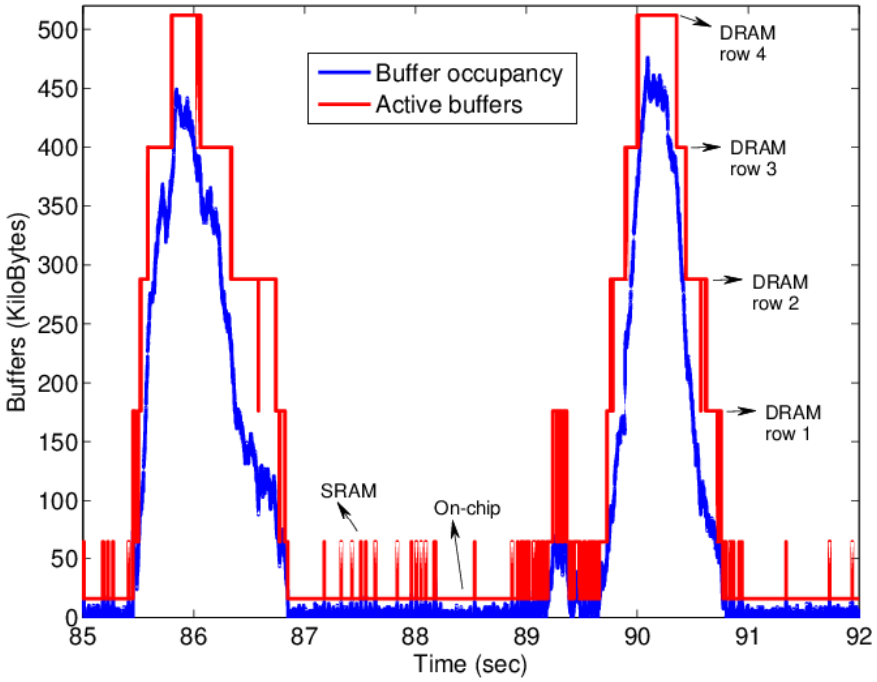
- Mix of long-/short-lived flows, mean RTT 250ms
- Simulation duration over 180s; results for $\alpha = 0.8$

Workload	Load	Time off-chip buffers used	Power saved	Packet loss
Low/Medium	21.5 - 41.1%	0.25%	97%	0
High	59.8%	12.3%	83.4%	10^{-7}
Heavy	78.6%	40%	52.9%	10^{-6}
Very Heavy	90.9%	82%	11.6%	10^{-6}

Demonstration using 1G NetFPGA card



UDP traffic burst



TCP traffic using Iperf

- Validation of hardware implementation with UDP traffic burst
- 150 TCP flows generated using Iperf
  - Algorithm reacts to buffer occupancy in real-time
  - 40% of off-chip buffering energy saved

ceet centre for energy-efficient telecommunications
BELL LABS AND UNIVERSITY OF MELBOURNE

Conclusions

- Buffering consumes around 10% of the power
- A very simple energy saving scheme
  - Only one control knob
  - Hardware changes minimal
  - No changes to network architecture/protocols
  - Can be deployed in routers today
- Large savings across hundreds of line-cards
- Transition path to 2020 line-cards

NetFPGA Implementation

http://netfpga.org/foswiki/bin/view/NetFPGA/OneGig/RouterBufferAdaptation