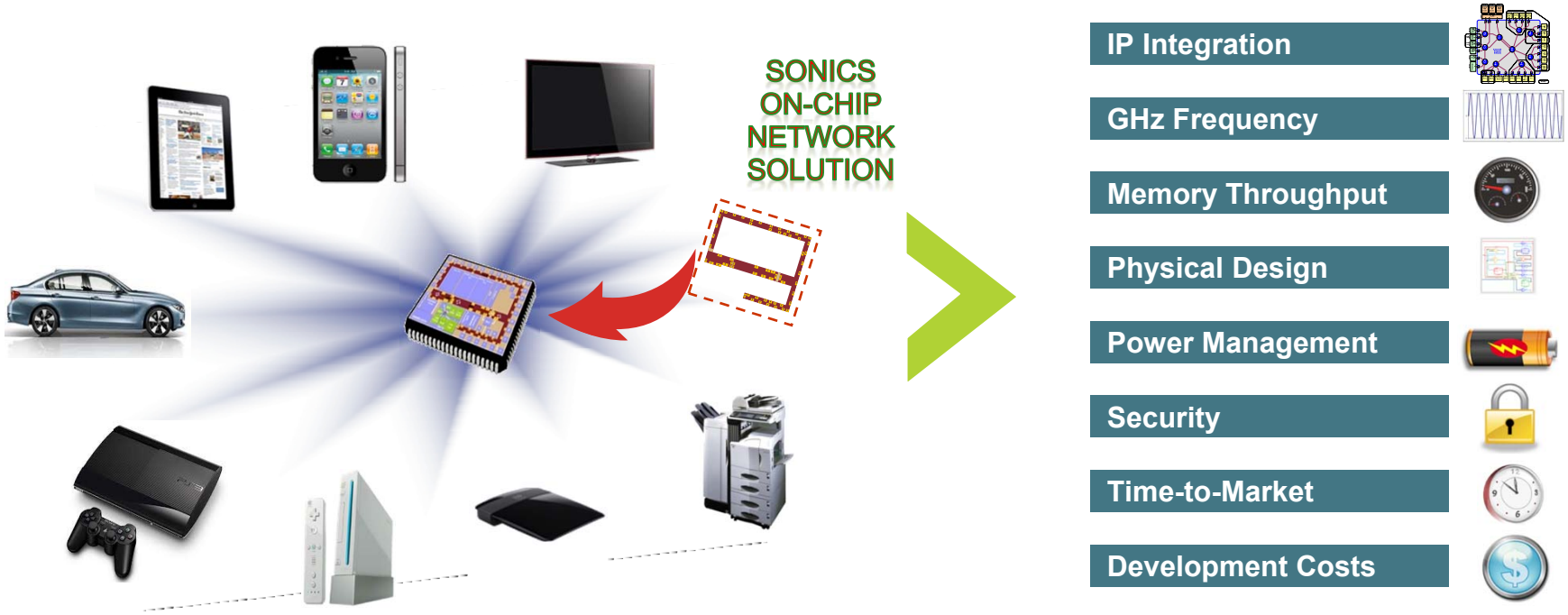
SONICS™

MemMax® Product Presentation

January 2018

NoC IP Vendor-of-Choice

Sonics' on-chip networks help leading SoC designers solve some of the most difficult challenges in SoC design



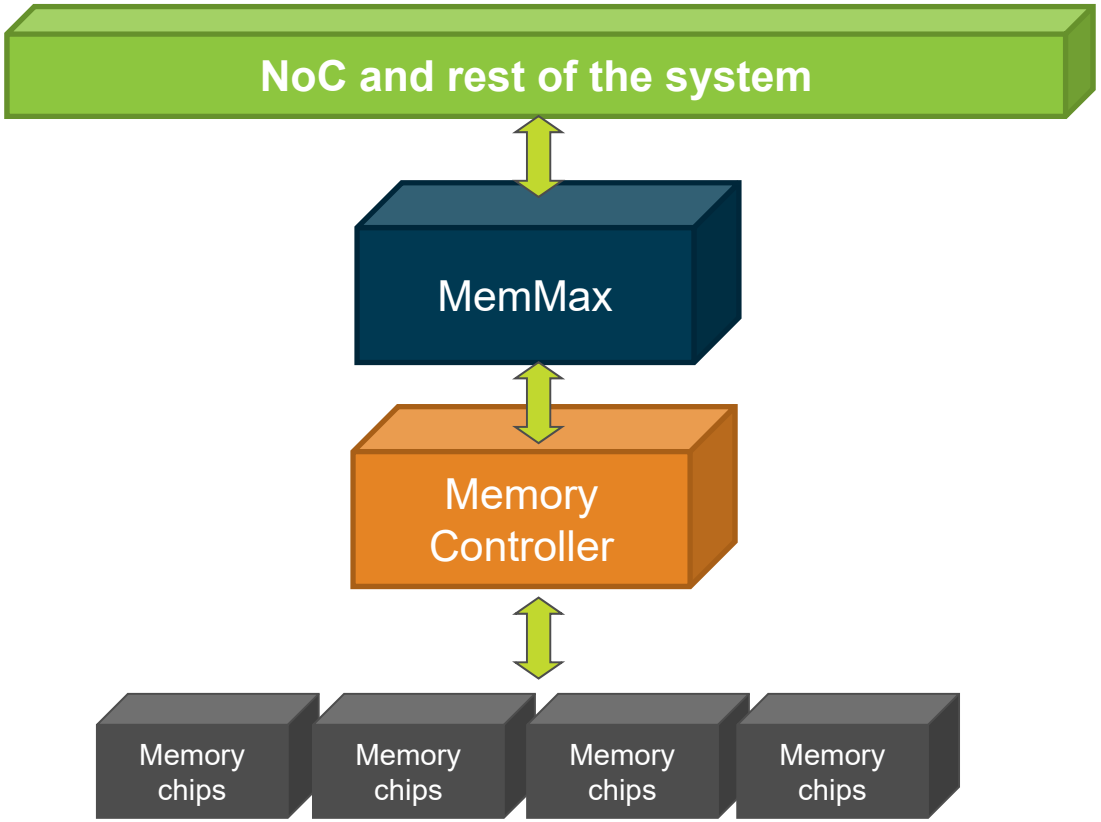
- **Sonics System IP:**
  On-chip Networks, Memory Subsystem, Power Partitioning & Management, Performance Monitor & Debug, Security Firewalls

Sonics NoC IP Suites

SonicsStudio® Director Development Environment			Fully Non-Blocking	Interface Protocols
On-Chip Network System IP	SonicsGN®	High-speed, serialized, router-based network with integrated power/clock domain management and scalability from IoT to servers	✓	ACE-Lite AXI AHB APB OCP
	SonicsSX®	Low-power, low-latency, high-bandwidth network featuring cascaded switching exchanges, multi-channel memory interleaving and a rich set of network services	✓	
	SonicsLX®	A limited version of the SonicsSX suitable for mid-range SoCs and IP subsystems requiring minimum latency to subsystem memory	✓	
	Sonics3220™	Non-blocking, peripheral interconnect that ensures power-efficient, low latency access to a large number of low bandwidth target cores while spanning large physical distances	✓	
	SonicsExpress™	A highly configurable bridge, with native OCP or AXI support, typically used for asynchronous clock, voltage, or power domain crossing.	✓	
DRAM Systems	MemMax®	A multi-threaded DRAM scheduler that optimizes memory throughput while protecting QoS	✓	
On-Chip Analysis	SonicsMT™	Performance monitoring and hardware trace IP provides real-time measurement of throughput and latency and tracing of transactional events for CoreSight®-compliant debug		

MemMax® At-a-Glance

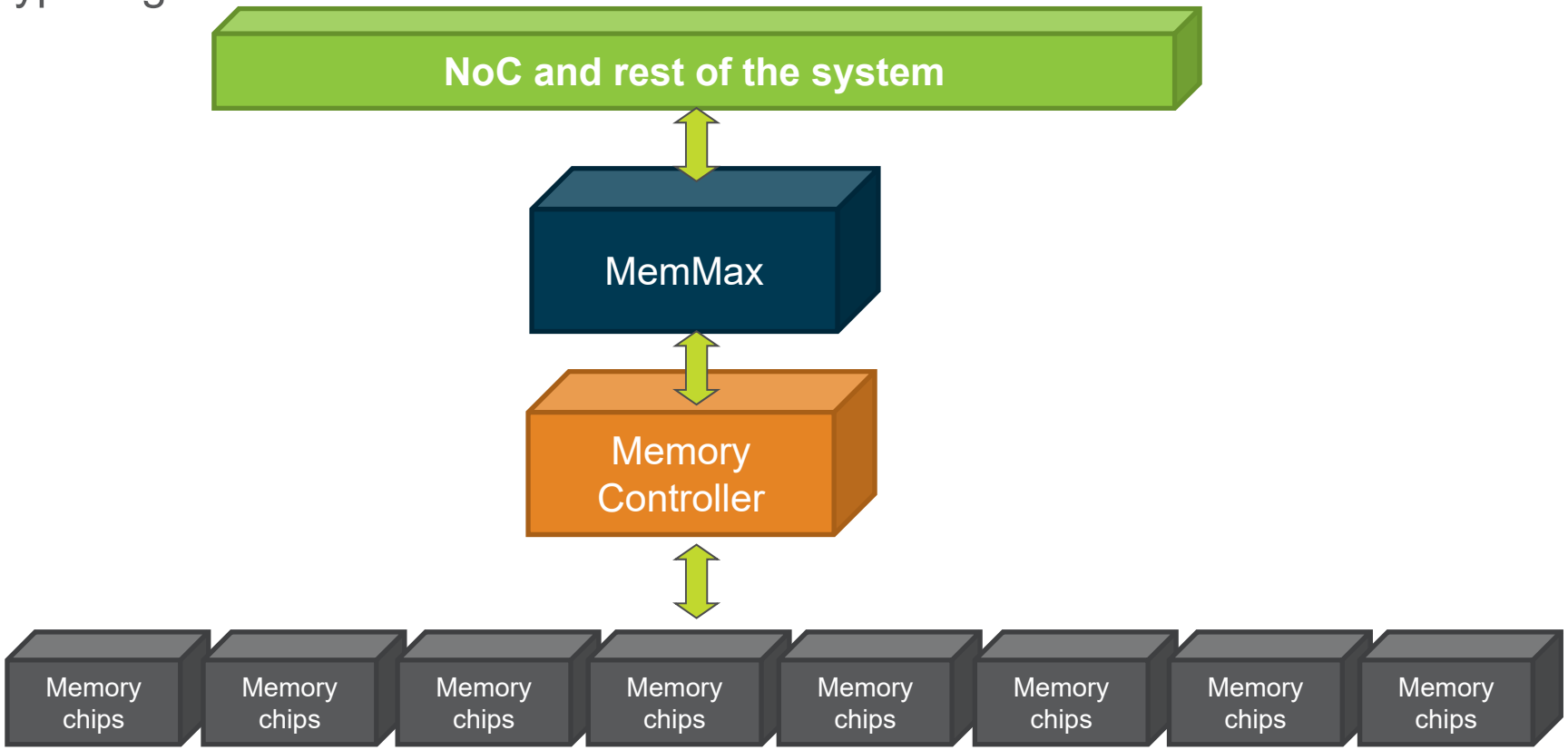
- Schedules traffic for high throughput and QoS
- Independent from DRAM type/organization



SONICS

MemMax® At-a-Glance

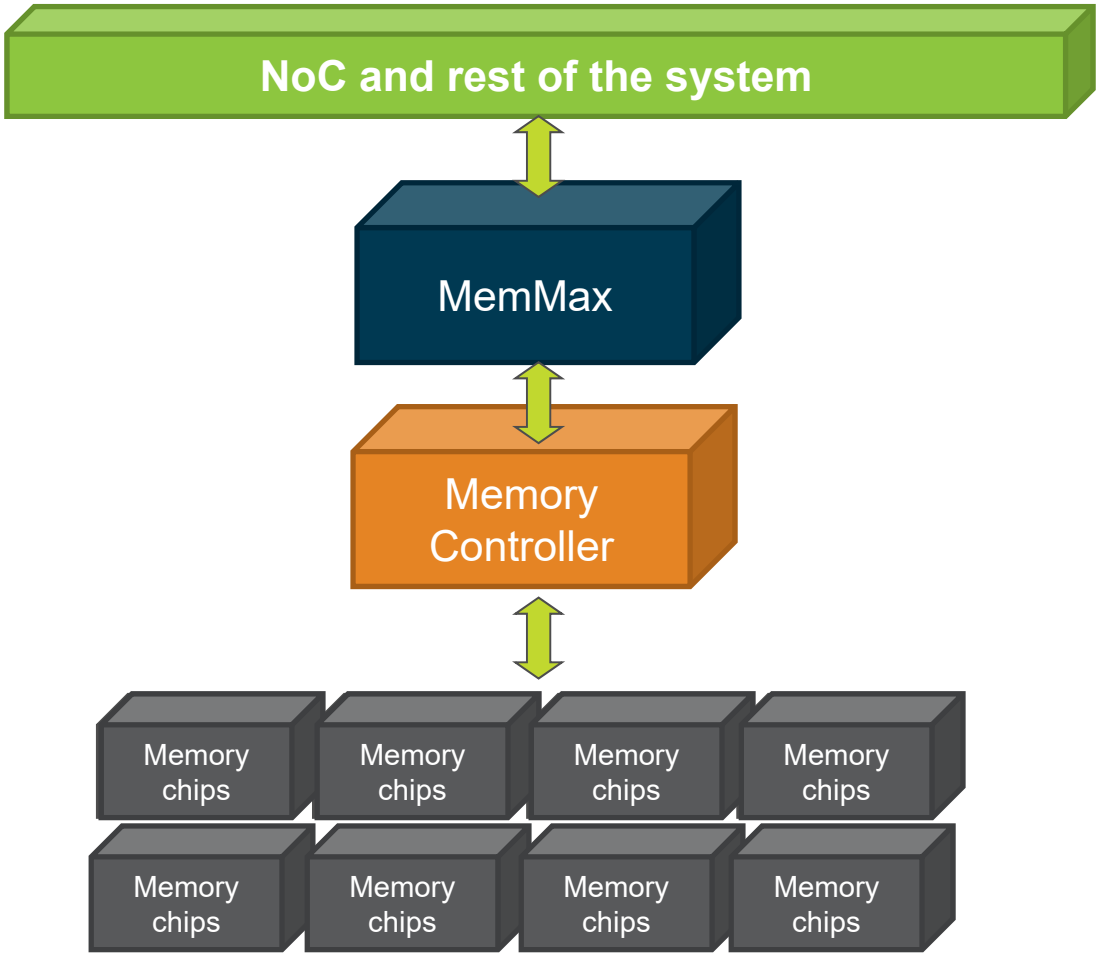
- Schedules traffic for high throughput and QoS
- Independent from DRAM type/organization



SONICS

MemMax® At-a-Glance

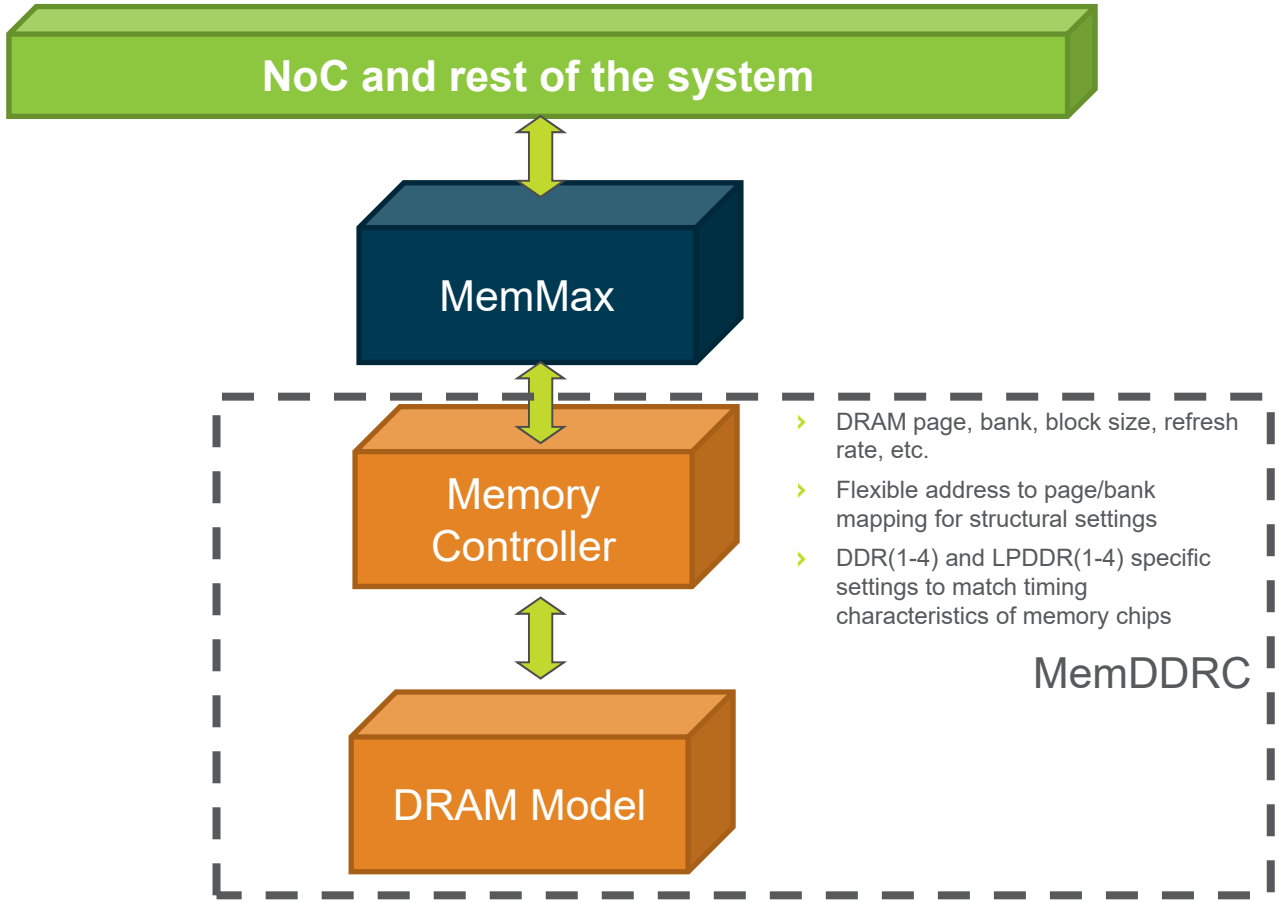
- Schedules traffic for high throughput and QoS
- Independent from DRAM type/organization



SONICS

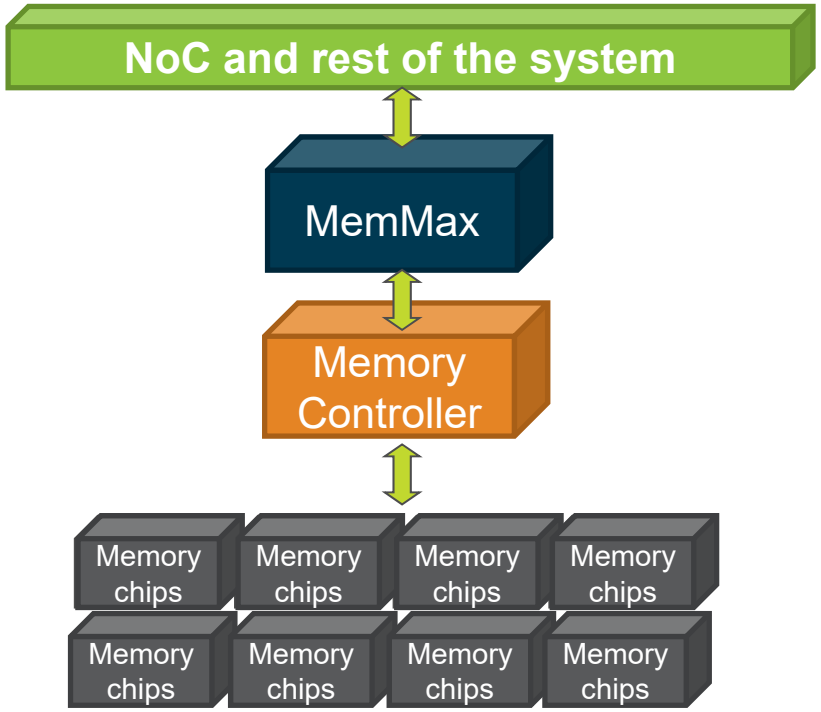
MemMax® At-a-Glance

- Schedules traffic for high throughput and QoS
- Independent from DRAM type/organization
- Behavioral MemDDRC matches almost any DRAM subsystem configuration or behavior



MemMax Benefits

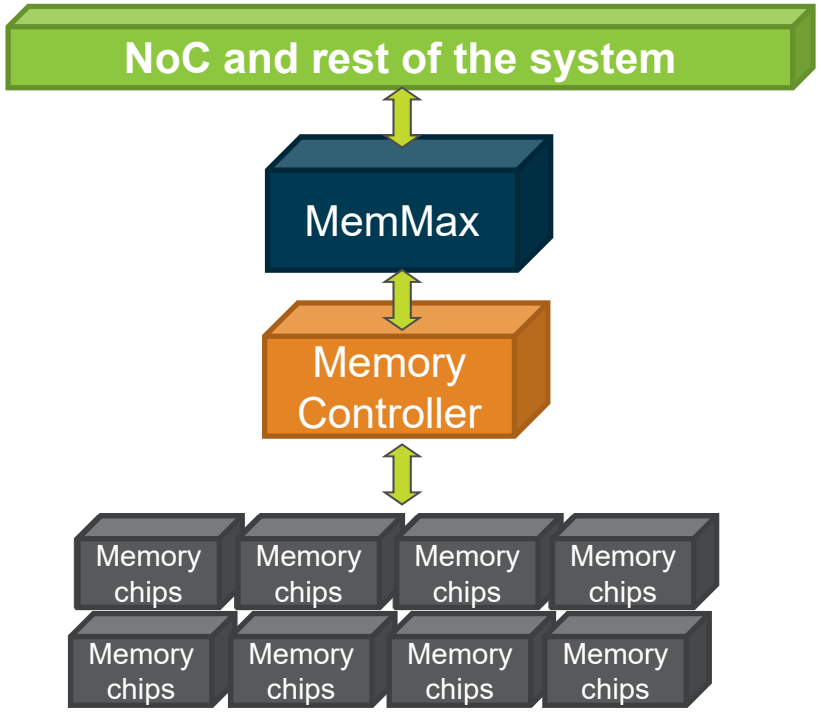
- High DRAM utilization (i.e. bandwidth) and Quality of Service (QoS) are always in conflict
  - High BW requires longer DRAM bursts for better page hit rate
  - QoS demands short DRAM bursts for better interleaving
- MemMax solution gives the designer the power to optimize DRAM efficiency and the freedom to tune the trade-offs between BW and QoS



MemMax Technology Overview

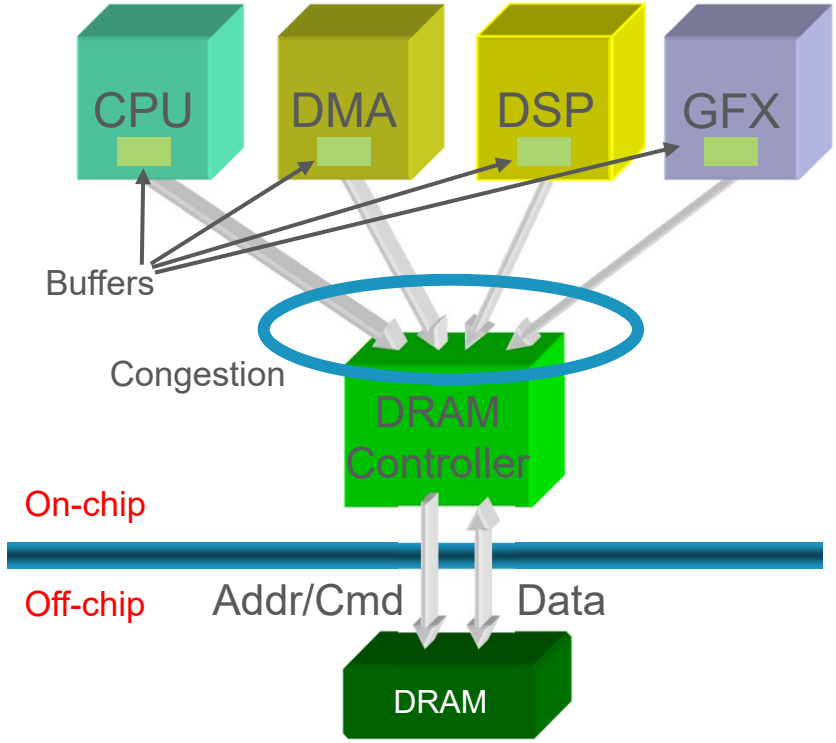
Agenda

- Configuration Challenges
- NoC + MemMax + Controller
- Weighted DRAM scheduling with QoS
- Flexible Address Tiling for Higher Efficiency
- Quality-of-Service (QoS)
- MemMax® Block Diagram
- Scheduler Block Diagram
- Memory State Awareness
- Efficient Transaction Grouping
- Summary



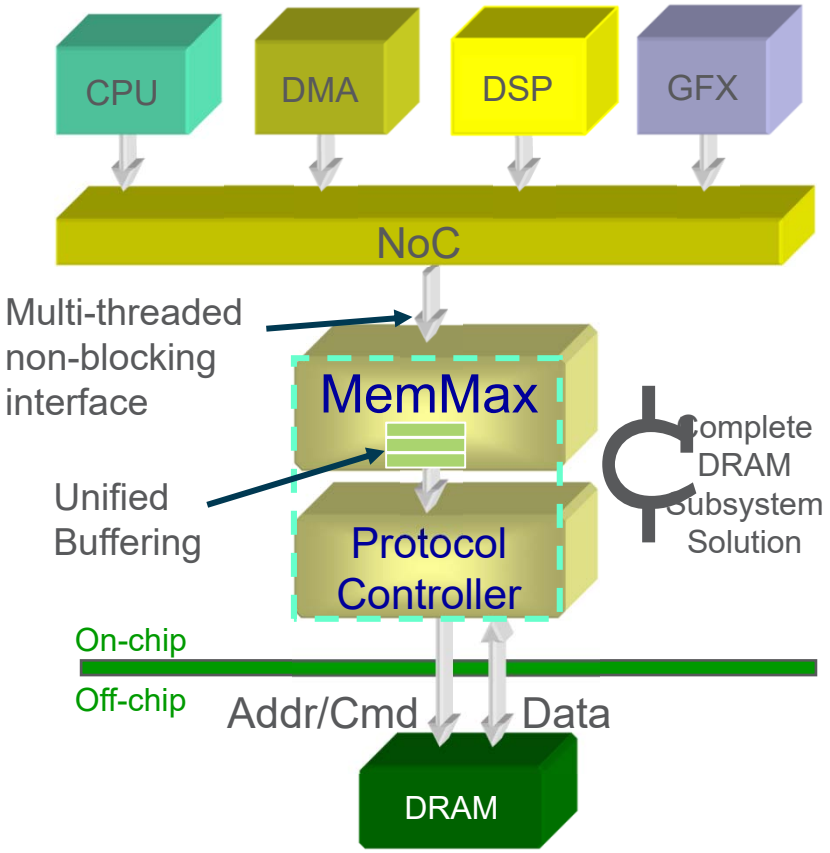
Configuration Challenges

- Lots of wires = congestion
- Must have Initiator FIFOs to match DRAM protocol and rate
- Difficult for placement
- Poor scalability
- Custom for each SoC design



Interconnect + MemMax + Controller

- NoC providing flexible mapping between initiator flows and MemMax threads → *Scalability*
- Unified multi-threaded non-blocking interface → *Throughput and Wire Minimization*
- Separation of memory scheduling from protocol control → *Decoupling*
- Unified buffering using compiled RAM → *Area Optimization*
- Programmable QoS → *Flexibility*
- Internal bank timers avoid sending commands to busy banks → *Efficiency*



SONICS

Weighted DRAM Scheduling with QoS

- Weighted scheduling algorithm

- QoS measured per-thread
  - QoS levels: Priority, Guaranteed Bandwidth and Best-effort
- Page Filter keeps page open to maximize in-page access
- Direction Filter groups RDs or WRs together
- Chip Filter minimizes rank crossing
- LRS ensures fairness / prevents starvation

SONICS

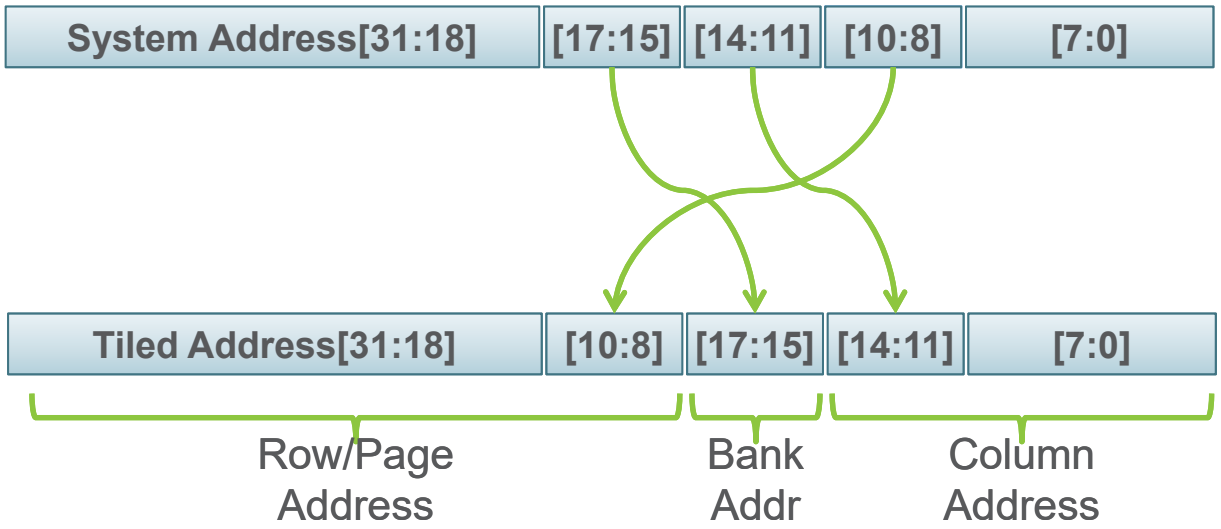
Quality-of-Service (QoS)

- Initiator data flow (threads) mapped to MemMax threads through interconnects
  - Multiple master can share one thread before coming into MemMax
- Each MemMax thread is assigned to one of the QoS levels
- Non-blocking, multi-threaded interface allows
  - Out of order acceptance for high priority traffic
  - Minimize buffer size w/ guaranteed throughput
  - Maintain best achievable DRAM efficiency

Thread QoS Level	Bandwidth Allocation?	QoS Model w/ Promotion and Demotion
Priority	Yes	Low latency while within BW allocation, best-effort otherwise
Bandwidth	Yes	Guaranteed BW while within BW allocation, best-effort otherwise
Best-effort	No	N/A

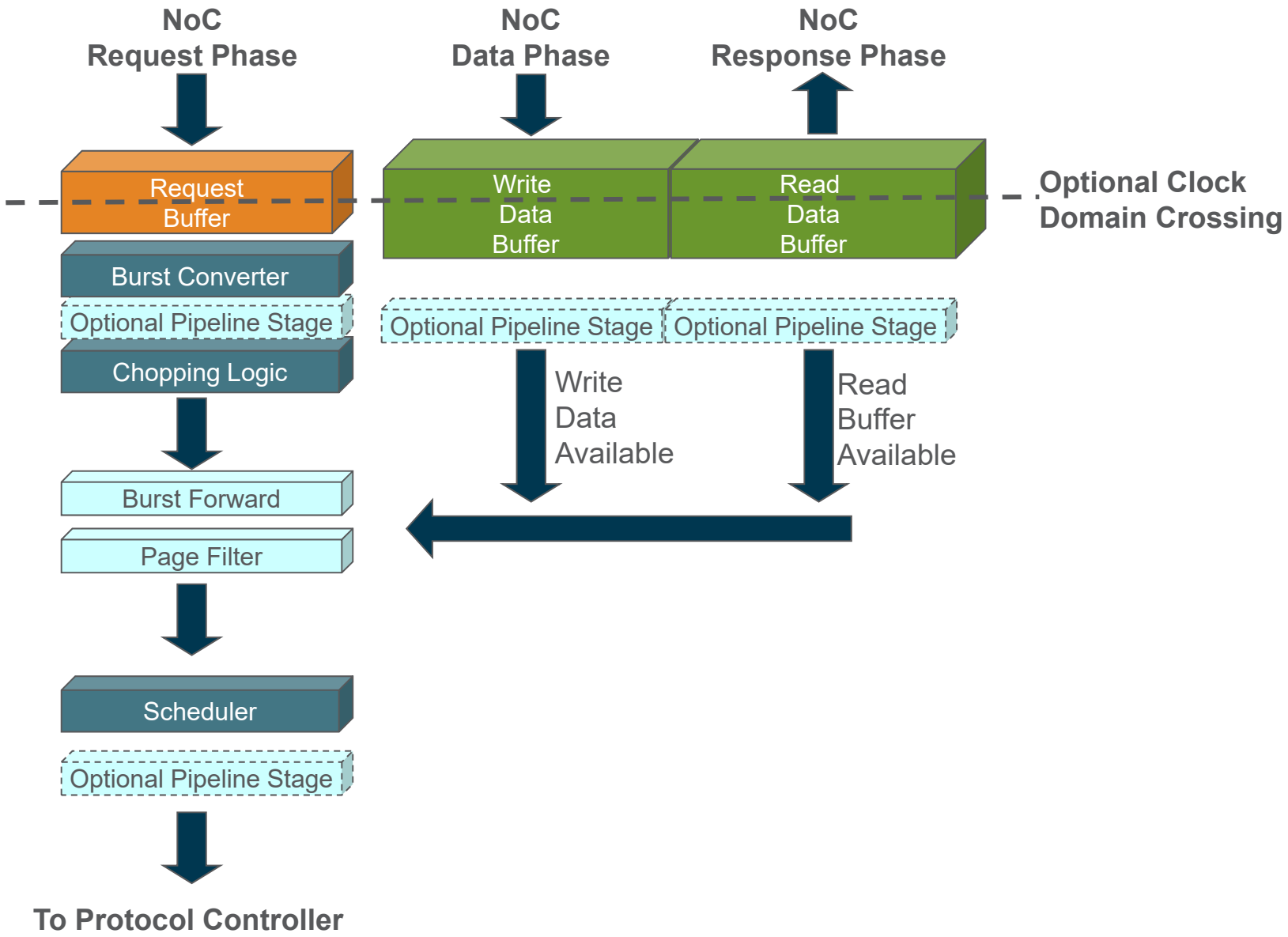
Flexible Address Tiling for Higher Efficiency

- Up to 7 different tiling regions simultaneously active
  - Triggered by NoC address space (associated with address regions)
- Frequently used for video frame buffers
  - Put rectangular region of pixels in same page, e.g. to speed macroblock fetch
- Support Bit Swapping as well as Bank Bit Transformation



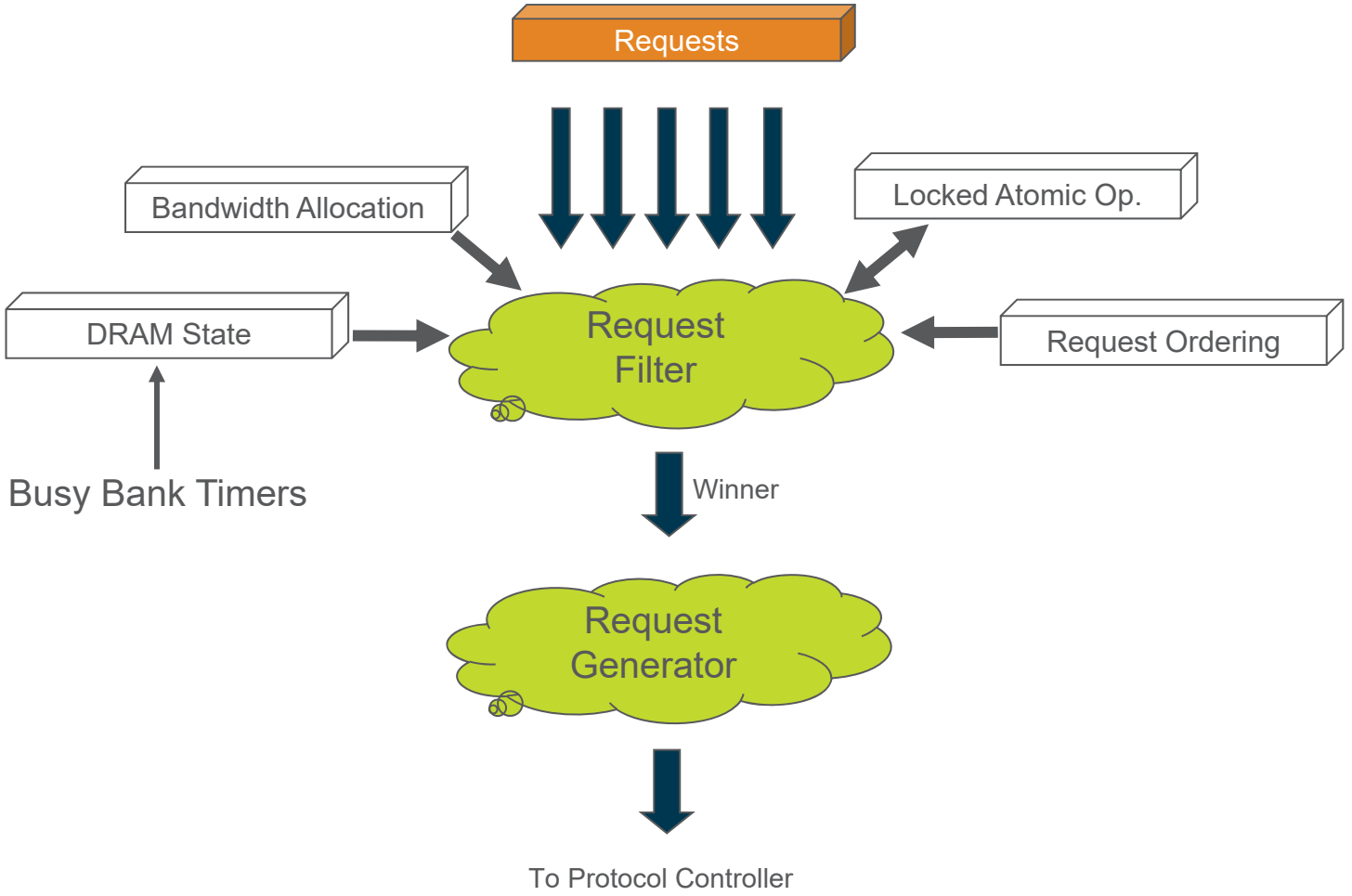
SONICS

MemMax Block Diagram

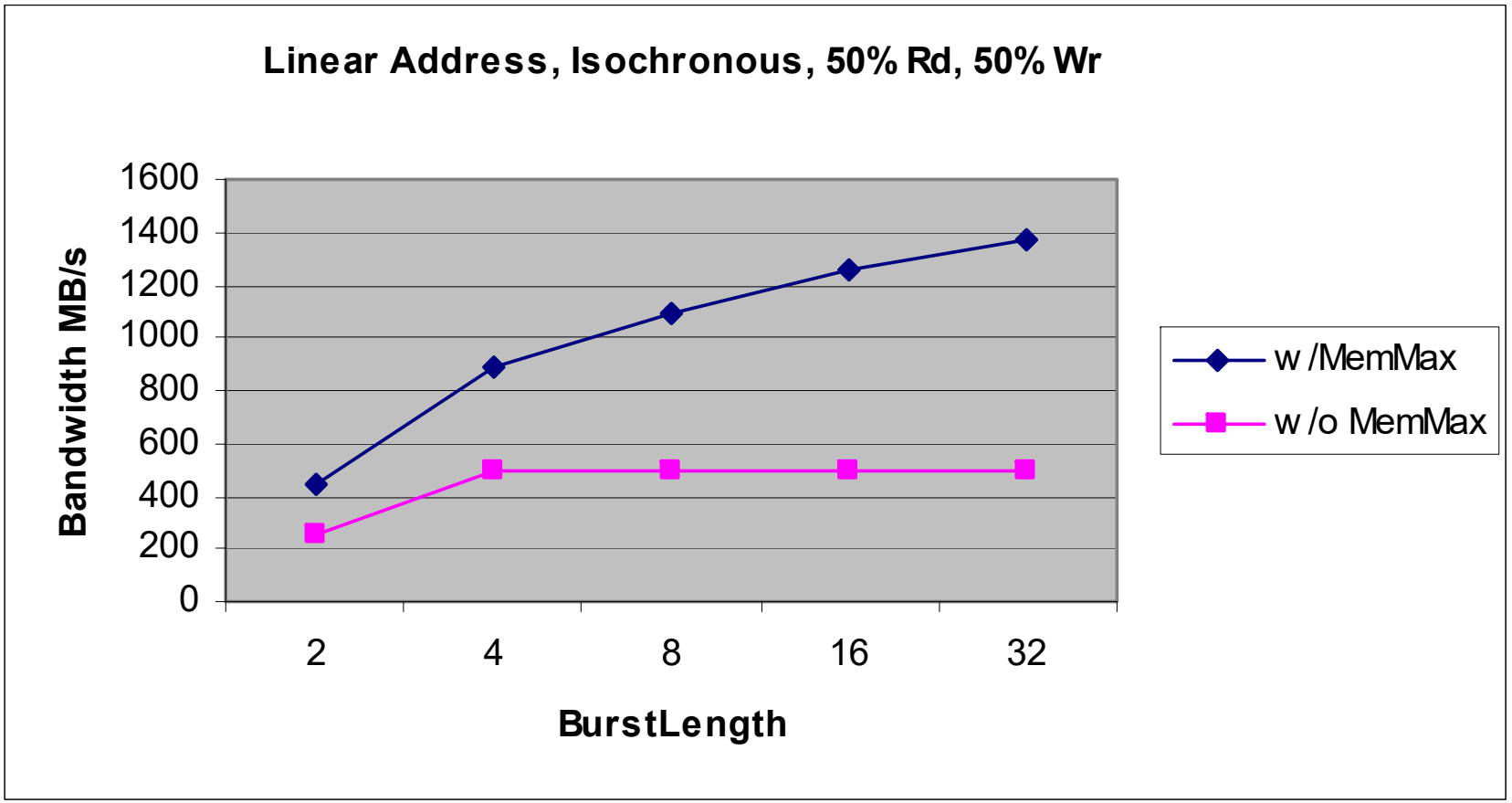


SONICS

Scheduler Block Diagram

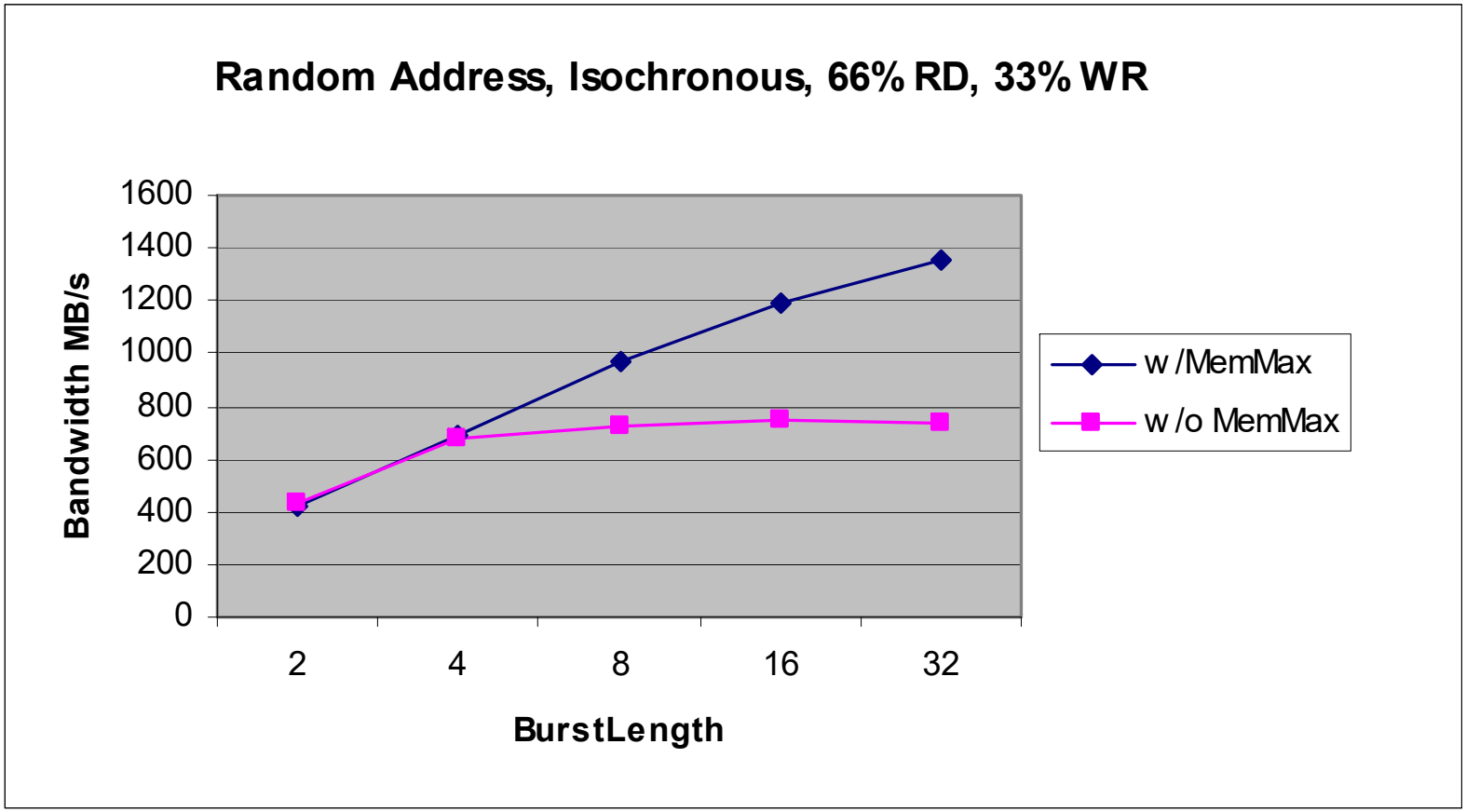


Memory State Awareness



SONICS

Efficient Transaction Grouping



SONICS

MemMax Memory System Summary

- Separation of system from memory speeds, protocols
- Achieve higher DRAM efficiency
- Multi-threaded non-blocking interface with end-to-end QoS
- Compile-time/runtime configurability maximize flexibility to match DRAM type and organization
- Centralized data buffers with compiled RAM minimizes area, including optional asynchronous clock crossing

Thank You!

Sonics Inc.

2570 N. First Street, Suite 100
San Jose, CA 95131