BROCADE

100GbE – THOUGHTS FROM A SYSTEM PERSPECTIVE

David Skirmont

March 7, 2012

It Is All About the I/O

- What does an Ethernet switch do? It's simple…
  - Takes in frames
  - Figures out where they need to go
  - Stores them for a while, maybe deletes a few
  - Sends them out a different interface
- Each step is limited by its own I/O requirement
  - Front panel I/O
  - Packet processor interface and lookup bandwidth
  - Traffic manager buffer memory bandwidth
  - Switching fabric interface bandwidth
- All need to grow to support higher capacity switches

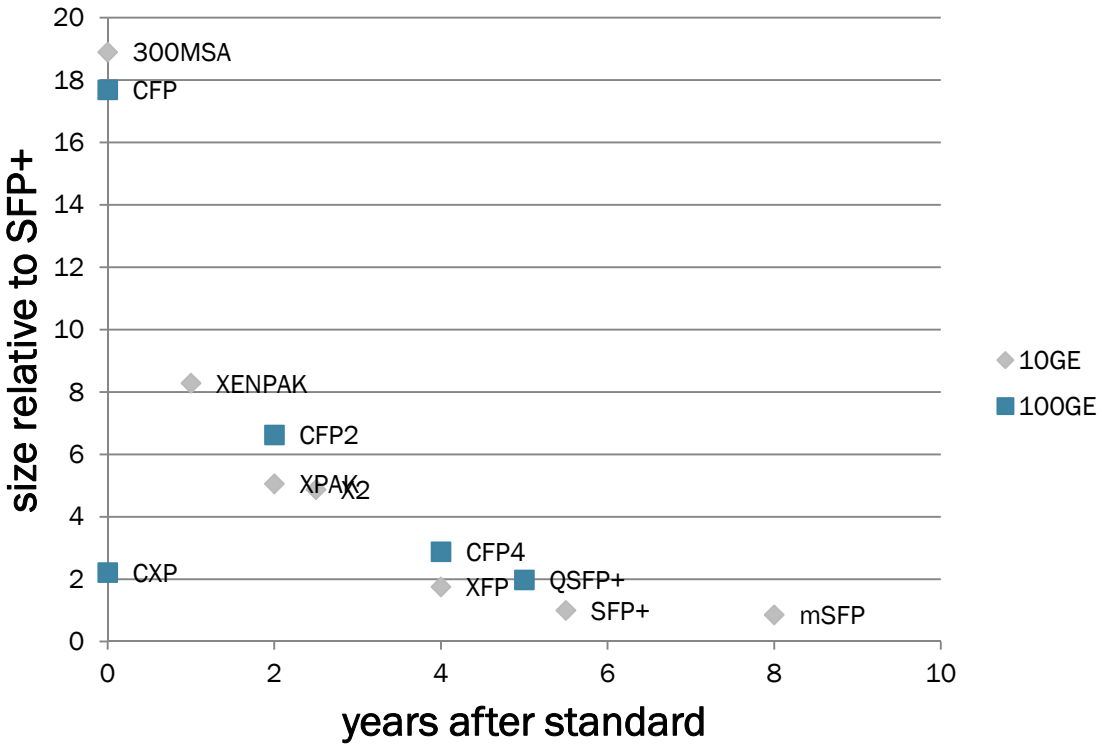
Front Panel I/O and Module Size

- Blade bandwidth is limited by the modules on the front panel

- Today, 4xCFP lags 48xSFP+ for total bandwidth/area
- 8xCFP2 will double bandwidth to 800Gb/s
- 16xCFP4 will be 4x the bandwidth at 1600Gb/s
- 2x18 double stack QSFP+ could give 3600Gb/s!
- CXP is also dense but can't fit all PMDs

Module Size Though the Years

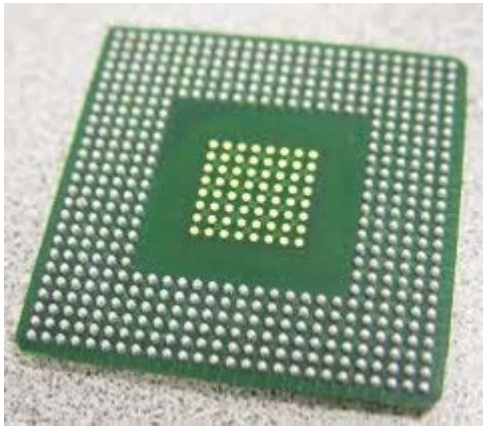
- Started off huge
- Very expensive
- Got smaller over time
- Got cheaper too!



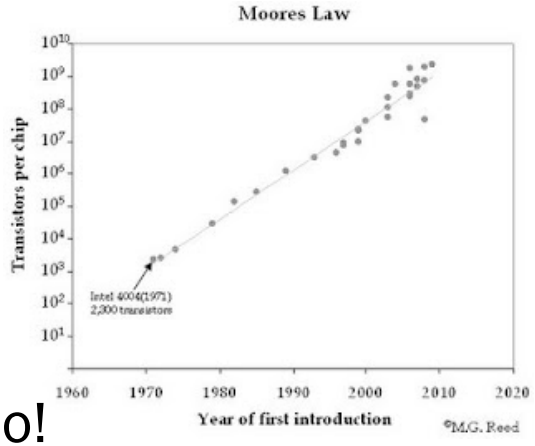
Box of lab junk: glue stick, CF card, 10GbE SFPs…

The Move to 4x25Gb/s Module Interfaces

- The current CAUI 10x10 interface is a limitation
  - A 100GbE port takes the same ASIC pin resources as 10x10GbE
  - Switching throughput remains the same
- Typical ASIC SERDES count tops out at ~96-128
  - Need those pins to carry more data!
  - 2.5x improvement with a CAUI-4 interface
- An ASIC with 100 SERDES has options
  - 100x10GbE for 1Tb/s of switching
  - 25x100GbE for 2.5Tb/s of switching
  - Or a mix and match



Addressing Other System Limitations



- ASIC technology improves at each node
  - Faster switching times
  - Greater gate density
  - Except for the masks, almost the same price too!
- External memories moving to serial I/O
  - Faster TCAM lookup rates
  - Fewer pins to support wider logical block sizes
  - Under the hood DRAM speeds are not really increasing ☹
- Backplane I/O getting faster
  - 25Gb/s capable connectors and topologies
  - IEEE 802.3bj 100Gb/s backplane and copper cable TF

Reversible Gearboxes for 10x10 PMDs

- There will still be 10x10G in a 4x25G world
- A reversible gearbox is needed for 10x10 PMDs
  - Need support for legacy 10x10G PMDs in a 4x25G module form factor
  - Must deal with lane-to-lane skew on 10x side
- A multi-link gearbox is also needed
  - Breakout a 4x25G module into 10 individual 10GbE interfaces
  - Different timing domains for each 10GbE interface
  - Increases 10GbE density on both the front panel and the ASIC
  - Already breaking out a 40G-SR4 QSFP into 4x10GbE
  - Addressed by the OIF MLG project

All of Those PMDs

- 10GbE was basically a serial PMD for MMF and SMF
- 100GbE complicates things with 10x10G PMDs moving to 4x25G PMDs
- 100G-SR4 makes sense for datacenter
- Parallel or different CWDM SMF PMDs less so
  - Confuses the market relative to LR4
  - Fragments the market for LR4
  - Is there a real need/benefit versus cost/confusion?
  - Lots of debate in 100GE_NGOTX study group
- How about a cheaper, shorter reach, yet still –LR4 compatible PMD

New Form Factors

- Modules are getting smaller, that is good!
- A given system should use the “current” form factor available at that time
  - This implies that all PMDs should be available in that form factor
  - Mix and match is difficult to support over multiple platforms
  - Default to lowest common denominator (CFP-CFP2-CFP4)
- Complicated with QSFP+ based PMDs
- 300pin MSA transponders still work with SFP+!
  - Still running the same 10GbE serial PMD on the fiber

BROCADE

Thank You