CDRs, FEC, power and reach

Piers Dawe IPtronics

Supporters

- Petar Pepeljugoski — IBM
- Oren Sela — Mellanox
- Phil McClay — TE Connectivity

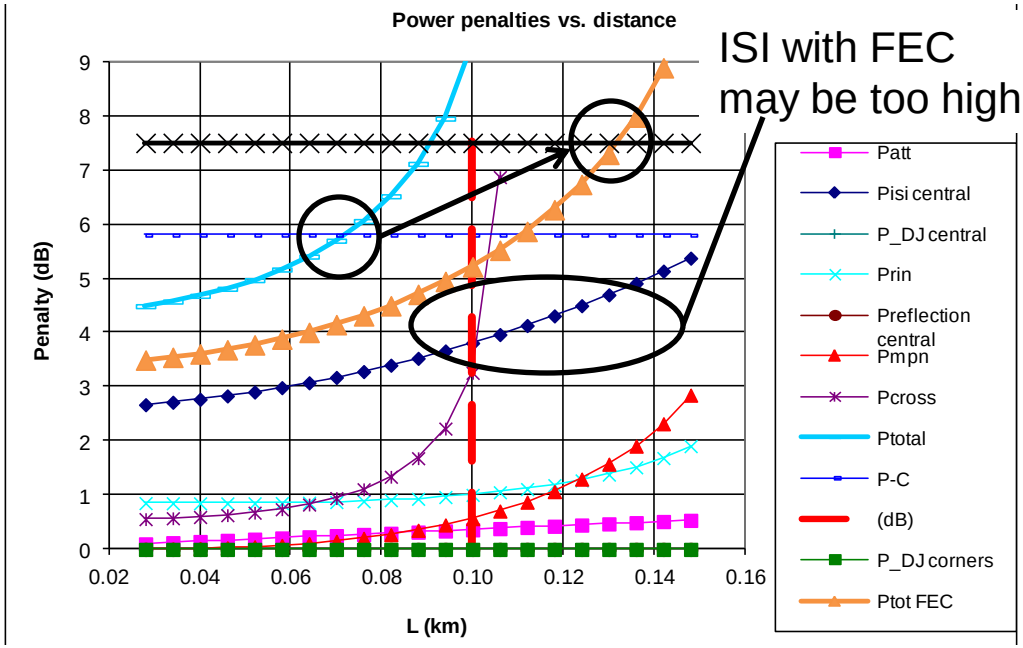
Contents

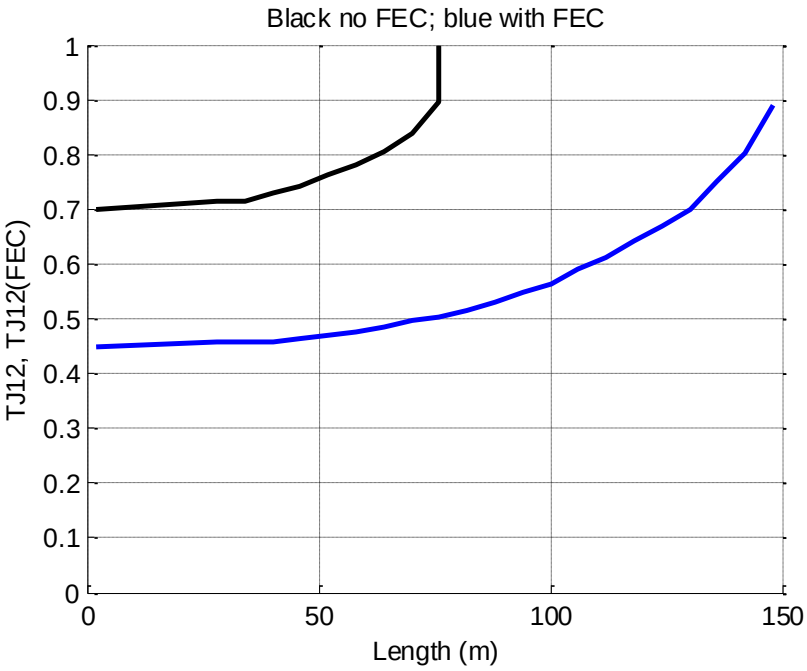
- Introduction
- Penalty vs. reach
  - 2 and 1 CDRs, FEC and no FEC
- Tx and Rx EQ
- Latency vs. reach
- Power vs. reach
- One PHY two options, or two PHYs?
- Compatibility with 100GBASE-CR4
- Conclusion

Introduction

- Most links are short
  - More so in future, with denser blade servers
- Low power is increasingly desirable
- What is the low cost/power/size option?
- Does it need the "extras":
  - CDRs in the module at each end of each lane?
  - Tx side and Rx side equalisation?
- Is it interoperable with the option with extras?
- The lowest power option is where the volume is: it's the most important to get right

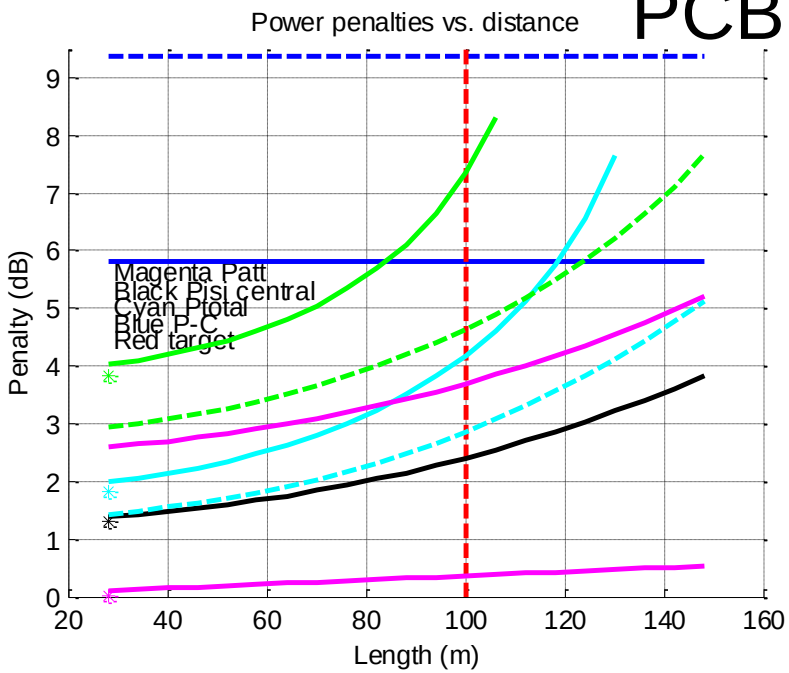
Penalty and jitter vs. reach

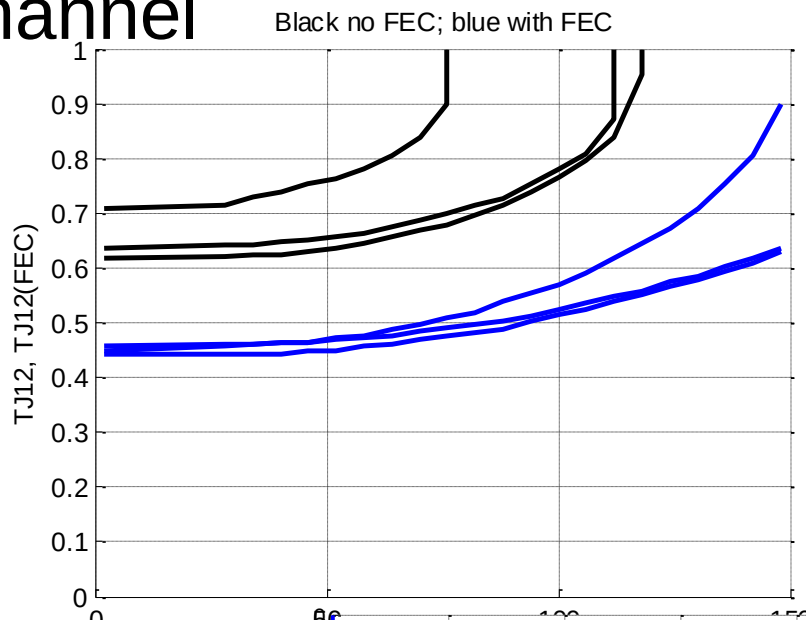


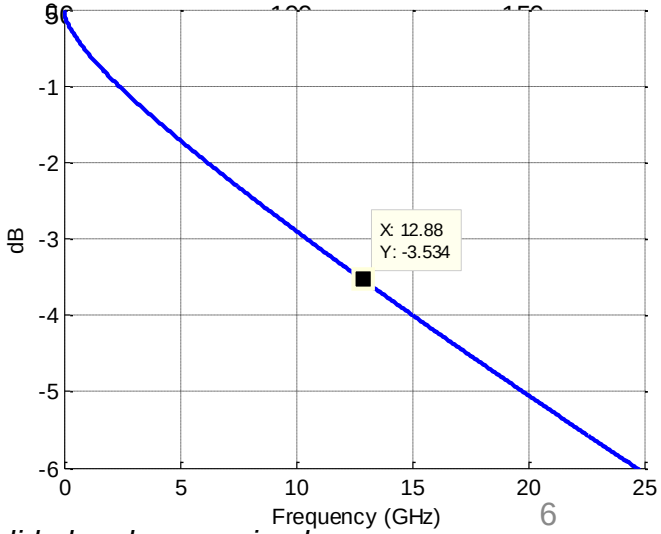


- Ideal case, OM4, "with CDRs". Standard link model equations on left, simulation on right
- Tx risetime 20 ps, zero sine jitter SJ
- RIN_OMA -130 dB/Hz
- Spectral width 0.6 nm
- Receiver bandwidth 20 GHz
- 6 ps pulse width shrinkage (PWS) ("DCD" in the link model)
- Uncorrected BER for FEC taken as 1e-6: leading candidates in gustlin_01_0112 show 2 to 5e-6 for corrected BER of 1e-15

Optical transmitter emphasis more than compensates for lack of one CDR with a clean PCB channel







- Left: magenta (ISI) and green (total penalty) fully retimed, no optical Tx emphasis, cyan and black semi-retimed with optical Tx emphasis. Clean 3.5 dB channel, compensated with 2-tap FFE
- Right: upper fully retimed, no optical Tx emphasis, lower semi-retimed with optical Tx emphasis. Clean 3.5 dB or 7 dB channel, compensated with 2-tap FFE

This slide has been revised

Tx and Rx EQ

- Note EQ is not as effective as faster laser
  - More susceptible to laser resonance, random noise from all causes, random jitter
  - Adaptive equalisation probably too power-hungry, lasers vary over temperature, EQ cannot be highly tuned

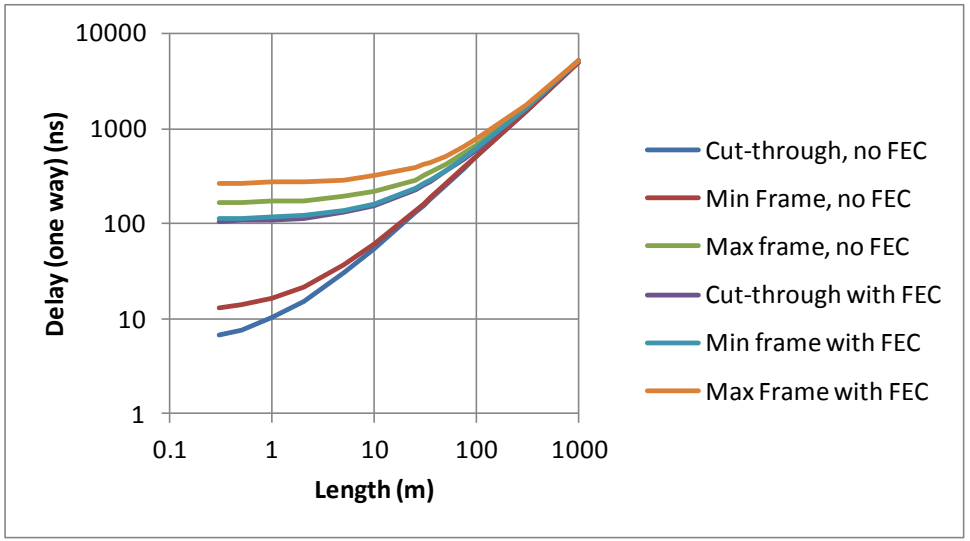
Latency vs. reach

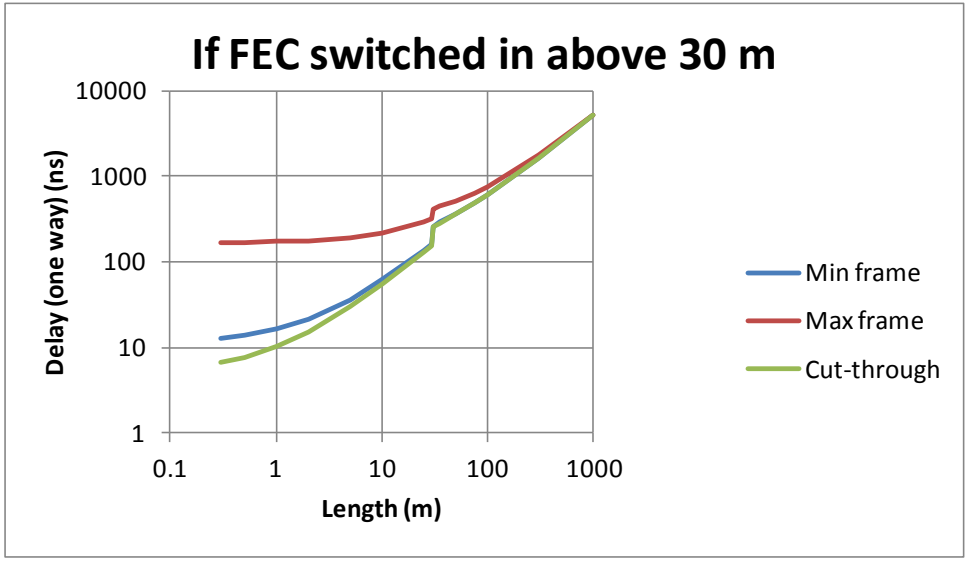
Latency (one way) vs. reach

Contributors are:

- MAC 64+14 bytes to 2000+14 bytes = 6.24 to 20.14 ns
  - Taken as zero for "cut-through switch"
- FEC 90 to 100 ns (one way)
  - per gustlin_01_0112
- Basic 64/66 coding: ~1 block on one lane = 2.56 ns
- 4"+4" host +2"+2" in QSFP+ = 2.44 ns
- E.g. 50 m fibre at 2e8 m/s = 250 ns
- CDRs: say 2 UI x 2 = 0.16 ns

- FEC could be switched off on benign links
  - If power is more important than latency, some CDRs might be switched off before FEC





Power vs. reach

- FEC power 100 mW gustlin_01_0112
- Unretimed module
  - 345 mW/lane => 1380 mW petrilla_01_0112_NG100GOPTX_MMFAdHoc
  - 1500 mW per port sela_01_0112
- 8 CDRs
  - 5 mW/Gb/s? => 1 W/module
  - 345 mW/lane => 1380 mW petrilla_01_0112_NG100GOPTX_MMFAdHoc
- Tx EQ ~30 mW/lane => 120 mW king_01_1111_NG100GOPTX
- Rx fixed peaking or CTLE
  - <30 mW/lane? => <120 mW king_01_1111_NG100GOPTX
  - 50 mW/lane petrilla_01_0112_NG100GOPTX_MMFAdHoc
- Rx DFE 175 mW/lane => 700 mW petrilla_01_0112_NG100GOPTX_MMFAdHoc
- Rx EQ and adaptation
  - ~150-200 mW/lane => 750 mW king_01_1111_NG100GOPTX
  - 350 mW/lane => 1400 mW petrilla_01_0112_NG100GOPTX_MMFAdHoc

Reach is hard to predict today

- Lasers are not as fast as we would wish
  - Speed and spectral width not yet published
- Link model's mode partition noise theory
  1. is an approximation
  2. is not valid with equalisation of the optical link
  3. matters more with OM4 than for previous (OM3) projects
     - king_02_0112_NG100GOPTX_MMFAdHoc shows the uncertainty of point 2 alone
- CDR jitter needs to be factored in
- **A single 100 m objective is missing the point: not the volume market, not known to be correct even for the high end**

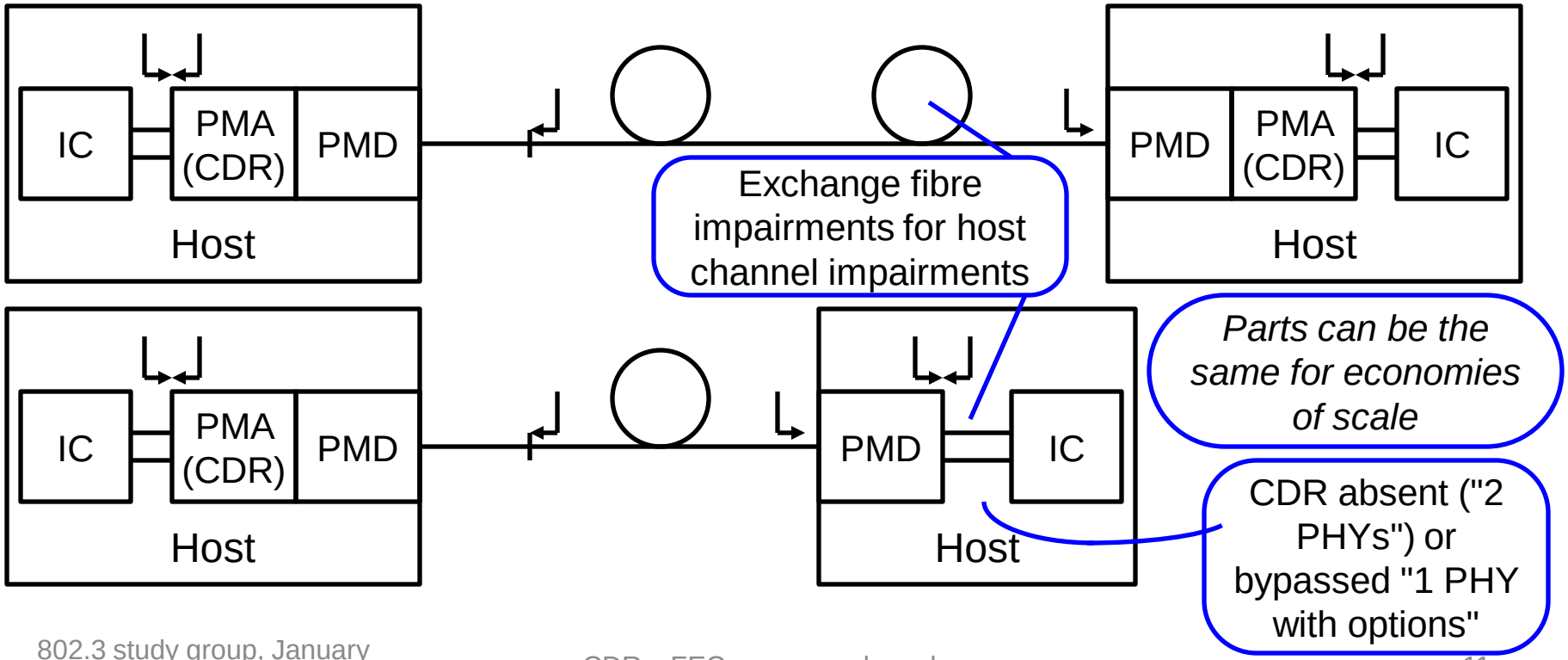
One PHY 2 options, or 2 PHYs?

Precedent for interoperable PHYs:

1000BASE-LX and 1000BASE-LX10

1000BASE-PX20-D with 1000BASE-PX10-U or 1000BASE-PX20-U

10GBASE–PR–U1 with 10GBASE–PR–D1 10GBASE–PR–D2



Compatibility with 100GBASE-CR4

- Must fit in same socket
- Limits power consumption

Conclusions

- Objective creep would cost considerable extra power
- With short links and/or FEC, PHYs with fewer than 8 CDRs are viable
- Two PHY types can be interoperable
  - Unretimed electrical interface "CPPI-4" is worth considering, especially with FEC
    - Both retimed and unretimed electrical interfaces should be part of this project
- Cannot establish actual reaches without more work on jitter, equalisation and MPN
  - Hard reach objectives are premature: should write objectives such as power, QSFP+ compatibility
  - Focus clearly on low cost for the majority of links
    - As well as 75-100 m links
  - Do not repeat the 300 m 10GBASE-SR mistake