

What kind of interface do we need
in HPC/server environment?

Petar Pepeljugoski

IBM

Server/Supercomputer environment

- Supercomputer applications for higher speed networks*:
 - Cluster: inside the cluster/parallel system
 - for executing parallel applications
 - Computer/Storage nodes
 - LAN (between compute nodes and external machines)
 - Visualization, FTP to other machines etc.

Cluster links are the most numerous ones

*Assume 25 Gb/s/lane

What kind of links do we need?

- Everyone needs optical interconnect with low power, low cost and high density
 - One size fits all does not meet the requirements in this environment
- Most links are very short (20-30m, some up to 50-60m)
 - We need to have very short optical PMD (few meters to <10m) for within the rack connections
 - We also need one PMD for the rest
- Some large data centers may need longer links
 - Number of such links is relatively low
 - SMF PMD(s) may service those needs

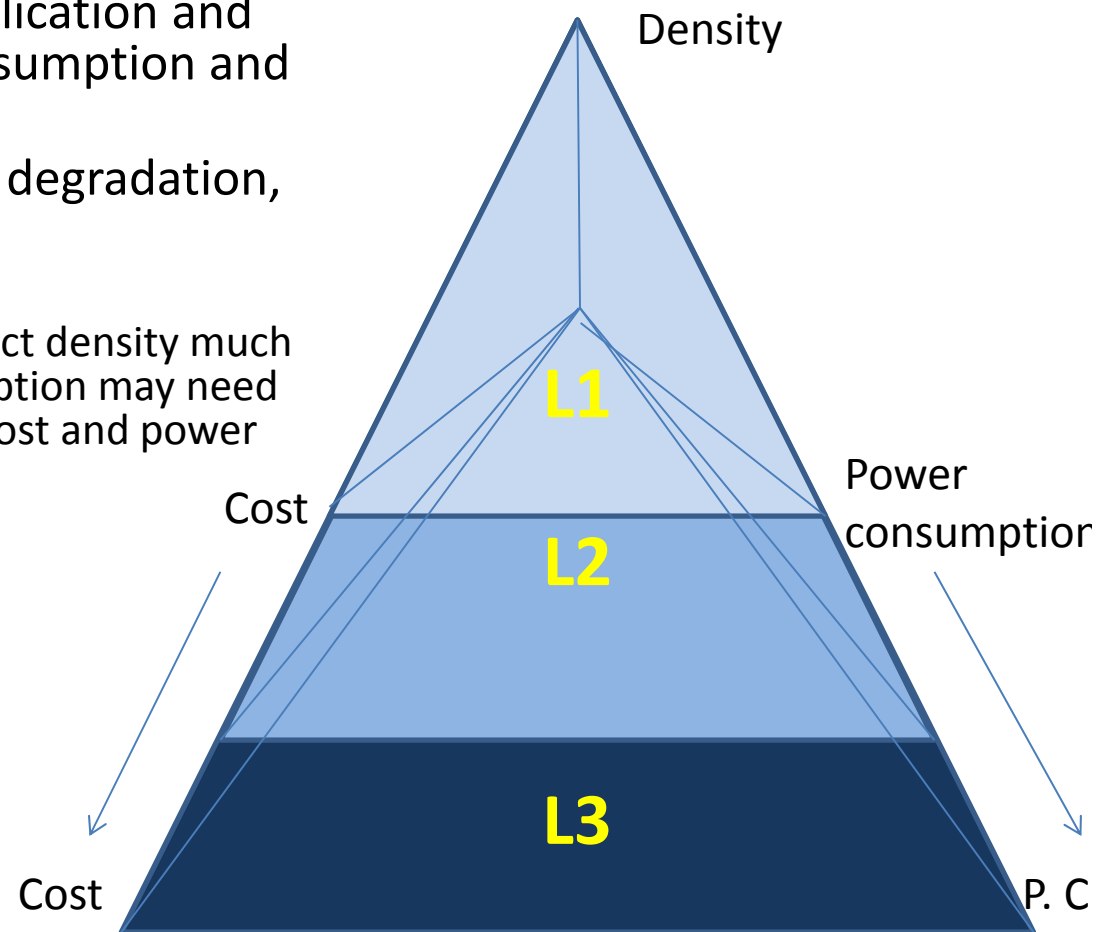
Why do we need multiple PMDs: optimizing link cost

- Link cost is affected by many parameters
 - Optimization depends on application and sensitivity to cost, power consumption and density
 - Length increase causes signal degradation, penalties, requires:
 - Equalization and/or CDR
 - CDR, Equalization do not impact density much (except higher power consumption may need larger package), but impacts cost and power consumption (~2x)

L1 – < 10m, small penalties (intra-rack, or to neighboring rack)

L2 – 50-75m, moderate penalties, may not need CDR

L3 - >100m, significant penalties, (CDR required, equalization, possibly FEC)



What else do we need?

- Can improve cost by increasing yield (i.e. not worrying about laser rms linewidth), no extra chips, simplifying testing
- Need to look for other options to reduce cost and power consumption
 - modules in some PMDs can be placed deep inside the board, providing savings in packaging, eliminating expensive electrical links and connectors (for example, no need for CDR, equalization in optical module, simplified host receiver)
 - Can specify only TP2 and TP3

Length Objectives

- Link lengths <50-75m (MMF based PMD)
- Link lengths <10m (MMF based PMD)
- Optional MMF based PMD with only TP2 and TP3 specified (optical modules deep inside, enables optical backplane)
 - or optical backplane (under consideration in other standards, many technology demos)
- Larger data centers which need longer lengths can be serviced by SMF PMD(s)
 - Si Photonics is coming, may eliminate need for multiple PMDs