Fast Readout Links

- Optical interconnects by network application
- Comparison with developments for LHC and HL-LHC
- What to expect for future High Energy Physics experiments
Optical Link Metrics

- **Bandwidth**
- **Energy (pJ/bit, mW/Gbps)**
  - $1\text{mW/Gbps} = 1\text{pJ/bit}$
    - 1 google search = 1kJ $\rightarrow$ 1 Pbits ($@1\text{pJ/bit}$) !!!
- **Density (mm$^2$/Gbps)**
- **Cost ($/\text{Gbps}$)**
- **Environmental Resistance**
Use and Misuse of Scaled Metrics

- mW/Gbps, mm²/Gbps, $/Gbps
  - Systematically favour very high bandwidth systems
  - Cannot be used to extrapolate likely figures for bandwidth-distributed systems
    - Per device (TRx) scaling more appropriate
    - mW/TRx, mm²/TRx, $/TRx
  - Nevertheless very useful to compare systems
Existing Optical Links in CMS (1)

- **CMS Tracker AOH**
  - 39k links
  - $40\text{MS/s} \times 8 \text{ bits} = 320 \text{ Mbps}$
  - 100 mW, 3 channels
    (dependent on laser bias)
  - $30\text{mm} \times 23\text{mm}, 3 \text{ channels}$
  - 300 CHF / channel

- **CMS ECAL GOH**
  - 10k links
  - 800 Mbps
  - GOL operates up to 1600 Mbps

Aggregate bandwidth ~ Tk: 12.5Tbps
ECAL: 8Tbps
### Existing Optical Links in CMS (2)

<table>
<thead>
<tr>
<th></th>
<th>AOH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bandwidth Gbps</strong></td>
<td>0.32</td>
</tr>
<tr>
<td><strong>Reach m</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Energy pJ/bit, mW/Gbps</strong></td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>* extrapolating to bidi link</td>
</tr>
<tr>
<td><strong>Density mm²/Gbps</strong></td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>* extrapolating to bidi link</td>
</tr>
<tr>
<td><strong>Cost $/Gbps</strong></td>
<td>900</td>
</tr>
</tbody>
</table>
Bandwidth and Capacity

- **Bandwidth**
  - \( \# \text{Symbols/s} \times \text{bits/symbol} \times \# \lambda \)
  - Only one fibre considered here

- **Capacity = Reach x Bandwidth**
  - Reach not an issue in typical detector applications (~100m)
  - Only bandwidth optimization considered here

---

**Diagram:**

- **TDM**
- **WDM**
- **Bits/Symbol**

**a)** TDM

**b)** WDM

**c)**
- QPSK
- 8-PSK
- 16-QAM

---

23 June 2011    francois.vasey@cern.ch
WAN Network features

- **Ultimate Capacity**
  - Design for minimum received energy
  - Exploit TDM, WDM, Advanced modulation formats

- **Single Mode**

- **Electronic Mitigation Techniques**
  - Ultra fast electronics operating at line rate (f.i. 50GS/s ADC)
  - FEC, equalizers, …

- **Metrics (very coarse guess)**
  - Bandwidth: 10Tbps
  - Reach: 100km
  - Energy: 10’000 pJ/bit
  - Density: 10’000 mm²/Gbps
  - Cost: 100$/Gbps

Not in our ballpark, but watch out for emerging technologies
**Access over Passive Optical Network**

- **GPON, EPON 1.25Gbps - 2Gbps**
  - Consumes 18x less energy per user than VDSL2
  - Sharing of this capacity among multiple users (time-division multiplexing)
  - Distance to central-office can go up to 60km when amplifier-assisted, not a few 100s meters from set-top box to DSLAM.

- **10GPON solutions recently proposed by system vendors**

- **WDM dimension can also be exploited to increase capacity,**
  - And provide Point to Point connections, capacity on demand...while preserving scalability
PON Access Network features

- Moderate speed, reach
- Scalability
- Shared resource
- High dynamic range
  - High split ratio
- High Volume, Low Cost
- Single Mode, bidirectional
- 10G PON coming up, WDM PON next

Metrics (coarse guestimates)
- Bandwidth: 2.5Gbps
- Reach: 10km
- Energy: 400pj/bit
- Density: 200mm²/Gbps
- Cost: 10$/Gbps

But high volume and low cost model in right bandwidth range

Traffic flows in wrong direction

Not in our ballpark,
Local Area Network (LAN)

- Data Center (Rack to Rack): 1-100 Meters
  - Ethernet
  - InfiniBand
  - Fibre Channel / SAS
  - Proprietary

- Chip to Chip: 1 – 50 cm
  - Memory
  - Tera-scale CPUs

- Intra-Rack: 10s of Millions
  - Ethernet
  - InfiniBand
  - Fibre Channel / SAS
  - PCIe
  - Proprietary

- Board to Board: ~0.5 - 1.5M

- Metro & Long Haul (0.1-80km)
  - SONET
  - WDM
  - Ethernet

- Backplane

- Volume
  - 100s of Thousands
  - 10s of Millions
  - 100s of Millions
  - 10s of Billions
  - 100s of Billions

23 June 2011

francois.vasey@cern.ch
# 40/100GbE based on 10G PHYs

**PHY Types to be added by 802.3ba**

<table>
<thead>
<tr>
<th></th>
<th>40GbE</th>
<th>40GBASE-</th>
<th>100GbE</th>
<th>100GBASE-</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m backplane</td>
<td>4 x 10.3125G</td>
<td>Optional FEC</td>
<td>KR4</td>
<td></td>
</tr>
<tr>
<td>10m Cu Cable</td>
<td>4 x 10.3125G</td>
<td>Parallel Coax Cable</td>
<td>CR4</td>
<td>10 x 10.3125G</td>
</tr>
<tr>
<td>100m OM3 MMF</td>
<td>4 x 10.3125G</td>
<td>0.8µm, Ribbon Fiber</td>
<td>SR4</td>
<td>10 x 10.3125G</td>
</tr>
<tr>
<td>10km SMF</td>
<td>4 x 10.3125G</td>
<td>1.3µm CWDM</td>
<td>LR4</td>
<td>4 x 25.78125G</td>
</tr>
<tr>
<td>40km SMF</td>
<td></td>
<td></td>
<td></td>
<td>4 x 25.78125G</td>
</tr>
</tbody>
</table>

**Nomenclature**

- **40G BASE** – L R 4
- **MAC Data Rate**
- **Media Type**
- **PCS Type**
- **PMD Lane**

- **10G**: directly modulated EE lasers or VCSELs
- **25G**: externally modulated lasers
- **40G-100G**: aggregation

23 June 2011  
francois.vasey@cern.ch
LAN Network Features

- Converging with MAN
- TDM limits reached
- Aggregation via parallelism
  - Physical
  - Wavelength
  - Inheriting WAN features
- High Density
- Low Power
- High Volume Low Cost
- Multimode OM3 & OM4 fibre
  - 10G over 100m & 300m
  - Single Mode next

Metrics (guestimates)
- Bandwidth: 100Gbps
- Reach: 100m
- Energy: 10pJ/bit
- Density: 2mm²/Gbps
- Cost: 1$/Gbps
From modules to chips: packaging issues

- **Module-Based**

- **Chip-Level Packaging**

*Courtesy Shigeru Nakagawa, IBM*
Electrical Interconnects are still alive

- For short distances, EOE cannot beat EE
- Short range low power electrical links are the subject of intense development efforts
  - Moving target

**Electrical Link Power Efficiency Progress**

<table>
<thead>
<tr>
<th>Year</th>
<th>mW/Gb/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>30</td>
</tr>
<tr>
<td>2002</td>
<td>35</td>
</tr>
<tr>
<td>2004</td>
<td>40</td>
</tr>
<tr>
<td>2006</td>
<td>45</td>
</tr>
<tr>
<td>2008</td>
<td>20</td>
</tr>
<tr>
<td>2010</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Link Type</th>
<th>Power Efficiency (pJ/bit)</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrical</td>
<td>3</td>
<td>&lt; 1 m</td>
</tr>
<tr>
<td>EOE</td>
<td>7</td>
<td>&lt; 100 m</td>
</tr>
</tbody>
</table>
## Optical network metrics

### A risky comparison* (1)

*This comparison expresses the views of the author at presentation time and do not necessarily represent an accurate vision of the current and future situation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bandwidth Gbps</strong></td>
<td>0.32</td>
<td>10’000</td>
<td>2.5</td>
<td>100</td>
</tr>
<tr>
<td><strong>Reach m</strong></td>
<td>100</td>
<td>100’000</td>
<td>10’000</td>
<td>100</td>
</tr>
<tr>
<td><strong>Energy pJ/bit</strong></td>
<td>150</td>
<td>10’000</td>
<td>400</td>
<td>10</td>
</tr>
<tr>
<td><strong>Density mm²/Gbps</strong></td>
<td>1000</td>
<td>10’000</td>
<td>200</td>
<td>2</td>
</tr>
<tr>
<td><strong>Cost $/Gbps</strong></td>
<td>900</td>
<td>100’s</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
Next for HEP: Front End Versatile-TRx for HL-LHC

- Low Mass & Volume
  - Minimize material, avoid metals
- Capable of operating in a magnetic field
  - Requires replacement of ferrite bead used in laser bias network
- Radiation Tolerant
- Bitrate determined by ASICs: 5 – 10 Gbps
- 850nm and 1310nm flavours
Front End Versatile-TRx Results at 5Gbps

- **Tx**
  - SM@1310nm
  - MM@850nm

- **Rx**

BER

Rx Sensitivity = -18 dBm
## Optical network metrics

### A risky comparison* (2)

*This comparison expresses the views of the author at presentation time and do not necessarily represent an accurate vision of the current and future situation*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bandwidth</strong></td>
<td>0.32 Gbps</td>
<td>10’000</td>
<td>2.5</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td><strong>Reach</strong></td>
<td>100 m</td>
<td>100’000</td>
<td>10’000</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Energy</strong></td>
<td>150 pJ/bit</td>
<td>10’000</td>
<td>400</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td><strong>Density</strong></td>
<td>1000 mm²/Gbps</td>
<td>10’000</td>
<td>200</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>900 $/Gbps</td>
<td>100’s</td>
<td>10</td>
<td>1</td>
<td>60 ?</td>
</tr>
</tbody>
</table>
A TRx for an all-in-one module

- 50% scaling required
  - Size
  - Power
**Optical network metrics**

**A risky comparison* (3)**

*This comparison expresses the views of the author at presentation time and do not necessarily represent an accurate vision of the current and future situation*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth Gbps</td>
<td>0.32</td>
<td>10’000</td>
<td>2.5</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>Reach m</td>
<td>100</td>
<td>100’000</td>
<td>10’000</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Energy pJ/bit</td>
<td>150</td>
<td>10’000</td>
<td>400</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Density mm²/Gbps</td>
<td>1000</td>
<td>10’000</td>
<td>200</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>Cost $/Gbps</td>
<td>900</td>
<td>100’s</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Let’s not forget radiation tolerance
Radiation Tolerance: Is there a limit?

- Active opto devices tested up to now do not resist fluences beyond $10^{16}$ cm$^{-2}$
  - Still OK for SLHC Tk
  - Tighter margins
  - Are there alternatives for fluences beyond $10^{16}$ cm$^{-2}$?
- (Re)-consider Passives?
  - modulators
Extrapolating I/O Bandwidth Requirements

Input/Output interconnect (I/O) rate does not keep up with ability of chip to calculate

Ideal of 1 Byte of memory access for each floating point operation (FLOP) cannot be retained

- Byte/FLOP gap

Need to go optical, on chip, with WDM

- Si – CMOS Photonics
Evolution of high speed interconnects
2010-2015

Electrical or Optical engines + Opto PCB MM
PIC TRx

2015-2020 Monolithic integration of TRx and Processing chip SM

Base circuit board
Power and slow signals

Will this paradigm shift occur?
Will it bring us benefits
(density, power, radiation tolerance)?
Conclusions

- Broadband optical communication is a commodity and is ubiquitous
  - Also in HEP

- Scaling bandwidth and connectivity remains a major issue at all levels of the network hierarchy
  - HEP has a distributed bandwidth need. It cannot simply aggregate.

- Current optical communication technologies will not allow to meet the challenges ahead without major breakthroughs:
  - Advanced modulation, coding and electronic mitigation schemes
  - High density parallel and wavelength multiplexed optics (optical engines)
  - Photonic integration with CMOS

- Pragmatically speaking, all incremental electrical chip to chip interconnection approaches will be exhausted before fast interconnects go optical
  - Will this kill photonics integration?

- In the short/medium term: laser-based EOE conversion such as proposed by the Versatile Link common project is the way forward
  - 10Gbps, 10’s pJ/bit, 10’s mm²/Gbps

- In the long term:
  - Worth watching photonics integration
  - Embracing a new technology is resource intensive