25GE host/cable options

Summary of thinking after talking to a number of people…

Mark Nowell & Gary Nicholl

From dudek_111914_25GE_adhoc:

Possible interesting configurations.

QLOGIC

Configuration	Switch loss	Cable (m)	Server /switch loss	FEC	Comments
1	mid	5	mid	Clause 91	Baseline 100GBASE-CR4
2	high	5	low	Clause 91	Higher loss switch still 5m cable to server
3	high	3	low	Clause 74	Lower latency still high loss switch to server
4	mid	3	low	None	Lowest latency switch to server
5	high	3	high	Clause 91	Higher loss switch to switch link
6	mid	3	mid	Clause 74	(added based on ad hoc discussion) lower latency

- Based on assumptions on next page all these are possible using 100GBASE-KR4 like ASICs

6 configs for AutoNeg to deal with.

Historical Note: Autoneg is used to advertise abilities only, not host information.

In this preso: mid=6.81dB (consistent with 802.3bj), low = mid - 3dB, high = mid + 3dB

But further discussion (smaller email thread) resulted in a larger list…

	Switch host loss	Cable length	Server host loss	FEC and channel loss budget	Notes
1	mid	5m	mid	RS FEC, 35 dB	same as 100G CR4.
2	mid	5m	low	RS FEC, 35 dB	must allow this based on above.
10	low	5m	low	RS FEC, 35 dB	must allow this based on above.
3	mid	3m	mid	KR FEC, 28 dB	lower latency for 3m cable.
4	mid	3m	low	No FEC, 25 dB	zero (lowest) latency for 3m cable.
5	low	3m	low	No FEC, 25 dB	must allow this based on above.
6	high	5m	low	RS FEC, 35 dB	higher loss for switch, with 5m cable.
7	high	3m	high	RS FEC, 35 dB	higher loss both ends, with 3m cable.
8	high	3m	mid	RS FEC, 35 dB	must allow this based on above.
9	high	3m	low	KR FEC, 28 dB	lower latency for 3m cable.

10 combinations is starting to get really ugly when considering mfg testing, compliance testing, usability issues, economies of scale… what is reasonable to manage?

Back to the original motivations and objectives

5m objective

Define a single-lane 25 Gb/s PHY for operation over links consistent with copper twin axial cables, with lengths up to at least 5m

Why: Industry interest in supporting ports/reaches that were backwards compatible to 100GBASE-CR4 and would enable breakout configurations of 4x25GE using same equipment

3m objective

Define a single-lane 25 Gb/s PHY for operation over links consistent with copper twin axial cables, with lengths up to at least 3m

Why: lower power and cost solution, optimized for intra-rack server to switch interconnect, lowest latency

More motivations

- Plug and play – a channel that is guaranteed by definition to come up working, between transmitter, receiver, and channel
- Avoid spec “sprawl” – identify a minimum-set, straightforward spec which limits implementation complexity and multiplicity of test modes (ideally one spec for cable vendors to comply to)

This could be achieved with two configurations

Config	Host#1 Loss	Cable (m)	Host#2 loss	FEC	Comments
1	mid	5m	mid	RS (CL91)	Meets 5m objective & industry goals
2	mid	3m	mid	No	Meets 3m objective & industry goals

Only one problem… configuration #2 doesn't initially appear to be technically feasible.

- 3m cable reduces channel loss but not enough to support no FEC operation
- 35dB starting channel budget (802.3bj w/ 5m) – needs to reduce to 25dB to support no FEC (is this verified?)

Questions:

- Are we able to meet MTTFPA without FEC?
- Are we absolutely sure we can't close the link with improvements in cable assemblies for 3m or COM model assumptions?

Options to consider

Config	Host#1 Loss	Cable (m)	Host#2 loss	FEC	Comments
1	mid	5m	mid	RS (CL91)	Meets 5m objective & industry goals
2	mid	3m	mid	No	Meets 3m objective & industry goals

We start exploring options:
Option #1: Adding a “light” FEC with low latency (but not zero) – KR (CL74)
Option #2: Lowering the channel loss spec to claw back dBs

High level option #1: Use KR (CL74) FEC

Config	Host#1 Loss	Cable (m)	Host#2 loss	FEC	Comments
2a	mid	3m	mid	KR (CL74)	Meets 3m objective & industry goals

- Interesting (but possibly not verified as feasible yet)
- Buys the necessary margin to allow 3m cable, with hosts consistent with 802.3bj (aka “mid”) loss
- KR (CL74) FEC implementation is easily available to anyone who has a 10G design.
- But it still isn’t a solution for the ultimate desire of a No FEC solution
  - Question: What is market impact of the extra 82ns latency?

High level option #2: Reducing overall channel budget further

Config	Host#1 Loss	Cable (m)	Host#2 loss	FEC	Comments
2b	mid	3m	low	No	Lower host loss @ one end
2c	mid	<3m	mid	No	Further reduce cable length until No-FEC achievable

2b) Lowering server side host loss budget by some number of dB. “Justified” since typically smaller boards (compared to switches)

- Do servers really have margin to spare? How much? Conflicts possibly with desire to move to lower cost board materials (with higher loss).
- If we reduce one end, then we “could” increase the other end host loss spec. Creates quickly broadening set of configurations to address…

2c) Lowering cable reach to < 3m could eradicate configuration explosion

- How much lower than 3m would we need to go? Is it a useable length?
- Would this limit broad market potential for server TOR applications?

Beyond options 2a, 2b, 2c

The first order options open up possibilities for other options – leading to the 10+ potential configurations discussed earlier. Some potentials and comments:

Options	Justification	Issues to to mitigated
Increased host loss	• Take advantage of low loss host loss implementations to fully use available budget. • Enable lower cost materials • Potential to reduce # retimers on switch designs	• Breaks compatibility with 100GBASE CR4 compliant products • Creates configurations that are not “supported” but easily implemented • How many high host loss variants above 6.81dB? • Is there significant impact to cost (savings)?

Increasing configurations options

Blue = difference from above
Red = unsupported option that is a possible user config

Number of Valid Configs	Host-cable-Host-FEC	Comments
2	Mid-5m-Mid-RS Mid-3m-Mid-(No/KR)	Meets all goals & objectives (assumes KR FEC acceptable if No FEC not feasible)
2	Mid-5m-Mid-RS Mid-<3m-Mid-No	Find the reach that supports goals of low-latency operation and addresses market need (might not be possible)
3	Mid-5m-Mid-RS Mid-3m-Mid-KR Mid-3m-Low-No	Add lower host loss variant • All subsets are valid and managed by AN
5	Mid-5m-Mid-RS Mid-3m-Mid-KR Mid-3m-Low-No High-5m-Low-RS High-3m-High-RS High-5m-High-RS High-5m-Mid-RS	Adding a higher host loss variant. • Valid configs starts to grow • Start to see invalid (but implementable) configs

Key Takeaways/Next Steps

- It should be possible to minimize configurations
- Higher host loss (than 802.3bj) brings a number of issues. Do we need it?
- Interest in exploring ways to achieve loss budget that might support no-FEC
  - Cable length, Cable improvements, lower host loss variant, COM model and/or parameters optimization
  - More 3m cable channel data contributions?
- “No FEC” MTTFPA analysis needed
- User feedback needed on whether there is an acceptable minimum reach <3m