The Case for Extended Reach MM Objectives

10-minute version: Talk fast, highlights only

Paul Kolesar

IEEE 802.3ba, Seoul Korea

September 2008

Outline

- **Market analysis**
- **5 Criteria examination**
- **XR Objective setting**

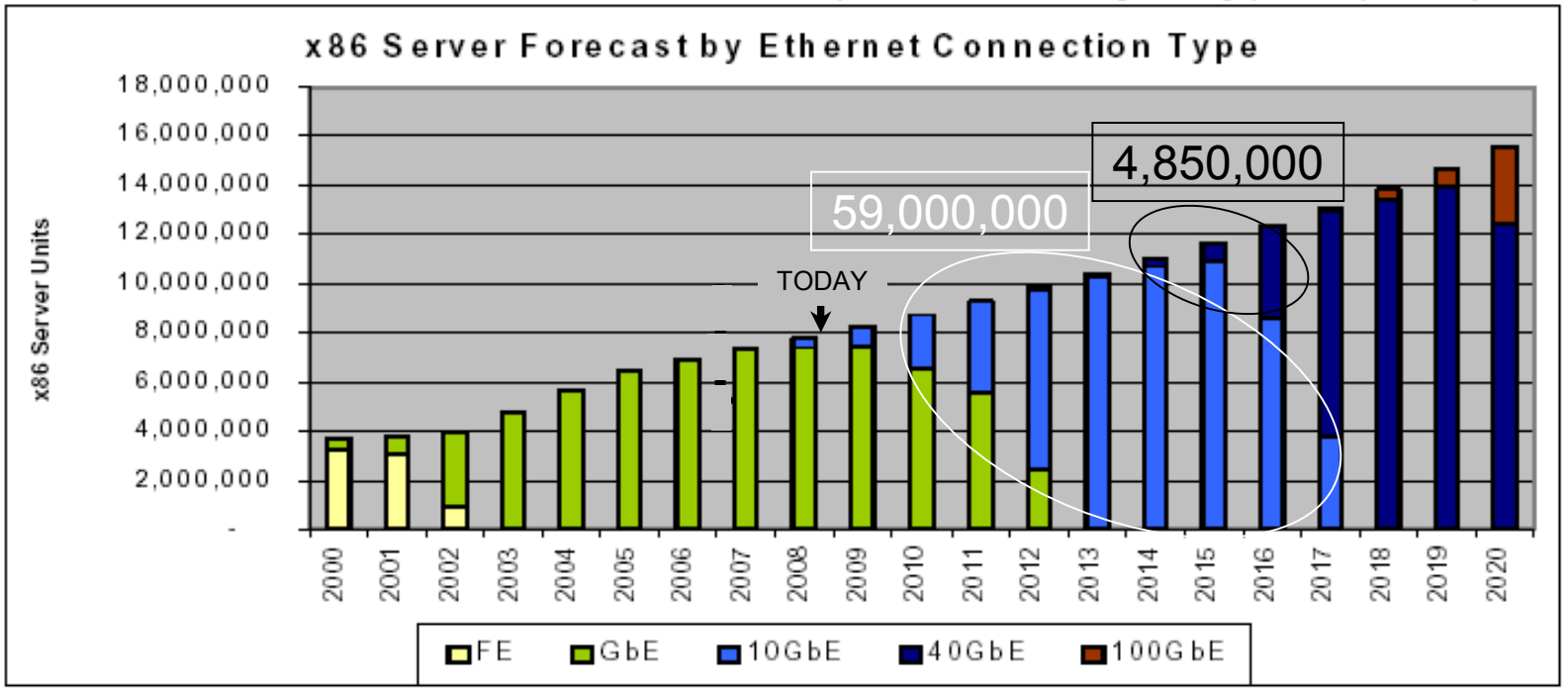
Background

- A major criticism to providing XR MM optics is that the market size does not justify another PMD type
  - Focus instead on addressing the <100m market, well suited to HPC environments and server access, at the lowest cost
- I have argued that the client-to-access (C-A) "horizontal" channels of the data center should be considered separately from the access-to-distribution (A-D) and distribution-to-core (D-C) "backbone" channels, just as we do for LANs within buildings
  - They serve different purposes
  - Their need for higher-rate services occurs at significantly different times
    - Aggregation in "backbones" years before access in "horizontal"

- This material will offer some analysis to show these differences are sufficient basis for adding XR MM optics to the standard, <u>focusing on the 100G case</u>

Server Market Evolution

from http://www.ieee802.org/3/hssg/public/apr07/hays_01_0407.pdf

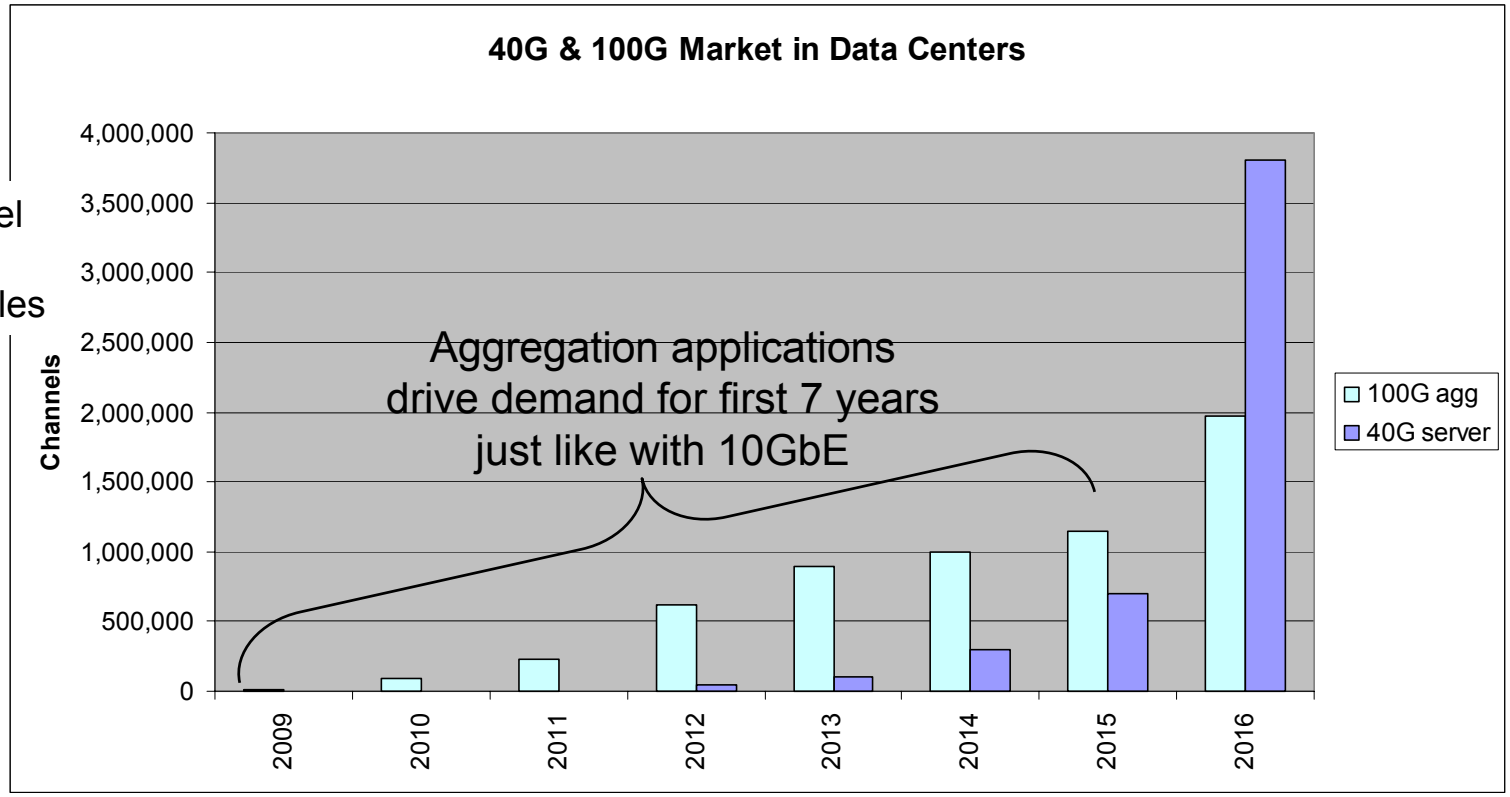


updated:
July 2007
to revise CAGR
from 10% to 6%
and presented in
HSSG_tutorial_1107.

- **10G servers dominate from 2012 thru 2016, 59 million units total**
  - **Drives potential 4.9 million 100G channels at 12:1 aggregation**
- **40G servers virtually non-existent until 2012, not dominant until 2017**
  - **This is viewed as the big market, supported by Flatman's data center survey that shows it outnumbers backbones by 13:1, but it is <u>not the primary market until 2016</u>. See next slide.**

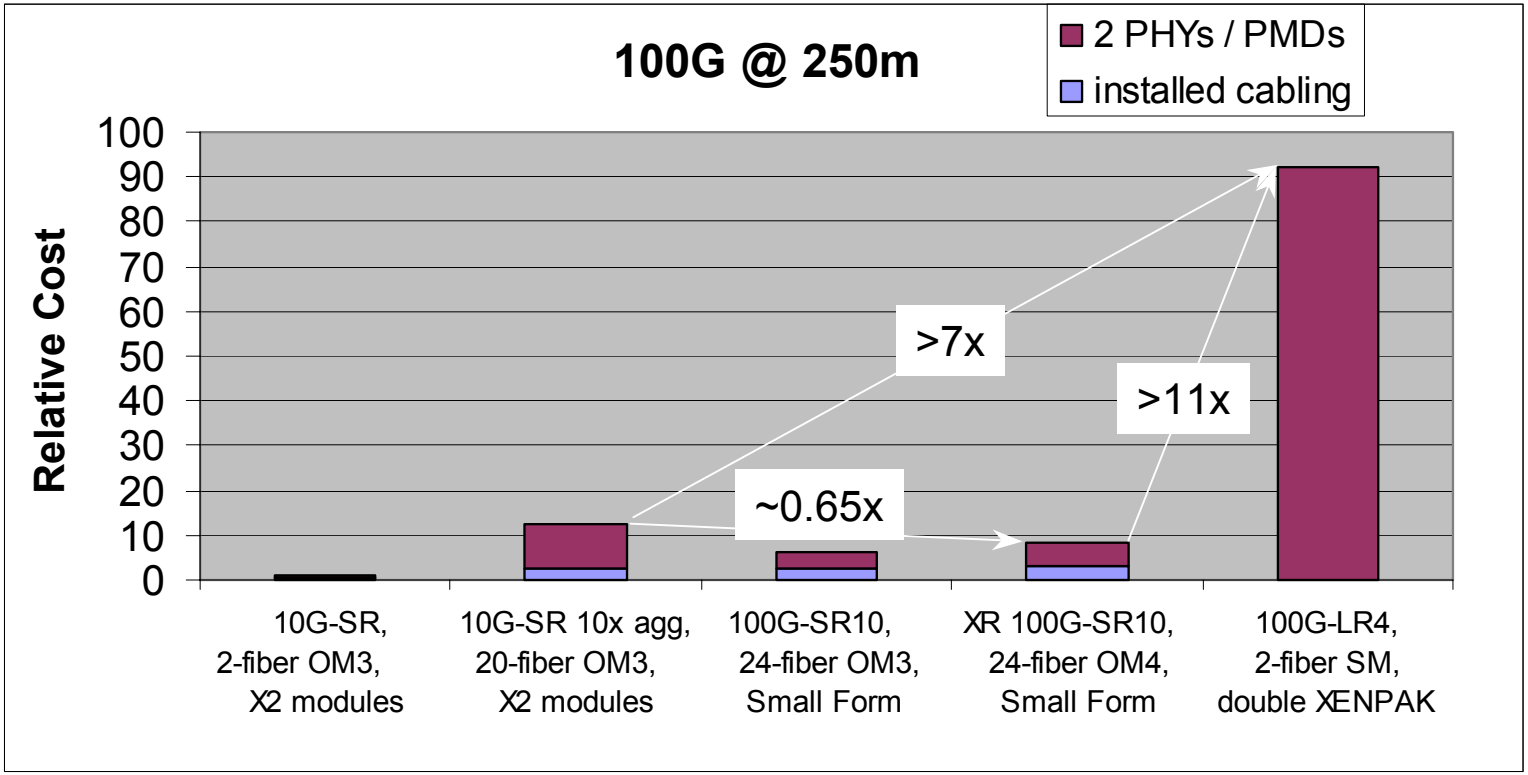
40G & 100G Market in Data Centers

Each channel requires two optics modules



- **100G market is based on aggregation of 10G and 40G server channels**
  - **Accounts for 10GbE LAG impact in early years from 2009 thru 2011**
- **XR market is ~12% of 100G aggregation market (weighted sum of A-D & D-C)**
  - **> 1,400,000 XR ports thru 2016**
  - **The XR market is similar to the total market for 10G switched ports year by year from the release of the 10GbE Standard in 2002 per barbieri_01_0107. See backup.**

100G Channel Cost Comparisons



- **SM channel is over 11 times the cost of XR 10G-SR10**
  - **Using 10x 40GBASE-LR4 cost factor for 100GBASE-LR4 PHY/PMD per cole_40_02_0208**
  - **Using 20% w.c. PMD cost premium per Jewell and 34% OM4 cabling cost premium**
- **Cost comparison does not account for module density impact on line card cost, which would amplify the illustrated differences**
  - **XR module can be physically the same size as SR 100m module ⇒ no loss of density**

Note: baseline 100G-SR10 cabling cost is calculated at 250m for like comparisons,but present baseline PMD is specified to only 100m on OM3

Economic Feasibility and Broad Market Potential Issues of Present Objectives

- **Given the cost disparity between SM and XR MM, the customer's cost to use 100G-LR4 in >100m DC channels is greater than the cost for all sub-100m 100G-SR10 backbone channels combined !!!**
- **This is neither economically feasible nor offers broad market potential. Instead it erects market barriers.**

Market Analysis Conclusions

- **The great cost disparity between SM and XR MM solutions, combined with the early need for aggregation in data centers, makes this an acute problem that must be addressed**
- **The market for XR MM aggregation channels develops years ahead of that for 40G servers and can broadly pave the way for server upgrades if a cost-effective solution is offered**
- **The market for XR MM PMDs is sufficiently large to warrant standardization and would address these issues with an interoperable solution**

5 Criteria Examination

- **Broad Market Potential**
- **Compatibility**
- **Distinct Identity**
- **Technical Feasibility**
- **Economic Feasibility**

- **Previous contribution to XR ad-hoc, kolesar_xr_01_0808, examined the existing 5 criteria statements and showed they**
  - **need no modification to include XR MM objectives**
  - **support the addition of XR objectives**
- **A condensed and enhanced Distinct Identity criteria matrix follows**

Distinct Identity Matrix

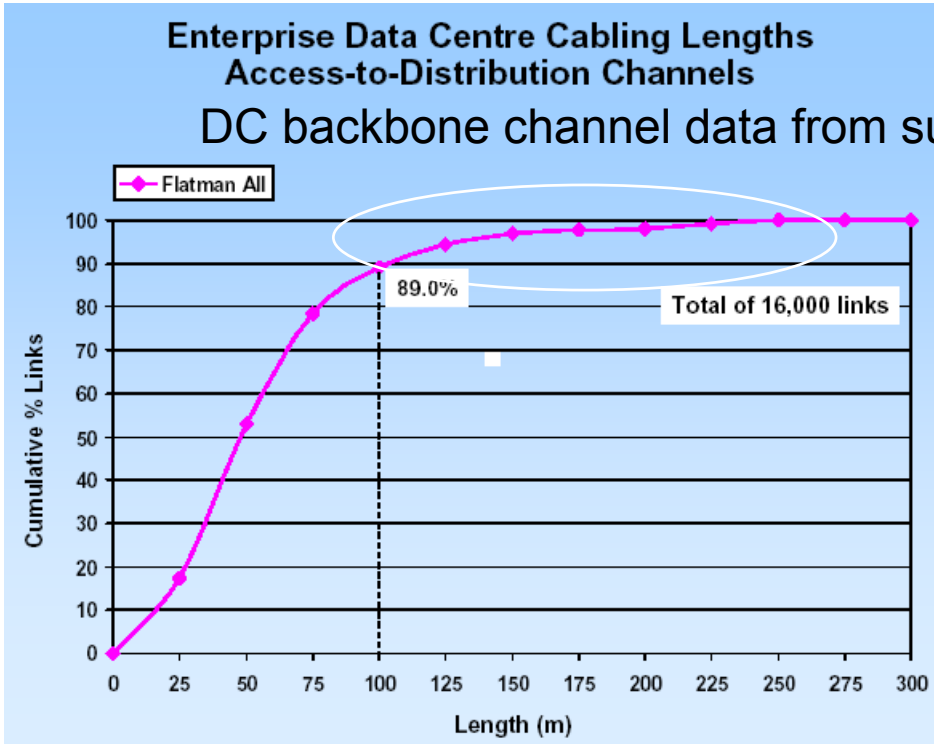
Distinct Identity Matrix			
Application	**SR4/10**	**XR SR10**	**40/100G-LR**
High Perf. Computing	from 2009		
Server/Storage Access	from 2012		
Agg. Acc.-to-Dist. in DC		from 2009	
Agg. Dist.-to-Core in DC		from 2009	
Agg/Transport in Campus			from 2009
Agg/Transport in Metro			from 2009

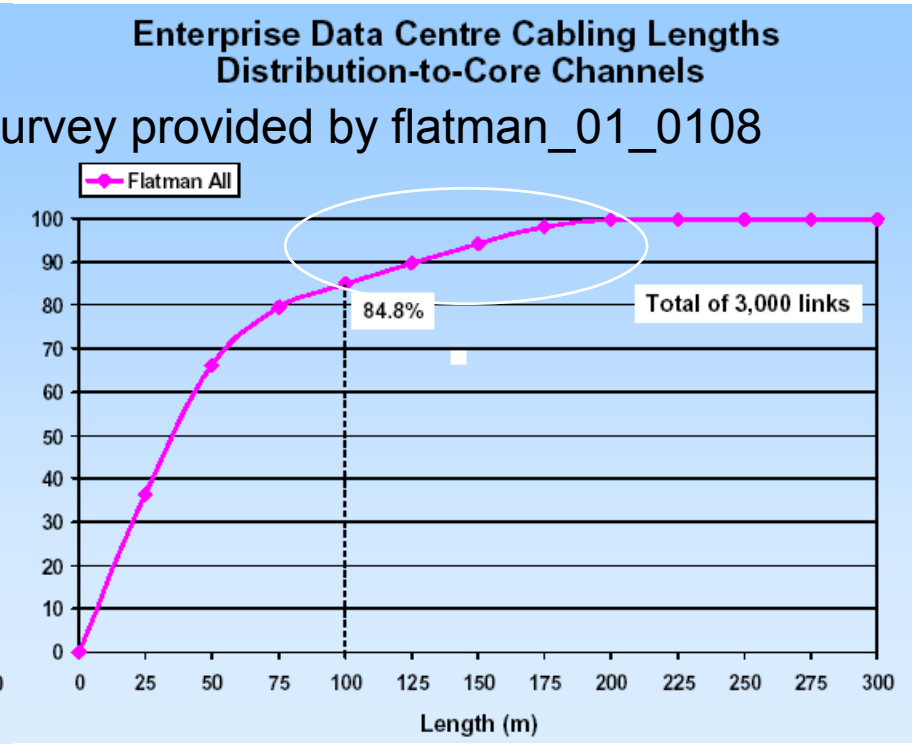
- **Each PMD optimally serves multiple distinct applications (see backup)**
- **XR distinct from LR based on optimized alignment of cost for distance served. XR can't serve metro, LR too costly for DC.**
- **XR distinct from SR based on distance/cost optimization for access vs aggregation, and market development timing**

Distilling XR MM Objective(s)

i.e. optimizing >100m DC channel coverage with XR proposal capabilities

Distance vs Coverage





DC backbone channel data from survey provided by flatman_01_0108

Coverage of >100m channels by infrastructure subsystem		
distance	access-to-distribution	distribution-to-core
150m	~70%	~60%
200m	~80%	~100%
250m	~100%	~100%

Capability of XR MM Proposals

Distance capability (two-sided)			
Proposal	**Description**	**OM3**	**OM4**
jewell_xr_01_0508	Enhanced laser spec	150m	250m
latchman_xr_01_0508	CDRs in module	208m	251m
dudek_xr_02_0708	Light weight equalization	250m	300m
petrilla_xr_01_0508	FEC in host "big chip"	183m	234m

- **All proposals support at least 150m on OM3**
  - **Covers 2/3 of channels longer than 100m in DCs**
- **Most proposals support at least 250m on OM4**
  - **250m covers virtually all channels longer than 100m in DCs**
  - **FEC distance can increase, depending on baseline spec**

New Objective(s) and Spec Approach

- **All things considered, the complete extended-reach objectives for both 40G and 100G are:**
  - **at least 150m on OM3 MMF**
  - **at least 250m on OM4 MMF (contingent on stable OM4 spec)**

- **But if only one new objective is favored by the TF and WG, then it should be the latter because**
  - **anything that meets 250m on OM4 also meets 150m on OM3**
  - **250m on OM4 is the goal of the ad-hoc (per telecon vote Sept 4)**

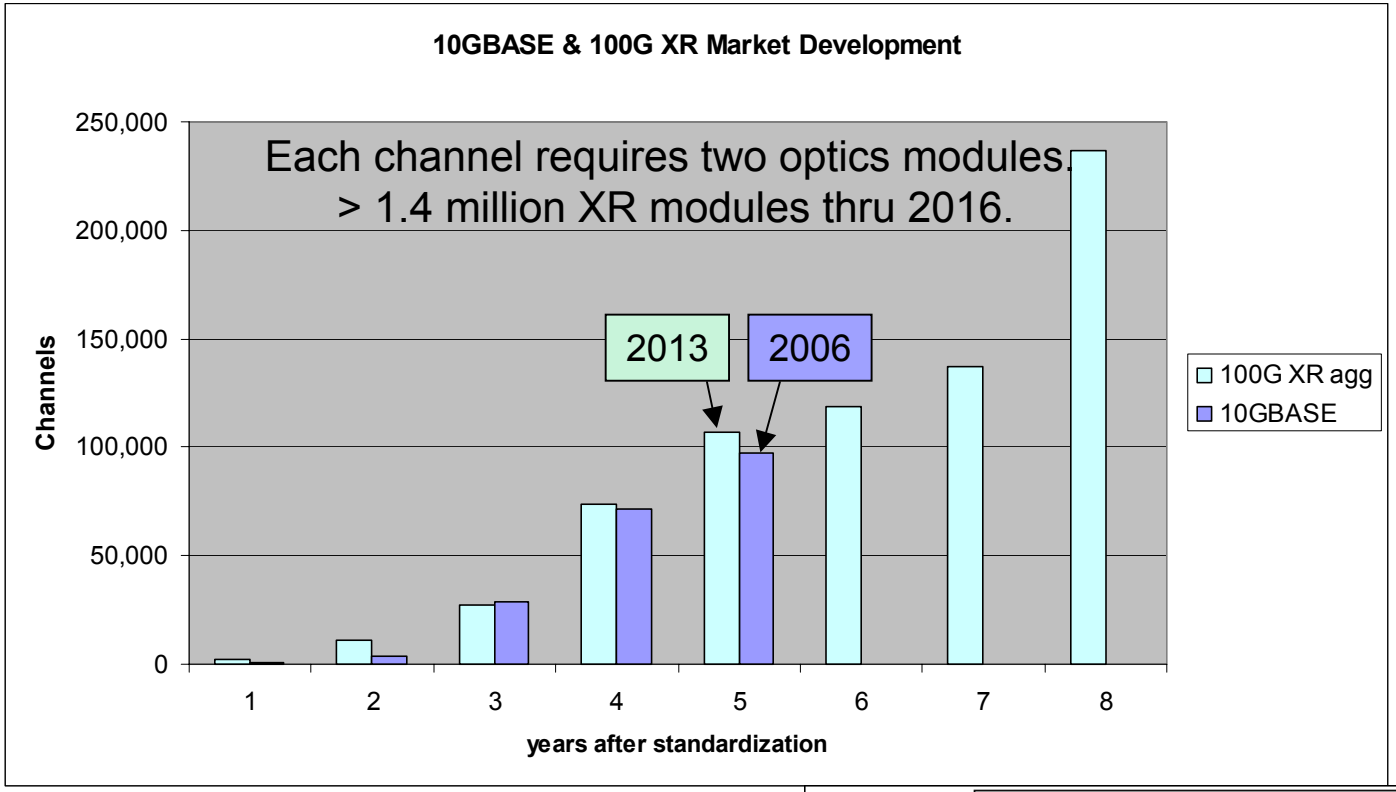
- **Specs would take the form of interface requirements**
  - **just like with the baseline proposals**
  - **not specific to particular implementation(s)**

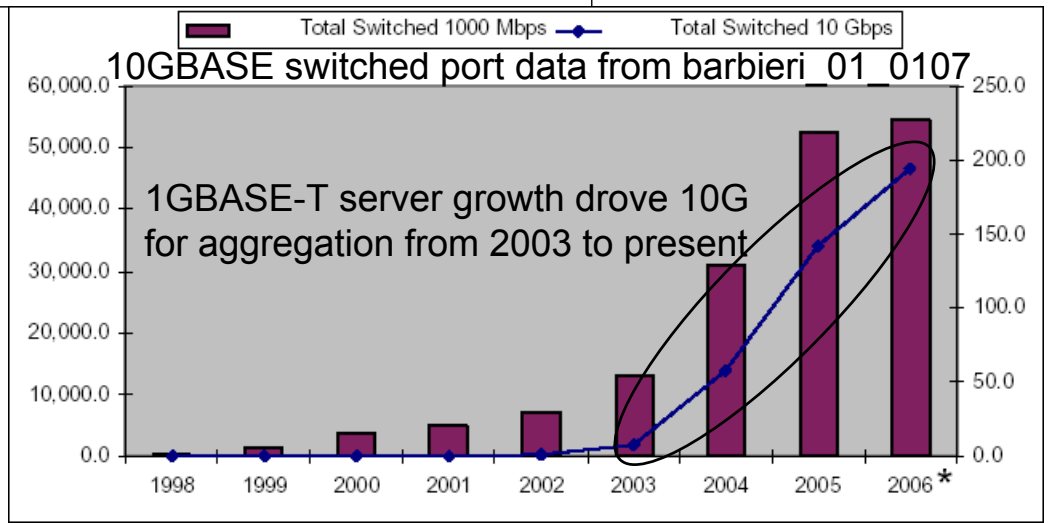
Summary

- Present course does not provide an economically feasible solution to address complete Data Centers
  - SM channel costs impose barrier to market acceptance
- To correct this issue, the minimum new 40G & 100G objective is:
  - at least 250m on OM4 MMF (contingent on stable OM4 spec)
- The addition of this XR objective meets the 5 Criteria:
  - enables Economical solution
  - Broadens Market Potential
  - retains Distinct Identity
  - is Technically Feasible in multiple ways (see backup)
  - is Compatible with previous work (see backup)
- No changes are needed to the PAR or 5 Criteria to add the new objectives
  - XR objectives are aligned with the PAR purpose and need
  - The PAR and 5C not only support, but mandate, their addition (see backup)

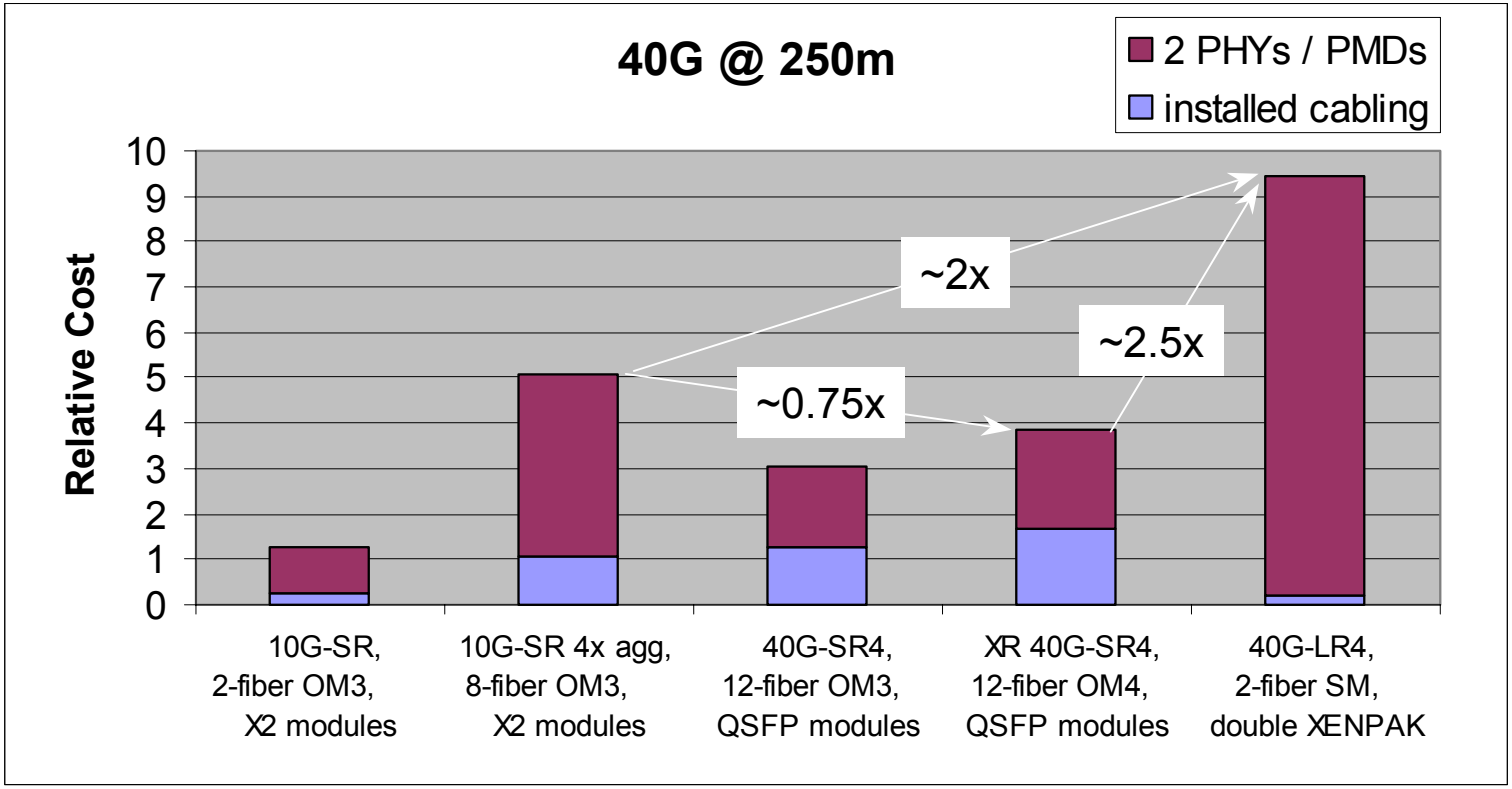
Backup Material

XR Market Development Parallels 10G





40G Channel Cost Comparisons



- **SM channel is ~2.5 times the cost of XR 40G-SR4**
  - **Using 6x 10GBASE-LR cost factor for 40GBASE-LR4 PHY/PMD per Traverso and Cole**
  - **Using 20% w.c. PMD cost premium per Jewell and 34% OM4 cabling cost premium**
- **Cost comparison does not account for module density impact on line card cost, which would amplify the illustrated differences**

Note: baseline 40G-SR4 cabling cost is calculated at 250m for like comparisons,but present baseline PMD is specified to only 100m on OM3

Consequences of Present Course

- Few customers will install 100GBASE-LR4 within Data Centers
  - Instead will choose aggregated 10GBASE for >100m channels
- High 100G SM channel costs erect substantial barriers for aggregation applications in Data Centers, causing ripple effects
  - Lack of economically advantaged “fat pipe” for aggregation within DCs will impede installation of 10G & 40G at server/storage access level
    - because DC backbones needed to support server/storage aggregation will not scale economically
  - General market penetration rate for 40G and 100G will be slower than with an economically advantaged “fat pipe” solution in DCs
    - because non-optimal DC solutions will drain funds needed for deployment of 100G in campus & metro
  - Almost no additional volume will be driven to the 100G-LR4 PMD by remaining with present course
    - because all but the customers with the most desperate aggregation problems will construct channels using multiple 10G PMDs instead
  - Increases risk of market failure
    - Incomplete economically-attractive solution becomes weakest link

Additional Cost Considerations – The Repeater Approach

- **The use of repeaters with baseline 40(100)G-SR4(10) modules could lengthen the reach between switches without incurring the cost of the switch fabric. To reach 250m, three baseline channels would be needed, tripling the number of optical modules.**
- **Some advantages (+) and disadvantages (-) of repeaters are:**
  - **(+) reuses the same module type**
  - **(+) repeaters cost less than switches**
  - **(-) requires retiming, power, space**
  - **(-) 2 to 3 times the cost of XR approach**
  - **(-) repeaters may require management interface**

An Economically Attractive Solution Enhances Broad Market Potential

- **Extended Reach MM optics on OM3 and OM4 fibers can address DC backbone distances at costs less than some 10G aggregation alternatives**
  - **Savings increase with higher aggregation rates**
- **Will drive market demand instead of thwarting it**
  - **Offers efficiencies in handling aggregated traffic while simultaneously providing cost reduction**
  - **Improvement in efficiency at lower cost is a hallmark of all successful Ethernet port types**
- **Broad Market Potential will be enhanced**
  - **Appeals to virtually any customer needing >10G rates**
  - **Accelerates market acceptance overall**
  - **Lowers risk**

Broad Market Potential and Economic Feasibility Criteria

- **The addition of the 40G ≥10km SM objective in March 2008 affected the content of only the Broad Market Potential and Economic Feasibility criteria**
  - **The phrase "including server traffic aggregation" was added to both**
- **This same rationale is driving the need for extended reach on MMF**
  - **Therefore, no additional changes are needed to these two criteria statements with the addition of XR MM objectives**
  - **These criteria support the addition of XR MM objectives**

Distinct Identity Criteria (1 of 3)

- **The 40G ≥10km SM objective was added to address server traffic aggregation**
  - **This same rationale is driving the need for extended reach on MMF, so raises question of distinctness**
- **The difference that provides distinct identity between these:**
  - **MM PMDs economically address aggregation within DCs where distances are relatively short**
  - **SM PMDs address aggregation (and transport) within campus and metro where distances are relatively long**
- **Server access and HPC applications require lowest cost PMDs, but these cover distances too short for the complete Data Center environment**
  - **Baseline MM PMDs provide lowest cost for access and HPC**
  - **XR MM PMDs optimally complete the DC space solutions by providing lowest cost for distribution and core years ahead of server access**

Distinct Identity Criteria (3 of 3)

- Requires “One unique solution per problem”
  - Previous slides show that each PMD type addresses a different problem set
  - Therefore adding XR MM PMDs complies with this criterion
- The approved Distinct Identity statements do not delve into PMD-specific items, other than justifying the need for two data rates
  - Therefore, no modifications are needed with the addition of XR MM objectives

Technical Feasibility Criteria

- Many proposals to the Extended Reach ad-hoc show there are multiple technically-feasible methods to achieve the needed reach extension
  - Jewell – enhanced laser specification (like SR)
  - Latchman – clock recovery (like SR/LR/etc.)
  - Petrilla – forward error correction (like KR)
  - Ghiasi & Dudek – simple equalization (like Fibre Channel)
- All are in practice today illustrating technical feasibility
  - The ad-hoc's efforts have concentrated on comparing these to optimize choice of distance goal(s)
  - The ad-hoc's goal is 250m on OM4 (per telecon vote September 4)
- The existing Technical Feasibility criteria statements need no modification with the addition of XR MM objectives

Compatibility Criteria

- **The current text most closely related to PMDs states:**
  - **“As was the case in previous IEEE 802.3 amendments, new physical layers specific to either 40 Gb/s or 100 Gb/s operation will be defined.”**
- **The addition of XR MM PMD objectives does not impact the existing Compatibility Criteria**

Further Objective Setting Considerations

Coverage of >100m channels by infrastructure subsystem and media			
distance	access-to-distribution	distribution-to-core	media
150m	~70%	~60%	OM3
200m	~80%	~100%	OM3/OM4*
250m	~100%	~100%	OM4

- **250m on OM4 is the design constraint (rather than 150m on OM3) for CDRs and light weight EQ**
  - **Both methods may see slightly easier implementation if OM4 distance objective is reduced below 250m**
- **Flatman survey indicates access-to-distribution channels outnumber distribution-to-core channels by more than 5 to 1 (16k to 3k)**
  - **Needs of access-to-distribution channels are primary concern**
  - **250m is needed to fully cover these channels**
- **OM4 specification must progress to a stable reference-able state**
  - **Objectives based on OM4 should be stated with that contingency**

* Media capability depends on choice of optical module interface specifications

PAR Impact (1 of 2)

- **Two statements that might be affected by objective changes are sections 5.2 regarding Scope and 5.4 regarding Purpose**
- **Present Scope statement**
  - Define 802.3 Media Access Control (MAC) parameters, physical layer specifications, and management parameters for the transfer of 802.3 format frames at 40 Gb/s and 100 Gb/s.
- **Present Purpose statement**
  - The purpose of this project is to extend the 802.3 protocol to operating speeds of 40 Gb/s and 100 Gb/s in order to provide a significant increase in bandwidth while maintaining maximum compatibility with the installed base of 802.3 interfaces, previous investment in research and development, and principles of network operation and management. The project is to provide for the interconnection of equipment satisfying the distance requirements of the intended applications.
- **Neither of these statements are affected by the addition of the proposed XR objectives**
  - **The underlined sentence supports, and mandates, the addition of the XR objectives to (optimally) satisfy the distance requirements of the intended applications**

PAR Impact (2 of 2)

- **A third statement that might be affected by objective changes is section 5.5 regarding Need**
- **Present Need statement**
  - The project is necessary to provide a solution for applications that have been demonstrated to need bandwidth beyond the existing capabilities. These include data center, internet exchanges, high performance computing and video-on-demand delivery. Network aggregation and end-station bandwidth requirements are increasing at different rates, and is recognized by the definition of two distinct speeds to serve the appropriate applications.
- **This statement is not affected by the addition of the proposed XR objectives**
  - **The underlined words support, and mandate, the addition of the XR objectives to (optimally) provide a solution for network aggregation in data centers**

- **Bottom line: No PAR changes are needed to add XR objectives**