Initial Thoughts on Architecture and PMD Objectives for 400GE

Gary Nicholl – Cisco
IEEE 802.3 400GSG, May 17, Victoria, Canada

1

Topics

- Applications

- Architecture Considerations

- PMD Considerations

- One Final Thought

- Summary

Applications

- Application space evolves over the life-cycle of a given Ethernet data rate

- Applications typically start at the core of the network (lower volume, strategic importance) and migrate to end user / compute (high volume, commodity) over time

- Architecture must support all applications over full life-cycle

- PMD requirements evolve with the application space
  - Initial PMDs only need to address initial applications
  - Don't need to (and shouldn't) define PMDs for applications that are 4+ years out (post ratification)
  - Need to be honest about the initial application space

3

Initial Application Space for 400GE

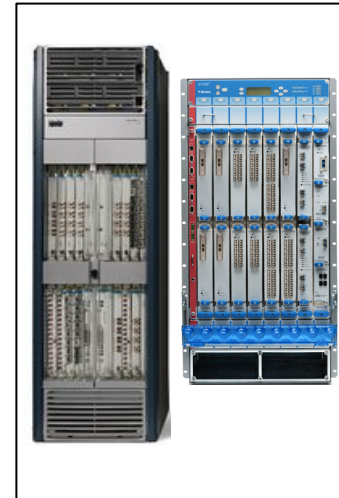Massive Scale Data Center (MSDC) Example
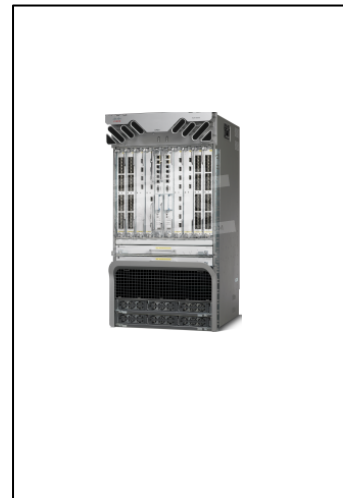
Servers	TOR Switches	Agg Switches	Routers	Transport

| 100,000 | 2000 | 100's (# Boxes) | <10 | 1 |

Photos courtesy of Cisco and Juniper

Initial Application Space for 400GE

Massive Scale Data Center (MSDC) Example
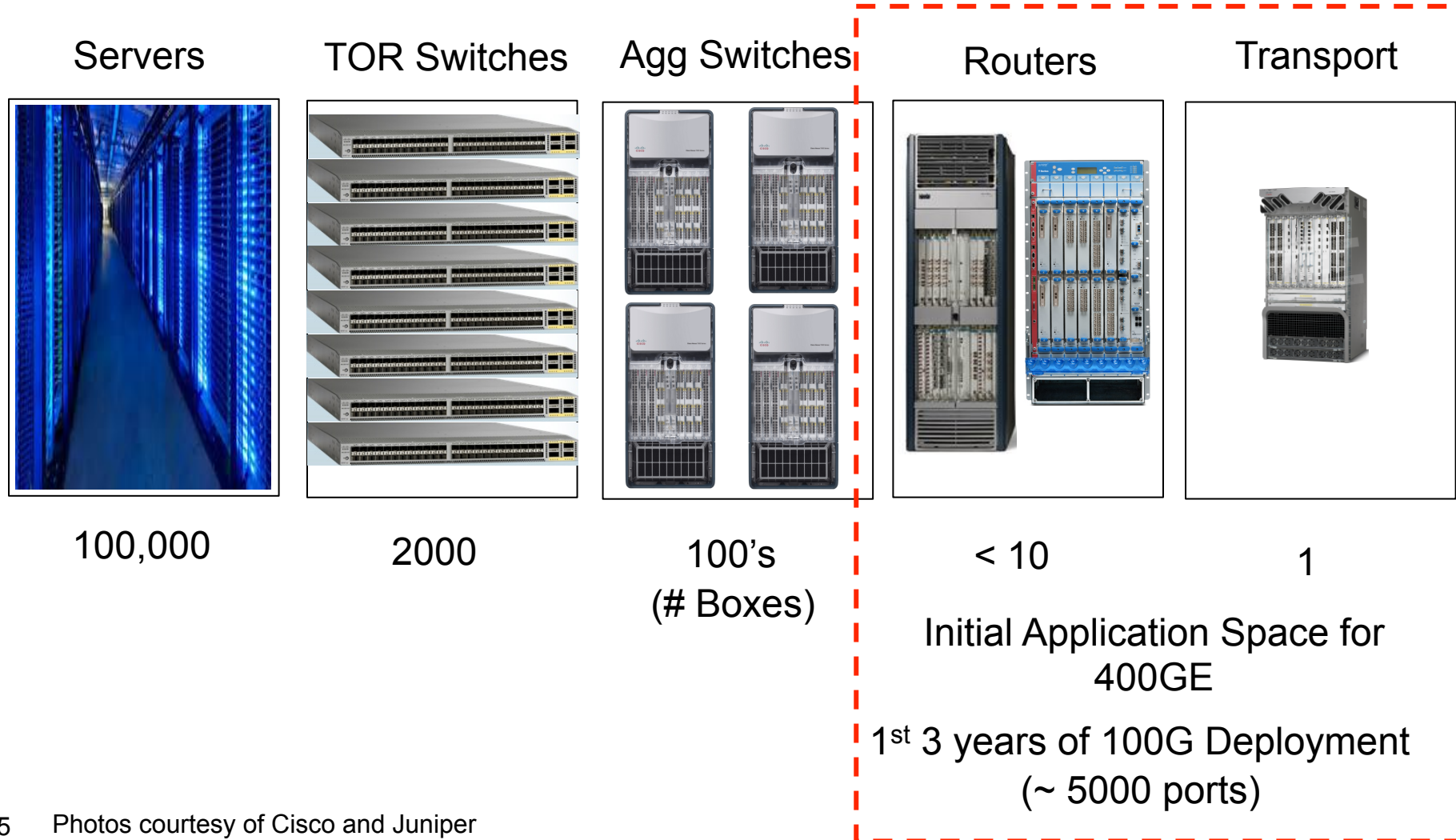
Servers	TOR Switches	Agg Switches	Routers	Transport
100,000	2000	100's (# Boxes)	< 10	1

Initial Application Space for 400GE

1st 3 years of 100G Deployment (~ 5000 ports)

Photos courtesy of Cisco and Juniper

Architecture Considerations

- It's all about the long game
- Should be scalable and flexible
  - Evolving PMD needs (elect / optical lane widths & rates)
- Well defined functional layers, with clearly defined interfaces
  - Allows different parts of the architecture to evolve independently, without impacting interoperability
- 802.3ba provides a very good base to build upon
  - Scalable MAC, MII, PCS, n-AUI
- Primary discussion will be around the role and impact of FEC
- Further reading: http://www.ieee802.org/3/hssg/public/jan07/muller_01_0107.pdf

6

Architecture Considerations

802.3ba Architecture
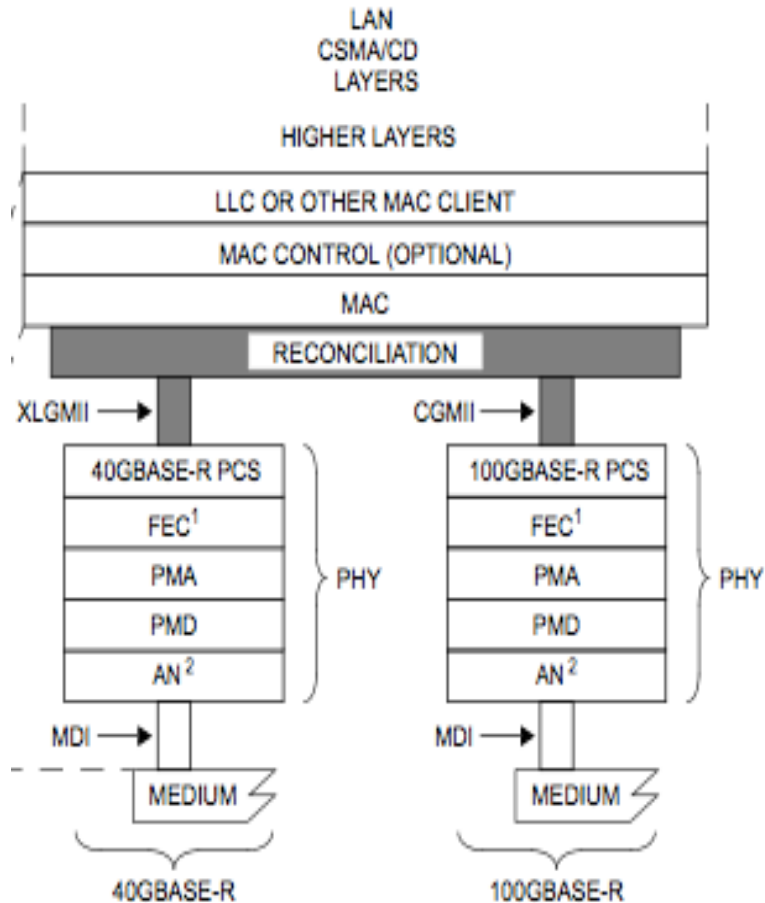
- MAC
  - speed-independent
  - just need to define bit time !

- MII
  - scalable (added in 802.3ba)
  - shouldn't have to touch

- PCS
  - scalable MLD (added in 802.3ba)
  - need to define # PCS lanes & rate
  - 16 x 25G ?

- FEC
  - ???

- PMA
  - simple, scalable, flexible bit muxing (.ba)
  - does FEC change this (word muxing?)

- PMD
  - see next section !

PMDs

- PMDs are potentially more challenging that the architecture
- A very large part of the industry effort and investment for a new Ethernet rate, goes into developing new PMDs
  - Architecture development can typical leverage the wider industry investment in Silicon, i.e. track Moore's Law
  - PMD development typically can't, and often requires dedicated, boutique technology development (slow and expensive)
  - Last point is compounded by new speeds being introduced on long haul transport and data networking at same time (see backup slide)

PMDs – What does this mean (1)

- "Laser" focus on initial applications.

- Limit PMD definition to only those required for the initial application.

  ➤ Be honest about the initial application ☺

- No point in defining/developing PMDs for applications that are 4+ years out. By the time the PMD is needed, the solution will likely be obsolete.

- Do we need a change to the IEEE process, to easily allow new PMDs to be defined/introduced over the lifecycle of a data rate, without the overhead of a new project each time ?
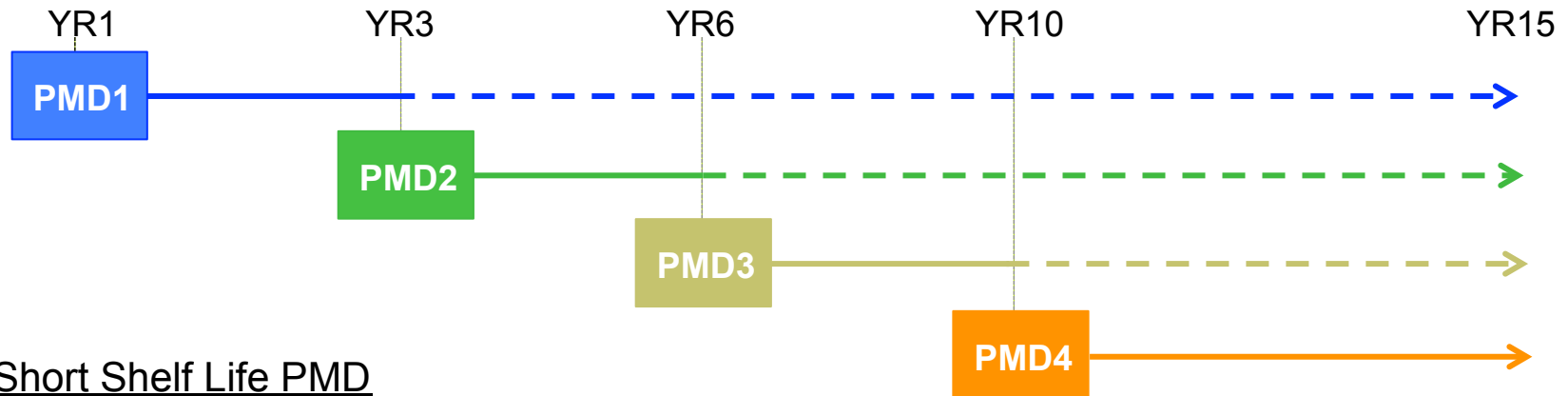
PMDs – What does this mean (2)

- Let's consider the PMD choice for a given reach objective ?
    - take 500m, SMF as an example
- People want a PMD that is cost effective at introduction
    - ~ same $/Gb as previous data rate
- People also want a PMD that has some legs to it …
    - meet cost/size/power requirements in 15+ years
- Can a single PMD meet both these requirements ?
- In the next slide we will introduce the concept of a "long shelf life" and a "short shelf life" PMD approach

Long Versus Short Shelf Life Approach

YR1

YR15

PMD1

Long Shelf Life PMD

- One reach objective. One PMD for all time (over several generations of form factor)
- Drives a more aggressive initial technology choice.
- Initial deployments higher $/Gb than previous rate. Cost comes down over time.

YR1　　　YR3　　　YR6　　　YR10　　　YR15

PMD1

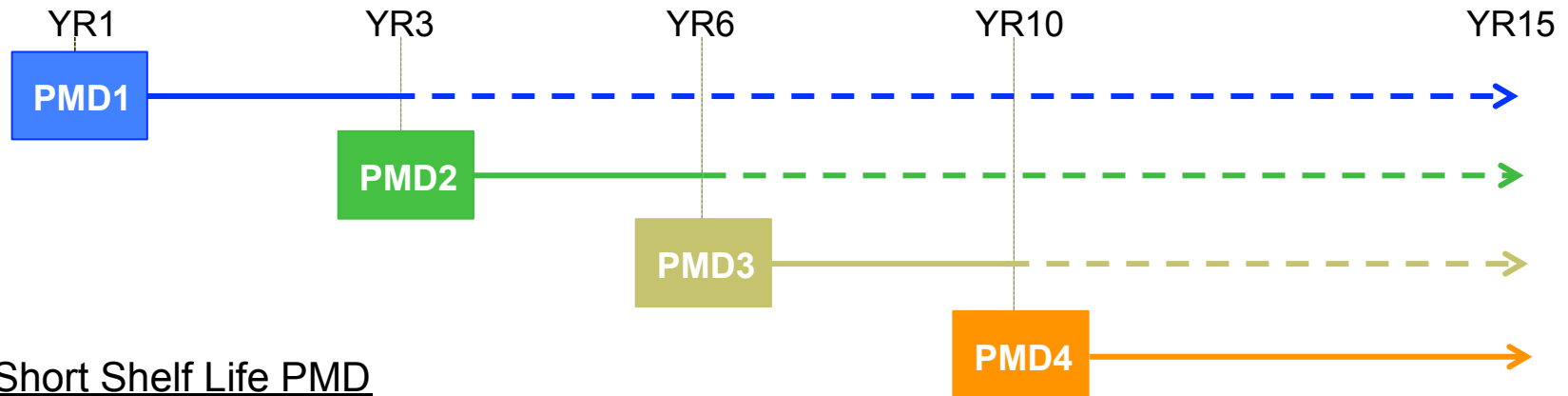**PMD2**

PMD3

PMD4

Short Shelf Life PMD

- One reach objective. Multiple different PMDs over time.
- More tactical technology choices. Each PMD is cost effective at introduction.
- Change PMD every 2-3 years, whenever a cheaper technology/solution is available
- No optical interoperability between PMDs. What are the real consequences ?

11

Long Versus Short Shelf Life Approach

YR1

YR15
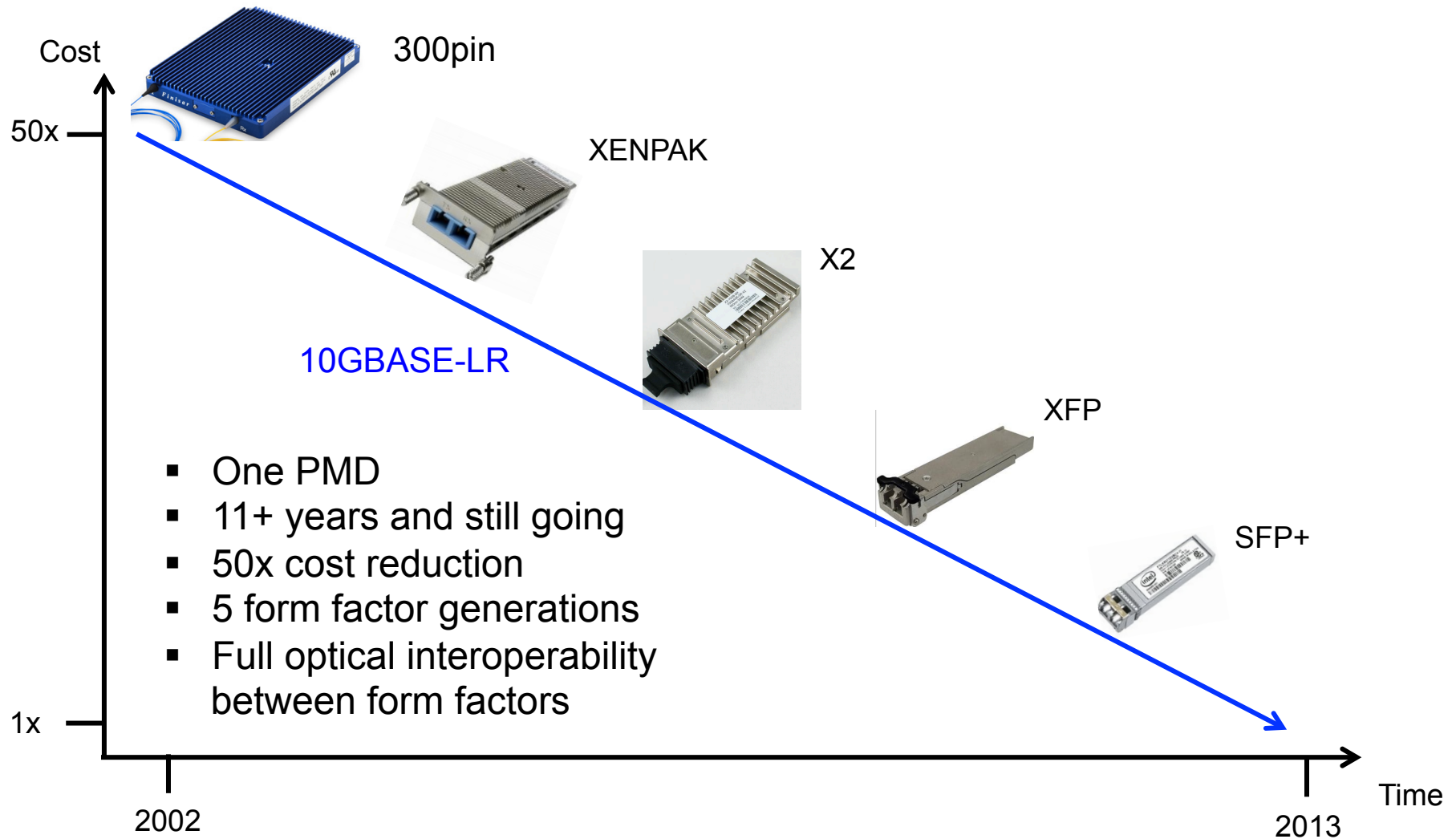
PMD1

Long Shelf Life PMD

- One reach objective. One PMD for all time (over several generations of form factor)
- Drives a more aggressive initial technology choice.
- Initial deployments higher $/Gb than previous rate. Cost comes down over time.

YR1   YR3   YR6   YR10   YR15
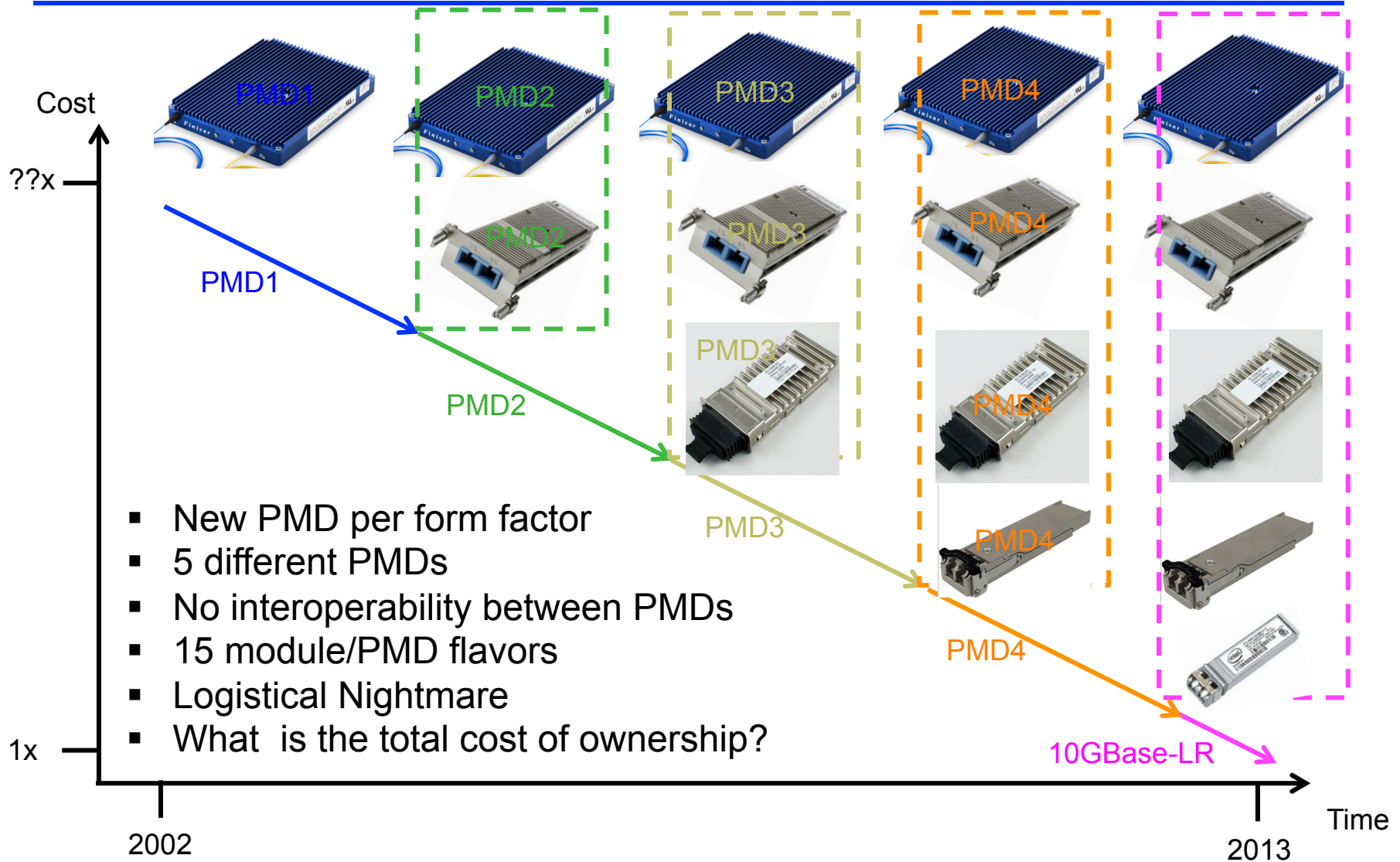
PMD1

PMD2

PMD3

PMD4

Short Shelf Life PMD

- One reach objective. Multiple different PMDs over time.
- More tactical technology choices. Each PMD is cost effective at introduction.
- Change PMD every 2-3 years, whenever a cheaper technology/solution is available
- No optical interoperability between PMDs. What are the real consequences ?

Long Shelf Life Example - 10GBASE-LR

Cost

50x

300pin

XENPAK

X2

10GBASE-LR

XFP

SFP+

- One PMD
- 11+ years and still going
- 50x cost reduction
- 5 form factor generations
- Full optical interoperability between form factors

1x

2002

2013

Time

What if we had taken a different (short shelf life) approach for 10G SMF ?

Cost

??x

1x

Time

2002

2013

PMD1

PMD2

PMD3

PMD4

10GBase-LR

- New PMD per form factor
- 5 different PMDs
- No interoperability between PMDs
- 15 module/PMD flavors
- Logistical Nightmare
- What is the total cost of ownership?

14

PMDs – Final Thoughts

- Perhaps the right solution is a compromise between these two extremes?

- One could define a short shelf life solution for the initial deployment (where volumes are likely to be low), and then focus on a single long shelf life solution for the second wave of deployment (where volumes are presumably picking up and warrant the investment)

- Examples could be:

  - Short Shelf life: 4 x 100GBASE-LR4 (Jeff's proposal)

  - Long Shelf life: 4 x WDM based on 100G Adv Mod
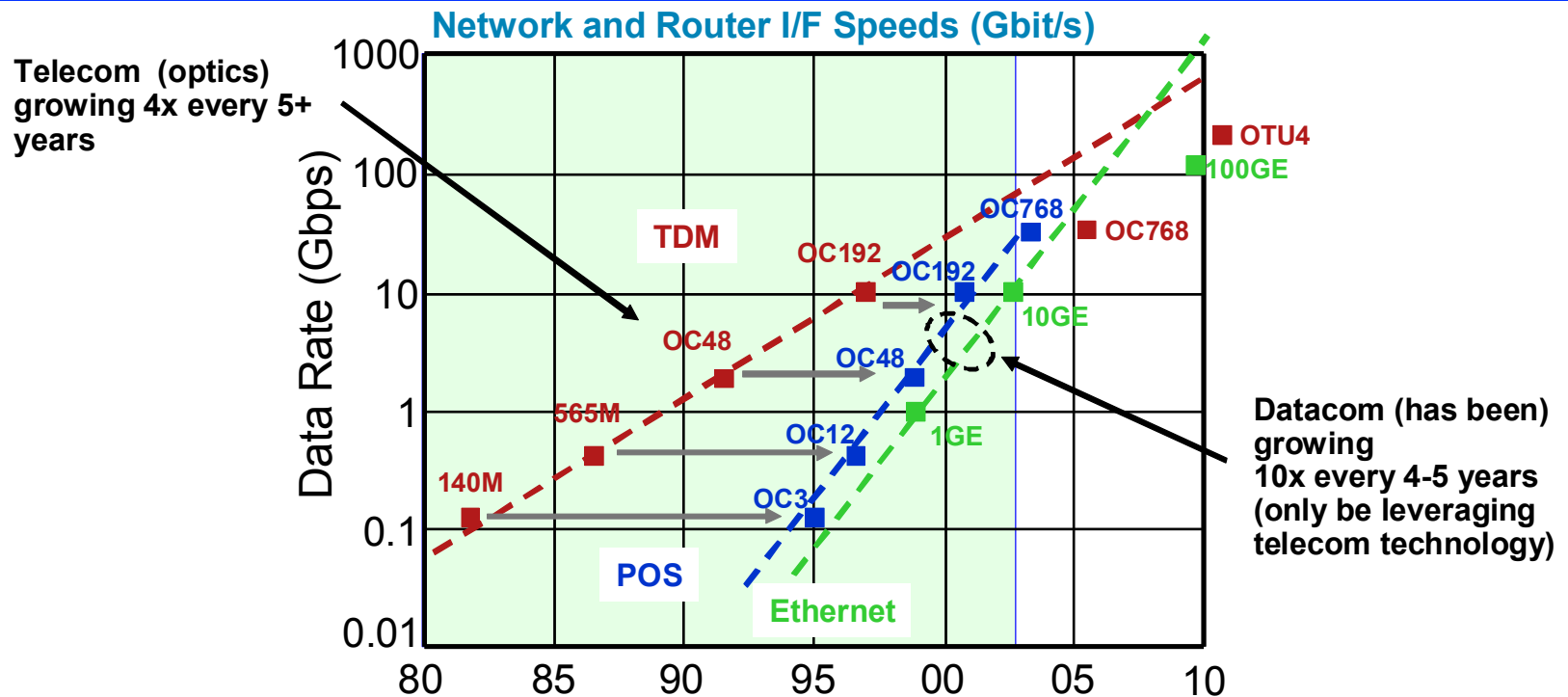
One Final Thought - BER Requirements !

- We all know that a BER objective of 10-12 just doesn't cut it these days

- Let's stop fooling ourselves and get it right this time !

- We need 10-15 or lower, maybe even 10-18 ?

- Perhaps FEC will be our savior ?

  - Allows us to operate at a low BER but test compliance at a much higher BER (less time)

Summary

- Architecture:
  - Long shelf life
  - 802.3ba provides a good foundation
  - Primary discussion will be around role/impact of FEC
- PMDs
  - "Laser" focus on initial applications
  - Don't define PMDs for applications that are 4+ years out
  - Long versus short shelf life PMD approaches for a given reach objective ?
  - Strategy for dealing with backwards/forwards compatibility (if going with a short shelf life PMD)

Backup

PMDs – Some historical concept

Network and Router I/F Speeds (Gbit/s)

Telecom (optics) growing 4x every 5+ years

Datacom (has been) growing 10x every 4-5 years (only be leveraging telecom technology)

TDM · OC192 · OC48 · 565M · 140M

POS · OC192 · OC48 · OC12 · OC3

Ethernet · OTU4 · 100GE · OC768 · 10GE · 1GE

Data Rate (Gbps): 1000, 100, 10, 1, 0.1, 0.01

Years: 80, 85, 90, 95, 00, 05, 10

- Historically new optical technology was developed first by long haul transport, and had gone through several iterations before being required on data networking equipment (client interfaces)

- This is very different to today, where new speeds are typically deployed on both long haul transport and data networking at the same point in time.