Server Bandwidth Scenarios

Signposts for 40G/100G Server Connections

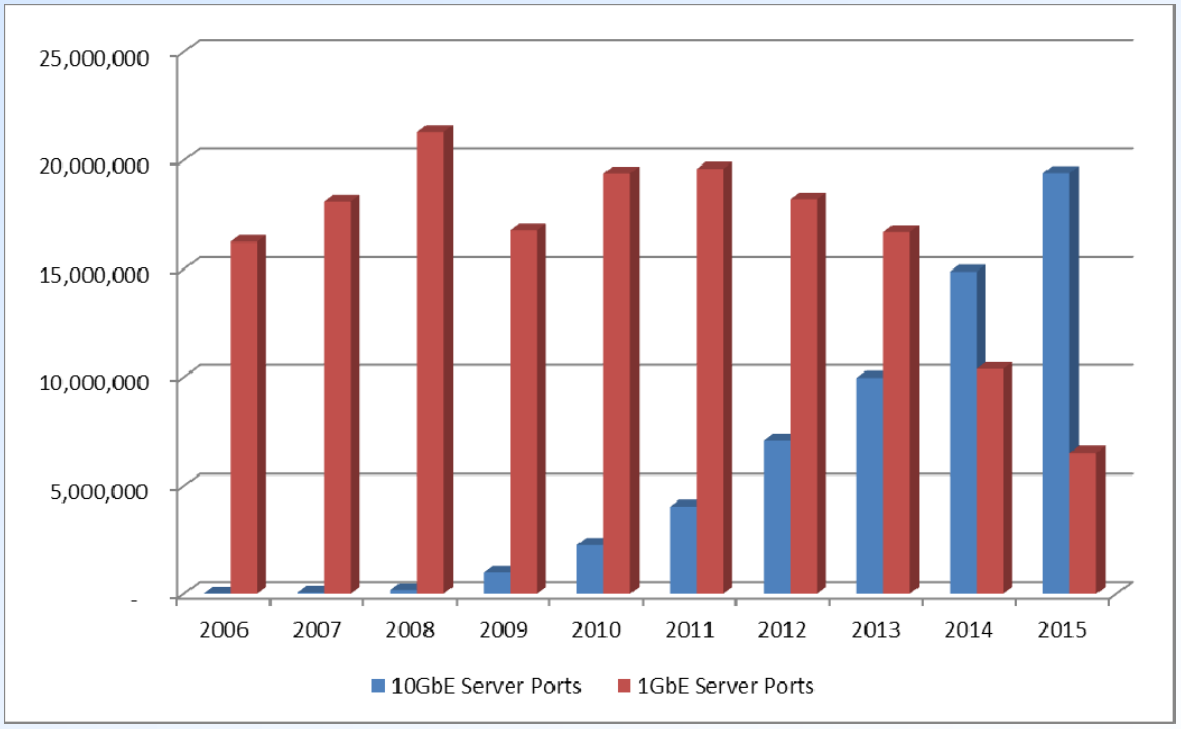
Presented by Kimball Brown
kimball@lightcounting

Server Breakdown

- x86 Servers garner about 2/3 of server revenues, but over 90% of server units
- Traditional premises software (i.e. OLTP, ERP, BI) is running on Mainframes and UNIX (they split about 1/3) and x86 (about 2/3)
- Web 2.0 (e.g. Google, Facebook) and financial trading companies (e.g. Goldman, Morgan Stanley) run x86 servers where Intel has about 80% share and AMD has about 18% share
- This presentation will concentrate on x86 servers

Servers Moving to 10Gbps

- Servers are beginning the transition to 10GbE



~17% of servers have 10GbE in 2011
~28% of servers have 10GbE in 2012
~37% of servers have 10GbE in 2013
~59% of servers have 10GbE in 2014
According to LightCounting

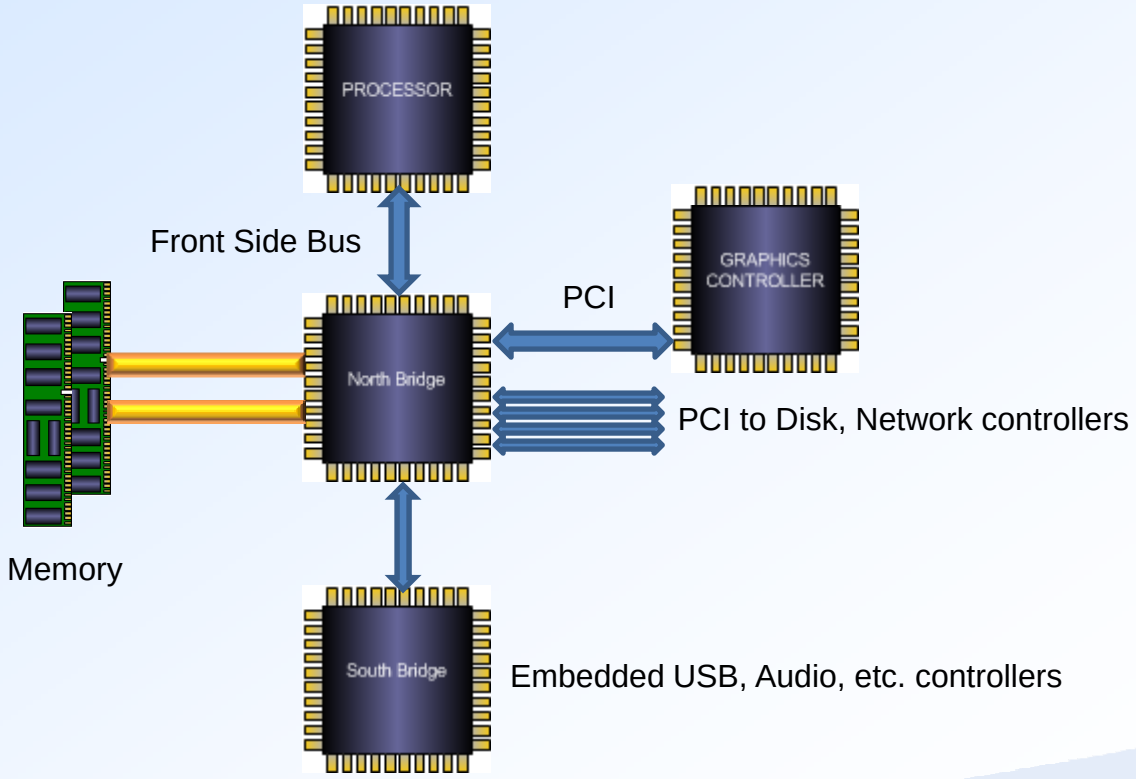
Source: LightCounting, July 2011.

Server Architectures

- Servers take data in, process it, and send data back out again
- Server Architectures relate to how the ever more powerful CPU(s), memory and I/O are connected
- x86 servers have evolved from a PC on its side into extremely powerful servers that rival UNIX RISC and Mainframe architectures

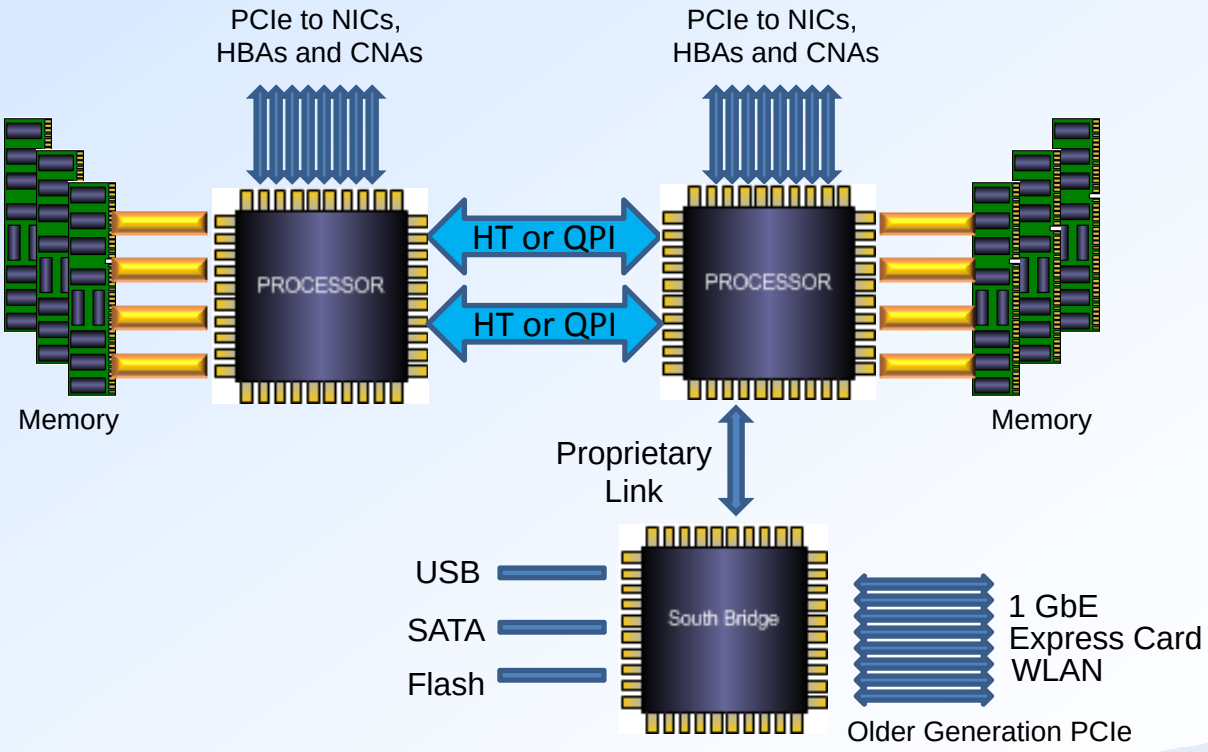
PC and early Server Architecture

North Bridge contained the Memory and High Speed I/O Controllers

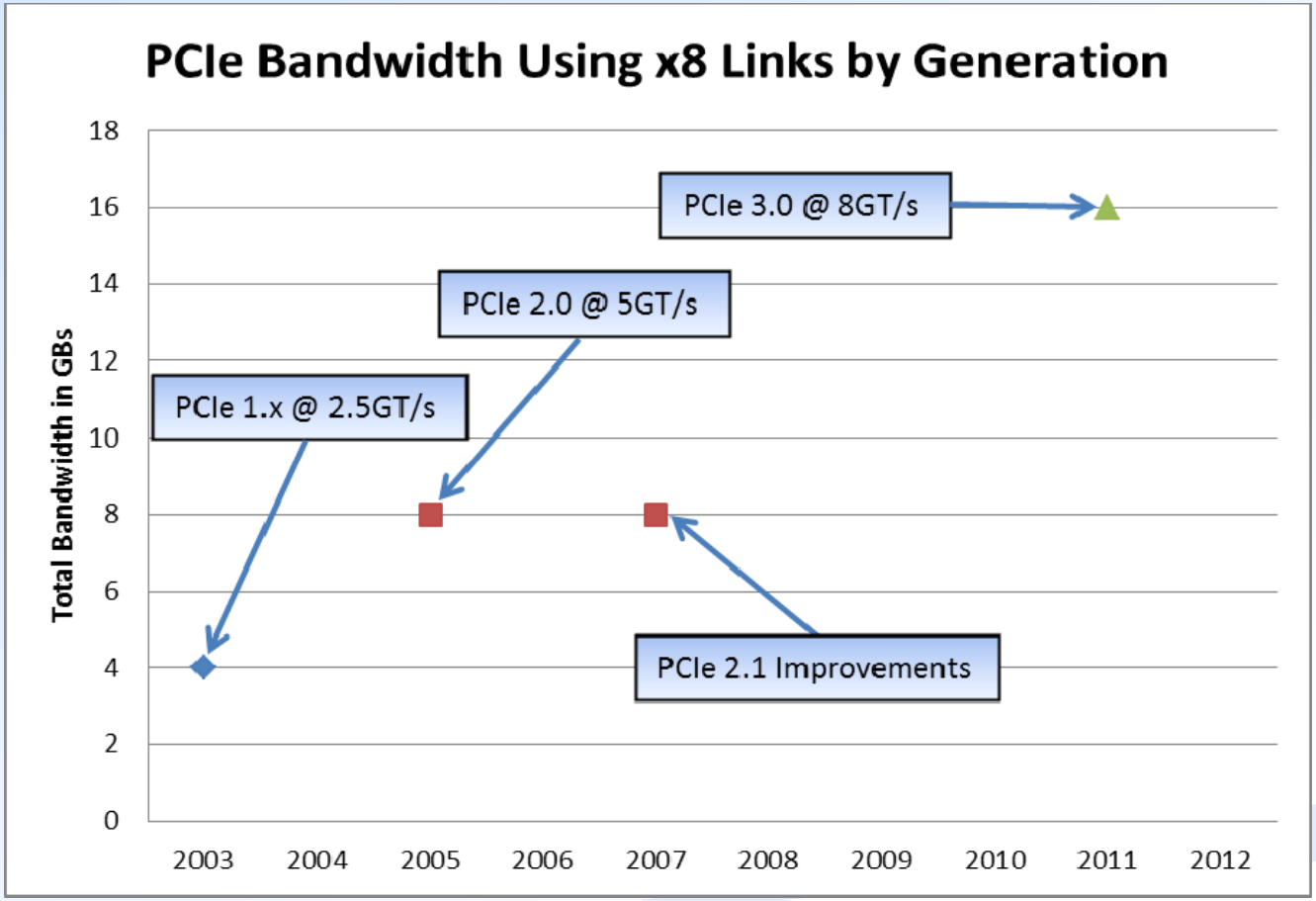


Memory and High Speed I/O Control Moves into the CPU

AMD via HyperTransport (2001) and Intel via Quick Path Interface (2009)



PCIe Generations



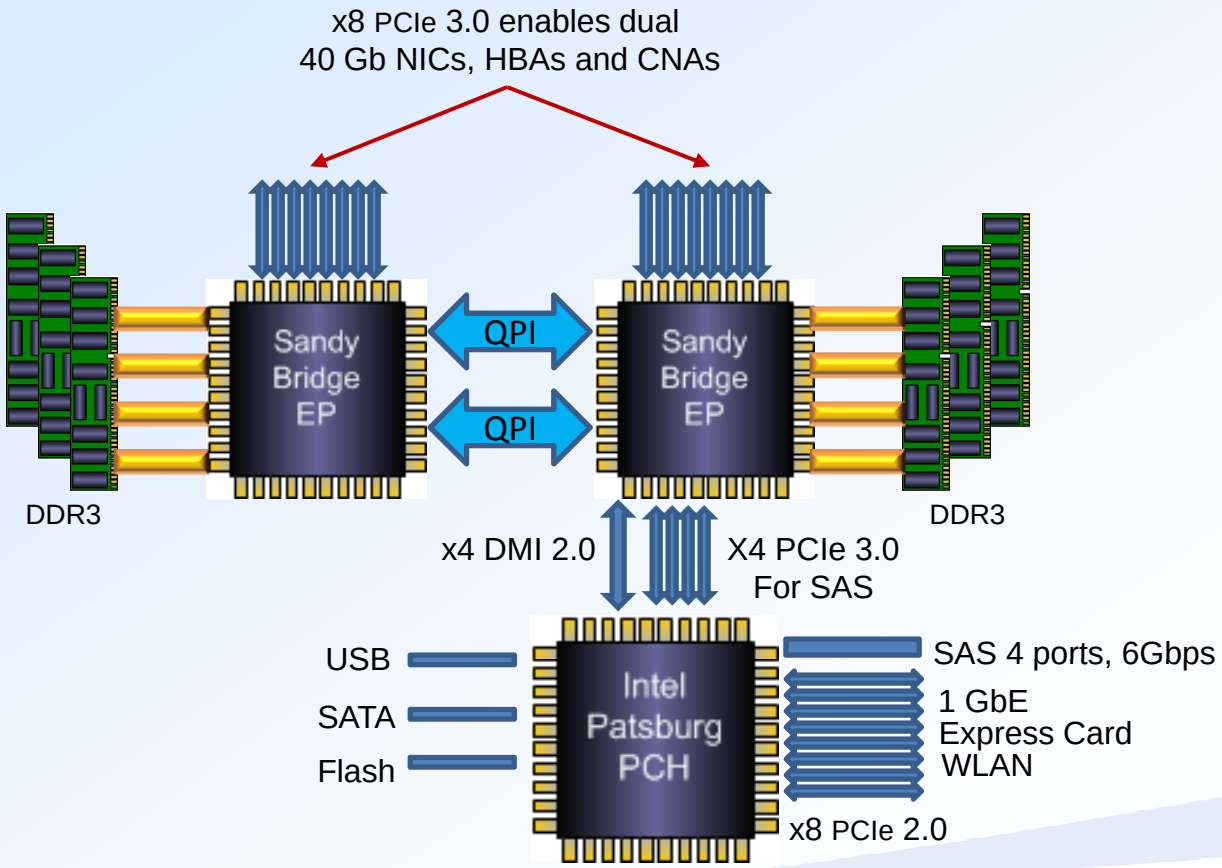
PCIe Bandwidth to Date

	Link Width				
	x1	x2	x4	x8	x16
PCIe 1.x Total Bandwidth (GB/s)	0.5	1	2	4	8
PCIe 2.x Total Bandwidth (GB/s)	1	2	4	8	16
PCIe 3.0 Total Bandwidth (GB/s)	2	4	8	16	32

- x8 is the most common high-end but low-cost server slot
- 16GB/s of total bandwidth translates to 8GB/s full duplex and to 64 Gbps full duplex
- Accommodates 40Gbps full duplex, but not 100Gbps

Intel PCIe 3.0 Server Architecture

If second CPU is installed, server can support dual port 40Gb connections



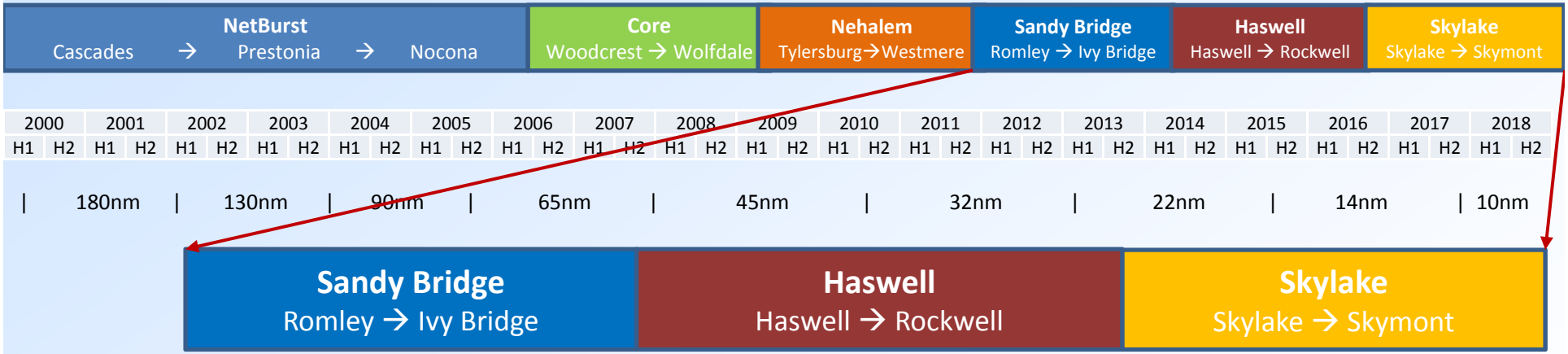
PCIe 4.0 Expected in 2015/2016

- PCIe 4.0 is still in the research stage
- Roadmap should be set by end of 2011
- PCI-SIG is driving the standard to double throughput again allowing 100GbE using x8
- If the PCI-SIG is successful, dual 100Gb connections will be in place by 2015 or 2016
- LightCounting believes PCI 4.0 will ship with Intel's Skylake Microarchitecture due in 2016

Moore's Law has Become Intel's Law

- Moore's Law: Transistor Counts Double Every 24 months
  - Enables Double the Performance or Half the Price
- Intel's Tick-Tock Model has Supplanted Moore's Law
  - Now, Every 24-30 Months, Transistors Double
  - One Year, a New Microarchitecture is Introduced
  - The Next Year, a CPU Shrink is Introduced

Intel Server Microarchitecture Roadmap



- PCIe 3.0 Ships with Romley
- Enables Dual Port 40Gb server connections

- 28nm 10GBASE-T LOM Option
- iSCSI moves rapidly to 10GbE
- FCoE grows rapidly with CNA LOM Option

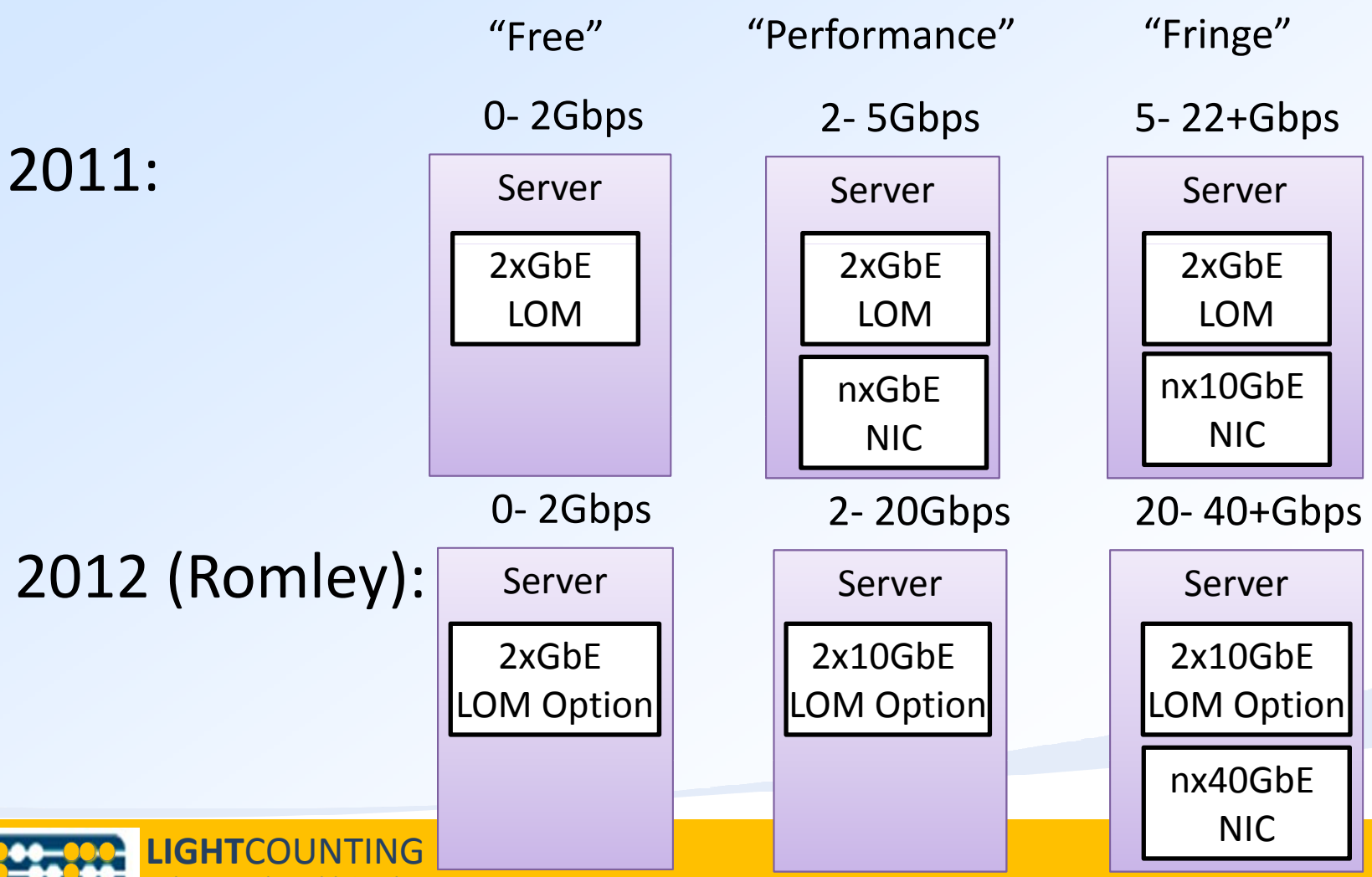
- PCIe 4.0 Ships with Skylake
- Enables Dual Port 100Gb server connections

Source: LightCounting, July 2011.

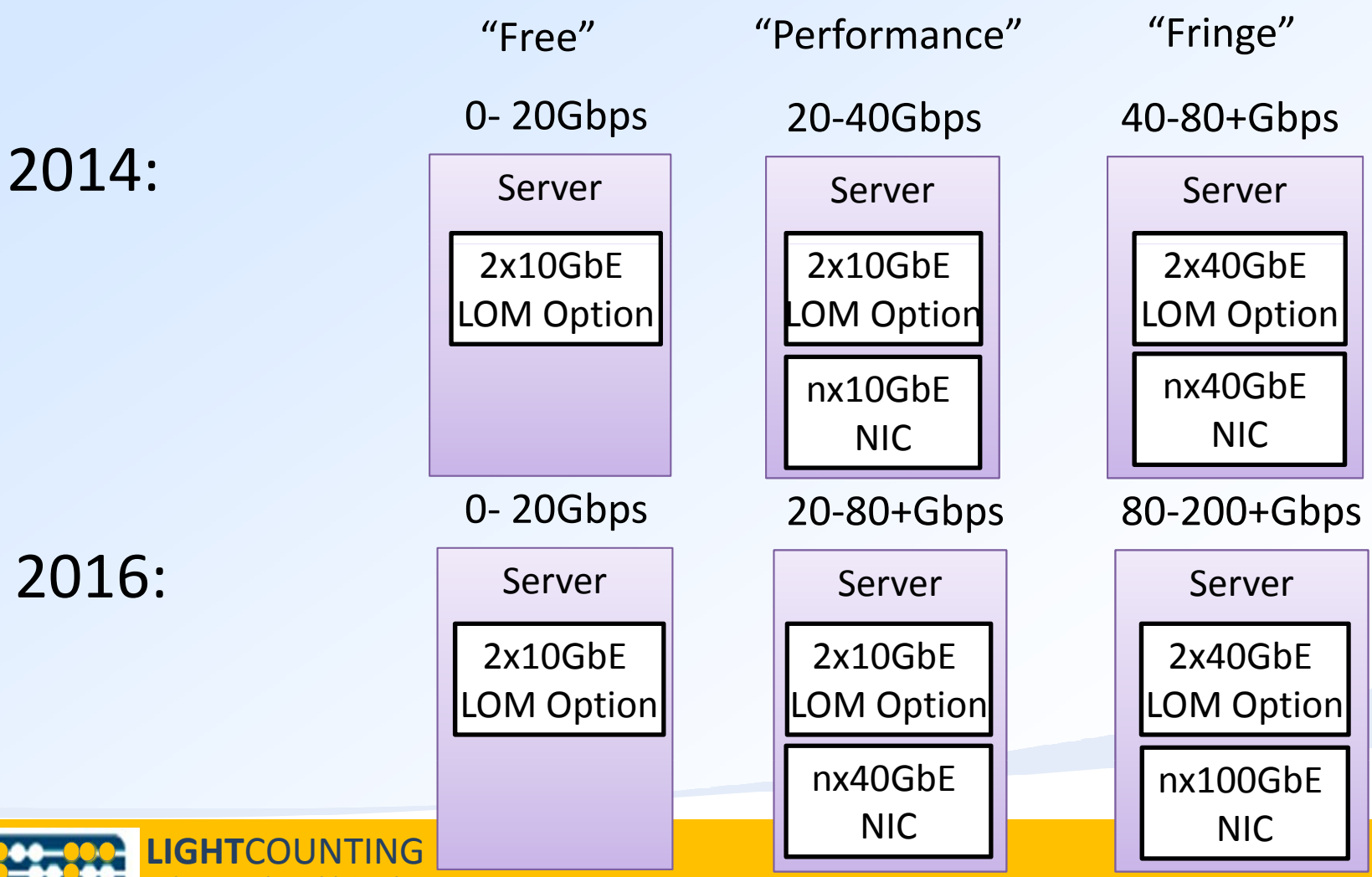
Next Gen Server I/O

- Ethernet needs to plan for the migration of servers from 1Gbps to 10Gbps to 40Gbps and then to 100Gbps
- The following slides will show how the I/O bandwidth of three classes of servers is expected to grow over the next few years
- Definitions of I/O Scenarios on next slide:
  - Free: A large portion of server buyers will only implement what is offered as the base configuration. These buyers would choose the "Free" option
  - Performance: Users who demand more I/O performance due to virtualization or, in some cases, the desire to converge the SAN and LAN networks within the rack
  - Fringe: Users who demand the most possible bandwidth. The servers that would need this bandwidth would typically be the high end 4 or 8 socket versions (translating into 40 or 80 cores in the Romley cycle, and huge amounts of memory).

Server Interconnect Scenarios (1 of 2)



Server Interconnect Scenarios (2 of 2)



Summary

- 10GbE LOM Option coming in Romley will drive adoption
  - Primarily used to enable up to 8 Virtual NICs facilitated by SR-IOV
  - Little adoption of storage over 10Gb Ethernet (iSCSI or FCoE) to date
- PCIe 3.0 in Romley generation enables dual 40GbE server ports via x8 PCIe links
  - Again, high end users are demanding 32 Virtual NICs over 40GbE ports to support highly virtualized environments
- PCIe 4.0 is expected to enable dual 100GbE server ports starting in 2015
  - The 2015 timeframe is an estimate, but we expect it will be in place by 2016
  - By 2015, we expect a majority of block storage (i.e. iSCSI or FCoE) traffic over Ethernet

Thank You

Kimball Brown

(408)799-3500

kimball@lightcounting.com