25 Gb/s Ethernet link configuration

Adam Healey
Avago Technologies

25 Gb/s Ethernet Study Group
8 October 2014

Introduction

- A number of interesting use cases for 25 Gb/s operation over a copper cable assembly have been discussed
  - 3 m cable assembly, PHY does not include Forward Error Correction (FEC)
  - 5 m cable assembly à la 100GBASE-CR4
  - 3 m cable assembly with higher [than 100GBASE-CR4] host insertion loss

- This presentation does not debate the merits of the various use cases
  - It only postulates that a different solution for each use case could exist

Implications of the multiple use cases

- Let us assume some future project defines a different solution for each use case

- The following building blocks would then be defined
  - 3 m cable assembly
  - 5 m cable assembly
  - PHY without FEC
  - PHY with FEC and host insertion loss per Clause 92
  - PHY with FEC and host insertion loss higher than Clause 92 allocation
  - There are a number of incompatible combinations

- Don’t forget about 25GBASE-SR!
  - Assuming that a future definition of 25GBASE-SR would require RS-FEC or its equivalent
  - Assuming some higher host insertion loss could be tolerable with a relaxed electrical BER target (assuming FEC)

Some [not quite] analogies

- BASE-R FEC is optional
  - The medium requirements are not dependent on the FEC mode
  - This option was included for additional margin and/or MTTFPA protection

- 100GBASE-KR4 receiver is allowed to bypass error correction for high performance channels
  - There is no change to the encoding
  - This feature is not allowed for 100GBASE-CR4

Sorting out the various combinations

- This presentation considers two approaches
  - Define multiple PHYs
  - Define a single PHY with multiple “modes” of operation

- This presentation assumes that a cable assembly could have the means to advertise specifications to which it is compliant
  - By the way, it is not just about insertion loss
  - Such capability is currently beyond the scope of IEEE 802.3

Observations on interoperability

- Let us begin with a definition of interoperability

  Transmitter A and receiver B, when connected through a medium M, can communicate with a defined maximum error ratio

- We already have a number of cases where interoperability is not guaranteed even though physical mating is possible
  - Ex. 1: 100GBASE-CR4 PHYs connected with a 40GBASE-CR4 cable assembly
  - Ex. 2: “25GBASE-CR” PHYs connected with an SFF-8431 direct attach copper cable assembly
  - Such connections will exhibit degraded performance or won’t link up at all

- Expectation is that the user knows the PHY type being connected and the appropriate (compliant) medium for said PHY
  - It can also be taken out of the user’s hands using means currently beyond the scope of IEEE 802.3

Define different PHYs

- Designate each variant as a different PHY
  - 25GBASE-CR-x x = <no FEC>
  - 25GBASE-CR-y y = <with FEC, but with higher host loss>
  - 25GBASE-CR-z z = <with FEC>
  - Obviously, this is not a nomenclature proposal

Configuration matrix

	-CR-x	-CR-y	-CR-z
-CR-x	3 m	[1]	[1]
-CR-y	[1]	3 m	3 m [2]
-CR-z	[1]	3 m [2]	5 m

[1] Incompatible encoding
[2] Longer (higher loss) cable assembly could be supported but let us keep this simple for now

Superset implementations

- In this framework, 25GBASE-CR-z would likely be a superset of 25GBASE-CR-y
  - During Auto-Negotiation, both capabilities could be advertised
- If the implementation supported multiple encoders, then it could advertise 25GBASE-CR-x, -CR-y, and -CR-z capabilities
- Despite the intrinsic capabilities of the implementation, one can advertise a subset based on preferred operating modes
  - This can be based on a priori knowledge of the channel
- The “down-side” to this approach is that FEC may be enabled over some links that did not require it
  - Would “latency sensitive” applications advertise a capability other than 25GBASE-CR-x?

Additional compatibility considerations

- 25GBASE-CR-x hosts are likely incompatible with 25GBASE-SR4 modules and links due to lack of FEC
  - Again, an implementation usually has the means to determine what has been plugged into a socket and take appropriate action

Define a single PHY, multiple operating “modes”

- Could use existing the FEC ability/enable bits in the Auto-Negotiation link codeword base page
  - If both ports advertise FEC ability and at least one port requests that it be enabled, then FEC is enabled
  - Otherwise FEC is not enabled

- Consider a “host ability” bit (for lack of a better name)
  - 0 = Host channel complies with Clause 92 requirements
  - 1 = Host channel does not comply with Clause 92 requirements but does comply to some relaxed set of requirements (TBD)

- Assumptions
  - A device with “host ability” = 1 must always advertise “FEC ability” = 1 and “FEC enable” = 1

Configuration matrix, 1 of 4

Port A			Cable	Port B			Result
FEC ability	FEC enable	Host ability		FEC ability	FEC enable	Host ability	
0	0	0	3 m	0	0	0	OK, no FEC
			3 m	1	0 [a]	0	OK, no FEC
			3 m	1	1 [a]	0	OK, no FEC
			3 m	1	1	1	Degraded, no FEC
			5 m	0	0	0	Degraded, no FEC
			5 m	1	0 [b]	0	Degraded, no FEC
			5 m	1	1 [b]	0	Degraded, no FEC
			5 m	1	1	1	Degraded, no FEC

[a] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 0
[b] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 1

Configuration matrix, 2 of 4

Port A			Cable	Port B			Result
FEC ability	FEC enable	Host ability		FEC ability	FEC enable	Host ability	
1	0	0	3 m	0	0	0	OK, no FEC
			3 m	1	0 [a]	0	OK, no FEC
			3 m	1	1 [a]	0	OK, FEC
			3 m	1	1	1	OK, FEC
			5 m	0	0	0	Degraded, no FEC
			5 m	1	0 [b]	0	Degraded, no FEC
			5 m	1	1 [b]	0	OK, FEC
			5 m	1	1	1	Degraded, FEC

[a] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 0
[b] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 1

Configuration matrix, 3 of 4

Port A			Cable	Port B			Result
FEC ability	FEC enable	Host ability		FEC ability	FEC enable	Host ability	
1	1	0	3 m	0	0	0	OK, no FEC
			3 m	1	0 [a]	0	OK, FEC
			3 m	1	1 [a]	0	OK, FEC
			3 m	1	1	1	OK, FEC
			5 m	0	0	0	Degraded, no FEC
			5 m	1	0 [b]	0	OK, FEC
			5 m	1	1 [b]	0	OK, FEC
			5 m	1	1	1	Degraded, FEC

[a] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 0
[b] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 1

Configuration matrix, 4 of 4

Port A			Cable	Port B			Result
FEC ability	FEC enable	Host ability		FEC ability	FEC enable	Host ability	
1	1	1	3 m	0	0	0	Degraded, no FEC
			3 m	1	0 [a]	0	OK, FEC
			3 m	1	1 [a]	0	OK, FEC
			3 m	1	1	1	OK, FEC
			5 m	0	0	0	Degraded, no FEC
			5 m	1	0 [b]	0	Degraded, FEC
			5 m	1	1 [b]	0	Degraded, FEC
			5 m	1	1	1	Degraded, FEC

[a] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 0
[b] Given a priori knowledge of the channel, it makes more sense to advertise “FEC enable” = 1

Summary

- Preference is to define the smallest number of variants that serves the application space
- If there are multiple variants, this presentation puts forth some ideas on how to reconcile inter-connections of the variants
- One could argue the variants considered are distinct PHYs
  - Different encoding (low-latency versus higher performance)
  - Different reach (intra-rack versus inter-rack)
  - Different electrical requirements (cost trade-offs)
  - Give them different names, user gets a clue about medium compatibility
  - Auto-Negotiation can reconcile inter-connection of variants

Summary, continued

- One could also consider a single PHY with multiple modes
  - They are all “25GBASE-CR” ports
  - Compatibility information is buried a bit deeper in PHY management
  - Auto-Negotiation can reconcile inter-connection of variants
  - Enables [slightly] finer control over the enabling of FEC
  - A “host ability” bit could be used to identify incompatible connections when combined with a priori knowledge of the cable assembly
- Regardless of the approach, physical connections that are not interoperable can still be made
- Fodder for discussion…