Higher-Speed Ethernet Study Group

Implementation Considerations

Sept 2006

Roger Merel

Topics

- Lane Rates
- Reach Applications
- Media Options
- Wavelength Plan
- Total Capacity for HS-Ethernet
- Capacity Scaling Options
- Take Aways

Lane Rate Permutations

Lanes vs Total Rate

	Number of Lanes									
Total Rate	1	2	3	4	5	6	8	10	12	16
40G	40G	20G	13.3G	10G						
80G	80G	40G	26.7G	20G	16.0G	13.3G	10G			
100G	100G	50G	33.3G	25G	20G	16.7G	12.5G	10G		
120G	120G	60G	40G	30G	24G	20G	15G	12G	10G	
160G	160G	80G	53.3G	40G	32G	26.7G	20G	16G	13.3G	10G

Lane Rate vs Technology

Technology	Relative Economics	Rate per Lane						
		<11.3G	12-14G	15-17G	20G	24-30G	33-40G	50G+
Avail CMOS IP	$	✓	?					
New CMOS IP	$$	✓	✓	✓	?			
SiGE IP	$$$	✓	✓	✓	✓	✓	?	?
III-V Elect	$$$$	✓	✓	✓	✓	✓	✓	✓
Cu Cable Drivers	$	✓	?					
850 VCSEL	$$	✓	?					
CMOS Photonics	$$	✓	?					
DFB DMLs	$$	✓	✓	✓	✓	✓	?	?
InP PICs	$$$	✓	✓	✓	✓	✓	✓	?
Discrete EMLs	$$$$	✓	✓	✓	✓	✓	✓	?
Discrete Photonics	$$$$$	✓	✓	✓	✓	✓	✓	✓
Twinax	$$$$$	✓	?					
OM3 Ribbons	$$$$	✓	✓	✓	?	?		
SMF Ribbons	$$$	✓	✓	✓	✓	✓	✓	✓
OM3 Pairs	$$	✓ +WDM	✓ +WDM	✓ +WDM	?	?		
OM2 Pairs	$$	✓ +WDM +EDC	✓ +WDM +EDC	✓ +WDM +EDC	?	?		
SMF Pairs	$	✓ +WDM	✓ +WDM	✓ +WDM	✓ +WDM	✓ +WDM	✓ +WDM	✓ +WDM

NOTE: While it is likely possibly to develop EDC >10G, not clear if there is a willingness to do so.

Lane Rate Consequences

- Choosing a lane rate for the PHY does NOT require system vendors to us this interface throughout their system.
  - System vendors may use 2:1 Mux and 1:2 Demux at the interface between the backplane & line cards. This eliminates objections regarding system density.
  - System vendors will need 10G interfaces between ASICs & NPUs.

- Choosing a rate above ~10G likely implies:
  - A need for New CMOS IP for chip I/O and more.
  - Eliminating Twinax Copper solutions for Adjacent rack (≤7m)
  - Eliminating VCSEL solutions for Datacenters (≤100m)
  - Eliminating CMOS Photonic solutions for Enterprise Reach (≤2000m)
  - Eliminating synergistic use of the PHY for other applications which could benefit from Nx10G solutions.

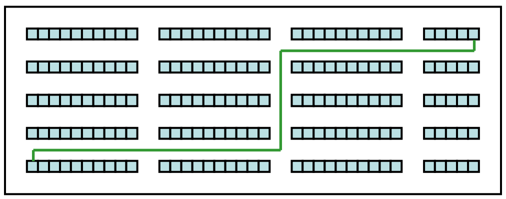
Reach Applications

- ✓ <7-10m is very useful for interconnecting “adjacent” rack and shelves.
- ✓ <100m will generally serve within a datacenter.
- ✗ 300m is **not** a clear priority as this only serves from Datacenter to remote closet which is only important for 10G to the desktop.
- ✓ <2km will serve to carry aggregate capacities in super-datacenters or connections in a meta-datacenter, enterprise, carrier hotel, or campus.
- ✗ 10km is **not** a clear priority as this only serves from Business Premises to the closest Service Provider’s Switching Office.
  - Should not expect that all Switching Offices will have HS-Ethernet
- ✓ <40-60km will generally serve to interconnect:
  - Customer Datacenters to a Service Provider’s major Switching Office
  - Between major Switching Offices
  - Between different Service Providers’ main facilities in a city
  - Between a POP facility and a major Switching Office
- ✗ >60km will be served by WDM Line Systems
  - We need **not** define HSSG here; they are not multi-vendor inter-operable.

We will require evidence to assure that these are the reaches needed.

Reach Applications

- Connections in a datacenter
  - For up to ~100m (or more), parallel VCSEL arrays are one option.
  - Objections focus on the cost of patch panels and of the 24-fibers for longer distances.
  - Concerns focus on yield & reliability of Arrays of 10G VCSELs.

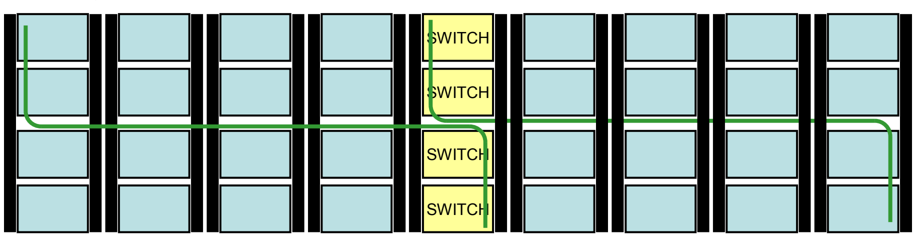


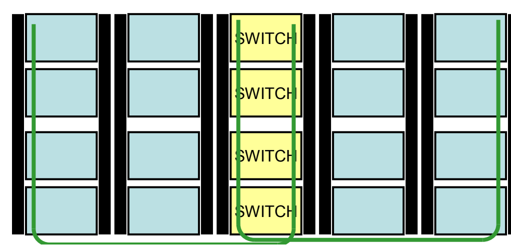
50’ x 125’ Datacenter w/ 175 Racks or Cabinets
(55m covers any-to-any in this sized datacenter)

100’ x 250’ Datacenter w/ 700 Racks or Cabinets
(55m covers any-to-middle)
(100m covers any-to-any)

200’ x 500’ Datacenter w/ 2800 Racks or Cabinets
(100m covers any-to-middle in this sized datacenter)

- Connections between “adjacent” equipment
  - For up to ~7m, copper will beat the economics of optics.
  - 10G Lanes over Twinax is being done now.
  - Shouldn’t select a format that precludes this (e.g. one using >15Gbps lanes).
  - While it doesn’t seem like a long reach, the following can be done with 7m depending on cable routing “rules”.





Media Options

- What was true at <4G is likely **NOT** true at HS-Ethernet speeds.
  - That MM solutions have favorable economics compared to SM.
- It has always been true that MM cable is considerably more expensive than SM cable. This “gap” has only widened with OM3/OM3+.
- Using a Ribbon cable of OM3 is **NOT** a legacy cable solution.
- Ribbon cable of OM3 is definitely a single generation media solution.
- Two cables of 12-strand OM3 and their patch panels is expensive.

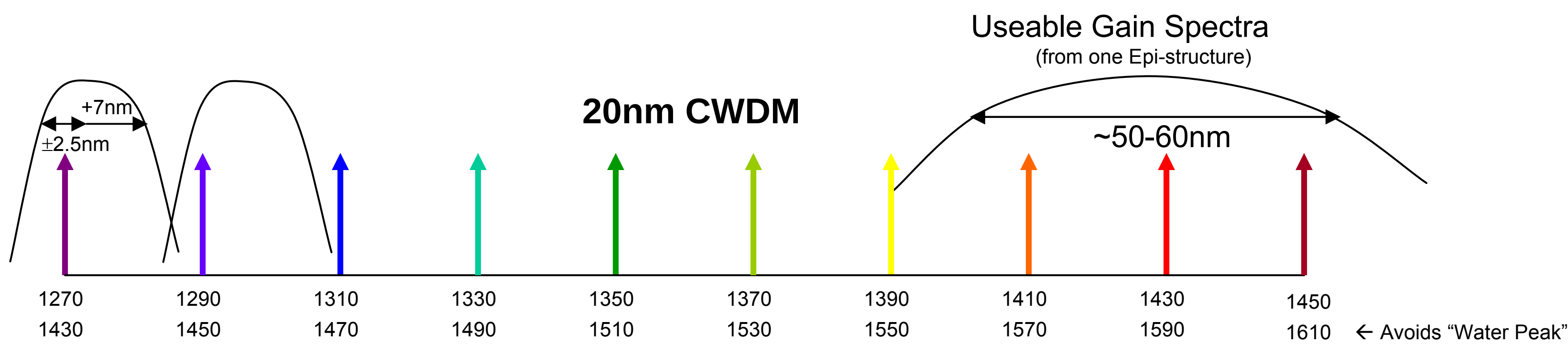
- SMF is the only future-proof media option.
- SMF can support PICs whereas MMF can not.

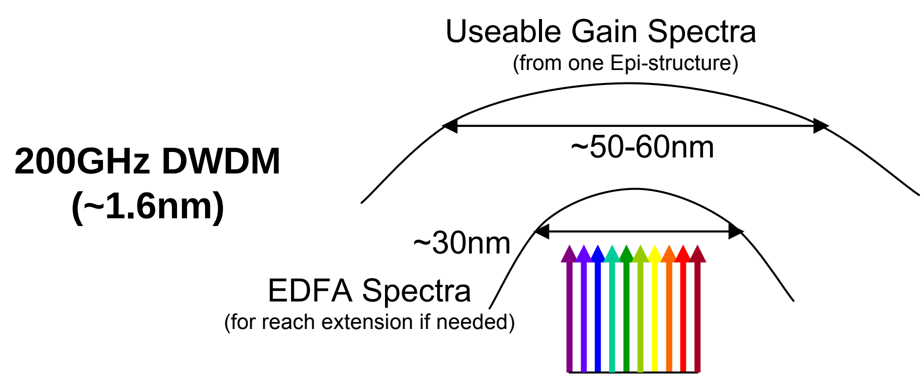
Wavelength Plan Permutations

Laser Configuration	Thermal Package	Fine Adjust	Absolute Fab Var.	Relative Fab Var.	Rel. Temp Shift	-3dB Passband	Channel Spacing	Gain Spectra	# Lasers / WDM Array	Filters	Supports PICs	Supports EDFAs
Indiv. Lasers & Indep. Temp. Environ.	Uncooled	n/a	±2.5nm	±2.5nm	7 nm	12 nm	20 nm	50-60nm	≤3	fixed	no	no
Indiv. Lasers & Shared Temp. Environ.	Uncooled	n/a	±2.5nm	±2.5nm	0.5 nm	5.5 nm	10 nm	50-60nm	≤5	fixed	no	no
Array Lasers & Shared Temp. Environ.	Uncooled	n/a	±2.5nm	±0.25nm	0.1 nm	0.8 nm	1.6 nm	50-60nm	»10	tracks	yes	yes
Array Lasers & Shared Temp. Environ.+Fine Adj.	Uncooled	incl	±2.5nm	±0.1nm	0 nm	≤0.4 nm	≤0.8 nm	50-60nm	»10	tracks	yes	yes
Array Lasers & Shared Temp. Environ.	Cooled	n/a	±2.5nm	±0.25nm	0.1 nm	0.8 nm	1.6 nm	50-60nm	»10	fixed	yes	yes
Array Lasers & Shared Temp. Environ.+Fine Adj.	Cooled	incl	±2.5nm	±0.1nm	0 nm	≤0.4 nm	≤0.8 nm	50-60nm	»10	fixed	yes	yes

Can use same PIC with variations on package.

Can use same PIC with variations on package.





NOTE: Gain spectra center varies more than DFB gratings and shifts faster with temp which means that in uncooled applications, even if spectra is wider, not all may be used. .

Wavelength Consequences

- Choosing a CWDM solution likely:
  - Requires lasers made from varied epi-wafers
  - Precludes monolithic integration which may be the most economical
  - Can't offer optical amplification for reach-extension.

- MMF CWDM technologies for LX-4 do not scale to $8^{+}\lambda$ x10G

  Particularly for the RxDEMUX:
  - Mechanical offsets increase by number of lambdas
  - Photodetectors areas shrink by the root of the lane rate.

  Potentially solvable but at higher cost.

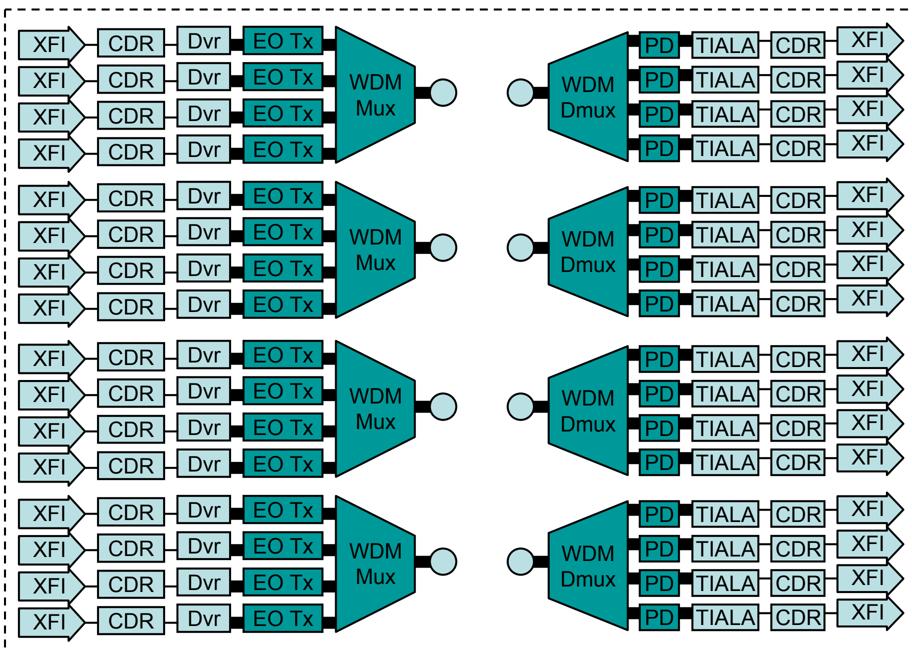
  Can <u>never</u> be a PIC/PLC! Must always be built from discrete filters.

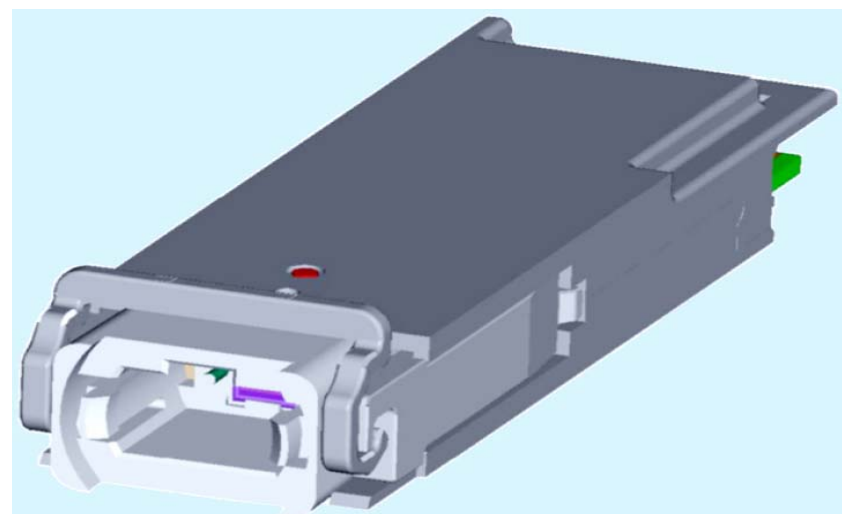
Total Capacity for HS-Ethernet

- HS-Ethernet must make a major leap in capacity.
  - Essential for ROI on development
- 10x is a completely arbitrary albeit historical increase
- Should consider rational reasons for the increase
  - 80G would allow packing two (2) OC768s into the same PHY & could allow packing two(2) 80G HS-Ethernet into a future OC3072 or possibly future SONET/SDH could next choose 80G (OC1536).
  - 80G would offer capacity interchangeability against the latest current system architectures having 8-port and 16-port 10G modularity and backplanes with lanes counts in multiples of 4.

Capacity Scaling for HS-Ethernet

- Not certain that Capacity Scaling is a good idea, however…
- If it is to be considered it must create benefits for PHY vendors, System vendors, & customers.
- Example solution: 8-fiber MPO connector interface which supports
  - Up to 4 Tx & 4 Rx fiber where each fiber supports 4x10G
  - Units can be populated with 1 Tx+Rx, 2 Tx+Rx, or 4 Tx+Rx
  - The 4 sets of 4x10G can be used to create HS-Ethernet ports which operate at 40G, 80G, 120G, or 160G. Thus a single module can be used as 4- 40G, 2-80G, 1-160G, 1-40G + 1-120G, or 2-40G + 1-80G ports.
  - Earlier versions can support fewer fibers pairs (e.g. 2 pairs for 80G).
  - Units which don't yield all 4 fiber pairs can be sold as lesser capacity units.
  - This won't work on a 8+ lambda plan





Take Aways

- There are many reasons to select a lane rate of ~10G.
- The same “Reaches” used by 10G are not necessarily applicable. We need evidence to support what reaches are required.
- 24-fibers of OM3 would be a expensive cable for a single technology generation.
- SMF cable for any reach is a future-proof media.
- CWDM is not the idea WDM format for shorter reaches.
- HS-Ethernet should have a reason for the selected rate and 100G may not be that rate.
- There are capacity scaling plans which can make sense.