2 Port types for 25GBASE-CR.

Mike Dudek QLogic

802.3by Berlin March 2015

Supporters of 2 Port Type Proposal.

Eric	Baden	Broadcom
Vittal	Balasubramanian	Dell
Tom	Brown	Vitesse
John	D'Ambrosia	Dell
Jeff	Maki	Juniper
Erdem	Matoglu	Amphenol
Richard	Mellitz	Intel
Gary	Nicholl	Cisco
Dave	Ofelt	Juniper
Tom	Palkert	Molex
Kapil	Shrikhande	Dell
Scott	Sommers	Molex
Rob	Stone	Broadcom
Nathan	Tracy	TE
Pavel	Zivny	Tektronix

Introduction.

- A version of this proposal was presented at the Architecture Ad-Hoc on 3/4/15. The major difference is re-naming CR-L from that proposal to CR. In addition there are consequential changes in the auto-negotiation, and a simplification to only show one option as an auto-negotiation proposal.

- **The 25G project has two objectives, 3m and 5m.**
  - Two objectives were established because there are two applications.
    - Within the rack. Only 3m required.
    - Adjacent rack. 5m needed.

- **Desirable characteristics that may be conflicting and require engineering trade-offs that may be different in different applications.**
  - Low Latency
  - Low power
  - Allow trade-off between power/latency and Frame Loss ratio (Error rate)
  - Small silicon area
  - Plug and Play (Simple for the customer to use).
  - Have minimal market confusion.

Solution.

- **Create two port types and label the port types differently.**
  - One optimized for the within rack application
  - One optimized for the Inter rack application

- **Ensure that both port types will inter-operate in a plug and play fashion in the “within rack” application. Use Auto-negotiation to achieve this.**

- **Enable Auto-negotiation to be set to prioritize latency/power or Frame Loss ratio.**

Two Port types proposed definition.

- **Port type 1 For within the rack. Label it 25GBASE-CR-S.**
  - Port is required to implement both Base-R FEC and “No FEC”. Auto-negotiation is required.
    - Base-R FEC is required to meet 3m application
    - Using Base-R enables plug and play for all CA-S and CA-N cables.
    - “No FEC” comes almost free. Will be used with No FEC (CA-N) cables by systems wanting minimum latency/power. (choose this preference in Auto-negotiation).
    - Port is not burdened with the extra power/silicon area of RS-FEC.

- **Port type 2 For Inter-rack. Label it 25GBASE-CR**
  - Port is required to implement RS FEC, Base-R FEC, and “No FEC”. Auto-negotiation is required. It is a superset of 25GBASE-CR-S plus RS FEC capability.
    - RS FEC is required to meet 5m application.
    - Using RS FEC with another 25GBASE-CR port enables plug and play for all 25G cables (CA-L, CA-S and CA-N)
    - Requiring Base-R FEC in addition to RS FEC is a small power/silicon area increase (much less than RS-FEC). Enables plug and play with 25GBASE-CR-S for CA-N and CA-S cables (ie within rack application). Provides option of lower power/latency for 3m cables with another 25GBASE-CR port
    - “No FEC” comes almost free. Will be used with No FEC (CA-N) cables by systems wanting minimum latency. (choose this preference in Auto-negotiation).

Advantages of Proposal.

- Provides optimum solution for the two different applications.
- Minimal market confusion
- Plug and Play if cable type is appropriate for the application.
- Depending on their applications customers can choose whether to deploy 25GBASE-CR or 25GBASE-CR-S (or mix, eg 25GBASE-CR in switches, 25GBASE-CR-S for in-rack servers).
- Enables application tuning for latency/power versus Frame Loss Ratio/cable selection.

Auto-negotiation Proposal

Supporters of Auto-negotiation Proposal.

Eric	Baden	Broadcom
Tom	Brown	Vitesse
Erdem	Matoglu	Amphenol
Richard	Mellitz	Intel
Gary	Nicholl	Cisco
Tom	Palkert	Molex
Scott	Sommers	Molex
Rob	Stone	Broadcom
Nathan	Tracy	TE
Pavel	Zivny	Tektronix

Auto-negotiation Proposal.

- **There are a number of different ways of handling Auto-negotiation with this proposal. The following is the recommended proposal however other options could be created, which achieve the same result.**
  - Use two entries for Technology Ability field. 25GBASE-CR and 25GBASE-CR-S (The same bits can also be used for 25GBASE-KR, but don't have to be).
  - Add an extra bit to be "25G Maximum FEC Requested". This would be default set. (Extra bit is added so that different requests can be made for 10G versus 25G).
  - Add one extra bit for "BASE-R FEC requested". This would only be used if an attempt is made to link up with no FEC but the frame loss ratio is too high.
- **Rules**
  - If "25G Maximum FEC requested" is asserted by either partner then either RS FEC or BASE-R FEC is enabled.
    - If both partners advertise "25GBASE-CR" then RS FEC Else BASE-R FEC.
  - If "25G Maximum FEC requested" is not asserted by either partner then RS FEC is not used.
    - If either partner asserts BASE-R FEC then BASE-R FEC is used. Else No FEC

Setting of Auto-Negotiation bits (option 1)

- **Based on Port type and preference for Low latency and power versus most plug and play and lowest Frame Loss Ratio the bits are set as below.**

Initial setting.					
Local information		Bits advertised			
Type of Port	Highest priority latency (set by management)	Technology Ability 25GBASE-CR-S	Technology Ability 25GBASE-CR	25G Maximum FEC requested	Base -R FEC requested
25GBASE-CR	N	Y	Y	Y	N
	Y	Y	Y	N	N
25GBASE-CR-S	N	Y	N	Y	N
	Y	Y	N	N	N

Priority Resolution

Priority Resolution.								
Near end advertisement				Far end information received				
Technology Ability 25GBASE-CR-S	Technology Ability 25GBASE-CR	25G Maximum FEC requested	Base -R FEC requested	Technology Ability 25GBASE-CR-S	Technology Ability 25GBASE-CR	25G Maximum FEC requested	Base -R FEC requested	FEC to use
Y	Y	X	X	Y	Y	Y	X	RS
Y	Y	Y	X	Y	Y	X	X	RS
Y	X	X	X	Y	N	Y	X	BASE-R
Y	X	Y	X	Y	N	X	X	BASE-R
Y	N	X	X	Y	X	Y	X	BASE-R
Y	N	Y	X	Y	X	X	X	BASE-R
Y	X	N	Y	Y	X	N	X	BASE-R
Y	N	X	X	Y	X	N	Y	BASE-R
Y	X	N	N	Y	X	N	N	No FEC

If Frame loss ratio is too high.

- **If the Frame loss ratio is too high when attempting low latency and power then the auto-negotiation can be retried with the following settings which will increase the level of FEC being used.**

Retry Setting to use if No FEC has failed for too high Frame Error ratio					
Type of Port	Highest priority latency	Technology Ability 25GBASE-CR-S	Technology Ability 25GBASE-CR	25G Maximum FEC requested	Base -R FEC requested
25GBASE-CR	Y	Y	Y	N	Y
25GBASE-CR-S	Y	Y	N	N	Y

Retry Setting to use if Base-R FEC has failed for too high Frame Error ratio					
Type of Port	Highest priority latency	Technology Ability 25GBASE-CR-S	Technology Ability 25GBASE-CR	25G Maximum FEC requested	Base -R FEC requested
25GBASE-CR	X	Y	Y	Y	N
25GBASE-CR-S	NA				

Backup

Gate estimates. (from slavick_022515_25GE).

PCS, FEC area cost and performance

	Gates	% of total	35 dB BP	3m Cable	5m Cable
Clause 49	45k	9%	No way	Possibly	No way
Clause 74	80k	15%	Doubtful	Likely	Doubtful
Clause 108	400k	76%	Likely	No problem	Likely
Total	525k				

	Area	PCS % of PHY
PMD/PMD	X	
Cl 49	X * 0.05	4.5%
Cl 49, 74	X * 0.13	11.7%
Cl 49, 74, 108	X * 0.65	34.8%

Other proposals from Nowell_022515_25GE.

802.3by PHY designation options

Single 25GBASE-CR PHY

- Optional RS-FEC implementation
- Mandatory KR-FEC implementation
- Optional no-FEC mode implementation
- Mandatory AN implementation

Two PHYs 25GBASE-CR-L & 25GBASE-CR-S

- 25GBASE-CR-L
  - Mandatory RS-FEC implementation
- 25GBASE-CR-S
  - Mandatory KR-FEC implementation
  - Optional no-FEC mode implementation
- Mandatory AN implementation

Pro/Con (not complete)

- Guaranteed LP interop (w/3m cable)
  - But not with 5m if one LP doesn't have RS-FEC
- Does it satisfy 5m objective?
- Optional to implement RS-FEC
  - Some prefer not to implement to save power/area
- Mandatory to implement KR-FEC
  - Some might prefer to not implement KR-FEC to save power/area

Pro/Con (not complete)

- Ability to implement either/both
- No interoperability between PHYs
  - With any cable reach
  - Risk to user experience?
- Clarity on cable reach that will work with each PHY
- Risk of lack of clarity by users on difference between PHYs

Auto-negotiation as presented at ad-hoc.
CR is labeled CR-L in the following

Auto-negotiation Proposal Option 1.

- **There are a number of different ways of handling Auto-negotiation with this proposal. Two ways are described. Other options could be created.**

- **Option 1.**
  - Use a single entry for Technology Ability field of 25GBASE-CR. (This could be the same as 25GBASE-KR, but doesn't have to be).
  - Add one bit for RS FEC ability. (All 25GBASE-CR Technology has both no FEC and Base-R FEC).
  - Change the definition of FEC requested bit to be "Maximum FEC requested". This would be default set.
  - Add one extra bit for "BASE-R FEC requested". This would only be used if an attempt is made to link up with no FEC but the frame loss ratio is too high.

Setting of Auto-Negotiation bits (option 1)

- **Based on Port type and preference for Low latency and power versus most plug and play and lowest Frame Loss Ratio the bits are set as below.**

Option 1. Initial setting.					
Type of Port	Highest priority latency (set by management)	Technology Ability	RS FEC Availability	Maximum FEC requested	Base -R FEC requested
25GBASE-CR-L	N	25GBASE-CR	Y	Y	Y
	Y	25GBASE-CR	Y	N	N
25GBASE-CR-S	N	25GBASE-CR	N	Y	Y
	Y	25GBASE-CR	N	N	N

Priority Resolution (option 1)

- **Based on Port type and preference for Low latency the FEC is set as below.**

Priority Resolution.									
Local information		Local advertisement			Far end information received				
Type of Port	Highest priority latency	RS FEC Availability	Maximum FEC requested	Base -R FEC requested	Technology Ability	RS FEC Availability	Maximum FEC requested	Base -R FEC requested	FEC to use
25GBASE-CR-L	X	Y	X	X	25GBASE-CR	Y	Y	X	RS
	N	Y	Y	Y	25GBASE-CR	Y	X	X	RS
	X	Y	X	X	25GBASE-CR	N	Y	X	BASE-R
	N	Y	Y	Y	25GBASE-CR	N	X	X	BASE-R
	X	Y	X	X	25GBASE-CR	X	N	Y	BASE-R
	Y	Y	N	N	25GBASE-CR	X	N	N	No FEC
25GBASE-CR-S	X	N	X	X	25GBASE-CR	X	Y	X	BASE-R
	X	N	X	X	25GBASE-CR	X	N	Y	BASE-R
	Y	N	N	N	25GBASE-CR	X	N	N	No FEC

Priority Resolution (option 1)

- **If the host has cable information available the FEC could be set as below which would reduce the number of iterations of auto-negotiation for longer cables when attempting to get low latency. However this would require us to document the cable information feature.**

Priority Resolution.							
Local information			Far end information received				
Type of Port	Highest priority latency	Cable type	Technology Ability	RS FEC Availability	Maximum FEC requested	Base -R FEC requested	FEC to use
25GBASE-CR-L	X	X	25GBASE-CR	Y	Y	X	RS
	N	X	25GBASE-CR	Y	X	X	RS
	Y	L	25GBASE-CR	Y	N	X	RS
	X	X	25GBASE-CR	N	Y	X	BASE-R
	N	X	25GBASE-CR	N	X	X	BASE-R
	X	X	25GBASE-CR	X	N	Y	BASE-R
	Y	S	25GBASE-CR	X	N	X	BASE-R
	Y	L	25GBASE-CR	N	N	X	BASE-R
	Y	N or No info	25GBASE-CR	X	N	N	No FEC
25GBASE-CR-S	X	X	25GBASE-CR	X	Y	X	BASE-R
	X	X	25GBASE-CR	X	N	Y	BASE-R
	Y	S or L	25GBASE-CR	X	N	X	BASE-R
	Y	N or No info	25GBASE-CR	X	N	N	No FEC

Auto-negotiation implications.

- **Option 2.**
  - Is the same as Option 1 except that instead of having technology ability as just 25GBASE-CR and a separate RS FEC ability, two different technology ability bits would be used. One for 25GBASE-CR-L and one for 25GBASE-CR-S. (A 25GBASE-CR-L would set both as it is a superset.). (The advertisement and priority resolution tables would look the same except the column labeled 25GBASE-CR would be re-labeled 25GBASE-CR-S and the column labeled RS FEC ability would be re-labeled 24GBASE-CR-L