



# 10x10 for Warehouse Scale Computing

Bikash Koley, Vijay Vusirikala, Ashby Armistead

Jan 25, 2011

10x10 MSA Meeting

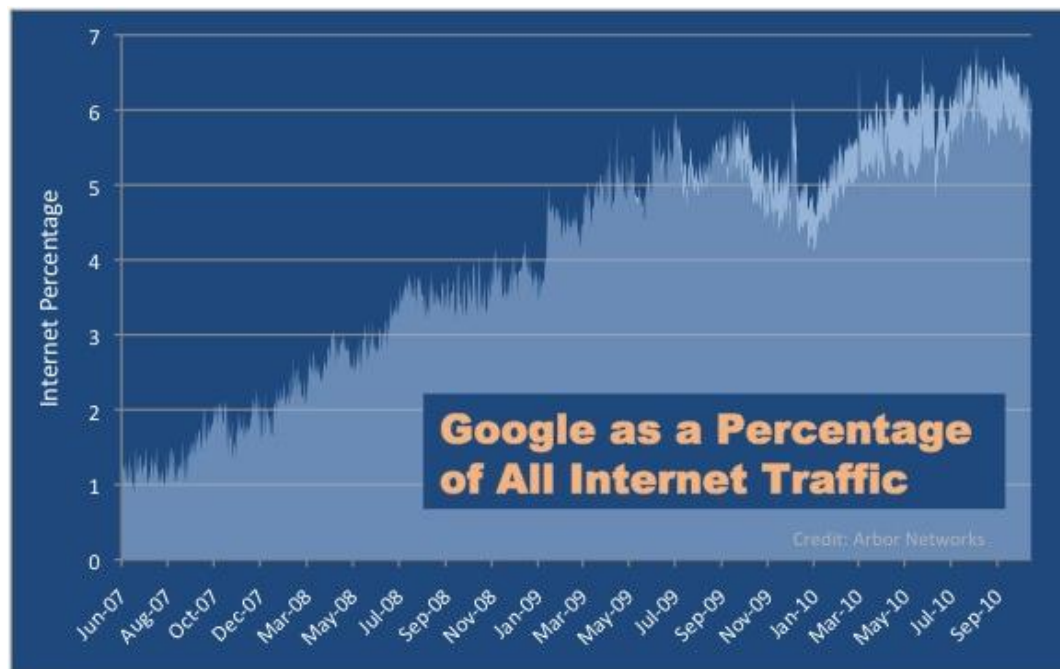
## The “ATLAS” Report

Posted on Monday, October 25th, 2010 | Bookmark on [del.icio.us](#)

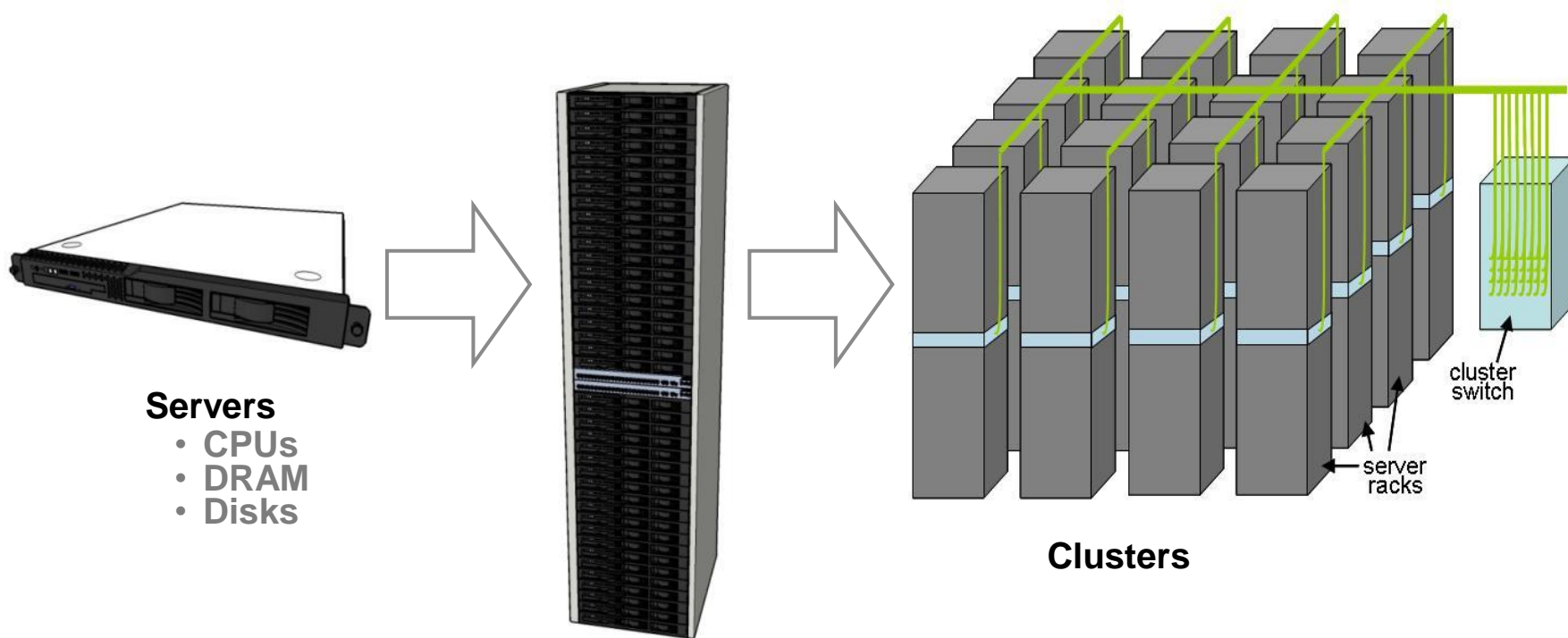
### Google Sets New Internet Traffic Record

by Craig Labovitz

This month, Google broke an equally impressive Internet traffic record — gaining more than 1% of all Internet traffic share since January. If Google were an ISP, as of this month it would rank as the **second largest** carrier on the planet.



# Warehouse Scale Computer: The Machinery



## Servers

- CPUs
- DRAM
- Disks

## Racks

- 40-80 servers
- Ethernet switch

## Clusters

# Why Adopt a Technology?

- **Utility**
  - Does it solve my problem?
- **Interoperability**
  - Can I combine vendors + products to solve my problems?
- **Independence**
  - Can I be totally free of a solution lock-in?

# Why Adopt a Technology?

- **Some technology adoptions need not go past utility**
  - 10GBASE-ZR
  - DWDM-XFP
  - SFP+
- **Utility drives Rapid Adoption**
  - FEC in transport preceded any OTN definition by several years
  - **100BASE-LX10** SFPs were widely used much before 802.3ah became a standard (because 100BASE-FX wasn't optimal)
- **Some technologies cannot achieve UTILITY without INTEROP**
  - TCP/IP
  - IEEE 802.3

# Lessons learnt from 10GbE

- Standard looking for **UTILITY**: Does any one remember 10GBASE-LX4? 4-lane 1310nm WDM over 10km of SMF. Except not very cost-effective!
- Proliferation of optical interfaces/ form-factors:

XENPAK  
X2 XPAK  
XFP  
SFP+

- Most network operators are still sparing 3 to 5 different types of 10G modules even after 5 generation of module form-factor iteration
- This is expensive!

# Issues with adopting the current 100GbE standardized solutions in WSC

- **100GBASE-LR4:**

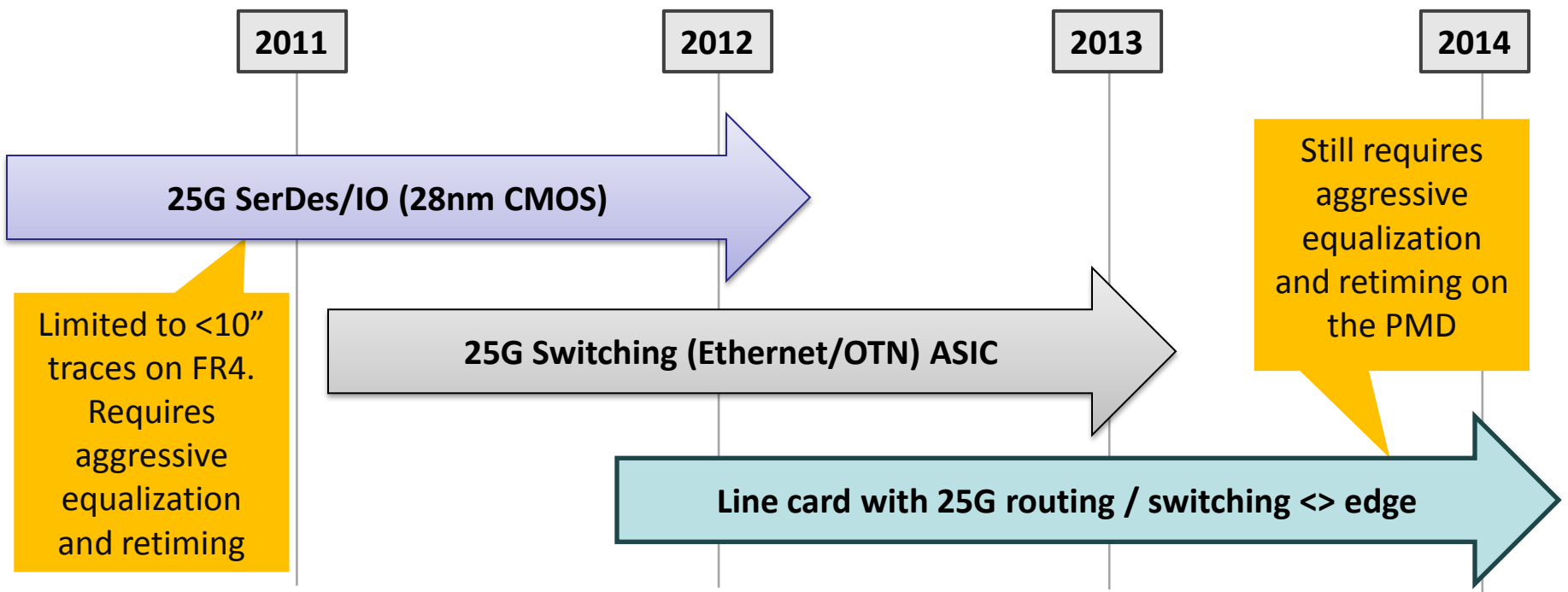
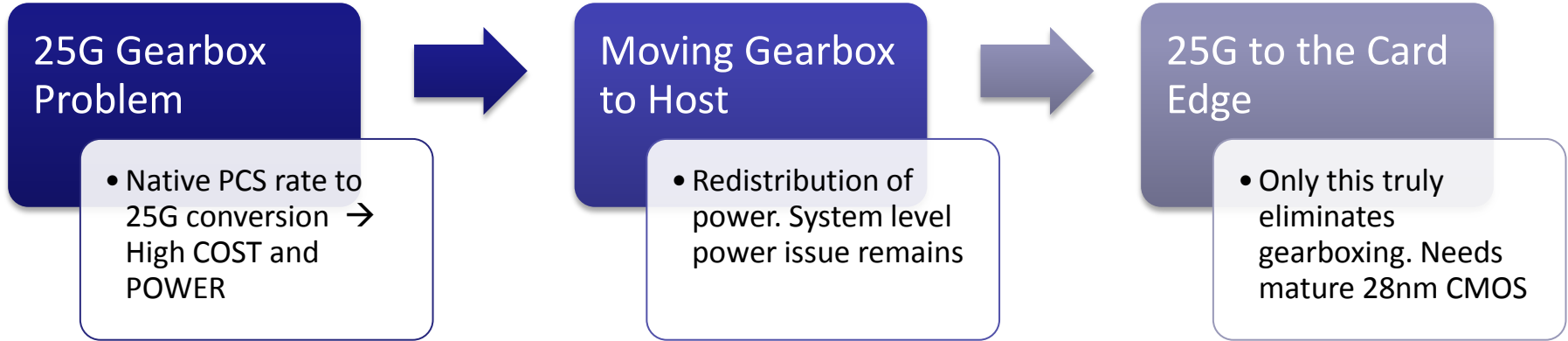
- Too expensive today
- Too power hungry (20W+, need net reduction compared to 10W for 10x SFP+)
- Does not support non-retimed interface for low-cost high-density WSC interconnect fabric (where is my SFP+ equivalent?)
- 0-2km is sufficient for intra-DC, need 40km+ for inter-DC: Won't pay the price/power premium to go 10km when all we need is 2km!

- **100GBASE-SR10**

- 100m reach is insufficient for WSC fabric
- Ribbon fibers are incompatible with large WSC fabrics: hard to deploy, manage, terminate etc etc

*Need a 2km low-cost, low-power SMF client interface that can also be non-retimed*

# So why can't LR4 Optics Address this?





# Addressing Key Issues

## Board Edge Density

- 25G I/O not needed for achieving board edge density
- Commercial Ethernet switches support  $48 \times 10\text{G}$  (with RXAUI - 6 Gb/s) today
- CXP (or equivalent) form-factor provides SFP+ type edge density

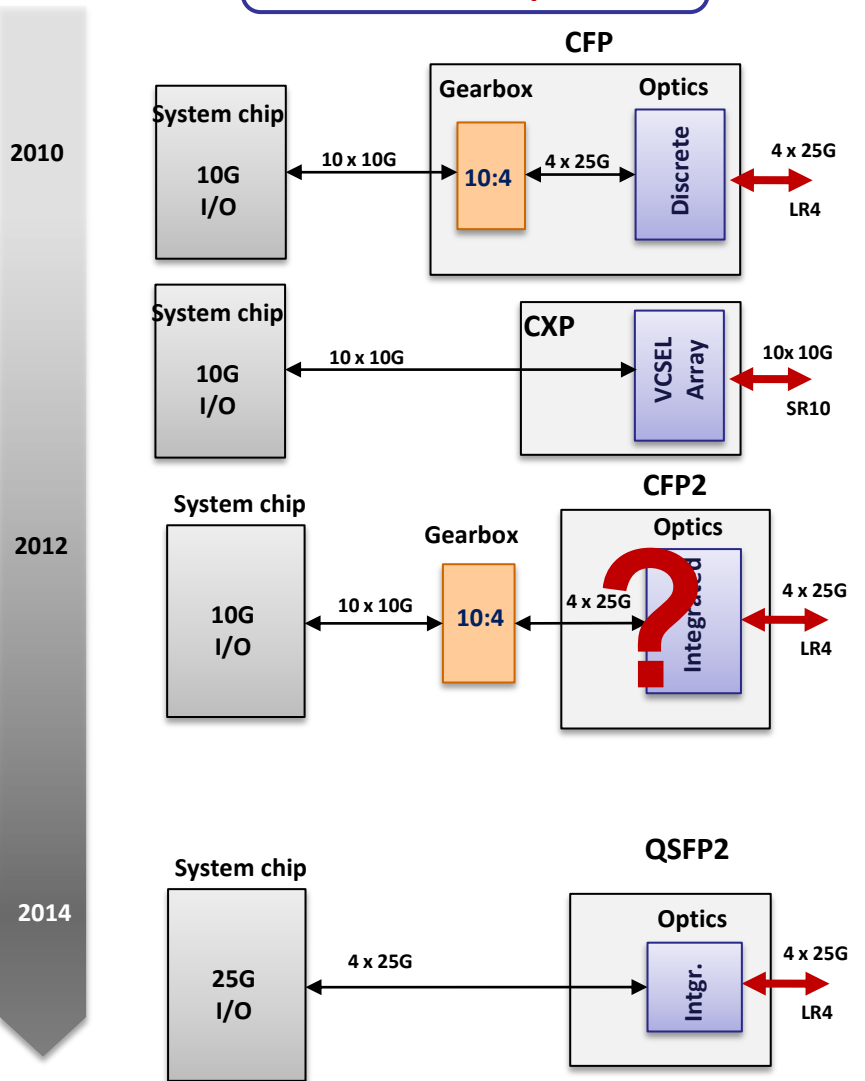
## Non-retimed Interfaces

- LR4 would require on-module CDR for next few years (even if gearbox is moved to host board)
- Power burden of CDR is significant
- 10x10G can support non-retimed interfaces for Ethernet applications (path to  $\sim 5\text{W}$  for 100G)

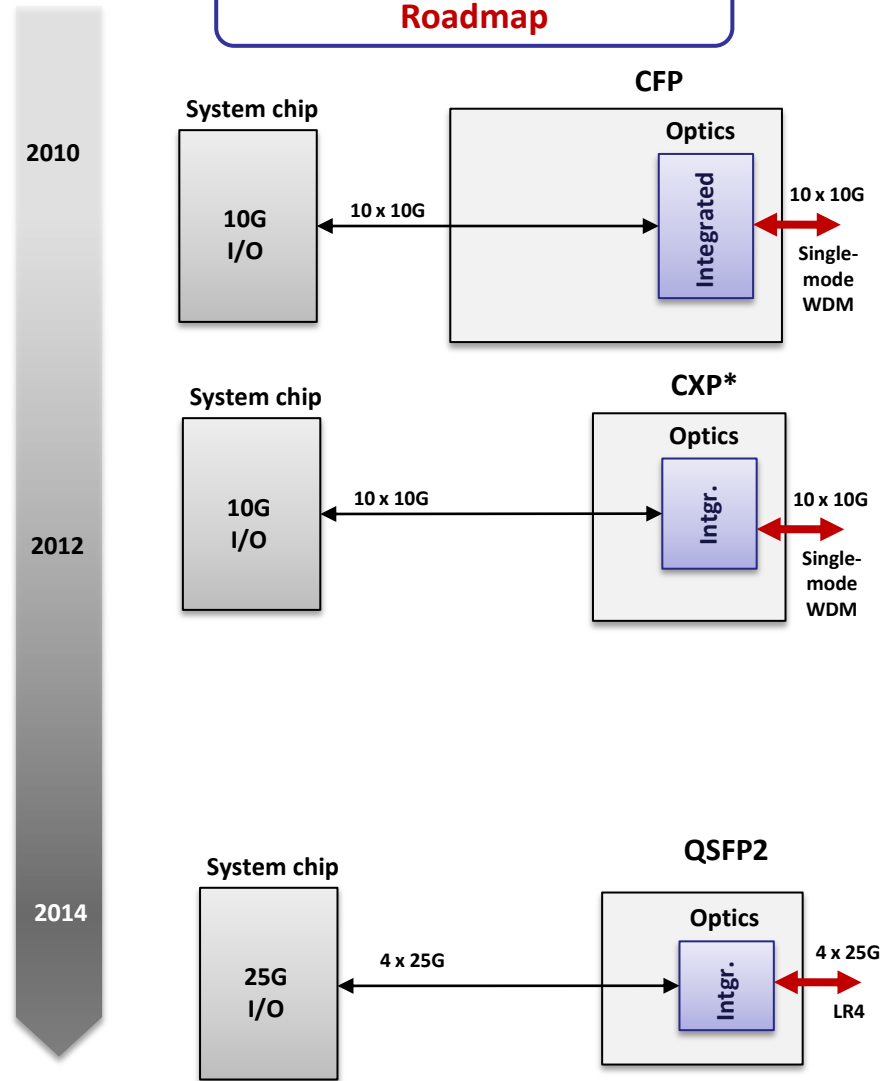
*LR4 : Standard ? (Yes), Interoperability ? (Yes) but Utility ? (No)  
No price-volume elasticity curve for next 3-4 years before true 25G ecosystem utility kicks in*

# 100G SMF Interface Evolution

## LR4 Sub-system / Module Roadmap

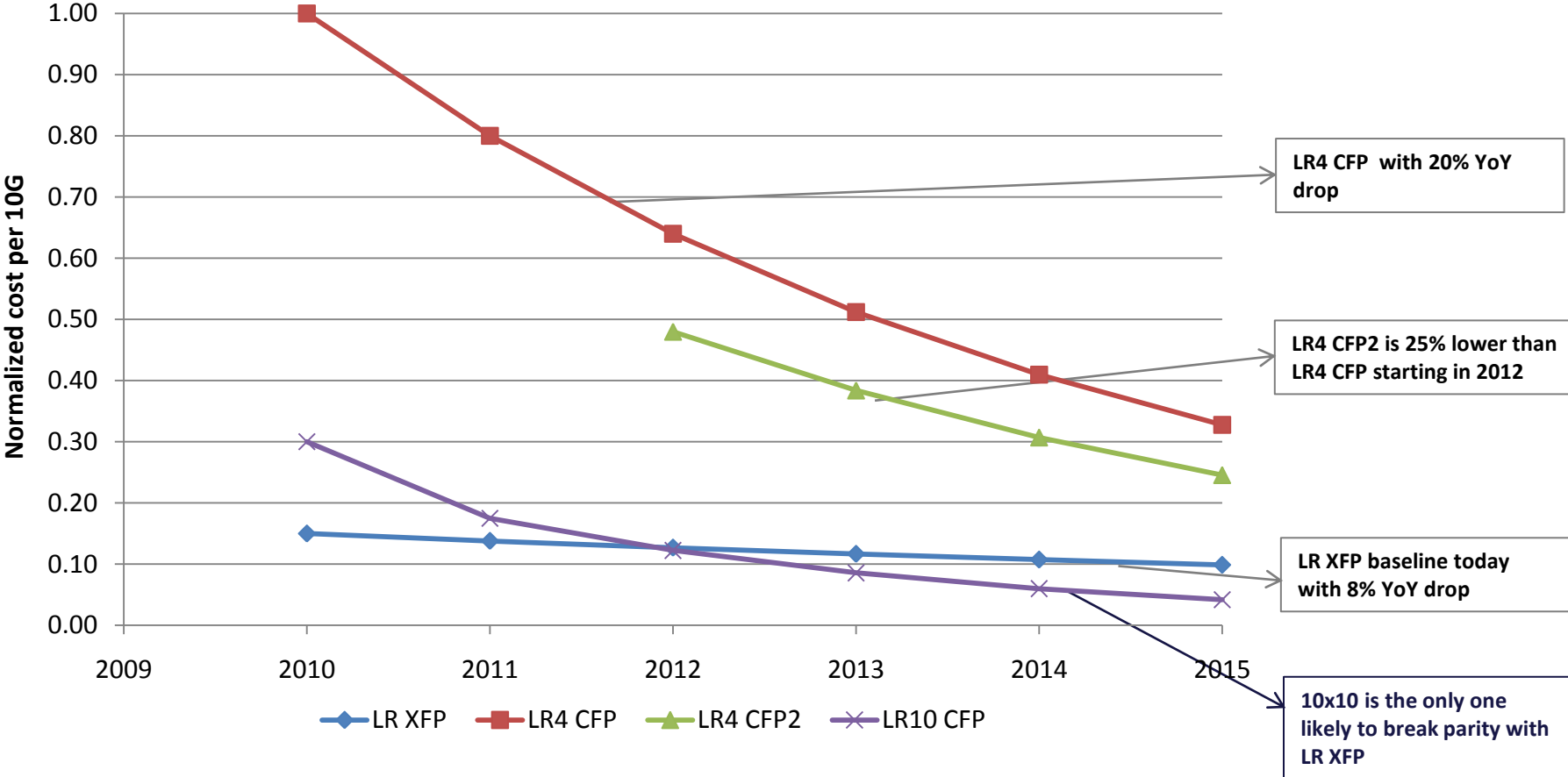


## High-volume User Aligned Roadmap



# Perspectives on Pricing - Comparison with XFP

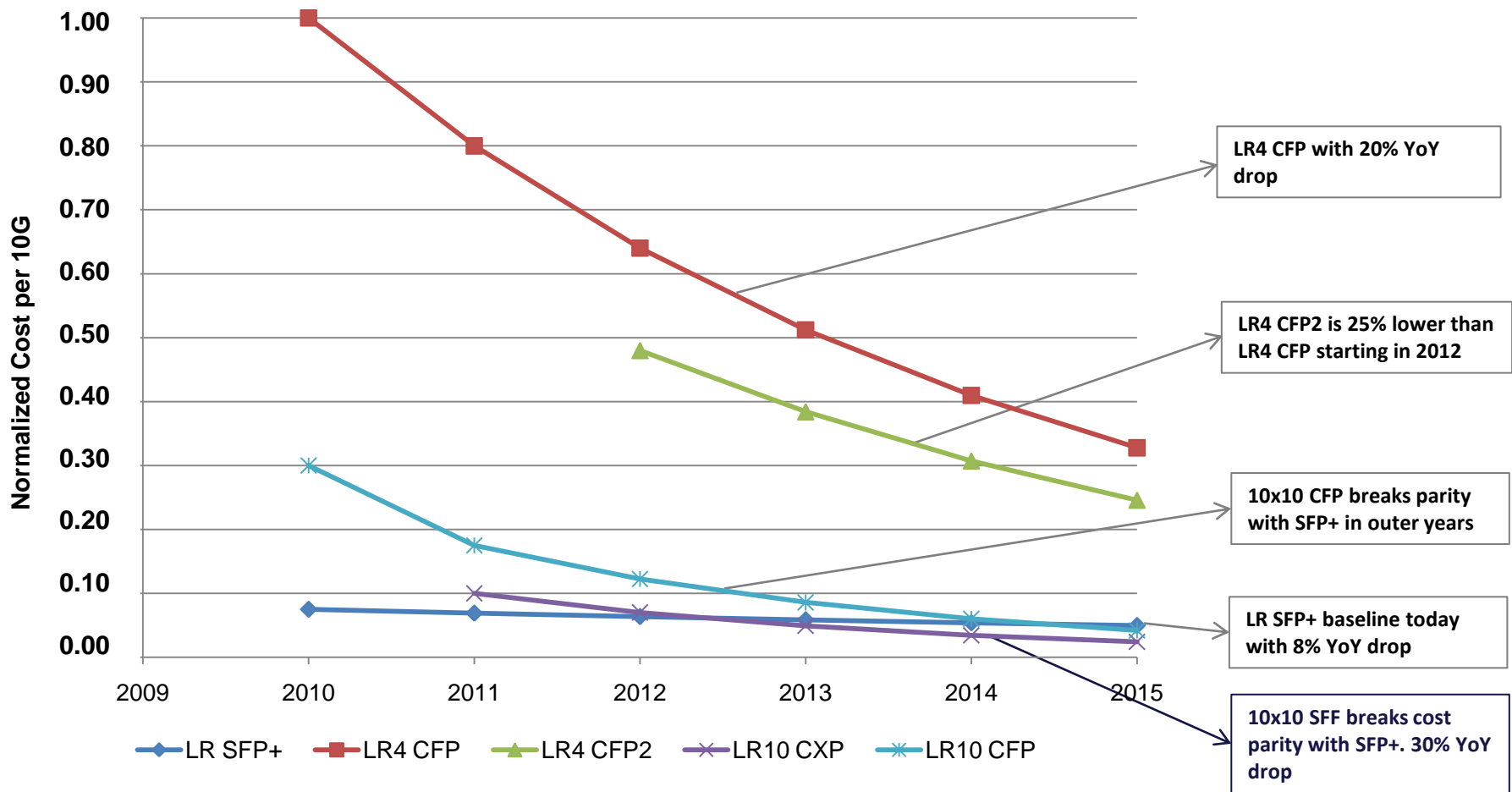
Pricing comparison for Multi-protocol support (Ethernet / OTN)  
Form-factor supports re-timing



# Perspectives on Pricing - Comparison with Google

## SFP+

Pricing comparison for Ethernet only  
Form-factor does not support re-timing



# Conclusions

- Utility + Interoperability → High Volumes
  - Price / volume elasticity and virtuous cycle
- LR4 has key fundamental issues (cost, power) for next 3-4 years
  - 25G is the right building block when switch ASIC have 25G I/O and can be routed to board edge
- 10x10 MSA leverages a mature, cost-effective 10G ecosystem
  - Directly addresses WSC community needs now
- High-density 10x10 optical modules are needed in high-volume in 2012, definition needs to be completed in 2011