



Multicore and the 32 Core Cavium OCTEON II 68xx

R. E. Kessler

Cavium, Inc.

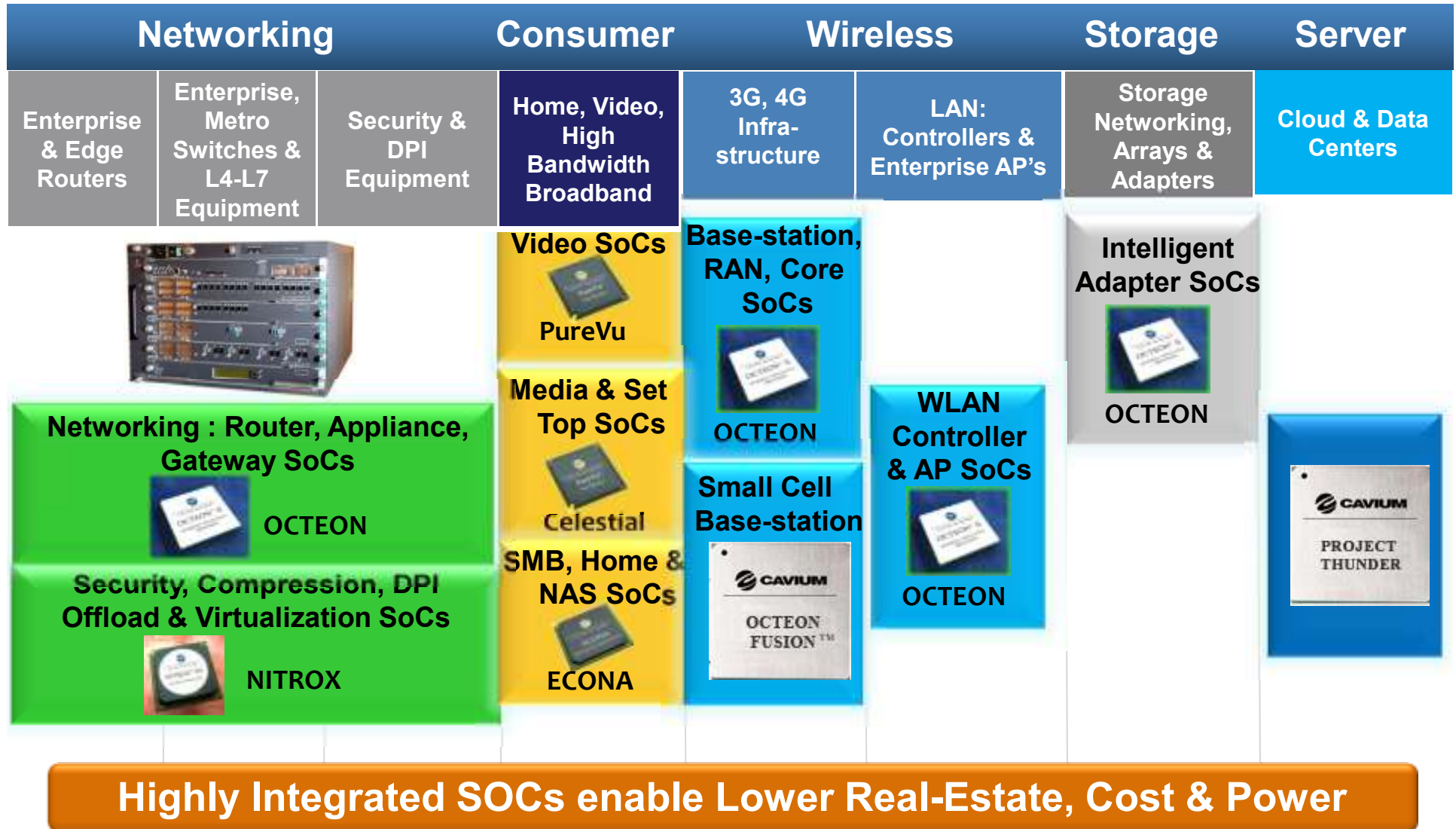
February, 2013

Agenda

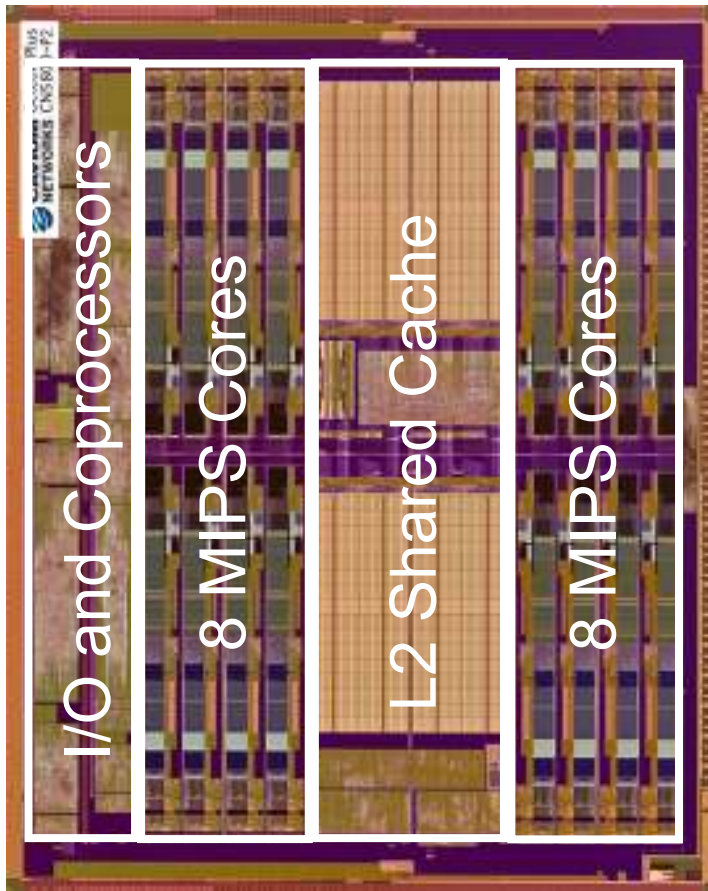


- Cavium Multicore Background and Upcoming
- 32-core OCTEON 68xx Architecture
 - Efficient CPU cores
 - Cache coherence, interconnect & memory bandwidth, chip floorplan
 - Power scalability
 - Hardware work queueing, scheduling, synchronization, and ordering assist
 - Coprocessor acceleration
- 32-core OCTEON 68xx Performance Results
- OCTEON 68xx Evaluation Board, Program, and Potential Projects

Cavium SoC's for Range of Target Markets



Some Existing OCTEON chips

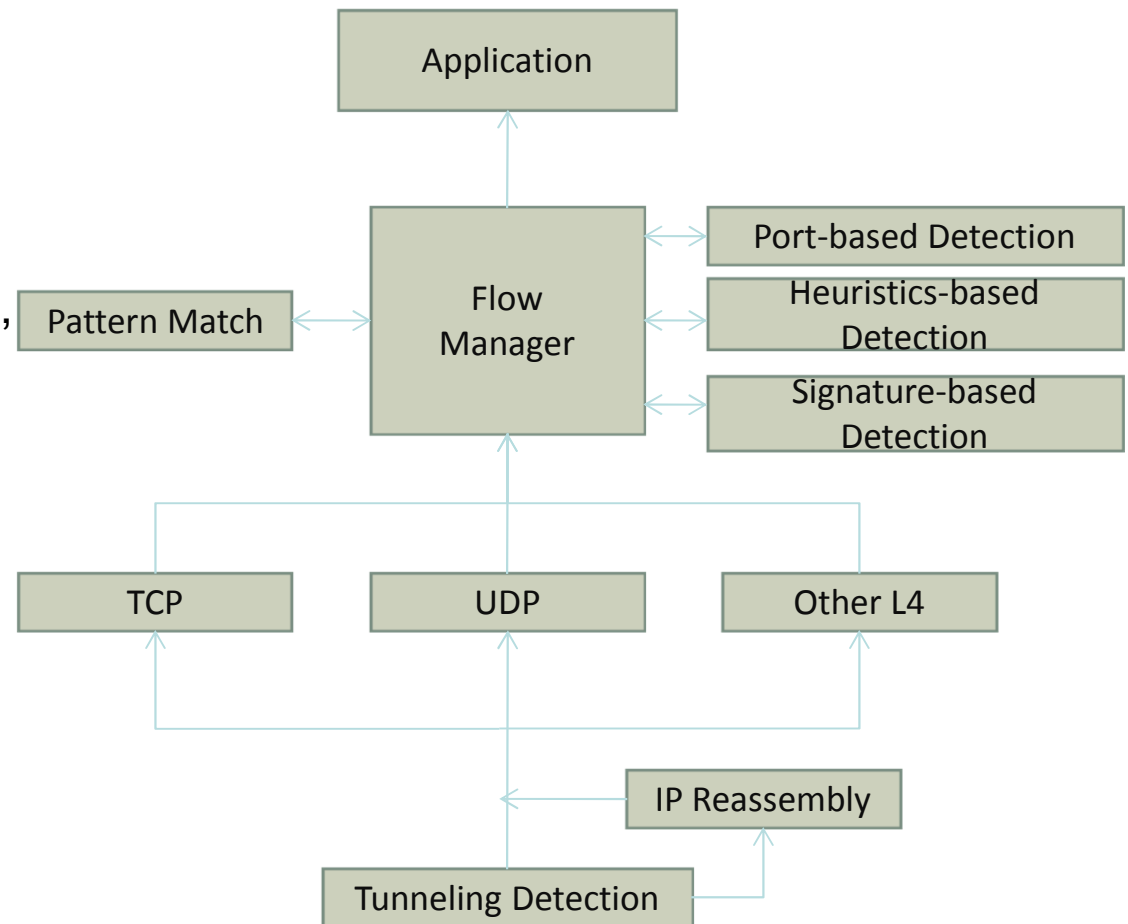


OCTEON: Defined the “Intelligent Networking Processor”? (2005+)

Example Usage: Application & Content Aware DPI Flow



- High throughput and Low latency L2 – L7 processing
 - Reassemble packet flows from individual packets
 - Detect high layer protocols and applications (e.g. Email, ftp sessions, VoIP, Messaging, Multimedia, Conferencing, etc.)
- Based on App/Content types
 - Enforce security, QoS
 - Scan for malware & attacks
- Require comprehensive set of HW acceleration
 - Packet, TCP, RegEx
 - Crypto, De/Compression



This application needs an intelligent processor tuned for networking

Upcoming: OCTEON III CN78XX



Cores:

- 48 cnMIPS III @ 2.5GHz
- Large shared L2 Cache w/ ECC
- Cores, Crossbar, L2 @ 2.5GHz

Memory Controllers:

- 4 x 72b DDR3-2133 & DDR4 w/ ECC
- Up to 256GB, 4-rank x4 DIMMs

OCTEON Coherent Interconnect

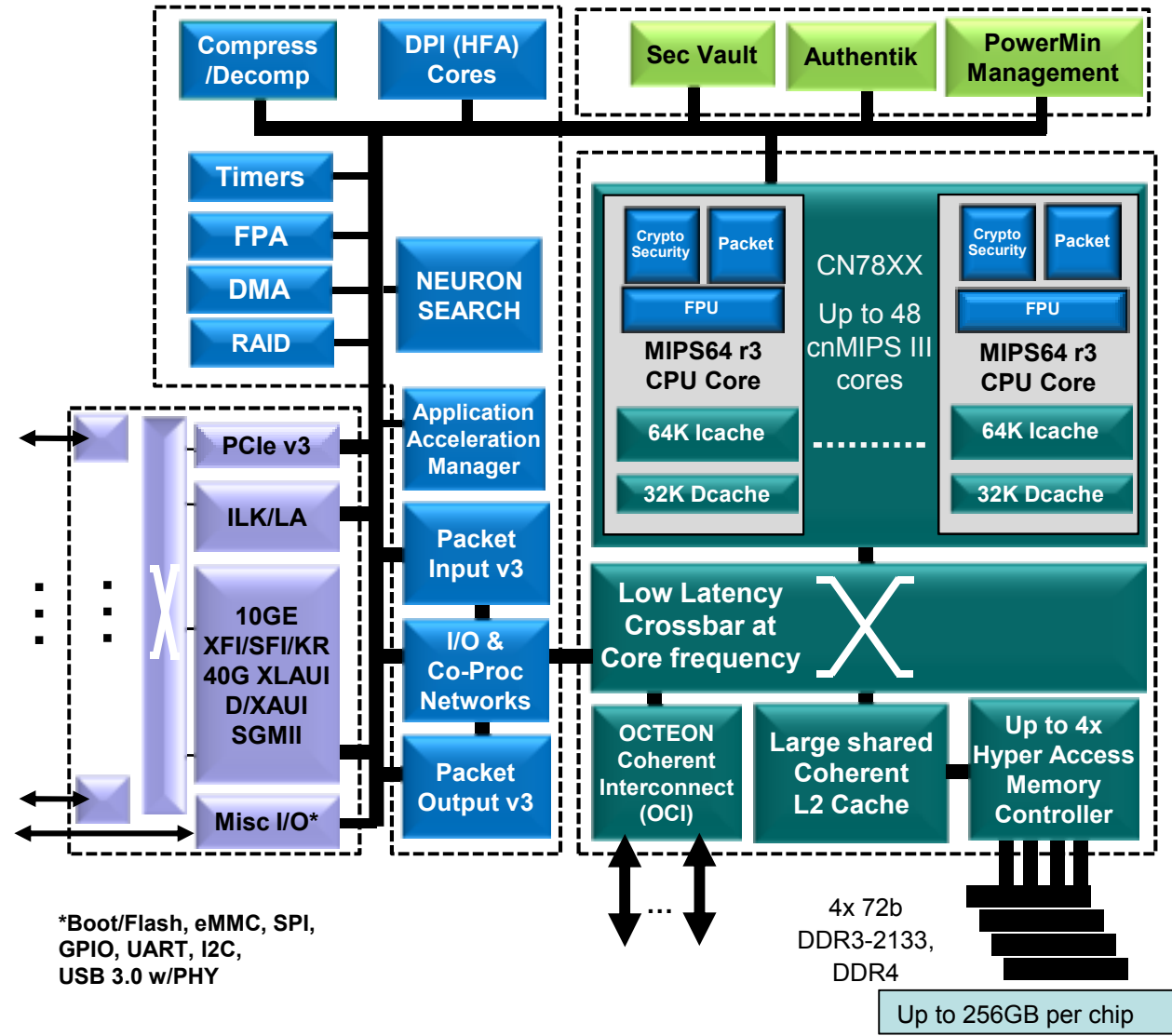
50+ lanes 10+ Gb Serdes

HW Acceleration (Up to 100G+)

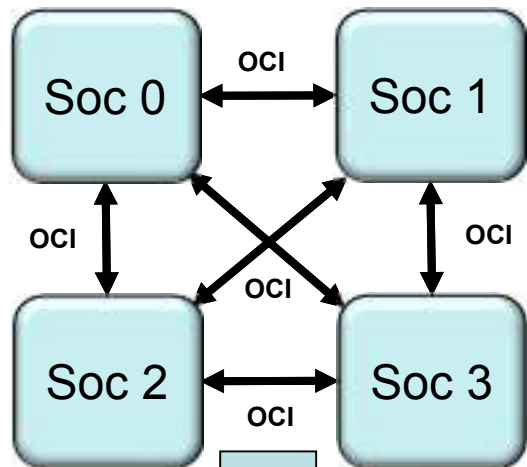
- Packet Processing, QoS, TCP, SCTP, MPLS, FCoE, iSCSI
- Packet Ordering, Schedule, Synchron.
- Security, Compression
- Deep Packet Inspection (HFA)
- Search and ACL Lookup (NEURON)
- RAID, De-Dup

Compatibility

- Backward and Software compatible with all OCTEON families



Upcoming: Multi-socket Solution (OCI)



Four Soc Example



- **OCI connects 2 to 8 OCTEON III SoCs to appear as a single logical multicore processor**
- **Coherency implemented across Cores, Memory, Network, I/O, and coprocessor**
 - Architecture eliminates unnecessary memory copies
 - All connected SoC's can use any coprocessor or I/O
- **OCI enables unmatched scalability along with very low latency. E.g. 4-socket system delivers**
 - Up to 192 cores
 - Up to 480 GHz of compute power
 - Up to 400 Gbps+ of application processing
 - Up to 1 Terabytes of memory
 - Lowest latency for embedded networking applications

Seamless expansion to 800 Gbps+ of packet processing through OCI

Upcoming: SoC's with 64-bit ARM cores

- Project Thunder

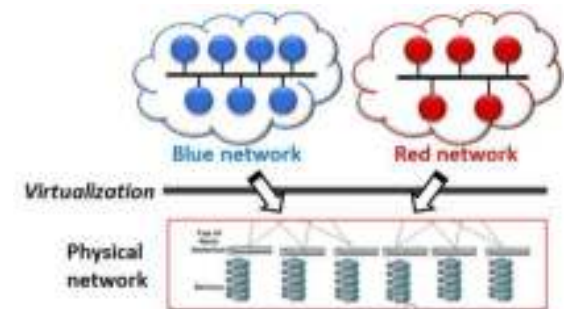
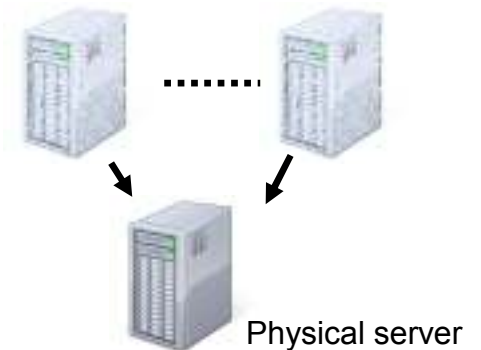


- Full custom cores built from the ground up based on 64-bit ARMv8 Instruction Set
- SoC architectures optimized for cloud and datacenter applications

Upcoming: Hardware for Virtualization

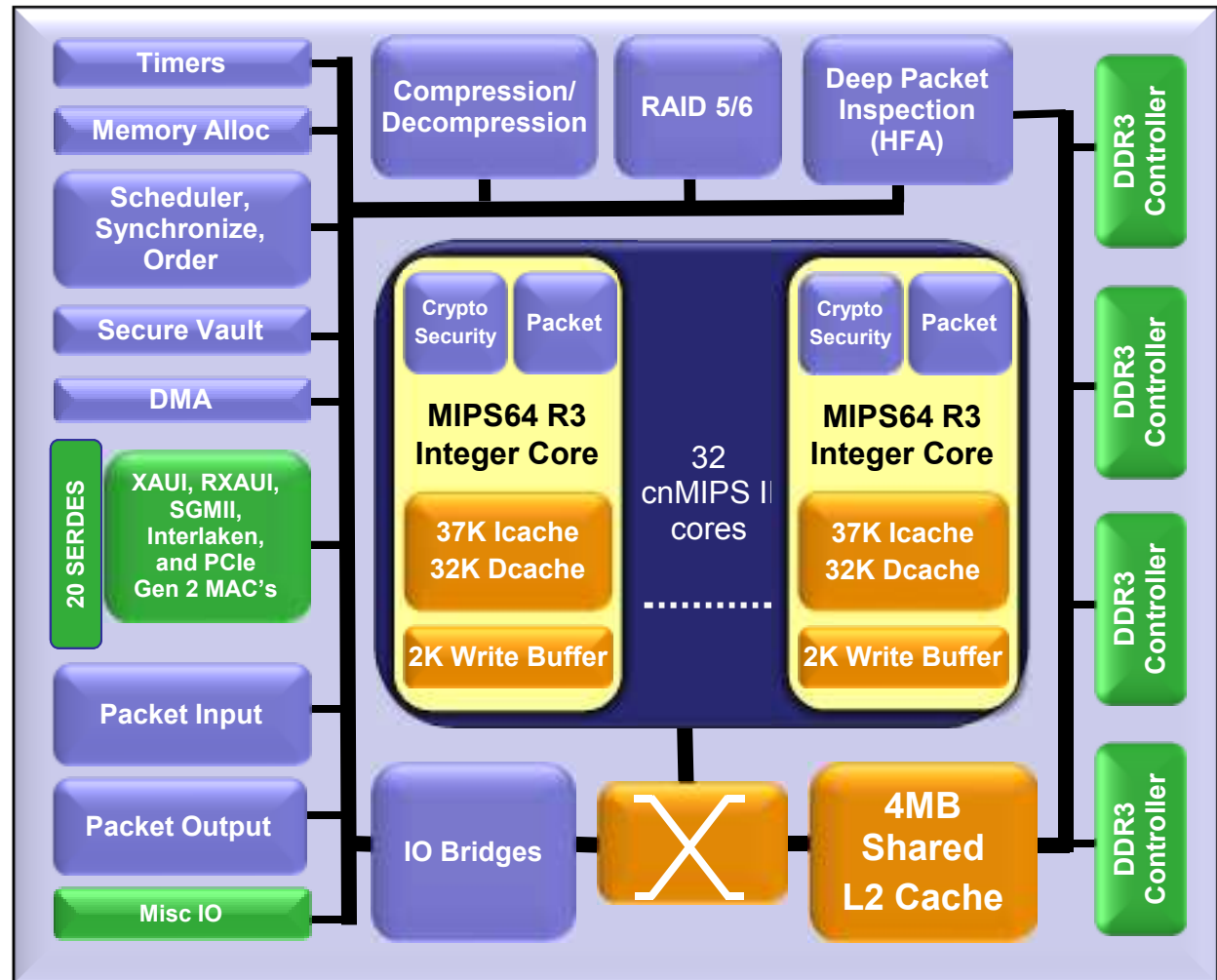


- Server virtualization
 - Virtual machines (VMs) share a physical server
 - Software running on a VM has the illusion that it runs on the entire physical server
- Network virtualization
 - Maintain network connectivity for VMs in a private cloud as if they were connected physical servers inside a campus
 - Software-Defined Networking
 - Overlay techniques, e.g. OpenFlow
- Isolation
 - More cores -> isolation more important



OCTEON II CN68XX Block Diagram

- 32 custom designed MIPS64 cores
- Up to 1.5 GHz
- Up to 96G inst/sec, 40+Gbps
- 4 72-bit DDR3 interfaces up to 1600 MHz data rate
- Optimized for service-rich networking, security, wireless, and storage apps
- HW Acceleration:
 - ✓ DPI acceleration with integrated HFA (RegEx Engine)
 - ✓ Comprehensive crypto algorithms and RNG
 - ✓ TCP, Packet Processing
 - ✓ Compression
 - ✓ RAID5/6, De-dup
 - ✓ Multi-core scaling



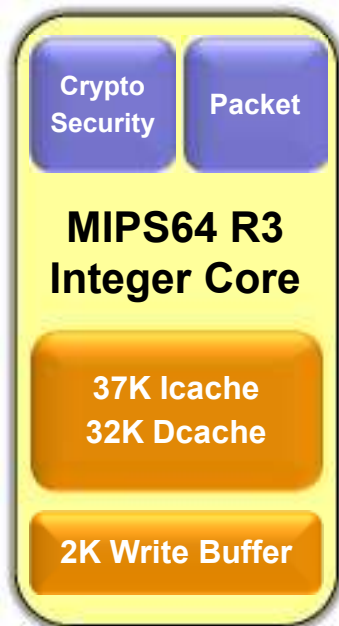
- Cavium Multicore Background and Upcoming
- 32-core OCTEON 68xx Architecture
 - Efficient CPU cores
 - Cache coherence, interconnect & memory bandwidth, chip floorplan
 - Power scalability
 - Hardware work queueing, scheduling, synchronization, and ordering assist
 - Coprocessor acceleration
- 32-core OCTEON 68xx Performance Results

Small CPU Core or Big CPU Core?

- Many potential Big Core features:
 - Huge caches
 - Very high frequency, deep pipeline
 - Many-way issue
 - Out-of-order issue
 - Floating-point
 - ...

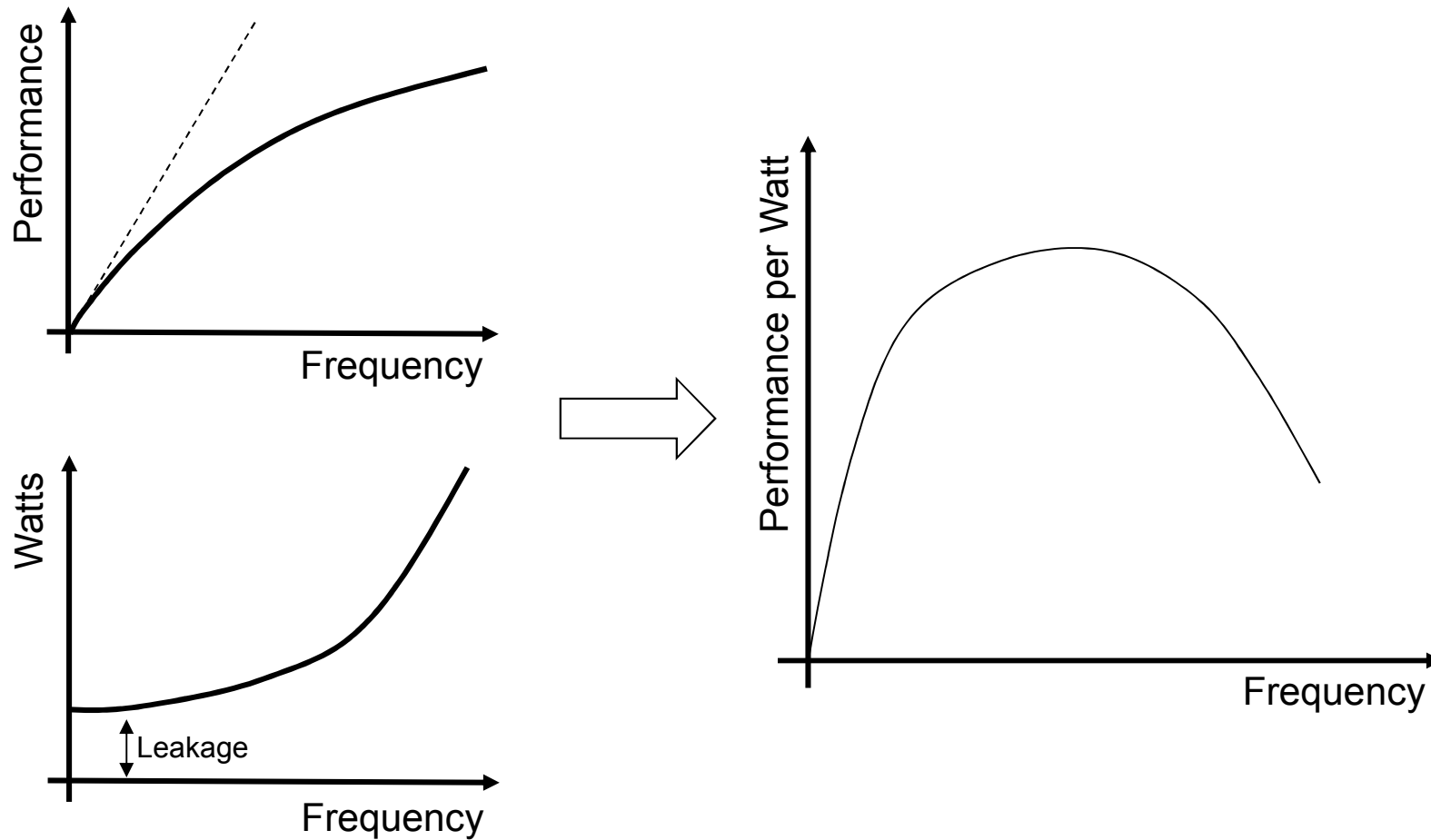
- Important questions:
 - Does the feature add more performance than area/power?
 - Is the feature difficult or expensive to implement, take to production, and support?

OCTEON II CPU (i.e. cnMIPS II) Design

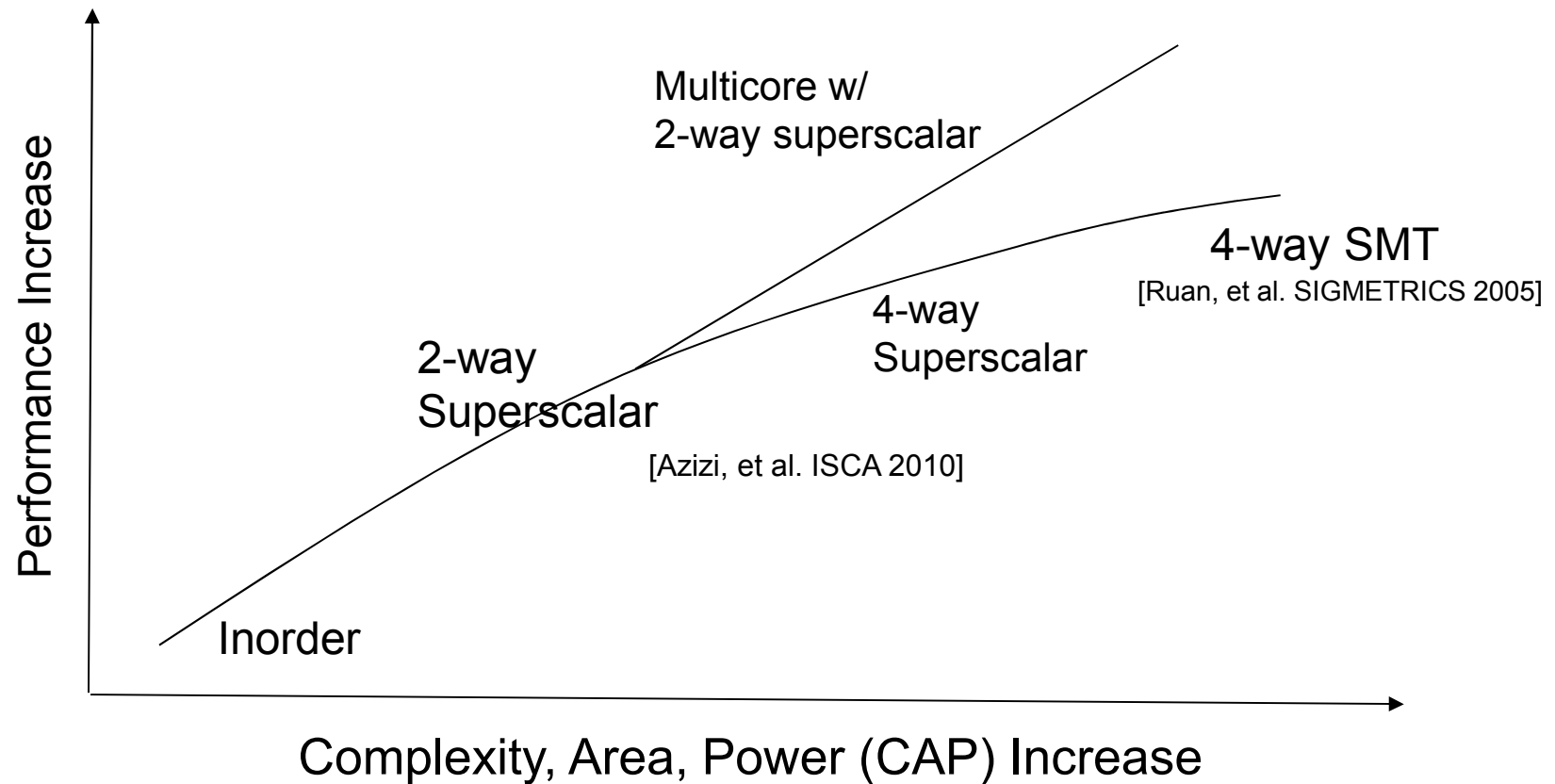


- cnMIPS II Core Goals:
 - General-purpose, industry-standard 64-bit ISA
 - Great fit for networking, security, wireless
 - Excellent MIPS/area & MIPS/watt
 - Use multi-core to scale product line up and down
- cnMIPS II Core Non-goals:
 - Highest power
 - Highest cost
 - Greatest complexity (longer implementation time)
 - Largest customer support cost
- Not directly mentioned:
 - Frequency
 - Single-thread performance
 - ...

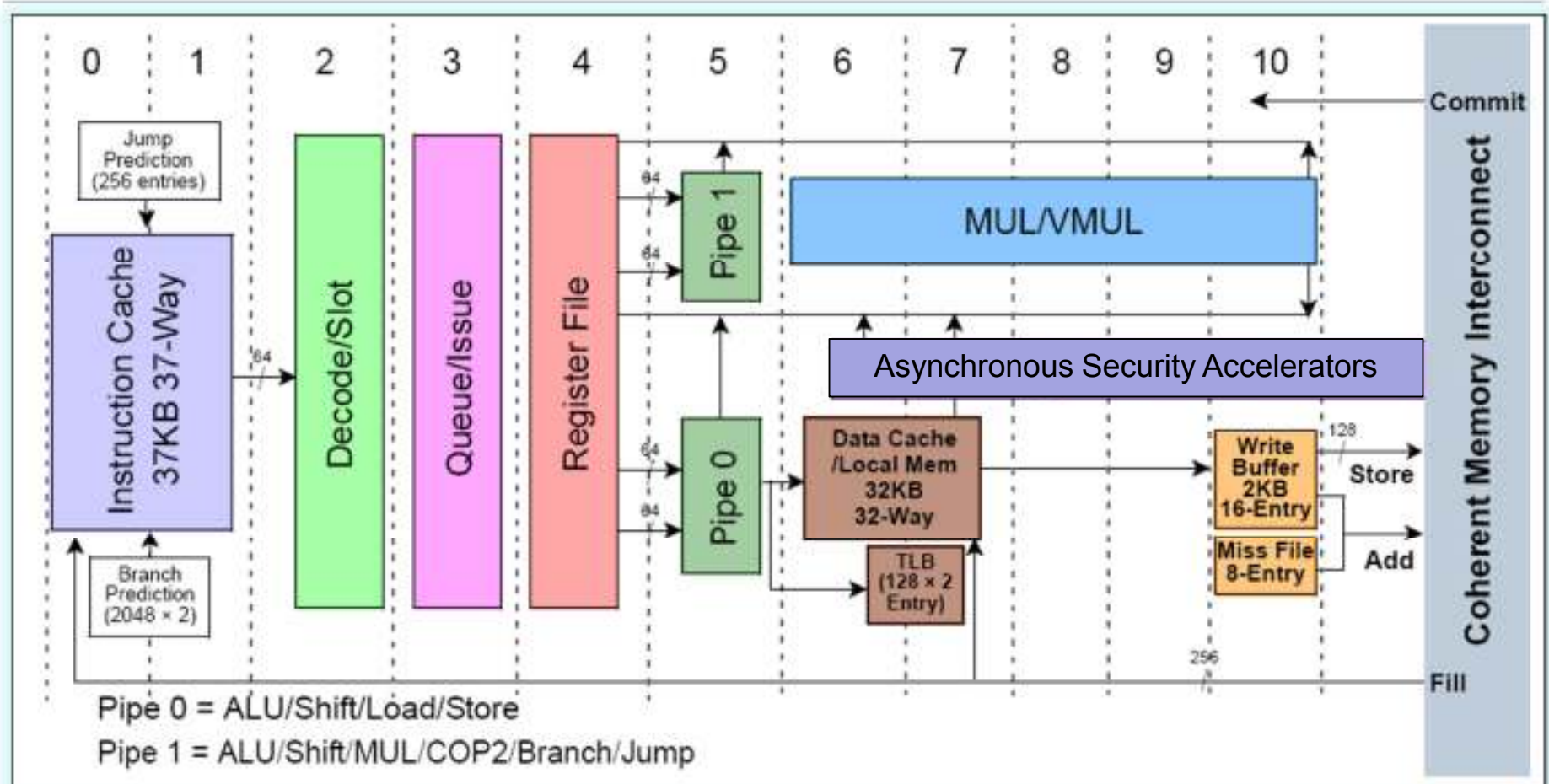
Frequency Effects



Choosing the Issue Width of Pipeline



cnMIPS II Core 8+ Stage Pipeline



- Shipping at up to 1.8 GHz in 65nm
- Thread-dedicated resources = very deterministic CPU performance
- Highly-associative L1 caches = equivalent miss rate to much larger caches

Optimized Instruction Set Architecture



64-bit Addressing

- Efficiently manage large data structures and tables for both data and control planes

Packet Processing Operations

- Efficient packet processing, 80+ new instructions added

CRC Instruction

- Calculate any 32-bit or shorter polynomials

Memory Atomic Operations

- Efficiently implement many statistics counters in memory

Prefetch, Cache hints, Zero Cache Block Instructions

- Optimize memory access performance
- Avoid or minimize bandwidth consumption

Cache Manipulation Instructions

- Lock or invalidate cache lines

Hash Instructions

- SHA1, SHA2, MD5 accelerate comparison of large data blocks

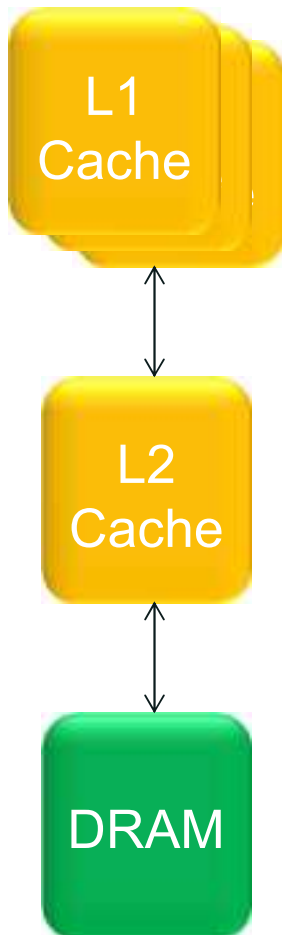
Conditional Move Instructions

- Avoids conditional branches

Optimized for Layer 2-7 Intelligent Networking Applications

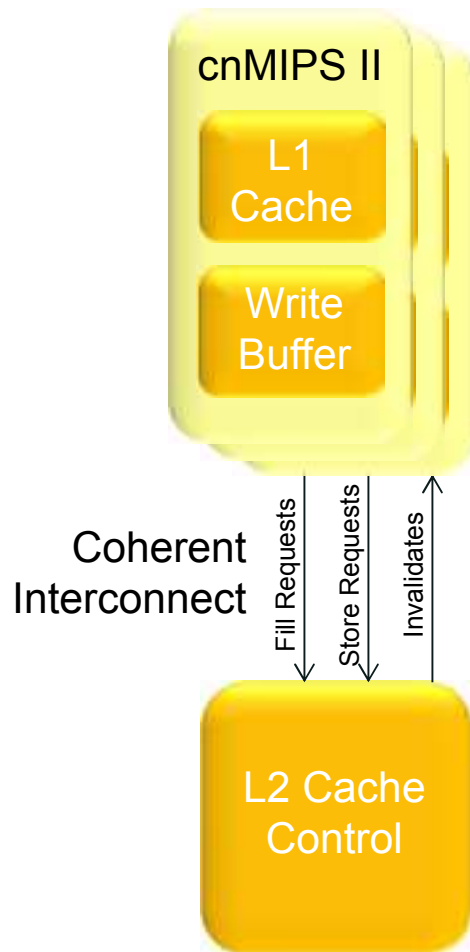
- Cavium Multicore Background and Upcoming
- 32-core OCTEON 68xx Architecture
 - Efficient CPU cores
 - Cache coherence, interconnect & memory bandwidth, chip floorplan
 - Power scalability
 - Hardware work queueing, scheduling, synchronization, and ordering assist
 - Coprocessor acceleration
- 32-core OCTEON 68xx Performance Results

OCTEON Cache Policies



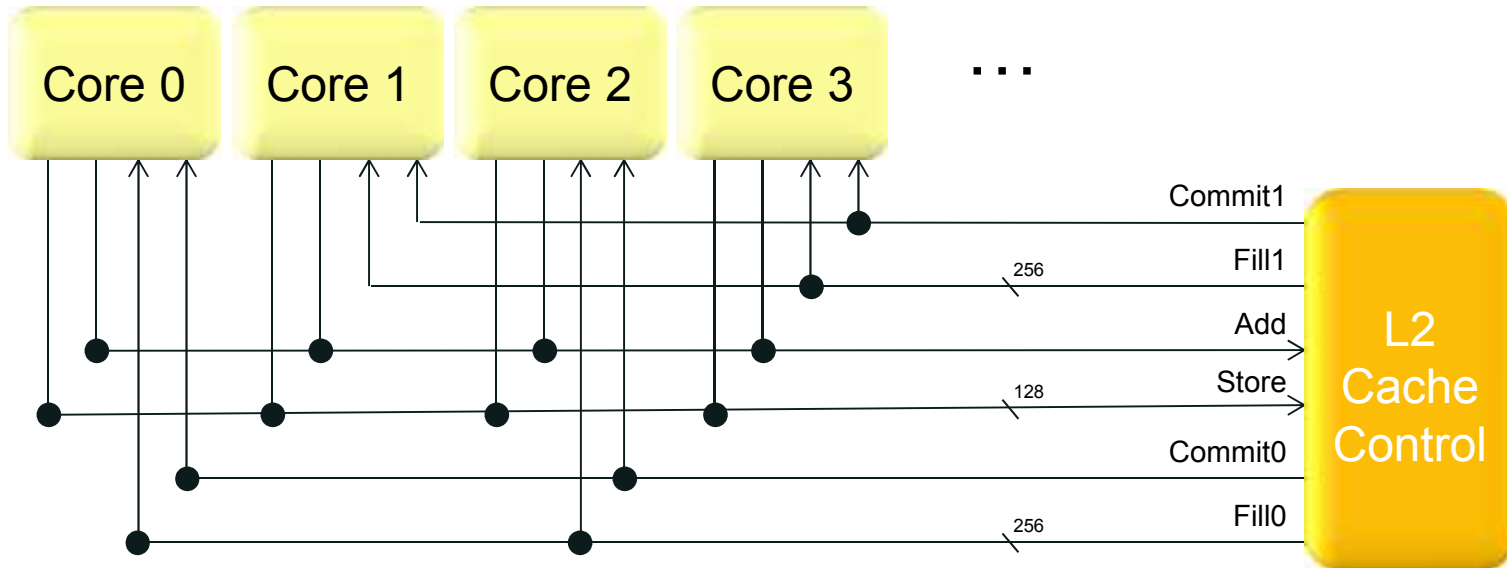
- L1 <-> L2 Cache: Write-through
 - Excellent performance for networking and mobile applications
 - Minimal per-CPU-core cost
 - Simple and highest performance
 - Lowest possible read latencies
 - Allows many outstanding stores, optimizations
 - Automatic L1 error correction
- L2 Cache <-> DRAM: Write-back
 - Standard DDR3 DRAM DIMM's are highest performance with block transfers
 - Minimizes required DRAM bandwidth
 - Don't-write-back feature (e.g. for most of packet data) plus additional cache hints

OCTEON L1<->L2 Coherence and Memory Model

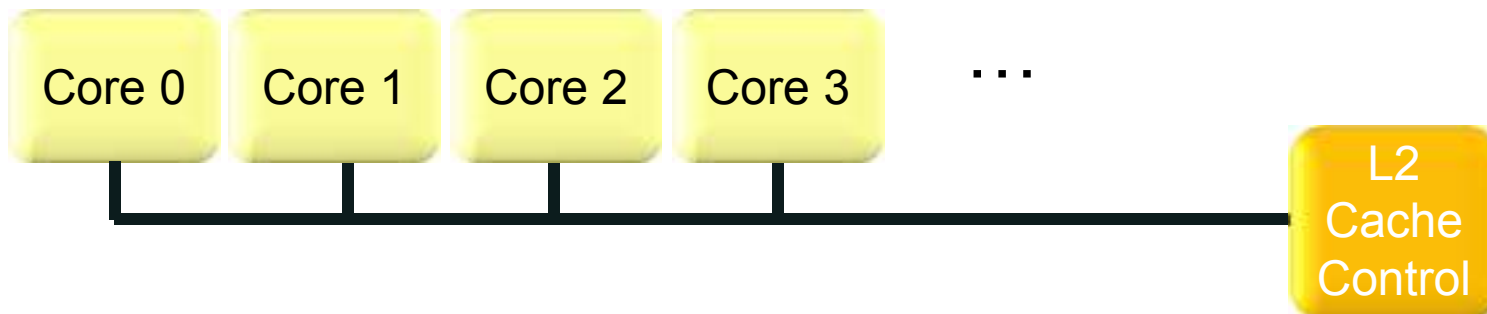


- Write-through, write-invalidate coherence protocol
- L2 Cache Controller is the coherence point
 - L2 controller tracks L1 cache contents
 - Invalidates to maintain L1 coherence
- Aggressive write-buffering in cnMIPS II cores eliminates writes
 - 2 KB merging write buffer
 - Fully-coherent, loosely-consistent memory model
 - Page-wise hints to eliminate write-buffer flushes of private data

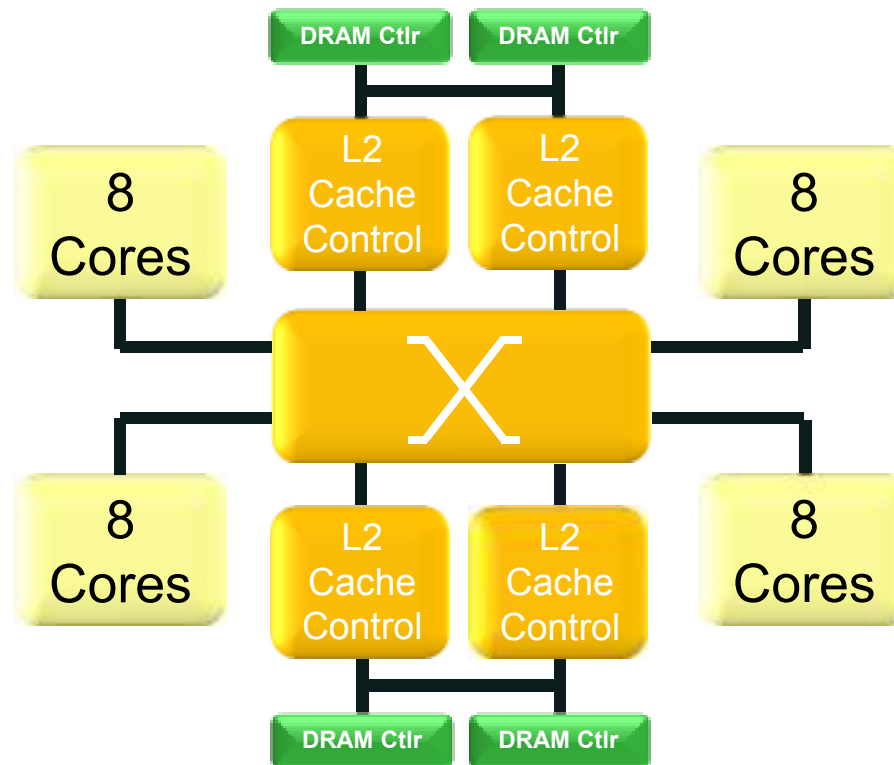
Lower Core Count OCTEON Coherent Interconnect



Redrawn:

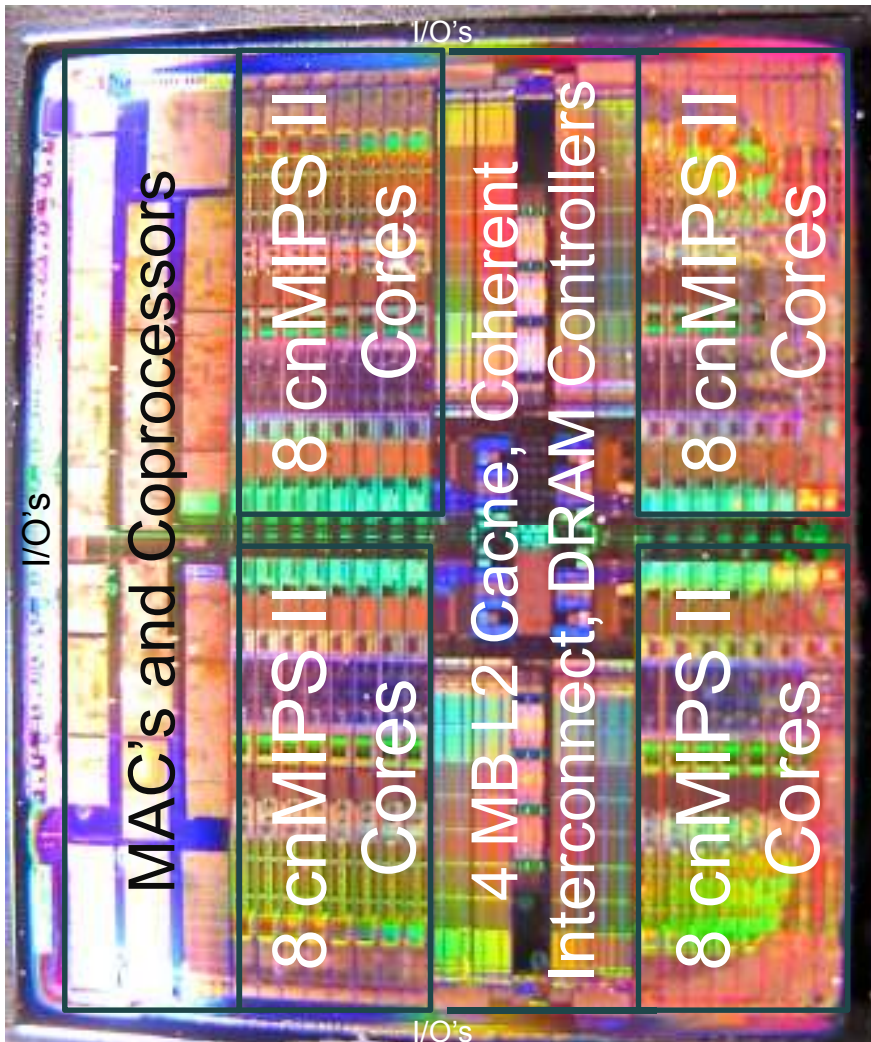


32 Core OCTEON Coherent Interconnect



- Crossbar interconnect easily scales to 32 cores
- Optimized for both low latency and high bandwidth
- Flat, deterministic latency profile
- Best combination of scalability and low-power

32-Core OCTEON



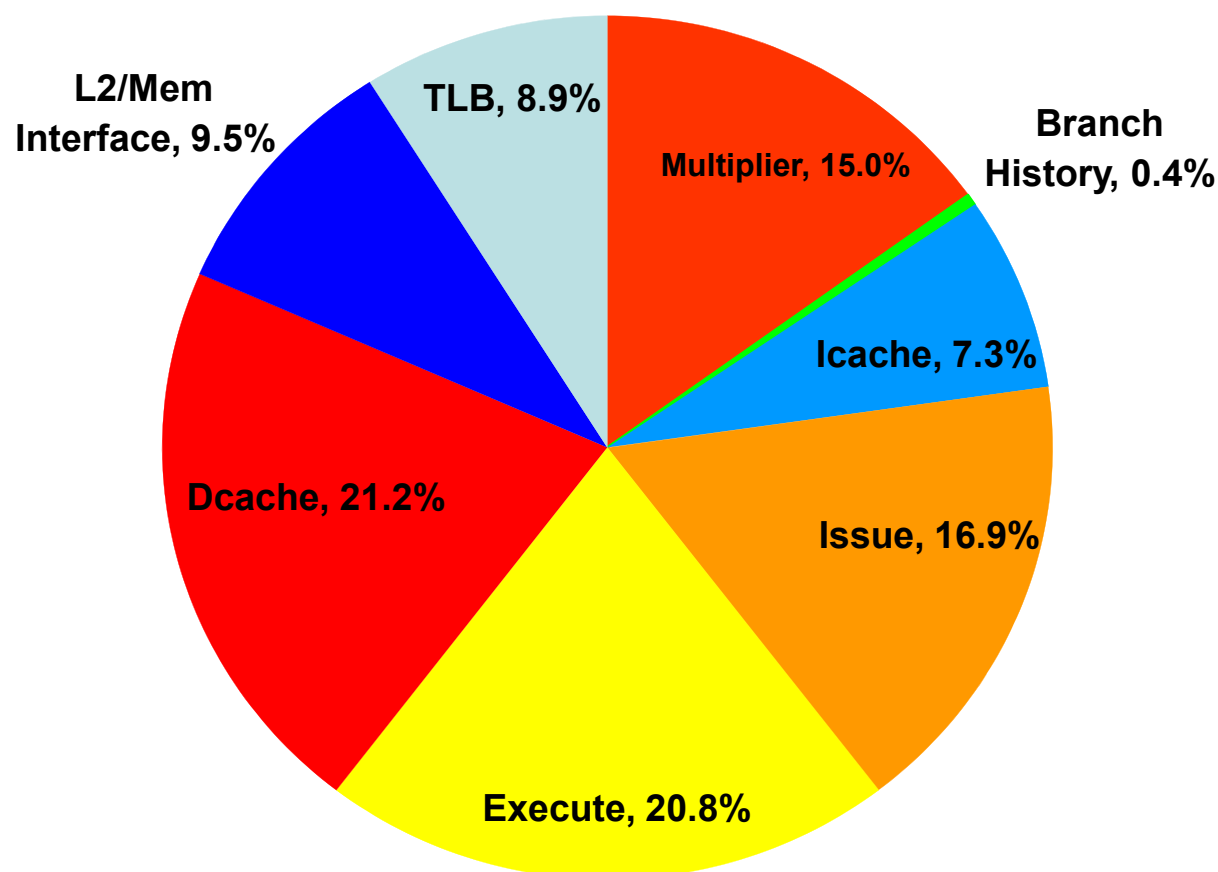
Transistors	700 Million
Power	40-65W
Frequency	1.6GHz
Voltage	1. 0V
Process	65nm CMOS
Metal Layers	10 copper
MIPS Cores	32
Icache (each core)	37KB
Dcache (each core)	32KB
L2 Cache	4MB

- Cavium Multicore Background and Upcoming
- 32-core OCTEON 68xx Architecture
 - Efficient CPU cores
 - Cache coherence, interconnect & memory bandwidth, chip floorplan
 - Power scalability
 - Hardware work queueing, scheduling, synchronization, and ordering assist
 - Coprocessor acceleration
- 32-core OCTEON 68xx Performance Results

Core Power Doing Powerloop



Core Peak Power: 1.28W @1.6GHz

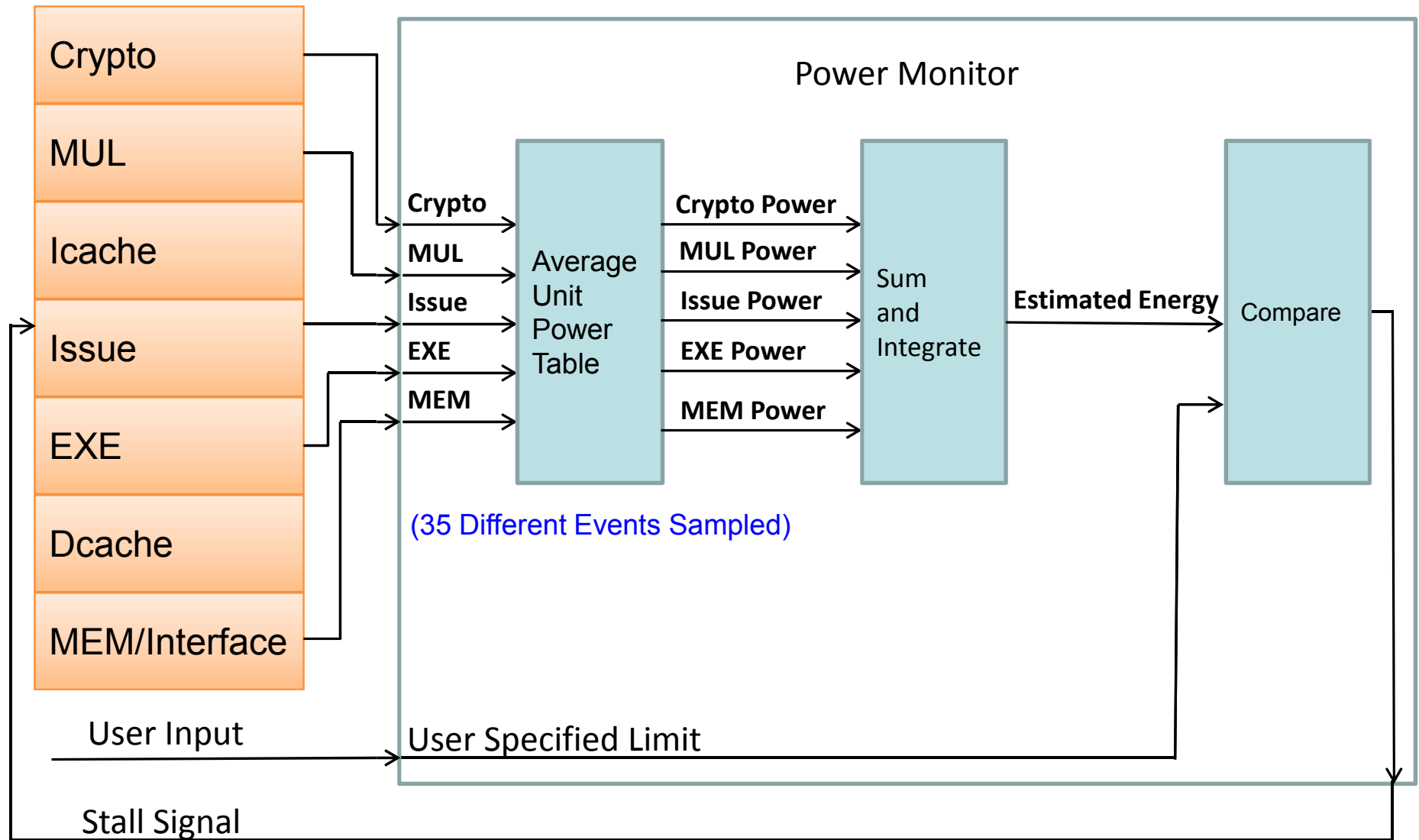


OCTEON 68xx Power Optimizer Technology



- Per-CPU dynamic power consumption estimates
- Per-CPU dynamic power threshold
 - Hardware threshold
 - Software threshold in a register
 - Software can quickly and easily change it
- Hardware threshold can limit worst-case power
- Software threshold can:
 - Reduce average power (open loop)
 - Reduce temperature (closed loop with thermal sensor)
 - Other ...

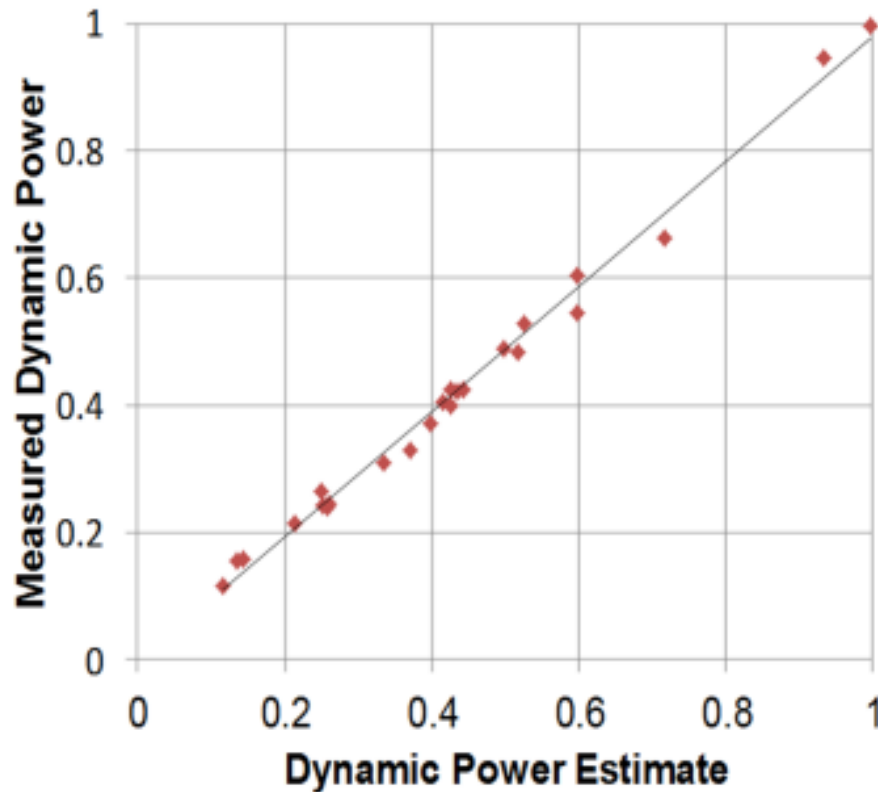
Core Power Management



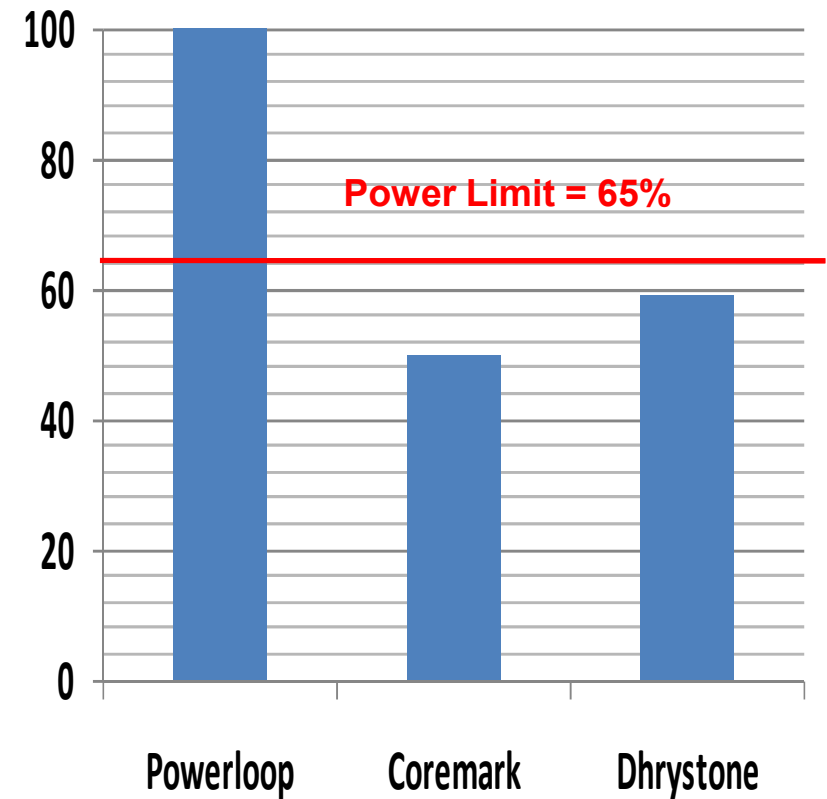
Power Management and Performance



Correlation Measured vs Estimated Power



Core Power as a Percentage of Max



OCTEON Power Management Comparison

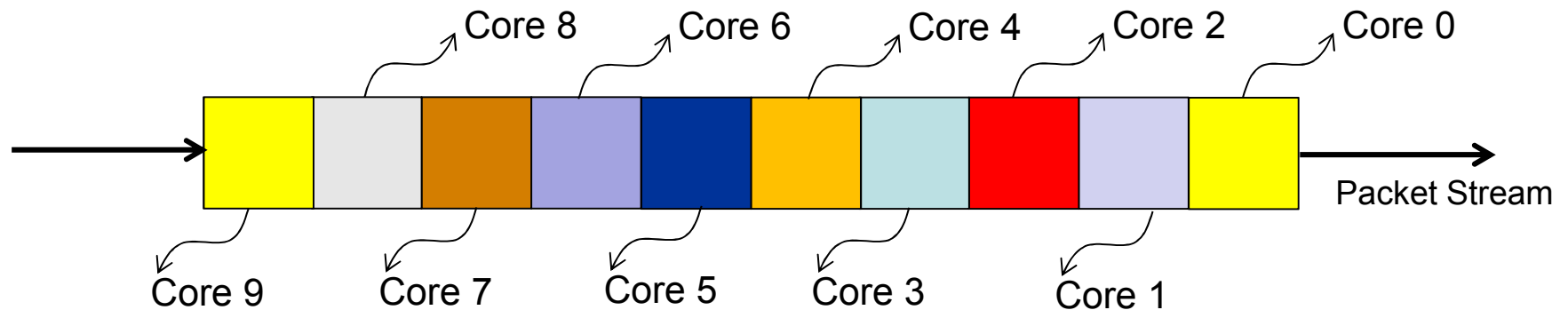


- Advantages compared to dynamic voltage and frequency scaling (DVFS):
 - Very fine-grained core-by-core power control
 - A low power application is not penalized
 - Frequency reduction affects all applications
 - 95+% of applications don't achieve 80% of max power
 - Power optimizer settings can change instantly with minimal software interruption
 - Simpler chip and system design
 - Voltage and frequency do not need to change

- Cavium Multicore Background and Upcoming
- 32-core OCTEON 68xx Architecture
 - Efficient CPU cores
 - Cache coherence, interconnect & memory bandwidth, chip floorplan
 - Power scalability
 - Hardware work queueing, scheduling, synchronization, and ordering assist
 - Coprocessor acceleration
- 32-core OCTEON 68xx Performance Results

Overcoming Amdahl's Law in Multicores

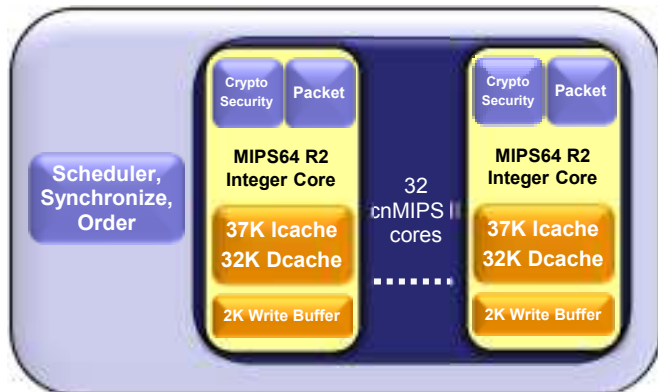
animated slide



Work Queue Creates Sequential Bottleneck

OCTEON's overcome this bottleneck:

- Hardware packet classifier that tag packets
- Hardware load balances packets among cores using tags
- Cores pick up packets for processing using tags

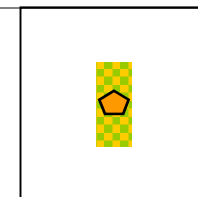
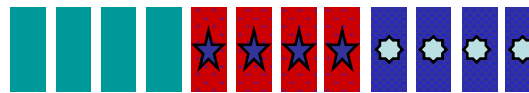


- Work queueing
 - Unlimited-size queues for work
 - Work can be created by software
 - Work can be created by hardware
 - e.g. packet arrival
- Work/Packet Ordering
- Automatic synchronization and lock-removal
- Dynamic work scheduling
 - Hardware selects from amongst input queues
 - Quality of service
 - Different cores can receive different work
 - Integrated with ordering and synchronization
 - Work proceeds only when ordering and synchronization allows

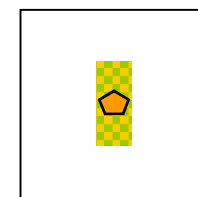
Packet Processing without Schedule/Sync/Order Hardware



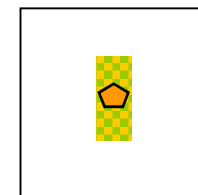
Packets from different TCP flows



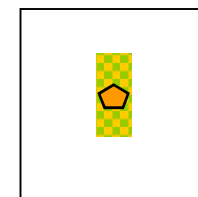
Core #1
Executing
Flow #1



Core #2
Stalled
Flow #1



Core #3
Stalled
Flow #1



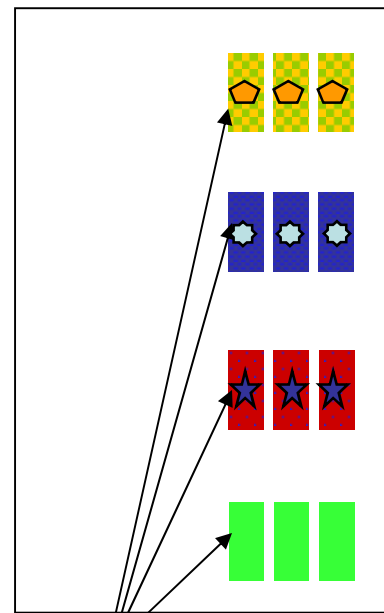
Core #N
Stalled
Flow #1

Packet Processing with Schedule/Synch/Order Hardware

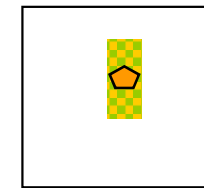


Packets from multiple TCP flows

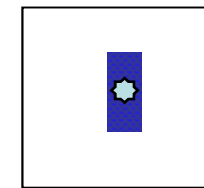
Packet Order Unit



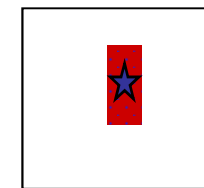
Multiple inflight queues



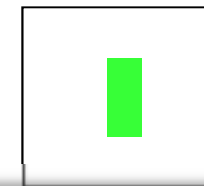
Core #1
Executing
Flow #1



Core #2
Executing
Flow #2



Core #3
Executing
Flow #3



Core #N
Executing
Flow #N

OCTEON provides lock-free access to packets, removing the sequential bottleneck for packet processing

Agenda



- Cavium Multicore Background and Upcoming
- 32-core OCTEON 68xx Architecture
 - Efficient CPU cores
 - Cache coherence, interconnect & memory bandwidth, chip floorplan
 - Power scalability
 - Hardware work queueing, scheduling, synchronization, and ordering assist
 - Coprocessor acceleration
- 32-core OCTEON 68xx Performance Results

OCTEON's Accelerator Usage



OCTEON's offload compute-intensive functions

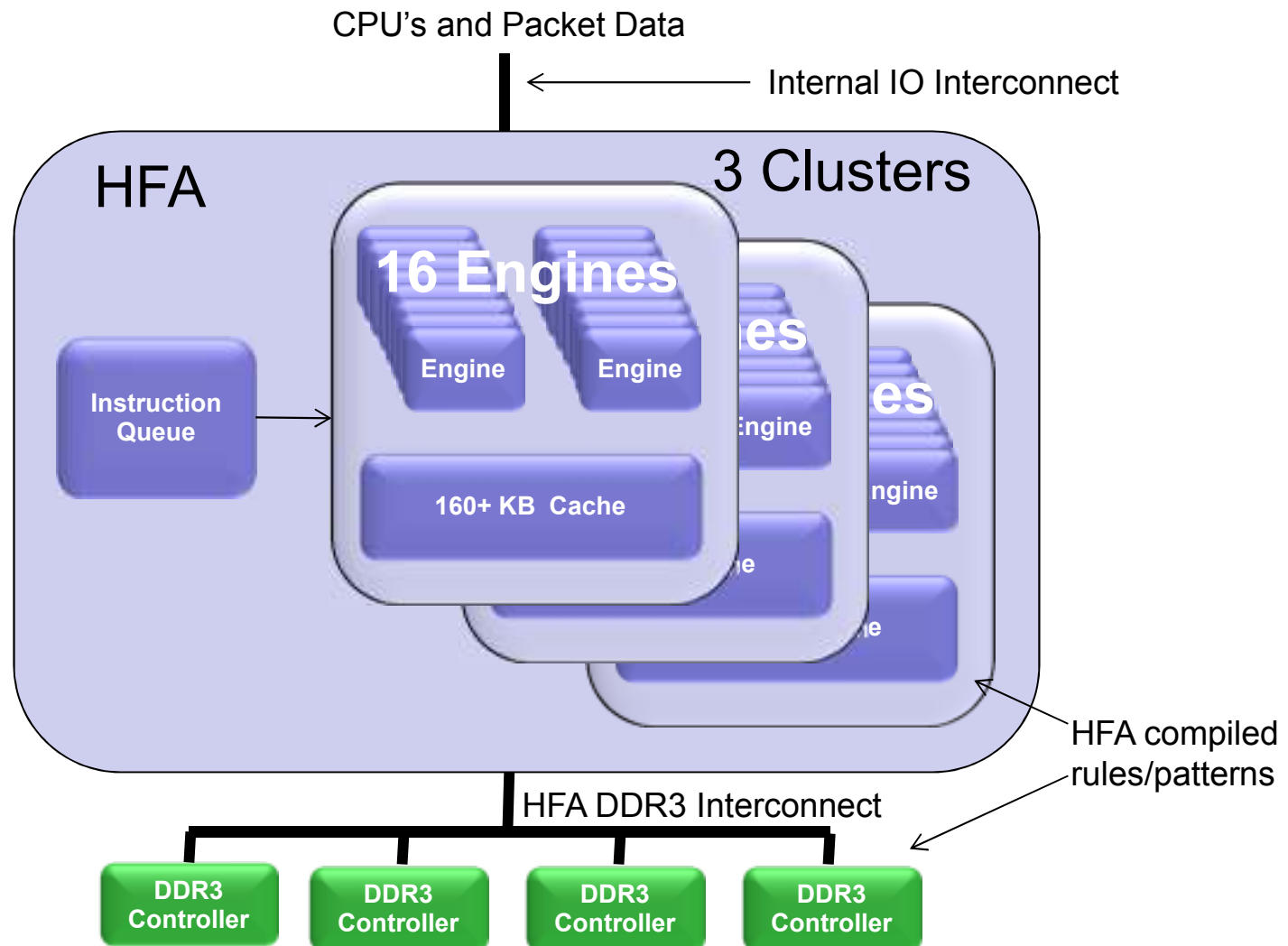
- Encryption & decryption
- SSL processing
- Memory allocation
- Compression & decompression
- Network search & lookup
- Regular expression processing
- ...

Implication for processor cores

- Accelerators reduce the need for power-inefficient complex cores

- Many applications require Deep-Packet Inspection (DPI):
 - Intrusion detection/prevention, Packet classification, ...
- We focus on pattern matching here
 - DPI may also require packet, TCP, and other processing that can be accelerated by other OCTEON coprocessors
 - The percentage of data scanned for matches varies for different applications
 - a few percent (e.g. Application Recognition) to most packet bytes (e.g. Anti-Virus, IPS)
- Patterns/rules are often regular expressions
 - Pre-compiled into hardware state machines
 - Deterministic Finite Automata (DFA) and Non-deterministic Finite Automata (NFA)
- Includes graph compression and caching to maximize coverage and performance

68xx Deep Packet Inspection HFA Hardware

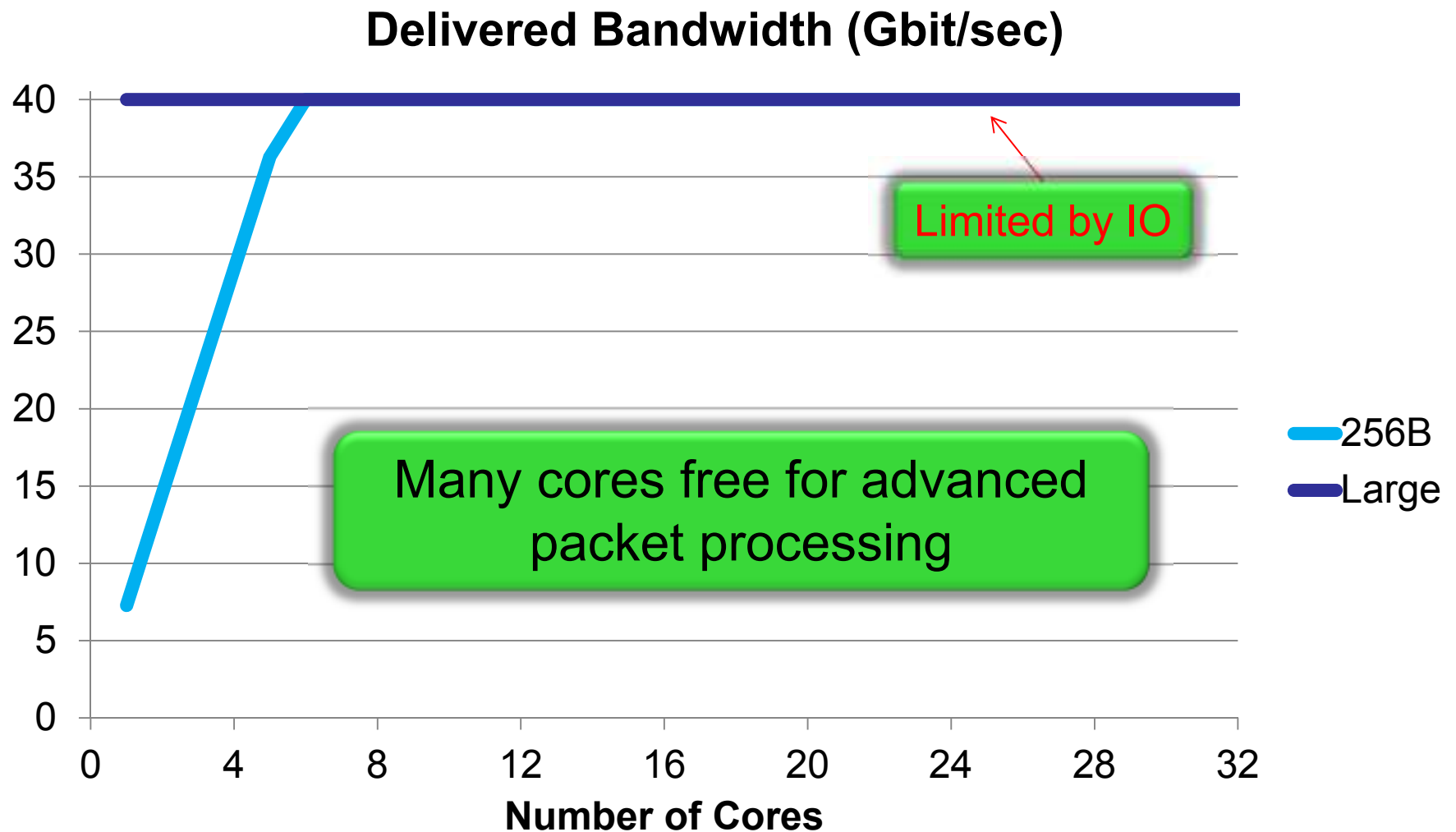


Agenda

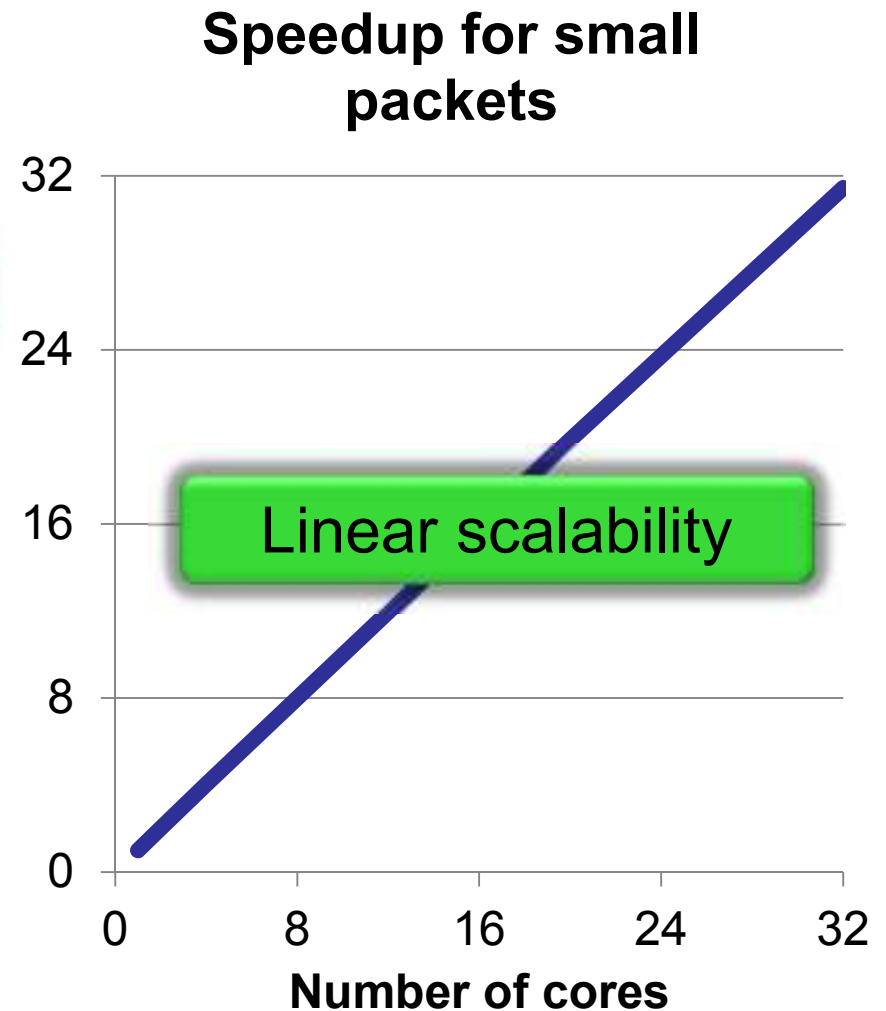
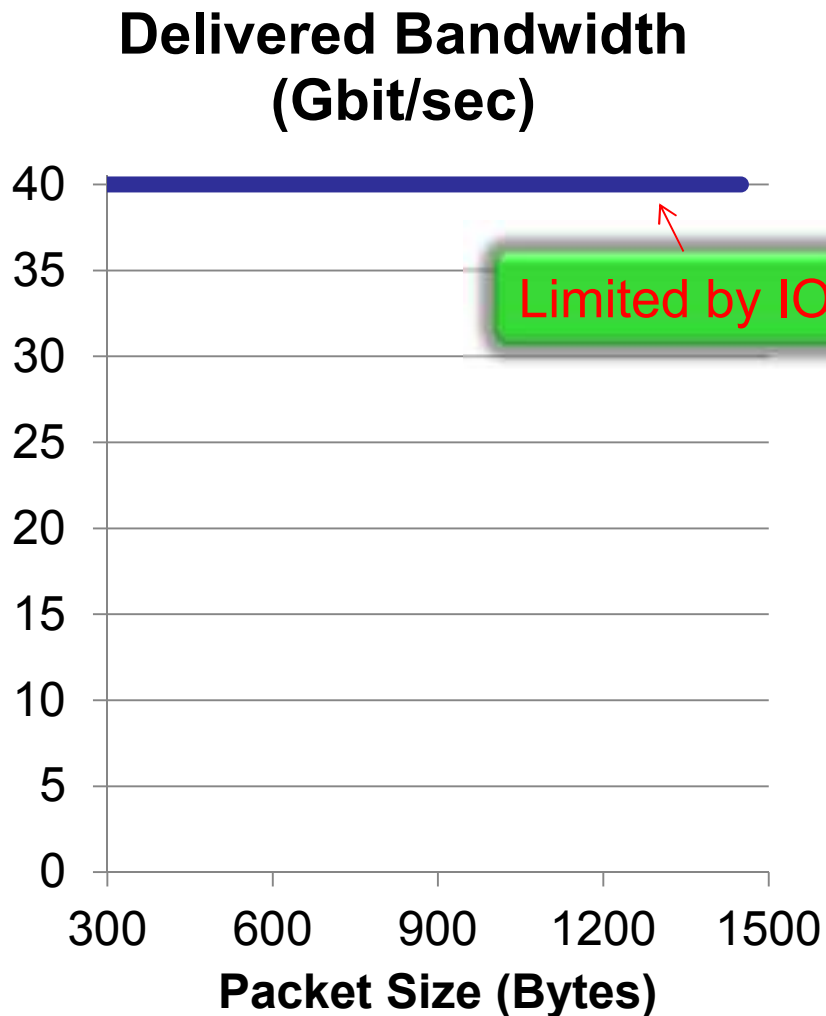


- Cavium Multicore Background and Upcoming
- 32-core OCTEON 68xx Architecture
 - Efficient CPU cores
 - Cache coherence, interconnect & memory bandwidth, chip floorplan
 - Power scalability
 - Hardware work queueing, scheduling, synchronization, and ordering assist
 - Coprocessor acceleration
- 32-core OCTEON 68xx Performance Results

IPv4 Packet Forwarding



Full IPSEC Application



Conclusions



- Cavium is applying multicore products to networking, server, and other markets

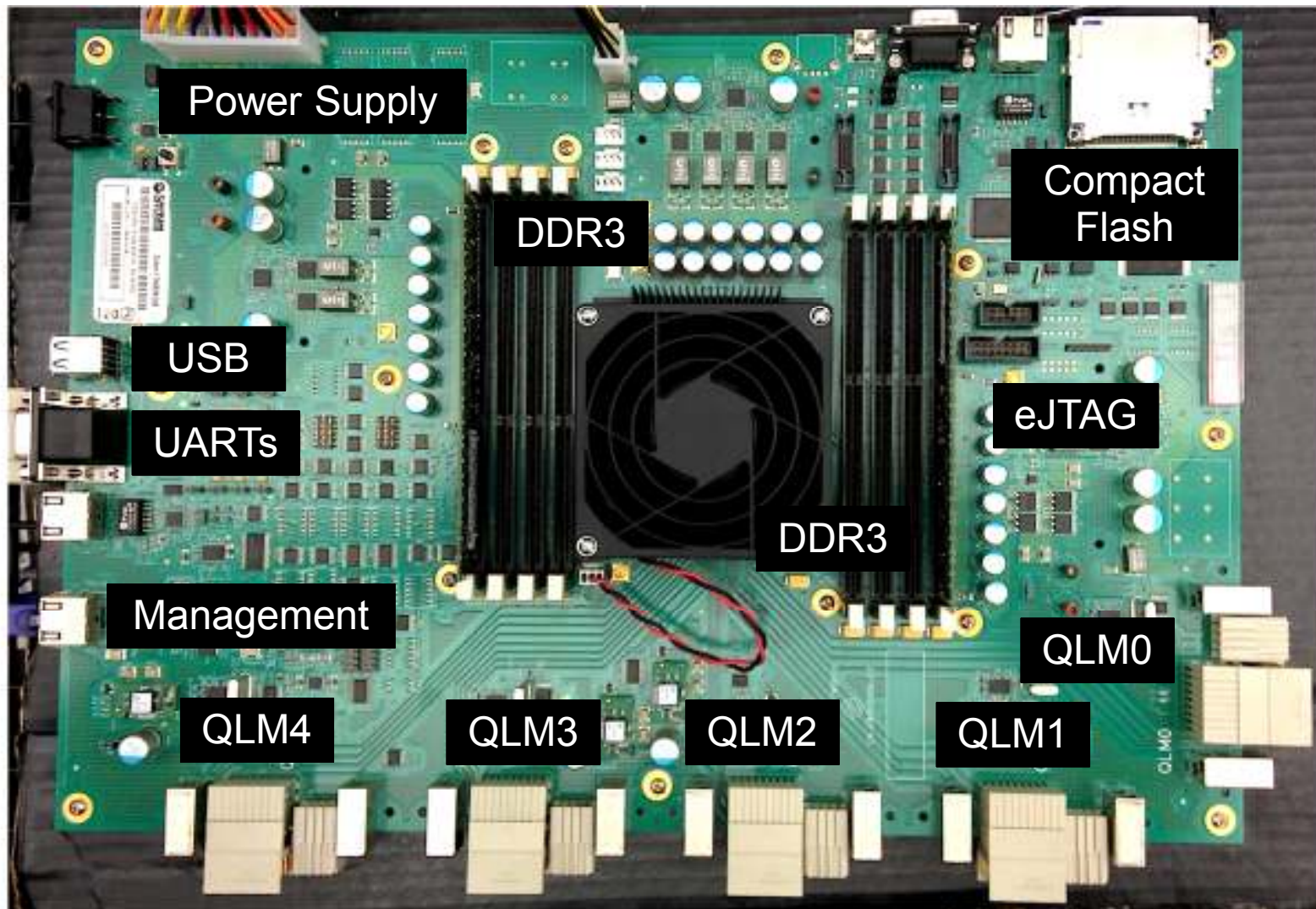
- The OCTEON architecture includes:
 - Efficient CPU cores
 - Scalable Interconnects
 - Power Scalability
 - Hardware Queueing, Scheduling, Synchronization, and ordering assist
 - Coprocessors suitable for target markets

Questions?



OCTEON 68xx Evaluation Board, Program, and Potential Projects

OCTEON II CN68XX

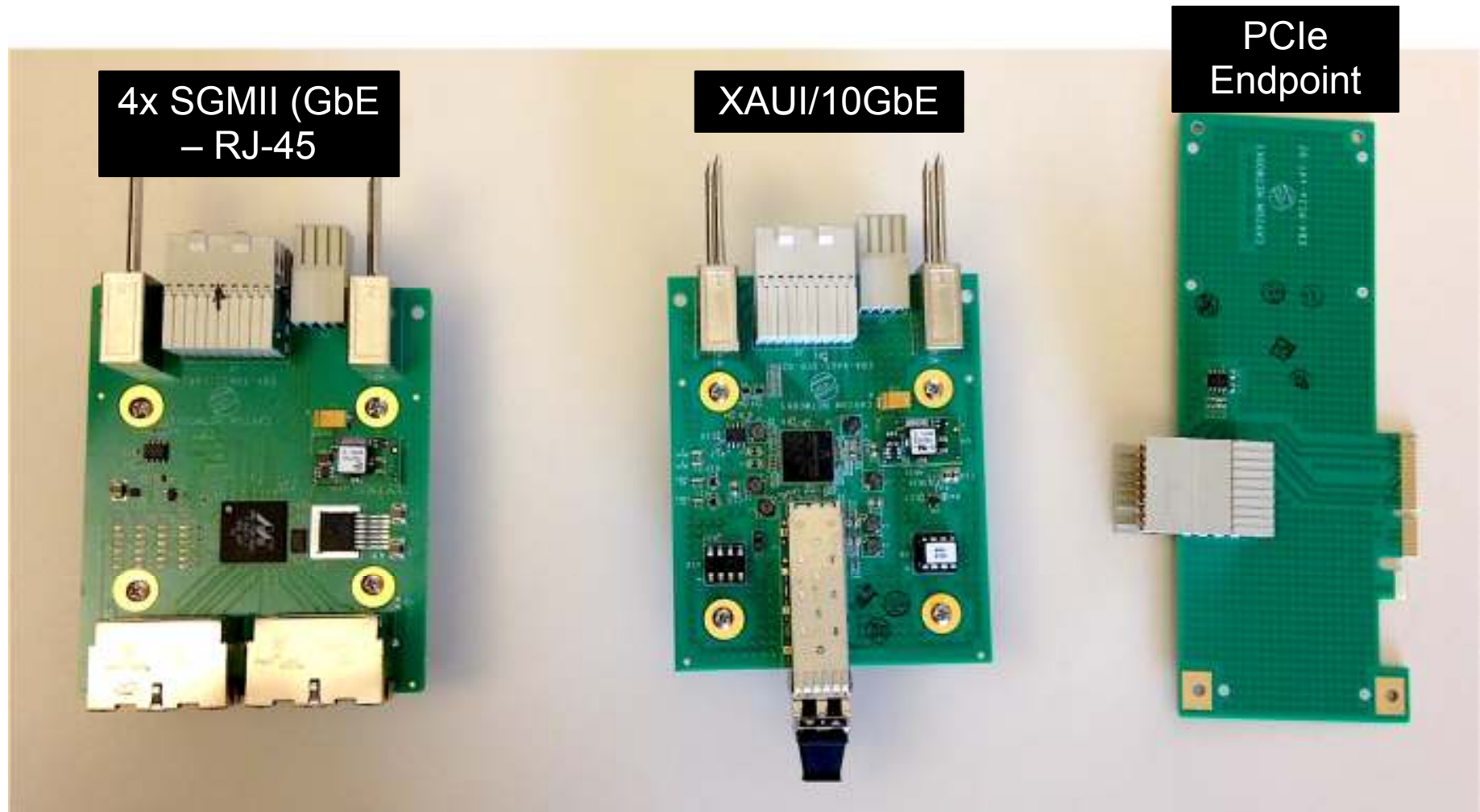


CN68XX EBB Specifications



- Powered via standard ATX power supply
- 5 QLM (Quad Lane Module) Connectors for I/O Modules
- Support for up to 8 DDR3 DIMMS
- 2 Serial Ports for console access and debugging
- 2 USB Host Ports
- Ethernet Management Port
- EJTAG Connectivity for debugging
- Compact Flash Storage

QLM SERDES Modules



QLM Module for PCIe Root Complex is also included (not shown)

OCTEON II Program – Spring 2013

- 5 Universities Using the OCTEON II for Class Projects
 - Cornell (Prof. Jose Martinez)
 - CMU (Prof. Onur Multu)
 - Harvard (Prof. David Brooks)
 - MIT (Prof. Daniel Sanchez)
 - Uppsala (Prof. Stefanos Kaxiras)
- OCTEON Workshop - Boston in May (Date/Venue TBD)
 - Student Presentations of Projects - OCTEON Trophy for Best Paper
- Annual “Global Multicore Challenge”
 - Programming Contest Open to All Undergraduates
 - Look for Details in March

Program Resources



- Visit the Cavium University Program Website
 - www.university.cavium.com
 - OCTEON Resources, Text Books, White Papers
- **Sign Up on the Cavium Users Group Website**
 - **www.cnusers.org**
 - Join the University Discussion Group to Collaborate with other Students
 - Blog about your project, share results, network with others
- Sign Up on the Cavium University Facebook Page
 - www.facebook.com/CaviumUniversityProgram
 - 1600 students from 100 universities
 - Industry and University Related Content, Feel Free to Comment/Post
- Contacts - Program People
 - Jim Ballingall, Program Director (jim.ballingall@cavium.com)
 - Binitha Surendran, Program Manager (binitha.surendran@cavium.com)
 - Gregg Bouchard, Program Engineer (gregg.bouchard@cavium.com)

Suggested Spring 2013 Projects



- “Bump in the Wire” Applications
 - Monitor N Packet Flows on N Processors
 - Use Accelerators for Packet Inspection and Pattern Matching
 - Intrusion and Malware Detection

- Performance/Power Scaling Tradeoffs and Management
 - Optimize Performance and Power for N Cores
 - Vary the Number of Cores and Workloads
 - Off-load Tasks to Accelerators
 - Optimize Parallel Program Code for Specific Workloads