

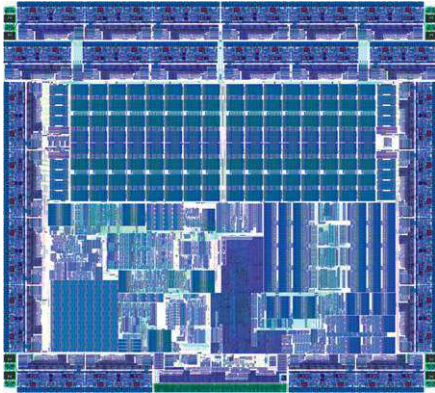


**FocalPoint**

**A Low-Latency, High-Bandwidth  
Ethernet Switch Chip**



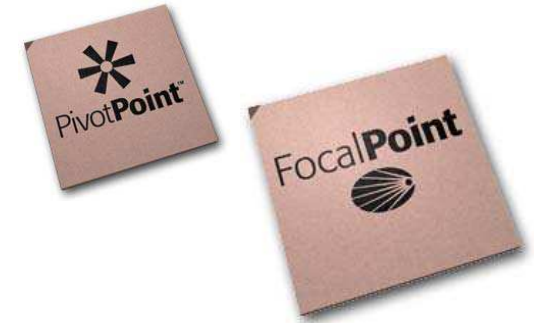
# Company Overview



**Fabless Semiconductor  
Company (50+ people)**



**Formed out of Caltech  
(1/00)**



**Shipping two low-latency  
product families today**

**NEA**  
NEW ENTERPRISE ASSOCIATES



**WORLDVIEW**  
TECHNOLOGY PARTNERS

**Backed by top-tier investors**



# FocalPoint: an Ethernet Switch Chip

*The world's most powerful Ethernet switch chip*



- **Highest port density** (24 10GE ports)
- **Lowest latency** (200ns)
- **Highest performance** (240Gbps)
- **Most power efficient** (<150mW/Gbps)
- **Most integrated** (single chip)
- **Most scalable** (fat trees, 1,000s of ports)



**FocalPoint Evaluation Platform**  
(The world's most integrated 10G Ethernet system)

# Agenda

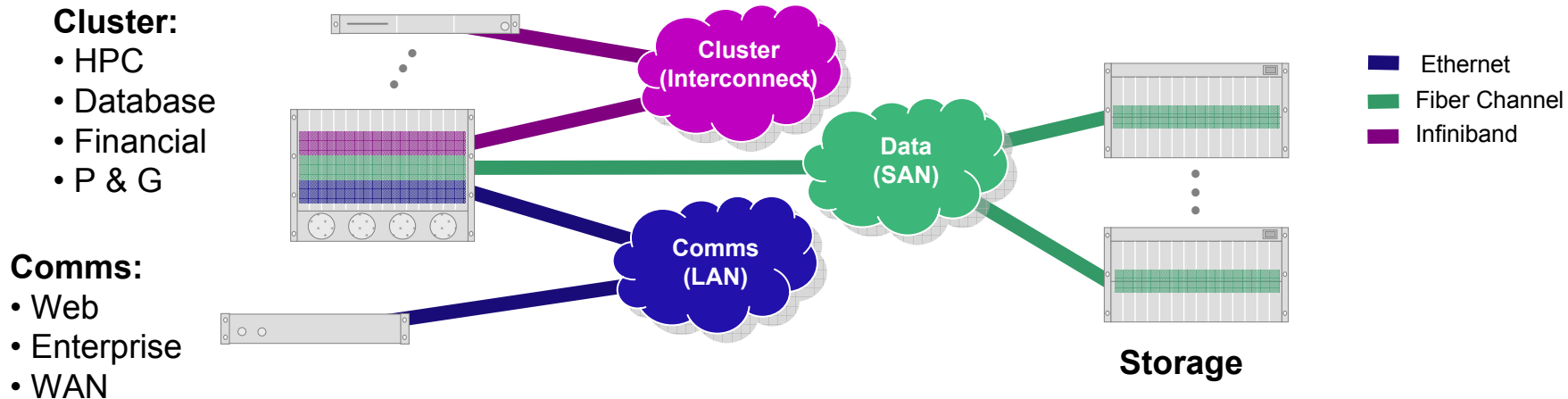
## **Datacenter Interconnect Requirements**

FocalPoint Chip

FocalPoint in Datacenter Applications

# Problem: Disjointed Datacenter Inhibits Scale

*Multiple interconnects create islands of specialization*



- **Network technologies in today's data center:**
  - Cluster: Optimized for low latency (Infiniband)
  - Data: Low latency, robust delivery (Fibre Channel)
  - Comms: Secure, flexible, cheap, interoperable (Ethernet)
- **Ethernet is the industry's preferred choice**
  - Poor latency characteristics led to specialized solutions

# Enabling Low-Latency Fabrics

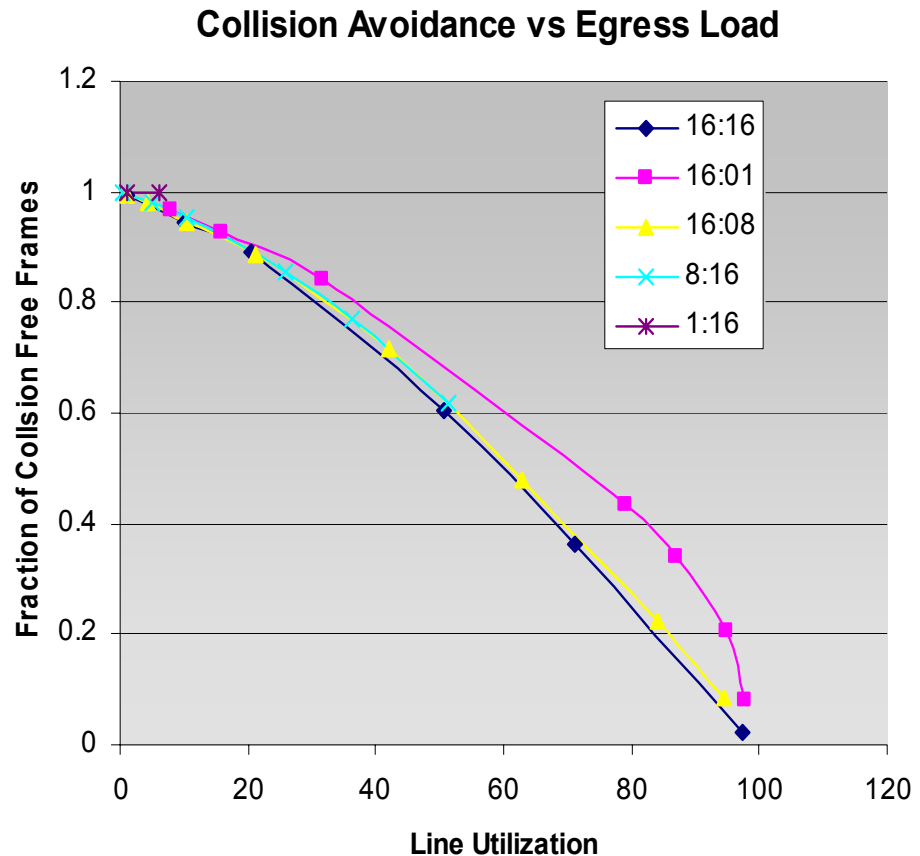
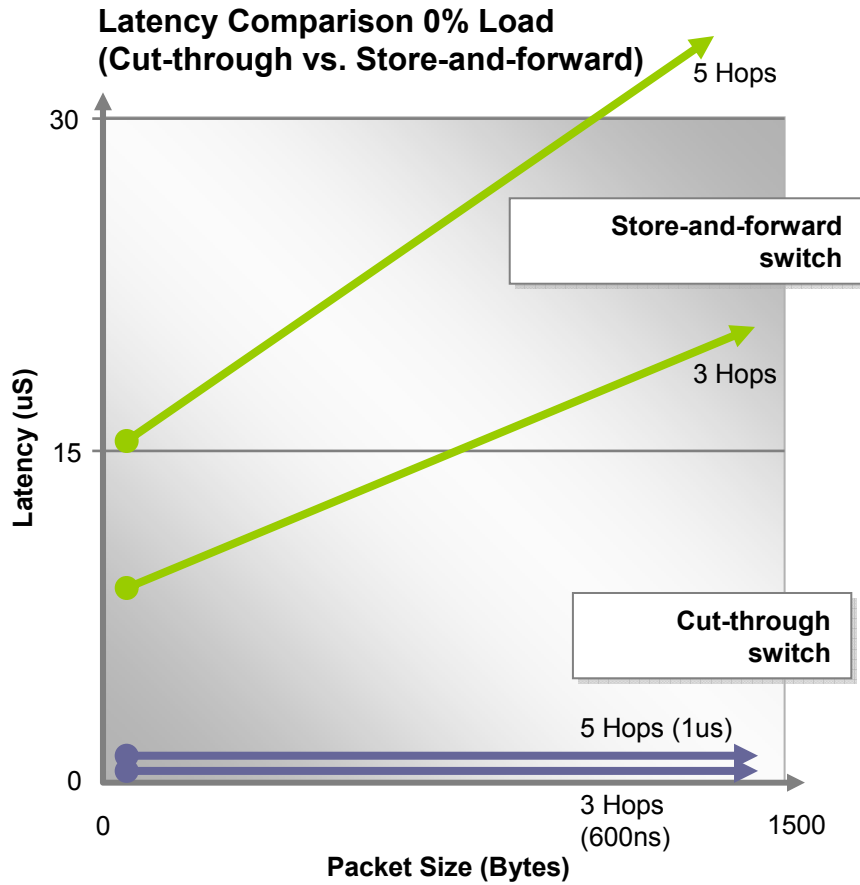
*Solutions balance additive contributors to latency*

## Three contributors to switch latency:

- 1. Store-and-forward latency (last bit in to first bit out)**
  - Typical vendor: 3 $\mu$ S per 10GE switch hop
  - FocalPoint: 150nS per 10GE switch hop
- 2. Packet serialization time**
  - Typical: 0.8nS/byte at 10GE and 8nS/byte at 1GE
  - FocalPoint cut-through: 50nS (packet independent)
- 3. Scheduling latency**
  - Effects store-and-forward and cut-through equally
  - Linearly dependent on egress port load
  - Solution: add more ports  
(FocalPoint has 24, others have 20 or less)

# Latency and Performance Under Load

*Functional Ethernet never much more than half loaded*



# Performance Comparison

*Switch latency should be 10-20% of system latency*

## Comparison Assumptions

- **System**
  - 16 servers per rack switch
  - 20P and 24P switches
  - 4 or 8 uplinks
  - 25G total uplink BW
  - 3 and 5 hop networks
- **Per-hop collision free**
  - 33% 16:4 configuration
  - 67% 16:8 configuration
- **Store-n-Forward Latency**
  - 3  $\mu$ S – standard vendor
  - 150 nS - FocalPoint
- **Traffic Profile**
  - 40% 64B
  - 40% 1500B
  - 20% Even (64B,1500B)

## 3 Hops

Frame Size	FP-CT Unloaded	FP-SF Unloaded	V-SF Unloaded	FP-CT Loaded	V-SF Loaded
Byte	( $\mu$ S)	( $\mu$ S)	( $\mu$ S)	( $\mu$ S)	( $\mu$ S)
64	0.6	0.6	9.2	1.5	15.6
512	0.6	1.7	10.2	1.5	16.7
1500	0.6	4.1	12.6	1.5	19.0
10000	0.6	24.5	33.0	1.5	39.4

## 5 Hops

Frame Size	FP-CT Unloaded	FP-SF Unloaded	V-SF Unloaded	FP-CT Loaded	V-SF Loaded
64	1.0	1.0	15.3	2.6	26.0
512	1.0	2.8	17.0	2.6	27.8
1500	1.0	6.8	21.0	2.6	31.7
10000	1.0	40.8	55.0	2.6	65.7

FP-CT: FocalPoint in cut-through mode

FP-SF: FocalPoint in store-and-forward mode

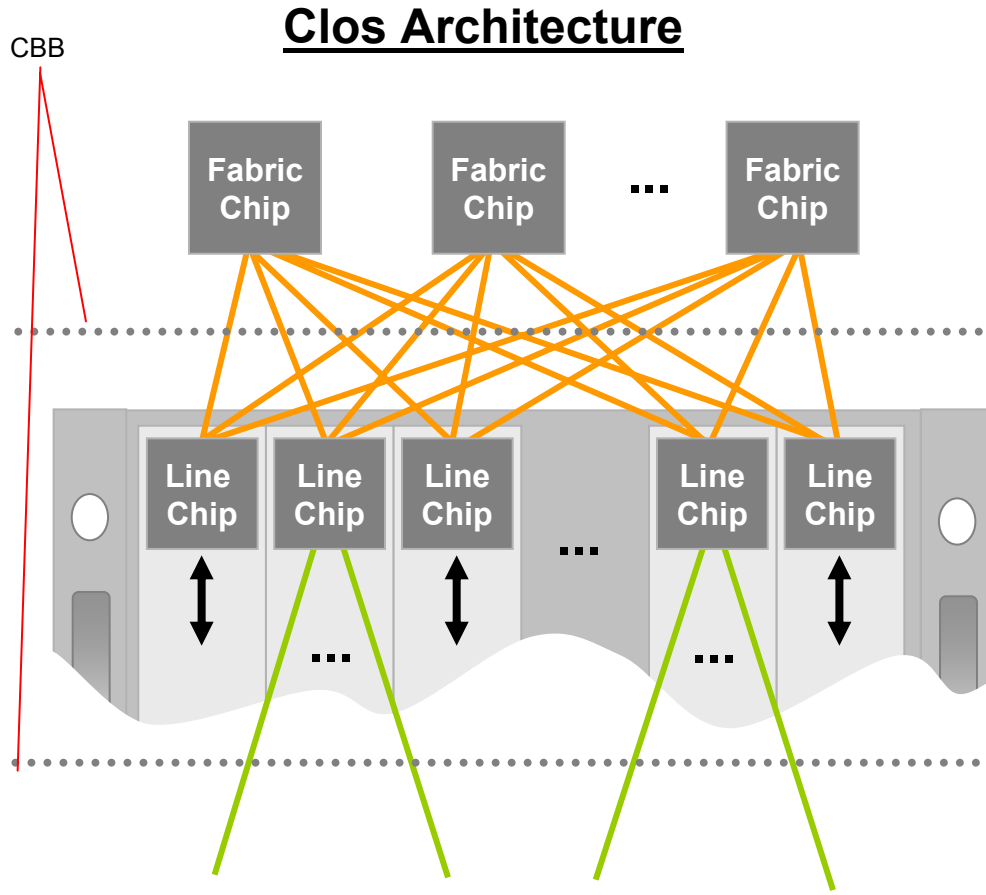
V-SF: Vendor (typical) 10GE product in store-and-forward mode

Unloaded: 0% load – a measure of fabric fall-through latency

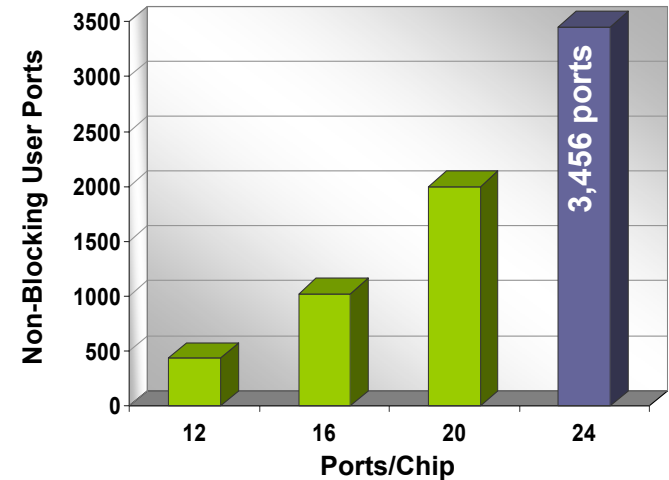
Loaded: 33% load for 8 uplinks, 66% load for 4 uplinks



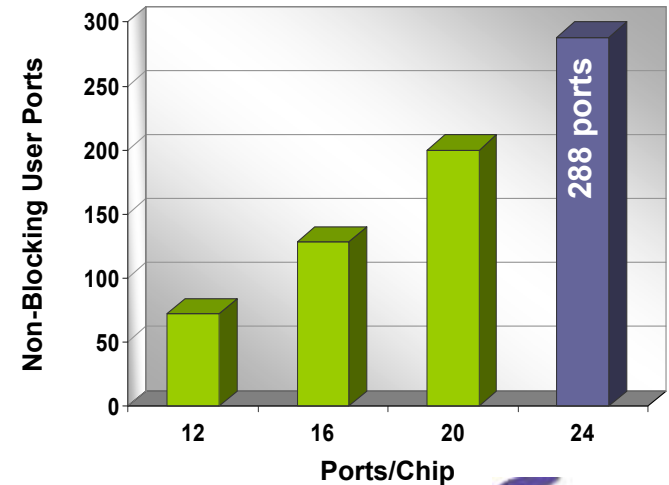
# Port Density Enables Cost Effective Scale



Three-Tier Fat Tree



Two-Tier Fat Tree



# Agenda

Datacenter Interconnect Requirements

**FocalPoint Chip**

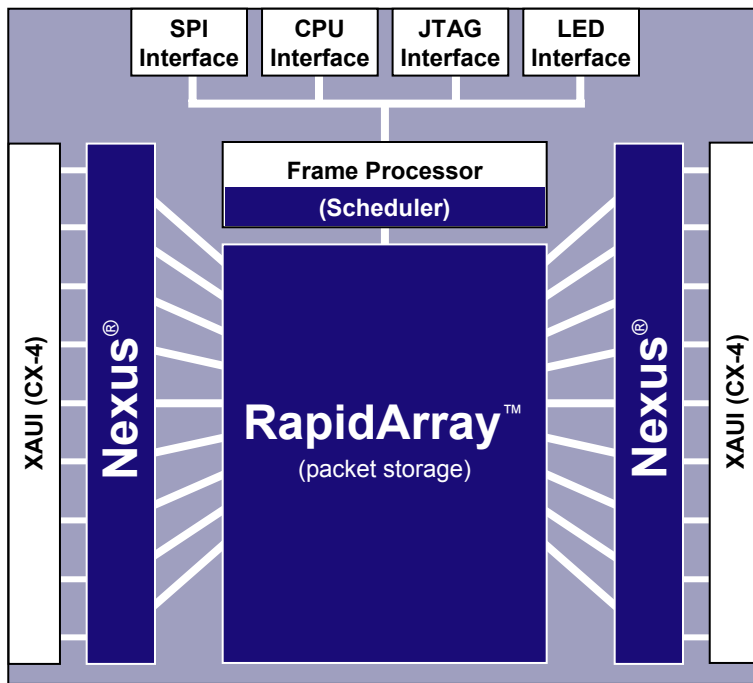
FocalPoint in Datacenter Applications




# FocalPoint Project Goals



*The only low-latency feature-rich 10GE switch*



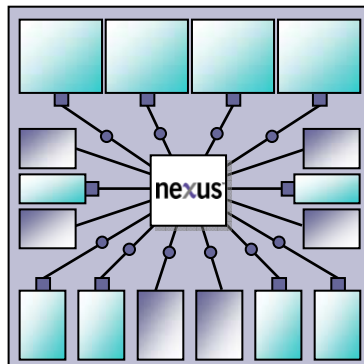
 Fulcrum proprietary IP

- **24 10G Ethernet ports**
- **200nS fall-through latency**
- **240Gbps shared memory fabric**
  - Fully non-blocking fabric
  - Full-rate multicast
- **Standards compliant, feature rich**
  - Good QoS and congestion mgmt
  - 16K MAC addresses
  - 4K VLAN and STP tables
- **Process**
  - TSMC 0.13 $\mu$ m FSG process
  - All standard flows
  - Fully outsourced GDS to customer ship
- **< 1W per port, typical**

# Architecture Enabling Circuits

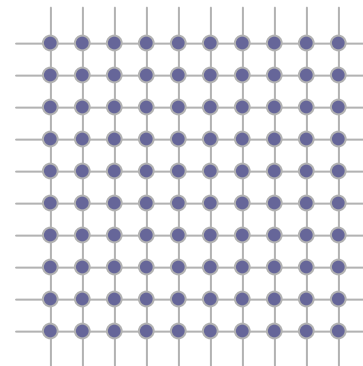
*Two key IP blocks differentiate the product*

**Nexus\***  
(Terabit Crossbar)



- Gigahertz performance
- Terabit capacity
- Nanosecond latency
- No power penalty

**RapidArray**  
(Packet Storage)

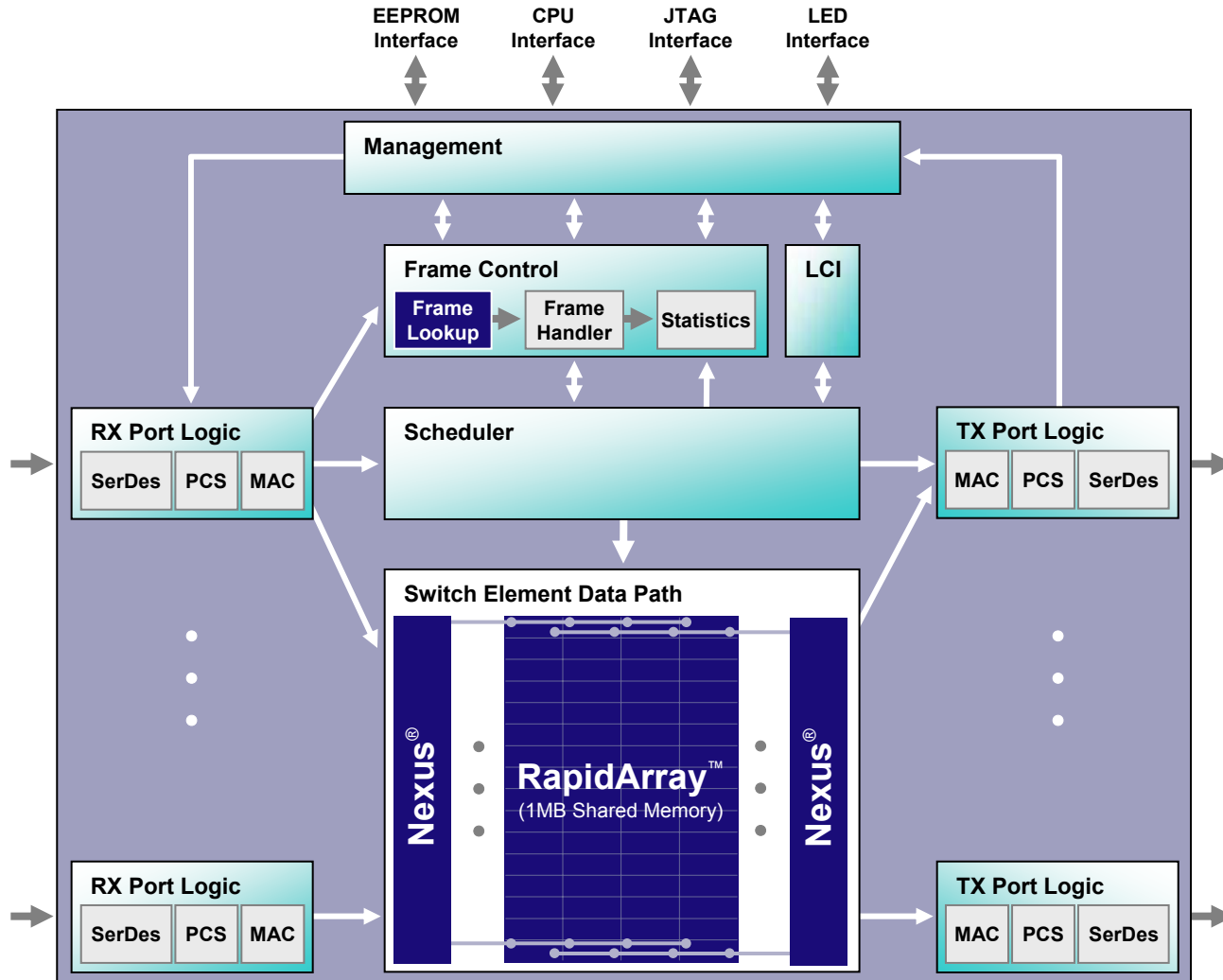


- 720 MHz SRAM
- 1200 MHz interconnect
- 76.8 GB/s throughput
- Scalable for larger designs

## Key Benefits:

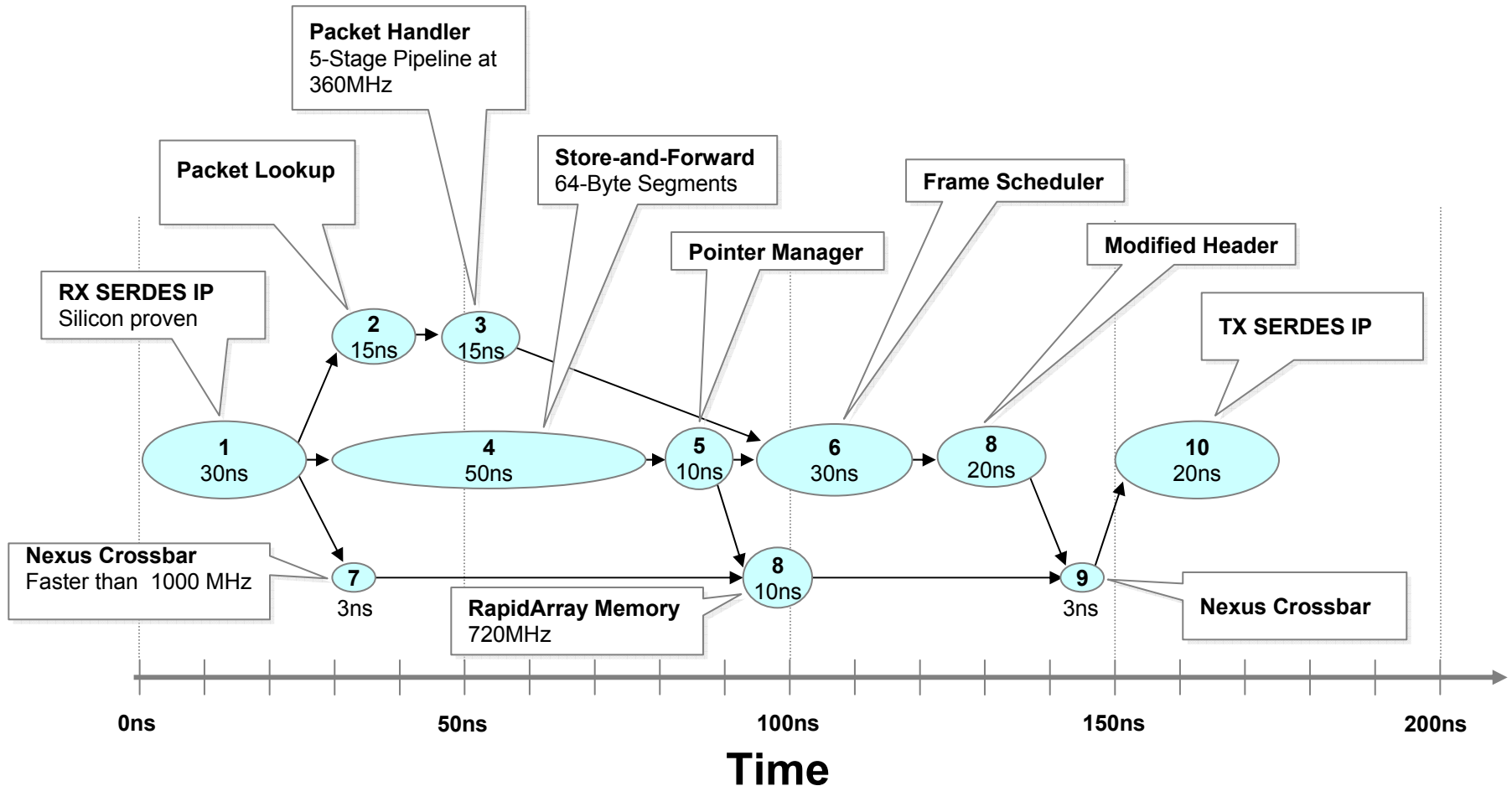
- **3 nS latency (including arbitration)**
- **Terabit(s) per square millimeter**
- **Usage based power consumption**
- **2x the speed of vendor cores (same size, density, yield)**
- **Small block optimized**

# FocalPoint Hardware Architecture



# FocalPoint Latency Detail

*Ball-to-Ball Latency is less than 200ns*



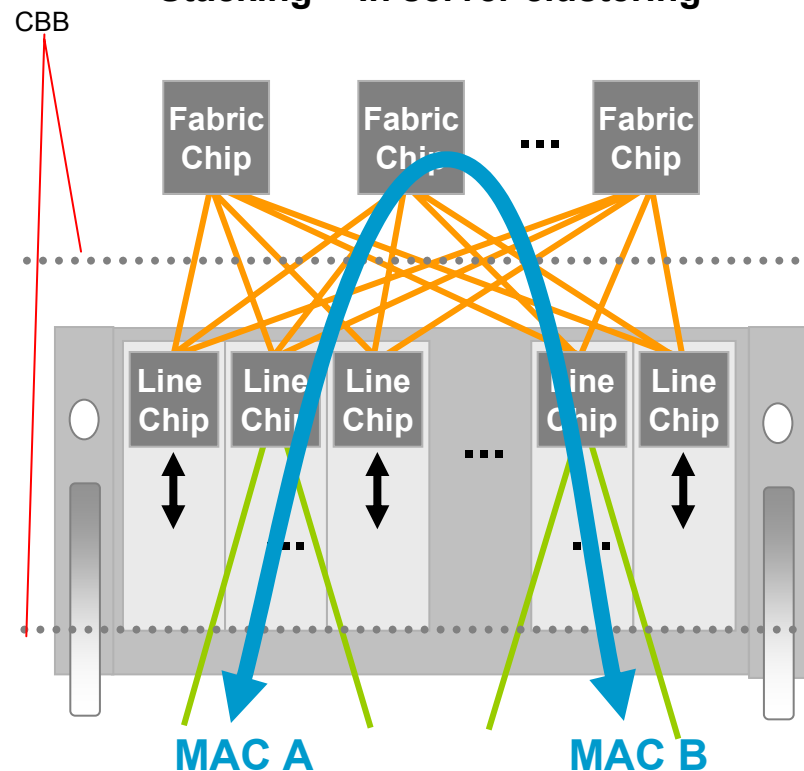
# Bridge Features in the Data Center

## Complete Ethernet Feature Set

- **Bridge Features**
  - 16k MAC address entries
  - All spanning tree variants
  - Learning and aging controls
- **VLANs (IEEE 802.1Q)**
  - 4k VLAN entries
  - Double tagging (Q-in-Q)
  - Port-based flood groups
  - 4k Spanning Trees (IVL)
- **QOS**
  - Per port and shared memory watermarks
  - 802.1p – 8 priorities per port
  - Pause & packet discard
  - 100 Queues
  - Transmission selection
- **Link Aggregation**
- **Security**
  - 802.1x & MAC Address Security
- **Layer 2 classification engine**
  - Drop, Mirror, change priority
- **Statistics**
  - >1,000 64 bit counters

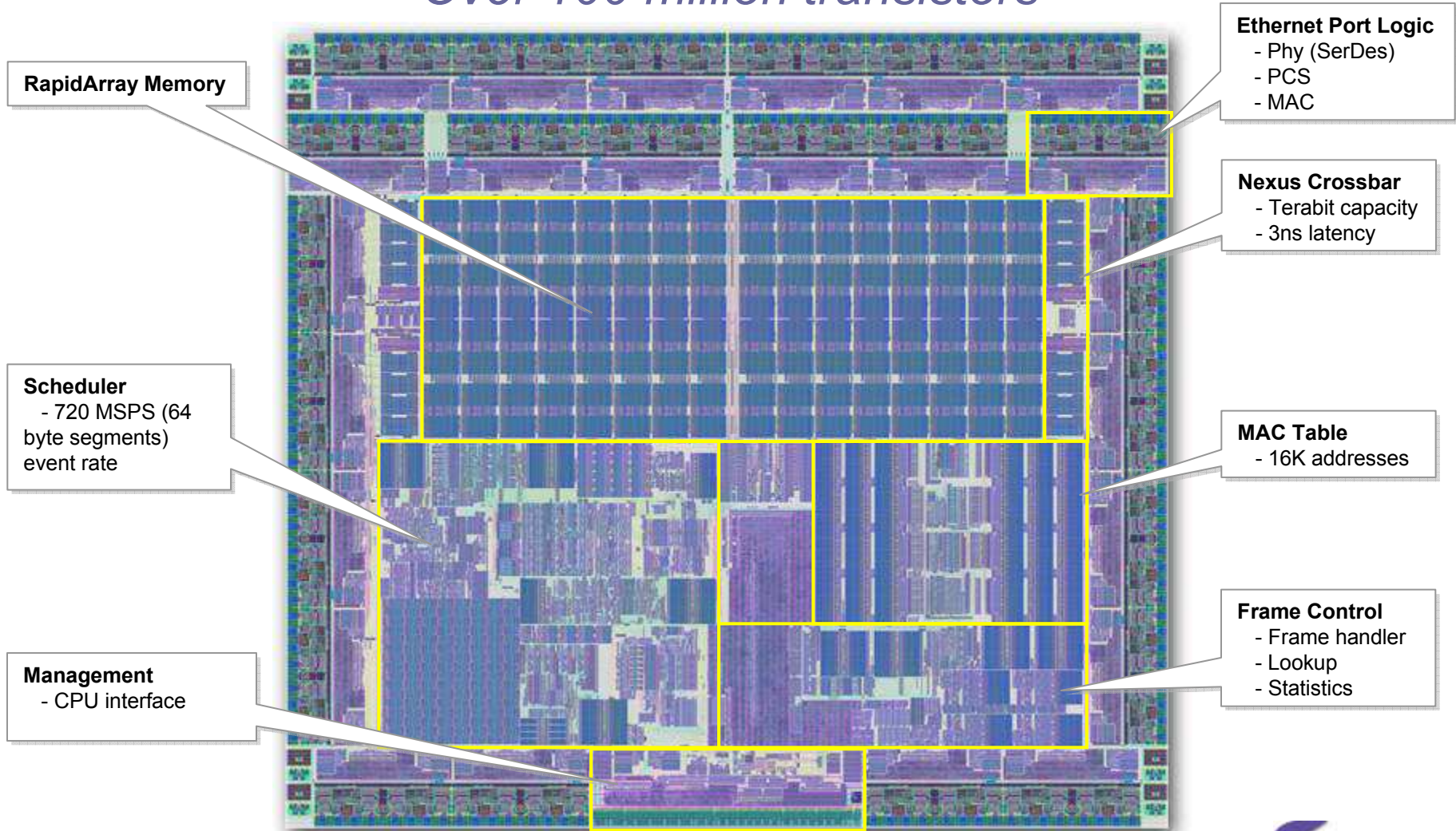
## Clustering enhancements

- Flexible link agg -- 12 port trunks
- Fat tree support -- HW learning, aging
- Stacking -- In server clustering



# FocalPoint Chip Plot

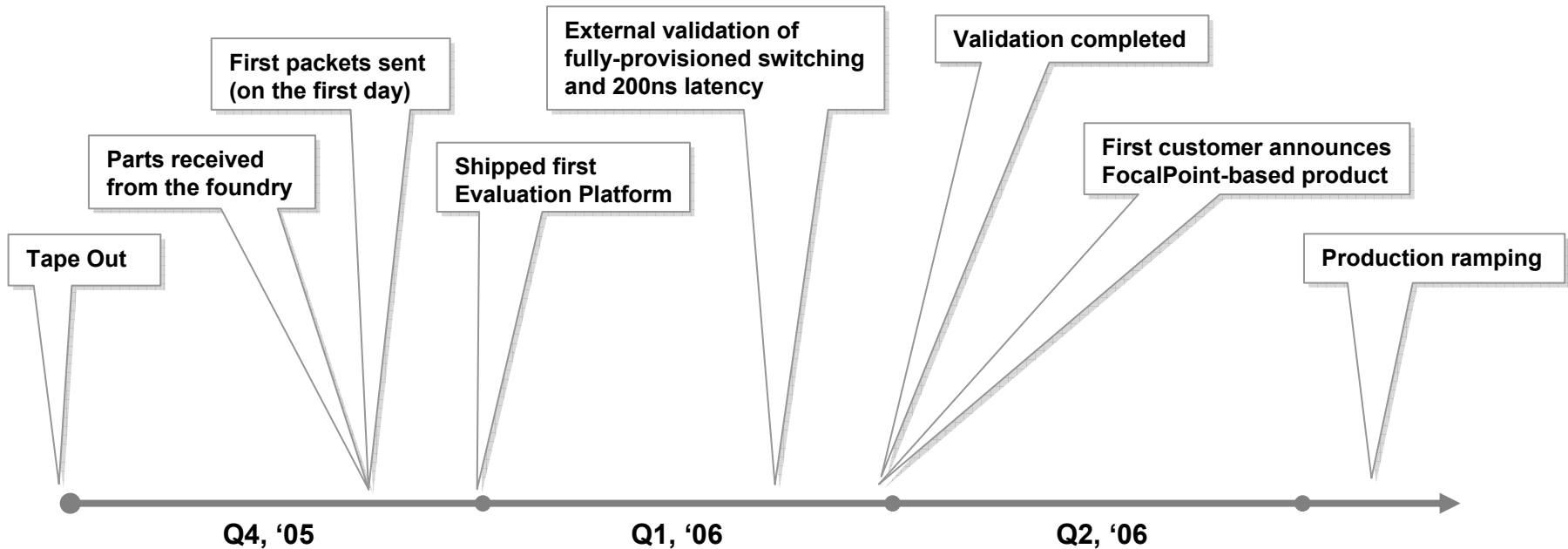
*Over 100 million transistors*





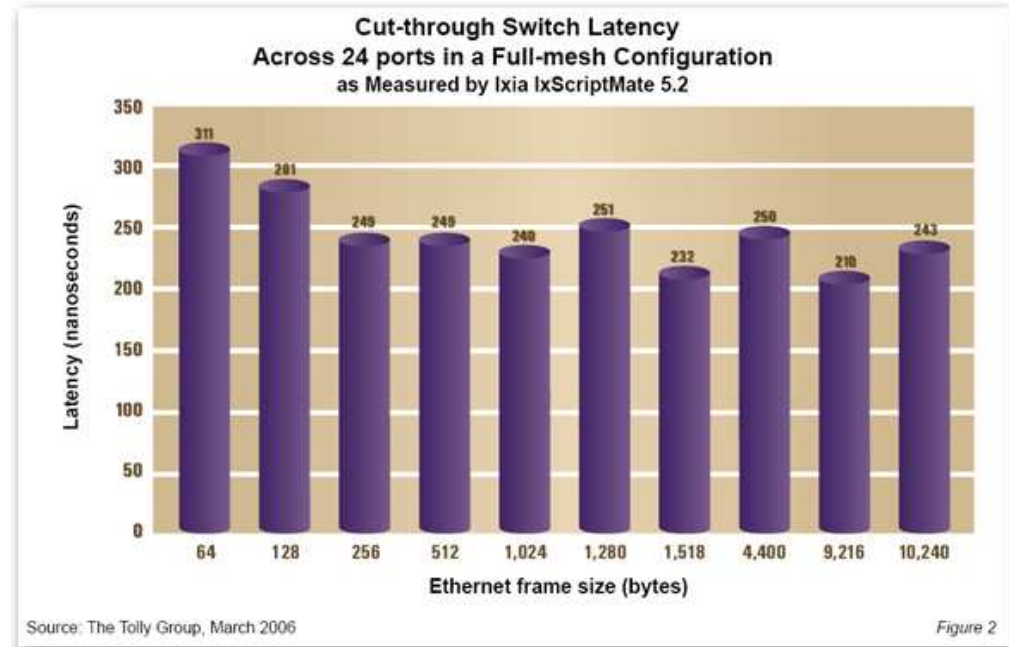
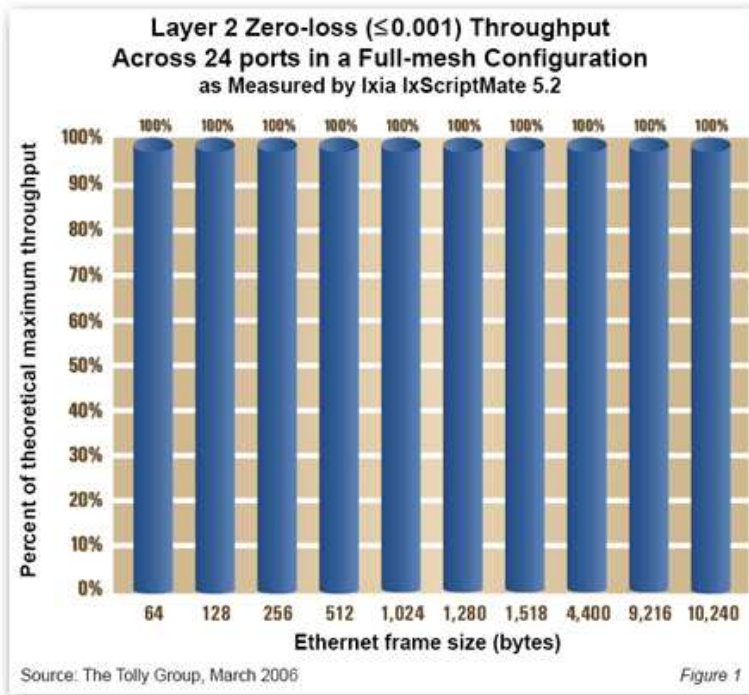
# FocalPoint Status Report

*FocalPoint is in production*



# Recent External Validation

*Industry-leading latency and performance, as expected*



# Agenda

Datacenter Interconnect Requirements

FocalPoint Chip Architecture

**FocalPoint in Datacenter Applications**



# Validated End-to-End Latency

*Latency comparable to specialty fabrics*



Lowest Ethernet latency – ever!

*2.4μs, application-to-application (MPI)*

	MX/Myrinet	MX/Ethernet	OpenIB/InfiniBand
Switch Vendor	Myricom	<b>Fulcrum</b>	Mellanox
Ping Pong Latency	2.4μs	<b>2.4μs</b>	4.0μs
Two-way data rate	2,397 MB/s	<b>2,162 MB/s</b>	1,902 MB/s



Lowest full iWARP latency – ever!

*<10μs, application-to-application (MPI)*



Lowest 1G Ethernet latency – ever!

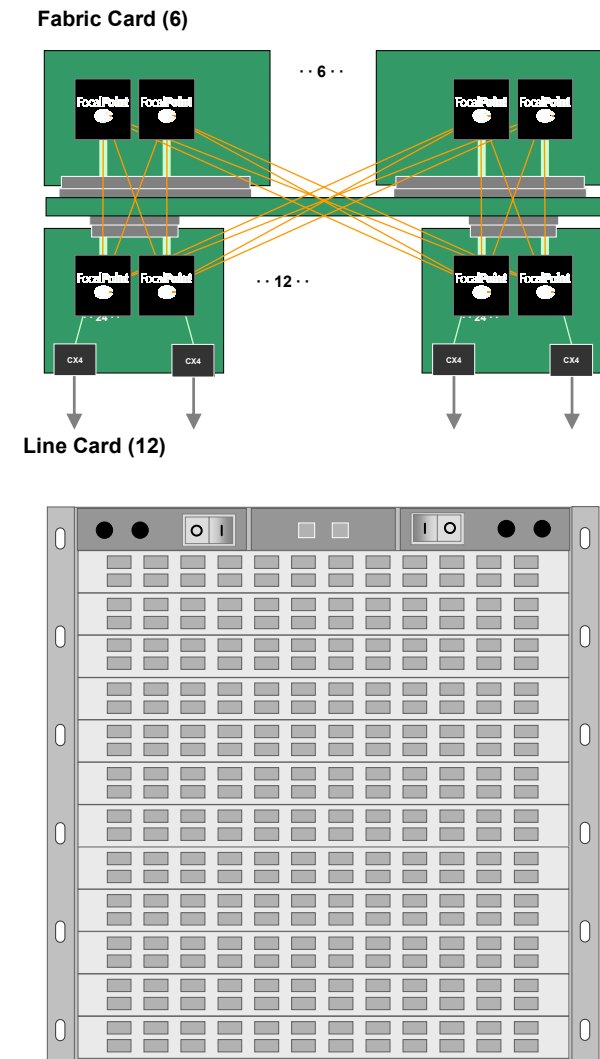
*<10μs to the application*

More headlines coming soon...



# Data Center Switch (Two-Tier Fat Tree)

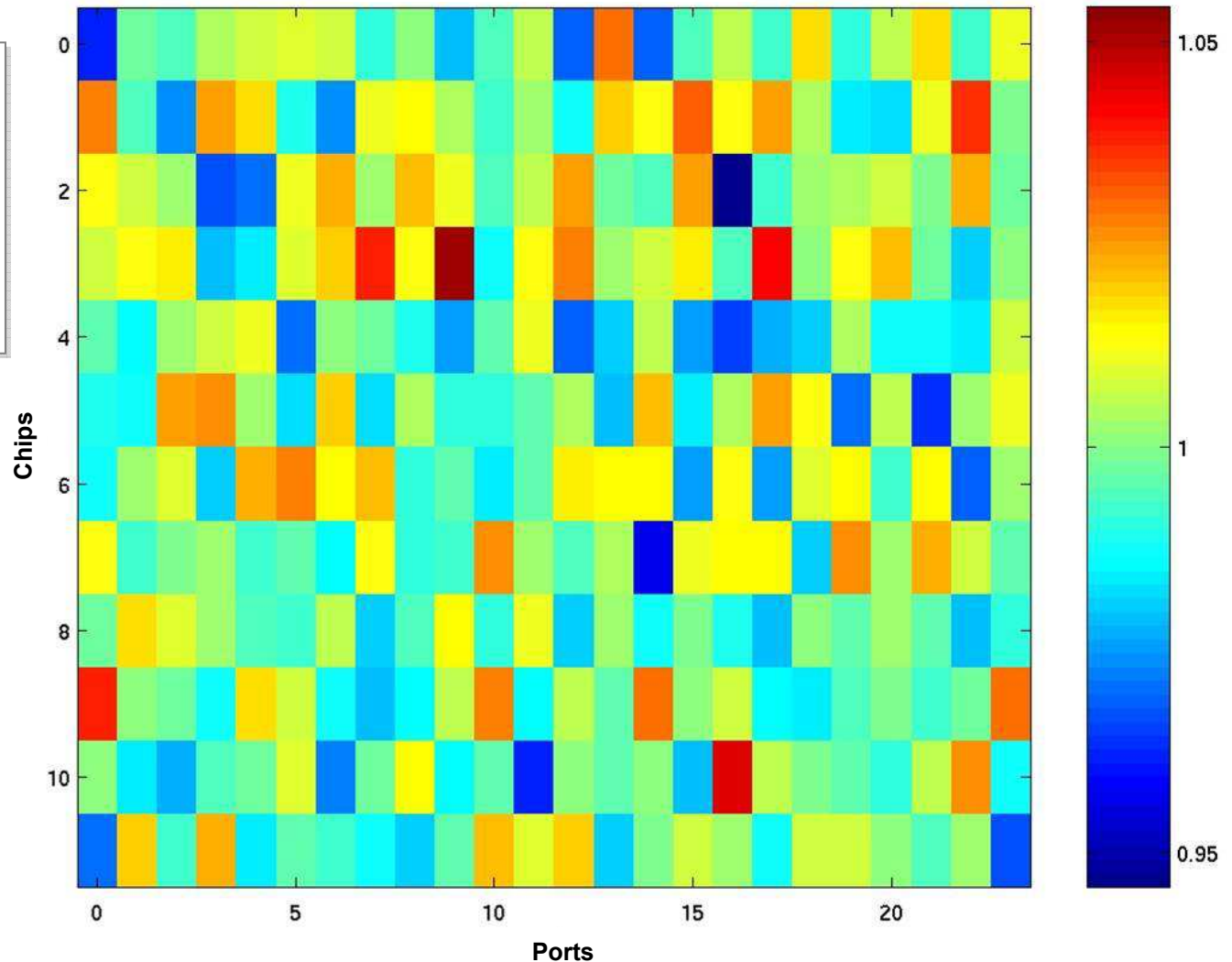
- **Features**
  - 288 10GE ports
  - CX-4 and XFP line cards
  - Non-blocking architecture
  - 0.6 $\mu$ S port-to-port latency
  - 192,000 MAC addresses (effective)
  - Single-switch software image
  - 100% multicast bandwidth
  - Rich Ethernet L2 feature set
- **Composition**
  - 24 ports per blade
  - 36 chips per chassis
- **Extremely cost effective**
- **Significant industry interest**



# Hash Efficiency (288-Port Switch)

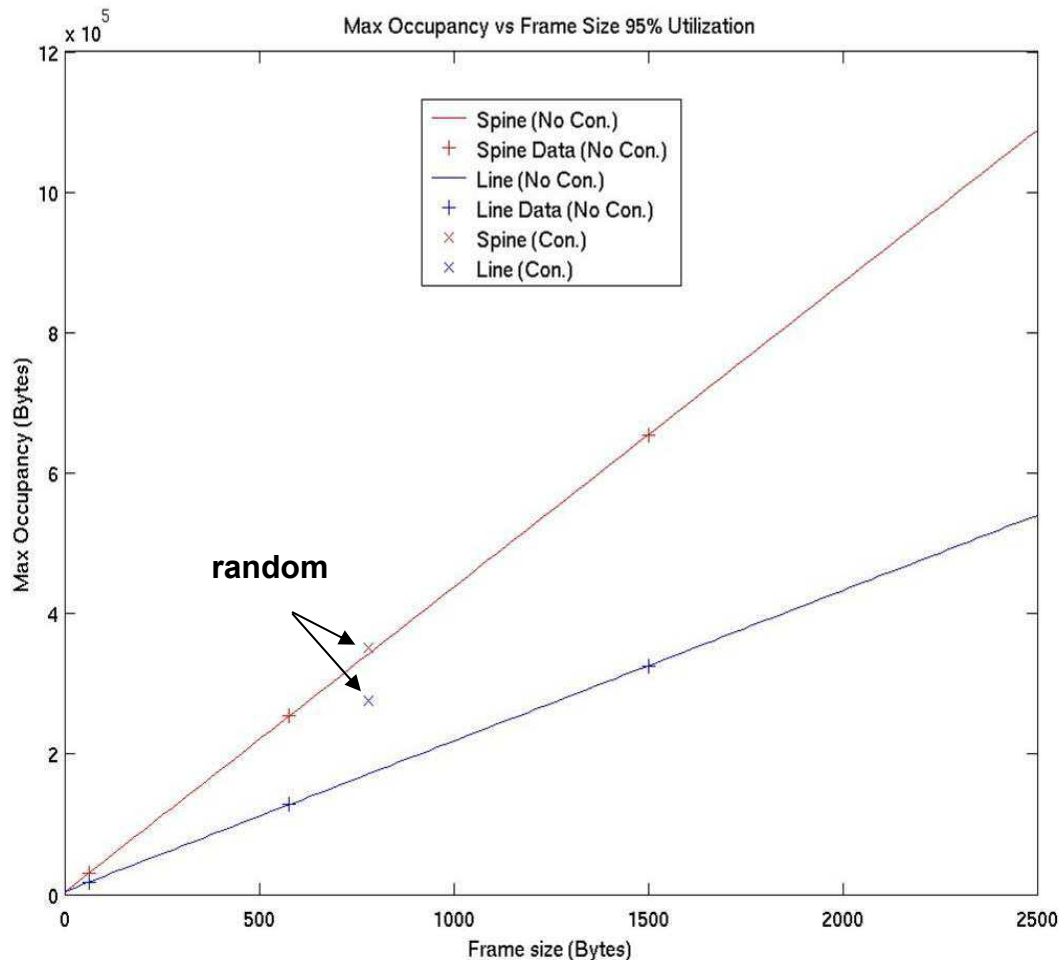
Spine Chip Load Balancing

- SA-DA hash for 8k MAC addresses
- Mesh round robin
- Each pixel is a port for 12 spine chips
- +/- 5% asymmetry
- Load independent



# Memory Utilization for Multiple Profiles

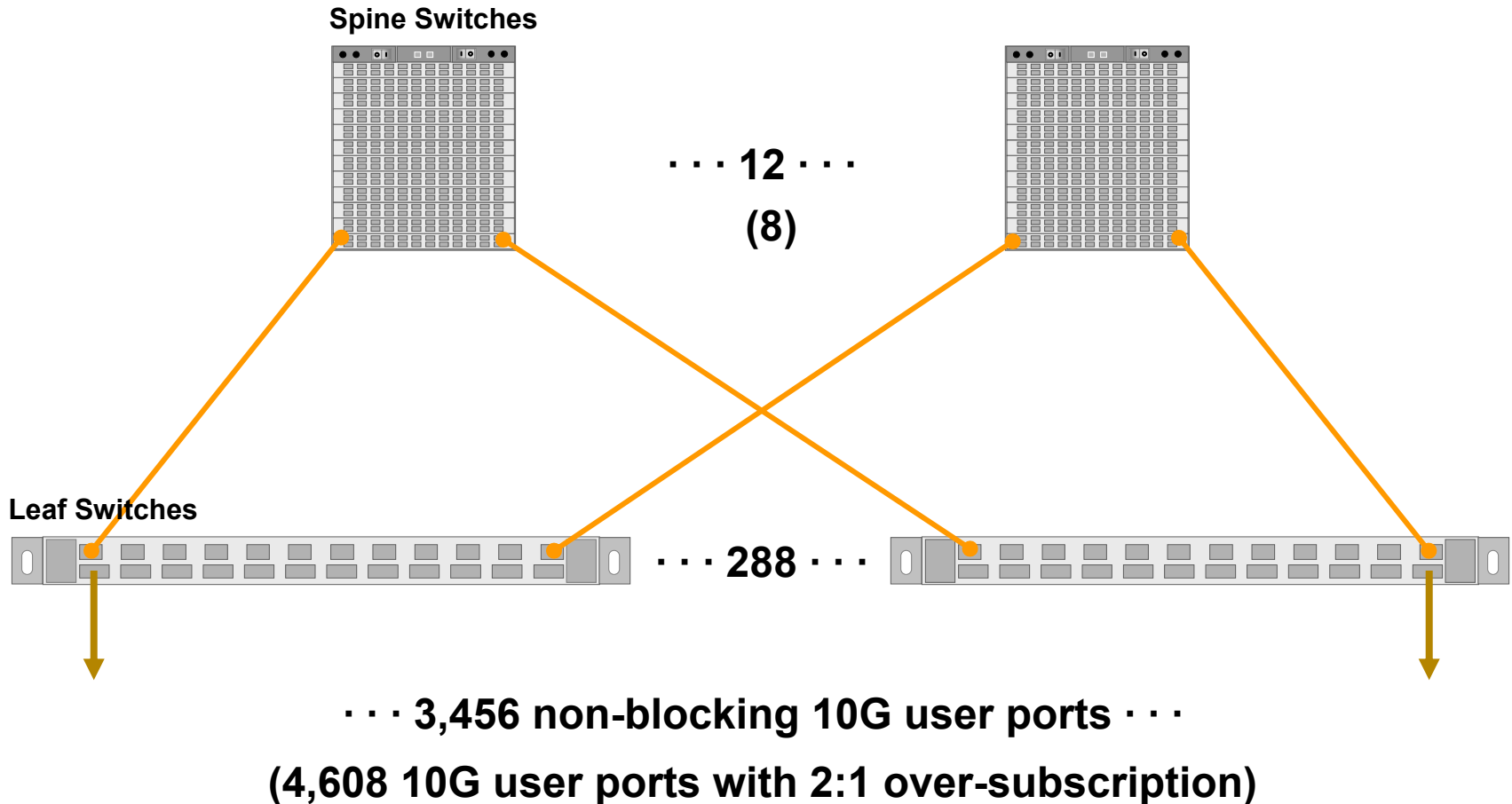
*Maximum memory of 36 chips*



- 64B, 576B, 1500B
  - Random
- Random profile
  - 40% 64B
  - 40% 1,500B
  - 20% flat distribution
- Even at 95% load, no drops of 1,500B frames

# Three-Tier Fat Tree Architecture

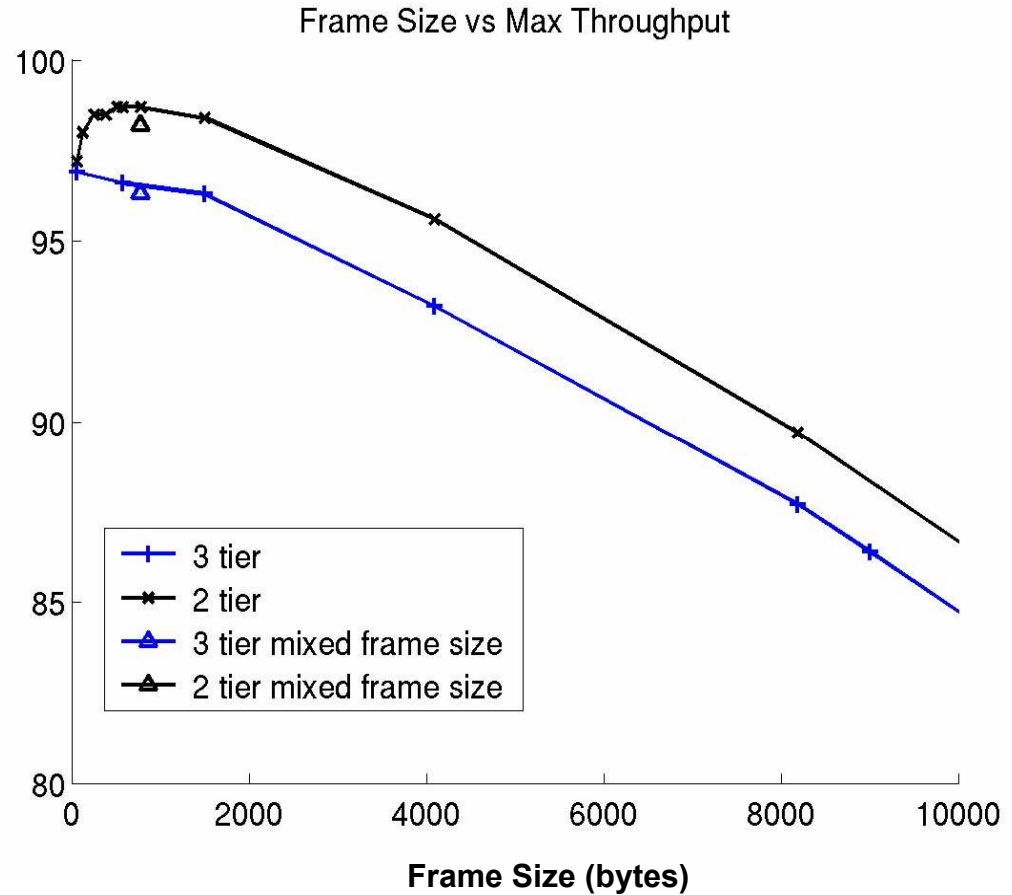
*~1 $\mu$ s latency from any port to any other port*





# Available Bandwidth in Multi-Tier Fat Trees

- 3-tier system performs within 2% of 2-tier system
- Larger frames → fewer hashes
  - Exposes hash inefficiencies



# Thank You!

**Uri Cummings**

*Founder, CTO*

uri@fulcrummicro.com



818.871.8100  
www.fulcrummicro.com

26630 Agoura Road  
Calabasas, CA 91302

"Fulcrum is betting that by eliminating the latency issues with Ethernet switching, the vast ecosystem that surrounds Ethernet will drive much-needed consolidation."

*Simon Stanley, Research analyst for Light Reading's Comm Chip Insider*