Low power 100G MMF PHY and electrical specs

Piers Dawe

IEEE P802.3bm November 2012

Supporters

Oren Sela	Mellanox
Hiroshi Hamano	Fujitsu
Marc Verdiell	Samtec
Kevin Burt	Samtec
Petar Pepeljugoski	IBM
Brad Booth	Dell
Robert Coenen	Reflex Photonics
Phil McClay	TE Connectivity

Contents

- Introduction
- Two MMF PHYs
- Host and module circumstances
- Common electrical connector, variants of electrical interface
- Why unretimed is reasonable with FEC
- Clause 91 FEC
- Strategy for proposed chip-module specs
- Conclusions
- Detailed specs at TP1a and TP4
- References
- Revision history

Introduction

- This presentation builds on September and earlier presentations exploring unretimed links (dawe_01b_0112_NG100GOPTX, dawe_01a_0912_optx), and Clause 91 (P802.3bj, 100GBASE-CR4) FEC always transmitted (petrilla_02a_0912_optx, P802.3bj)
- Addresses the 20 m on MMF objective
- Relates to the "re-timed 4-lane 100G PMA to PMA electrical interfaces for chip to module applications" part objective
- Leverage already paid-for data centre host features
  - FEC, FFE driver, high performance electrical receiver
  - As well as transforming the jitter budget, FEC reduces test time, hence cost
- Allowing unretimed modules for low cost high volume short links
  - Specs for a CPPI-4 function
  - Also allows retimed modules
  - Retimed and unretimed 20 m modules are interoperable (across the fibre)
    - Coordinated specs for a CAUI-4 function
- 20 m PHY and 100 m PHY are interoperable over 20 m MMF
- Hosts that support unretimed modules will support retimed modules

Module > Host \/	Unre-timed	Re-timed
CPPI-4	Y	Y
CAUI-4	X	Y

Two MMF PHYs, 100GBASE-UR4 and 100GBASE-SR4

- Low power, relaxed (MMF) mechanics and low test time => Low cost
- Reduce power in the module by keeping it simple
  - Keep: laser, photodiode, amplifiers
  - Avoid: AGC, equalisers, DFE
  - Optional: CDRs
- 100GBASE-UR4 for the 20 m objective

- For the 100 m objective, 100GBASE-SR4 includes more of these things for more reach on MMF – see other presentations

- 100GBASE-UR4 and 100GBASE-SR4 are interoperable
  - Over the shorter (100GBASE-UR4) MMF channel
  - Like 100GBASE-LR4 and 100GBASE-ER4
  - Like XFP 10GBASE-SR and SFP+ 10GBASE-SR
  - Identical line rate and coding – no Auto-Negotiation

	UR4	SR4
UR4	20 m	20 m
SR4	20 m	100 m
(Showing minimum reaches in objectives)		

What do the modules fit into?

- High density data centre equipment
  - Focus on short reach
  - QSFP28 ports that support 100GBASE-CR4
    - Stacked connectors
  - Clause 91 256b/257b line code and FEC for 100GBASE-CR4
  - Adjustable 3-tap FFE driver
  - Sensitive, adaptive receiver
- General purpose or telecoms-oriented equipment
  - Supports long reach
  - CFP/CFP2/CFP4 ports that support e.g. 100GBASE-LR4, 100GBASE-ER4
  - Clause 82 64B/66B line code, probably no FEC
  - Simple (OIF VSR – no FFE) driver
  - Simple (OIF VSR - CTLE) receiver
    - Some hosts may have OIF SR, MR or LR grade I/O

Thermal budget

- QSFP28 could support 4-lane retimed MMF module
- But lower power than QSFP's 3.5 W max. is essential for fully populated high density cards
  - Require 1.5 W to 2 W per module: see sela_01a_0112 (802.3bj) and dawe_01_0312_NG100GOPTX

- CFP family can support 100GBASE-ER4, so with 100GBASE-UR4 or 100GBASE-SR4, there is thermal budget to spare, allowing FEC in the module
  - Could use a 100GBASE-CR4 PMA/PMD IC as CDR/transcoding/FEC chip
  - (and reduce its power because don't need high output swing or all of its long equaliser)

What goes over the common electrical connector?

- In QSFP, the electrical connector carries:
  - the 4-lane PMD service interface (100GBASE-UR4, unretimed)
  - or a 4-lane PMA service interface (100GBASE-UR4 or -SR4, retimed)
  - just like QSFP/XLAUI/40GBASE-SR4 or QSFP/XLPPI/40GBASE-SR4 with 10G lanes
    - XLPPI is the 4-lane version of nPPI
- In CFP, the electrical connector carries:
  - a 10-lane CAUI PMA service interface (100GBASE-UR4 or -SR4, retimed)
  - as for 100GBASE-LR4 or -ER4
- In CFP4, the electrical connector could carry a retimed 4-lane CAUI-4 PMA service interface (100GBASE-UR4 or -SR4), as for 100GBASE-LR4, -ER4
- or, as CFP4 is allowed for 100GBASE-CR4, some implementations could be unretimed with FEC in the host, as QSFP

Family of electrical chip-module interfaces

	Not relying on FEC	Relying on FEC
Retiming in module ("CAUI-4")	A Moderate	B Easiest, more power
Not retiming in module ("CPPI-4")	C Difficult, less power	D Moderate
Passive copper cable		100GBASE-CR4, different

- It looks like there is really a family of electrical interfaces for consideration
- Not yet clear what the cost of using spec A for application B is
- C is controversial
- A, B, D, C? need to be compatible
  - Also need compatibility with 100GBASE-CR4, but very different channel leads to different electrical specs
- Look for common spec items for A and D
- Strategy: address the hardest ones (D, maybe C) first, then we can dumb them down to create the easy ones (A and B), assuring compatibility

Unretimed electrical spec has to be chosen to be suitable for unretimed optical spec

- Remember SFP+'s difficulties, grafting an unretimed electrical spec onto a difficult, frozen, optical spec
- We should investigate the unretimed specs (D) immediately, with the 20 m optical specs
- Retimed electrical spec relying on FEC (B) is easier – less constraints
  - Can be adjusted for commonality with unretimed electrical spec
- Retimed electrical spec without FEC (A) must also be compatible
  - With unretimed
  - Preferably, with OIF VSR
  - Can be adjusted for commonality with other electrical specs
- All must be compatible with 100GBASE-SR4
  - E.g. crosstalk: avoid excessive amplitude and edge speed

- Next two slides show that unretimed is reasonable with FEC, and slide 14 compares unretimed ("CPPI-4") and retimed ("CAUI-4") proposals with nPPI and OIF VSR

Why unretimed is reasonable with FEC 1/2
Example spreadsheet link model jitter budgets

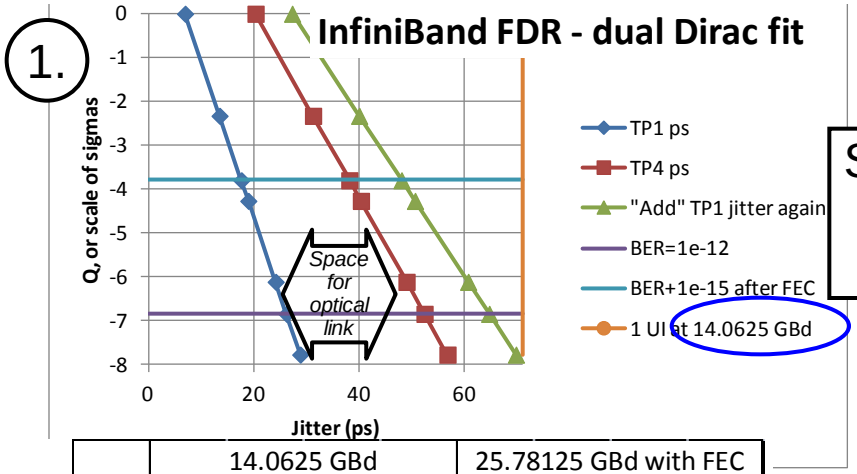
- From Jonathan King, using an enhanced spreadsheet link model similar to ExampleMMF_LinkModel_12_09_18.xlsx (see petrilla_01a_0912_optx)

Parameter	Retimed, FEC	Un-retimed, FEC
Tx_OMA min at max TDP	-3.2 dBm	
RMS spectral width	0.6 nm	
Rise fall time	20.75 ps	
$RIN_{12}OMA$	-130 dB/Hz	
Receiver noise limited sensitivity, BER =10^{-12} BER =$5x10^{-5}$	-7.8 dBm (-10.5 dBm)	
TP1: J2, J9	0.08, 0.23 UI	0.17, 0.29 UI
TP4: TJ at $5x10^{-5}$	0.7 UI	0.61 UI
OM4 distance	120 m	30 m

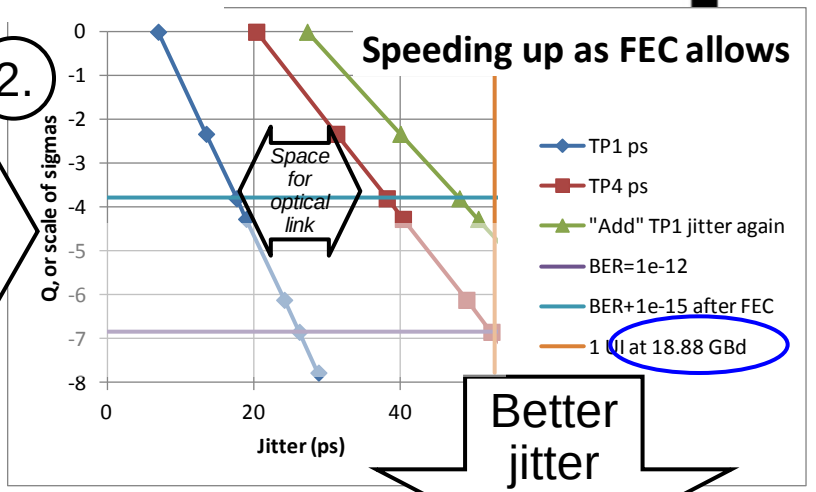
Similar Tx, Rx values as 100G -SR4

For similar Tx and Rx values being proposed for 100GBASE-SR4, link model analysis shows that TP1 and TP4 jitter specs comparable to 802.3ba (nPPI) may be possible

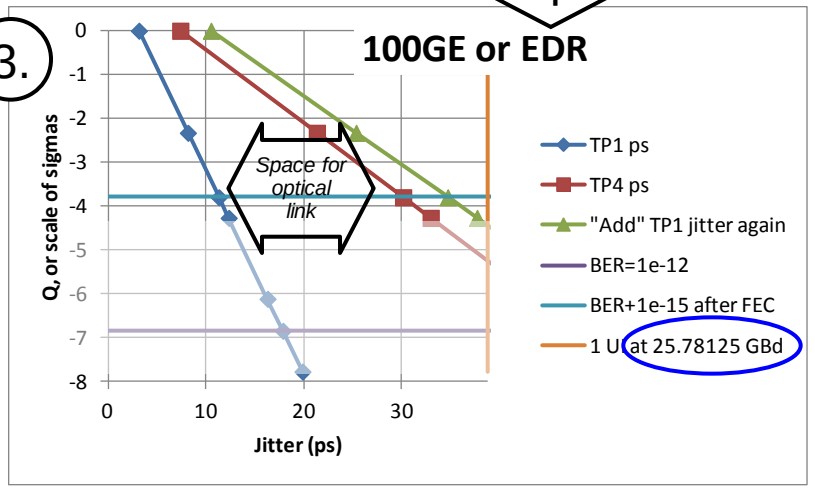
Why unretimed is reasonable with FEC 2/2 Dual Dirac analysis



Same jitter in ps



Better jitter in ps



	14.0625 GBd				25.78125 GBd with FEC			
	TP1		TP4		TP1		TP4	
	UI	ps	UI	ps	UI	ps	UI	ps
J2	0.19	13.5	0.44	31.3	0.21	8.1	0.55	21.3
J5	n/a				0.32	12.3	0.85	33.0
J9	0.34	24.2	0.69	49.1	0.42	16.3	n/a	n/a

Compare nPPI J2 0.17, 0.42

For retimed non-FEC use?*

TP4 jitter seen through reference equaliser e.g. OIF's CTLE

1. Unretimed 14G today
2. Using FEC lets us run at 19G for same jitter as 14G in ps, AND move delivered BER from 1e-12 to 1e-15 while PMA and PMD below FEC run at ~1e-5
3. 37% faster line rate required – e.g. with jitter limits in table

* Compare OIF VSR: EW15 min 0.46 UI

Not shown: faster rates need faster silicon – feasible

For FDR spec (which is for active optical cable) see references

Host loss: FDR 5.26 dB, -CR4 6.81 dB

Clause 91 RS(528, 514) FEC

- Clause 91 has two Reed-Solomon FEC options
- RS(528, 514) is used for 100GBASE-KR4 and 100GBASE-CR4
- Choose RS(528, 514) which runs at the same line rate as 64B/66B
- At Tx, 100GBASE-CR4 has FEC transmitted always, at Rx FEC-checked always, FEC-corrected by default
- FEC is allowed 409.6 ns latency, or the same as ~80 m of fibre
  - An implementation can be significantly better
  - Latency can be further reduced by implementing the "PCS" and the "FEC" in one IC
- Assuming that 5.28e-5 => 1e-12 (coding gain is ~5.2 dB)
  - Slide 12 assumes 1.84e-5
- By the way, FEC is desirable for 100GBASE-SR4 (100 m objective) as well as 100GBASE-CR4 and 100GBASE-UR4 because it mitigates MPN and RIN well
- FEC power is remarkably low (if in the host ASIC)

Strategy for proposed chip-module specs for 100GBASE-UR4, SR4 Unretimed and retimed

1. VSR does not use host's FFE, eye can be closed at connector/TP1a. nPPI has open eye at connector/TP1a
- Proposal follows nPPI, using host's FFE. Host knows its channel loss, eye is open at connector/TP1a

2. VSR does not plan for mix of copper and optical ports. nPPI did, somewhat. IB FDR went further, with much lower signal swings. Also reduces power.
- Proposal follows IB FDR

3. nPPI does not assume any equalisation ability in host Rx – "leaves performance on the table". VSR assumes a CTLE but sets a very narrow range for module's electrical output state of emphasis
- Proposal builds on VSR or XFP, allows a wider range for module's electrical output state of emphasis, expects host will adapt to it. Use software equaliser or equivalent for spec at TP4. 100GBASE-UR4 or 100GBASE-SR4, using FEC, don't need VSR's extrapolation
  - Retimed interface for 100GBASE-LR4 or 100GBASE-ER4 still would

- No Auto-Negotiation, no need for a Training phase
  - Unlike 100GBASE-CR4, host Rx knows loss between source of electrical signal and itself
- See two slides after conclusions for possible detailed specs at TP1a and TP4

Conclusions

- Leverage already paid-for data centre host features
  - FEC, FFE driver, high performance electrical receiver
    - FEC is always transmitted, like 100GBASE-CR4
    - FEC enables 100 m+ reach or unretimed modules, and reduces test time, hence cost
    - Use host FFE to open the eye at TP1
    - Assume host Rx is adaptive
- Power in the module is a cost
  - Specify for unretimed for high volume short links
- Retimed and unretimed are interoperable
  - UR4 and SR4 can be connected over UR4 (short) MMF
  - Retimed module will work in non-retimed host
  - SR4 is retimed
  - UR4 can be retimed or not

Straw man electrical specifications follow

- These are as shown in September (dawe_01a_0912_optx) with additional material from Cu ad hoc
  - There are also straw man optical specifications in dawe_01a_0912_optx

Table xA-1 CPPI-4 host electrical output specifications at TP1a

Table xA-1 CPPI-4 host electrical output specifications at TP1a

	nPPI				CPPI-4		CAUI-4		CEI-28G-VSR		latchman_01_102	
Parameter description	**Min**	**Max**	**Units**	**Condition**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**		
BER	n/a				n/a			1e-6? and 1e-13?		1e-15?		
Single ended output voltage	–0.3	4	V	Referred	–0.3	4	–0.3	4				
Common Mode Voltage common mode voltage is generated by host			V						-0.3	2.8		
Differential peak-to-peak output voltage (max) with Tx disabled			mV									35
AC common-mode output voltage	—	15	mV	RMS	—	20	—	20		17.5		17.5
Termination mismatch at 1 MHz	—	5	%		*(n/a)*		*(n/a)*			*10*		
Differential output return loss	See 86A.4.1.1	—	dB		Eqn. A-1	—	Eqn. A-1	—		Eqn. 1-2		
Common-mode output return loss	See 86A.4.1.2	—	dB		Eqn. A-2	—	Eqn. A-2	—				
Common-mode to differential output return loss					Eqn. A-3	—	Eqn. A-3	—		Eqn 1-3		
Output transition time, 20% to 80%	28	—	ps		Around 10 TBD		Around 10 TBD		10		10	
J2 Jitter output	—	0.17	UI		—	0.19	—	0.45				
J9 Jitter output	—	0.29	UI		—	0.31	—	0.63				
Data Dependent Pulse Width Shrinkage (DDPW	—	0.07	UI		—	0.10	*(n/a)*					
Equalized J2 Jitter output					—	0.10	—	0.3				
Equalized J9 Jitter output					—	0.22	—	0.52				
Equalized DDPWS					—	0.05	*(n/a)*					
Eye width at 10-15 probability (EW15)1			UI			see J9		see J9	0.46	at 10^-15	0.46	at 10^-12
Eye height at 10-15 probability (EH15)1			mV						100		100	at 10^-13
Q_{sq} for XLPPI	45	—	V/V		45	—	45	—				
Q_{sq} for CPPI	43	—	V/V				*(n/a)*					
	Specification values											
Eye mask coordinates: X1, X2 Y1, Y2	0.11, 0.31 95, 350		UI mV	Hit ratio = 5 ×	0.13, 0.33 95, 350		0.24, 0.45 95, 350		-, -, -, 450		-, -, -, 450	
Crosstalk source VMA, each input lane	700		mV	At TP4	470		470					
Crosstalk source transition times, 20% to 80%	34		ps	At TP4	Around 8 to 10 TBD		Around 8 to 10 TBD					

1.Open eye is generated through the use of a reference Continuous Time Linear Equalizer (CTLE) Use compliance board methodology, and observation bandwidth 25 to 33 GHz TBD

Table xA–3 CPPI-4 module electrical output specifications at TP4

Table xA–3 CPPI-4 module electrical output specifications at TP4

	nPPI				CPPI-4		CAUI-4		CEI-28G-VSR		latchman_01_10	
Parameter description	**Min**	**Max**	**Units**	**Condition**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**		
BER	n/a				n/a			1e-6? and 1e-13?		1.E-15		1.E-12
Single ended output voltage tolerance	–0.3	4	V	Referred	–0.3	4	–0.3	4				
Common Mode Voltage common mode voltage is generated by host									-0.3	2.8		
AC common-mode output voltage (RMS)	—	7.5	mV		—	17.5	—	17.5		17.5		17.5
Termination mismatch at 1 MHz	—	5	%		*(n/a)*		*(n/a)*			10		10
Differential output return loss	ee 86A.4.2.	—	dB	10 MHz to	Eqn. A-1	—	Eqn. A-1	—		Eqn 1-2		
Common-mode output return loss	ee 86A.4.2.	—	dB	10 MHz to	Eqn. A-2	—	Eqn. A-2	—				
Common-mode to differential output return loss					Eqn. A-3	—	Eqn. A-3	—		Eqn 1-3		
Output transition time, 20% to 80%	28	—	ps		Around 8 to 10 TBD		Around 8 to 10 TBD		9.5		9.5	
J2 Jitter output	—	0.42	UI		—	0.6	—	0.42				
J9 Jitter output	—	0.65	UI		*(n/a)*		—	0.6				
Equalized J2 Jitter output (when used without FEC)					—	0.5	—	0.28				
Equalized J9 Jitter output (when used without FEC)					—	0.7	—	0.5				
Equalized J2 Jitter output (when used with FEC)					—	0.5	*(n/a)*					
Equalized **J5** Jitter output (when used with FEC)					—	0.7	*(n/a)*					
	Specification values											
Eye mask coordinates: X1, X2 Y1, Y2	0.29, 0.5 150, 425		UI mV	Hit ratio = 5 × 10–5	Around 0.45, 0.5 40, 250		Around 0.22, 0.43 50, 250		-, -, -, 450		-, -, -, 450	
Eye width at 10-15 probability (EW15)									0.57	at 1e-15	0.57	at 1e-12
Eye height at 10-15 probability (EH15)									240	at 1e-15	240	at 1e-12
Crosstalk source VMA, each lane	700		mV	At TP1a	660		660					
Crosstalk source transition times, 20% to 80%	37		ps	At TP1a	Around 10 TBD		Around 10 TBD					

References

- Piers Dawe, CDRs, FEC, power and reach, http://ieee802.org/3/100GNGOPTX/public/jan12/dawe_01b_0112_NG100GOPTX.pdf
- Piers Dawe, Coordinated proposal for 100GBASE-SR4, 100GBASE-UR4, CPPI-4 and CAUI-4 http://ieee802.org/3/bm/public/sep12/dawe_01a_0912_optx.pdf
- John Petrilla, 100G SR4 & RS(528, 514, 7, 10) FEC, http://ieee802.org/3/bm/public/sep12/petrilla_03a_0912_optx.pdf
- Oren Sela et al., Power Considerations for 100GBASE-KR4 PHY, http://ieee802.org/3/bj/public/jan12/sela_01a_0112.pdf
- Piers Dawe et al., Low power MMF objective for High Performance Computing and End-of-Row applications, http://ieee802.org/3/100GNGOPTX/public/mar12/plenary/dawe_01_0312_NG100GOPTX.pdf
- John Petrilla, 10GE Link Model Extensions, http://ieee802.org/3/bm/public/sep12/petrilla_01a_0912_optx.pdf and http://ieee802.org/3/bm/public/sep12/ExampleMMF_LinkModel_%2012_09_18.xlsx
- InfiniBand™ Architecture Specification Volume 2, Release 1.3, November 2012 http://www.infinibandta.org/
- Ryan Latchman, CAUI-4 Consensus Building, Specification Discussion, latchman_01_1022_caui4.pdf

Revision history

Slide	Change
2	Updated supporters list
4, 5	Mini tables added
5	Example: 100GBASE-LR4 and 100GBASE-ER4
6	Stating which FEC
8, 9	Terminology corrections (XLAUI, XLPPI, CPPI)
10	Corrected cross-references
12	"receiver front end runs" -> "PMA and PMD below FEC run"
13	Stating which FEC
14	Clarification of applicability: retimed, FEC, extrapolation
19	Updated last reference and added URL
20	Added revision history slide