100G Electrical Interfaces in the Datacenter
Desirable Solution Attributes

Rob Stone

5/24/17

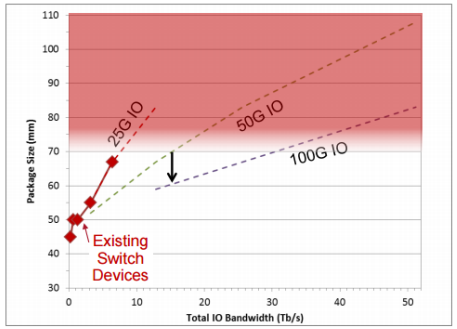
BROADCOM®

This is a non-scientific presentation intended to spur discussions!

Why will datacenters migrate to 100G Electrical IO?

- Quick recap: *(see http://www.ieee802.org/3/ad_hoc/ngrates/public/17_03/goergen_nea_01a_0317.pdf)*
- Increasing system bandwidth demands increased electrical lane speeds -
- Why? → primarily IO density driven:
  - ASIC (package escape)
  - Backplane (limited number of connector conductors, PCB routing)
  - Front Panel Module electrical Connector Density (current state of art is 36 x 8 lane modules)

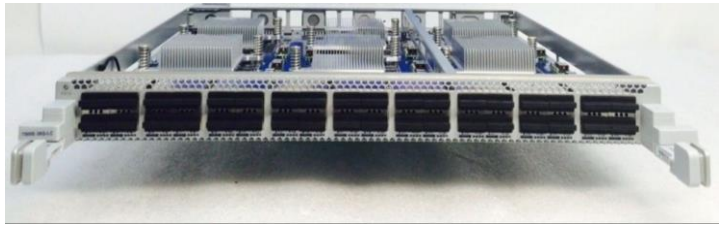
IO Escape forcing transition to higher lane speeds



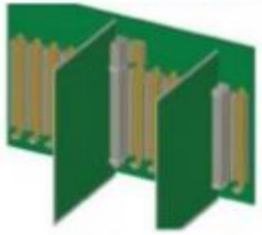
- ~ 70mm package is a current BGA practical maximum (due to coplanarity / warpage)
- This will force BGA devices with > 14Tb/s of aggregate bandwidth to transition to lane rates of higher greater than 50G (possibly 100G?)

IEEE 802.3 NEA Ad hoc, IEEE 802 March 2017 Plenary, Vancouver, BC, Canada

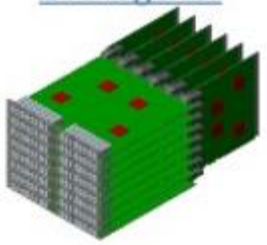
Cabled Backplane



Traditional Backplane



Orthogonal



BROADCOM

Datacenter End User Wants

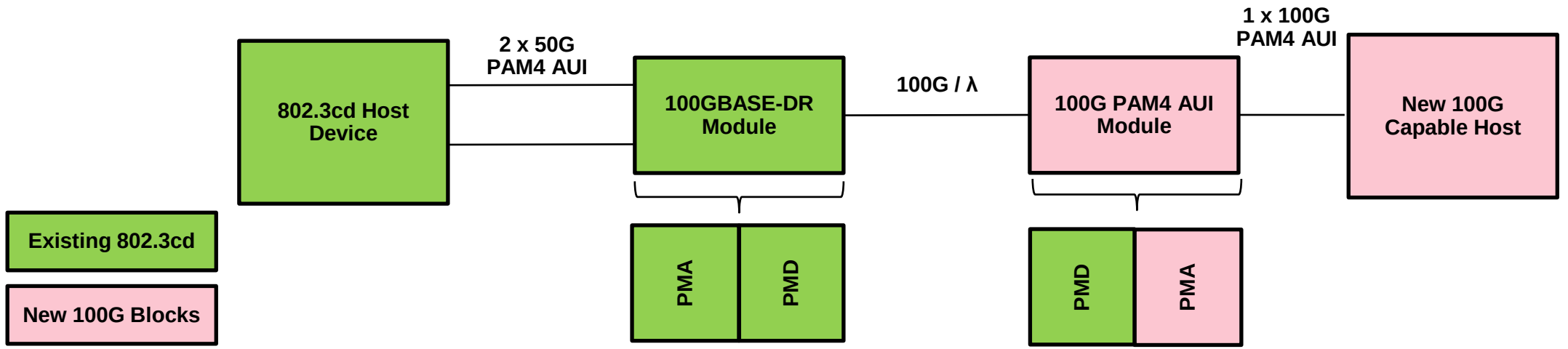
- Compatibility
  - 100G / λ Optical PMDs are already in development (paired with 50G AUIs)
  - Connection of 50G IO systems to 100G IO systems should be expected (suggests the same PCS would be a sensible choice)
- Low Fabric Latency
  - Machine Learning / Alternative Computing / Public Cloud demand low latency for best application performance
  - Latency needs to be same (or lower) as 50G IO generation (at same port speed)
- Low Power
  - Datacenter PUE remains a key metric for end users
  - Energy per bit needs to fall consistent with increasing system BW (otherwise network power outpaces supply)
- Lower Cost
  - Suggests higher fabric system density (less total components)
  - Desire for simplified system designs

Switch Silicon Implementer Wants

- Low area overhead coexistence of 100G serial IO in a multi-rate high density ASIC with existing port types
  - 100G PAM4 is the obvious modulation format choice (common AFE)
  - Hard to introduce a radically different serdes architecture and keep the silicon area compact
- PCS compatibility
  - If we can achieve this, the change could be as simple as different PMA gearing only for the logic side (serdes changes obviously)
  - No requirement for yet another FEC / PCS to be supported
  - Enables simple backwards compatibility to existing 50G PAM4 based systems

Backwards Compatibility Considerations – ideal situation

- Example – based on 802.3cd (apologies to 802.3bs! – same thinking applies..)
  - Effectively change module PMA gearing from 2:1 to 1:1
  - No changes to PCS, or PMD required
  - Would have to ensure each segment remains within the current DER budget



Oversimplified Electrical Channels

Interconnect Type	50G PAM4 Loss (dB)	100G PAM4 Loss (dB)
Traditional Backplane	30	~ 60 ?
Passive Copper Cable (DAC)	30	~ 60 ?
AUI (C2C)	20	~ 40 ?
AUI (C2M)	10	~ 20 ?

- Unlikely we can support existing channels at the same DER for backplane, DAC and C2C
- Looks promising we can support the C2M with an end – end architecture
- Possible Options to address backplane, DAC, C2C
  1. Use Extender FEC (segment by segment protection)
  2. Use stronger end – end FEC with a new PCS
  3. Change the channels which are desired to be supported… (cables vs PCB, new materials)

Summary

- We should make efforts to reuse the existing RS(544) based PCS architectures
  - Addresses majority of end-user and silicon implementer wants
    - Avoids stranding 50G based systems, maximizes compatibility
    - Offers same latency as 802.3bs / 802.3cd PCS
    - No changes required to optical PMDs
  - Offers system designers maximum flexibility:
    - Allows use of improved channels with end – end FEC (Cabled solutions, new PCB materials)
    - Allows use of extender FEC or retimers for longer channel support
    - Allows use of active copper cables

Thanks!