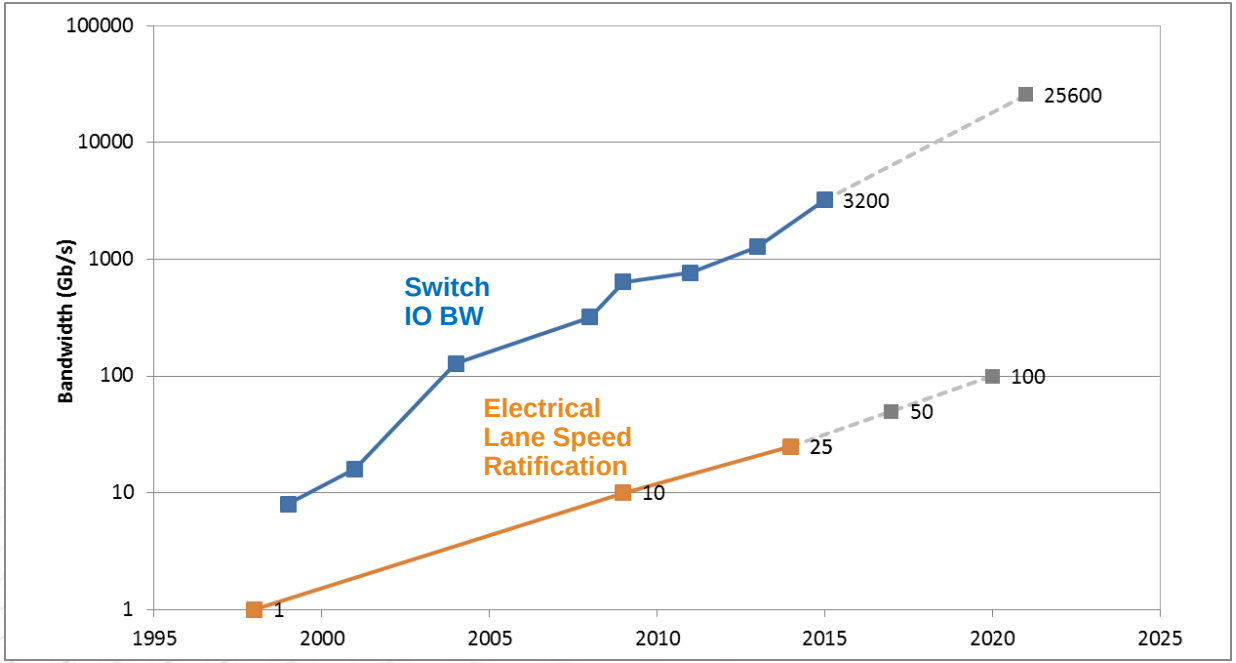
THE RATE DEBATE: SWITCH PERSPECTIVE

Rob Stone

DRIVERS FOR ADOPTION OF NEW RATES IN DATACENTERS

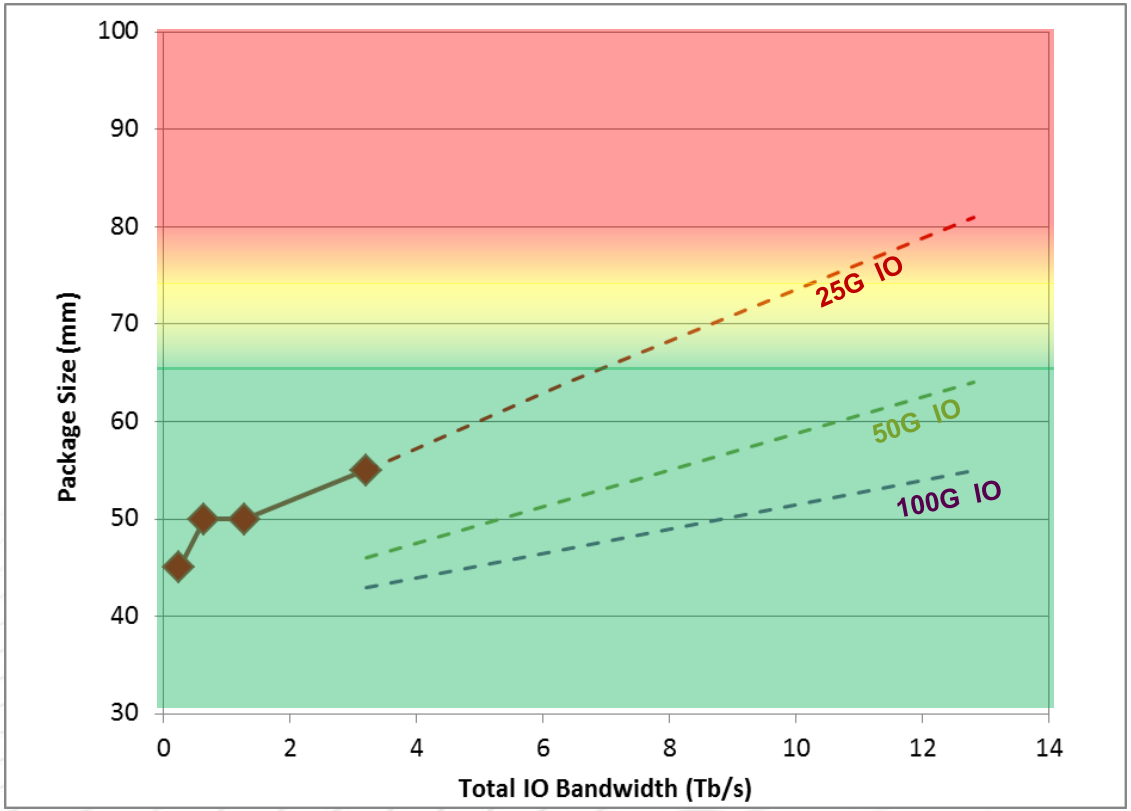
- **Package escape**
  - Practical limitations of BGA number of balls, and PCB wiring
- **Box escape**
  - Trace loss limitations
  - Front panel size constraint for pluggable interfaces
- **Power efficient**
  - Eliminate redundant gearboxing and retiming
- **Maximize switch radix**
  - Don’t strand IO (use all the available lanes)
- **Economic**
  - Less lanes is *eventually* less expensive

ETHERNET SWITCH HISTORY



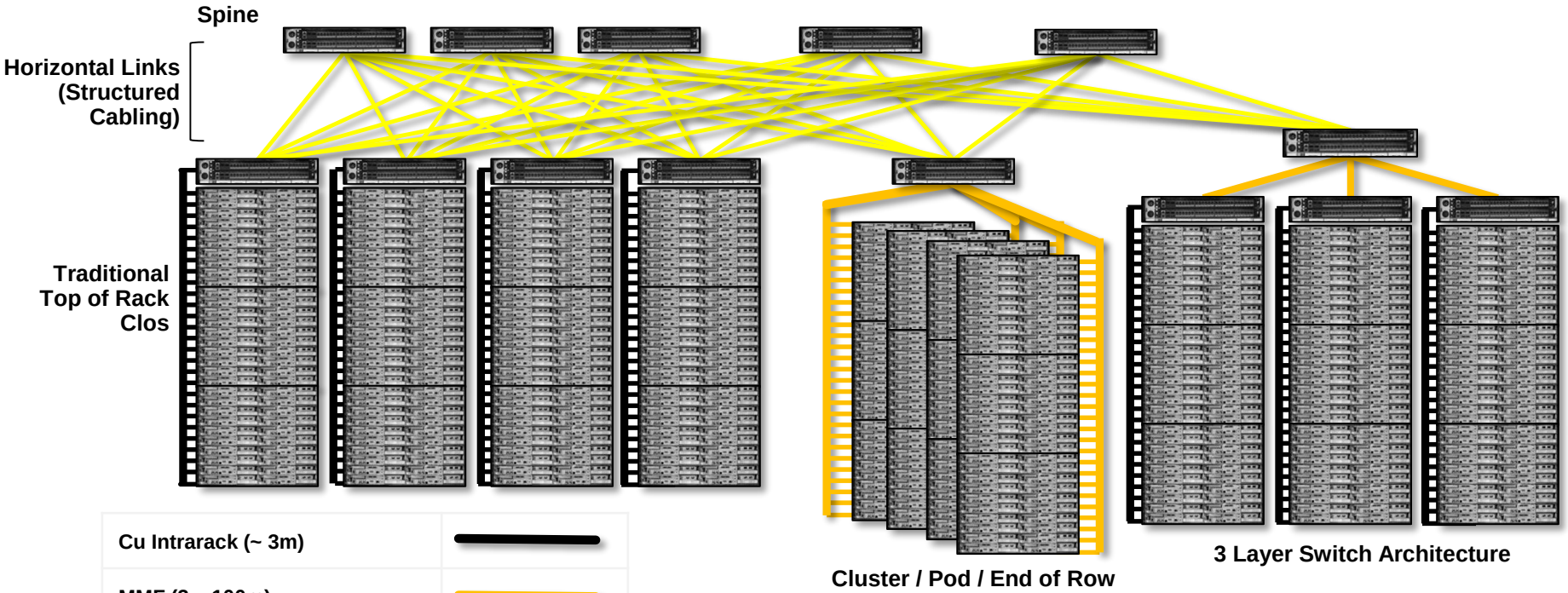
- **Will Moore's law for switching continue?**
  - New tools: HOM, 16nm, 10nm…
- **Disclaimer: Broadcom will continue to develop higher rate switches, but this is not a timeline for product availability, or specific configuration**

SWITCH PACKAGE ESCAPE – EXTRAPOLATION FROM TODAY

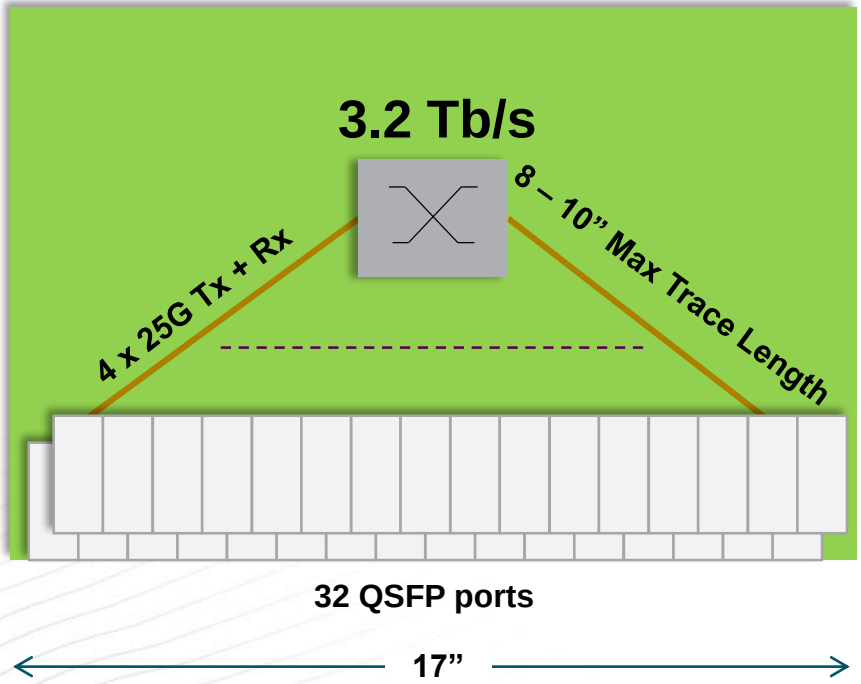


- **(Assuming 1mm ball pitch)**
- **Ball # limitations motivates migration to higher lane rates**
  - (Moves bottleneck downstream!)
- **Smaller packages are enabled by higher speed more complex IO**

DATACENTER INTERCONNECTS – UNIVERSAL SWITCH DESIRED

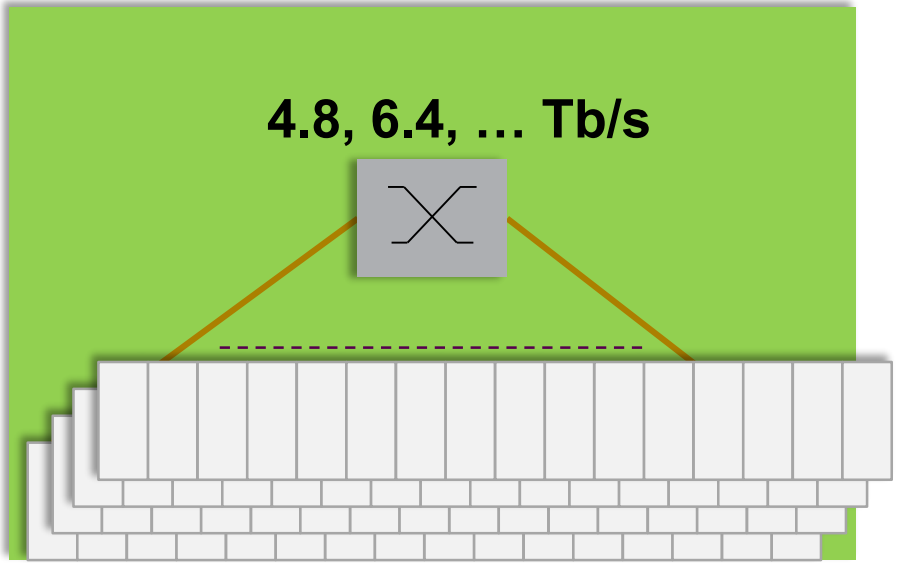


IDEALIZED SWITCH DESIGN AND CHALLENGES



- **Universal switch, many compatible interface types**
- **Enables homogeneity – can be used as ToR, Spine…**

IDEALIZED SWITCH DESIGN AND CHALLENGES (2X BW)



- **>32 QSFP28 Ports won't fit in 1RU**
- **Potential Solutions:**
  - Higher density pluggable module
  - Mid-board optics
  - Larger Box

- Limitation of module connector IO # lanes still exists
- Higher lane speeds → shorter electrical board reach
- Not front panel size limited, but less universal design
- 2RU → May need additional retimers added

FUTURE IO (PERSONAL PREDICTIONS)

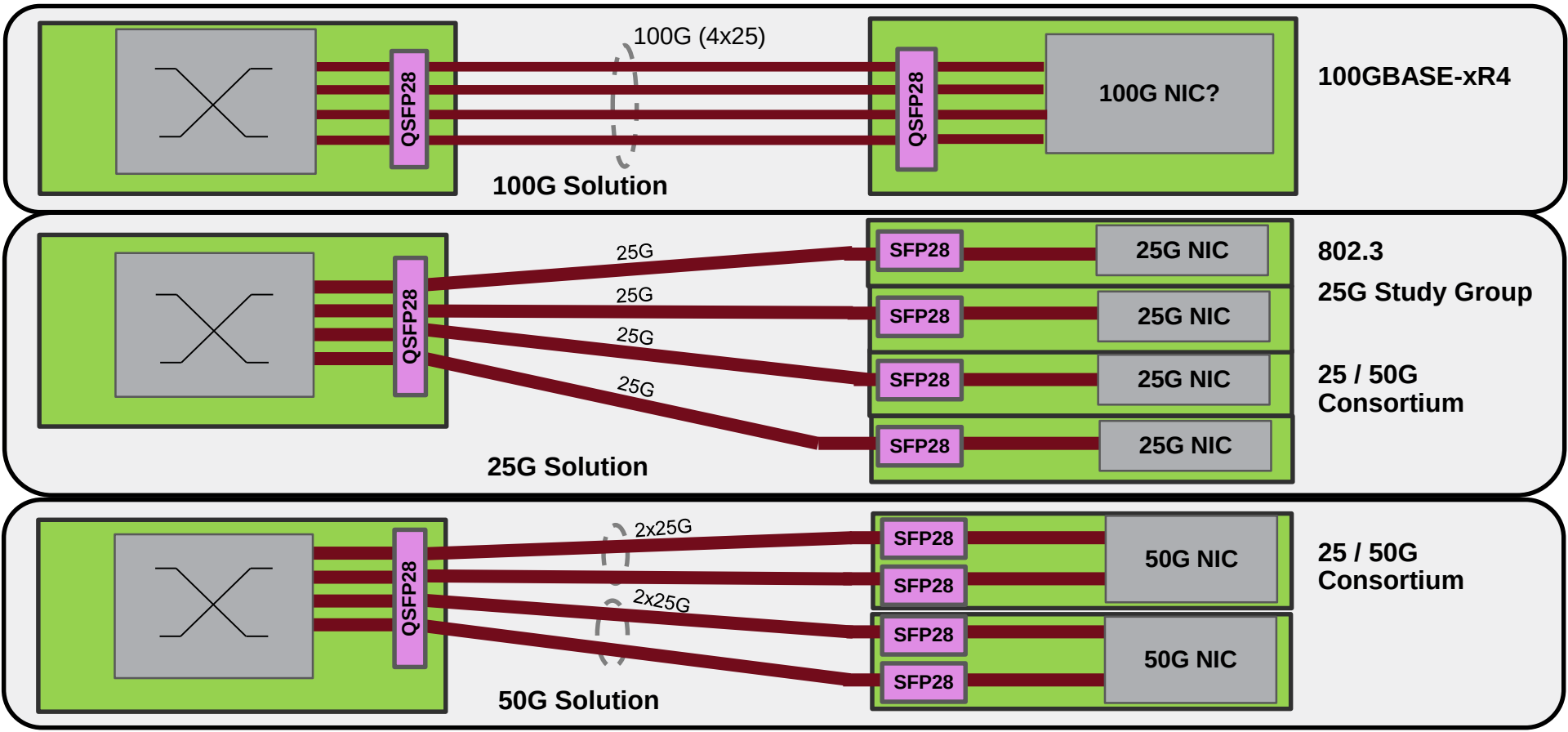
Total IO BW (Gb/s)	Lane Speed (Gb/s)	Symbol Rate (Gbd)	Front Panel Pluggable	Mid-board OE	Integrated Optics
3.2 Tb/s	25	25	1 RU	1 RU	X

Decreasing accuracy Axis

- **Required Solutions**
  - Higher lane speeds
  - Better cooling architectures
  - Higher density pluggables
  - Adoption of board mount optics
  - Integrated optics (2.5D / 3D)

Increasing Pain Axis

BREAKOUT TOR – SERVER – OPTIMAL LANE USE AT 100G



BREAKOUT AND HOMOGENEITY FOR HIGHER RATE PORTS

Native Switch Port Speed	# Lanes	Lane Speed (Gb/s)	Native Supported Modules	Breakout Options (no gearbox)	Breakout Solution Availability
100 GbE	4	25	100GBASE-CR4, SR4, LR4, PSM4…	25GbE 50GbE (2x25)	**25G TF** **25 / 50G Consortium**
400 GbE	16	25	See 802.3bs	16 x 25GbE (1x25) 8 x 50GbE (2x25) 4 x 100GbE (4x25)	**25G TF** **25 / 50G Consortium** **100GBASE-x**
	8	50	(In process)	8 x 50GbE (1x50) 4 x 100GbE (2x50)	**TBD** **TBD**

- **Need a flexible breakout capability for switch - server interconnects**
- **Goal: universal switch design for homogeneity within datacenters**
- **Maximize switch utilization – don't strand IO**
- **FEC implementations need to support breakout cases too!**

CONCLUSION

- **Increased lane rates are required to continue to drive higher IO density**
  - Switches & modules
- **Infrastructure homogeneity is a goal for management of large scale datacenters**
- **Efficient server breakout is a motivator for adoption of lower speed rates (25 / 50GbE)**
  - Needs to be cognizant of switch lane rates to minimize gearboxing
- **Challenge to the community: We need flexible, efficient end – end architectures**
  - Breakout capability (FEC / Lanes)
  - Higher IO density PHYs (Pluggable Modules / Mid-board, Copper & Optical)

Thanks!