Module and PMA delay limits

Piers Dawe, Nvidia

Introduction

- While a high speed Ethernet spec should allow or enable very low delay (latency), the requirements should be set at a more mundane level
- maki_3df_01a_230523* considered the delay of an optical module. Changes were made: see Table 169-4
- Even after these changes, delay allocation is not enough to enable a broad market in 8:8 PMAs
- Often an 8:8 PMA is implemented as two back-to-back SerDes, while 32:8 or 8:32 is implemented as one SerDes
- This problem would become more obvious for "on-board repeaters"
- Relates to comments 13, 9, 22, 3 (withdrawn), 50 and 15

* https://ieee802.org/3/df/public/23_0523/maki_3df_01a_230523.pdf , and see https://ieee802.org/3/dj/public/adhoc/optics/0423_OPTX/maki_3dj_optx_01_230427.pdf

What is the division between PMD and PMA?

- 120.1.3 Summary of functions "the PMA ... Provide per input-lane clock and data recovery"
- P802.3cw 156.2.1.2.1 Semantics of the primitive:
  The PMD_UNITDATA.indication primitive conveys four analog signals, representing the in-phase (I) and quadrature (Q) components for each of the polarizations...
- 3cw is not binding here, but it is a reasonable way of describing the component parts, that works when more sophisticated signal processing techniques are used
- So, the PMD does optical to electrical conversion, and may provide some continuous-time equalization (which adds very little delay) and limiting (for PAM2 not PAM4). These take a handful of UI or O(1 ns). In addition, some implementations may include fiber pigtails behind the MDI (several ns) and possibly a small PCB (very few ns)
- The PMA does clock recovery, A to D and any DSP, and might include some PCB

More on PMD and PMA(s)

- There is one PMD (per end of a link) in the stack
  - For BASE-R optical: one PMD delay limit, typically the same for all PHY types at a rate, including both directions at one end and 2 m of fiber (10 ns)
  - For BASE-R copper: one higher limit, typically the same for all PHY types at a rate, including AN **and the medium** (20 ns at 50G/lane, 14 ns at 100G/lane)
- There can be several PMAs in the stack. Traditionally one aggregate limit; D2.1 introduced per-instance limits
- "A PMA" may actually be two PMAs in series
- For a typical retimed module and for copper PHYs, the PMA-PMD interface is internal so spare PMD delay allocation might be given unofficially to the PMA in the module
- There are PMA-PMA interfaces (AUIs) and as linear optics becomes more popular, the PMA-PMD interface is accessible
- We should divide the allocations realistically

Details and numbers

Item or group of items	ns	Reference	Comments	More comments
100G/lane BASE-R PMAs, nearly all PHY types	92.16		up to four PMA stages	Seems to be the sum of PMAs in a stack What if there are more than 4? (e.g. 5 in Figure 135A-8, Example CAUI-4 chip-to-chip and 100GAUI-n chip-to-module with Inverse RS-FEC)
100G/lane 800GBASE-R PMA sublayer	46.08		As in D2.1 (previously as other 100G/lane)	A PMA sublayer means one instance in a stack
8:8 PMA, proposed	81.92	Comments 9, 15	Increase to allow for back-to-back SerDes implementation	
8:8 PMA, proposed	92.16	Comment 3 (withdrawn)	Considering 8:8 PMA as 8:32 + 32:8	
Nearly all 100G/lane BASE-R optical PMD, proposed in comment 13	20.48	Table 116-7	PMD extends from MDI to PMA, and 2 m of fiber for measurement convenience	One end, both directions, including 2 m of fiber (~10 ns) Includes an allowance for a pigtailed PMD (fiber behind the MDI) Leaves a handful of ns for the PMD electronics (analog), which is enough
800GBASE- VR8, -SR8, -DR8 or -DR8-2 PMD	40.96		As in D2.1 (previously 20.48	as other 100G/lane optical)
100G/lane BASE-R electrical PMD, AN and medium	40.96	Table 162-4, Table 163-4	AN and PMD extend from MDI to PMA	Assumes that one-way delay through the medium is <= 14 ns (was 20 at 50G/lane)
Auto-Negotiation			Included within the delay of the PMD and medium	
Retimed module: optical PMD + one 8:8 PMA	66.56	Base standard, D2.0		If there is no XS. Too little
	200	Proposed in maki_3df_01a_230523		More than enough?
	87.04	P802.3df D2.1		Slightly too little
	102.4	Proposed in comments 9, 13 15		About right
	133.12	Proposed in comment 3		
	112.64	Combining comments 13 and 3		About right
Standalone 8:8 PMA	46.08	Base standard, D2.0		E.g. on-board repeater between C2M and C2C (if no XS). Too little
	46.08	D2.1		Too little
	81.92	Comments 9, 15		About right
	92.16	Comment 3 (withdrawn)		About right

Also, comments 22 and 50 propose changes to clarify the text in D2.1

Conclusion

- D2.1 has addressed an issue in D2.0, does not go quite far enough for 8:8 PMAs, made a change for optical PMDs that sort-of helps some but not all 8:8 PMDs
- Delay allocation for 8:8 PMA should be increased to 92.16 ns, double that for 32:8 or 8:32
- After allowing appropriately for 8:8 PMA, the delay allocation for optical PMD should be reverted