MMF PMD for Short Distances in Data Center and High Performance Computing Environments

Petar Pepeljugoski

IBM Research

Collaborators and supporters

- Mike Dudek, Picolight
- Jack Jewell, Picolight
- John Abbott, Corning
- Steve Swanson, Corning
- Eric Grann, Omron
- Joel Georgen, Force10Networks
- Greg Hankins, Force10Networks
- Paul Kolesar, CommScope
- Mike Hughes, USConec
- Russ Granger, USConec
- Robert Lingle, OFS
- Ken Jackson, Emcore
- John Dallesasse, Emcore
- William Ring, WSR Optical Device Solutions
- Jim Tatum, Finisar
- George Oulundsen, OFS
- Andrew Oliviero, OFS
- Petar Pepeljugoski, IBM
- Wenbin Jiang, JDSU
- Jan Peeters Weem, Intel
- Brad Booth, AMCC
- Matt Traverso, Opnext
- Jeff Lynch, IBM

Outline

- Applications & market potential for MMF
- Distance needs for MMF solutions
- MMF Technology Options
- Why MMF – key trade-offs between MM and SM PMD
- Conclusions

Applications

- Data center
  - Zone-to-zone
  - Rack-to-rack
- HPC
  - CPU-CPU interconnect
  - I/O links
- In-building backbone
  - MDF-to-IDF
  - IDF-to-IDF

MDF – Main Distribution Frame
IDF – Intermediate Distribution Frame

Market Potential

- Data Center (today, Ethernet links)
  - more than 90% are MMF, less than 10% SMF
- HPC
  - thousands to tens of thousands of CPU-CPU links in a single supercomputer
    - Number depends on size (number of CPUs) and architecture
  - I/O links
    - Can be substantial fraction of the total number of interconnects
    - Ex.: Mare Nostrum uses ~5000 links
      - runs 4X SDR IB to transmit Fibre Channel traffic
        - » Takes advantage of density of parallel links
        - » Density and power consumption is very important in this environment

Distance Distributions for MMF

- Distances depend on platform
  - In HPC (to 100m)
  - In data center (to 150m)
  - In-building backbones (to 300m)
- Distance distributions*
  - 70-80% less than 100m
- Cost optimized 150m
  - Cover the majority of distances in all applications while optimizing the cost
- Performance optimized 300m
  - Required for future in-building backbone upgrades

* Based on ~several million MMF links in IBM installations

Technology Options*

- 850 nm MMF parallel
  - Strong history with 4 wide transceivers and 12 wide Tx, Rx modules
    - Proprietary links since 2000
    - InfiniBand SDR shipping since 2002 ?, now DDR, QDR planned
- 850 nm MMF CWDM over duplex
  - Considered during 10GbE
  - Shipping in high-def video markets today
- 850 nm Combination parallel and CWDM
  - Possibility to reduce cabling cost
  - Possibility to minimize total cost of ownership (TCO)
- All technology options above operate optimally on OM3 MMF

SDR – Single Data Rate (2.5 Gbd, 2 Gb/s/lane)
DDR – Dual Data Rate (5 Gbd, 4 Gb/s/lane)
QDR – Quad Data Rate (10 Gbd, 8 Gb/s/lane)

* See list of references at back of presentation

Key Trade-offs: multi-fiber MM vs. WDM SM for <150m

Benefit	MM	SM	Comment
Lower power consumption	+ +		Smaller optical power budget, no temp. control
Smaller footprint size	+		Smaller # of components, lower heat dissipation
Higher linear edge density	+		If module edge mounted, otherwise SM advantage
Ability to place close to logic ICs	+		Implies shorter el. links – better signal integrity
Lower Tx / Rx complexity	+ +		No temp. control, no mux/demux required
Lower laser cost	+ +		VCSEL arrays
Lower detector cost	+		GaAs vs. InP
Lower medium (fiber) cost		+	Significant for distances close to 300m
Lower connector cost		+	Only if mux/demux internal
Commonality with 10km solutions		+ -	+ One PMD type - Costly for short distances
Lower cost for added channels/features	+		e.g. up to 12 channels or simple FEC mechanism
More robust handling	+		Greater tolerance to dirt, lower bending loss

References

- TIA/EIA-604-5 Fiber Optic Connector Intermateability Standard - Type MPO
- TIA/EIA-568B.1-7 Guidelines for Maintaining Polarity Using Array Connectors
- InfiniBand™ Architecture Release 1.1 Volume 2 - Physical Specifications
- ANSI INCITS 364-2003 Fibre Channel - 10 Gigabit (10GFC)
- OIF-VSR4-01.0 Very Short Reach (VSR) OC-192/STM-64 Interface Based on Parallel Optics
- OIF-VSR4-03.0 Very Short Reach (VSR) OC-192 four fiber Interface Based on Parallel Optics
- OIF-VSR5-01.0 Very Short Reach Interface Level 5 (VSR-5): SONET/SDH OC-768 interface for Very Short Reach (VSR) applications
- Pepeljugoski et al., IEEE JLT, vol. 21, no. 5, 2003
- Kolesar_1_0906, Data Center Fiber Cabling Topologies and Lengths, IEEE 802.3 HSSG
- Kuchta et al., IEEE JLT, vol. 22, no. 9, 2004
- J. Dallesasse, Technologies for 100 Gigabit Ethernet: Components and Modules for the Physical Layer, OIDA Forum, August, 2006
- K. Jackson, Packaging for Next generation Optical data Links, OIDA Forum, August, 2006
- J. Jewell, Metrics and Technologies for 100GbE Components, OIDA Forum, August, 2006
- P. Pepeljugoski, 100Gb/s Ethernet Applications in High Performance Servers and Data Centers, OIDA Forum, August, 2006
- J. Tatum, VCSELs and Detectors for 100 Gb/s Ethernet, OIDA Forum, August, 2006
- Swirhun et al., proceedings of 46th ECTC, 1996.
- Brian E. Lemoff*, et. al.,* JLT, VOL. 22, No. 9, Sept. 2004, MAUI: Enabling Fiber-to-the-Processor With Parallel Multiwavelength Optical Interconnects

Conclusions

- The collective advantages of MM solutions position them as the optimal choice for short-distance applications
- The HSSG should set an objective to deliver a solution for a distance with a lower bound of 100-150m, and an upper bound of 300m, on OM3 fiber