1001101100101

010100

010010111

101001010010



2 Port types for 25GBASE-CR. Mike Dudek QLogic 2/26/15

1991/1100101/100101/10/00/10/10



Introduction.

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. ۲



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-L
 - Port is required to implement RS FEC, Base-R FEC, and "No FEC". Auto-negotiation is required.
 - RS FEC is required to meet 5m application.
 - Using RS FEC with another 25GBASE-CR-L 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-L 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 \bullet 25GBASE-CR-L or 25GBASE-CR-S (or mix, eg 25GBASE-CR-L in switches, **25GBASE-CR-S** for in rack servers).
- Enables application tuning for latency/power versus Frame Loss Ratio/cable selection.



Auto-negotiation Proposal

dudek_Port type draft 2-26-15



Auto-negotiation Proposal Option 1.

There are a number of different ways of handling Auto-negotiation with this \bullet 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.

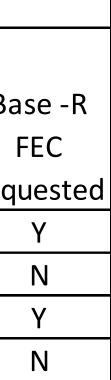


Setting of Auto-Negotiation bits (option 1)

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. | | | | | | |
|----------------------------|---------------------|------------|--------------|-----------|-----|--|
| | Highest priority | | | Maximum | Ba | |
| | latency (set by | Technology | RS FEC | FEC | | |
| Type of Port | management) | Ability | Availability | requested | rec | |
| 25GBASE-CR-L | Ν | 25GBASE-CR | Y | Y | | |
| ZJGDAJE-CN-L | Υ | 25GBASE-CR | Y | N | | |
| 25GBASE-CR-S | Ν | 25GBASE-CR | Ν | Y | | |
| | Y | 25GBASE-CR | Ν | N | | |





Priority Resolution (option 1)

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

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

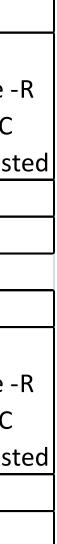


If Frame loss ratio is too high. (option 1)

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 | | | | | | |
|--|-------------------|------------------|----------------|-------------|--------|--|
| | | | | | | |
| | Highest | | | Maximum | Base - | |
| | priority | Technology | RS FEC | FEC | FEC | |
| Type of Port | latency | Ability | Availability | requested | reques | |
| 25GBASE-CR-L | Y | 25GBASE-CR | Y | N | Y | |
| 25GBASE-CR-S | Y | 25GBASE-CR | N | N | Y | |
| | | | | | | |
| Retry Setting to | o use if Base-R F | EC has failed fo | r too high Fra | me Error ra | itio | |
| | | | | | | |
| | Highest | | | Maximum | Base - | |
| | priority | Technology | RS FEC | FEC | FEC | |
| Type of Port | latency | Ability | Availability | requested | reques | |
| 25GBASE-CR-L | Х | 25GBASE-CR | Y | Y | Y | |
| 25GBASE-CR-S | NA | | | | | |



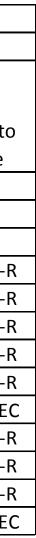


Priority Resolution (option 1)

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.

| | | Pr | iority Resolutio | on. | | | |
|-------------------|----------|------------------------------|------------------|--------------|-----------|-----------|--------|
| Local information | | Far end information received | | | | | |
| | Highest | | | | Maximum | Base -R | |
| | priority | | Technology | RS FEC | FEC | FEC | FEC to |
| Type of Port | latency | Cable type | Ability | Availability | requested | requested | use |
| | Х | Х | 25GBASE-CR | Y | Y | Х | RS |
| | N | Х | 25GBASE-CR | Y | Х | Х | RS |
| | Y | L | 25GBASE-CR | Y | N | Х | RS |
| | Х | Х | 25GBASE-CR | Ν | Y | Х | BASE-F |
| 25GBASE-CR-L | Ν | Х | 25GBASE-CR | Ν | Х | Х | BASE-F |
| | Х | Х | 25GBASE-CR | Х | N | Y | BASE-F |
| | Y | S | 25GBASE-CR | Х | N | Х | BASE-F |
| | Y | L | 25GBASE-CR | Ν | N | Х | BASE-F |
| | Y | N or No info | 25GBASE-CR | Х | N | Ν | No FE |
| 25GBASE-CR-S | Х | Х | 25GBASE-CR | Х | Y | Х | BASE-F |
| | Х | Х | 25GBASE-CR | Х | N | Y | BASE-F |
| | Y | S or L | 25GBASE-CR | Х | N | Х | BASE-F |
| | Y | N or No info | 25GBASE-CR | Х | N | N | No FE |





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



Backup

dudek_Port type draft 2-26-15



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 | Х | |
| CI 49 | X * 0.05 | 4.5% |
| CI 49, 74 | X * 0.13 | 11.7% |
| CI 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

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

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)

- 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



