Computer Systems, Inc.
MERCURY
Challenges Drive Innovation™

Using Layer 2 Ethernet For High-Throughput, Real-Time Applications

High-Performance Embedded Computing (HPEC) Conference – September 24, 2008

Robert Blau and Dr. Ian Dunn
Mercury Computer Systems, Inc.

Packets Over Ethernet (POET) Topics

- Ethernet Trends
- POET Overview
- POET Rationale
- POET Results
- Conclusion

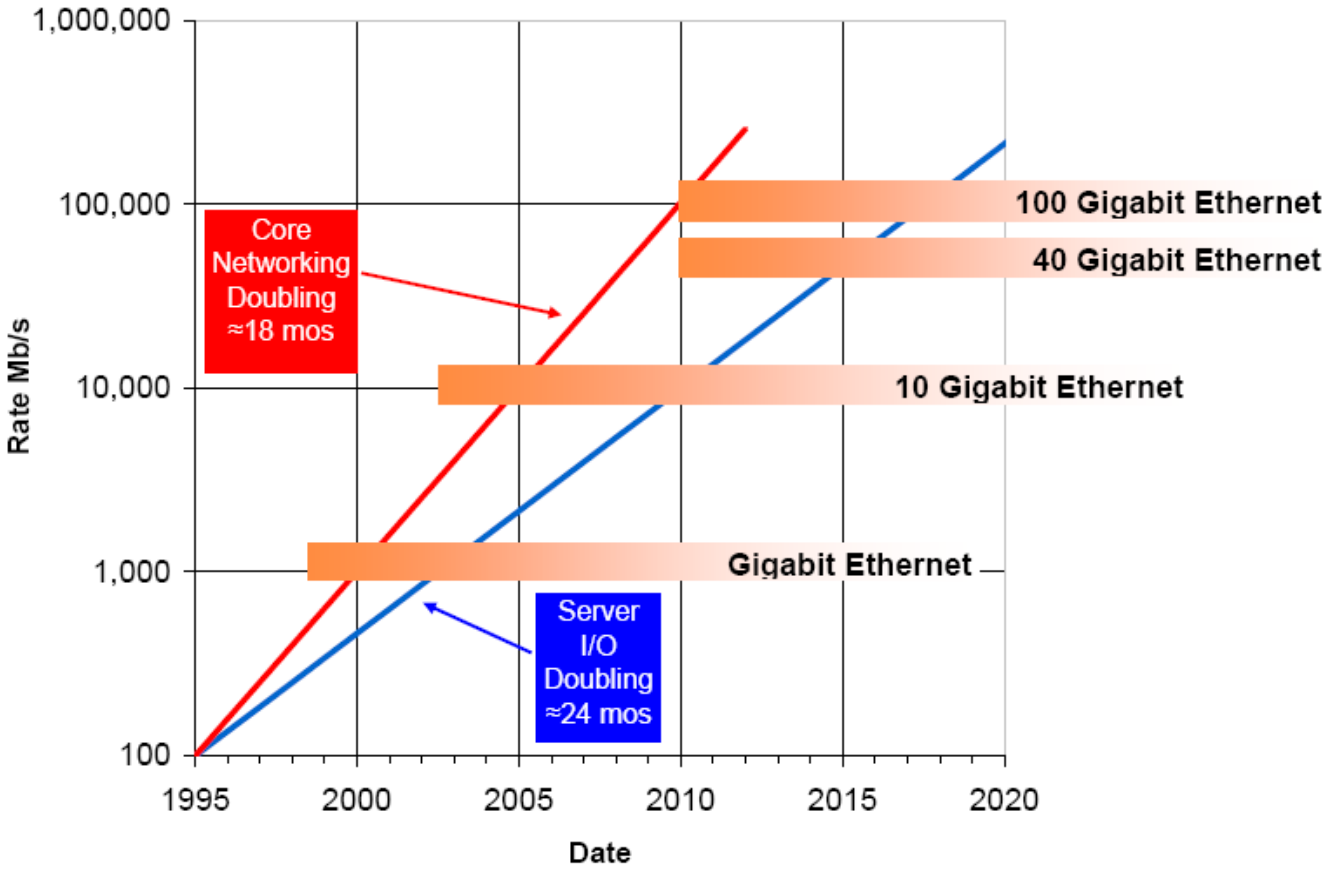
Problem/Opportunity Statement

Never bet against Ethernet	**Ethernet's ability to prevail over seemingly superior technologies has made this into a networking axiom** • Seamless interoperation at 1, 10, 40, & 100 Gbps standards • Location-agnostic operation – applications can reside wherever optimal, depending on performance, productivity, and life-cycle considerations • Low cost of ownership
Ethernet is the catalyst behind the growing pervasiveness of compute clusters	**Corporations/individuals looking for highest performance, most productive, and cost-efficient compute platforms to help translate research into value, deployed capabilities** • Spawning new repositories of innovative software-based capabilities that rely on Ethernet as the communications substrate • Clustering/networking will become dominant components of any good system engineering toolkit
Ethernet is the antithesis of real time	• Unbounded latency, power consumption • Industry-wide reliance on inefficient protocols in both software and hardware from a real-time perspective

No Longer an Inferior Technology

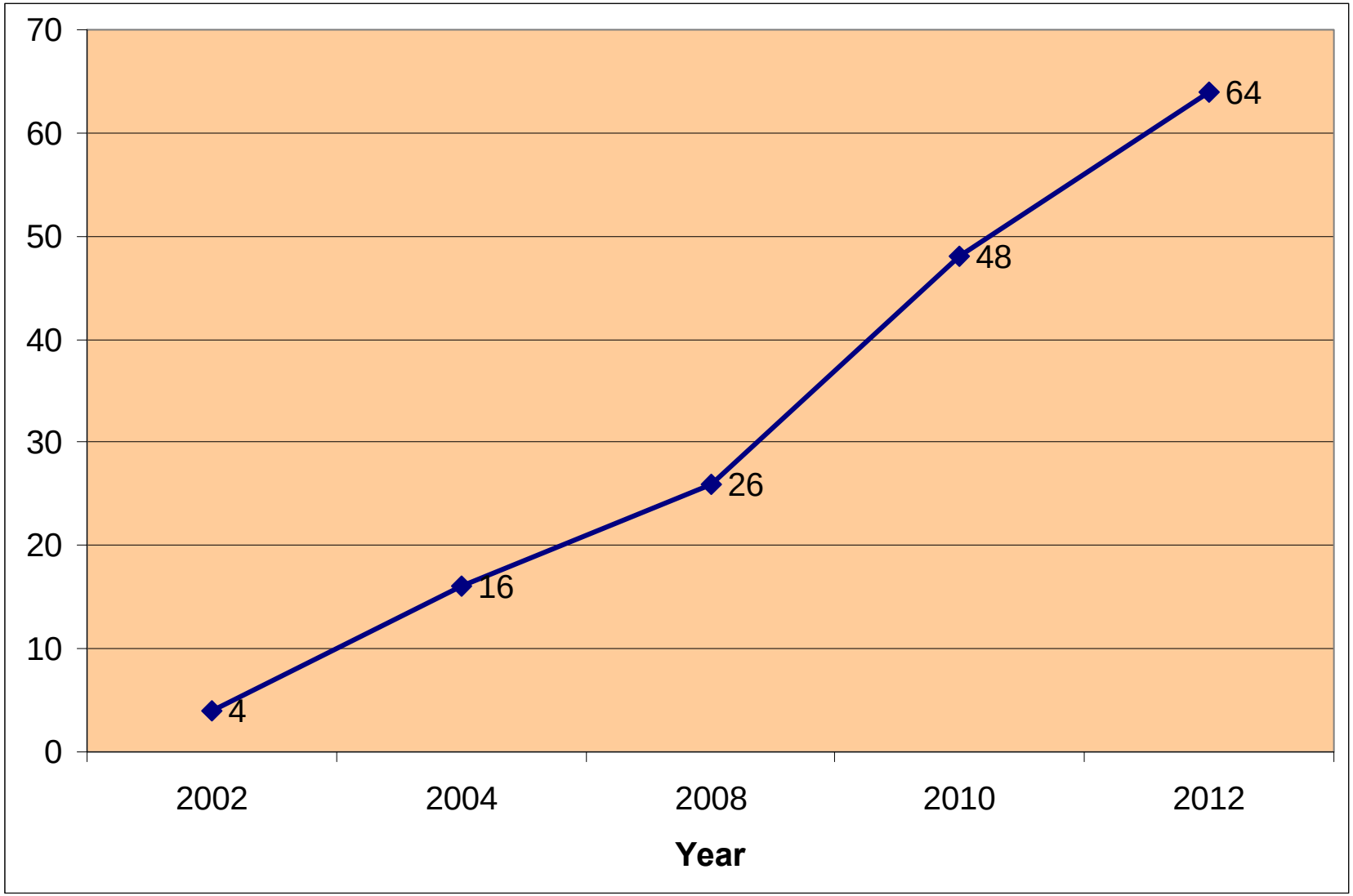
Catching up rapidly to embedded fabrics

40GbE and 100GbE: Computing and Networking



IEEE 802.3 Higher Speed Study Group - TUTORIAL

Impressive 10GbE Switch ASIC Port Counts



POET

- **Layer 2 Solutions Make Sense**
  - Ride the Ethernet technology curve
  - ATAoE, FCoE, IBoE
- **POET is Architected with Real-Time Requirements**
- **Simple Principles:**
  - Encapsulate an existing packet protocol using L2 frames
  - Add H/W-based guaranteed delivery
  - Add H/W-based traffic policing and shaping

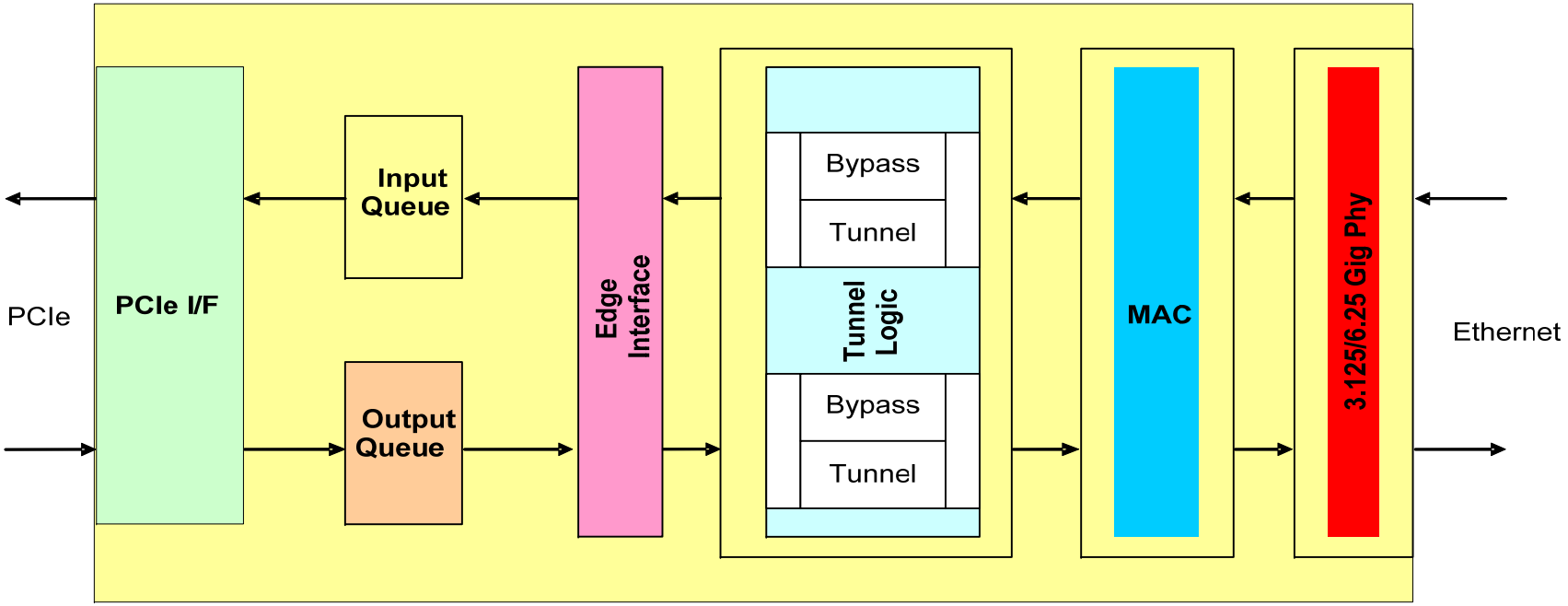
    Priorities

    Flow control

    Adaptive routing

    Channel bonding

POET Internals



TCP / IP Ethernet Format

0	Ethernet Preamble (PRE)			
4	Ethernet Preamble (PRE)			Ethernet Start Delimiter (SFD)
8	Ethernet Destination MAC Address (DA)			
12	Ethernet Destination MAC Address (DA)	Ethernet Source MAC Address (SA)		
16	Ethernet Source MAC Address (SA)			
20	Ethernet Type –VLAN = 0x8100	PRI	0	VLAN ID (VID)
24	Ethernet Type (IP)	IP Header		
28	IP Header			
32	IP Header			
36	IP Header			
40	IP Header			
44	IP Header	TCP Header		
48	TCP Header			
52	TCP Header			
56	TCP Header			
60	TCP Header			
64	TCP Header	Data		
...	Data			
N	Ethernet Frame CRC			

POET Format

Example of grouped packets encapsulated in the layer 2 Ethernet format – a small write packet and a read request in the same number of bytes as a TCP/IP header

0	Ethernet Preamble (PRE)				
4	Ethernet Preamble (PRE)				Ethernet Start Delimiter (SFD)
8	Ethernet Destination MAC Address (DA)				
12	Ethernet Destination MAC Address (DA)		Ethernet Source MAC Address (SA)		
16	Ethernet Source MAC Address (SA)				
20	Ethernet Type –VLAN = 0x8100	PRI	0	VLAN ID (VID)	
24	Ethernet Type (POET)	Payload Type		Ver	SeqID
28	Flow Control	Header CRC			
32	Packet 0 Write Header				
36	Packet 0 Header				
40	Packet 0 Data				
44	Packet 0 Data				
48	Packet 0 Data				
52	Packet 0 CRC				
56	Packet 1 Read Header				
60	Packet 1 Header (No Data)				
64	Ethernet Frame CRC				

Why Use Layer 2 Ethernet for Transport?

- **High Performance**
  - **Layer 2 avoids TCP/IP overhead**
  - **Latency and throughput very competitive with embedded fabrics**
- **Complete Functionality**
  - **Ability to provide all necessary functions and services**
  - **Scalable to 64K+ nodes**
  - **Multicast capability using VLANs**
- **Robust**
  - **Lossless transport layered on top in hardware**
  - **No TCP/IP (S/W Stack) overhead**
  - **Hardware-based proactive traffic policing and shaping to deal with contention and priority**
- **Compatible with Regular Networking Protocols**

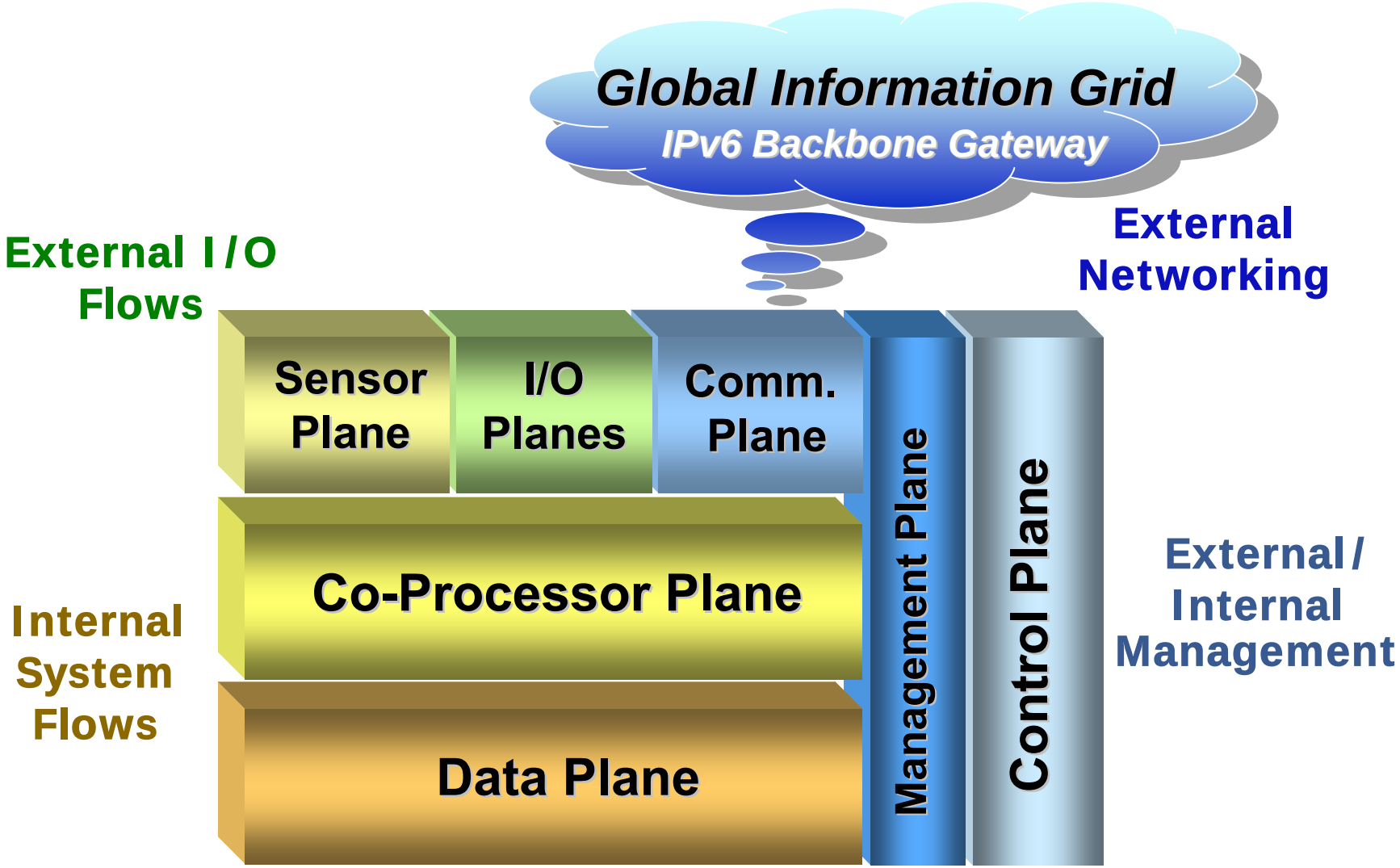
Key Requirements

- **Logical Layer Semantics**
  - **Memory ops: writes, reads, atomic ops**
  - **Messaging ops: mailbox, doorbell**
- **Performance**
  - **32+ Gbps point-to-point**
  - **Lightweight protocol**
  - **Channel bonding capability**
  - **Rapidly improving switch ASIC data rates**
- **Latency**
  - **1 us end-to-end**
  - **Minimal S/W involvement**
  - **Cut-through operation**
  - **Rapidly improving switch ASIC latencies**
    - **Fewer hops as # switch ports increase**
    - **Faster switch latencies**

Key Interconnect Features

- **Reliability**
  - Guaranteed delivery
  - Datagram (send and forget)
- **Contention Management**
  - Per-flow bandwidth control
  - Adaptive routing capability
- **Security**
  - DestinationID and address translation for memory region protection
  - Switch-based ACL, DOS features
- **Interoperability**
  - POET, TCP/IP, UDP, FCoE…

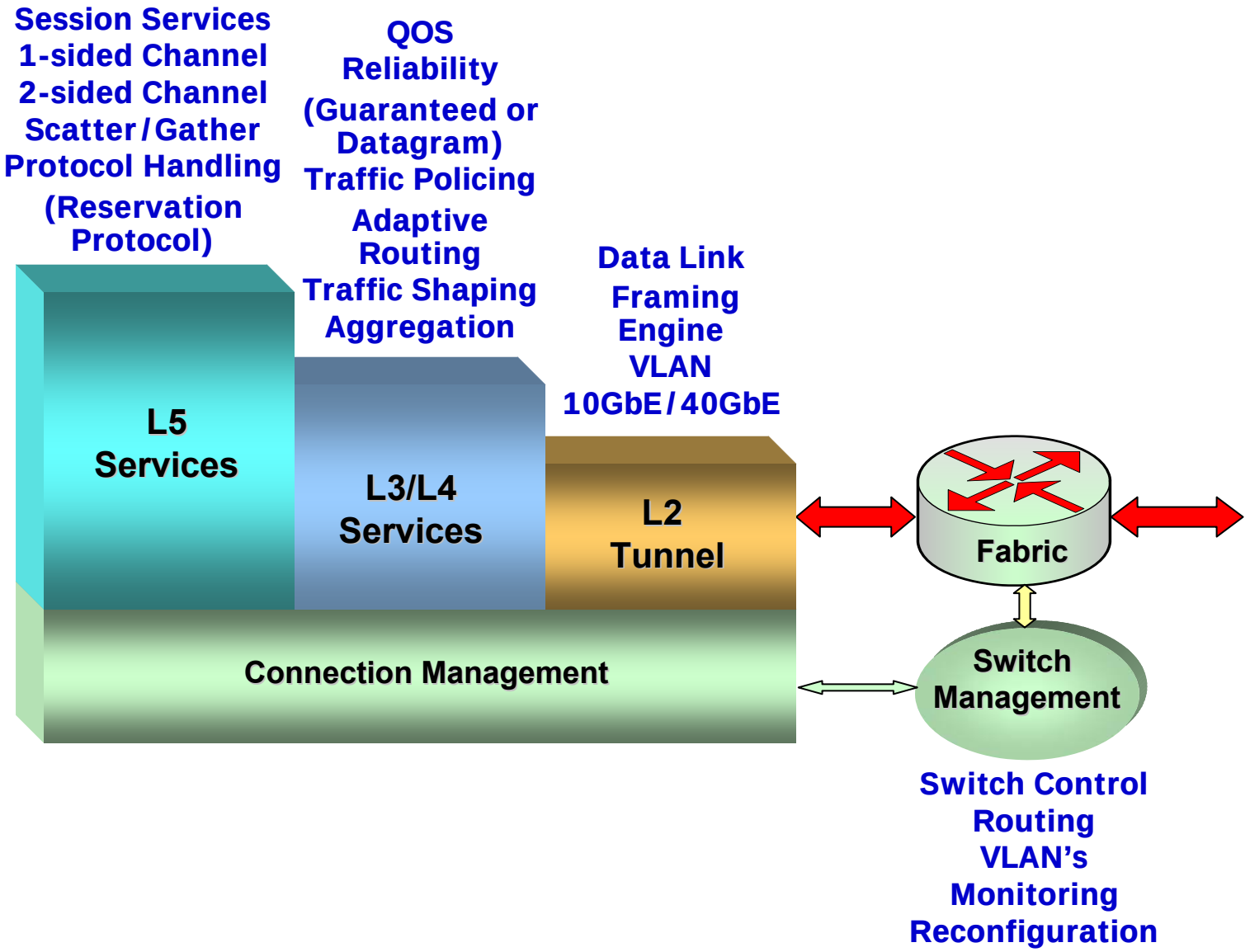
Mercury's Communications Model



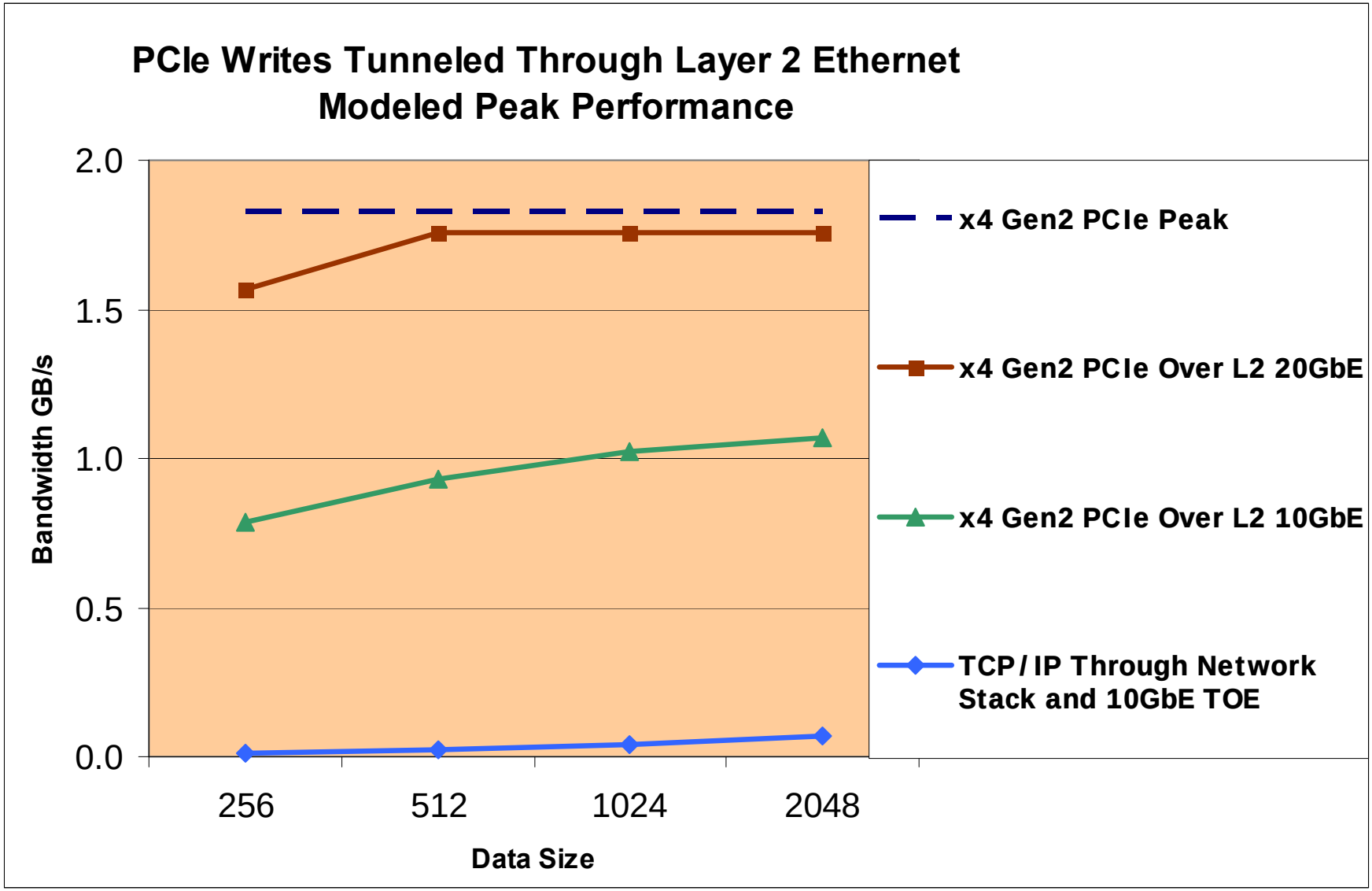
Nominal Prioritization Scheme

Priority	Traffic Class	Comments
3	**Sensor I/O, Management**	Acquires, conditions, and forwards resulting digitized data to data plane for segmentation, re-assembly, processing, and/or storage. Failure to service this communications plane properly adversely affects all downstream processing.
2	**Data, Co-Processing**	Transforms digitized data into information symbols and exploits those symbols to create information. As an adjunct to the data plane, encompasses data communication and processing used in the acceleration of data plane functions, as such it must have an elevated priority over data.
1	**Control**	Performs initialization, configuration, and synchronization of sensor processing components in data plane, including mapping and/or routing of data through sensor, data, and co-processing planes.
0 (low)	**Communications**	Provides external data communication, which includes, but is not limited to, network backhaul, wireless interfaces, and free-space optical transmission schemes. Control traffic associated with external interfaces is assumed to be part of control plane in this model.

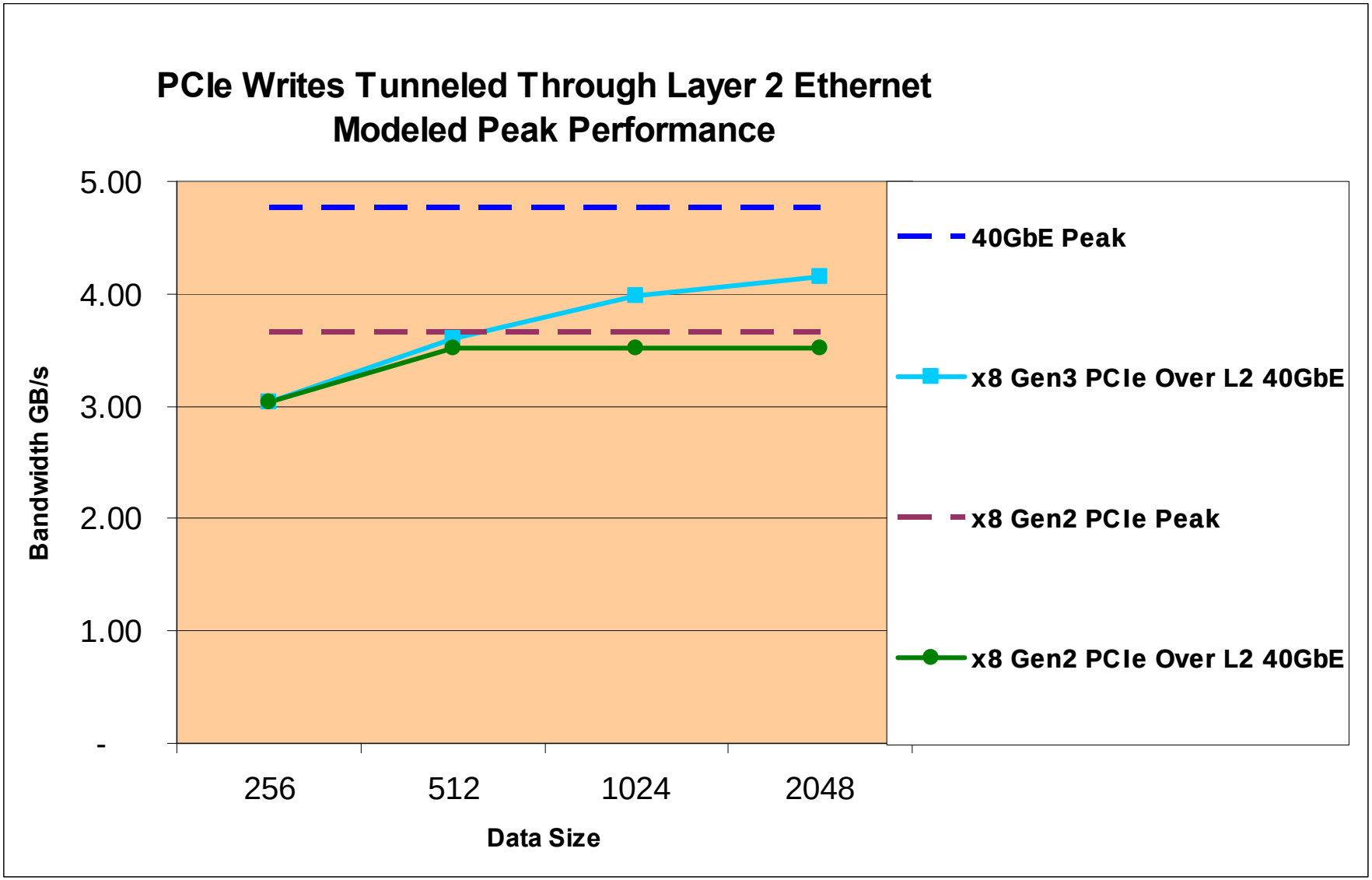
Layered Interconnect Architecture



POET 10GbE Throughput

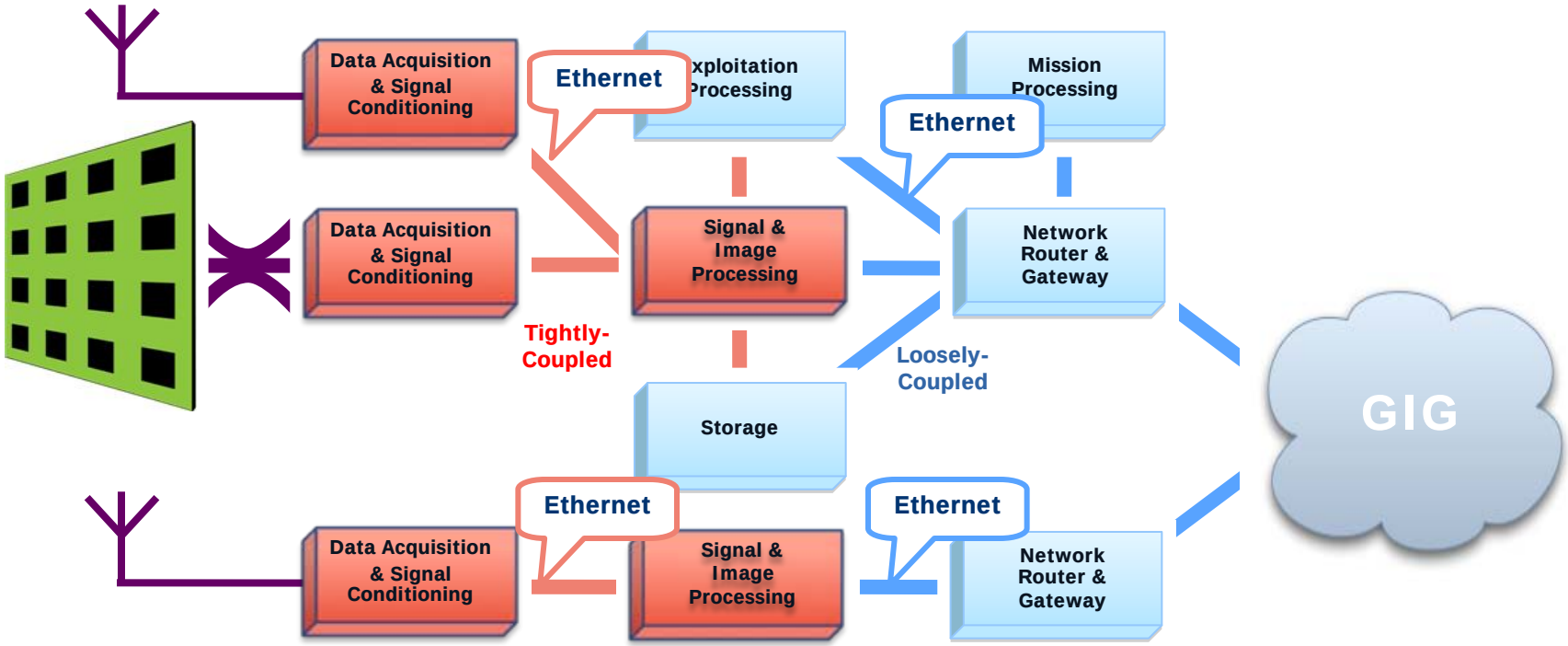


POET 40GbE Throughput



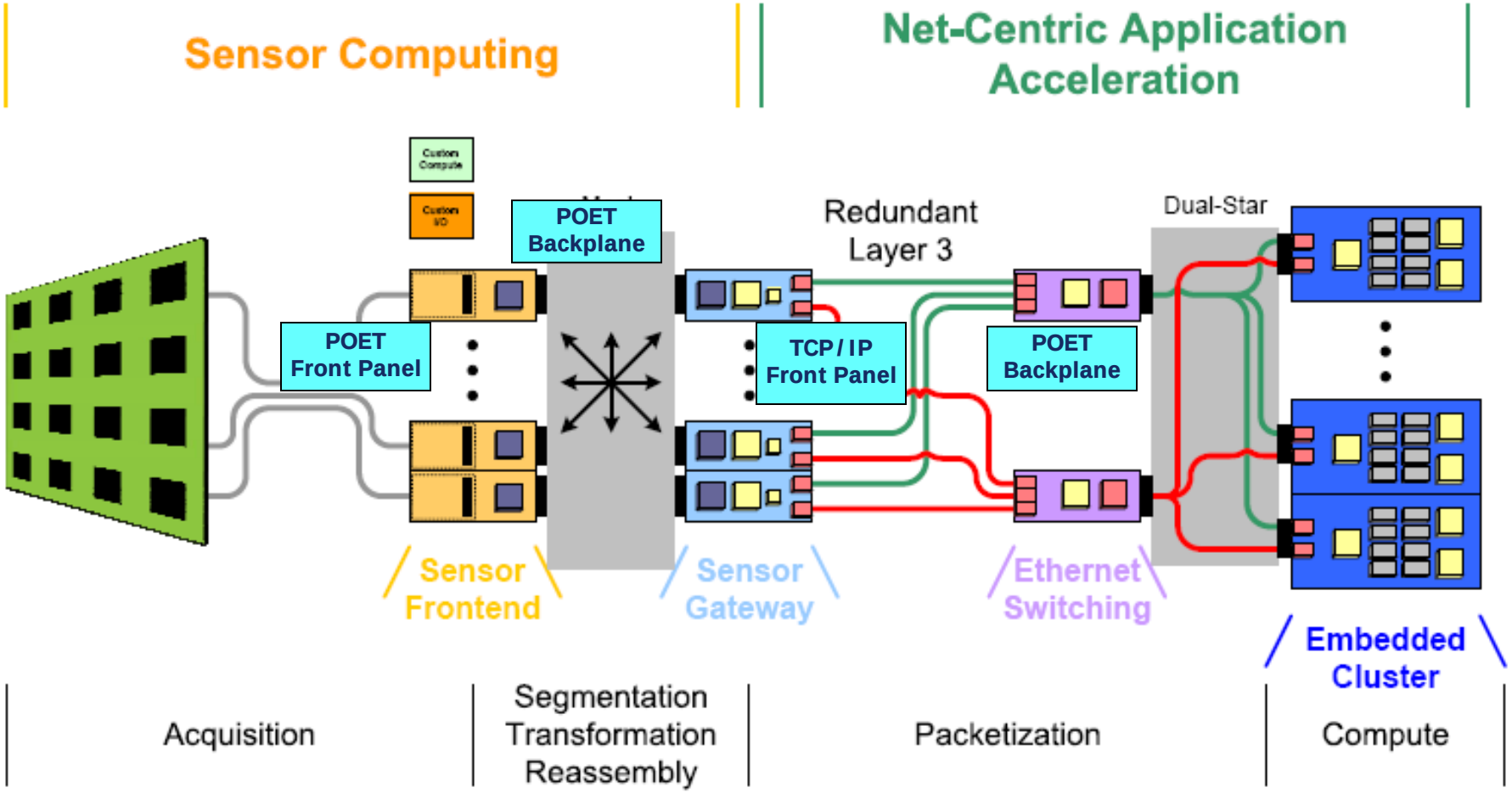
Summary: A New Level of Convergence

Bringing together signal and image processing, information exploitation, and information management



- **Integrated, optimized for low latency, high throughput, and SWaP**
- **Designed to deliver an "embedded" Quality-of-Service that supports convergence of processing and net-centric capabilities**

Typical System Architecture



Questions?