C2M CDAUI-8: considerations and proposals on P802.3bs 400 Gb/s Ethernet draft.

Authors and supporters

Authors

Marco Mazzini, Cisco
Rick Brooks, Cisco
Upen Reddy Kareti, Cisco
Joel Goergen, Cisco

Supporters

Gary Nicholl, Cisco
Stephane Dallaire, Inphi
Benjamin Smith, Inphi
Pirooz Tooyserkani, Cisco
Vineet Salunke, Cisco
Jane Lim, Cisco
Kelvin Qiu, Cisco
Elizabeth Kochuparambil, Cisco
Ken Ly, Cisco

Overview

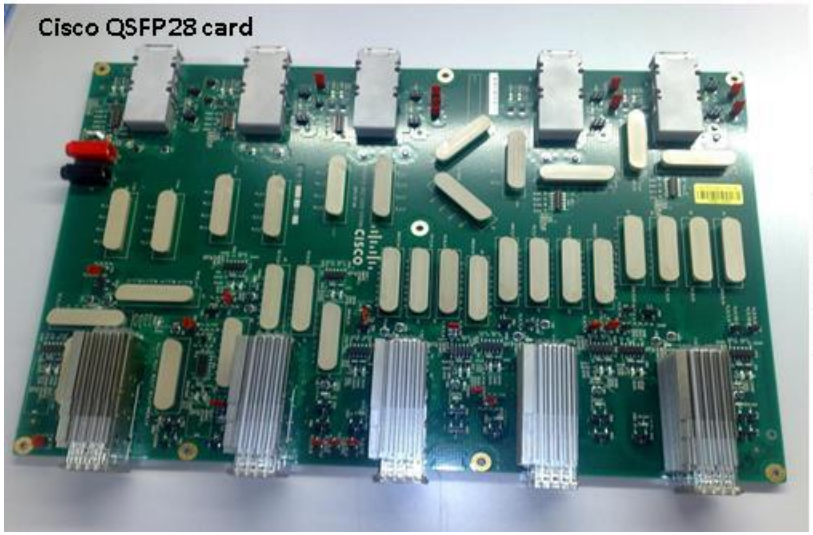
- P802.3bs 400 Gb/s Ethernet Draft 1.0 under review.
- This presentation refers also to some comments submitted by G.Nicholl (#185), P. Tooyserkani (#187, #188) and J.Maki (#191, #192)

Current draft defines a reference receiver with a continuous time linear equalizer (CTLE, two-poles, one-zero) to measure Eye Width and Eye Height at 1E-6 BER confidence.

Some contributions during ad-hoc electrical meetings were pointing to change some of the TX and RX characteristics.
Some options for improvement were indicated into mazzini_01_082415_elect (slide 19).

After more recent results and discussions, here we present some proposals to improve channel characteristics, besides re-state recommendations for CDAUI-8 C2M link robustness and tests improvement.

Our CDAUI-8 (CEI-56G-VSR) reference: Cisco QSFP28 card.



QSFP28 test card

Length	Cal trace	QSFP28 SMT	QSFP28 stacked
1.6	yes	yes	yes
2	yes	yes	yes
3		yes	yes
4	yes	yes	yes
5	yes	yes	yes

Cisco contributed sharing some of the QSFP28 test card (TU872SLK material, RTF copper) parameters, which were made available into http://www.ieee802.org/3/bs/public/adhoc/elect/index.shtml.

S-parameters are relative to 3, 4 and 5” channels stacked connector design, and are terminated on the other side with MXP 1x8 connectors (s2p of this available too, 3” and 4” channels characteristics shown into slide 6).

With respect to channels shared into mazzini_01_082415_elect these shown better ILD and loss characteristics.

1 – Improve Reference RX equalizer: from mazzini_01_082415_elect (Mazzini, Brooks, Reddy Kareti) and further work (Brooks).

Further work done over these new channels still showing that there's the need to improve Draft 1.0 RX reference equalizer.

Both experiments and simulations shown 2p1z module's RX CTLE it's not enough to close the link at 1E-6 BER.

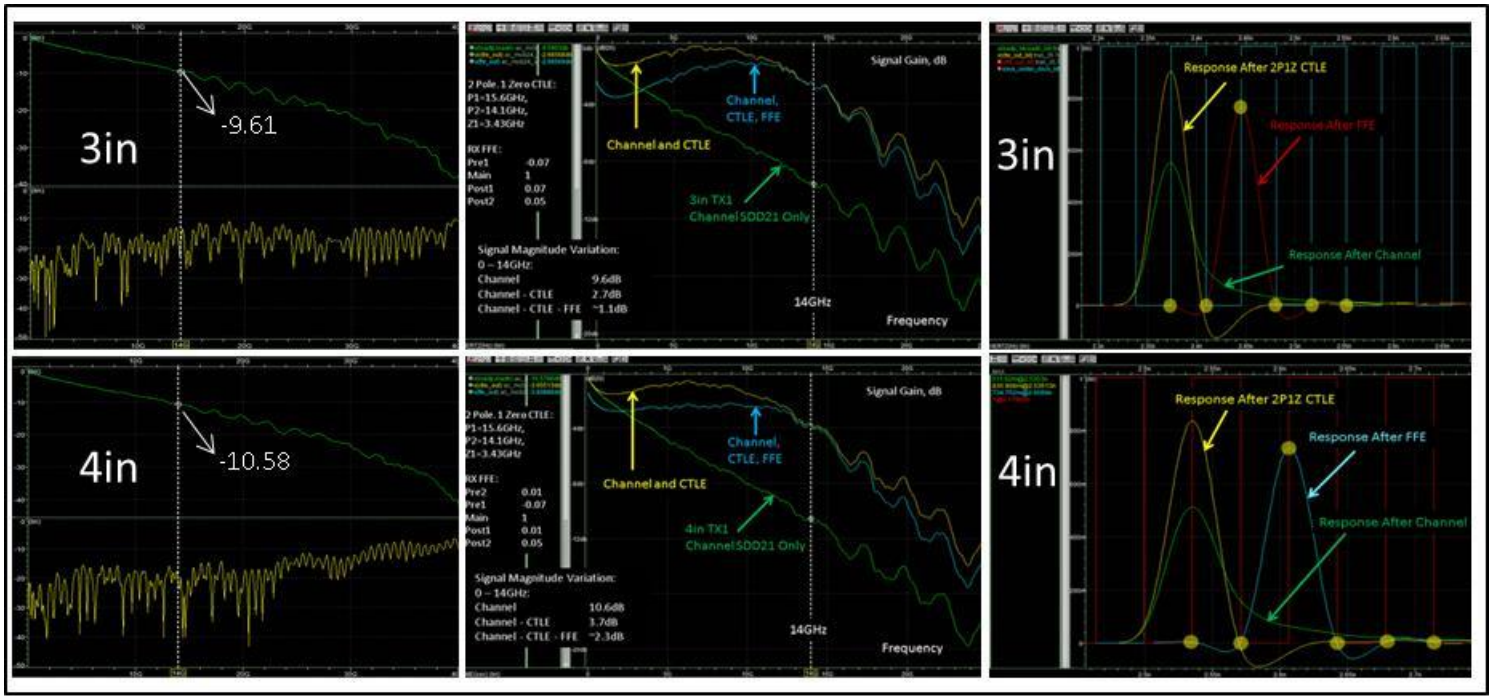
- **Simulation showing that even a 3p2z RX CTLE equalizer can be tight for 1E-6 BER requirements.**
- **The PAM-4 EQ is much more critical than NRZ for the same symbol time.**
  - **Adjacent Eyes ➔ Very Little Over or Under EQ Allowed!**
  - **Channel variations: worse impact than NRZ ➔ Need of slow (but continuous) adaptation.**

Investigation on long data pattern over 5" Channel:

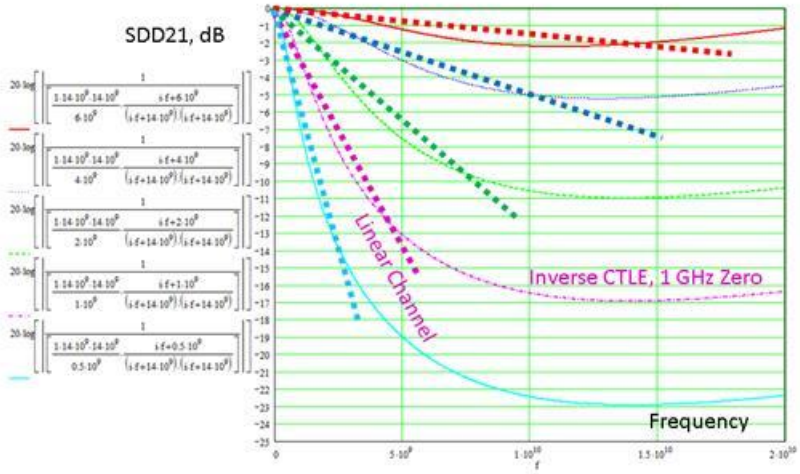
- Benefit of adding a Low Frequency De-emphasis Filter to Correct the Step Response
  Best Case: No RJ, DJ, Noise, Crosstalk.

- **A Low Frequency filter is needed to help equalize long data patterns.**
  - **LF time constant: 0.1ns to 0.5ns for C2M**
  - **LF De-emphasis amount: 0 to 25% for C2M**

1 – Improve Reference RX equalizer: from mazzini_01_082415_elect (Mazzini, Brooks, Reddy Kareti) and further work (Brooks).

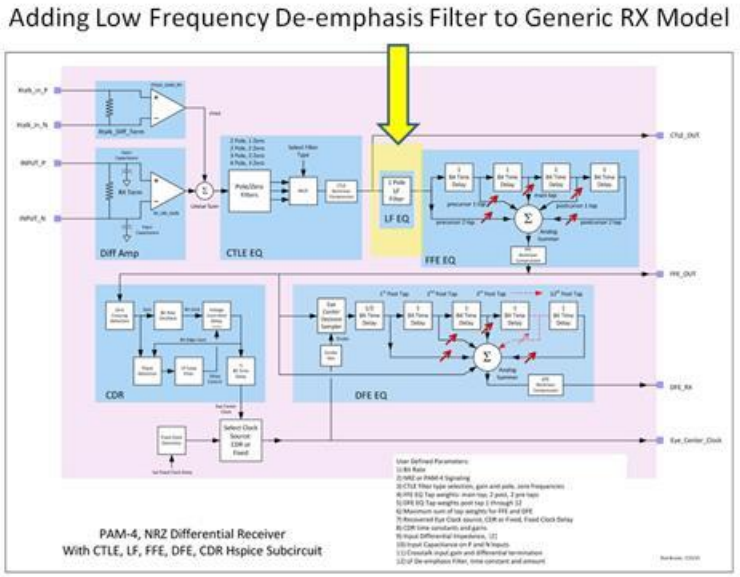


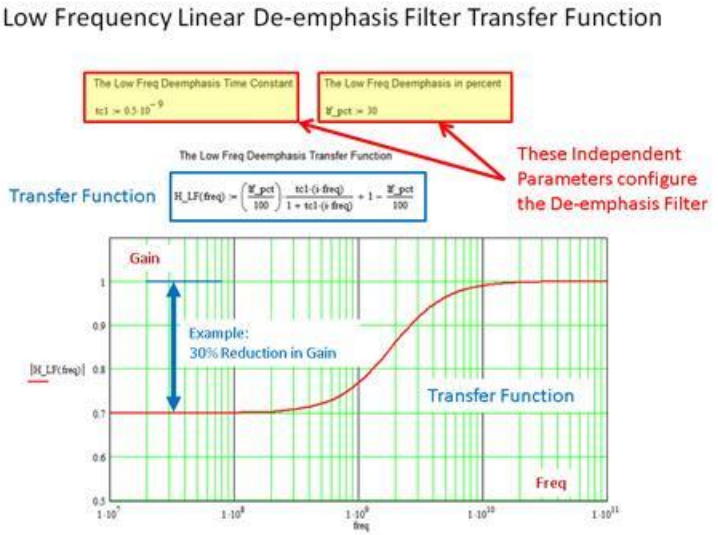
2p1z CTLE seems not enough to ensure the correct target BER over any of the QSFP28 card tested cases.

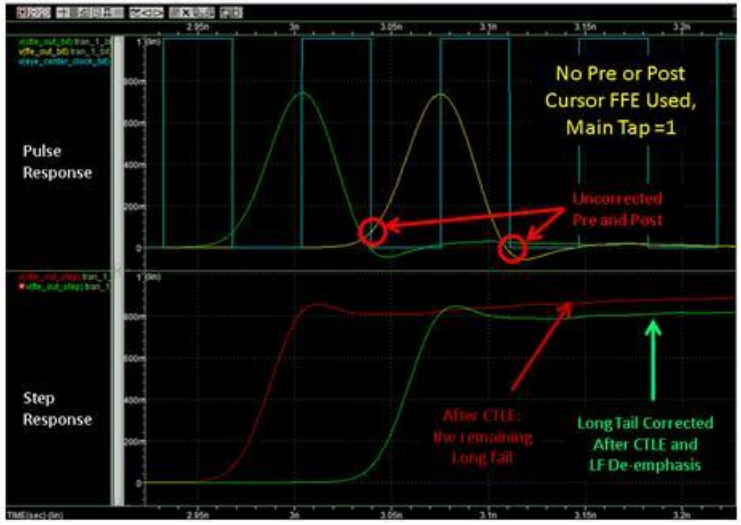


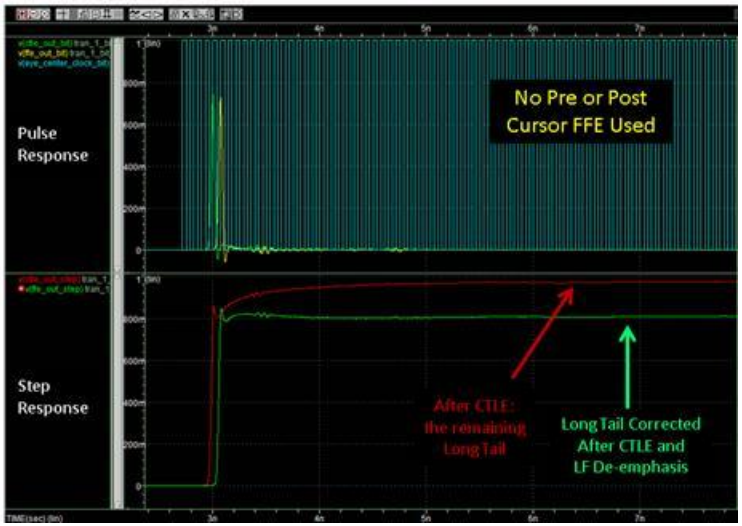
- **Channels are Roughly Linear on a dB vs Linear Frequency Plot.**
- **With CTLE Alone, Channels can be Equalized only up to a certain frequency.**

2 – Add LF filter: from mazzini_01_082415_elect: investigation on long data pattern over 5” Ch. (Brooks).



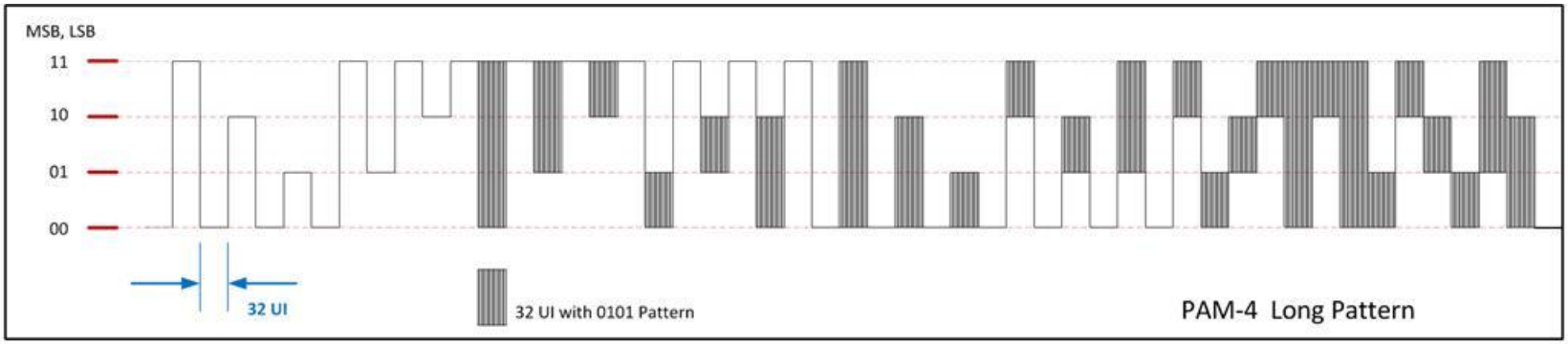




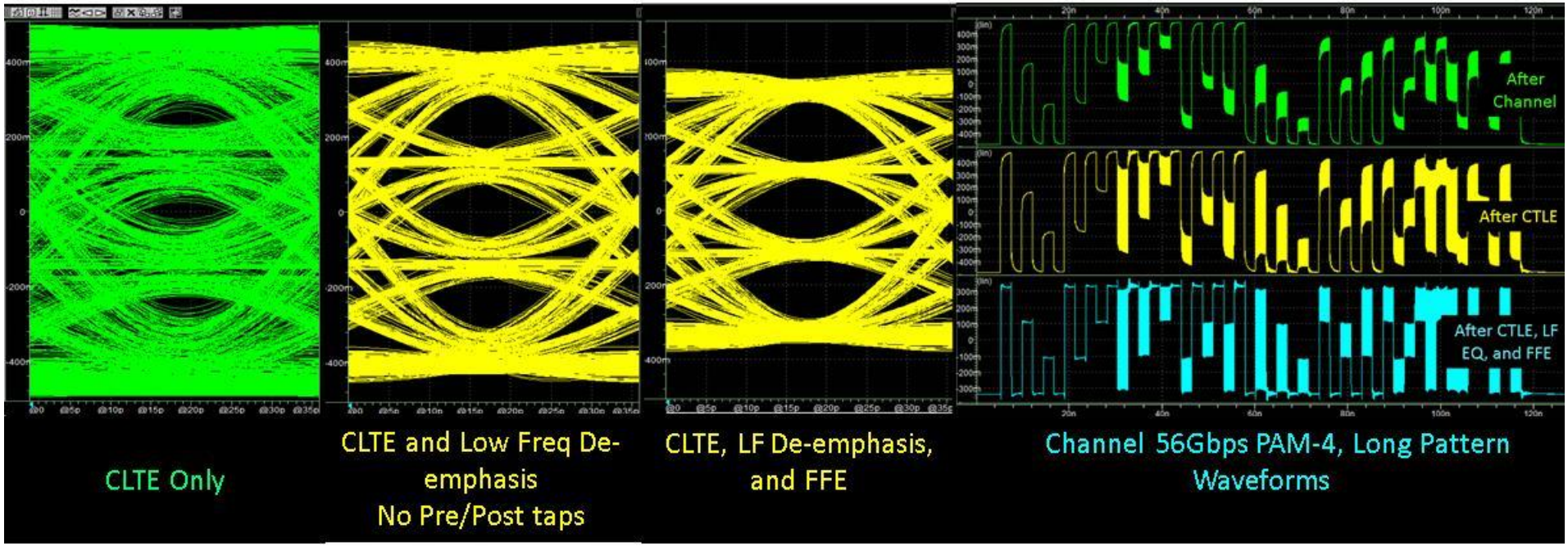


5” Channel Pulse Response – CTLE, Low Freq De-emphasis, No FFE

2 – Add LF filter: from mazzini_01_082415_elect: investigation on long data pattern over 5” Ch. (Brooks).



- In typical applications, Serdes run scrambled data patterns of 64B/66B encoded data.
- Long run lengths of zeros or ones are possible, surpassing 32 UI in length at times.
- The proposed data pattern uses 1600 UI total to allow simulation of possible wide. bandwidth signals with up to 32 UI duration of the same bit values (LSB and MSB).
- An alternate long pattern is to used 64UI lengths instead of 32UI lengths.

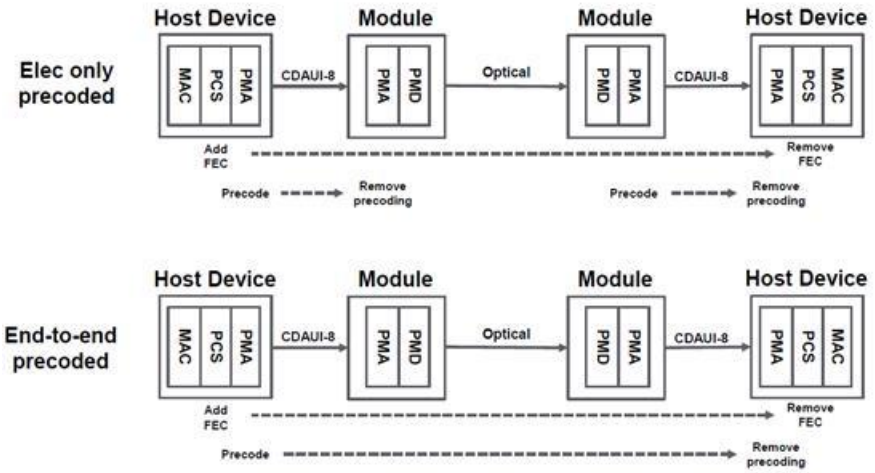


3 – Allow BER relaxation on CDAUI-8 C2M.

According to Pete Anslow's work (anslow_01_0815_logic), is would be possible to investigate if the BER of the electrical sub-link for a penalty of ≈ 0.1dB optical in the optical sub-link can be relaxed from 1E-6?

Precoded multi-part links analysed

Two different schemes for precoded links have been analysed in the next slide:



Equivalent to BER of 1E-13

	At slicer output for FLR = 6.2E-11			
	Total electrical		Optical	
Same cwd (1), a = 0.75	Burst	2.9E-7*	Random	2.4E-4
Same cwd, symbol interleave (2), a = 0.75	Burst	7.5E-7*	Random	2.4E-4
Same cwd (1), a = 0.5	Burst	1.6E-5*	Random	2.4E-4
1:4 Pre-interleaved (4), a=0.75	Burst	2.2E-5*	Random	2.4E-4
1:2 Pre-interleaved (8), a=0.75	Burst	3.5E-5*	Random	2.4E-4
Diff cwd (FOM) (7), a = 0.75	Burst	4E-5*	Random	2.4E-4
Same cwd elec only precoded, a=0.75	Burst	5.1E-5*	Random	2.4E-4
Same cwd end-to-end precoded, a=0.75	Burst	6.9E-5*	Random	4.9E-5
1:4 Pre-interleaved (6), a=0.75	Burst	5.7E-5*	Random	2.4E-4
1:2 Pre-int, sym mux (10), a=0.75	Burst	7.6E-5*	Random	2.4E-4
1:4 Pre-int, sym mux (9), a=0.75	Burst	1E-4*	Random	2.4E-4
Random errors	Random	8.2E-5	Random	2.4E-4

Note – these values are the BER **including** the additional errors due to the bursts. To account for burst errors, the values marked with "*" have been multiplied by 4 when a = 0.75 and 2 when a = 0.5.

For DFE-less links (like the current proposed C2M), should be a correct assumption to say we should be closer to the random error model rather then the burst one? If yes, the BER of the CDAUI-8 could be more like 1E-5, which is still with good margins with respect the 8.2E-5 bugdet for the cascade of electrical links in random errors environment.
Which would be the total electrical for C2C-C2M to C2M-C2C (four electrical and one optical links) case ?

4 - Improve TX SNDR and allow “coarse” or “fixed” host TX pre-cursor tap (dallaire_01_090415).

CHANNEL	FEXT	NEXT	IL @ 13.28125 GHz (dB)	ILD (dBrms)
From IEEE 802.3bs shanbhag_3bs_14_0823:				
(1) Nelco 4000-13SI Host PCB + next gen 28Gb/s high density SMT IO	5	0	8.7	0.110
(2) EM-888 Host PCB + next gen 28Gb/s press-fit stacked IO	7	0	8.9	0.051
From IEEE 802.3bs shanbhag_3bs_01_1014:				
(3) 4in Megtron6 Host PCB + next gen 28Gb/s high density SMT IO	5	0	4.3	0.110
(4) 10in Megtron6 Host PCB + next gen 28Gb/s high density SMT IO	5	0	8.8	0.106
(5) 4in Megtron6 Host PCB + next gen 28Gb/s press-fit stacked IO	7	0	4.5	0.051
(6) 10in Megtron6 Host PCB + next gen 28Gb/s press-fit stacked IO	7	0	9.0	0.052
Cisco Channels:				
(7) Cisco 2in Stacked	0	0	8.5	0.237
(8) Cisco 5in Stacked	0	0	11.3	0.245

Different channels were investigated (7, 8 being older Cisco channels too).

An Improved Reference RX/TX

- The following (crudely) improved reference RX/TX provides nearly all of the gain:

TX FIR	LFEQ: (Z1,P1) (GHz)	CTLE: (Z1,P1,P2) (GHz)
[-0.05,0.95]	(1,1.2)	(8.31,14.1,18.6)
[-0.05,0.95]	(1,1.2)	(7.10,14.1,18.6)
[-0.05,0.95]	(1,1.2)	(5.68,14.1,15.6)
[-0.05,0.95]	(1,1.2)	(4.98,14.1,15.6)
[-0.1,0.9]	(1,1.2)	(4.35,14.1,15.6)
[-0.1,0.9]	(1,1.2)	(3.82,14.1,15.6)
[-0.1,0.9]	(1,1.2)	(3.43,14.1,15.6)
[-0.1,0.9]	(1,1.2)	(3.00,14.1,15.6)
[-0.1,0.9]	(1,1.2)	(2.67,14.1,15.6)

Channel	1	2	3	4	5	6	7	8
CTLE	-0.07	-0.04	1.01	-0.45	1.24	-0.13	-1.37	-2.65
CTLE + TXFIR	1.47	1.53	1.43	0.84	2.08	1.35	0.84	0.55
CTLE + LFEQ	0.45	0.50	1.39	-0.14	1.92	0.27	-1.37	-2.49
CTLE + TXFIR + LFEQ	2.26	2.50	2.13	1.28	2.95	2.14	1.43	0.84
Reference RX/TX	2.22	2.47	2.13	1.28	2.95	2.14	1.18	0.19

- The degradation on channels 7 and 8 is due to insufficient pre-cursor equalization in the reference TX FIR

Analysis run assuming TX SNDR = 29dB shown margins defining a reference RX/TX, with a fixed TX FIR improves margins (Note: in terms of COM new Cisco channels are close to highligthed cases 1, 3 and 4 for the 3, 4 and 5 inches channels respectively).

But, since the CDAUI-8 C2M has to be a symmetrical channel, we have also to limit module’s power consumption.
For 400GE operation then, the eight reference 2-tap TX FIR inside the module towards the host should be turn on only if strictly needed.

- Improve CDAUI-8 C2M link robustness – Definition of continuos adaptation of module's CTLE RX.

- Currently C2M CDAUI-8 requires "Module receiver to be self-adaptive and autonomous".
- This keeps open the opportunity to develop "adaptive at start-up", but then "static" module's receivers, which we believe can have hard life to compensate CDAUI-8 channel variations.
  - Note that this item was already addressed during CAUI-4 definition with a couple of contributions from Cisco (see mazzini_01_042414_caui and alessandro_01_07032014_caui).
- **Since we do expect that channel variations should have worse impact than NRZ on PAM4, we believe there'll be the need of a slow (but continuous) adaptation of the CTLE receiver.**
  - For this CDAUI-8 should include a mention to «continuos» adaptation.
  - Our NEBS/DVT tests assume 2C/min temperature ramp variation, so we should cover these variations with the proposed adaptation method.
  - Our first estimation is +/- 1dB variation with respect to the previous CTLE gain setting with a minimum 1Hz frequency rate, but we're open to discussion.
- A further benefit would be to allow RX thresholds to optimize continuously.

- Improve CDAUI-8 C2M test – Define Vertical eye closure (VEC) at TP1a.

120E.3.1 CDAUI-8 host output characteristics

A CDAUI-8 host output shall meet the specifications defined in Table 120E–1 if measured at TP1a.

Table 120E–1—CDAUI-8 host output characteristics (at TP1a)

Parameter	Reference	Value	Units
Eye width[a] (min)	120E.4.2	0.25	UI
Eye height A[b], differential (min)	120E.4.2	50	mV

[a]All 3 PAM4 eyes, at 10^{-6} BER
[b]All 3 PAM4 eyes, at 10^{-6} BER

Table 120E–6—CDAUI-8 module input characteristics

Parameter	Reference	Test point	Value	Units
Signaling rate per lane (range)	120E.3.1.1	TP1	26.5625 ± 100 ppm	GBd
Differential pk-pk input voltage tolerance	120E.3.1.2	TP1a	900	mV
Module stressed input test[a]	120E.3.4.1	TP1a	See 120E.3.4.1	

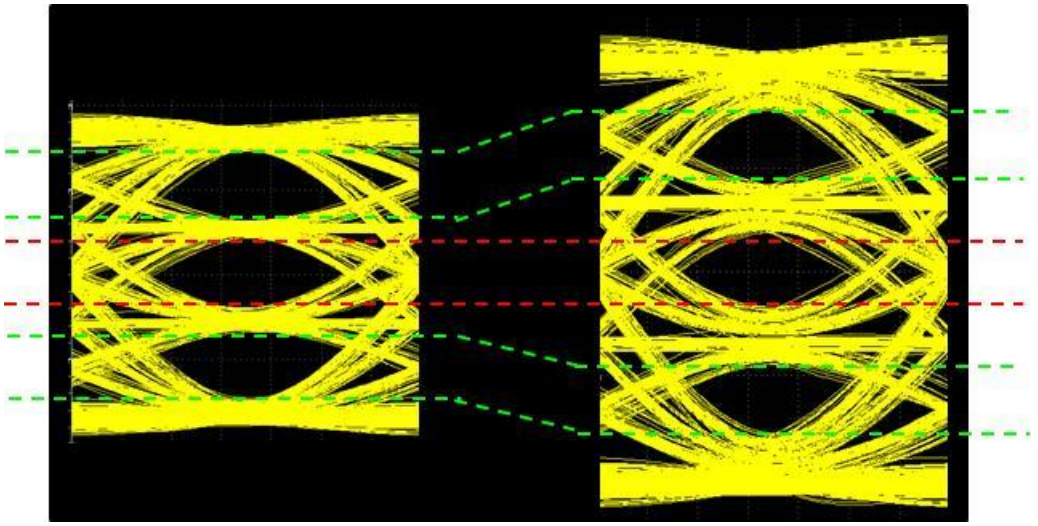
Random jitter and the pattern generator output amplitude are adjusted (without exceeding the differential pk-pk input voltage tolerance specification as shown in Table 120E–6) to result in the eye height and eye width given in Table 120E–7 using the reference receiver with the setting of the CTLE that maximizes the product of eye height and eye width.

Vertical eye closure is calculated using Equation (120E–2).

$$VEC = 20\log_{10}\left(min\left[\frac{AVupp}{Vupp}, \frac{AVmid}{Vmid}, \frac{AVlow}{Vlow}\right]\right)$$

Same EW, EH can be achieved with different VEC at TP1a.

Define a maximum VEC at TP1a too would help to constrain host output and calibrate the TP1 input stressor to the CDAUI-8 module.



SUMMARY – proposals for CDAUI-8 C2M, link robustness and tests improvements.

Proposals for spec improvement.

- (Items 1-2) Improve Reference receiver: Draft 1.0 CTLE (2p1z) + LFEQ (1p1z, divided by xx).
  - We are open to agree the frequency range of the LF filter based on contributions.
- (Item 3) Verify if possible to relax BER requirement from 1E-6 to 1E-5.
- (Item 4) Define an informative minimum TP0a SNDR of 29dB (TP0a only measurable point).
- Introduce a channel ILD deviation (#187,#188) mask requirement or limit ILD value – we propose 0.25dBrms (further validation is needed).
- (Item 4) Allow an optional (not mandatory) "fixed and universal" TX FIR.

Proposal for link robustness improvement.

- (slide 11) Define of continuous adaptation of module's CTLE RX (#185,#191#192).
  - Proposed +/- 1dB variation with respect to the previous CTLE gain setting with a minimum 1Hz frequency rate (see mazzini_01_042414_caui and alessandro_01_07032014_caui).

Proposal for test improvement.

- (slide 12) Include TP1a VEC requirement.
  - Value to be agreed.

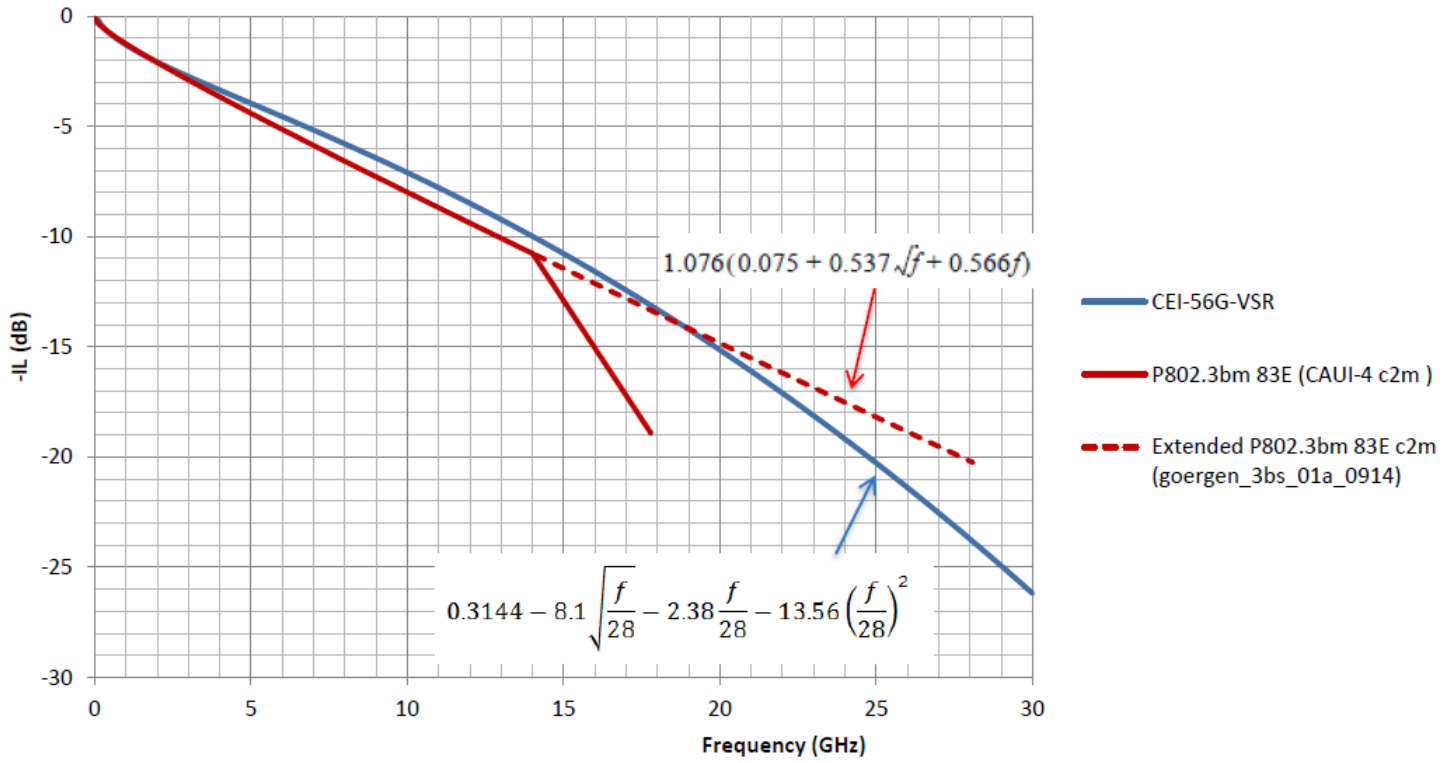
Back-up slides

Options to be considered for CDAUI-8 C2M (mazzini_01_082415_elect).

Six items (last one if strickly needed) that should be considered into C2M CDAUI-8, to allow TP1a EW/EH compliance and safe C2M link closure.

1. Define a more complex standard CTLE RX equalizer.
   - 3p2z or better.
   - Introduce a LF filter on the standard reference RX equalizer too?
2. Definition of continuos adaptation of module's CTLE RX.
   - Proposed +/- 1dB variation with respect to the previous CTLE gain setting with a minimum 1Hz frequency rate
3. Introduce a channel ILD deviation mask requirement.
   - Value to be agreed.
4. Improve TP0 TX SNDR requirement.
   - Value to be agreed (from 27 to 29dB ?).
5. Include TP1 VEC requirement.
   - Value to be agreed.
6. Allow "Coarse" host TX pre-cursor tap for long channels only (off for short ones) – this would be a fixed and "static" value.
   - Value and loss range to be eventually defined.
   - Symmetrically the same tap should be needed on module's transmitter too so should be writable by the host into the module.

CDAUI-8 c2m PAM4 Channel Insertion Loss



- Target IL curve is "extended CAUI-4 c2m" as adopted at Jan/2015 Interim
  - Loss at Nyquist (13.28GHz) = 10.27 dB
  - Working assumption: all IL curves shown are suitable for PAM4 signaling at 28GBd
  - Consider potential to operate over legacy CAUI-4 c2m channels

Host-to-Module Electrical Specifications at TP0a from OIF (not currently defined into Draft 1.0)

Note: A 2 tap FIR filter may be advantageous in meeting the TP1a requirements.

Table 16-9. Host-to-Module Electrical Specifications at TP0a

Parameter	Symbol	Min.	Max.	Units	Conditions
Baud Rate		19.5	28.0	GBd	
Differential Voltage, pk-pk	T_Vdiff	800	-	mV	Note 1
DC Common Mode Voltage	T_Vcm	0	1900	mV	Note 2
Differential resistance	T_Rd	80	120	ohms	
Differential Termination Resistance Mismatch	T_Rdm	-	10	%	at 1 MHz
Differential Return Loss	T_SDD22	-		dB	
Transition Time: 20 to 80%	T_tr, T_tf	10	-	ps	
Common-mode return loss				dB	
Common Mode Noise, RMS	T_Ncm	-	12	mV	See 12.3
Output waveform					
Level separation mismatch ratio, R_{LM}		0.92		-	
Steady-state voltage, v_f		0.4	0.6	V	
Linear fit pulse peak		$0.85*v_f$		V	
Normalized coefficient step size		0.0083	0.05	-	
Pre-cursor full-scale range		1.54		-	
Post-cursor full-scale range		4		-	
Output Jitter and Linearity					
Clock random jitter			0.005	UI_{RMS}	

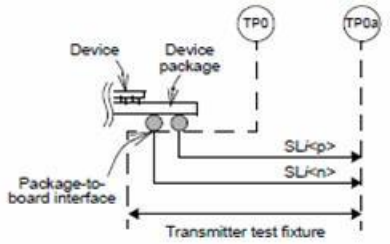


Figure 93–5—Transmitter test fixture and test points

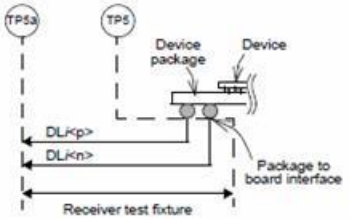


Figure 93–10—Receiver test fixture and test points

Table 16-9. Host-to-Module Electrical Specifications at TP0a

Parameter	Symbol	Min.	Max.	Units	Conditions
Clock deterministic jitter			0.05	UI	
Even-odd jitter			0.019	UI	
Signal-to-noise-and-distortion ratio			27	dB	

Notes 1:Max voltage is limited by specifications at TP1a. Minimum voltage can be lower for low loss channels.

Note 2: Load type 0 with min. T_Vdiff, AC-Coupling or floating load.