Short Host Channel System Implications

Rob Stone

5/9/18

Use of DAC
## Current generation technologies

- Server (or more generally endpoint) to ToR connections
  - Low cost, low power, used in close to all use cases today

- Fixed box "virtual chassis"
  - Example is Facebook "Fabric Aggregator" with DAC based "sideplane"

- Relevant attributes of these designs for this discussion:
  - All IO is from the front of the box (no backplanes)
  - Long PCB traces from the most distant ports to the switch (~ 9")

- Will this scale to 100G / lane DAC?
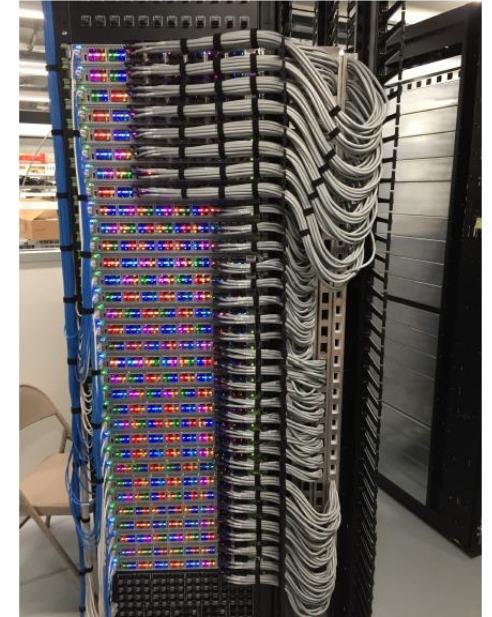
From: FacebookFabricAggregatorOCP_Spec_v1.0
*(OCP)*

BROADCOM®

Architectural changes to ToRs due to reduced physical VSR reach

- Hypothetical Example:
  - 25.6T, 256 x 100G
  - 1RU box, Single ASIC (ToR design profile, also used as virtual chassis, aka "Fixed Box")
  - 32 x 800G module cages, all front panel IO

- Using Rosemont budget proposal from Jane Lim:
  - http://www.ieee802.org/3/100GEL/public/18_03/lim_100GEL_01b_0318.pdf
  - [~ 5" Host trace supported for VSR channels]
  - Approximately 12 / 32 module cages cannot accommodate the proposed host budgets (VSR or CR), requiring either intermediate retimers, or intra-box cabling
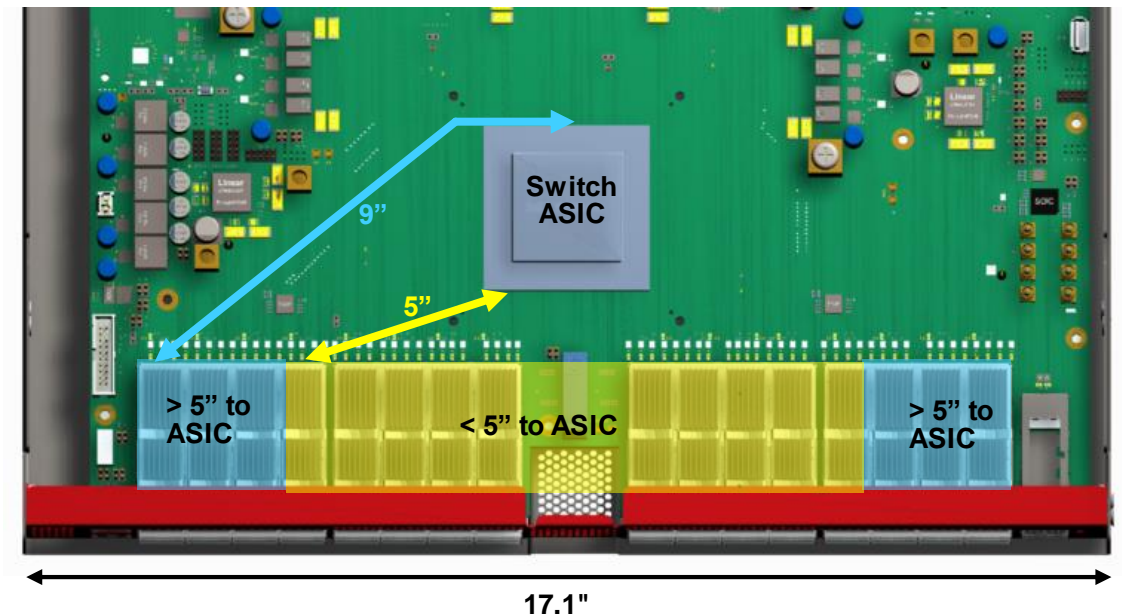
How Shorter Host Loss Maps to Possible Solutions

| Add retimers	Intra-box cables	Multi-ASIC Linecards (Chassis Systems)	MR Capable Modules
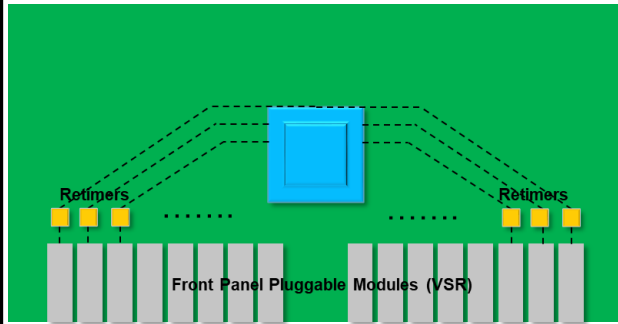
Add retimers
- Middle ports within proposed VSR budget do not require additional retimers
- Edge ports use additional retimers (shown in yellow) to enable longer overall host channels
- **Pros:** similar architecture to prior generation systems
- **Cons**: Cost and power of additional retimers
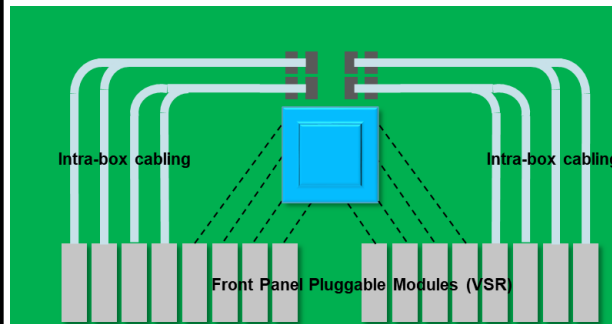
Intra-box cables
- Edge channels use intra-box cables to enable longer physical reach, but staying within proposed VSR budgets
- **Pros:** System does not incur cost or power of additional retimers, commonality with existing "PHYless" designs
- **Cons**: Increases mechanical complexity, may impact airflow, cost of cable and associated mechanicals
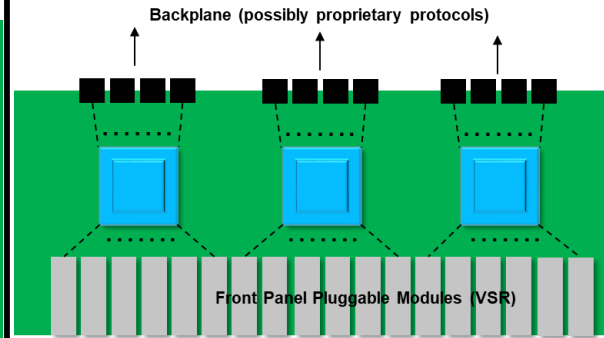
**Multi-ASIC Linecards (Chassis Systems)**
- Each ASIC can connect to fewer, closer module ports, which are supported within VSR proposed budget
- **Pros:** Similar "PHYless" design to current generation systems
- **Cons:** Does not address single ASIC "fixed box" designs forecast to be the dominant volume of the datacenter market
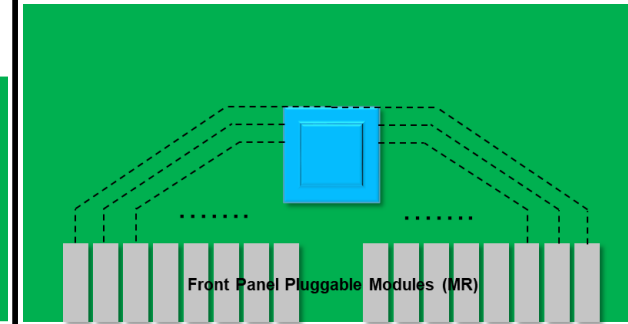
MR Capable Modules
- Enable modules with MR capability
- **Pros:** Similar "PHYless" design to current generation systems
- **Cons:** Requires MR support in modules, potentially increasing module power. Serdes may require training, and appropriate management support.

BROADCOM®

Datacenter Switch Market Architectural Forecast
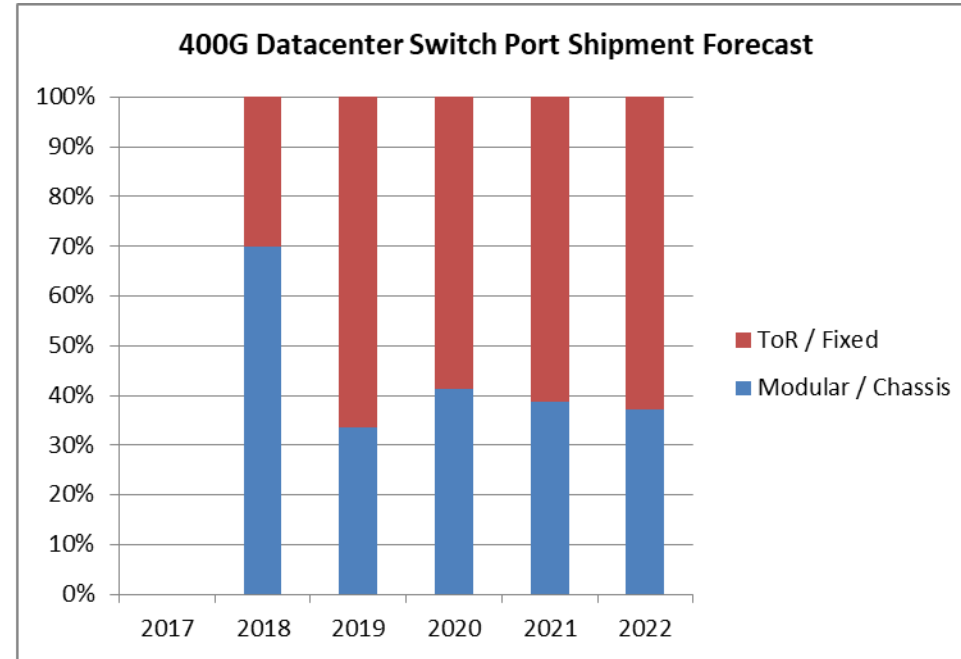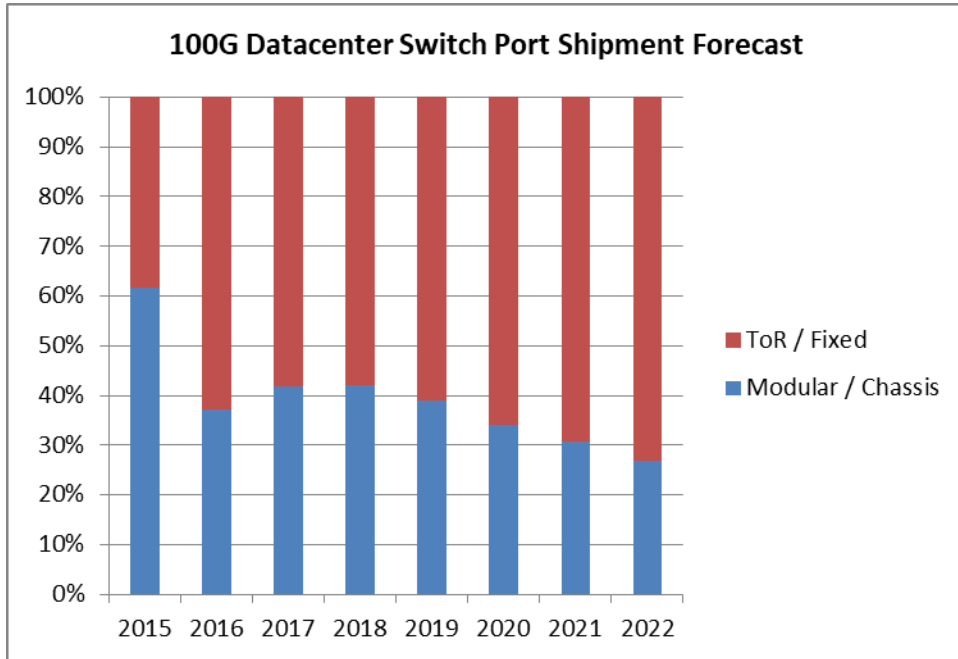
- Aggregate port shipment data presented for 100G and 400G[1]

- Datacenter switches are forecast to migrate from chassis to majority fixed / ToRs

- Important to find a low power, cost effective solution which supports the majority of this market!
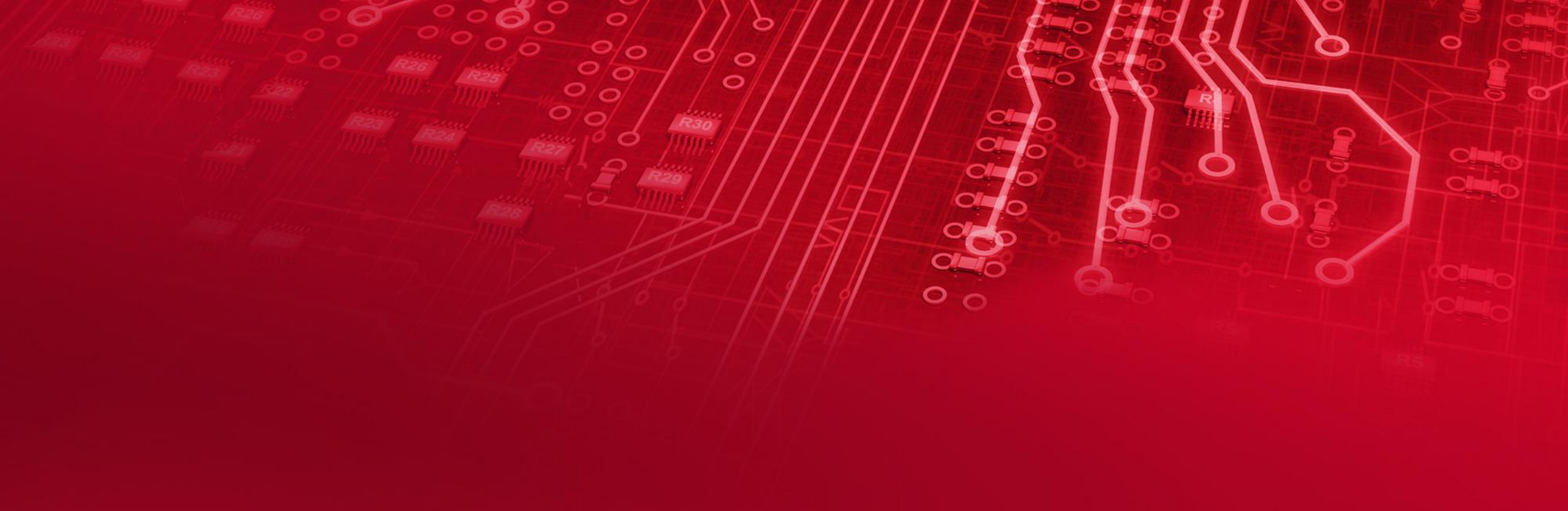
[1]Source: *January 2018 CREHAN Long-range Forecast Data Center Switch* (used with kind permission)

BROADCOM®

Possible Additional Work Needed for Different Architectures

- Retimers – nothing (VSR budget as proposed appears OK for this approach)
- Cables – physical reach – will this enable PHYless systems?
  - Asymmetric is an option for endpoints connections perhaps, but not applicable for single ASIC switch – switch DAC links
- Multi-ASIC Chassis Linecards – nothing (VSR budget as proposed)
- MR based modules
  - Training required? In-band or out-of band?
  - Additional management complexity?
  - Increase in module power?
- Cost and power models?
  - Always a bit of a minefield..

BROADCOM®

Summary

- Short host channels have implications on single ASIC solutions which are forecast to be the dominant part of the datacenter deployments at 100 and 400GE

- Economic and system trade-offs need to be understood to ensure we develop a viable solution for this market segment

BROADCOM®

Thank You

BROADCOM®