SCHEDULE 3

25G & 50G Specification

Web: http://25gethernet.org/

Revision History

Revision	Who	Date	Change Description
1.0	HMF/BRCM	02/07/2014	Initial Release
1.1	HMF/BRCM	03/13/2014	Added information about CL91 FEC
1.2	HMF/BRCM	03/19/2014	Updated description of FEC control bits
1.3	HMF/BRCM	05/13/2014	Updated description of AM insertion in 3.2.1.3 p 8
1.4	OW/MLNX	08/21/2014	Updated the detailed implementation in 3.2, updated 3.2.5 with the 25G Ethernet Consortium CID
1.5	RJS/BRCM	11/03/2014	Updated with considerations for member comments, added asymmetric TP definitions and additional detail on AN for FEC
1.51	RJS / BRCM	7/20/2015	Corrected typo in Fig. 11 with the correct bit positions for the KR1 / CR1 abilities (correct in r1.4 and earlier)
1.6	RJS / BRCM	8/18/2015	Changed AM0 in RS FEC mode to 100G, modified AN priority table (IEEE > Consortium for 25G), corrected typo in next page Figure 11 (D23 → 0), rationalized twinax cable section, added informative section on 25/50G SFP / QSFP cable management. Added references to 802.3by project where applicable.
2.0	RJS / BRCM	6/14/16	Final Version, no changes

Definitions

25GBASE-R: An Ethernet Physical Coding Sublayer based on Clause 49 of IEEE Std 802.3, operating at a data rate of 25 Gb/s.

50GBASE-R: An Ethernet Physical Coding Sublayer based on Clause 82 of IEEE Std 802.3, operating at a data rate of 50 Gb/s.

25GBASE-CR1/KR1: An Ethernet Physical layer operating at 25 Gb/s on twin-axial copper cable/backplane traces.

50GBASE-CR2/KR2: An Ethernet Physical layer operating at 50 Gb/s on twin-axial copper cable/backplane traces.

Table of Contents

1 Overview 2
2 Standards Reference 3
3 25 Gb/s and 50 Gb/s Ethernet Specification 4
3.1 Architectural Overview 4
3.1.1 Blade server topologies 5
3.1.2 ToR switch topologies 6
3.1.3 Leveraging Existing Standards 7
3.2 Detailed Specification 9
3.2.1 Physical Coding Sublayer (PCS) 10
3.2.2 BASE-R FEC (Clause 74) Sublayer 12
3.2.3 RS-FEC (Clause 91) Sublayer 13
3.2.4 PMA Sublayer 19
3.2.5 Auto-negotiation 19
3.3 Electrical Specification 22
3.4 Channel Specifications 22
3.4.1 Copper cable channels 22
3.4.2 Backplane channels 22
3.4.3 Twin-axial cable channel loss allocation (informative) 22
3.5 Interconnect Annex A. 50G over QSFP28 MDI connector 22
3.5.1 50GBASE-CR2 / 50GBASE-SR2 over QSFP28 lane assignment 23
3.5.2 Use case example – 2 x 50GBASE-CR2 splitter cable 24
3.6 Interconnect Annex B. 25G, 50G QSFP28/SFP28 connectors management interface considerations (Informative) 25
3.6.1 Relevant memory map fields 25
3.6.2 25G/50G Cables Compliance Codes and Connectivity Examples 28

List of Tables

Table 1 – 25G 50G Solutions Summary 5
Table 2 – Link Priority List 20

List of Figures

Figure 1 – SerDes Bonding Schemes 4
Figure 2 – Mapping of SerDes Lanes with Mid-Plane to Permit 25 Gb/s Links 6
Figure 3 – Mapping of SerDes Lanes with Mid-Plane to Permit 50 Gb/s Links 6
Figure 4 – Open Compute server using N x 25G lanes to connect to TOR switch 7
Figure 5 - 25G and 50G modes of operation 9
Figure 6 - Alignment marker insertion period 11
Figure 7 - 25G RS-FEC Alignment Markers Mapping 14
Figure 8 - 50G RS-FEC Alignment Markers Mapping 16
Figure 9 - 50G RS-FEC Symbol Distribution 18
Figure 10 – Auto-negotiation Next Page – OUI extended 20
Figure 11 – Auto-negotiation OUI Extended Next Page Codes 20

1 Overview

The 25G & 50G Ethernet Consortium standard provides specifications for implementation based on single and dual lane 25 Gb/s technology, enabling adopters to deploy cost effective interoperable Ethernet technologies. The initial motivation for the 25 & 50G Consortium was to fill an important gap in the standards coverage afforded by 802.3 to enable rapid adoption of these important new speeds. Since the first draft of the Consortium specification was released in 2014, the IEEE also began work on standardization of 25 Gb/s over a single lane in the 802.3by project, with an anticipated release date in 2016. It should be noted that the Consortium specification is based on 802.3 clauses which pre-date the 802.3by project.

The capability to interconnect devices at 25 and 50 Gb/s Ethernet rates becomes especially relevant for next-generation data center networks where:

(i) In order to keep up with increasing CPU and storage bandwidth, rack or blade servers need to support aggregate throughputs faster than 10 Gb/s (single lane) or 20 Gb/s (dual lane) from their Network Interface Card (NIC) or LAN-on-Motherboard (LOM) networking ports;

(ii) Given the increased bandwidth to endpoints, uplinks from Top-of-Rack (TOR) or Blade switches need to transition from 40 Gb/s (four lanes) to 100 Gb/s (four lanes) while ideally maintaining the same per-lane breakout capability;

(iii) Due to the expected adoption of 100GBASE-CR4/KR4/SR4/LR4, SerDes and cabling technologies are already being developed and deployed to support 25 Gb/s per physical lane, twin-ax cable, or fiber.

The 25 Gb/s and 50 Gb/s Ethernet specifications described in this document allow existing rack and blade system architectures to support link speeds faster than 10 Gb/s and 20 Gb/s respectively, with no increase in cable/trace interconnect density, while keeping pace with the growth trajectory of networking bandwidth as faster and richer server CPUs are introduced. The 25 Gb/s and 50 Gb/s networking links defined herein act and behave like most other links already defined by IEEE 802.3.

IEEE defines 40 Gb/s as the next faster speed after 10 Gb/s. 40 Gb/s assumes four SerDes lanes available to end points to establish link between them. The proposed 25 Gb/s solution in this document requires only a single SerDes lane and potentially offers lower cost per unit bandwidth versus 40 Gb/s for rack server connectivity, while delivering a 2.5X speedup over current 10 Gb/s solutions. On the other hand, certain blade server chassis solutions today are limited to only two SerDes lanes for their LOM networking ports and therefore cannot implement a four lane 40 Gb/s interface. The 50 Gb/s solution described in this document overcomes this limitation and actually helps deliver 20% faster link speed.

The 25 Gb/s and 50 Gb/s PMDs defined in this document additionally support operation on low cost, twin-axial copper cables, requiring only two twin-axial cable pairs for 25 Gb/s operation,

and only four twin-axial cable pairs for 50 Gb/s operation. Links based on copper twin-axial cables can be used to connect servers to Top of Rack (ToR) switches, and as intra-rack connections between switches and/or routers.

2 Standards Reference

References are made throughout this document to IEEE 802.3-2012 Ethernet Access Method and Physical Layer [base standards]. It should be noted that at the time of publication of the Consortium Specification, the 802.3by specification was still in draft form (2.0), and so reference to this standard is mostly omitted from this document.

PCS

- Clause 49 PCS
- Clause 82 PCS and MLD

Auto-negotiation

- Clause 73, OUI Next Page (Using the 25G Ethernet Consortium CID). Base page FEC selection applies after PCS is selected.

FEC

- Clause 74 Fire code FEC
- Clause 91 RS FEC

Training

- IEEE 802.3bj-2014
- Clause 93 Link Training may be optionally applied lane by lane.
- If Clause 93 is not selected, Clause 72 Training may be optionally applied lane by lane.

Electrical

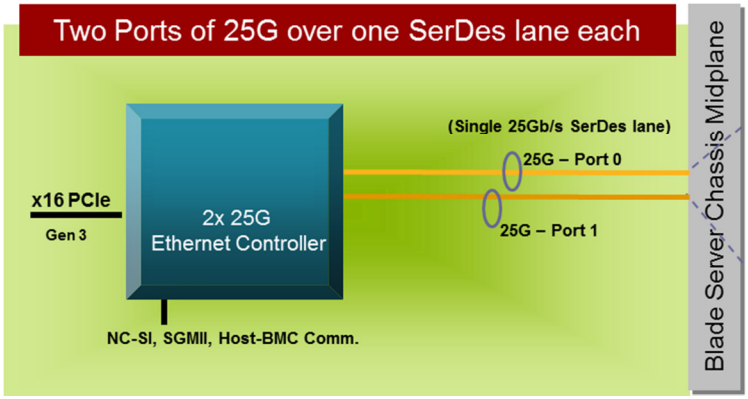
- IEEE 802.3bj-2014 Clauses 92, and 93.
- Clause 93 Link Training may be optionally applied lane by lane.

Note: IEEE 802 (e.g. IEEE 802.3-2012, etc) standards documents are available free through IEEE's Get program including IEEE 802 from http://standards.ieee.org/about/get/, six month after publication of each.

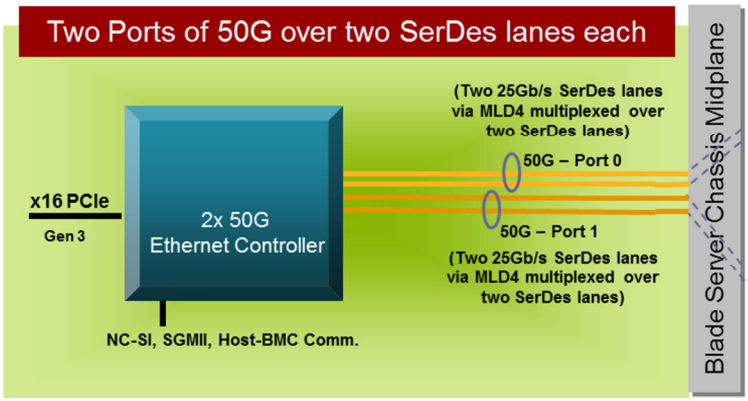
3 25 Gb/s and 50 Gb/s Ethernet Specification

3.1 Architectural Overview

25 & 50Gb/s Ethernet technology is designed for networking purposes that uses (a) one 25 Gb/s lane using a Clause 49 PCS to connect a single MAC operating at 25 Gb/s, (b) two 25 Gb/s lanes and combines them together using a Clause 82 PCS and MLD4 or RS-FEC over 2 FEC lanes to create a single logical MAC operating at 50 Gb/s. The following diagram depicts ways that 25 Gb/s and 50 Gb/s ports can be enabled:



A. 25 Gb/s SerDes Single Lane



B. 50 Gb/s SerDes Bonding Scheme

Figure 1 – SerDes Bonding Schemes

For the configurations shown in Figure 1, 25 Gb/s and 50 Gb/s interfaces represent themselves as networking ports. They can auto-negotiate link speed by proprietary bit locations in extended capabilities registers. The link can also be forced to run at 25 Gb/s or 50 Gb/s without advertising auto-negotiation capability.

Auto-negotiation advertising is conducted on lane 0 and is in accordance with IEEE 802.3 specifications. Distinction is made in the auto-negotiation bits that help the far end to identify whether 25 Gb/s or 50 Gb/s is being advertised and over what number of SerDes lanes.

25G and 50G auto-negotiation integrates with the Physical Coding Sublayer (PCS) defined by IEEE 802.3 for standardized speeds (10G, 40G). This allows the device to advertise 25G and 50G capability along with other supported speeds thereby permitting to link at any of the advertised speeds under IEEE 802.3 clause 73 auto-negotiation. For link partners that may not recognize 25G or 50G speed indication, they could choose to link at one of the other speeds that they are able to support.

Table 1 captures the port characteristics for two topologies.

Port Mode	SerDes (or PMD) Lane Count	Speed per Lane (GHz)	PCS Lane Bonding	Auto-negotiation
A. Single Lane 25G	1	25.78125	N/A	Yes
B. Two Lane 50G	2	25.78125	MLD4	Yes

Table 1 – 25G 50G Solutions Summary

A 25 Gb/s port can be bonded with 25 Gb/s, and a 50 Gb/s port can be bonded with 50 Gb/s, using any of the commonly used bonding protocols used to team Ethernet ports (e.g. IEEE 802.1ax Link Aggregation).

1. Benefits of Figure 1, Scheme A:

   25 Gb/s over one SerDes lanes running at 25 Gb/s allows a user to send data at a 250% faster rate than if the same lane was running at 10 Gb/s.

2. Benefits of Figure 1, Scheme B:

   50 Gb/s utilizing only two SerDes lanes provides a means to overcome the link speed limitation of lower speed Ethernet ports.

3.1.1 Blade server topologies

Figure 2 and Figure 3 provide examples of blade server topologies using 25 Gb/s and 50 Gb/s links.

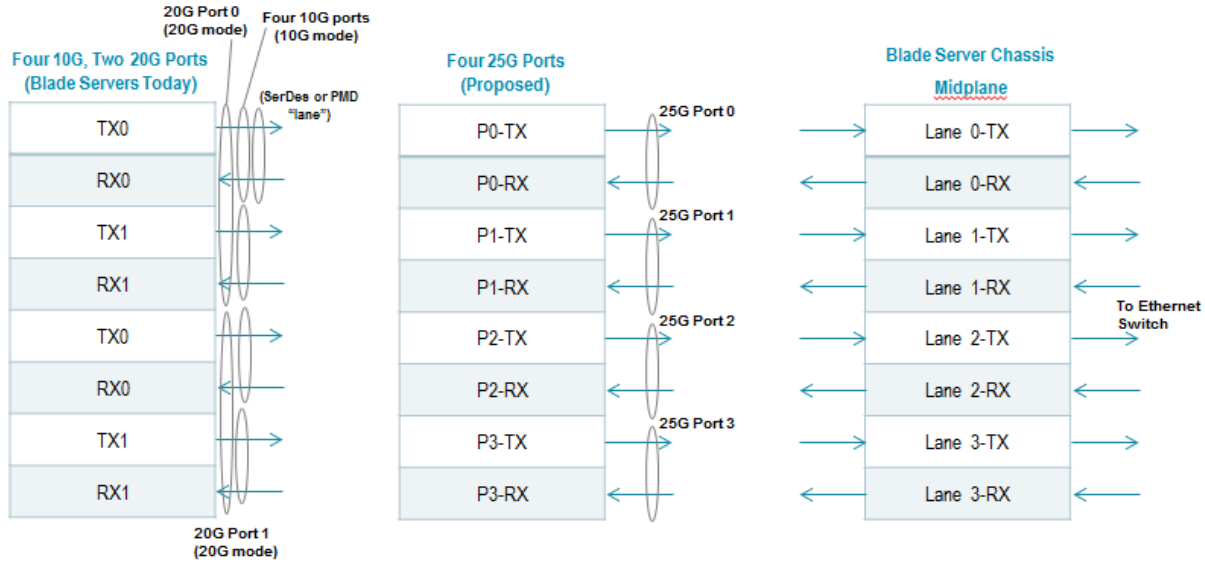


Figure 2 – Mapping of SerDes Lanes with Mid-Plane to Permit 25 Gb/s Links

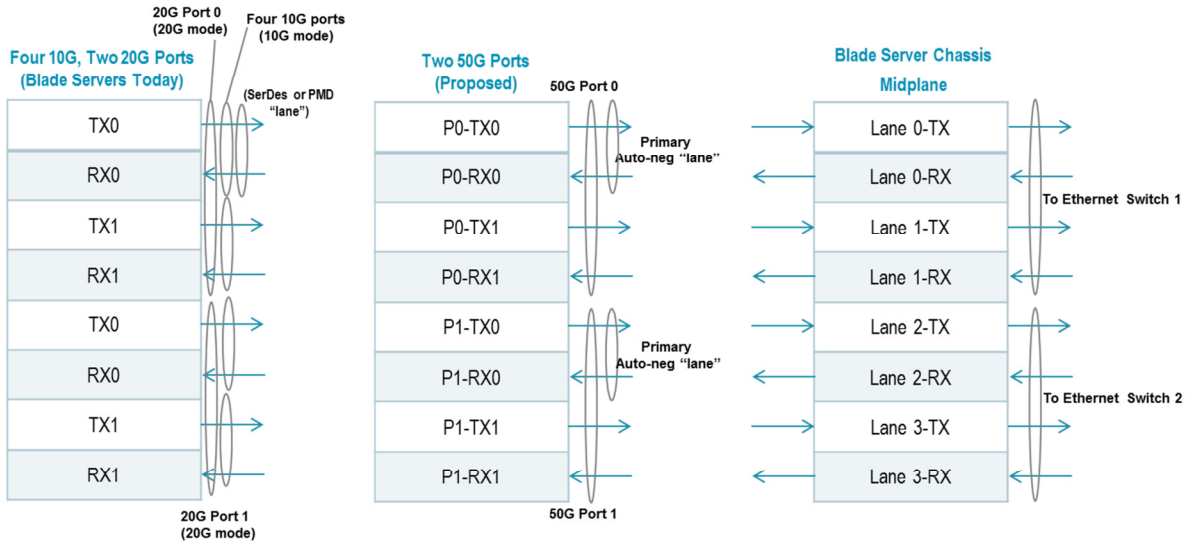


Figure 3 – Mapping of SerDes Lanes with Mid-Plane to Permit 50 Gb/s Links

3.1.2 ToR switch topologies

Figure 4 provides an example Open Compute Platform server rack using 25 Gb/s links.

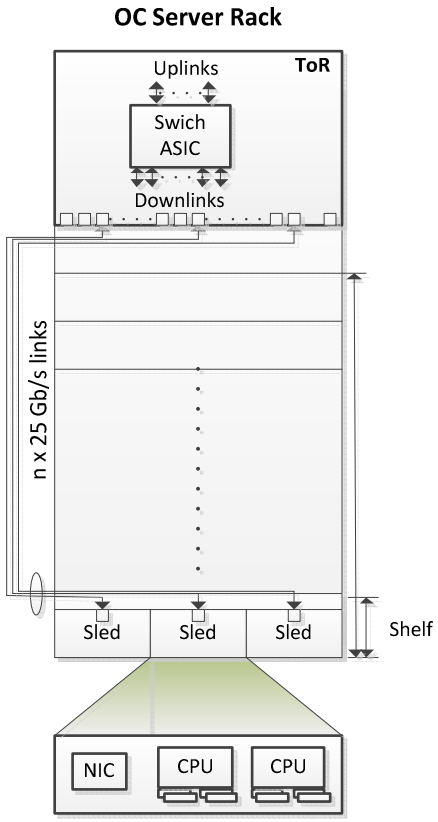


Figure 4 – Open Compute server using N x 25G lanes to connect to TOR switch

3.1.3 Leveraging Existing Standards

25 Gb/s capability can be supported by splitting four 25 Gb/s SerDes lanes of a 100G port into four 25 Gb/s ports with one SerDes lane each for an economical four-port 25G implementation that can serve networking port redundancy or multipathing needs.

50 Gb/s capability can be supported by splitting four 25 Gb/s SerDes lanes of a 100G port into two 50 Gb/s ports with two SerDes lanes each for an economical dual-port 50G implementation that can serve networking port redundancy or multipathing needs. A 100 Gb/s interface can be split into two physical and logical ports, each operating at up to 50 Gb/s.

The IEEE 802.3 standard for 40 Gb/s and 100 Gb/s Ethernet employs multi-lane distribution (MLD) to distribute data from a single Media Access Control (MAC) channel across a number of virtual lanes. For a given operating speed, the number of virtual lanes (also known as Physical Coding Sublayer (PCS) lanes) is determined by the least common multiple (LCM) of the desired range of Physical Medium Dependent (PMD) lanes. In the case of 100 Gb/s, the desired range of PMD lanes is 1, 2, 4 and 10, which yields an LCM of 20. Thus, the IEEE 802.3 standard for 100 Gb/s Ethernet uses MLD striped across 20 virtual lanes. This can be referred to as MLD20. In the case of 40 Gb/s, the desired range of PMD lanes is 1, 2 and 4, which yields an LCM of 4.

Thus, the IEEE 802.3 standard for 40 Gb/s Ethernet uses MLD striped across 4 virtual lanes, and this can be referred to as MLD4.

An important aspect of the MLD striping technique is the use of a unique alignment marker (AM) for each virtual lane. The AMs are inserted into the striped data stream every 16,383 codewords, where every codeword employs 64b/66b encoding. The use of a unique AM for each lane supports three important functions of the MLD striping technique, namely lane identification, lane alignment, and bit level multiplexing/demultiplexing. In the case of 100 Gb/s using MLD20, there are 20 unique AMs, one for each virtual lane. In the case of 40 Gb/s using MLD4, there are 4 unique AMs.

3.2 Detailed Specification

25G and 50G support three link modes of operation (See Figure 5 - 25G and 50G modes of operation):

1. Operation with no FEC
2. Operation with BASE-R (Clause 74 Firecode) FEC
3. Operation with Reed-Solomon (Clause 91) FEC

Auto-negotiation can be used to determine whether Firecode FEC, Reed-Solomon FEC, or no FEC is employed on the link. The PCS sublayer operation for the various modes at 25G and 50G are detailed in the following subsections.

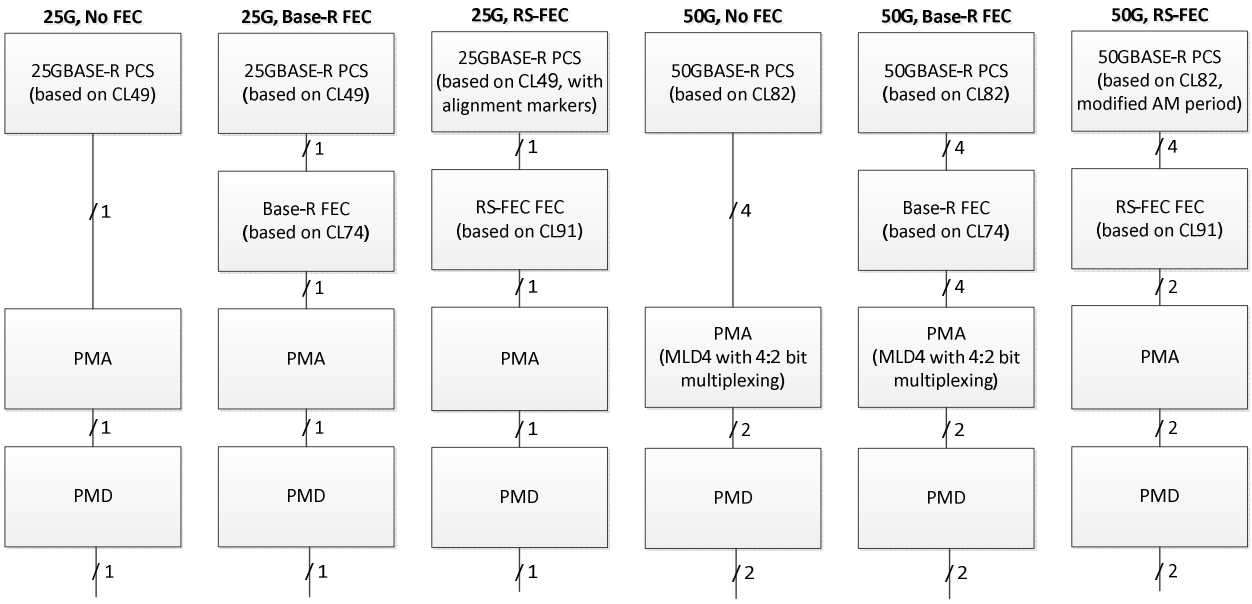


Figure 5 - 25G and 50G modes of operation

3.2.1 Physical Coding Sublayer (PCS)

3.2.1.1 *PCS sublayer for 25G*

For all 25G links, regardless of FEC mode, The PCS sublayer operates similarly to the IEEE 802.3 Clause 49 BASE-R PCS operating at x 2.5 rate, allowing data to be transmitted and received at 25.78125 Gb/s over a single lane.

3.2.1.1.1 25G PCS sublayer operation for links that use BASE-R FEC

For 25G links that use BASE-R (Clause 74 Firecode) FEC, the PCS lower interface connects to the BASE-R FEC sublayer over a single lane as defined in section 3.2.1.1 above. The PCS operation is similar to the operation without FEC where alignment markers are not inserted / removed.

3.2.1.1.2 25G PCS sublayer operation for links that use RS-FEC.

For 25G that uses RS-FEC, the PCS lower interface connects to the RS-FEC sublayer. When RS-FEC is used, the 25G PCS implements an alignment markers insertion and removal function. The alignment markers are used by the RS-FEC in order to achieve FEC block lock.

Alignment markers insertion and removal

Room for the alignment markers is created by periodically deleting IPG (inter-packet gap) from the XGMII data stream. Other special properties of the alignment markers are that they are not scrambled and do not conform to the encoding rules. The transmit function removes idle control characters, if necessary, to accommodate the insertion of the 4x66-bit alignment markers. The Idle deletion for the alignment markers insertion must match the requirements for Idle deletion for clock compensation (as defined in IEEE 802.3 49.2.4.7 Idle (/I/)) to guarantee that the data can be encoded in valid Clause 49 64B/66B control blocks.

The 4 alignment markers are inserted *after* every ([16,384 x 5 – 4] or 81,916) 66-bit blocks on the 25G single PCS lane. This enables the RS-FEC sublayer to detect and map the alignment markers to the *beginning* of every 1024 RS-FEC block.

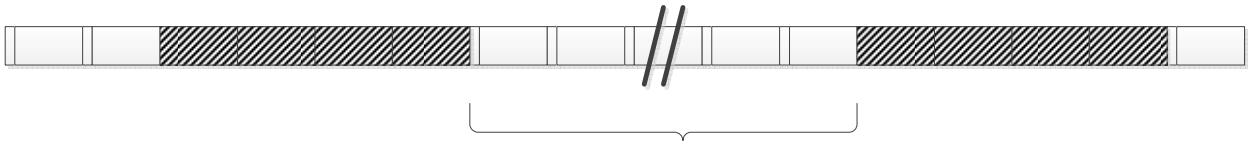


Figure 6 - Alignment marker insertion period

The alignment markers format resembles a combination of the clause 82 100G and 40G alignment markers format where the BIP fields are replaced with a reserved field.

On the receive side, the alignment markers are deleted from the data stream. The difference in rate from the deleted alignment markers is compensated for by inserting idle control character by a function in the Receive process.

Alignment markers format

The alignment markers insertion function inserts 4 consecutive alignment markers. The alignment markers are constructed by concatenating the clause 82 100G AM0 (PCS lane 0) followed by the clause 82 40G AM1, AM2, and AM3 (PCS lanes 1, 2 and 3 respectively). The

AM0 alignment markers format is as described in IEEE 802.3 Table 82-2 (0xC1, 0x68, 0x21, BIP_3, 0x3E, 0x97, 0xDE, BIP_7), and AM1, 2 and 3 are as described in IEEE 802.3 Table 82-3 (AM1 = 0xF0, 0xC4, 0xE6, BIP_3, 0x0F, 0x3B, 0x19, BIP_7), (AM2 = 0xC5, 0x65, 0x9B, BIP_3, 0x3A, 0x9A, 0x64, BIP_7) and (AM3 = 0xA2, 0x79, 0x3D, BIP_3, 0x5D, 0x86, 0xC2, BIP_7)

BIP3 and BIP7 are treated as reserved fields RSVD3 and RSVD7, where RSVD7 is the bit-wise inversion of RSVD3. RSVD3 value on transmission is implementation specific, however it is recommended that BIP3 is set to a constant value of 0x33 and BIP7 is set to 0xCC following 802.3by clause 108 in order to preserve DC balance. RSVD3 and RSVD7 are ignored by the receiver

3.2.1.2 *PCS sublayer for 50G*

For all 50G links, regardless of FEC mode, the 50G PCS sublayer operates similarly to the IEEE 802.3 Clause 82 40GBASE-R PCS operating at x 1.25 rate allowing data to be transmitted and received to/from the PMA/FEC sublayer at 51.5625 Gb/s over 4 lanes at 12.89063 Gb/s each. For operation with no FEC, or with base-R FEC, alignment markers insertion and removal period is as defined in IEEE 802.3 clause 82.2.7.

3.2.1.2.1 50G PCS sublayer operation for links that use BASE-R FEC

For 50G links that use BASE-R (CL74 Firecode) FEC, the PCS lower interface connects BASE-R FEC sublayer over 4 parallel lanes as defined in section 3.2.1.2 above.

3.2.1.2.2 50G PCS sublayer operation for links that use RS-FEC

For 50G that uses RS-FEC, the PCS lower interface connects to the RS-FEC sublayer. When RS-FEC is used, the 50G PCS inserts and removes alignment markers every 20479 66-bit blocks on each PCS lane. On the PCS receive, alignment markers appear 20480 66-bit blocks apart on every PCS lane. On the 4 lanes aligned data stream, alignment markers appear 16384x5 blocks apart which implies that an alignment markers transcoding block appears at the beginning of the RS-FEC codeword every 1024 RS-FEC codewords.

3.2.2 BASE-R FEC (Clause 74) Sublayer

Both the Clause 74 Fire code FEC and the Clause 91 Reed-Solomon FEC are supported by this specification. Auto-negotiation can be used to determine whether Clause 74 FEC, Clause 91FEC, or no FEC is employed on the link.

3.2.2.1 *BASE-R FEC Sublayer for 25G*

25G links can optionally instantiate a FEC sublayer. When the link is configured to use Firecode FEC , the FEC sublayer operates similarly to Clause 74 FEC at x2.5 rate, transmitting and receiving data at 25.78125 Gb/s over a single lane.

3.2.2.2 *BASE-R FEC Sublayer for 50G*

50G links can optionally instantiate a FEC sublayer. When the link is configured to use Firecode FEC, the FEC sublayer operates similarly to Clause 74 FEC at x1.25 rate, transmitting and receiving data at 51.5625 Gb/s over 4 lanes at 12.89063 Gb/s each.

3.2.3 RS-FEC (Clause 91) Sublayer

3.2.3.1 *RS-FEC Sublayer for 25G*

25G links can optionally instantiate a RS-FEC sublayer. When the link is configured to use RS-FEC, the RS-FEC sublayer operates similarly to the Clause 91 RS-FEC sublayer where the alignment markers mapping and the transmit side symbol distribution are modified for 25G operation over a single FEC lane and the transcoding function is extended to map the Clause 49 PCS Control codes.

The 25G RS-FEC sublayer sends to and receives from the PCS sublayer a single 25.78125 Gb/s bit stream. The RS-FEC sublayer sends to and receives from the PMA sublayer a single 25.78125 Gb/s bit stream.

3.2.3.1.1 Transmit side alignment markers mapping

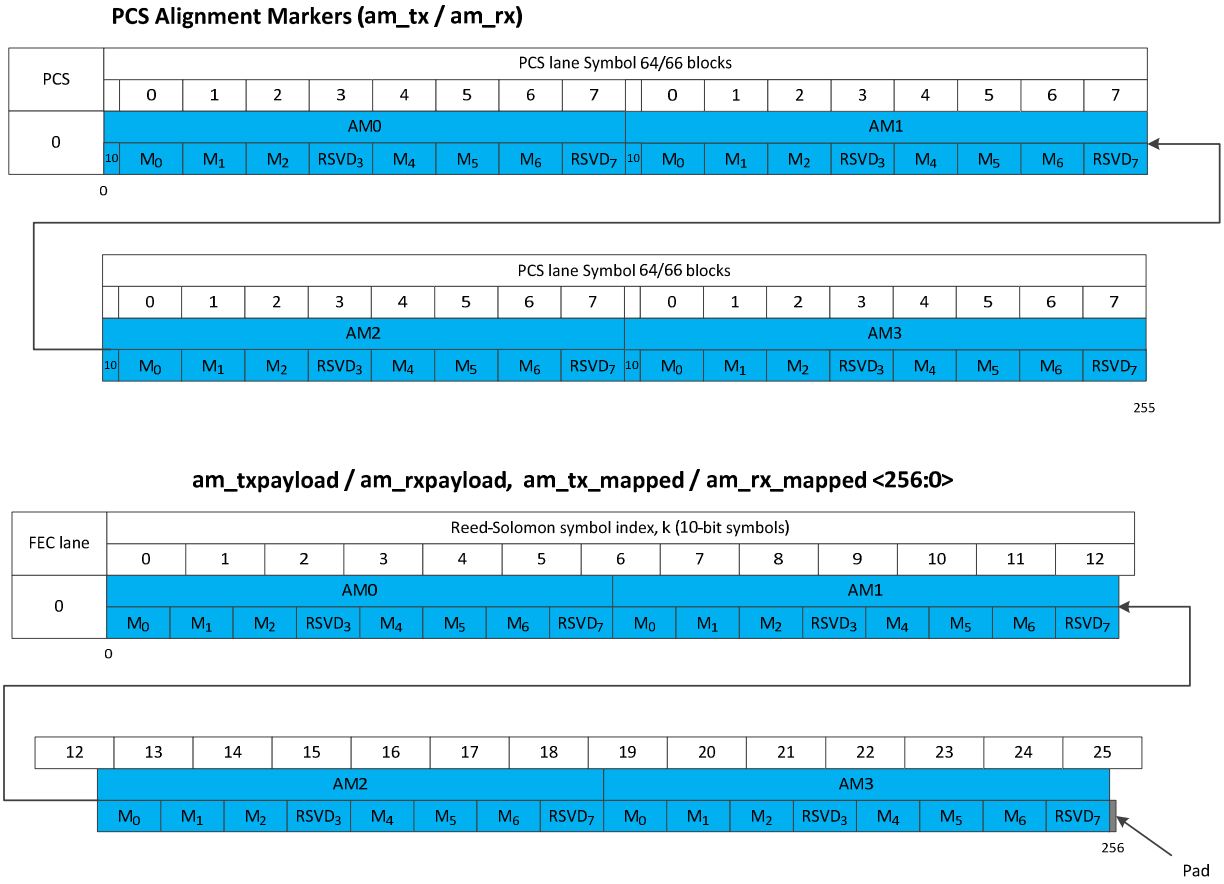


Figure 7 - 25G RS-FEC Alignment Markers Mapping

On the transmit side alignment markers are inserted every 81916 - 66 bit blocks, such that they appear every 81920 – 66 bit blocks on each PCS lane (which corresponds to 1024 Reed-Solomon codewords).

Let am_tx_x<65:0> be the alignment marker for PCS lane x, x=0 to 3, where bit 0 is the first bit transmitted. The alignment markers shall be mapped to an alignment markers transcoding codeword am_txmapped<256:0>, where bit 0 is the first bit transmitted, in a manner that yields the same result as the process defined below. (See Figure 7 - 25G RS-FEC Alignment Markers Mapping)

am_txmapped <256:0> is constructed as follows:

For x=0 to 3

a) am_txmapped<(64x+63):64x> = am_tx_x <65:2>

am_txmapped<256> = 1b’1 (pad bit).

3.2.3.1.2 Receive side alignment markers mapping

On the receive side the first 257 message bits in every 1024 codeword is the vector am_rxmapped<256:0>.

The alignment marker mapping function operates on the RS-FEC alignment markers transcoding codeword: am_rxmapped<256:0>. The alignment markers shall be mapped to am_rx_x<65:0> for PCS lane x, x=0 to 3, in a manner that yields the same result as the process defined below.

For x=0 to 3, am_rx_x <63:0> is constructed as follows:

a) am_rx_x<65:2> = am_rxmapped<(64x+63):64x>

b) am_rx_x<0>=1 and am_rx_x<1>=0.

Consistent with IEEE 802.3bj, the value of am_rxmapped<256> is ignored by the receiver.

3.2.3.1.3 256B/257B to 64B/66B transcoding

The 25G RS-FEC transcoding is done as described in Clause 91 RS-FEC while extending the control blocks support to the Clause 49 control blocks.

The transcoding function identifies the block type and completes the most significant nibble (h<3:0>) of the first control block based on its least significant nibble (g<3:0>), see IEEE 802.3 91.5.3.5 256B/257B to 64B/66B transcoder. The transcoding function for 25G is modified to support Clause 49 control blocks. The transcoding function shall identify the Clause 49 64B/66B control blocks described in IEEE 802.3 Figure 49-7.

3.2.3.1.4 Transmit side symbol distribution

Once the data has been Reed-Solomon encoded, it is transmitted over a single FEC lane, one 10-bit symbol at a time.

3.2.3.2 *RS-FEC Sublayer for 50G*

50G links can optionally instantiate a FEC sublayer. When the link is configured to use RS-FEC, the RS-FEC sublayer operates similarly to the Clause 91 RS-FEC sublayer where the alignment markers mapping and the transmit side symbol distribution are modified for 50G operation over 2 FEC lanes.
The 50G RS-FEC transcoding is done as described in Clause 91 RS-FEC transcoding function.

The RS-FEC sublayer operates over four parallel bit streams. The RS-FEC sublayer sends to and receives from the PCS sublayer 4x12.89063 Gb/s bit streams. The RS-FEC sublayer sends to and receives from the PMA sublayer 2x25.78125 Gb/s bit streams.

3.2.3.2.1 Transmit side alignment markers mapping

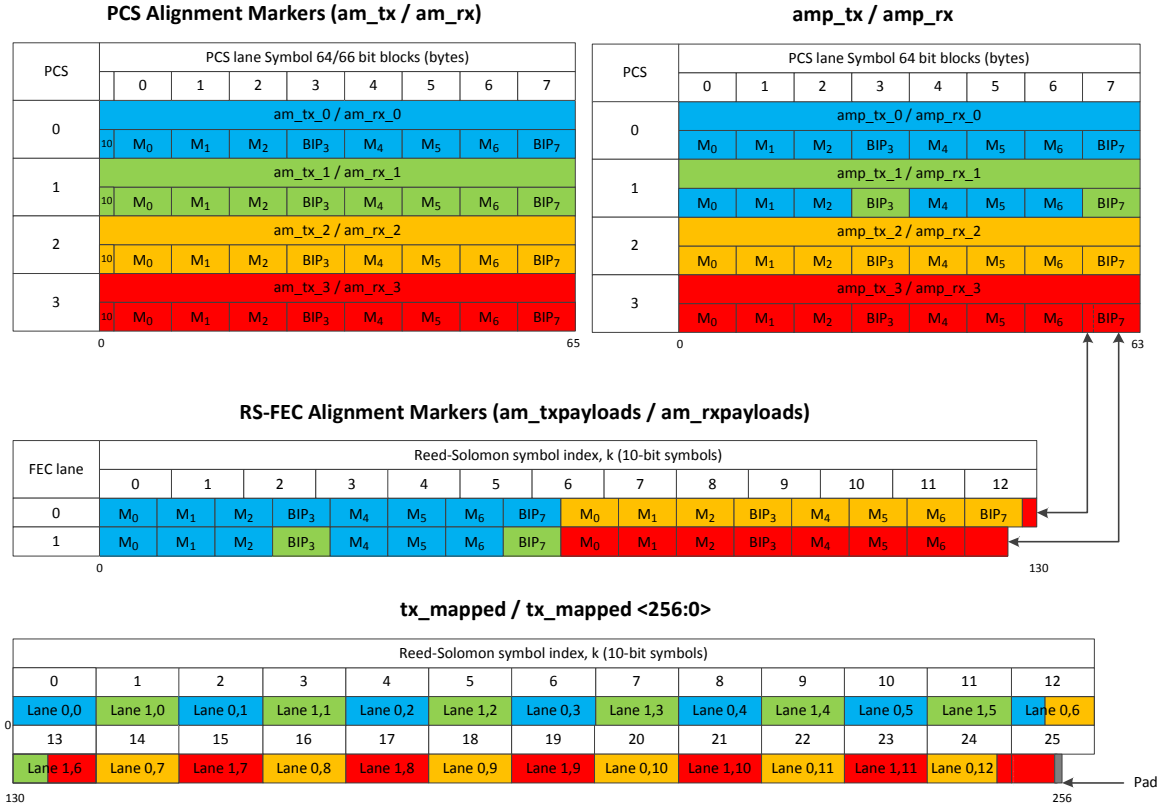


Figure 8 - 50G RS-FEC Alignment Markers Mapping

On the transmit side alignment markers appear every 20480 – 66 bit blocks on each PCS lane (which correspond to 1024 Reed-Solomon codewords).

Let am_tx_x<65:0> be the alignment marker for PCS lane x, x=0 to 3, where bit 0 is the first bit transmitted.

The alignment markers shall be mapped to an alignment markers transcoding codeword am_txmapped<256:0> in a manner that yields the same result as the process defined below. (See Figure 8 - 50G RS-FEC Alignment Markers Mapping)

For x=0 to 3, amp_tx_x<63:0> is constructed as follows:

a) set y = 0 when x ≤ 1, otherwise set y = x.

b) amp_tx_x<23:0> is set to M0, M1, and M2 using the values in Table 82–3 for PCS lane number y.

c) amp_tx_x<31:24> = am_tx_x<33:26>

d) amp_tx_x<55:32> is set to M4, M5, and M6 using the values in Table 82–3 for PCS lane number y.

e) amp_tx_x<63:56> = am_tx_x<65:58>

am_txpayloads<0, 63:0> = amp_tx_0<63:0>

am_txpayloads<0, 127:64> = amp_tx_2<63:0>

am_txpayloads<0, 129:128> = amp_tx_3<57:56>

am_txpayloads<1, 63:0> = amp_tx_1<63:0>

am_txpayloads<1, 119:64> = amp_tx_3<55:0>

am_txpayloads<1, 125:120> = amp_tx_3<63:58>

Given i=0 to 1, k=0 to 12, and y=i+2k, am_txmapped may then be derived from am_txpayloads per the following expression.

If (y < 25) then am_txmapped<(10y+9):10y> = am_txpayloads<i, (10k+9):10k>

am_txmapped<255:250> = am_txpayloads<1, 125:120>

am_txmapped<256> = 1b'1 (pad bit).

3.2.3.2.2 Receive side alignment markers mapping

On the receive side the first 257 message bits in every 1024 codeword is the vector am_rxmapped<256:0>.

The alignment marker mapping function operates on the RS-FEC alignment markers transcoding codeword: am_rxmapped<256:0>. The alignment markers shall be mapped to am_rx_x<65:0> for PCS lane x, x=0 to 3, in a manner that yields the same result as the process defined below.

Given i=0 to 1, k=0 to 12, and y=i+2k, am_rxpayloads may then be derived from am_rxmapped per the following expression.
If (y < 25) then am_rxpayloads<i, (10k+9):10k> = am_rxmapped<(10y+9):10y>

am_rxpayloads<1, 125:120> = am_rxmapped<(255:250>

Consistent with IEEE 802.3bj, the value of am_rxmapped<256> is ignored by the receiver.

a) amp_rx_0<63:0> = am_rxpayloads<0, 63:0>
b) amp_rx_1<63:0> = am_rxpayloads<1, 63:0>
c) amp_rx_2<63:0> = am_rxpayloads<0, 127:64>
d) amp_rx_3<55:0> = am_rxpayloads<1, 119:64>
e) amp_rx_3<57:56> = am_rxpayloads<0, 129:128>
f) amp_rx_3<63:58> = am_rxpayloads<1, 125:120>

For x=0 to 3, am_rx_x<63:0> is constructed as follows:

a) am_rx_x<0>=1 and am_rx_x<1>=0.

b) am_rx_x<25:2> is set to M0, M1, and M2 using the values in Table 82–3 for PCS lane number x.

c) am_rx_x<33:26> = amp_rx_x<31:24>

d) am_rx_x<57:34> is set to M4, M5, and M6 using the values in Table 82–3 for PCS lane number x.

e) am_rx_x<65:58> = amp_rx_x<63:56>

3.2.3.2.3 Transmit side symbol distribution

Once the data has been Reed-Solomon encoded, it shall be distributed to 2 FEC lanes, one 10-bit symbol at a time in a round robin distribution from the lowest to the highest numbered FEC lane.

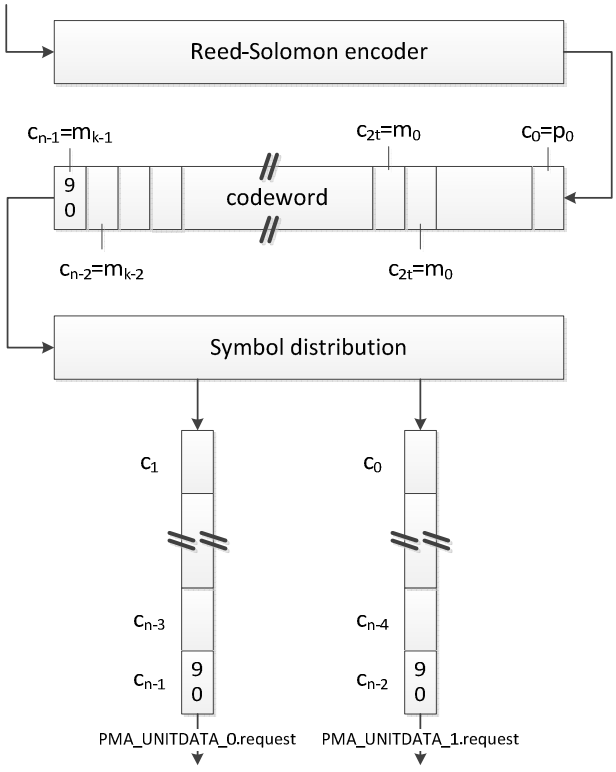


Figure 9 - 50G RS-FEC Symbol Distribution

3.2.4 PMA Sublayer

3.2.4.1 *PMA Sublayer for 25G*

The PMA sublayer for 25G operates similarly to 100G Clause 83 PMA with a single lane of 25.78125 Gb/s on the PMA client (PCS, BASE-R FEC or RS-FEC) and a single lane of 25.78125 Gb/s towards the PMD.

3.2.4.1 *PMA sublayer for 50G*

The PMA sublayer for 50G operates similarly to Clause 83 PMA.

If the 50G PMA client is the PCS or a BASE-R FEC sublayer (Clause 74), the PMA (or PMA client) continuously sends 4 parallel bit streams to the PMA client (or PMA), each at the nominal signaling rate of 12.89063 Gb/s.

When the 50G PMA client comprises of 4 lanes, it implements an MLD4 which performs bit-level multiplexing function that sends and receives the 4 virtual lanes over two 25.78 Gb/s backplane or copper cable SerDes communicating across two physical backplane or copper cable channels as shown in Figure 1B.

The PCS receiver shall support all combinations of virtual lane to physical lane distribution.

If the 50G PMA client is the RS-FEC (Clause 91) sublayer, the PMA (or PMA client) continuously sends 2 parallel bit streams to the PMA client (or PMA), each at the nominal signaling rate of 25.78125 Gb/s.

3.2.5 Auto-negotiation

3.2.5.1 *Speed selection*

For 25 Gb/s and 50 Gb/s operation, the Auto-negotiation base pages are exchanged between the two ends of the backplane or copper cable channel, with the exchange taking place on physical lane 0. After the exchange of the base page, the link partners exchange an OUI tagged formatted Next Page (using message code #5) and then the link partners exchange an OUI tagged unformatted Next Page with an extended technology abilities field, as detailed below. The link operating speed is determined by the highest common denominator advertised by the link partners, and resolved according to the priority table shown in Table 2.

Priority	Technology	Capability	Note
1	100GBASE-R4	100 Gb/s, 4 lane	IEEE mode
2	50GBASE-R2	50 Gb/s, 2 lanes	Consortium mode
3	40GBASE-R4	40 Gb/s, 4 lanes	IEEE mode
4	25GBASE-R1	25 Gb/s, 1 lane	IEEE mode
5	25GBASE-R1	25 Gb/s, 1 lane	Consortium mode
6	10GBASE-R	10 Gb/s, 1 lane	IEEE mode

Table 2 – Link Priority List

The Organization Unique Identifier (OUI) tagged formatted Next Page, with the OUI set to the 25G Ethernet Consortium CID of 0x6A737D(h) and a message code of 5 is shown in the figure below:

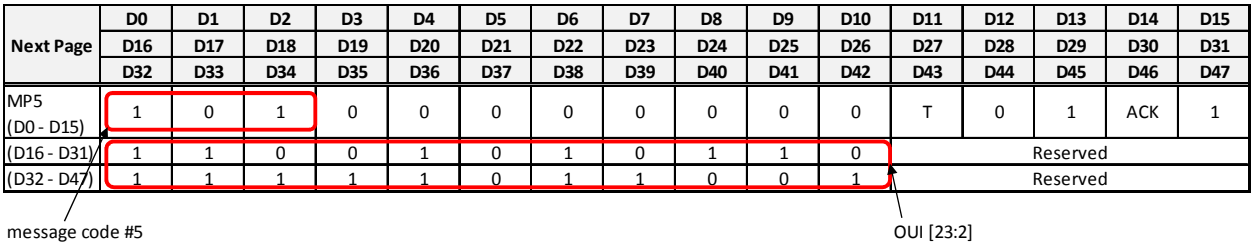


Figure 10 – Auto-negotiation Next Page – OUI extended

The OUI tagged unformatted Next Page, with the OUI set to the 25G Ethernet Consortium CID of 0x6A737D(h), a code for extended technology abilities, and the appropriate code bits for 25 GE CR1/KR1 and 50 GE CR2/KR2 operation, as well as four FEC control bits, is shown in the figure below:

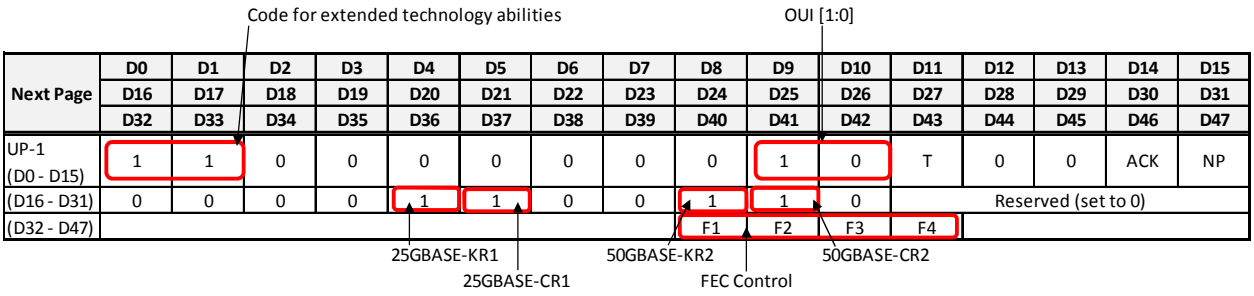


Figure 11 – Auto-negotiation OUI Extended Next Page Codes

Note: Implementers are encouraged to make the OUI value programmable.

3.2.5.2 *FEC control*

In addition to existing Clause 73 Base Page FEC control bits, four bits are added to the UP-1 (Unformatted Next Page 1) to determine FEC usage. These are F1, F2, F3, and F4. These bits only apply if the port speed selected by the AN process is either 25G or 50G.

Note the following references:
BP.F0 Refers to the Clause 73 Base Page (BP) F0 (10 Gb/s per lane FEC ability) bit
BP.F1 Refers to the Clause 73 Base Page (BP) F1 (10 Gb/s per lane FEC requested) bit
F1 advertises Clause 91 FEC ability
F2 advertises Clause 74 FEC ability
F3 requests Clause 91 FEC
F4 requests Clause 74 FEC
LD = local device
RD = remote device

Algorithmically, the meaning of these FEC control bits can be expressed as follows:

```
If the speed Highest Common Denominator is either 25 Gb/s or 50 Gb/s {
        If ((LD.F3 | RD.F3 ) & LD.F1 & RD.F1) {          //  If either link partner requests the
Clause 91 FEC, and BOTH support it.
                Use Clause 91 FEC ;
        }
        Else if ((LD.F4 | RD.F4) & LD.F2 & RD.F2) {      // If either link partner requests the
Clause 74 FEC, and BOTH support it.
                Use Clause 74 FEC;                      //  Regardless of the Clause 73 Base
        Page FEC bits
        }
        Else if (( LD.BP.F1 | RD.BP.F1 ) & LD.BP.F0 & RD.BP.F0) {    //  Use Clause 73 10G Base
Page FEC control bits to determine whether CL74 FEC is used
               Use Clause 74 FEC;
        }
        Else     NO FEC;
}
```

NOTE: To guarantee proper FEC operation, it is recommended to program the UP-1 FEC bits and the Clause 73 Base Page FEC bits in a consistent manner.

3.2.5.3 *Training*

Once Auto-negotiation completes, the hardware will automatically and seamlessly switch to link training.

1. Clause 93 link training may be optionally applied on a lane by lane basis.

2. If Clause 93 training is not selected, the Clause 72 link training may be optionally applied on a lane by lane basis.

3.3 Electrical Specification

The 25G and 50G PMDs shall conform to all of the electrical specifications defined in IEEE Std 802.3bj, Clause 92, and 802.3by Clause 110 “Physical Medium Dependent (PMD) sublayer and baseband medium”, specifically those defined in 92.8, and 110.8 where applicable.

3.4 Channel Specifications

3.4.1 Copper cable channels

The 25G and 50G channels shall conform to all of the channel characteristics defined in IEEE Std 802.3bj, Clause 92 and 802.3by Clause 110 “Physical Medium Dependent (PMD) sublayer and baseband medium”, specifically those defined in 92.9 and 110.9. There are a number of connector and cable combinations which can meet these requirements.

3.4.2 Backplane channels

The 25G and 50G channels shall conform to all of the channel characteristics defined in IEEE Std 802.3bj, Clause 93 and 802.3by Clause 111 “Physical Medium Dependent (PMD) sublayer and baseband medium”, specifically those defined in 93.9 and 111.9.

3.4.3 Twin-axial cable channel loss allocation (informative)

Detailed signal integrity and channel budgeting is not within the scope of this document. For more information, please refer to 802.3by annex 110, and 802.3bj annex 92.

3.5 Interconnect Annex A. 50G over QSFP28 MDI connector

This annex defines an MDI connector for 50GBASE-CR2 / 50GBASE-SR2 links based on a QSFP+ 28 Gb/s 4X Pluggable (QSFP28) connector. It is based on the Style-1 100GBASE-CR4 connector defined in IEEE 802.3 92.12.1.1. The annex does not preclude definition of additional connectors for 50GBASE-CR2 / 50GBASE-SR2 links.

3.5.1 50GBASE-CR2 / 50GBASE-SR2 over QSFP28 lane assignment

Each QSFP28 connector can accommodate 2x 50GBASE-R2 logical ports: Port 0, Port1. When only a single port is in use (depopulated QSFP) then port 0 shall be connected.

The lane assignment of Port0, Port1 is as follows:

1. Port 0, lane 0 – 100GBASE-CR4 lane 0, QSFP lane 1 (TX1/RX1)
2. Port 0, lane 1 – 100GBASE-CR4 lane 1, QSFP lane 2 (TX2/RX2)
3. Port 1, lane 0 – 100GBASE-CR4 lane 2, QSFP lane 3 (TX3/RX3)
4. Port 1, lane 1 – 100GBASE-CR4 lane 3, QSFP lane 4 (TX4/RX4)

The lane mapping results in the following lane to MDI connector contact mapping:

Tx lane	100GBASE-CR4 corresponding Tx lane	MDI connector contact	Rx lane	100GBASE-CR4 corresponding Rx lane	MDI connector contact
signal gnd	signal gnd	S1	signal gnd	signal gnd	S13
Port0_SL1<n>	SL1<n>	S2	Port1_DL0<p>	DL2<p>	S14
Port0_SL1<p>	SL1<p>	S3	Port1_DL0<n>	DL2<n>	S15
signal gnd	signal gnd	S4	signal gnd	signal gnd	S16
Port1_SL1<n>	SL3<n>	S5	Port0_DL0<p>	DL0<p>	S17
Port1_SL1<p>	SL3<p>	S6	Port0_DL0<n>	DL0<n>	S18
signal gnd	signal gnd	S7	signal gnd	signal gnd	S19
signal gnd	signal gnd	S32	signal gnd	signal gnd	S20
Port1_SL0<p>	SL2<p>	S33	Port0_DL1<n>	DL1<n>	S21
Port1_SL0<n>	SL2<n>	S34	Port0_DL1<p>	DL1<p>	S22
signal gnd	signal gnd	S35	signal gnd	signal gnd	S23
Port0_SL0<p>	SL0<p>	S36	Port1_DL1<n>	DL3<n>	S24
Port0_SL0<n>	SL0<n>	S37	Port1_DL1<p>	DL3<p>	S25
signal gnd	signal gnd	S38	signal gnd	signal gnd	S26

Table 1 – 50GBASE-CR2 / 50GBASE-SR2 lane to MDI connector contact mapping

Note:
The source lanes: SLi<p> and SLi<n> are the positive and negative sides of the transmitter's differential signal pairs for lane i (lanes 0-1 for 50G and lanes 0-3 for 100GBASE-CR4).
The destination lanes DLi<p> and DLi<n> are the positive and negative sides of the receiver's differential signal pairs for lane i (lanes 0-1 for 50G and lanes 0-3 for 100GBASE-CR4).

3.5.2 Use case example – 2 x 50GBASE-CR2 splitter cable

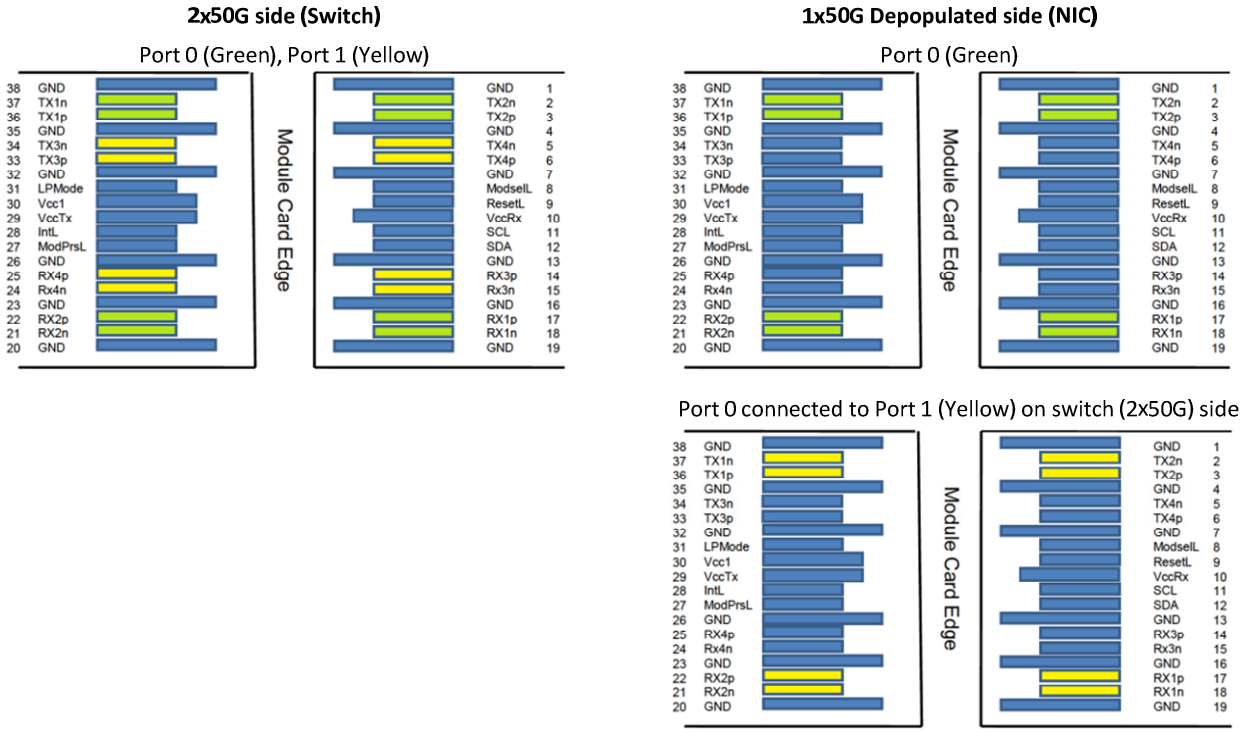


Figure 1 - 2 x 50GBASE-CR2 splitter cable

3.6 Interconnect Annex B. 25G , 50G QSFP28/SFP28 connectors management interface considerations (Informative)

The annex provides definition and examples for the management interface relevant to 25G and 50G interconnect using QSFP28/SFP28 connectors. For further information refer to http://sffcommittee.org

Document	Title	Revision	Relevant connectors
SFF-8024	SFF Committee Cross Reference to Industry Products (Includes cross references for various connectors, common fields, used for both QSFP28 and SFP28 connectors)	3.2	QSFP, SFP
SFF-8636	Management Interface for Cabled Environments (Defines the memory map for 4-lane connectors such as QSFP)	2.6	QSFP
SFF-8472	Diagnostic Monitoring Interface for Optical Transceivers (Defines the memory map for the SFP connector)	12.2	SFP

3.6.1 Relevant memory map fields

3.6.1.1 **Extended Specification Compliance Codes**

QSFP – address 192, SFP – address 36

Identifies the electronic or optical interfaces. For 25G / 50G interconnect the electrical or optical interfaces are identified on a per lane basis. For example a 50G copper cable with electrical lanes that comply with 25GBASE-CR CA-S will advertise 25GBASECR CA-S.

Note: for QSFP connectors, when the Extended Specification Compliance Codes is used, bit 7 of address 131 - Extended specification compliance codes should be set.

Relevant Extended Compliance Codes Values:*

Value	Description
01h	100G AOC (Active Optical Cable) or 25GAUI C2M AOC. Providing a worst BER of 5 × 10^(-5)
02h	100GBASE-SR4 or 25GBASE-SR
08h	100G ACC (Active Copper Cable) or 25GAUI C2M ACC. Providing a worst BER of 5 × 10^(-5)
0Bh	100GBASE-CR4 or 25GBASE-CR CA-L
0Ch	25GBASE-CR CA-S
0Dh	25GBASE-CR CA-N
18h	100G AOC or 25GAUI C2M AOC. Providing a worst BER of 10^(-12) or below
19h	100G ACC or 25GAUI C2M ACC. Providing a worst BER of 10^(-12) or below

* Additional compliance codes may be relevant, in case of misalignment between the annex and SFF-8024, the value from SFF-8024 should be used.

** C2M – Chip to Module interface as defined in IEEE 802.3 annex 83E, IEEE 802.3by annex 109B

3.6.1.2 **Free Side Device Properties**

QSFP – address 113, SFP – N/A

The free side devices properties is used to describe the interconnect connectivity. Cables/modules that implement only a subset of the lanes or breakout cables should use the field to describe the connectivity. A value of 0 is backward compatible to existing 4 lanes devices with unspecified far end implementation.

Note: in the field description and examples below that refer to SFF-8636, a channel (1-4) describes a lane of the QSFP connector.

Near End Implementation:

Bitmask, specifies which channels of the free side device at the near end are implemented. 0 indicates that a channel is implemented and 1 indicates that a channel is not implemented.

Bit	Description
0	0 - Channel 1 is implemented, 1 - Channel 1 is not implemented
1	0 - Channel 2 is implemented, 1 - Channel 2 is not implemented
2	0 - Channel 3 is implemented, 1 - Channel 3 is not implemented
3	0 - Channel 4 is implemented, 1 - Channel 4 is not implemented

Far End Implementation:

Specifies the type of the devices or devices that are implemented at the far end side.

Value	Description
000b	Far end is unspecified
001b	Cable with single far end with 4 channels implemented, or separable module with 4-channel connector
010b	Cable with single far end with 2 channels implemented, or separable module with 2-channel connector
011b	Cable with single far end with 1 channel implemented, or separable module with 1-channel connector
100b	4 far ends with 1 channel implemented in each (i.e. 4x1 break out)
101b	2 far ends with 2 channels implemented in each (i.e. 2x2 break out)
110b	2 far ends with 1 channel implemented in each (i.e. 2x1 break out)

3.6.2 25G/50G Cables Compliance Codes and Connectivity Examples

1. 50G Depopulated QSFP with CA-S cable spec

Cable end	Compliance Code	Far End Implementation	Near End Implementation
Both ends	25GBASE-CR CA-S	010 (Cable with single far end with 2 channels implemented)	1100b (channels 1,2 implemented)

2. 1 <=> 2 50G splitter cable with 100GBASE-CR4

Cable end	Compliance Code	Far End Implementation	Near End Implementation
Fully populated QSFP	100GBASE-CR4	101b (2x2 break out)	0000b (all channels implemented)
Depopulated QSFPs	100GBASE-CR4	001b (Cable with single far end with 4 channels implemented)	1100b (channels 1,2 implemented)

3. 1 <=> 4 25G SFP splitter cable with CA-N cable spec

Cable end	Compliance Code	Far End Implementation	Near End Implementation
QSFP	25GBASE-CR CA-N	100 (4x1 break out)	0000b (all channels implemented)
SFPs	25GBASE-CR CA-N	N/A	N/A

4. 1 <=> 2 25G SFP splitter AOC with 5 x 10^-5 BER

Cable end	Compliance Code	Far End Implementation	Near End Implementation
QSFP	100G AOC or 25GAUI C2M AOC. BER <= 5 ×10^(-5)	110 (2x1 break out)	1100b (channels 1,2 implemented)
SFPs	100G AOC or 25GAUI C2M AOC. BER <= 5 ×10^(-5)	N/A	N/A

5. 1 <=> 4 25G depopulated QSFP splitter cable with CA-N cable spec

Cable end	Compliance Code	Far End Implementation	Near End Implementation
Fully populated QSFP	25GBASE-CR CA-N	100 (4x1 break out)	0000b (all channels implemented)
Depopulated QSFPs	25GBASE-CR CA-N	001 (Cable with single far end with 4 channels implemented)	1110b (channel 1 implemented)