



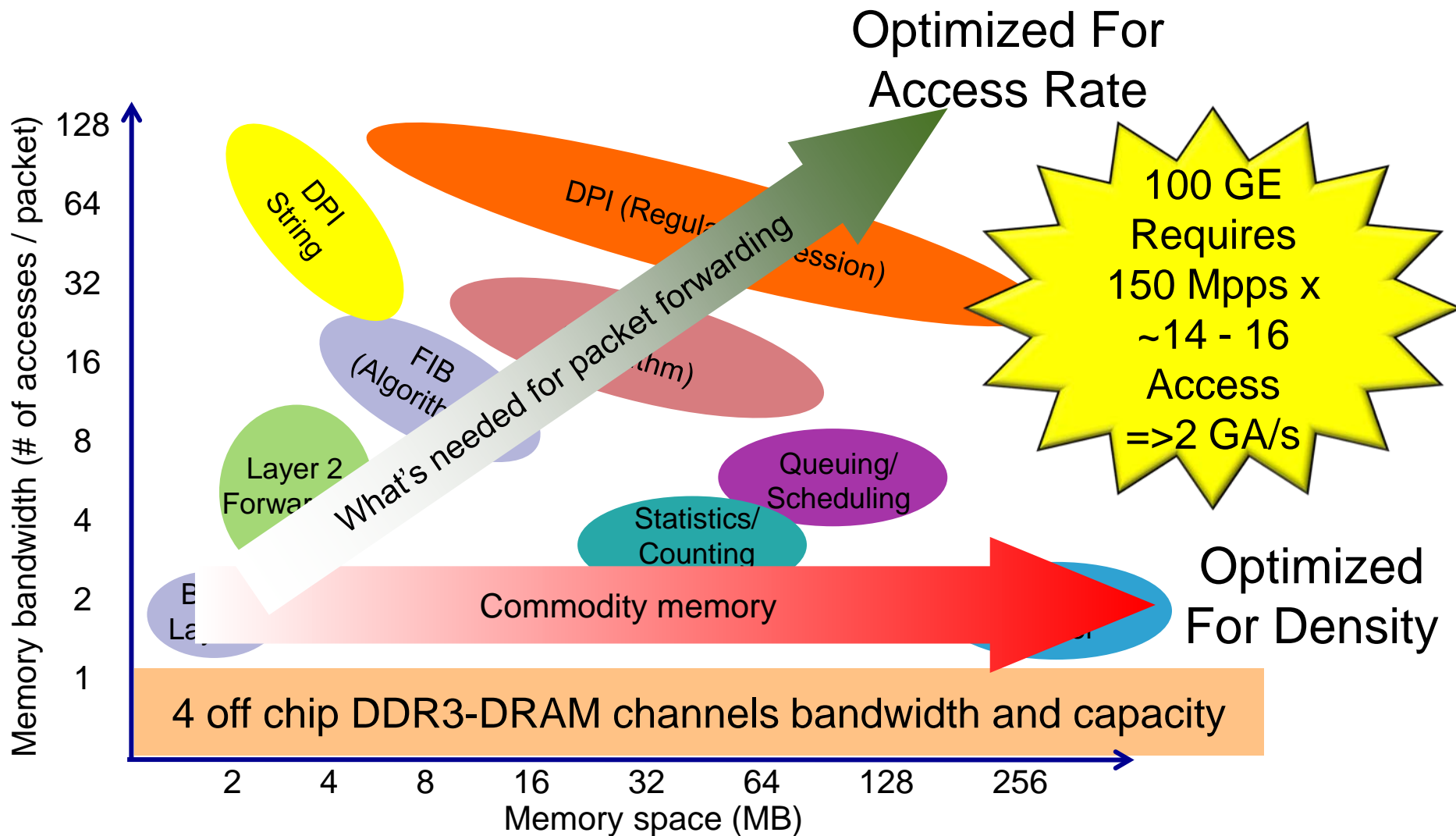
# **Bandwidth Engine<sup>®</sup> Serial Memory Chip Breaks 2 Billion Accesses/sec**

**Michael J. Miller**

**VP Technology Innovation & Systems Applications , MoSys**

**August 2011**

# Network Memory Access Requirements



Source: HotChips 2010 Huawei

## ❖ Networking memory characteristics

- Synchronous interface
- Modulo x9 accesses
- Small quanta (36b to 72b per access)
- High access rate

## ❖ Challenge: Create a 2+ GigaAccess networking memory device

- High access availability (4x existing devices)
- Minimize system power per access
- Utilize existing electrical interfaces
- Support 100G designs and scale to 400G

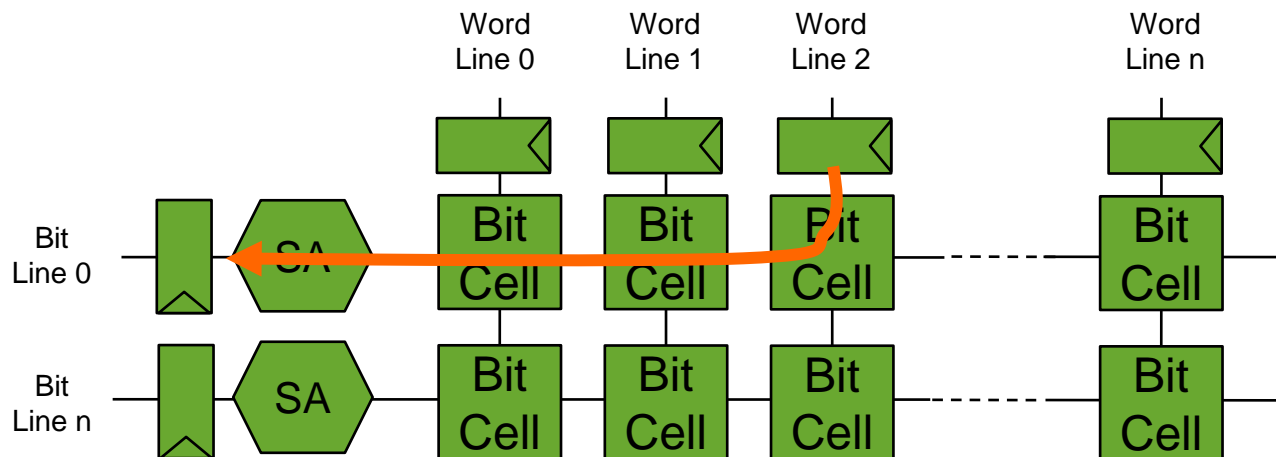
# Definition: GA/s and tRC

## ❖ GA/s is the number of billions of unique access to memory

- Access is a unique read or write “Transaction”
- Depends on: the bandwidth, the cycle rate (tRC) and the transfer size...
- Maximum GA/s = (I/O bandwidth: Gbps) / (Access size: bits)
- Sustained GA/s = (# simultaneous bank accesses) x (memory cycle rate: 1/tRC)

## ❖ tRC is the amount of time to cycle a memory bit for read or write

- Depends on: bitline RC & power/area allotted to Sense Amp
- Bitline RC  $\cong$  (# bit cells) x (RC per bit cell)



## ❖ Memory Interface

- Parallel interfaces are becoming bottlenecks, scaling slow down
- Serial interconnect is already everywhere except memory

## ❖ Memory Core Performance

- Cycle the memory faster
  - Vs. Power/Density/Refresh/... tradeoffs
- 400GE -> .8 ns or 1.2 GHz memory cells

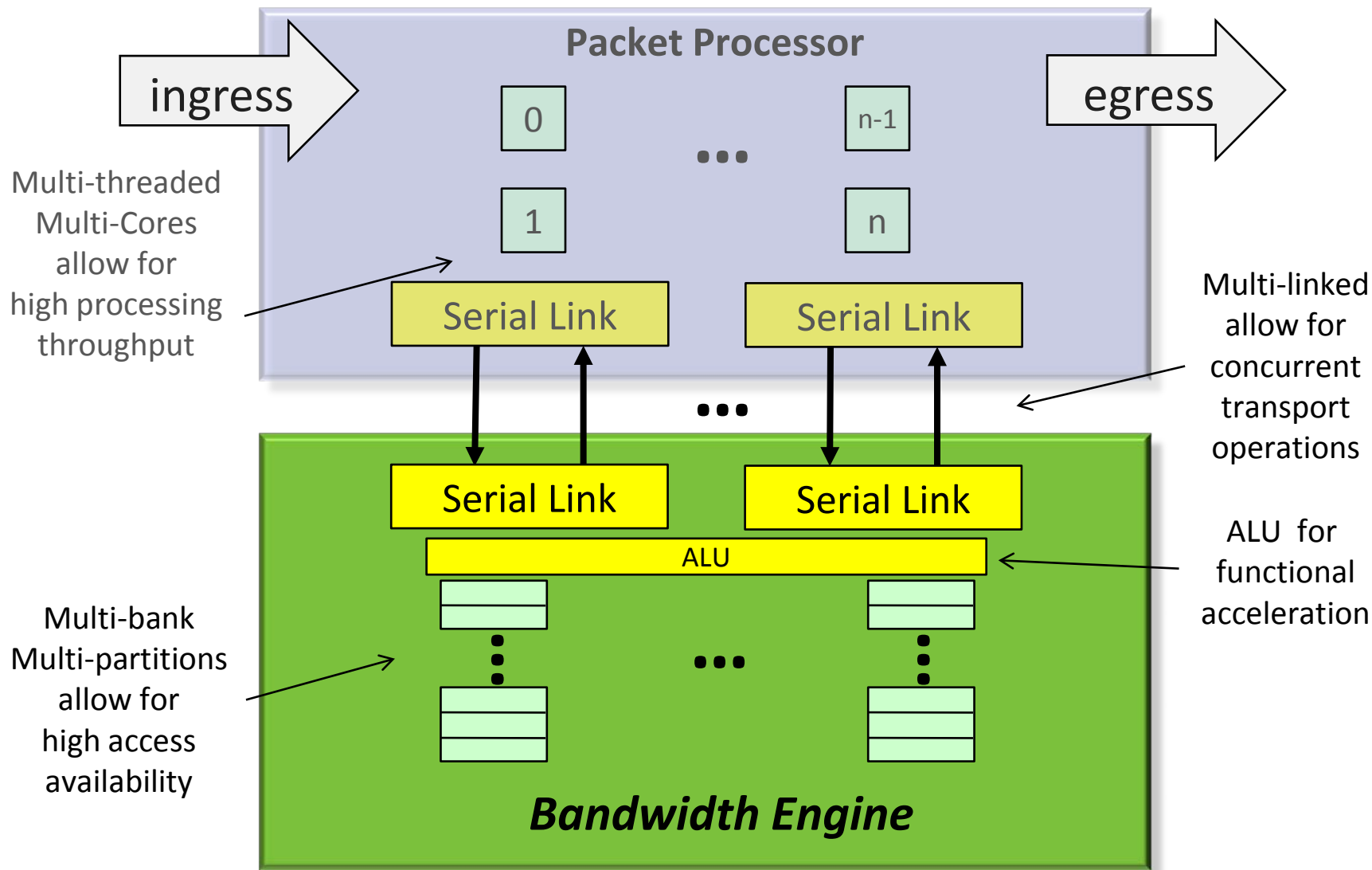
## ❖ Memory Architecture

- Run multiple banks in parallel
- Use “Round Robin/Ping Pong” algorithms scale up effective access rate
  - $t_{RC}/n$  but also Mbits/n

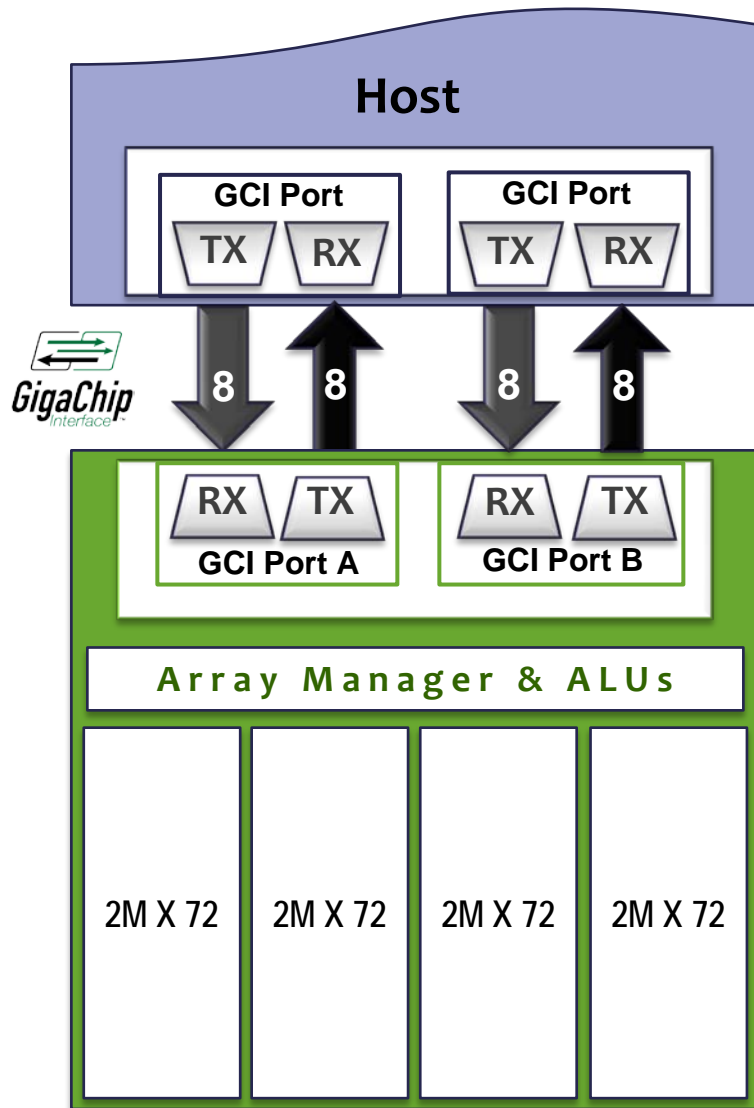
## ❖ Or a combination thereof

- There will have to be tradeoffs inevitably

# Multi Core => Multi-Partition & Multi-bank



# Bandwidth Engine IC Sampling Now



## ❖ Breakthrough Performance

### ...4X Throughput of RLDRAM

- 2.75GA : >2 billion reads/sec (2 GA), 1B writes
- 72b words each access
- 15.9 ns roundtrip latency
- Up to 16, 10G CEI 11+ serial lanes
- Macro operations (RMW, Inc/Dec..)
- <7W worst case system power

## ❖ High Density

### ...4-8X Density of QDR SRAM

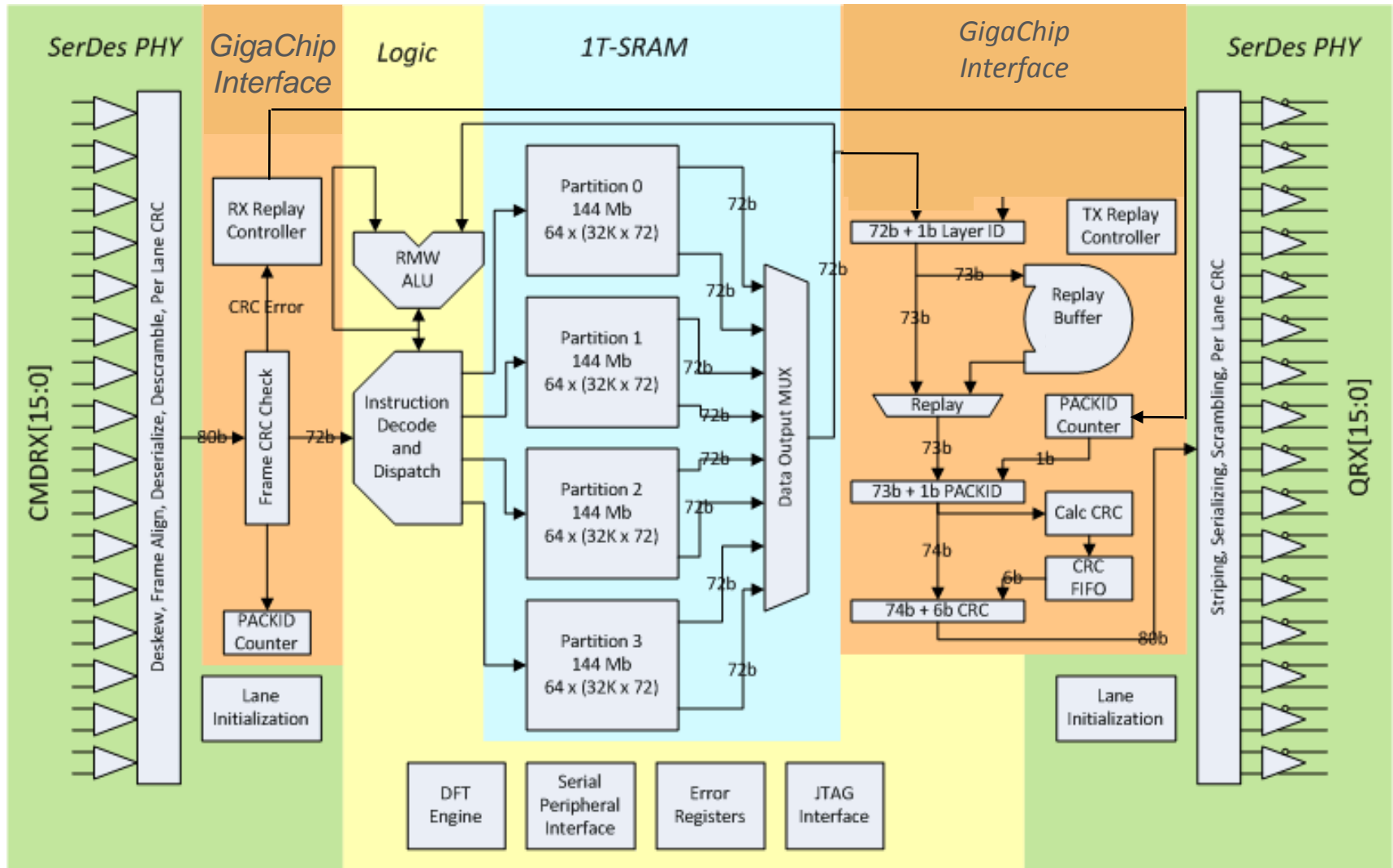
- 576Mb 1T-SRAM
- 3.9ns bank tRC (1T-SRAM performance)

## ❖ High Reliability

### ...70X better SER than 6T embedded SRAM

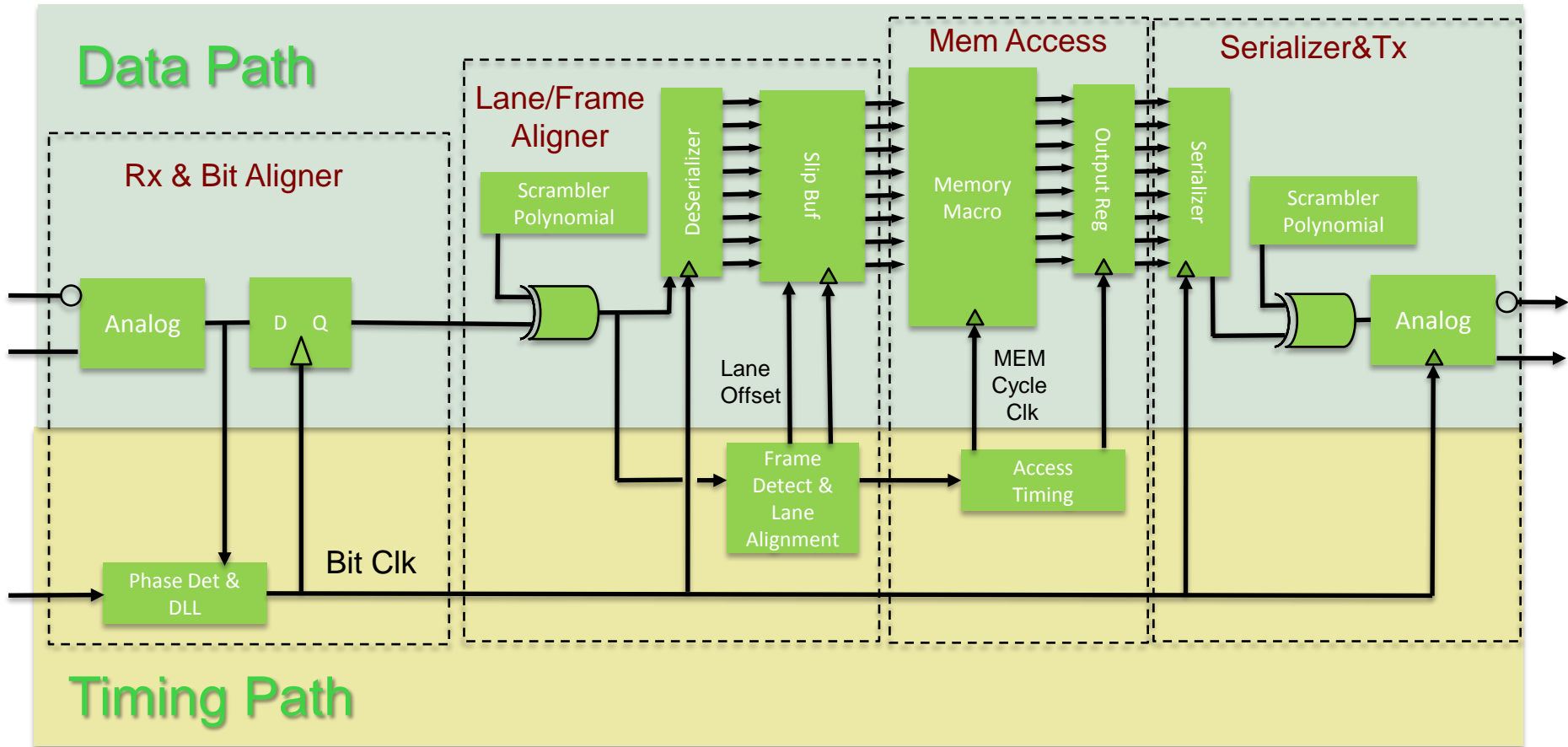
- Memory core: < 10 FIT/Mb native
- Interface: < 1 FIT

# Bandwidth Engine Block Diagram

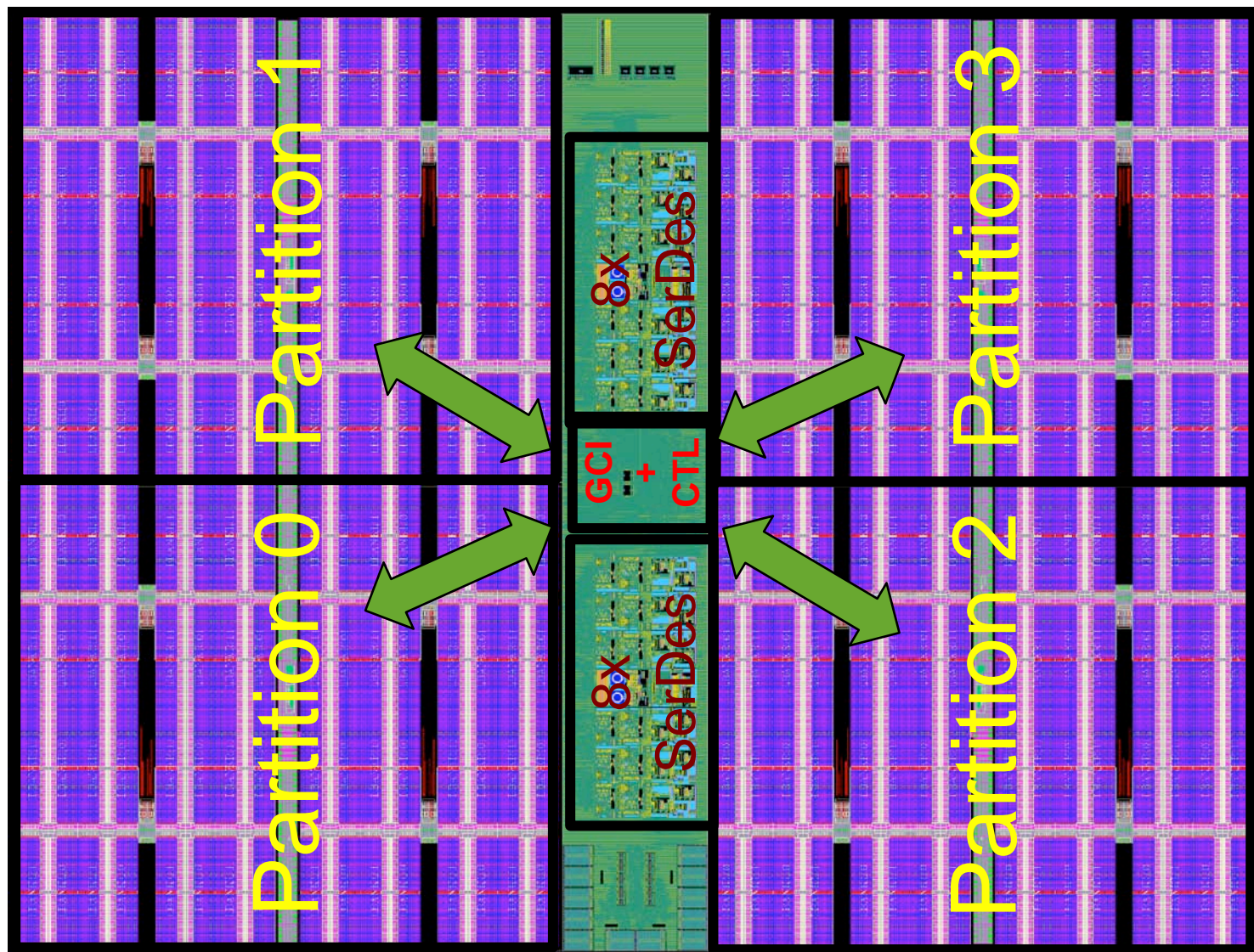


# Conceptual Timing & Data Access Control

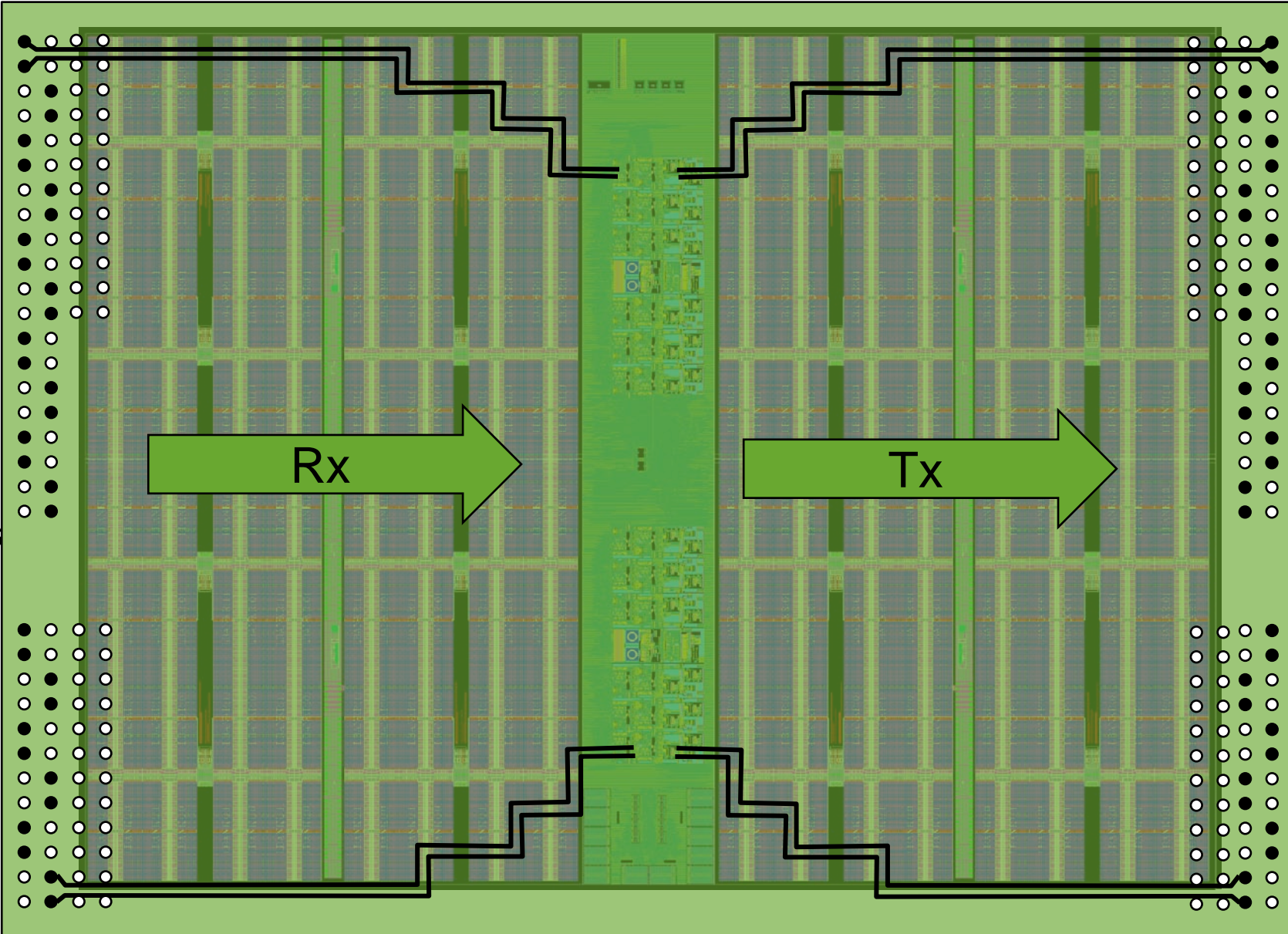
## Read Latency of ~16ns



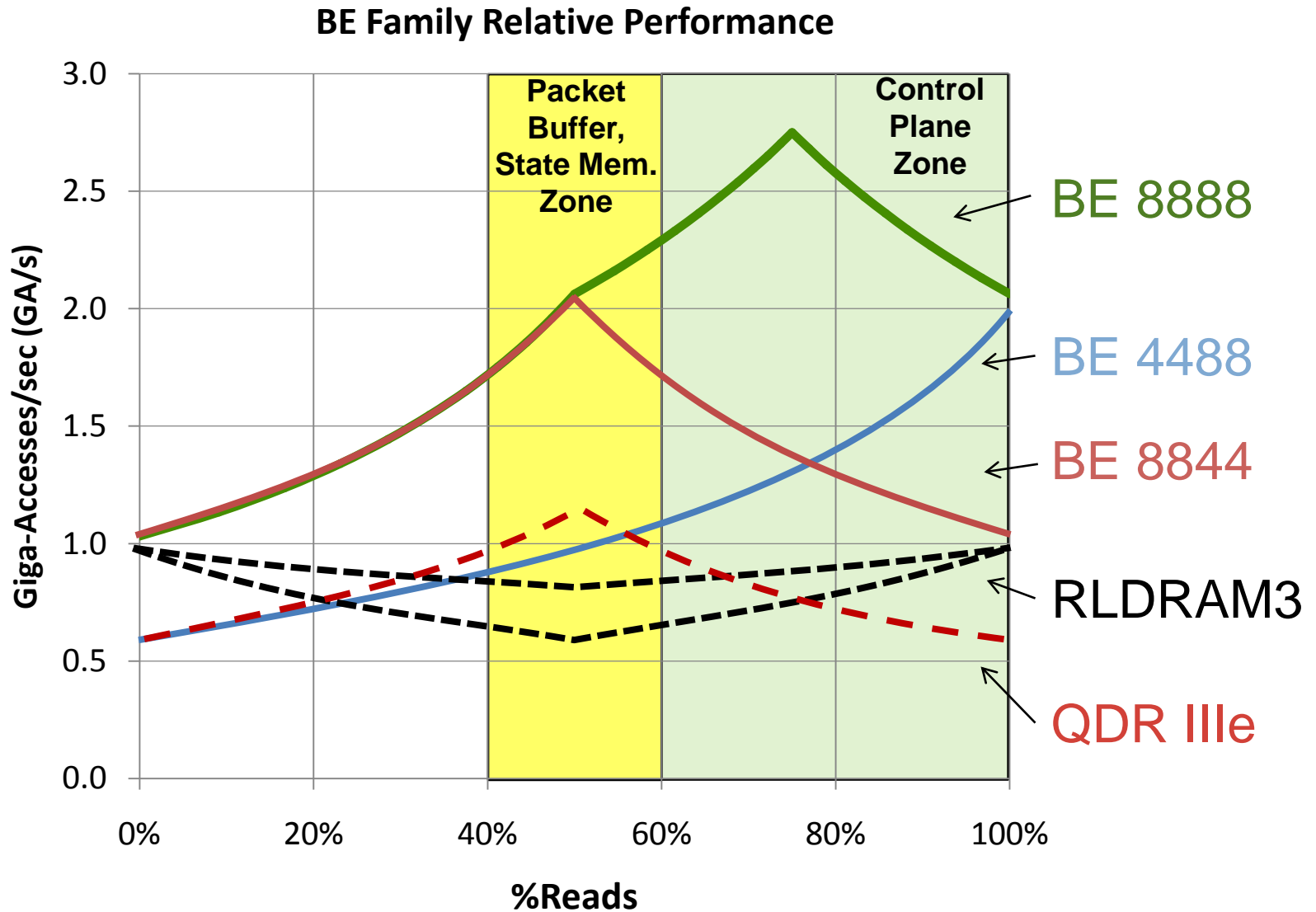
# Chip Layout Balances Memory Access Paths



# Package Layout Minimizes Rx/Tx Xtalk

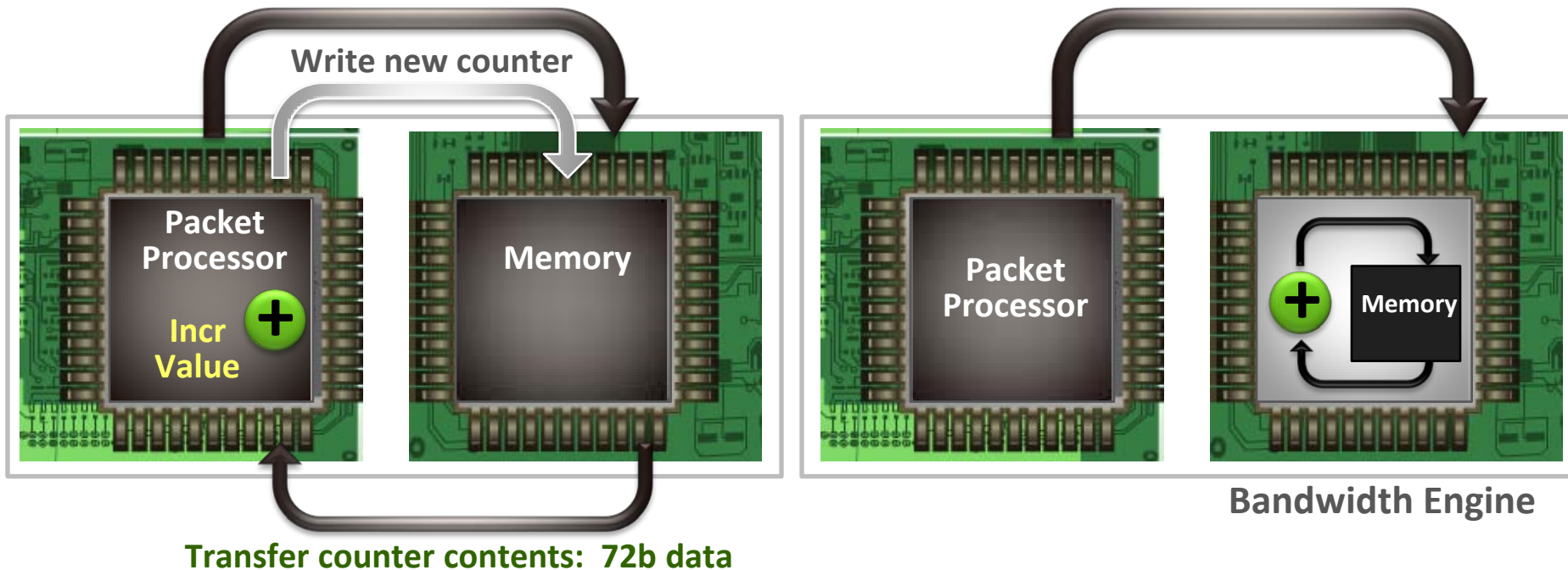


# Optimizing Read/Write Bandwidth



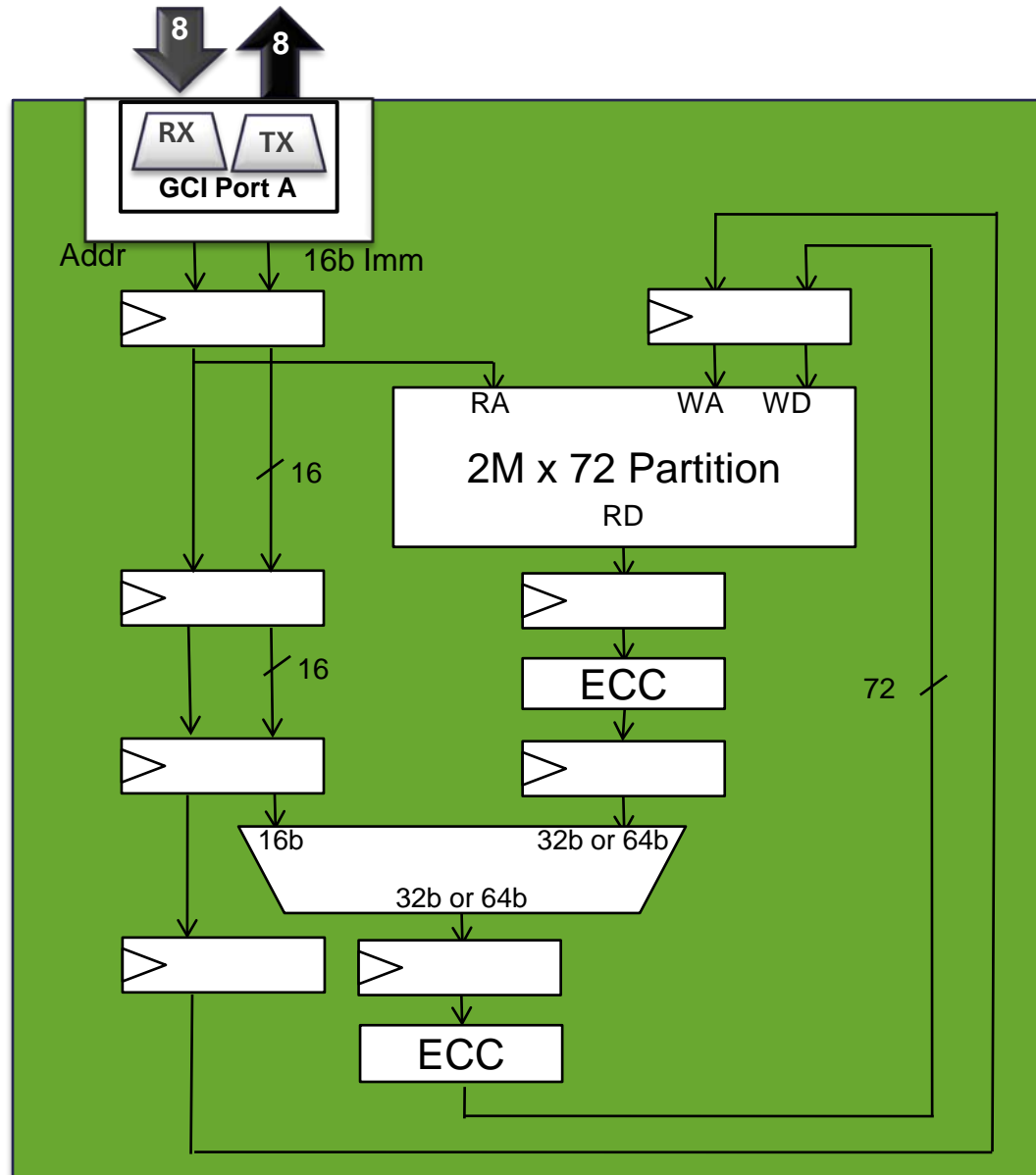
# Bandwidth Engine On-chip Macro Operations

## Initiate Read of statistics counter



- ❖ 2X or Better I/O Performance, saves Power & Pins
- ❖ BE-1: 16b, 32b and 64b Add/Subtract
  - 4 SerDes lane BE-1 => 4M flows @ 2 counters per flow @ 250 Mpps,
- ❖ Possible Future Macro Operations
  - Data manipulation: Fully flexible increment/decrement, semaphore (R-M-W)
  - Pointer indirection for data structure walking
  - Data packing/unpacking

# Four Stage Partition Macro Op Pipeline



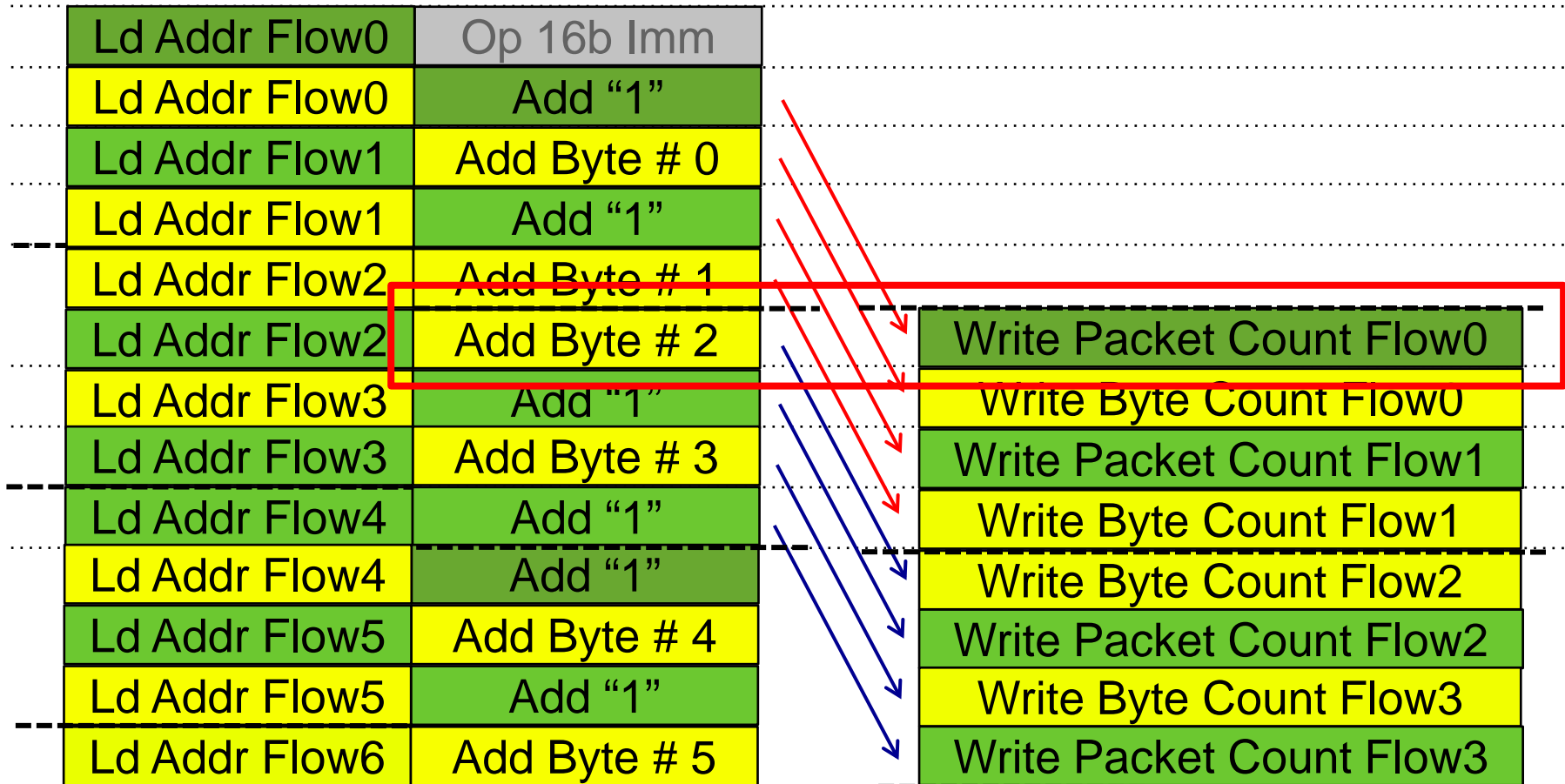
Note:  
Conceptual  
pipeline

# Avoiding Collisions

258MHz => 129Mpps Per Partition x 4 => 516 Mpps Per BE

## GCI Command Interface

## Write Port





**Tradeoff**  
**“tRC” vs Density**  
**Minimizing tRC can save power**

# Density vs tRC For Control Plane

## ❖ tRC is the amount of time to cycle a memory bit for read or write

- Depends on: bitline RC & power/area allotted to Sense Amp
- Bitline RC  $\cong$  (# bit cells) x (RC per bit cell)

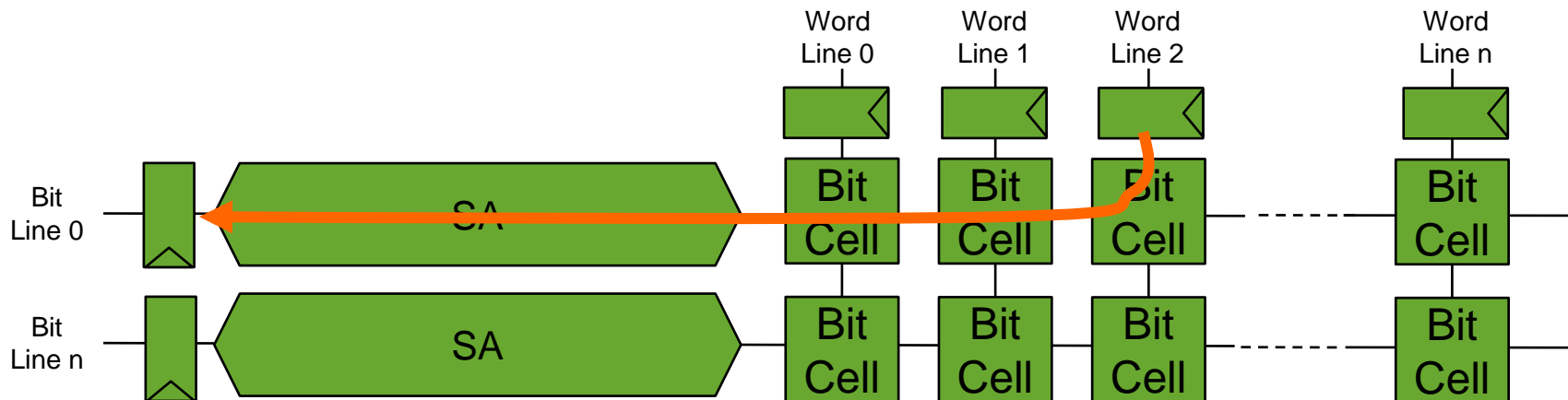
## ❖ Density is a function of:

- Size of bit cell  $\Rightarrow$  Function of process node & drive
- Ratio of #bit to Sense Amp  $\Rightarrow$  Length of bitlines

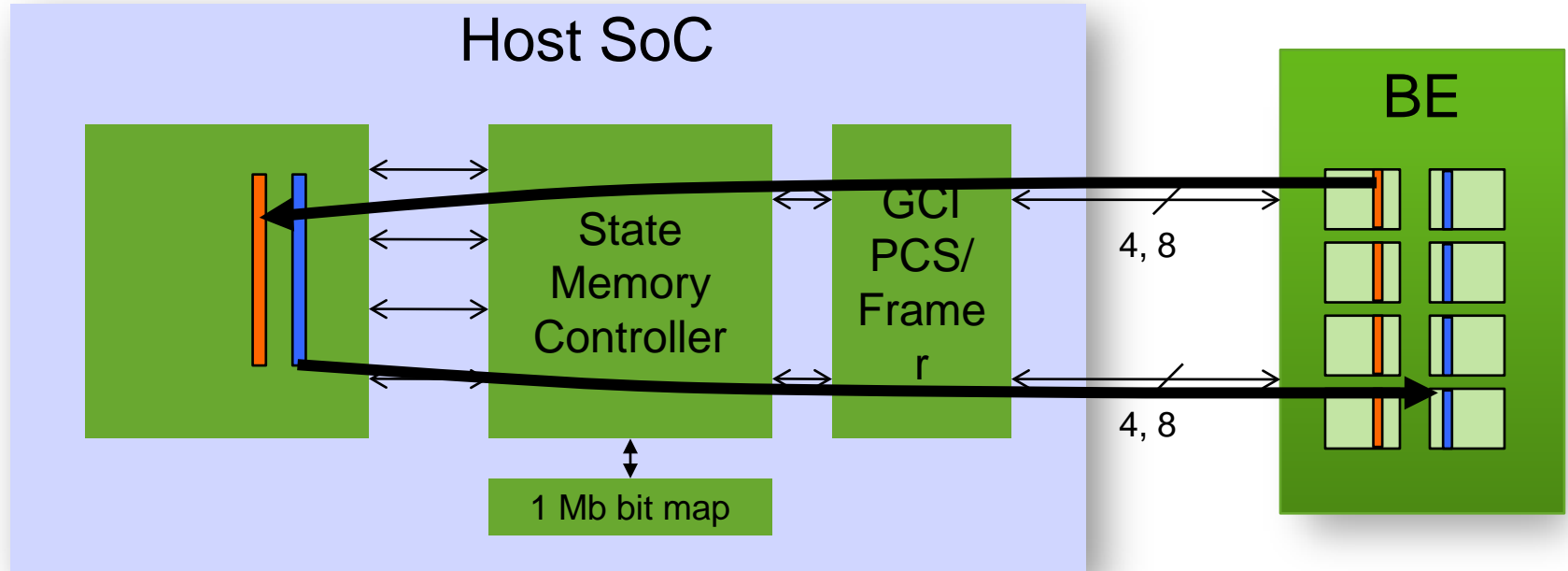
150Mpps  
 $\Rightarrow$  6.7ns  
Optimal  
tRC  
 $\sim$  3.3ns

## ❖ Proportional relation between density and tRC

- Sense Amp area  $\sim$  10x the area of 1 bit cell



# Speed Up Using Multi-Bank



## ❖ Ping Pong Algorithm for 2x throughput

- Read and write in same  $3.9\text{ns } t_{RC}$  cycle  $\rightarrow$   $1.9\text{ns}$  effective  $t_{RC}$
- Read gets priority in case of bank conflict
- Read from bank with most recent data according to bit map
- Write to the other bank and update bit map
- Bit map keeps record of most recent data

# Effective Table Size Using Multiple Copies



Case Study assumptions @ 100GE:  
Read, modify & write 288b every packet time  
(Packet count, Byte Count, Next schedule, etc.)

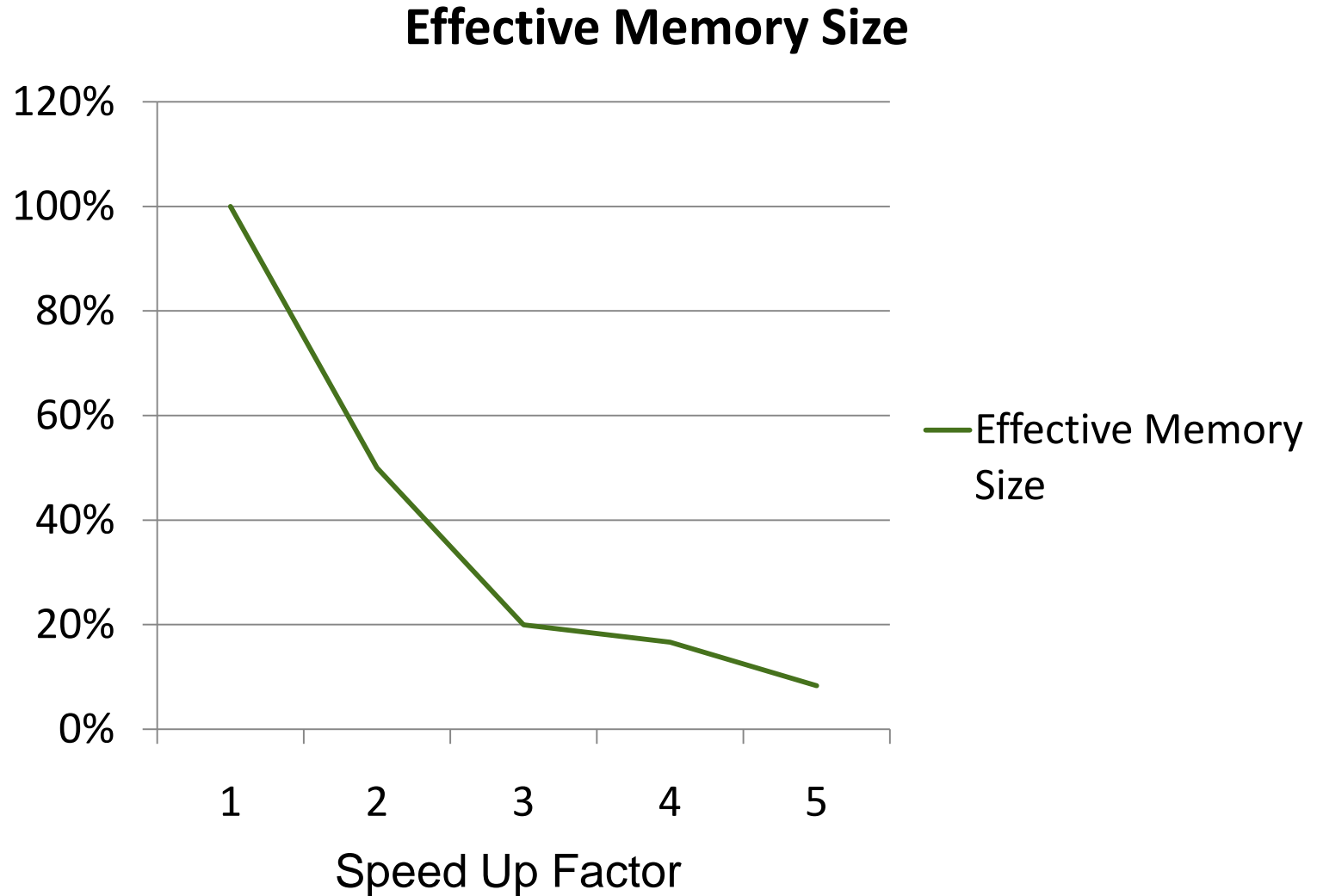
Packet arrival  
period in ns: 6.67

Record entry  
size bits: 288

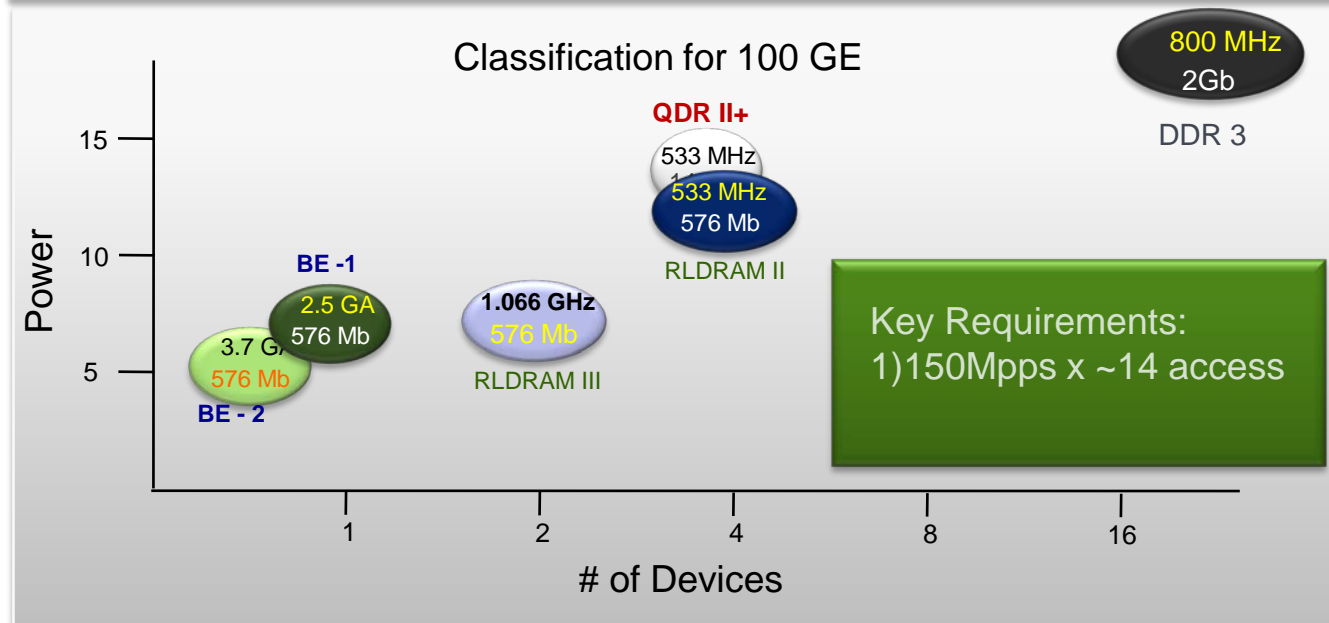
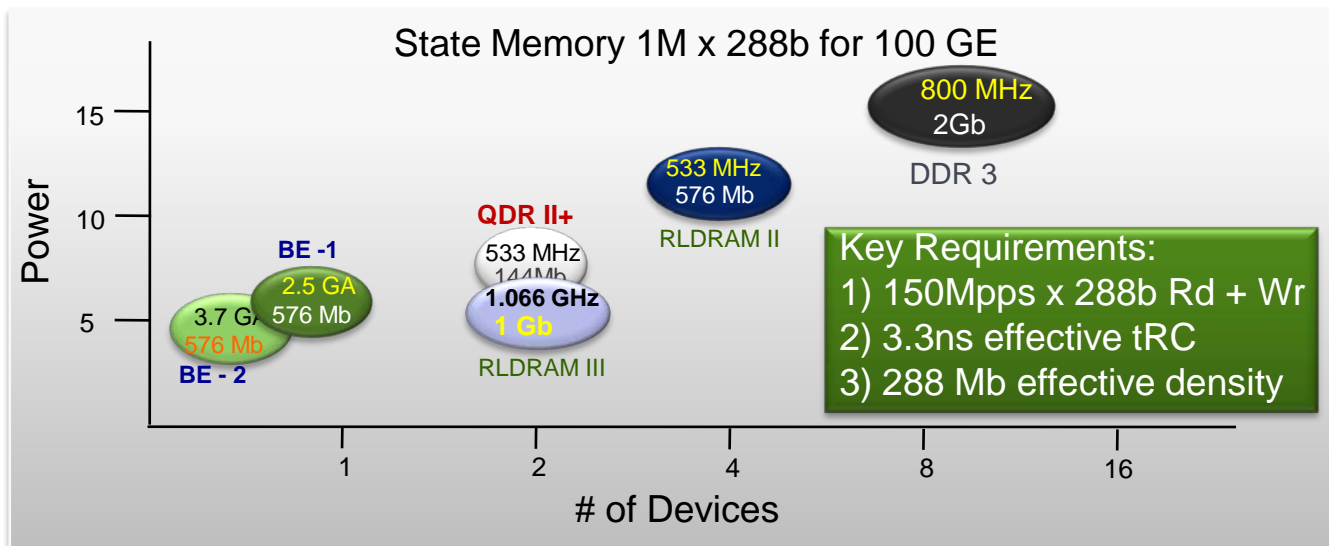
Device	Device Size in Mb	Rd/ Wr $t_{RC}$	Required Speedup	# Banks Per Entry	# Updates per $t_{RC}$	Table Size in M
QDR	144	2	1	1	0.5	0.50
BE	576	3.9	2	2	1	1.00
BE 2	576	3.1	1	1	0.5	2.00
RLDRAM II	576	15	5	9	4.5	0.22
RLDRAM III	576	10	3	5	1.5	0.4
RLDRAM III	1152	10	3	5	1.5	0.8
DRAM	2048	45	14	56	28	0.13

Source for speed up ratio: PhD Dissertation: "Load Balancing & Parallelism for the Internet"  
Stanford University, Sundar Iyer

# Effective Memory size vs Speed Up



# Comparing Performance and Power for 100G



## ❖ Serial I/O

- Reuse the same electrical I/O as network interfaces
- Scales same as network interfaces
- Room for improvement on very short reach power

## ❖ Multi-Bank Multi-Partition

- Increase access availability

## ❖ Optimize for cycle time

- Achieves better system density for networking applications

## ❖ Onchip ALU + Macro Operations

- Minimize I/O requirements for commands & data
- Lower pin counts & system power



**Thank You**

**Michael J. Miller**

**VP Technology Innovation & Systems Applications , MoSys**