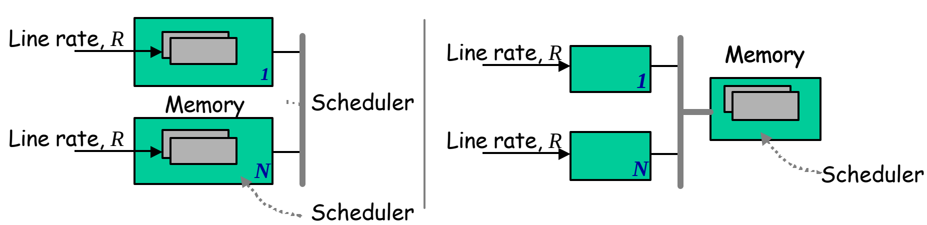
Statistical Analysis of Packet Buffer Architectures

Gireesh Shrimali, Isaac Keslassy, Nick McKeown
E-mail: gireesh@stanford.edu

Packet Buffering



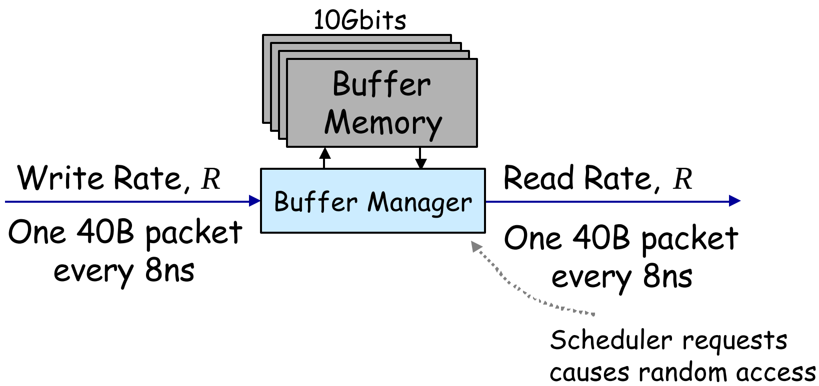
Input or Output Line Card

Shared Memory Buffer

- **Big:** For TCP to work well, the buffers need to hold one *RTT* (about 0.25s) of data.
- **Fast:** Clearly, the buffer needs to store (retrieve) packets as fast as they arrive (depart).

An Example

Packet buffers for a 40Gb/s line card



Problem is solved if a memory can be (random) accessed every 4 ns and store 10Gb of data

3

Key Question

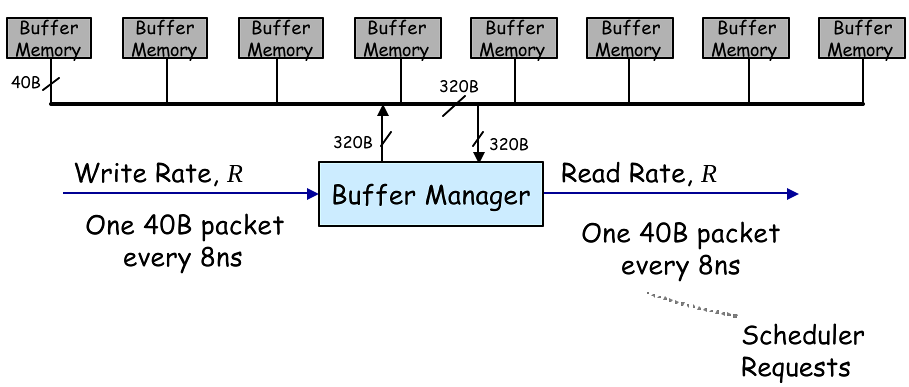
How can we design high speed packet buffers from commodity available memories?

4

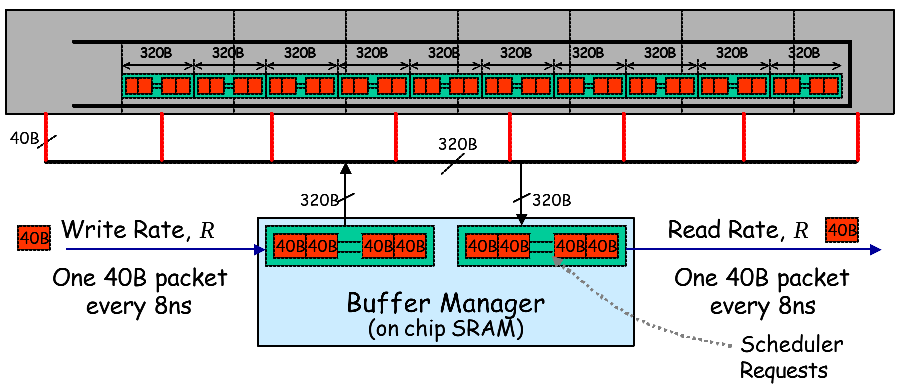
Available Memory Technology

- Use SRAM?
  - \+ Fast enough random access time, but
  - \- Too low density to store 10Gbits of data.
- Use DRAM?
  - \+ High density means we can store data, but
  - \- Can't meet random access time.

Can't we just use lots of DRAMs in parallel?



Works fine if there is only one FIFO queue



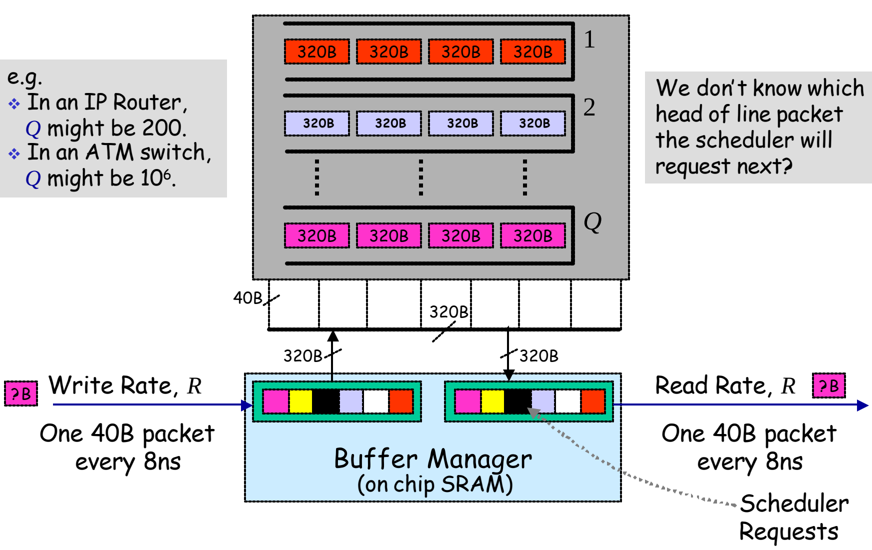
Aggregate 320B for the queue in fast SRAM and read and write to all DRAMs in parallel

7

In practice, buffer holds many FIFOs

e.g.
- In an IP Router, Q might be 200.
- In an ATM switch, Q might be 10^6.

We don't know which head of line packet the scheduler will request next?

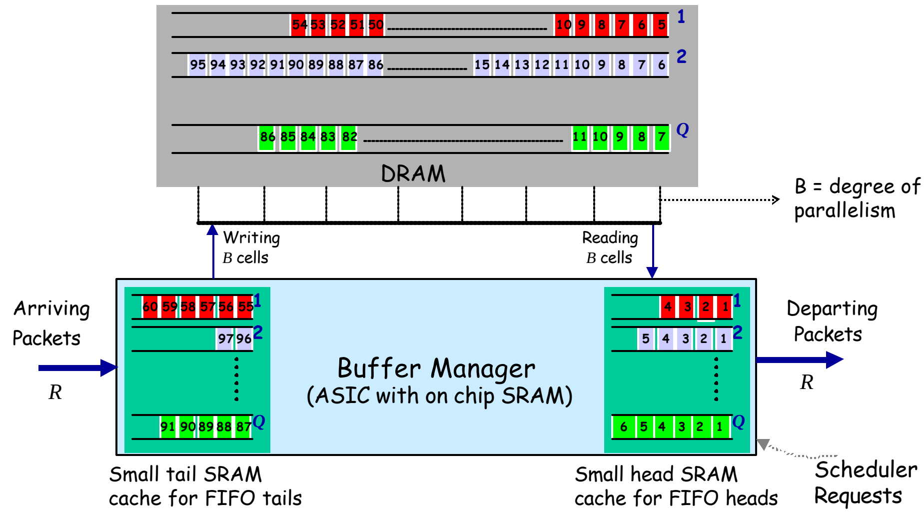


8

Parallel Packet Buffer

Hybrid Memory Hierarchy

Large DRAM memory holds the body of FIFOs



Objective

- Would like to

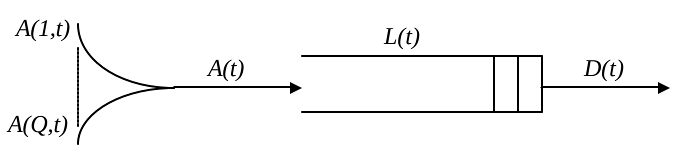
  Minimize the size of SRAM while providing reasonable guarantees
- So, ask the following question

  If the designer is willing to tolerate a certain drop probability then how small can the SRAM get?

Memory Management Algorithm

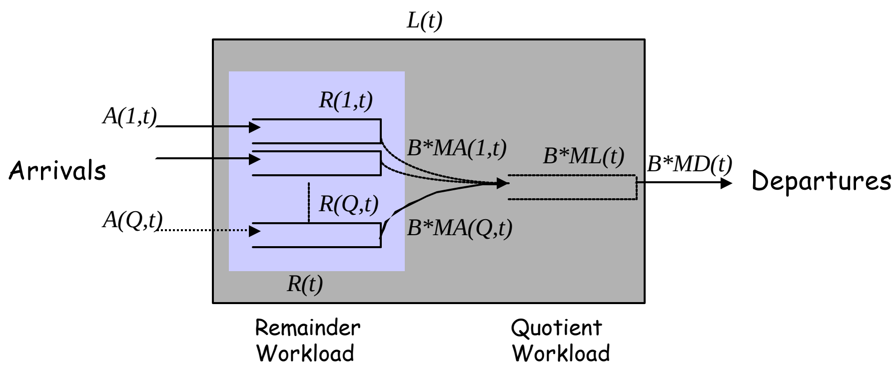
- Algorithm: At every service opportunity serve a FIFO from the set of FIFOs with occupancy greater than or equal to B
  - B-work conserving - thus minimizes SRAM size
  - Round-robin performs as well as largest FIFO first
- Some definitions
  - FIFO occupancy counter: L(i,t)
  - Sum of occupancies: L(t)

Model



- Model SRAM as a queue
  - Arrival process A(t) superposition of Q sources A(i,t) with rates [illegible]
  - Deterministic service at rate 1
  - Queue is stable, i.e., $\sum l_i \leq 1$
- Approach: assume A(i,t) are independent of each other
  - Step 1: Analyze for IID sources
  - Step 2: Show that the IID case is the worst case
- Tools used
  - Analysis in continuous time domain
  - Use $P_{overflow}(S) = P(L > S)$

Fixed Batch Decomposition



13

Assumptions

A(i,t) are

1. independent of each other
2. stationary and ergodic
3. simple point processes

14

PDF of SRAM Occupancy

- Theorem:

  The quotient workload and the remainder workload are independent of each other
- Thus

  The distribution of SRAM occupancy is the convolution of the distributions of the quotient and remainder workloads

PDF of Remainder Workload

- Theorem:

  For large Q, PDF of remainder workload approaches a Gaussian distribution with mean $Q(B-1)/2$ & variance $Q(B^2-1)/12$
- Intuition:

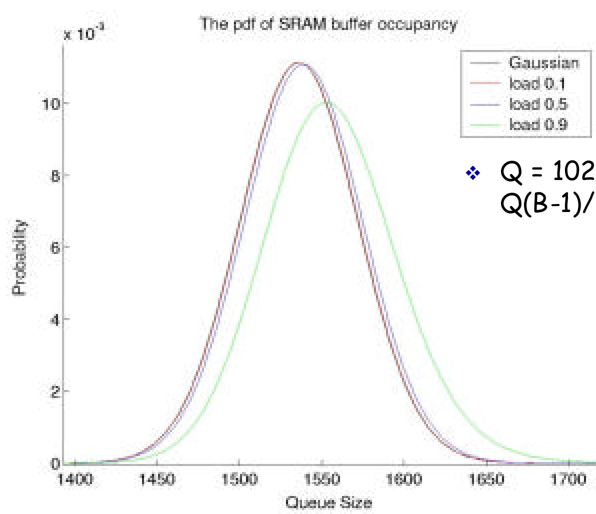
  Application of central limit theorem

PDF of Quotient Workload

- Theorem [Cao, Ramanan INFOCOM 2002]: For large Q, the behavior of the quotient FIFO approaches the behavior of an M/D/1 queue with the same load
  - Numerical solution through recurrence relations
- Depends only on load
  - Independent of Q and B
  - Close to impulse at low loads

17

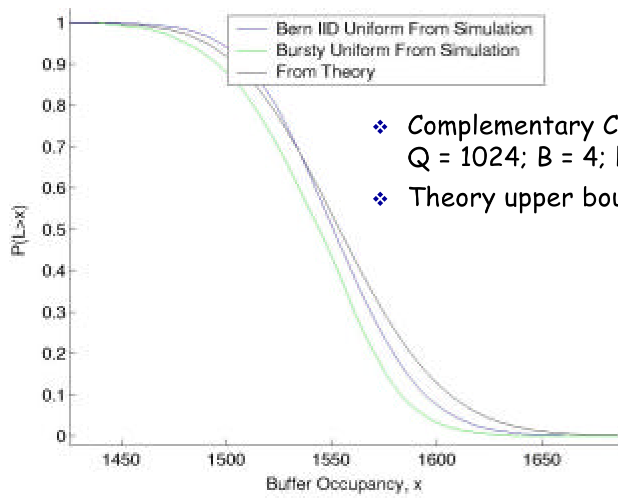
PDF of Buffer Occupancy



- Q = 1024; B = 4; Q(B-1)/2 = 1536

18

Simulations (load=0.9)



- Complementary CDF for Q = 1024; B = 4; load = 0.9
- Theory upper bounds simulations

Conclusions

- Established exact bounds relating the drop probability to the SRAM size
- Model may be applicable to many queueing systems with batch service
- Compared to deterministic guarantees ([Iyer, McKeown HPSR 2001]), an improvement by at most a factor of two
- O(QB) a hard lower bound for this architecture