The Case for Lower Cost Channel Support

IEEE P802.3bj **100Gb/s Backplane and Copper Cable Task Force**

Atlanta

November 2011

Kent Lusted, Intel Corporation

Dave Chalupsky, Intel Corporation

Supporters and Contributors

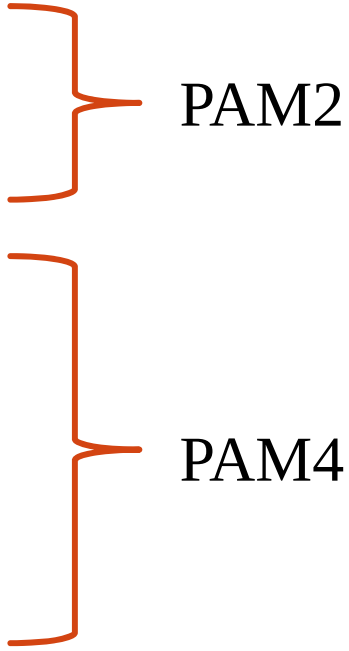
- Matt Brown, APM
- Venkatesh Nagapudi, APM
- Howard Frazier, Broadcom
- Vasudevan Parthasarathy, Broadcom
- Brad Booth, Dell
- Bhavesh A. Patel, Dell
- Ilango Ganga, Intel
- Bob Grow, Intel
- Rich Mellitz, Intel
- Adee Ran, Intel

Key Points for 100Gb in the x86 Server Market

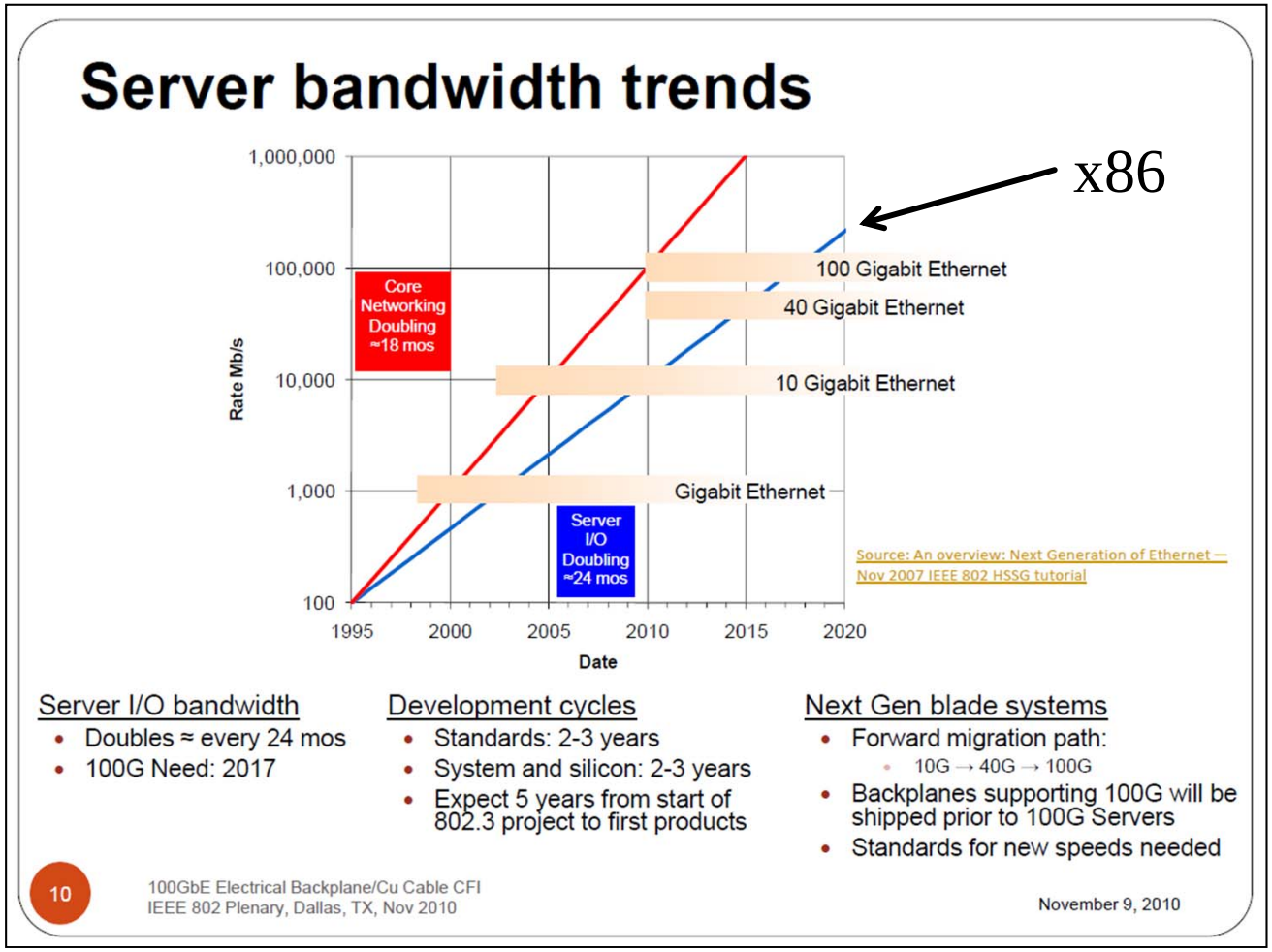
- Port mix: 100G will coexist with 10G & 40G
- Ethernet is just 1 of the many interfaces on PCB
- Corporate environmental and social responsibility is driving changes to PCB materials

100G Backplane Applications

- Edge/Core routers & switches
  - “Forklift” upgrade path – rip and replace
  - Demands high performance today!
- End point x86 servers
  - Modular upgrade path
    - upgrade components over the system/server lifecycle to maximize ROI
    - Leverage KR/KR4 era channels (frazier_01_0911.pdf)
  - Needs cost effective performance tomorrow
- 2 PHYs could address these applications
  - Each has strengths and weaknesses
    - (Brown_01_0911.pdf and hatab_01_0911.pdf)

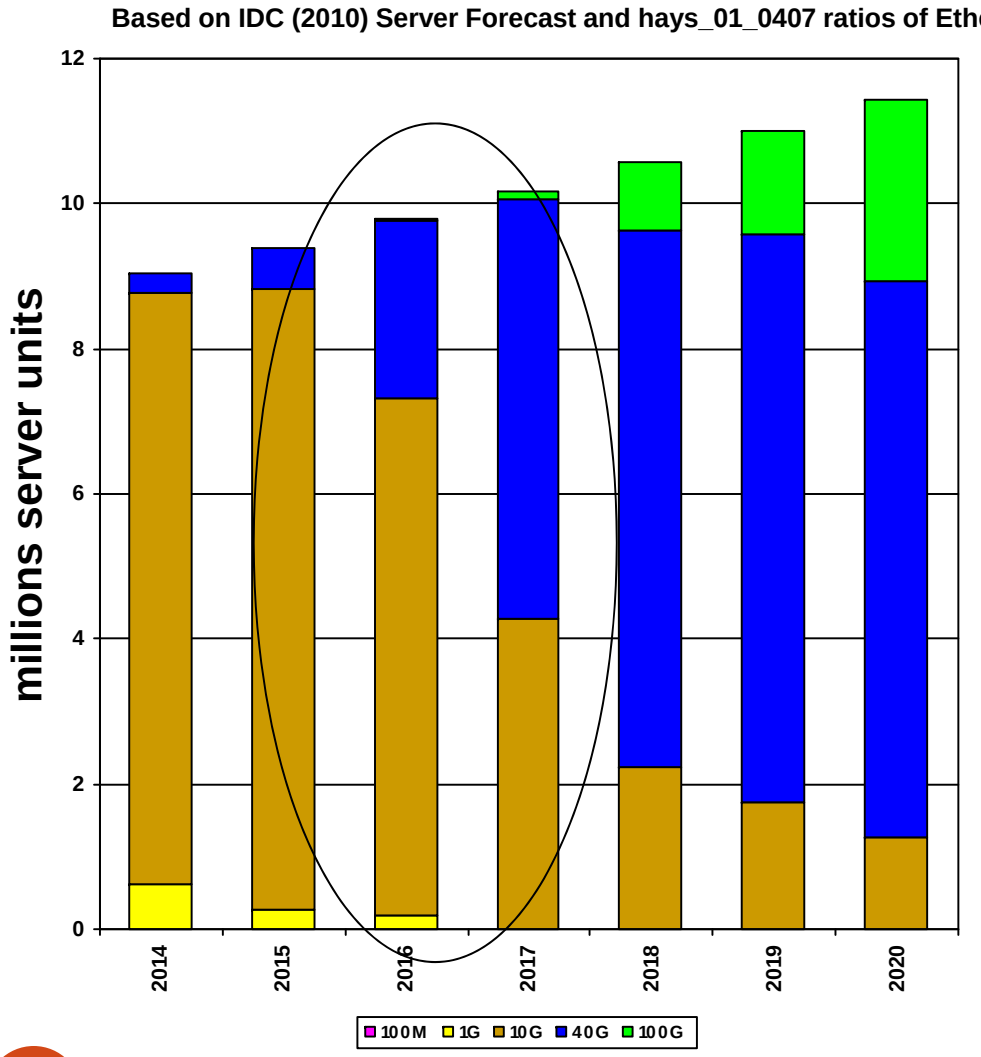


Core vs. x86 Server Trends



Source: http://www.ieee802.org/3/100GCU/public/nov10/CFI_01_1110.pdf

X86 Server Port Mix at Introduction



At introduction, 100G server ports will coexist with 10G & 40G…
Even some 1G

Blade and Rack Servers should support all these speeds

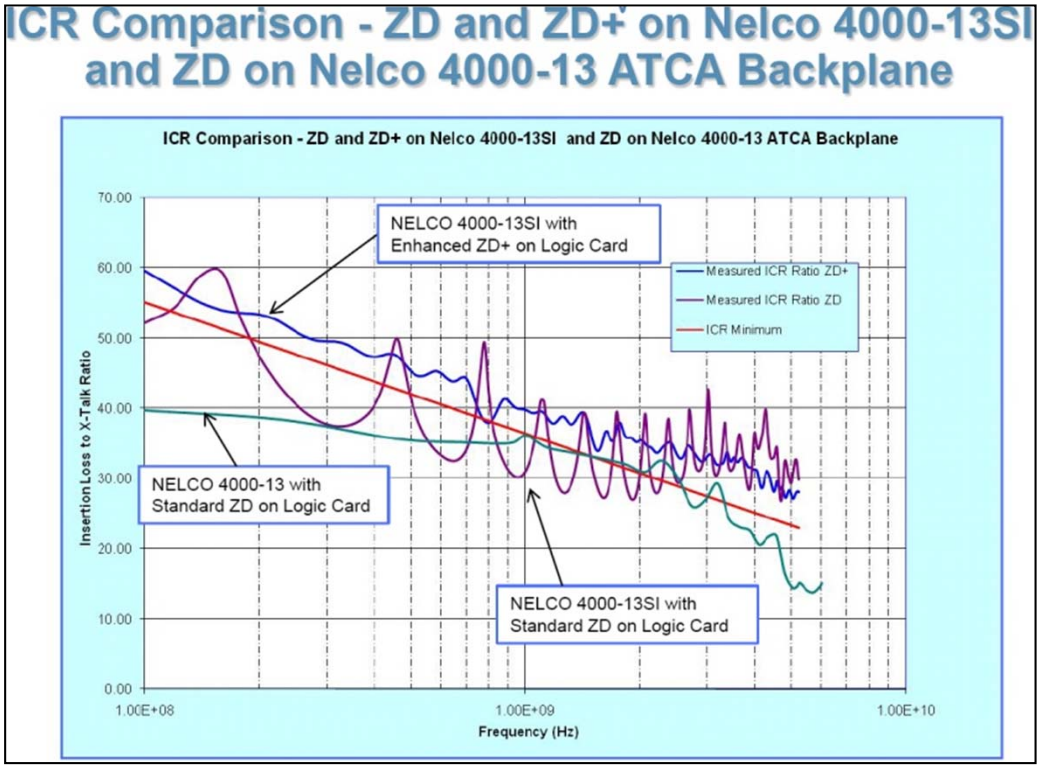
Avoid putting a cost burden on 10G/40G ports

Broad Market Potential - ATCA

- ATCA = Advanced Telecom Computing Architecture
  - Created to meet requirements of “carrier grade” comms equipment → Telco
  - Part of PICMG (PCI Industrial Computer Manufacturers Group)
    - http://www.picmg.org
- ATCA will soon add formal support for 10GBASE-KR!
  - PICMG 3.1 R2.0 ECN
  - Also adds 40GBASE-KR4
  - “Beyond integration of 10Gb Ethernet, one of the primary goals of the PICMG 3.1 Revision 2.0 subcommittee was interoperability and backward compatibility with existing ATCA equipment.”
    - http://blog.radisys.com/2011/02/picmg-tackles-interoperability-and-backward-compatibility/
- “Backward compatibility becomes more crucial as we can see a subset of platforms scaling from 200W in legacy platforms with 1G and 10G to beyond 200W platforms with 1, 10 & 40G support
    - http://www.advancedtcasummit.com/English/Collaterals/Proceedings/2010/20101111_SpecTutorial_Freudenfeld.pdf

ATCA Connector Migration

- PICMG vendors are migrating to enhanced ATCA Zone2 Fabric connectors such as the ZD+
  - Footprint compatible was a requirement
  - ZD did not meet performance
  - ZD+ series created for 10G/40G



Source: http://blog.radisys.com/wp-content/uploads/2011/02/ICR-Comparison.jpg

Open Architecture of ATCA

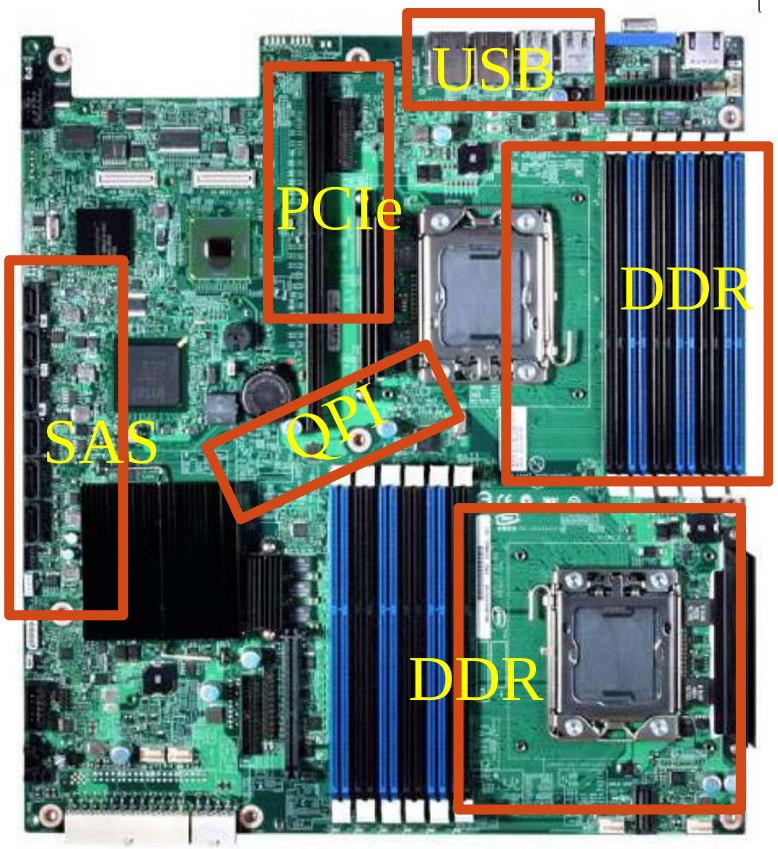
- Many suppliers of the different subcomponents
  - Decouples development schedules of blades and backplanes
  - All must operate seamlessly
  - Different from closed architecture of most blade server systems
- “IEEE defined the characteristics of the channel based on hypothetical test points at either end, but did not address the details and complexities of applying that channel model to an open, multi-vendor, bladed platform ecosystem such as ATCA.”
  - http://blog.radisys.com/2011/02/ethernet-on-the-40g-backplane/

X86 Server Development Environment

- High volume server motherboards are very cost sensitive
  - PCB technology is still standard FR4-class materials
  - Typical server motherboard = 130-150 sq. inches
  - It is a significant evolution to transition to 802.3ap spec'd "improved" FR4 materials
- Most volume server designs are outsourced to keep development costs low
  - Server platform enablement teams distill complex design problems to design rules/guidelines
    - CPU/Memory core layout is typically "copy exact" from a reference design
    - LAN is routed in remaining space ☹

Typical X86 Server Topology

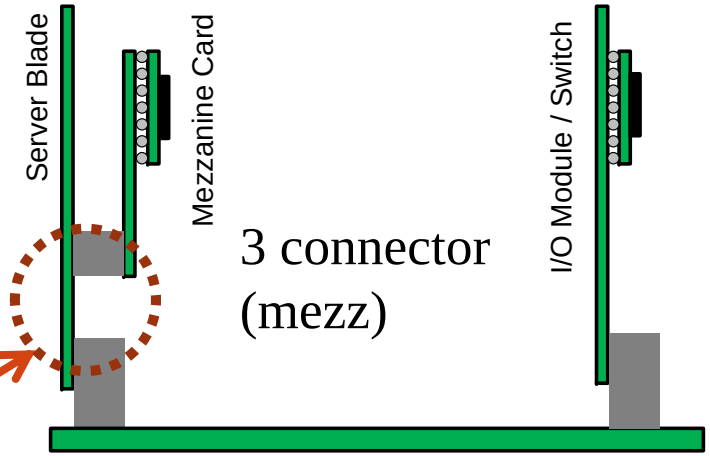
Source: http://download.intel.com/support/motherboards/server/s5520ur/sb/e44031012_s5520ur_s5520urt_tps_r1_9.pdf



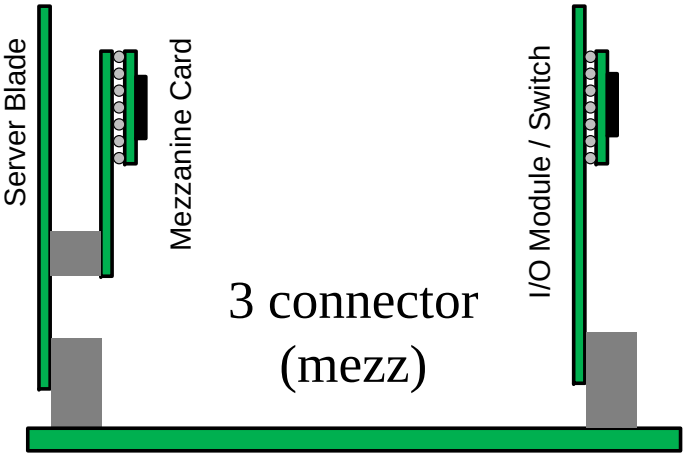
- LAN Routing is not a priority
  - Ethernet is just 1 interface on the x86 server
- Other key interfaces drive PCB requirements:
  - DDR – memory interconnect, 75-95 ohm Zdiff
  - QPI –CPU interconnect, 85 ohms Zdiff!
  - PCIe – expansion card, 85 ohms Zdiff!
  - SAS – to mass storage
  - USB – peripherals interconnect, 90 ohms Zdiff
- QPI & DDR get highest priority
  - If they don't need higher cost materials, then none get them
  - Future platform DDR4, QPI and PCIe requirements encourage use of lossy materials to mitigate reflections on short channels.

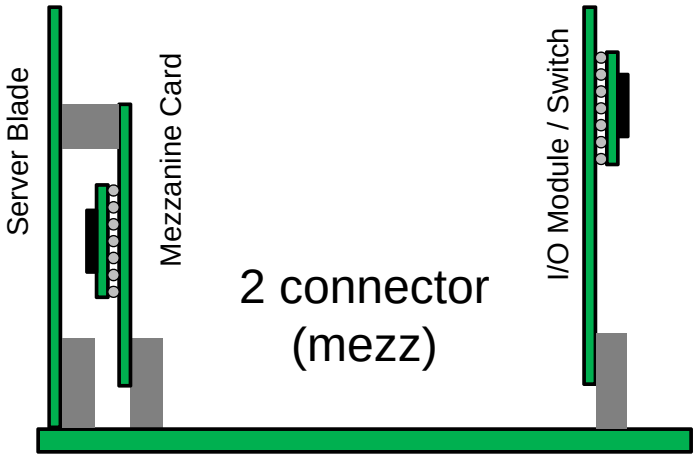
Two Common X86 Blade Server LAN Topologies

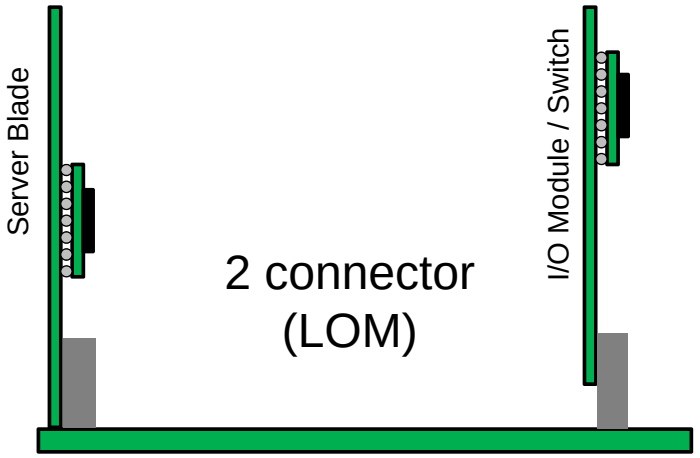
- Mezz concept brings flexibility and versatility of interfaces in a deployed system
  - Existing server & midplane can support 100G Ethernet Mezz card & switch upgrade
- **3 connector:**
  - LAN signals route back to mother board,
  - P802.3ap IL budget of 25dB @ 5GHz facilitated system vendor innovation
    - trade channel length for 3rd connector and FR4
  - **~60%** by vendor unit share
- **2 connector:**
  - LAN signals direct to midplane,
  - **~30%** by vendor unit share



Topology Details





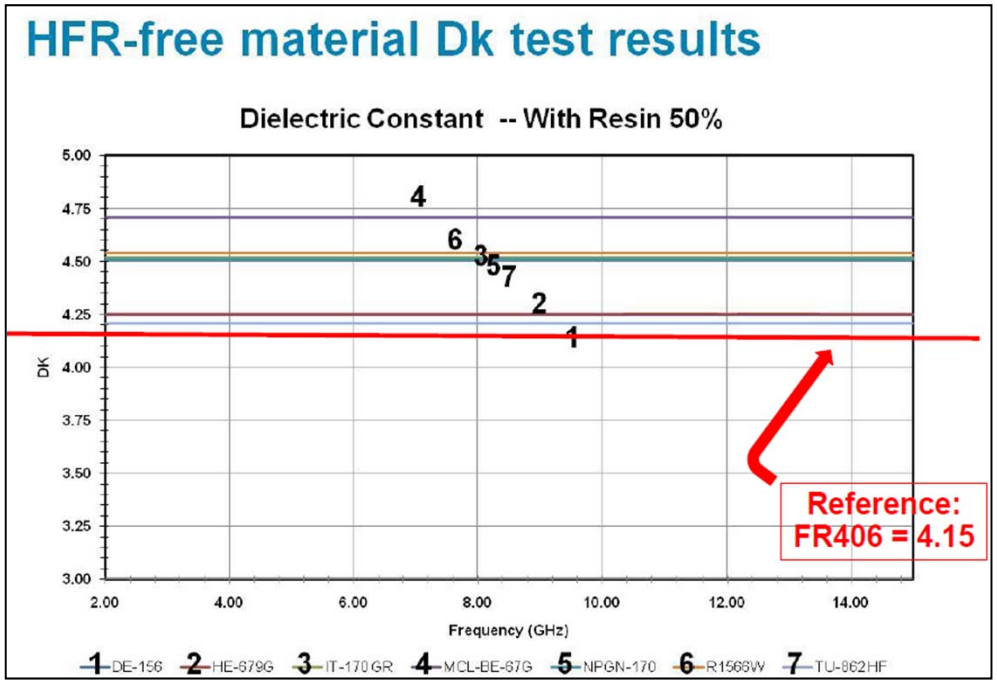


Some Options for PAM2 on 3-connector Backplane Designs

- Stay with 3 connector architecture
  - Reduce channel reach or limit # of slots supported
  - Use higher cost, low loss materials for server motherboard
    - Low Loss dielectrics 2-6x higher cost than basic FR-4
      - http://www.ieee802.org/3/100GCU/public/mar11/goergen_02b_0311.pdf
    - High speed, low loss PCB material is only 4.1% of the WW PCB production (by area) Prismark Printed Circuit Report, Q3'2010
  - Add retimer(s) in path
    - (mohan_01_0911.pdf)
- Migrate architecture to 2-connector solution for 100G deployment
  - Re-architect and re-partition the product line
    - Design 2 full product lines in parallel?
  - No easy upgrade path for customers
- Skip 100Gb 4x25 technology

PCB Regulatory Hurdles for x86 Server Market

- There are no low loss material options for servers
- RoHS: Restriction of Hazardous Substance
  - “Lead Free” materials today
  - Many compute products have made the transition to Lead Free
- Next big challenge is Halogen Free
  - HF is higher Dk than FR4
    - Best current HF material is similar to standard FR4…
  - No Low Loss or Ultra Low Loss Df equivalent yet
- Ultra low loss PCB is not suitable for servers



Source: http://thor.inemi.org/webdownload/newsroom/Presentations/HFR-Free_Signal_Integrity/HF_Test_Proposal.pdf

Summary Points for 100Gb in the x86 Server Market

- Port mix: 100G will coexist with 10G & 40G
  - Mezz provides upgrade versatility
  - Open architectures (ATCA) necessitate compromises; bandwidth limited channels are prevalent
- Ethernet is not the priority route on the x86 server platform
  - Other interfaces drive PCB material selection
  - QPI and DDR target lossy materials to attenuate reflections from packages/connectors
- Corporate environmental, social responsibility and government regulations are changing PCB materials
  - Halogen Free materials are higher Dk that standard FR4
  - There are no low loss Df material options for servers yet
- Consider 2 PHY solution: PAM2 and PAM4

Thank You!