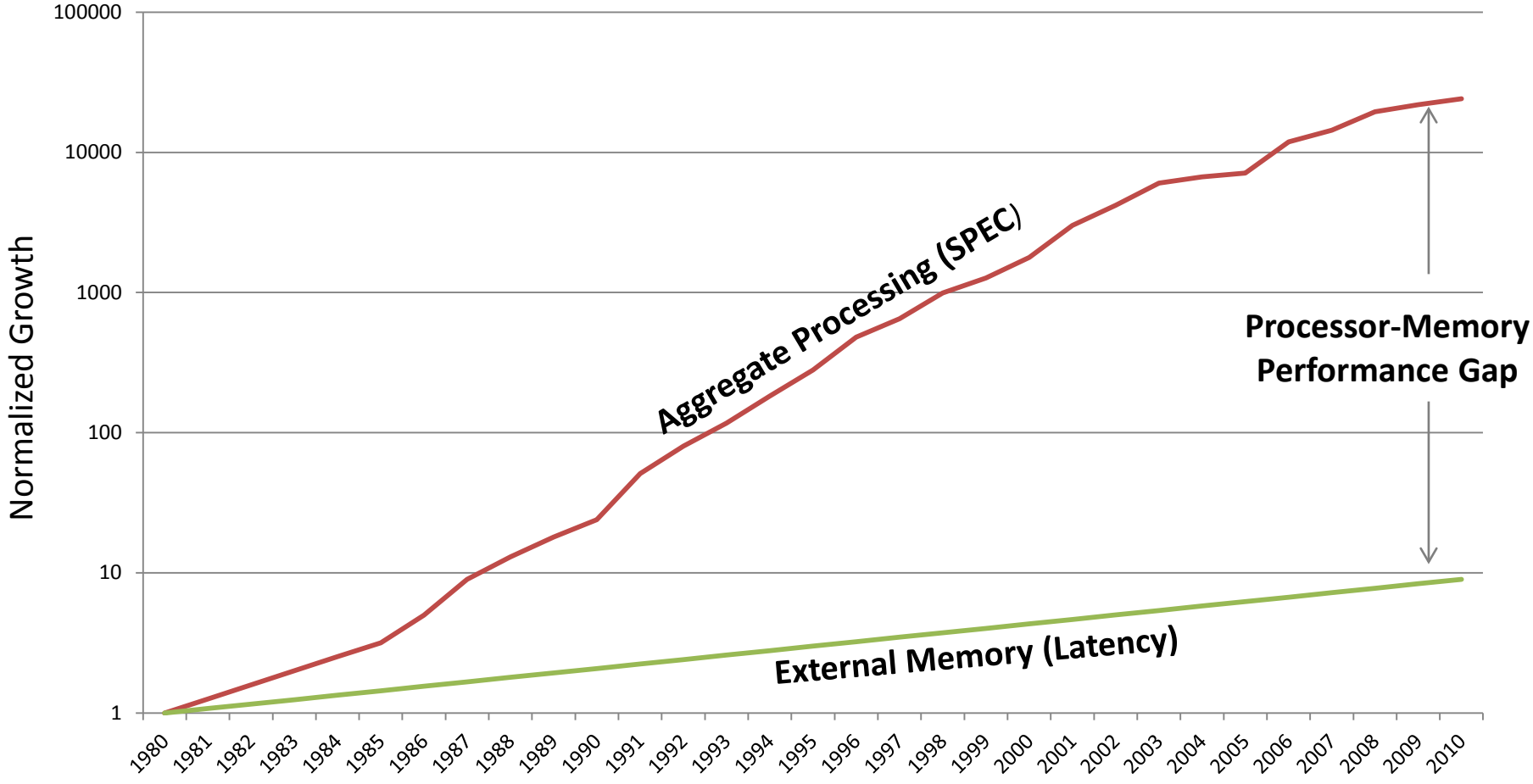
MEMOIR SYSTEMS

Algorithmic Memory – An Order of Magnitude Increase in Next Generation Embedded Memory Performance

Part I: Memory Performance
Part II: Other Benefits and Applications

Sundar Iyer – Co-founder and CEO, Memoir Systems

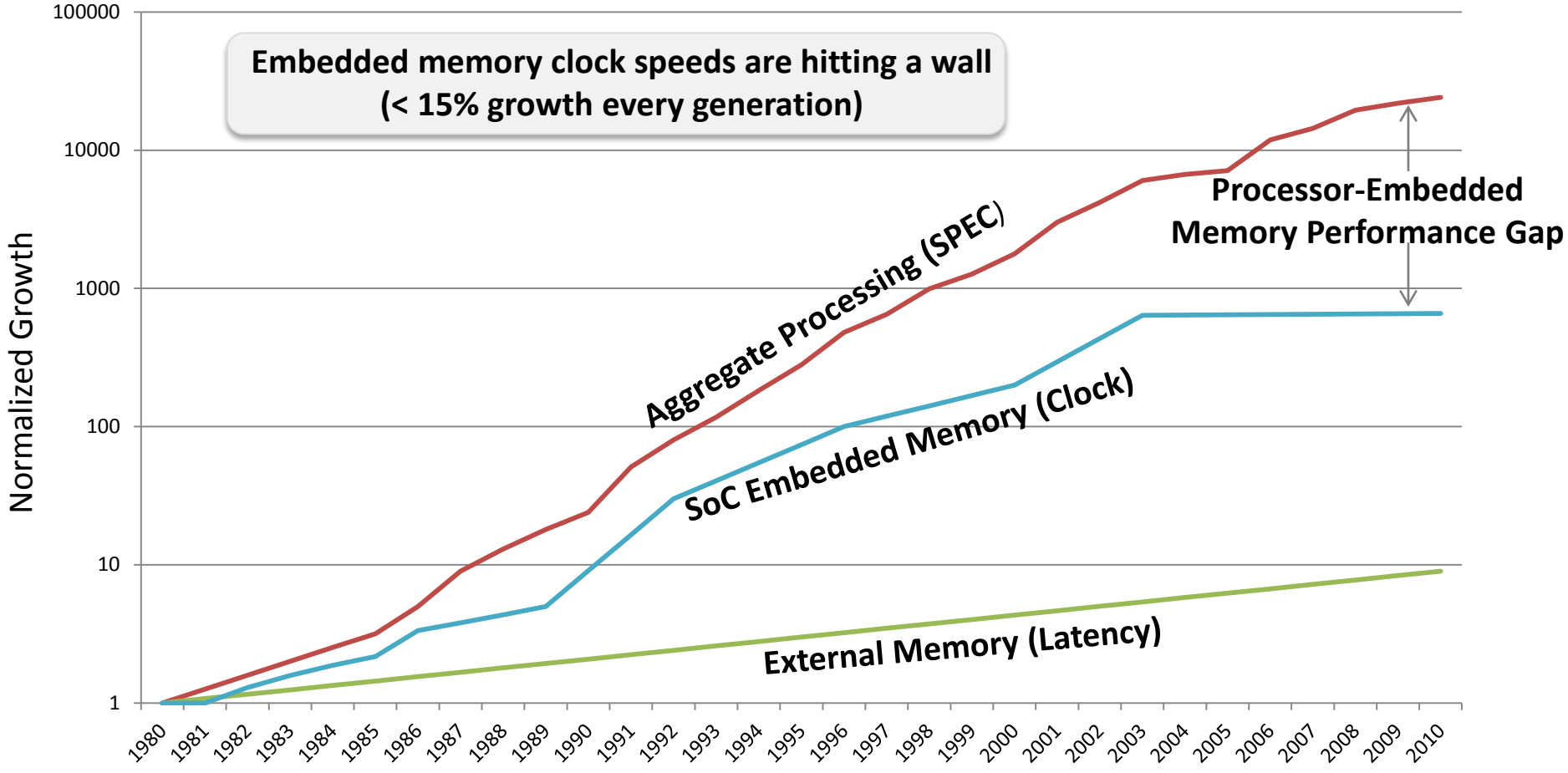
Problem: Processor-External Memory Performance Gap**



Solution: System On Chip (SoCs)

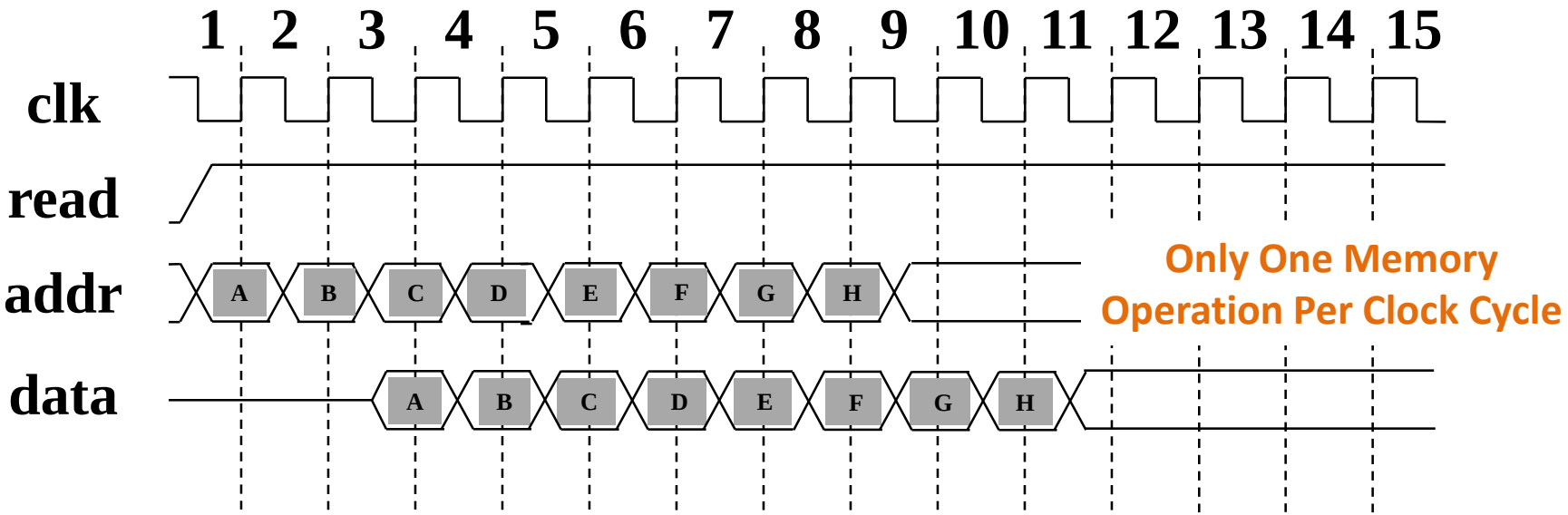
***Source: Hennessy Patterson, "Computer Architecture," 5th Edition*

New Problem: Processor-Embedded Memory Performance Gap



New Performance Gap: ***Processor/Aggregated Processors Access Embedded Memory At A Rate Faster Than It Can Handle. The Gap is Getting Worse …***

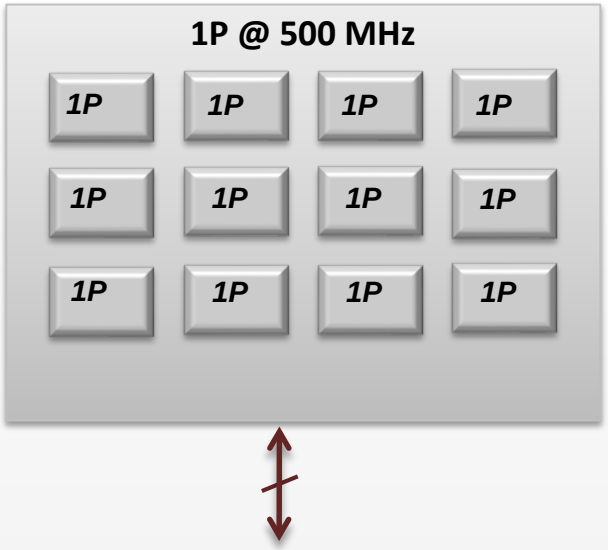
Why is Embedded Memory Slow?



How can we Increase MOPS (Memory Operations Per Second) Without Increasing Memory Clock Speed?

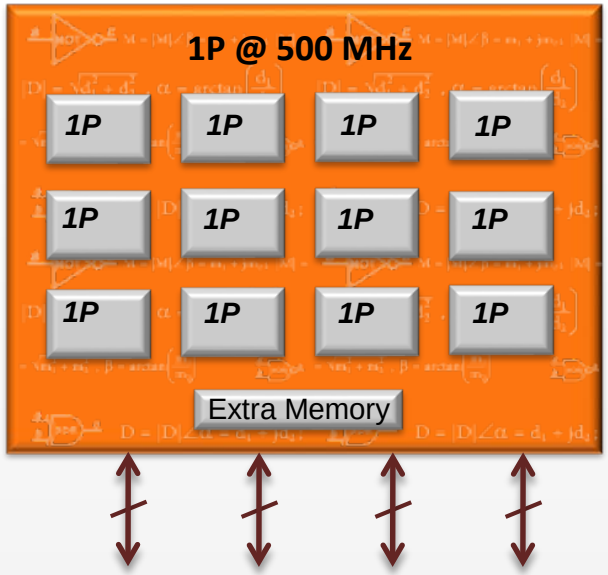
Solution: Algorithmic Memory® = Memory Macros + Algorithms

Physical Memory (12K x 144)



1P @ 500 MHz
Allows 500 Million MOPS[1]
([1] Memory Operations Per Second)

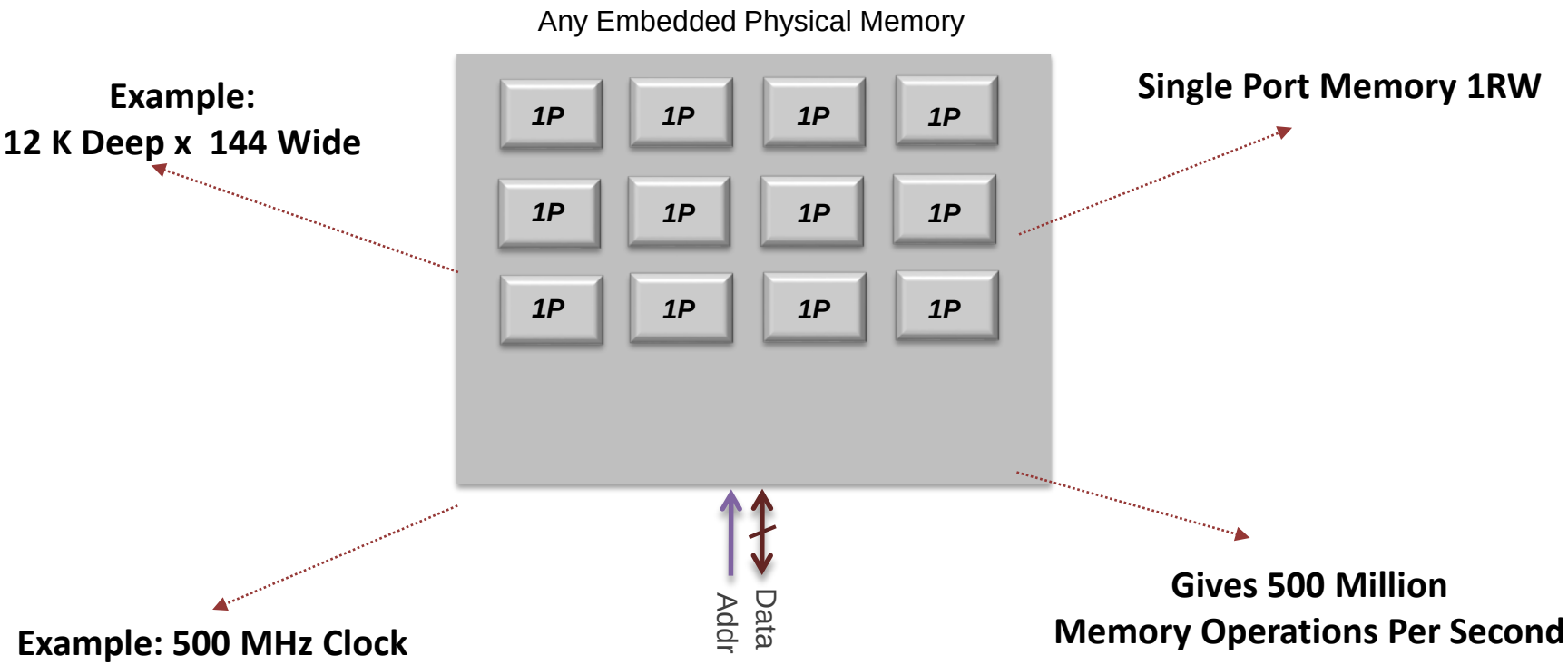
Algorithmic Memory (12K x 144)



4P @ 500 MHz
Allows 2000 Million MOPS
More Ports, Same Clock

NOTE: First Focus on Increasing True Random Access MOPS

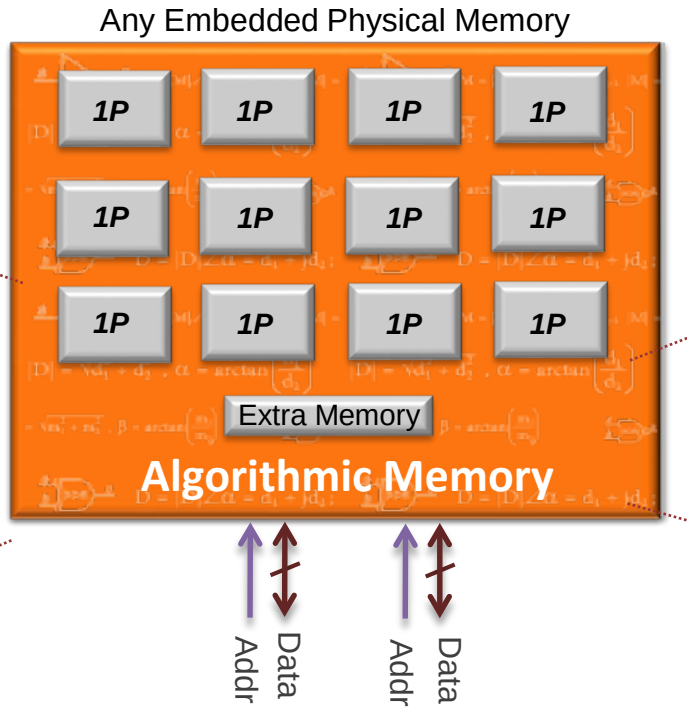
Solution: We Start with Physical Memory …



Memory Operations Per Second Limited By Clock Speed

Solution: … Transform To Algorithmic Memory

Algorithmic Memory = Generated from Existing Physical Memory



RTL Based: No Circuit or Layout changes

Simultaneous Accesses to the same Address, Row, Column, or Bank (no exceptions)

Each Port can access the entire Memory Address

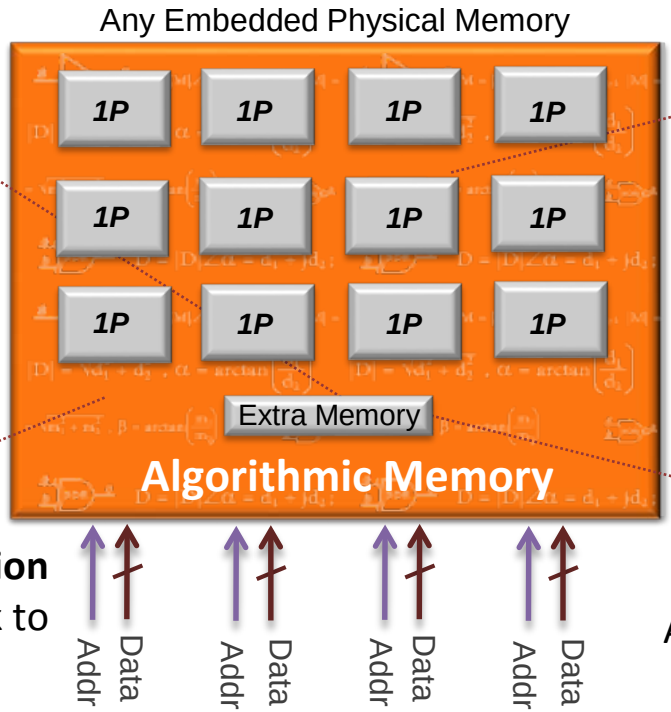
Exhaustively Formally Verified & Transparent to end-user

Complements Physical Memory & Increases Performance up to 10X more MOPS

Using Existing Physical Memory to Build any Multiport Functionality

Algorithmic Memory Technology: Explanation for Writes

Algorithmic Memory = Generated from Existing Physical Memory



(1) Extra Cache Bits: Sufficient Bits to Hold Burst of Writes in Cache

(3) Correct Load Balancing Algorithm: Move Data to Different Address During Write Congestion

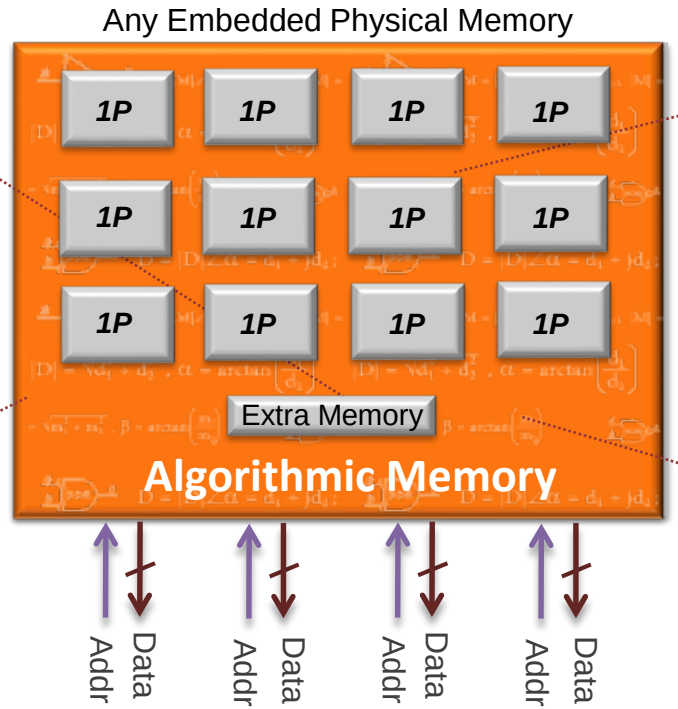
(4) Correct Garbage Collection Algorithm : Move Data Back to Original Location

(2) Cache Eviction: Correct Algorithm to Decide When to Evict Data from Cache

One can Mathematically Prove that With the Correct Steps in 1, 2, 3 and 4, all Patterns of Writes Are Covered

Algorithmic Memory Technology: Explanation for Reads

Algorithmic Memory = Generated from Existing Physical Memory



(1) Extra Bits: Sufficient Bits to Encode Data on Writes

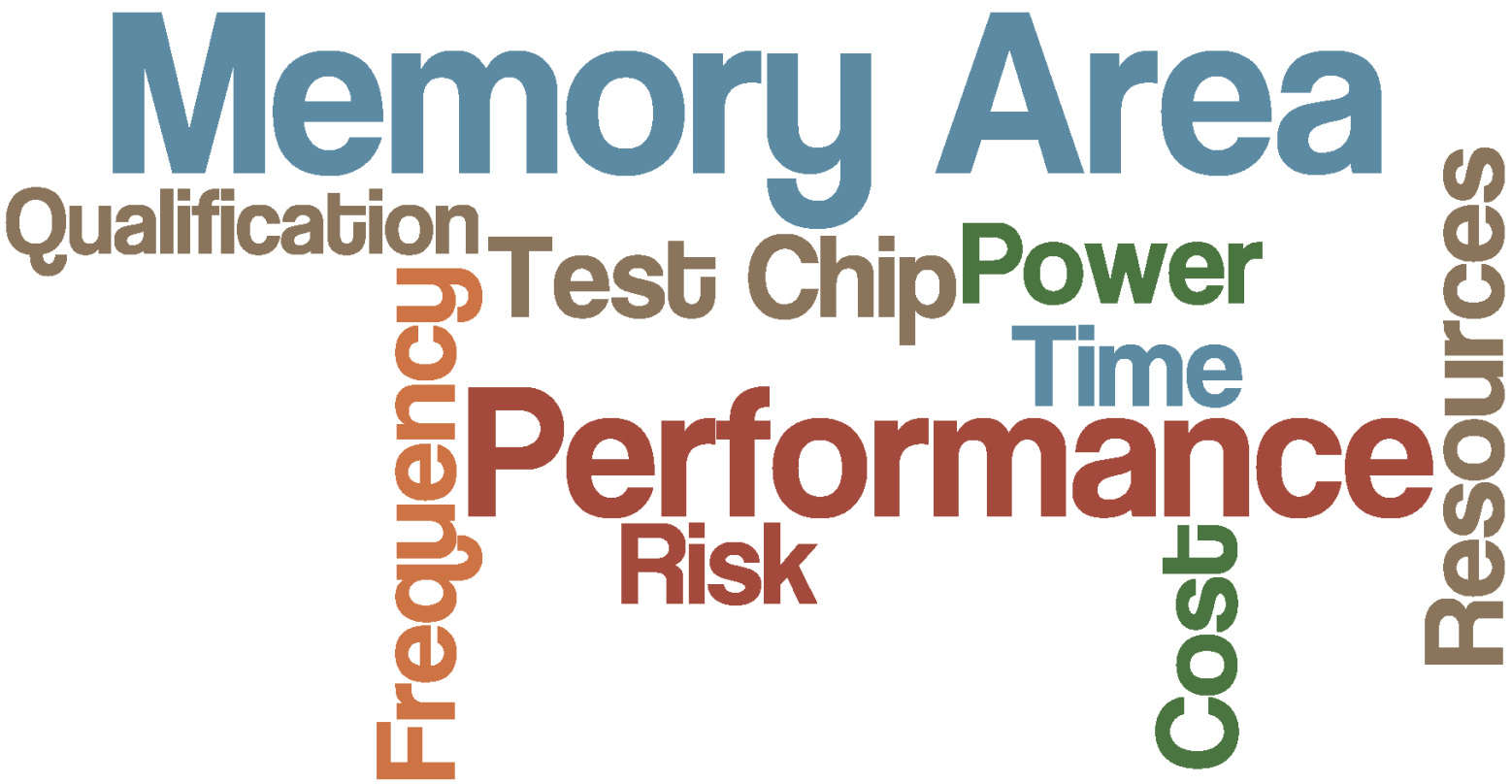
(3) Decoding Algorithms: Decode Time Affects Latency; Use Non-Optimal Codes for Faster Decode

(4) Multi Erasure Codes For More Reads: Supports Multiple Read Bank Conflicts

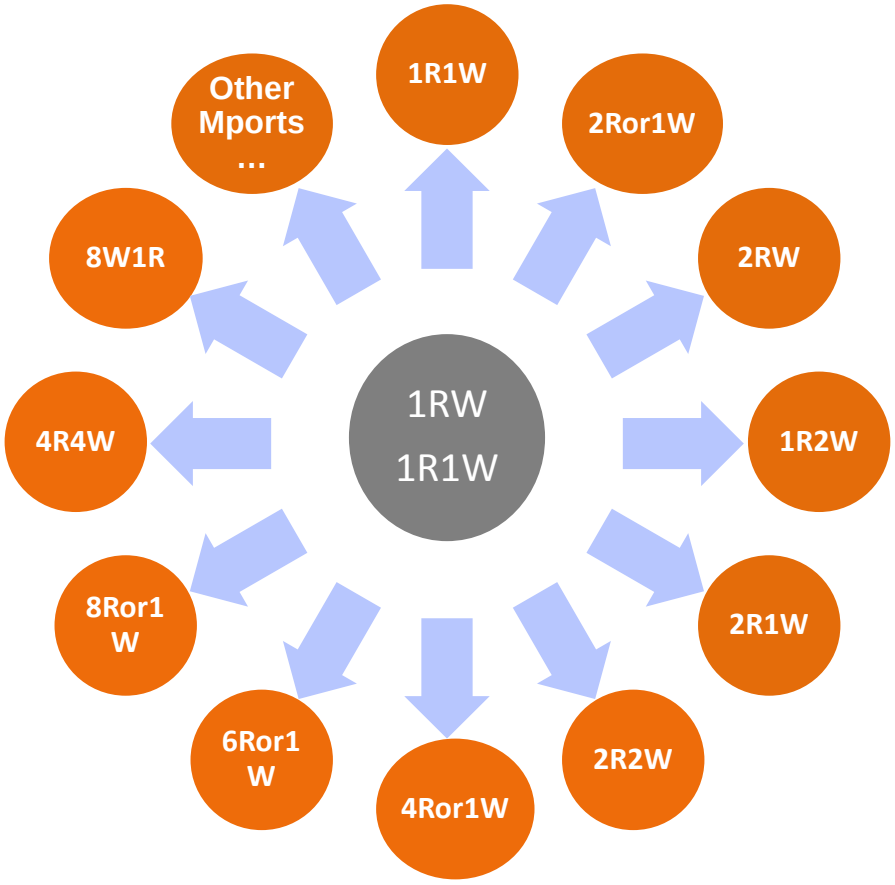
(2) Erasure Codes: Codes are Based on Erasure Coding; Treat Bank Conflicts as Erasures!

When 1, 2, 3 and 4, are Correctly Implemented All Read Conflicts Can be Resolved

Who Let the Dogs Out? …



Extend Performance, Power, Area of Physical Memory



Physical Memory
Algorithmic Memory

Built from Existing Embedded Memory

- Reduces cost, time to market,
- Reduces risk to build physical compilers

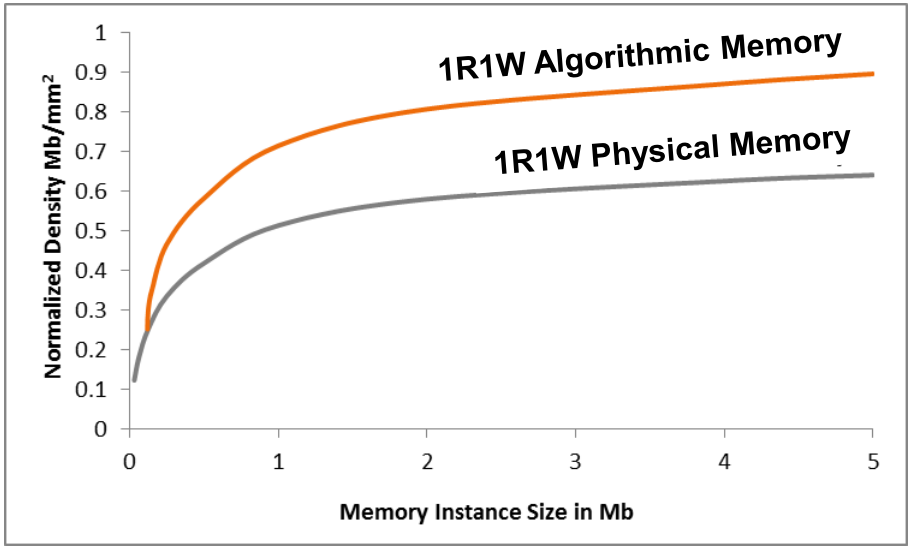
Better Performance, Power and Area

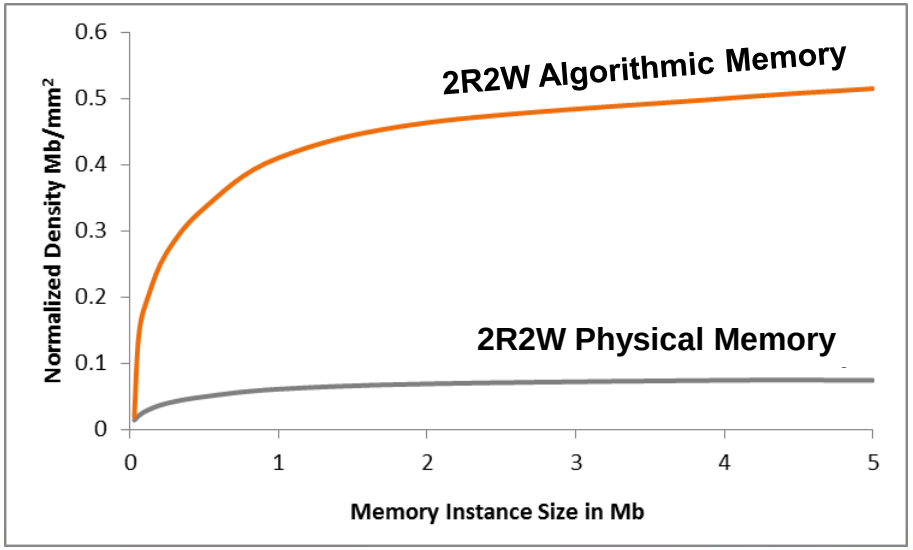
- Lowers area, power (medium, large size mem.)
- Increases clock frequency up to 30%

Integrates Seamlessly into ASIC Flows

- Exhaustively formally verified
- Supports standard SRAM interface
- Adds No additional clock cycle latency

Reduces SOC Memory Area

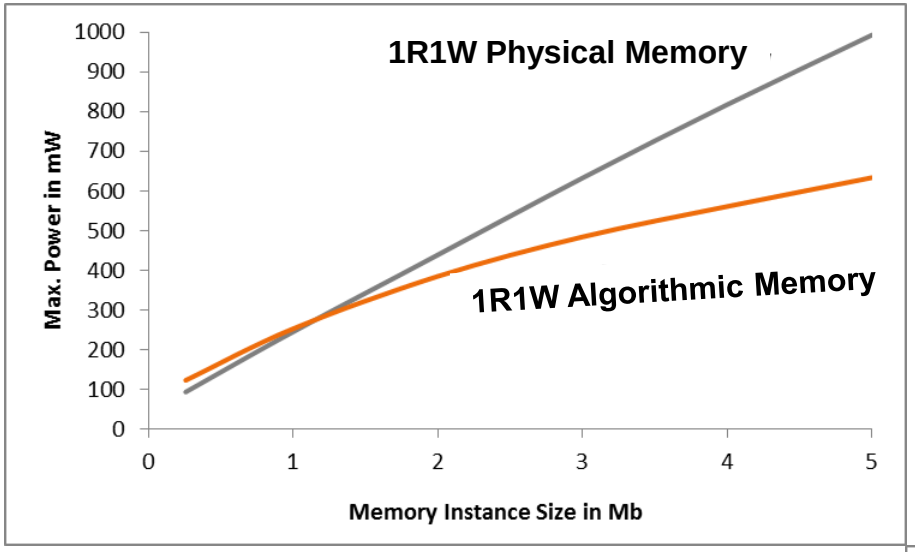


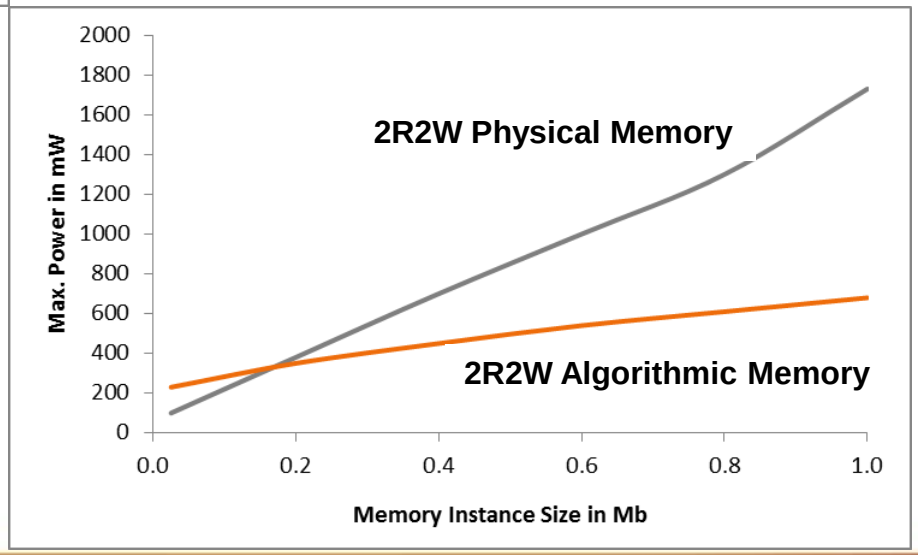


Normalized: Physical 1-Port Memory 1RW = 1 Mb/mm²

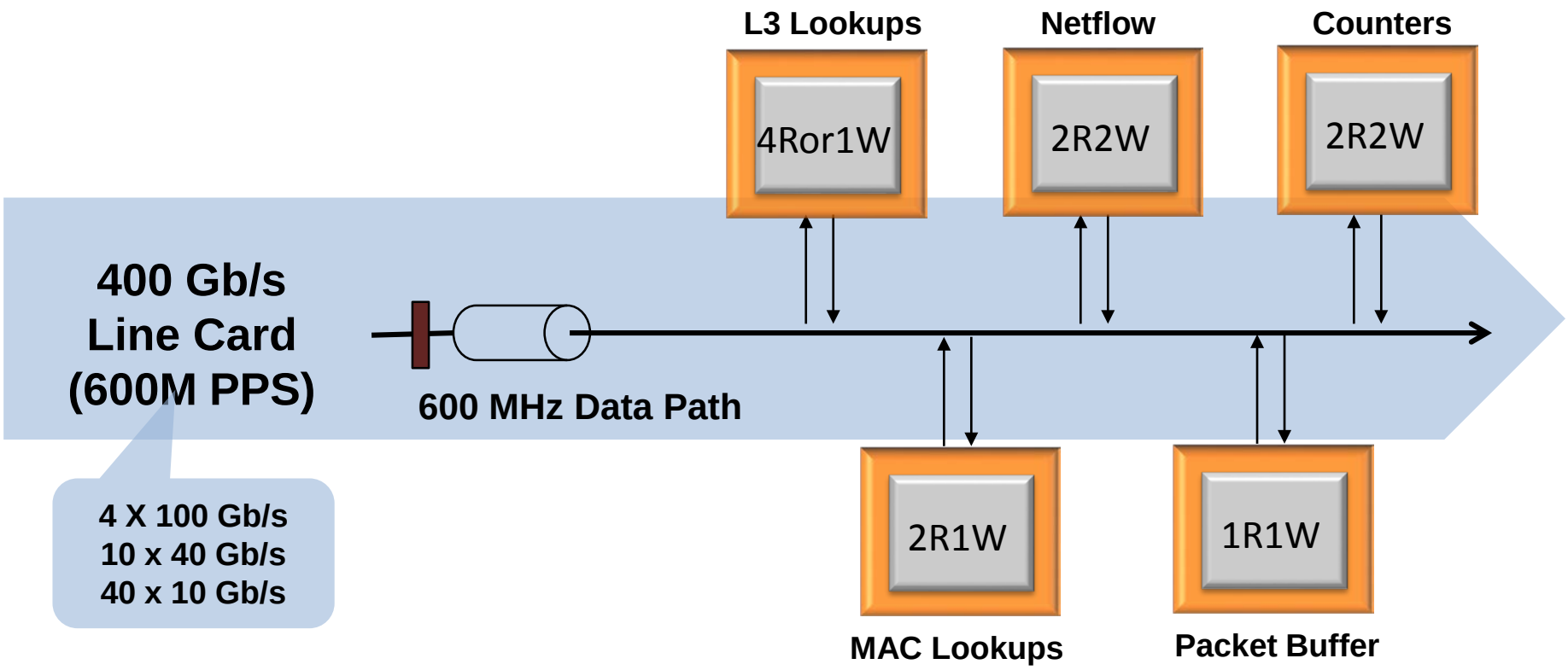
- Physical 2-port Memory, 1R1W = 0.6 Mb/mm²

Reduces SOC Memory Power



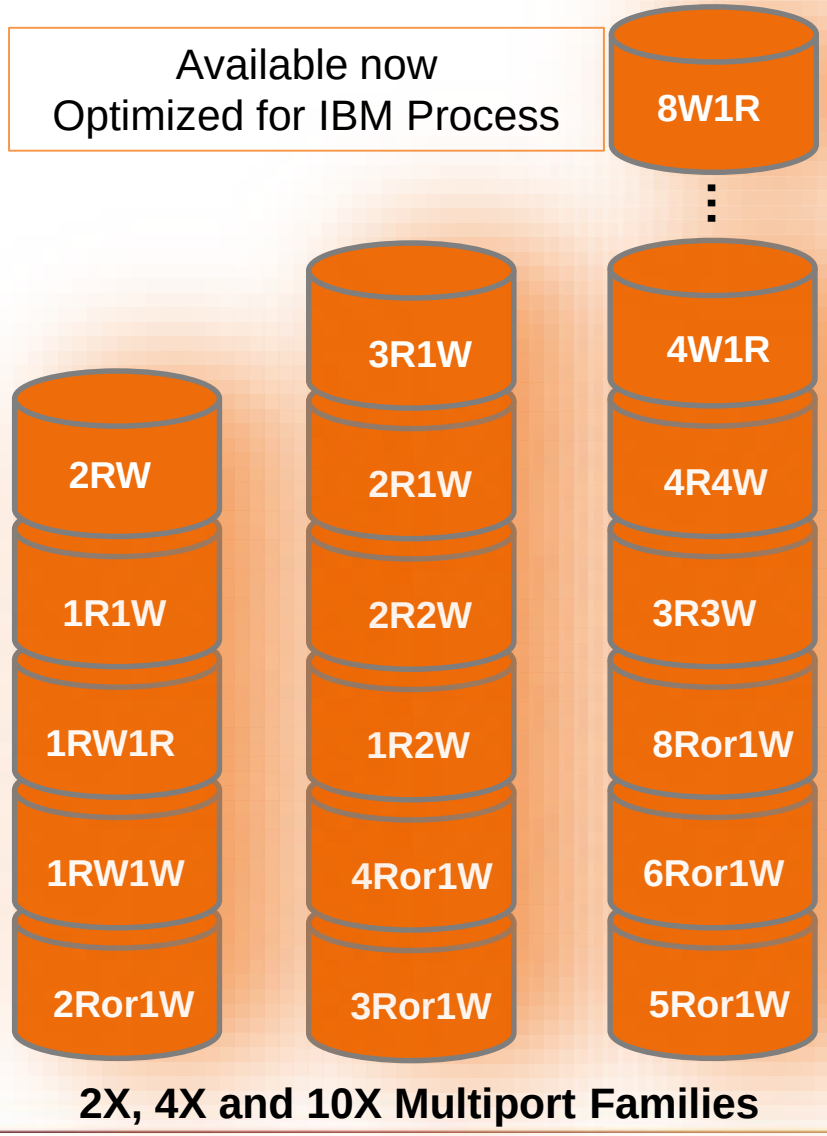


Algorithmic Memory Usage for Datacom Applications



Algorithmic Memory offers 2400 Million MOPS at 600 MHz Clock Speeds for Next Generation Aggregated 10G/100G Ethernet

Common Applications For Next Generation SoCs

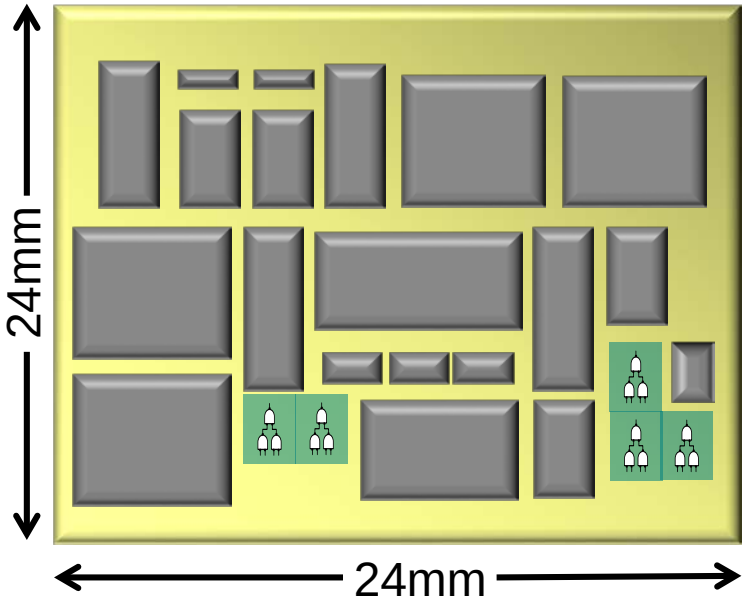


2X, 4X and 10X Multiport Families

- ***Data Comm: Networking/SDN/Storage Mobile Infrastructure/HPC***
  - Ingress buffers, egress buffers
  - Multicast descriptor lists
  - L2 MAC lookups, HPC lookups
  - Free lists for multicast buffers
  - Data and tag arrays for L2/L3 caches
  - Netflow, counters, state tables, linked lists
  - Route lookup tables
  - ACL tables
- ***HD Video, Automotive***
  - Frame Buffers, FIFOs
- ***High Performance Processors***
  - Multiprocessor L2/L3 tags/Caches
  - Mobile, Application Processor SoCs
  - DSP load/store units
  - Graphics SIMD Register files
  - Video pixel structures

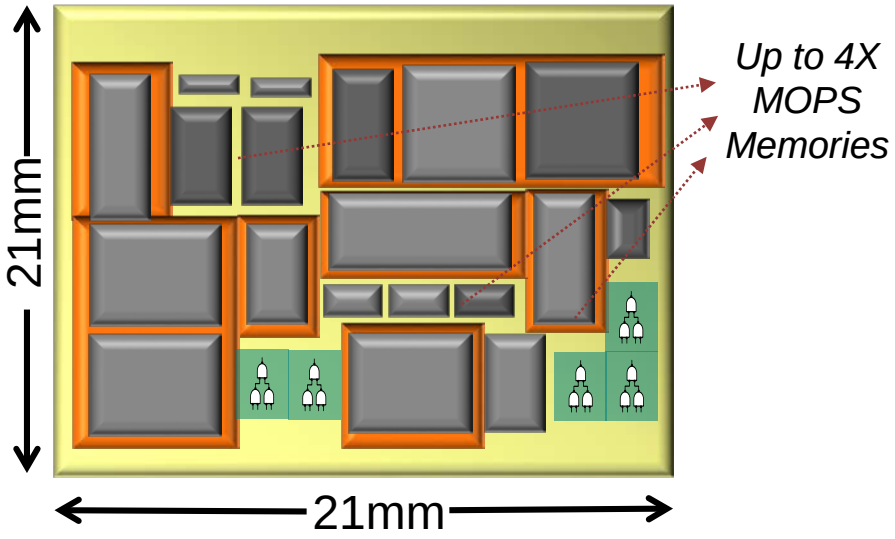
Tier-1 OEM Vendor Evaluation – PPA Benefits

Large ASIC with Physical Memory



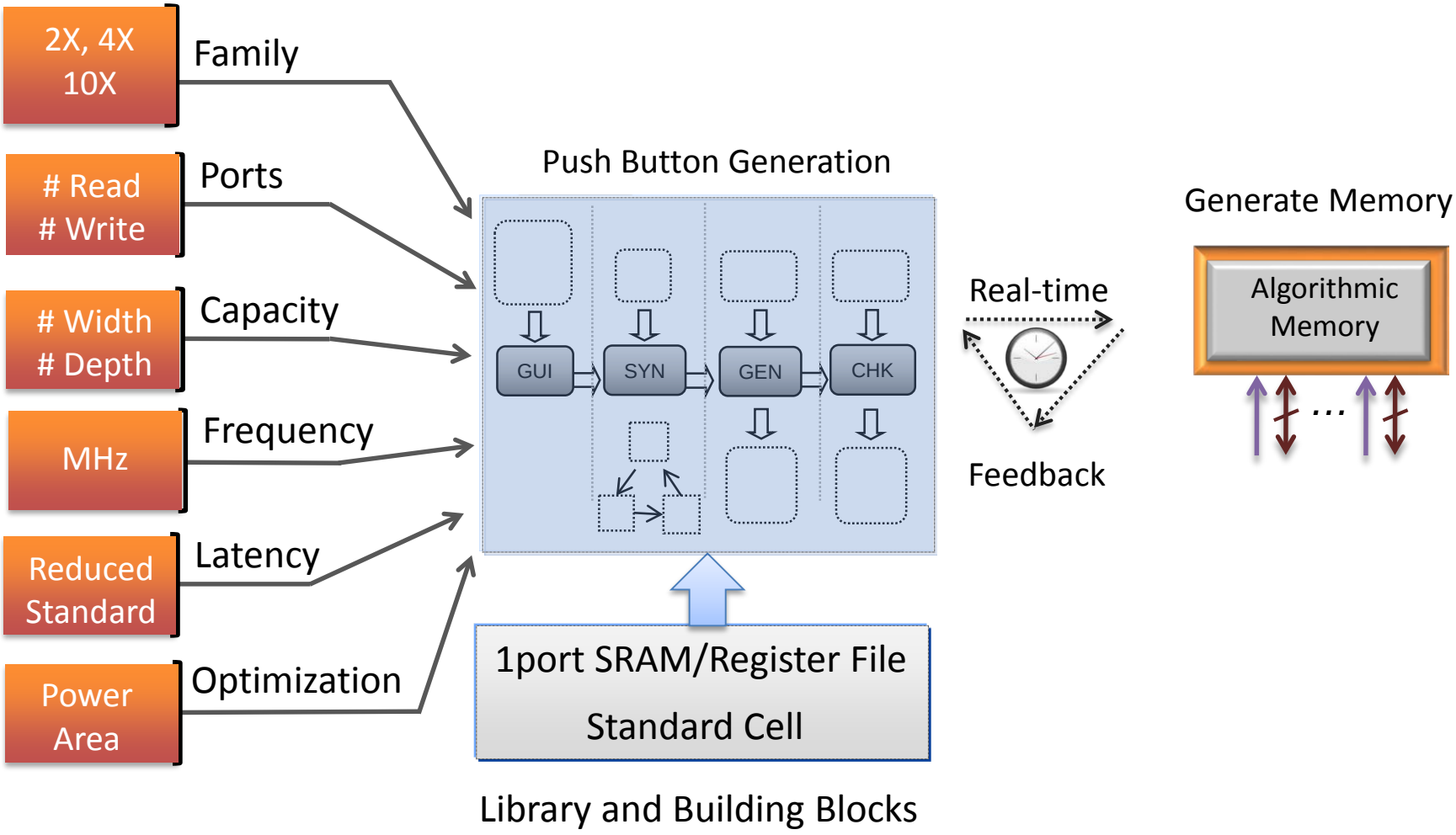
- Area 576 mm^2
  - 800 Mb of total memory
  - 165 Memory Instances
  - SRAM, RF, **eDRAM**
- Versatile memories required
  - 4R/1W, 2R1W, 1R2W memories

ASIC with Algorithmic Memory



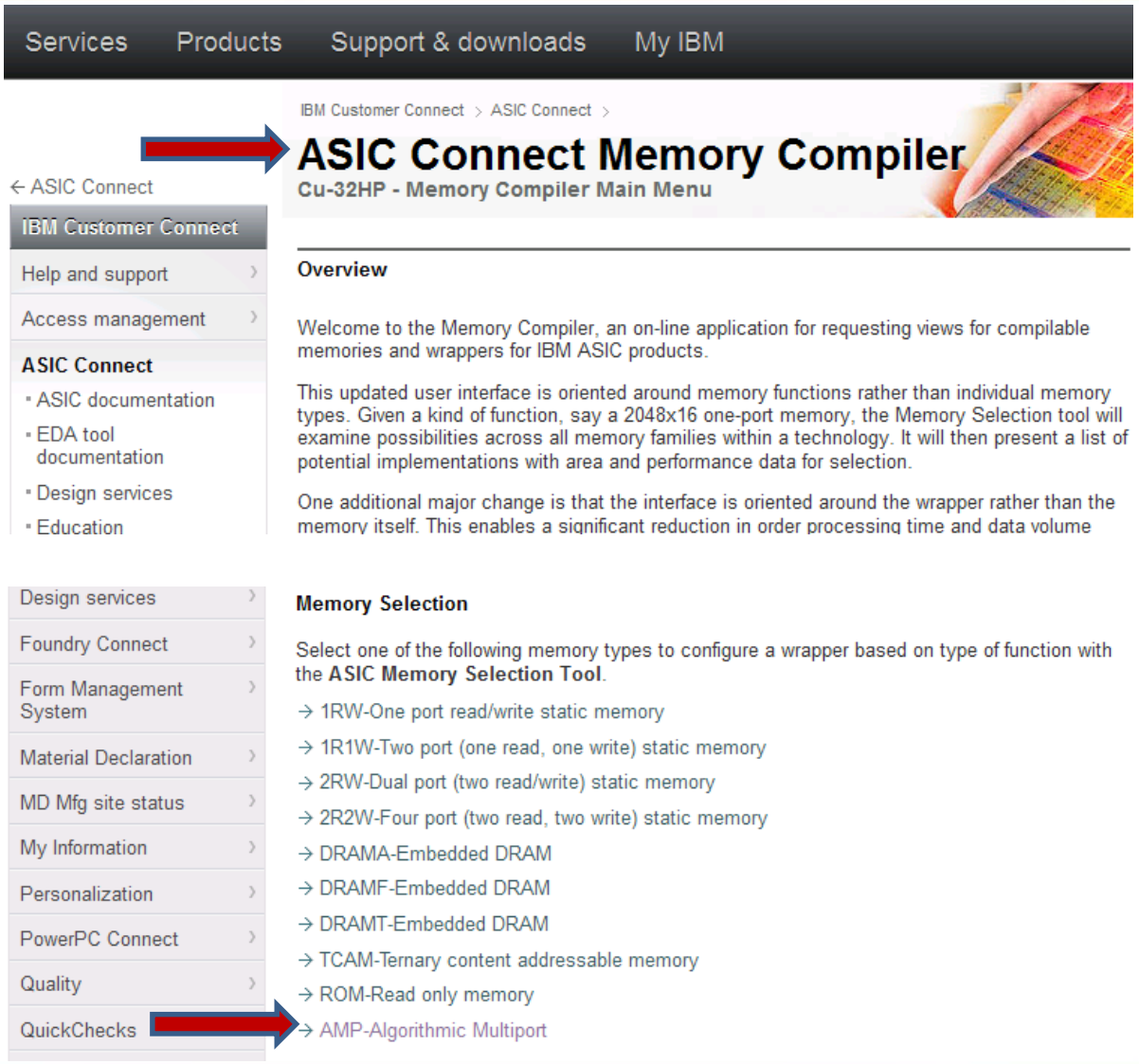
- Area 441 mm^2
  - Area Savings of 135 mm^2
  - Decreased die size by 23%
  - 136 Memory Instances Accelerated
- Power Savings > 12W
- 4X MOPS for select memories

Rapid Memory Analysis and Generation



Algorithmic MultiPort (AMP) Memories on IBM

- Login to IBM Customer Connect
  - https://www-03.ibm.com/services/continuity/resilience.nsf/pages/connect
- Accessing AMP
  - Navigate to CU-32HP
  - Libraries and Toolkits
  - AMP – Algorithmic Multiport



AMP: Offering On IBM 32nm, 14nm Process

IBM Physical Compilers

Single Port SRAM
Two Port SRAM
Dual Port SRAM
Four Port RF
1 Port eDRAM

Both SRAM and eDRAM Multiport Memories are Available on IBM Process

AMP eDRAM Multiports
2RW
1R1W
1R2W
2R1W
2R2W
3R1W
4R1W
1RW1W
2Ror1W
3Ror1W
4Ror1W

AMP SRAM Multiports
2RW
1R1W
1R2W
1R3W
1R4W
2R1W
2R2W
2R3W
2R4W
3R1W
4R1W
4R4W
1R1RW
1RW1W
2Ror1W
3Ror1W
4Ror1W

Conclusion

1. Summary of Benefits
   - Increases Memory Ports and Clock Performance
   - Lowers Area and Power
   - Easy Interface, Integration and Implementation
   - Creates Versatile Memory Portfolio
   - Reduces Cost, Risk and Time to Market

2. Algorithmic Pattern-Aware Memory
   - Not all Applications Require Random Access MOPS
   - Optimize Memory for Specific Access Patterns
   - Sequential, Read-modify-write, Counters, Allocation, Strides …

Algorithmic Memories are not a panacea, but present a new solution to alleviate the memory performance gap

MEMOIR
SYSTEMS

Questions & Answers