Document Number	**EDCS-1155168**
Based on Template	**EDCS-189229**
Created By	**Amrik Bains**

USGMII Specification

The Universal Serial Gigabit Media Independent Interface (USGMII) is an extension of current SGMII and QSGMII. USGMII provides flexibility to add new features while maintaining backward compatibility. Previous definition/implementations cover single (SGMII) and quad (QSGMII) options. This specification define USGMII option to support 4x1GE/8 x1GE network ports and 5G/10G PHY/MAC SERDES speed respectively:

- Convey 4 to 8 ports of network data and port speed between a 10/100/1000 PHY and a MAC with significantly less signal pins than required for GMII & SGMII.
- Utilize a 8B/10B PCS to maintain compatibility with SGMII/QSGMII
- USGMII interface operates as Full Duplex, while network interface supports both half and full duplex modes.
- Ability to send Packet Channel Header to support programmable information exchange between PH and MAC, for example PTP time stamp from PHY to MAC to improve accuracy/jitter on encrypted PTP packet and MACSec is in the ASIC
- Can be configured to support backward compatibility with SGMII and QSGMII

Cisco Systems Intellectual Property

Cisco encourages others to adopt this specification. To the extent that Cisco has proprietary rights to information contained in this specification, any company wishing to use this specification may do so under reasonable, non-discriminatory terms, with reciprocity, to implement and fully comply with the specification.

The reasonable non-discriminatory terms are:

Cisco will not assert any claims of any patents that are owned or controlled by Cisco essential for implementing this specification against any party for making, using, selling, importing or offering for sale a product that implements the specification, provided, however, that Cisco retains the right to assert its patents (including the right to claim past royalties) against any party that asserts a patent it owns or controls (either directly or indirectly) against any product that implements this specification. Cisco retains the right to assert its patents against any product or portion there of that is not necessary for compliance with the specification.

Royalty-bearing licenses are available to anyone who prefers that option.

For information contact:

Standards-ipr@cisco.com

Modification History

Revision	Date	Originator	Comments
3.0	11/12/2015	Amrik Bains	First general release
3.1	05/12/2016	Amrik Bains	Typo Fix
3.2	09/19/2016	Amrik Bains	Typo fix for Q-USGMII speed from 10G to 5G
3.3	11/02/2016	Amrik Bains	Typo correction for Extension Field Type to: 01: Extension Field contains Tag/Signature associated with for time-stamp to be taken by the PHY
4.0	02/10/2017	Amrik Bains	Added Pre-emption Features – PCH changes
4.1	05/11/2017	Amrik Bains	With pre-emption, additional PTP enable is required on egress – see table 5
4.2	08/17/23	Dale Mohlenhoff	Clarification on Intellectual Property and remove Cisco Confidential footer

Table of Content

Table of Figures

Table of Tables

Definitions

MII - Media Independent Interface: A digital interface that provides a 4-bit wide datapath between a 10/100 Mbit/s PHY and a MAC sublayer. Since MII is a subset of GMII, in this document, we will use the term "GMII" to cover all of the specification regarding the MII interface.

GMII- Gigabit Media Independent Interface: A digital interface that provides an 8-bit wide datapath between a 1000 Mbit/s PHY and a MAC sublayer. It also supports the 4-bit wide MII interface as defined in the IEEE 802.3z specification. In this document, the term "GMII" covers all 10/100/1000 Mbit/s interface operations**.**

SGMII- Serial Gigabit Media Independent Interface: A digital interface that provides a 1.25 Gbps serial dual-data-rate datapath between a 1000 Mbit/s PHY and a MAC sublayer. Refer to ENG-46158 or *ftp://ftp-eng.cisco.com/smii/smii.html* for details.

QSGMII- Quad Serial Gigabit Media Independent Interface: A digital interface that provides a 5.0 Gbps serial datapath between four 1000 Mbit/s PHY ports and a MAC sublayer. Refer to EDCS-540103 or *ftp://ftp-eng.cisco.com/smii/smii.html* for details.

USGMII - Universal Serial Gigabit Media Independent Interface: A digital interface that provides capability to carry multi-port/multi-rate serial datapath between PHY ports and a MAC sublayer using 8B/10B coding. Refer to EDCS 1155168

USXGMII- Universal Serial 10 Gigabit Media Independent Interface: A digital interface that provides capability to carry multiport/multi-rate serial datapath between PHY ports and a MAC sublayer using 64B/66B coding. Refer to appropriate specification based on media type, network data rate and PHY/MAC SERDES speed.

LPI- Low Power Idle: An alternative form of idle signaling that is used by the MAC to indicate that the PHY may enter a low power state and signal this change of state to the link partner; and is used by the PHY to signal to the MAC that the link partner has entered a low power state. The functions are defined by IEEE 802.3az in IEEE 802.3 clauses 22, 24, 25 (for 100Mb/s); 35, 36, 40, 70 (for 1Gbps); 46, 48, 49, 55, 71, 72 (for 10Gbps); and 78 (for overall descriptions).

1 Overview

USGMII uses two data signals in each direction to convey frame data and link rate information between a multi-port 10/100/1000 PHY and Ethernet MAC. From architecture perspective, the data rate using CDR technology to recover the clock at the MAC and PHY interfaces can be any rate depending on number of ports and maximum speed per port. The SERDES rate can be as high as 10Gbps (8x1GE). Due to the high speed of operation, each of these signal pairs are realized as differential pairs thus optimizing signal integrity while minimizing system noise. The table 1, below show current and next generation interface in term of number of port, port speed and SERDES data rate.

Number of Data ports	**Data Speed per port**	**Number of Part**	**Maximum SERDES Speed (Gbps)**	**Comment**
1-port 10/100/1000M (SGMII)	10/100/1000M	1	1.25	One Port, no PCH
4-port 10/100/1000M (QSGMII)	10/100/1000M	4	5.0	Maximum of 4 ports, no PCH
4-port 10/100/1000M (Q-USGMII)	10/100/1000M	4	5.0Gbps	Maximum of 4 ports with PCH
8-port 10/100/1000M (O-USGMII)	10/100/1000M	8	10.0Gbps	Maximum of 8 ports with PCH

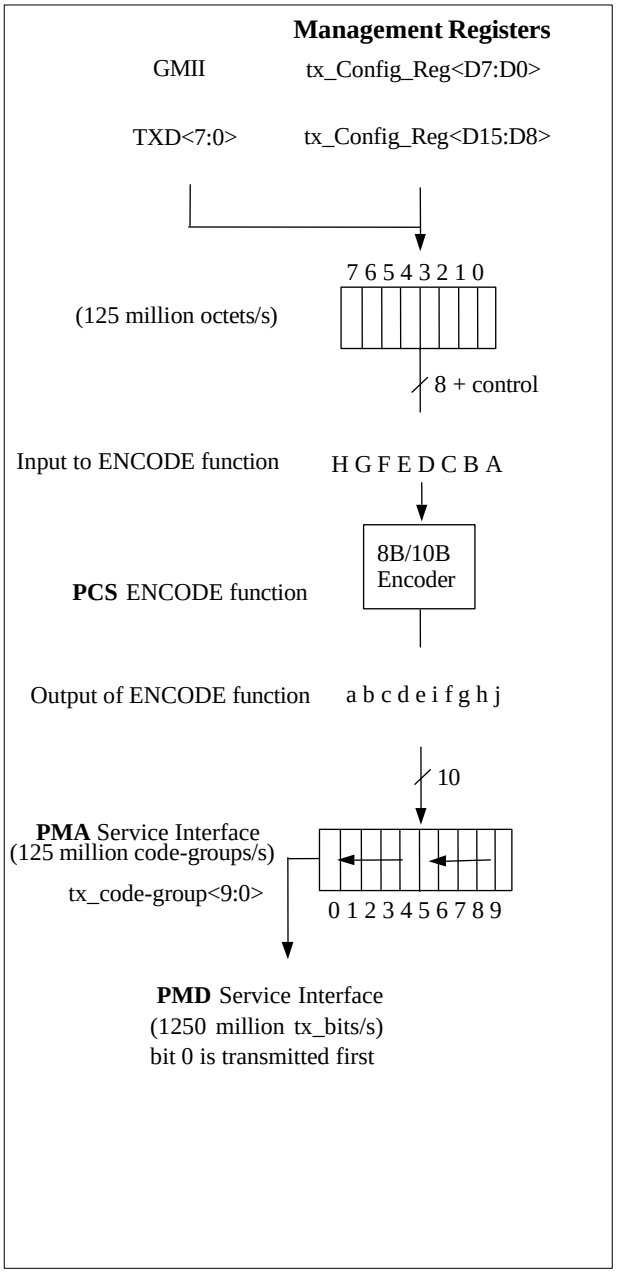
Table 1: Examples of Current and Next Generation Interface Options

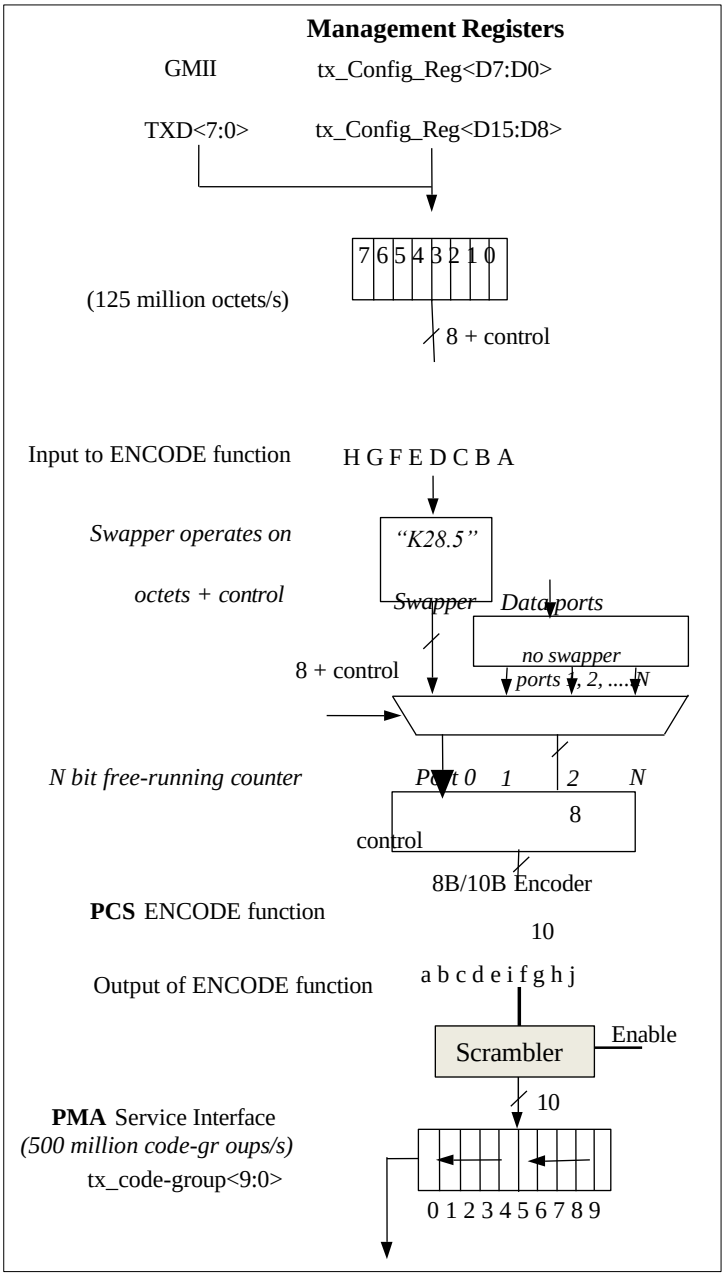
The other SERDES rates could be supported based on technology capability and network port data rate requirements

A simple round-robin scheme is used to interleave 8-bit words from each port and multiplex to create data stream that is encoded using 8B/10B PCS.

For 10Gbps operation, an optional/configurable scrambler/de-scrambler is placed at after the Encoder and before Decoder to reduce effect of repetitive patterns that may occur with 8B/10B coding. Refer to section 2.2.3 for more details.

Figure 1, compares the IEEE 802.3 PCS reference diagram before and after the USGMII modification.



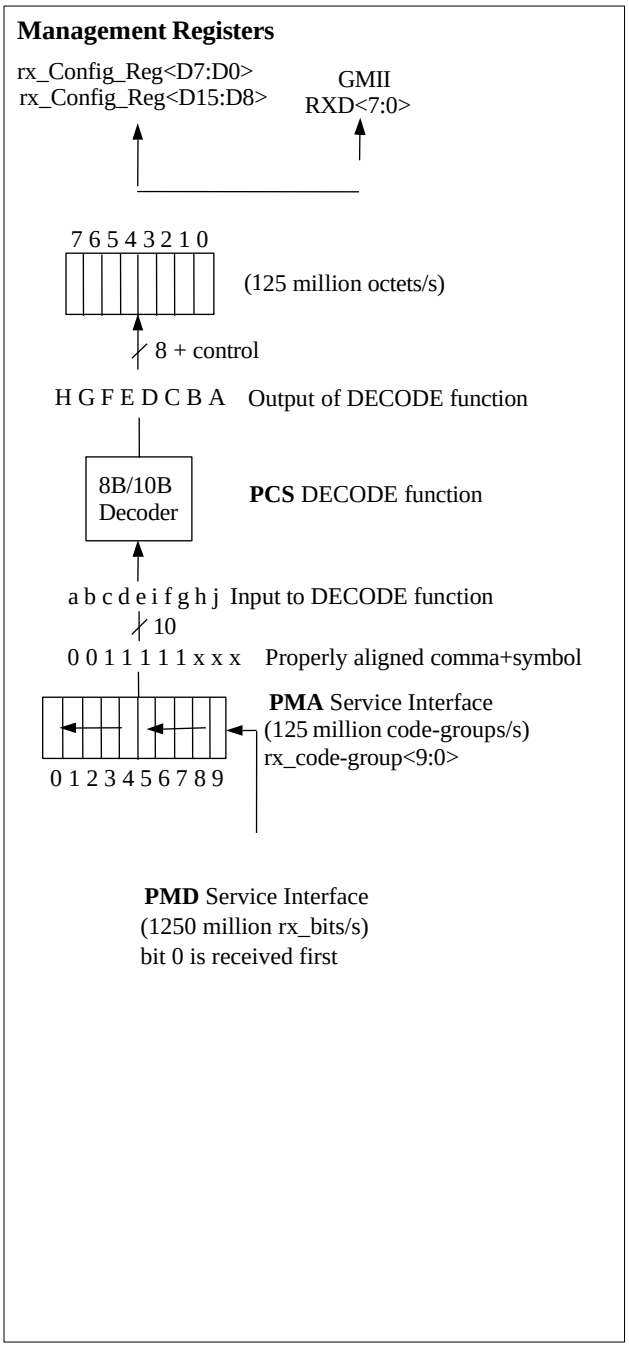


IEEE 802.3 Figur e 36-3-PCS r eference diagram

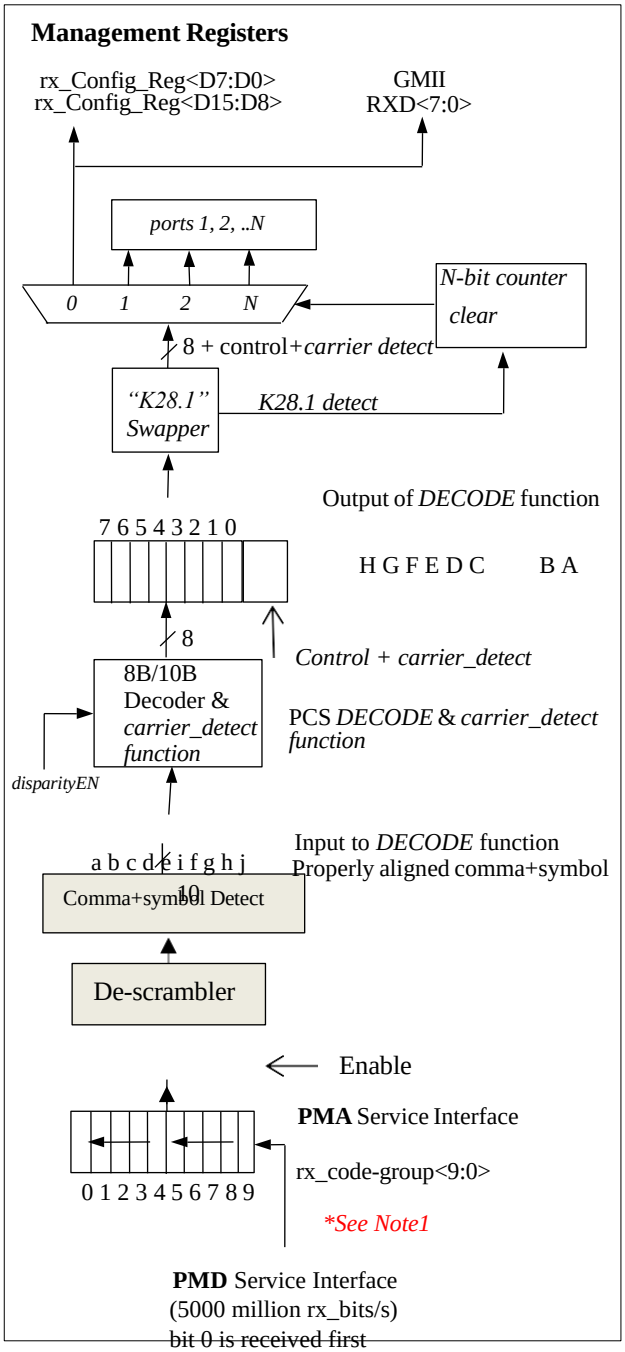
USGMII *Modified* Figure 36-3-PCS r eference diagram
(changes are shown with it alicized text)

Figure1: Standard (IEEE 802.3 PCS/PMA) and Modified Transmit Path Diagram

The IEEE 802.3 reference diagrams in Figure 2, shows where the modification to the receive PCS function occurs.



IEEE 802.3 Figure 36-3-PCS reference diagram



USGMII *Modified* Figure 36-3-PCS reference diagram
(changes are shown with italicized text)

Figure 2: Standard (IEEE 802.3 PCS/PMA) and Modified Transmit Path Diagrams

The transmit and receive data paths leverage the 1000BASE-X PCS defined in the IEEE 802.3z specification (clause 36). Traditional GMII data signals (TXD/RXD), data valid signals

(TX_EN/RX_DV), and error signals (TX_ER/RX_ER) are muxed, encoded, and serialized. Carrier Sense (CRS) is derived/inferred from RX_DV and collision (COL) is logically derived in

the MAC when RX_DV and TX_EN are simultaneously asserted. There is a small block in the PHY transmit path to suppress TX_ER in full duplex mode when TX_EN is not asserted. For example, four 1.25Gbps SGMII ports are interleaved onto a single link, the data-rate becomes 5.0 Gbps. Refer to Table 1 for more details

Figure 3, illustrates the resulting bit times on the SGMII, Quad-SGMII and Octal-SGMII options.

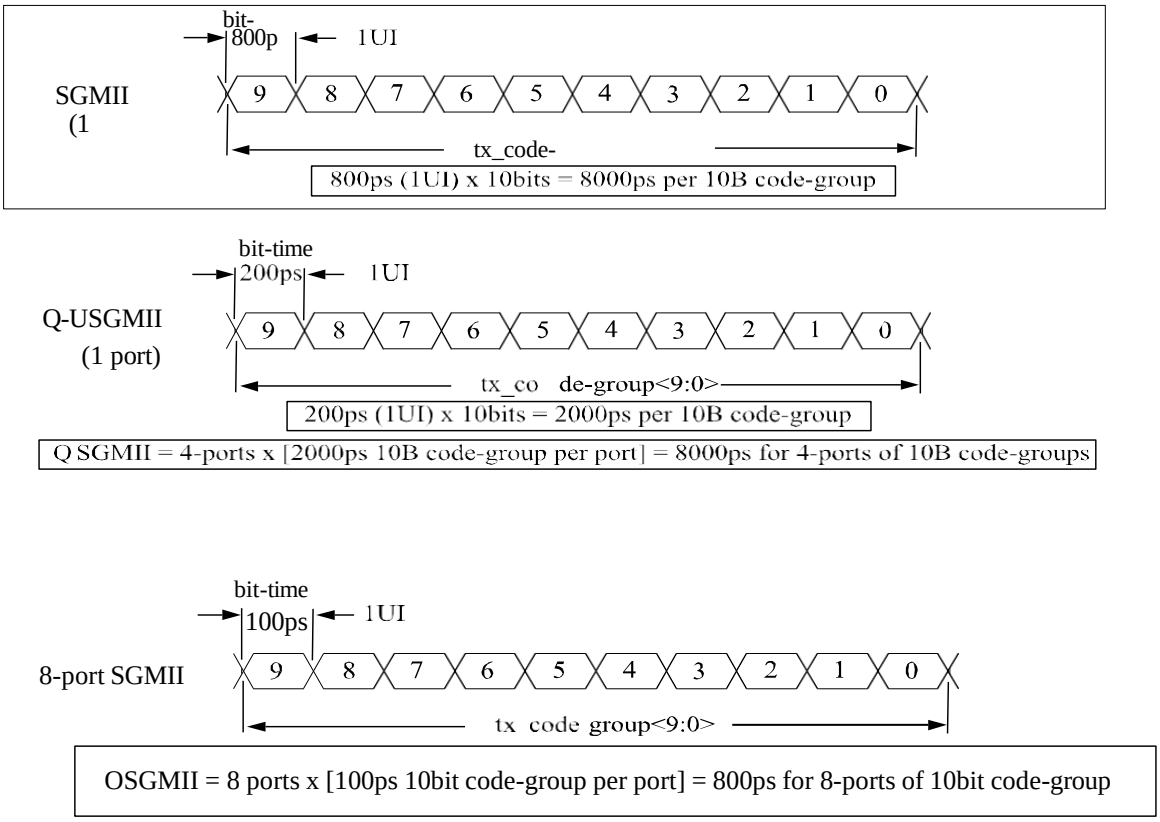


Figure 3: Examples of Bit Positions across Different Interface Mode

In order to determine the port number based time slots the port 0 transmit side incorporates a “K28.5” swapper function that modifies the IDLE /I/ and Configuration /C/ ordered_sets by replacing /K28.5/ with /K28.1/ every time /K28.5/ occurs as shown in table 2. Note that the swapper operates on the GMII octets (8 + control), not on the 10B code-group directly.

The data will appear on the USGMII link in the order: port 0 first, then port 1, then port 2, and lastly port N. Port 0 data then appears on the link again, and so on.

802.3z Section 36.2.4.12 explains the rules for running disparity by sending out one of the two IDLE /I/ ordered_sets whenever the GMII is idle. However, since 8B/10B encoder is detached from PCS Transmit Function, it is no longer feasible to use /I1/ and /I2/ ordered_sets to force the disparity. Therefore, the transmitter may be simplified to only generate /I1/ ordered_sets. This change requires more functionality from the framer as documented in Note1.

Note1: USGMII Receivers should not rely upon receipt of /I2/ ordered_sets for proper operation.

Code	Ordered_Set	Number of Code Groups	Port 0 "pre-swapper" Encoding	Port 0 "post-swapper" Encoding
/C/	**Configuration**		Alternating /C1/ and /C2/	Alternating /C1/ and /C2/
/C1/	Configuration 1	4	/K28.5/D21.5/Config_Reg	/K28.1/D21.5/Config_Reg
/C2/	Configuration 2	4	/K28.5/D2.2/Config_Reg	/K28.1/D2.2/Config_Reg
/I/	**IDLE**		Correcting /I1/	Correcting /I1/
/I1/	IDLE 1	2	/K28.5/D5.6/	/K28.1/D5.6/

Table 2: Port 0 "K28.5" Swapper Definition

The receive 10B code-groups pass through a "K28.1" swapper that undoes the modification of the IDLE /I/ and Configuration /C/ ordered_sets by replacing /K28.1/ with /K28.5/ for every occurrence. The K28.1 swapper also clears the de-mux (sets the counter to 0) in order to determine the port 0. Note that the swapper operates on the GMII octets (8 + control + carrier_detect), not on the 10B code-group directly.

On the receive side, carrier_detect function is done on 10B code-groups and this is shown in Figure 2. Please note that, there is a new carrier_detect function that needs to operate on K28.1 for Port0. K28.1 can be locally converted to K28.5 to generate the carrier_detect function. Please refer to 802.3z Section 36.2.5.1.4 for the carrier_detect function that operates on K28.5.

Due to the nature of USGMII, bit errors on the link may cause a running disparity error to propagate across ports. A software register bit that would enable/disable running disparity checking at the receiver is required. Disabling running disparity checking at the receiver prevents error propagation to other ports. It is not necessary to disable ALL disparity checking in the decoder to prevent error propagation to other ports. It is only necessary to disable the disparity checks that rely on the running disparity value from the previous symbol. Note that code violations due to invalid code-words (and current symbol running disparity errors) should continue to be detected regardless of the state of running disparity checking. Please also note that, 802.3z Section 36.2.4.6 and DECODE ([/x/]) function in Section 36.2.5.1.4 will be affected by this requirement.

Code	Ordered_Set	Number of Code Groups	Port 0 ”pre-swapper” Encoding	Port 0 “post-swapper” Encoding
/C/	**Configuration**		Alternating /C1/ and /C2/	Alternating /C1/ and /C2/
/C1/	Configuration 1	4	/K28.1/D21.5/Config_Reg	/K28.5/D21.5/Config_Reg
/C2/	Configuration 2	4	/K28.1/D2.2/Config_Reg	/K28.5/D2.2/Config_Reg
/I/	**IDLE**		Correcting /I1/	Correcting /I1/
/I1/	IDLE 1	2	/K28.1/D5.6/	/K28.5/D5.6/

Table 3: Port 0 “K28.1” Swapper Definition

Network Port definition is unchanged from SGMII specification. A number of ports per interface are statically configured on PHY and MAC device after power up. After synchronization is completed, auto-neg message of each port within USGMII can be shared between PHY and MAC device.

Auto-neg control information is embedded into the GMII at each port to perform auto-negotiation.

The Packet Control Header is inserted in GMII in place of Pre-amble to provide programmable control/status between MAC and PHY. One use case is PTP Time Stamps and other features can be implemented in the future using the Extension Field (as defined by Cisco). This requires the Port ASIC and PHY to be able to operate with pre-amble defined in section 2.2.3.

1.1 Auto-neg Mechanism

Auto-neg Control information, as specified in Table3, is transferred from the PHY to the MAC to signal the change of the link status. This is achieved by using the Auto-Negotiation functionality defined in Clause 37 of the IEEE Specification 802.3z. Instead of the ability advertisement, the PHY sends the control information via its tx_config_Reg [15:0] as specified in Table 3, whenever the link status changes. Upon receiving control information, the MAC acknowledges the update of link status by asserting bit 14 of its tx_config_reg [15:0].

The link_timer inside the Auto-Negotiation block has been changed from 10 msec to 1.6 msec to ensure a prompt update of the link status.

Bit Number	tx_config_Reg[15:0] sent from the PHY to the MAC	tx_config_Reg[15:0] sent from the MAC to the PHY
15	Link: 1 = link up, 0 = link down	Same function as PHY to MAC
14	Reserved for Auto-Negotiation acknowledge as specified in 802.3z	1
13	0: Reserved for future use	0: Reserved for future use
12	Duplex mode: 1 = full duplex, 0 = half duplex	Same function as PHY to MAC: Duplex mode: 1 = full duplex, 0 = half duplex
11:9	Speed: Bit 11, 10, 9: 1 1 1 to 011 = Reserved 0 1 0 = 1000 Mbps: 1000BASE-TX, 1000BASE-X 0 0 1 = 100 Mbps: 100BASE-TX, 100BASE-FX 0 0 0 = 10 Mbps: 10BASET, 10BASE2, 10BASE5	Same function as PHY to MAC: Speed: Bit 11, 10, 9: 1 1 1 to 011 = Reserved 0 1 0 = 1000 Mbps: 1000BASE-TX, 1000BASE-X 0 0 1 = 100 Mbps: 100BASE-TX, 100BASE-FX 0 0 0 = 10 Mbps: 10BASET, 10BASE2, 10BASE5
8	EEE capability:1= supported, 0 = not supported	EEE capability:1= supported, 0 = not supported
7	EEE clock stop capability: 1= supported, 0 = not supported	EEE clock stop enable: 1= enabled, 0 = not enabled
6:1	0: Reserved for future use	0: Reserved for future use
0	1	1

Table 4: Definition of Control Information passed between links via tx_config_Reg [15:0]

USGMII's rate can be much higher than total required by network port. For example Q-USGMII /OUSGMII 5.0/10 Gbps transfer rate is excessive for PHYs operating at 10 or 100 Mbit/s. When these situations occur, the interface "elongates" the frame by replicating each frame byte by appropriate ratio.

For Q-USGMII (4x1GE), frame is elongated by 10 times for 100 Mbit/s and 100 times for 10 Mbit/s. For O-USGMII (8x1GE), frame is elongated by 10 times for 100 Mbit/s and 100 times for 10 Mbit/s. This frame elongation takes place "above" the 802.3z PCS layer, thus the start frame delimiter only appears once per frame. The 802.3z PCS layer may remove the first byte of the "elongated" frame.

Auto-neg feature must be compatible with current SGMII and QSGMII specs.

1.2 Packet Control Header

Packet Control Header provides optional features such as PTP Time Stamp and capability to add features in the future via re-use of Extension Field described below.

PHY communicates with a port MAC (ASIC) through Packet Control Header (PCH). PCH is 8 bytes and it replaces the preamble of the frame.

.

The details of PCH and fields are shown in *Figure4.*

Figure 4: Packet Control Header (PCH) Format

Bit	UsgmiiPCH[47:0] sent from the PHY to the MAC	UsgmiiPCH[47:0] sent from the MAC to the PHY	Default
47:46	Packet type: 00: Ethernet Packet with PCH (packet information) 01: Ethernet packet, without PCH (packet information) 10: Idle Packet – Contains status data for a port – no packet data 11: Reserved	Packet type: 00: Ethernet Packet with PCH 01: Ethernet packet, without PCH (packet information) 10: Idle Packet – Contains status data for a port – no packet data 11: Preemption Frame, aka Interspersing Express Traffic (IET) frame	0
45:42	Subport ID (Channel number) 3:0: Port 0 to 15	Subport ID (Channel number): MAC to PHY 3:0: Port 0 to 15	0
41:40	Extension Field Type 00: Ignore Extension Field 01: Extension Field contains 8-bit Reserved + 32-bit Timestamp 10: Extension Field: For devices that support Preemption aka Interspersing Express Traffic (IET) frames, otherwise Reserved. It also contains fields related to PTP Timestamp	Extension Field Type 00: Ignore Extension Field 01: Extension Field contains Extension Field contains Tag/Signature associated with for time-stamp to be taken by the PHY 10: Extension Field: For devices that support Preemption aka Interspersing Express Traffic (IET) frames, otherwise Reserved. It also contains Tag/Signature associated with for time-stamp to be taken by the PHY.	0

Bit	UsgmiiPCH[47:0] sent from the PHY to the MAC	UsgmiiPCH[47:0] sent from the MAC to the PHY	Default
	11: Reserved	11: Reserved	
39:0	Extension Field Extension Type = 00: Extension Field is ignored Extension Type = 01: • 39:32: Reserved. • 31:0: PTP Timestamp Extension Type = 10: Preemption and PTP • 39:38: Preemption (IET) Frame Type/PTP 00: High-priority frame (eMAC) 01: Low-priority pre-emptable frame or initial fragment (pMAC) 10: Low-priority intermediate fragment or last fragment (pMAC) 11: Verify/Respond frame • 37:36: Type of Verify/Respond frame 00: Verify frame received 01: Respond frame received 10/11: Reserved • 35:34: Frame Count is a 2-bit binary Fragment Count of low-priority Fragment • 33:32: FRAG_COUNT is a 2-bit binary Fragment Count of low-priority Intermediate or Last Fragment • 31:0: PTP Timestamp Extension Type = 11 are reserved	Extension Field Extension Type = 00: Extension Field is ignored Extension Type = 01: • 39:32: Reserved • 31:16:Resvered • 15:0 Signature associated with the egress PTP Timestamp to be done in PHY Extension Type = 10:Preemption and PTP • 39:38: Preemption (IET) Frame Type/PTP 00: High-priority frame (eMAC) 01: Low-priority preemptable frame or initial fragment (pMAC) 10: Low-priority intermediate fragment or last fragment (pMAC) 11: Verify/Respond frame • 37:36: Type of Verify/Respond 00: Verify 01: Respond 10/11: Reserved • 35:34: Frame Count is a 2-bit binary Fragment Count of low-priority Fragment • 33:32: Fragment Count is a 2-bit binary Fragment Count of low-priority Intermediate or Last Fragment • 31: PTPTimestampEnable – when enabled, PHY adds egress PTP Timestamp to the associated frame • 30:16: Reserved • 15:0 Signature associated with the egress PTP Timestamp to be done in PHY	Refer to PHY spec for full details

Bit	UsgmiiPCH[47:0] sent from the PHY to the MAC	UsgmiiPCH[47:0] sent from the MAC to the PHY	Default
		Extension Type = 11 are reserved	

Table 5: Definition of channel control information passed between links via UsgmiiPCH [47:0]

Six byte UsgmiiPCH message is sent in the pre-amble of the packet. It is inserted between SOP (0xfb) and Header CRC. Refer to section 2.2.3 for more details.

2 Implementation Specification

This section discusses how USGMII interface shall be implemented by incorporating and modifying the PCS layer of the IEEE Specification 802.3z. Figure 5, illustrates the connections/features between the MAC and the PHY in a system utilizing USGMII, as well as the overall scheme of muxing, demuxing N-ports (including status/control channels) into a single USGMII channel and scrambler/de-scrambler. N is configured via control interface (e.g. MDIO) based on PHY and MAC device capabilities.

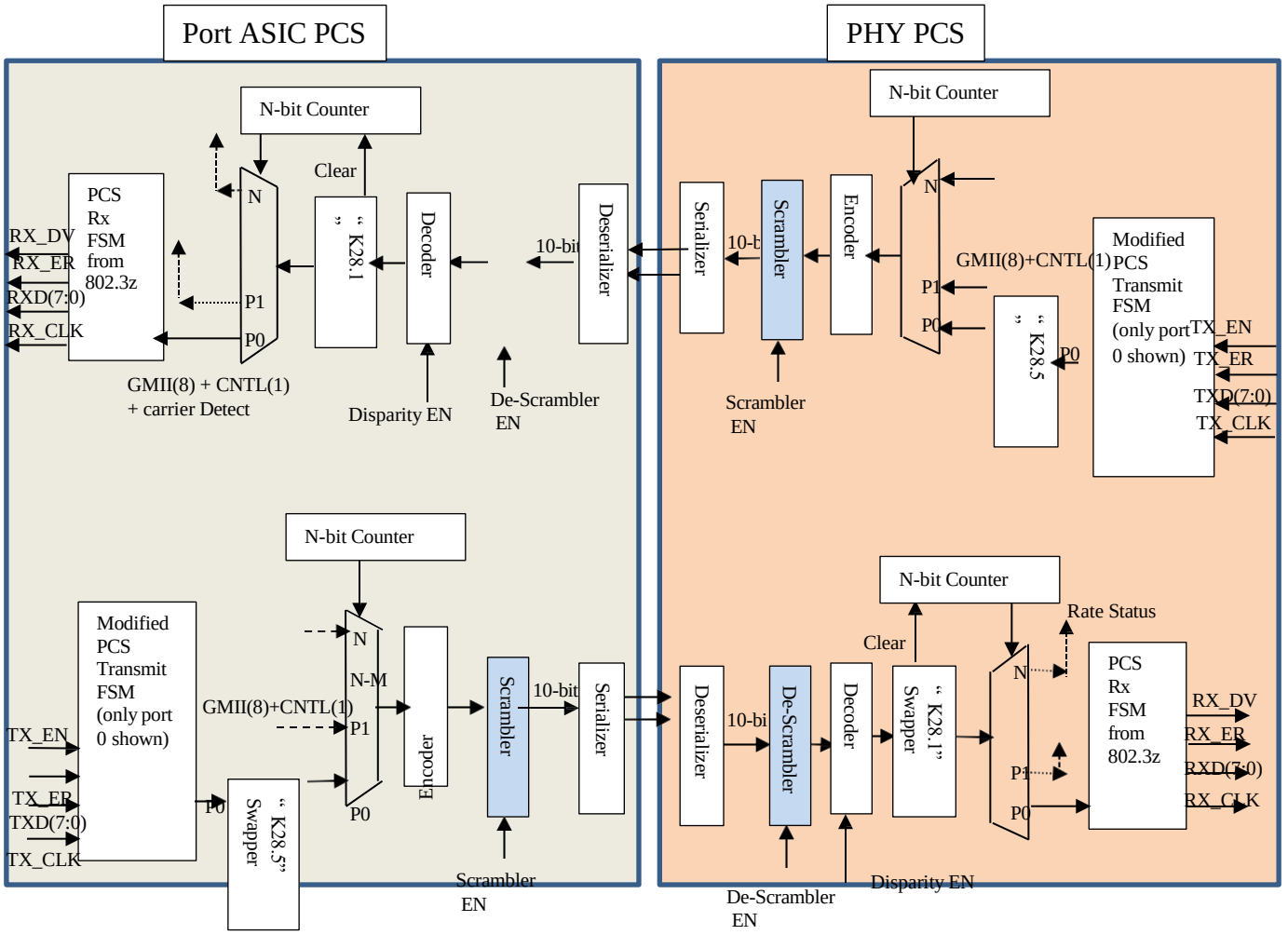


Figure 5: USGMII Features/Connectivity

2.1 Signal Mapping at the PHY side

Figure 6, shows the PHY functional block diagram. It illustrates function required per port as well as the common functions between ports such as Mux/Demux, Scrambler/De-scrambler and the PCS layer shall be modified and incorporated at the PHY side in the USGMII interface.

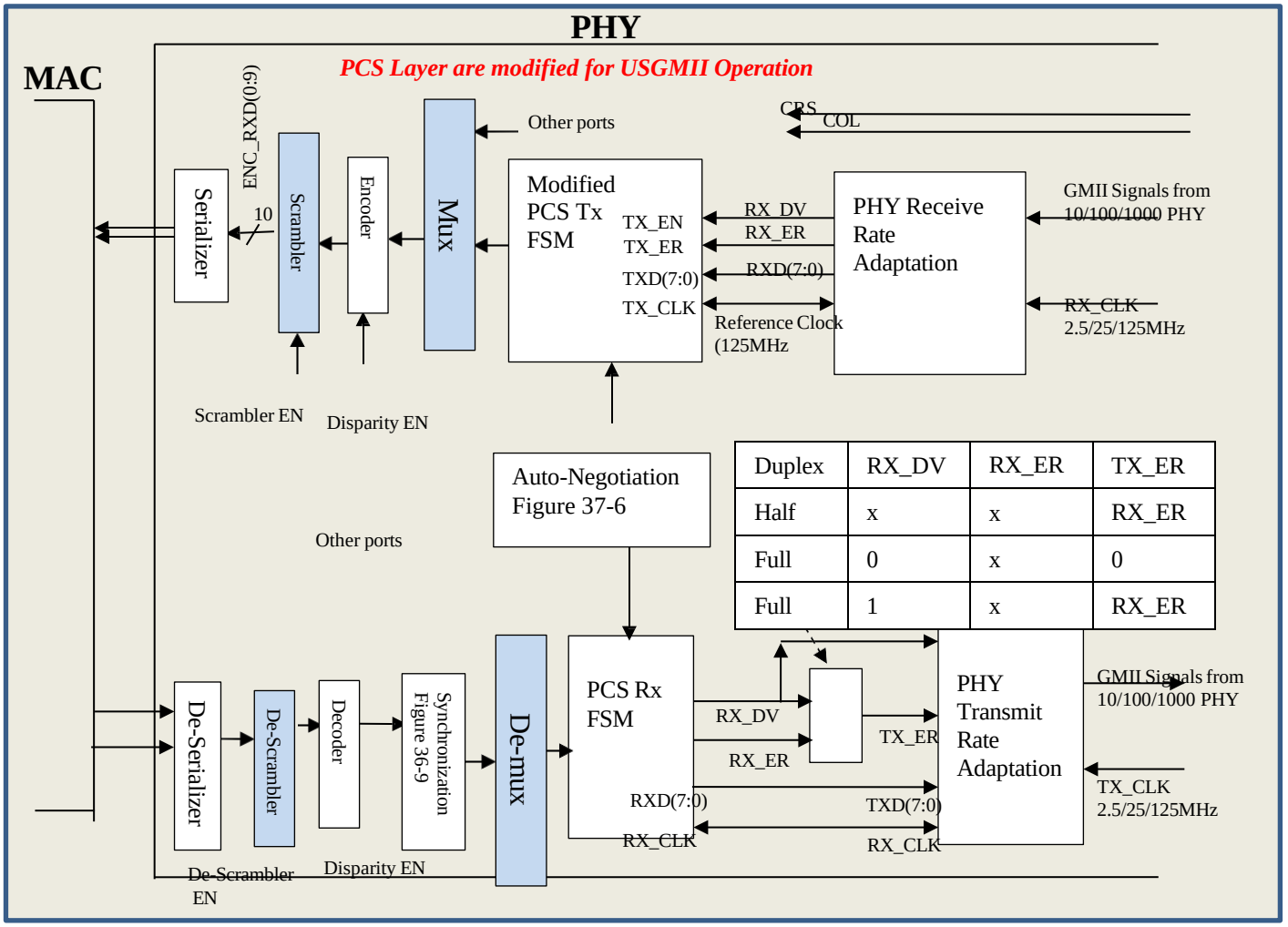


Figure 6: Phy Functional Block (One Port)

At the receive side (network Rx), GMII signals come in at 10/100/1000 Mbit/s clocked at 2.5/25/125 MHz The PHY passes these signals through the PHY Receive Rate Adaptation to output the 8-bit data RXD[7:0] in Reference Clock domain. Please note that since N ports are multiplexed, recovered RX_CLK cannot be used to output the 8-bit data RXD [7:0]. RXD is sent to the PCS Transmit State Machine which performs control symbols and auto-neg features. Tx

FSM sends RxD and control symbols to the Mux block.

Mux block performs muxing of ports data/control symbols, padding according to N/M configuration. Output of Mux (data/control symbols) is sent to the Encoder. 802.3z Section 36.2.4.12 explains the rules for running disparity by sending out one of the two IDLE /I/ ordered_sets whenever the GMII is idle. However, since 8B/10B encoder is detached from PCS Transmit Function, it is no longer feasible to use /I1/ and /I2/ ordered_sets to force the disparity. Therefore, the transmitter may be simplified to only generate /I1/ ordered_sets. This change requires more functionality from the framer as documented in Note1.

Due to the nature of USGMII, bit errors on the link may cause a running disparity error to propagate across ports. A software register bit that would enable/disable running disparity checking at the receiver is required. Disabling running disparity checking at the receiver prevents error propagation to other ports. It is not necessary to disable ALL disparity checking in the decoder to prevent error propagation to other ports. It is only necessary to disable the disparity checks that rely on the running disparity value from the previous symbol. Note that code violations due to invalid code-words (and current symbol running disparity errors) should continue to be detected regardless of the state of running disparity checking. Please also note that, 802.3z Section 36.2.4.6 and DECODE ([/x/]) function in Section 36.2.5.1.4 will be affected by this requirement.

The 8B/10B encoding scheme makes it possible to create sequences of symbols that have long runs of alternating 1s and 0s (i.e. 1010101...) followed by patterns with lower frequency contents (e.g. a repeated sequence of 01111000111000011100). Due to differential group delay (i.e. the difference in propagation delay between high-frequency and low-frequency components of the signal) such a transition (from a fast changing serial stream to slow changing serial stream or vice versa) may show up as a phase shift in the recovered clock (a.k.a. data dependent jitter). At 10Gbps, the bit period is only 100ps and the eye opening is in the 20-50 ps range. Obviously, it does not take a huge phase shift to sample outside the eye opening potentially causing a bit error.

To avoid pattern based signal integrity issues the output of the Encoder are scrambled for 10Gbps operation. Scrambler is enabled (for 10Gbps), the code-group from Encoder are scrambled using IEEE 802.3 Clause 49 self-synchronous LSFR (see section 2.2.3).

The PHY serializes ENC_RXD [0:9] or Scrambled (10Gbps) to create RX and sends it to the MAC at N x Max Port Speed. For example, QSGMII speed is 5.0 Gbps and OSGMII is 10Gbps.

At the transmit side, the PHY de-serializes and de-scrambles (if enabled) TX to recover encoded ENC_TXD [0:9]. The PHY passes ENC_TXD [0:9] through the PCS Decoder to recover data/control symbols. The data symbols are De-muxed to data and control per port. The Receive State Machine to recover the GMII signals. In the meantime, Synchronization block checks ENC_TXD [0:9] to determine the synchronization status between links, and to realign if it detects the loss of synchronization. The decoded GMII signals have to pass the PHY Transmit Rate Adaptation block to output data segments according to the PHY port speed.

To make the PCS layer from 802.3z work properly, the PHY must provide a frame beginning with pre-amble defined in section 2.2.3.

Some legacy end points will drop frames when RX_ER asserts during the first clock after a frame ends. The receive PCS state machine generates this signaling at the end of certain frames.

To avoid this problem, there is a small block in the PHY transmit path to suppress TX_ER in full duplex mode when RX_DV (from the PHY receive PCS state machine) is not asserted.

Low Power Idle operation requires that the LPI symbols are passed across the USGMII in the manner defined in 802.3 Clause 36. If the EEE clock stop capability is asserted (by the PHY) then the MAC may stop the signals on the USGMII may be stopped in the same manner as 1000BASE-KX (defined in 802.3 clauses 36 and 70), if and only if all N transmit channels are in transmitting LPI. If the EEE clock stop enable is asserted (by the MAC) then the PHY may stop the signals on the USGMII in the same manner. In cases where the MAC or PHY requires constant signaling on the USGMII (i.e. EEE clock stop enable or capable = 0), it is assumed that the LPI signal latency is the same as the datapath signal latency and therefore the PHY timing constraints remain unchanged. In cases where the MAC allows the USGMII clocking to be stopped during Low Power Idle, the system must negotiate an extra 20 uS of wake time via LLDP (see 802.3 clause 78) in the receive direction on all N channels to allow time for the USGMII to wake and resynchronize after the PHY has woken. Similarly, if the MAC stops the USGMII clocking in the transmit direction, then the system must wait for an additional 20uS between sending normal IDLE on the first channel to wake and sending data on any channel to allow extra time for the USGMII to wake and resynchronize in the transmit direction.

2.2 Signal Mapping at the MAC Side

Figure 7, shows the MAC functional block diagram. It illustrates that the PCS layer shall be modified and incorporated at the MAC side in the USGMII interface.

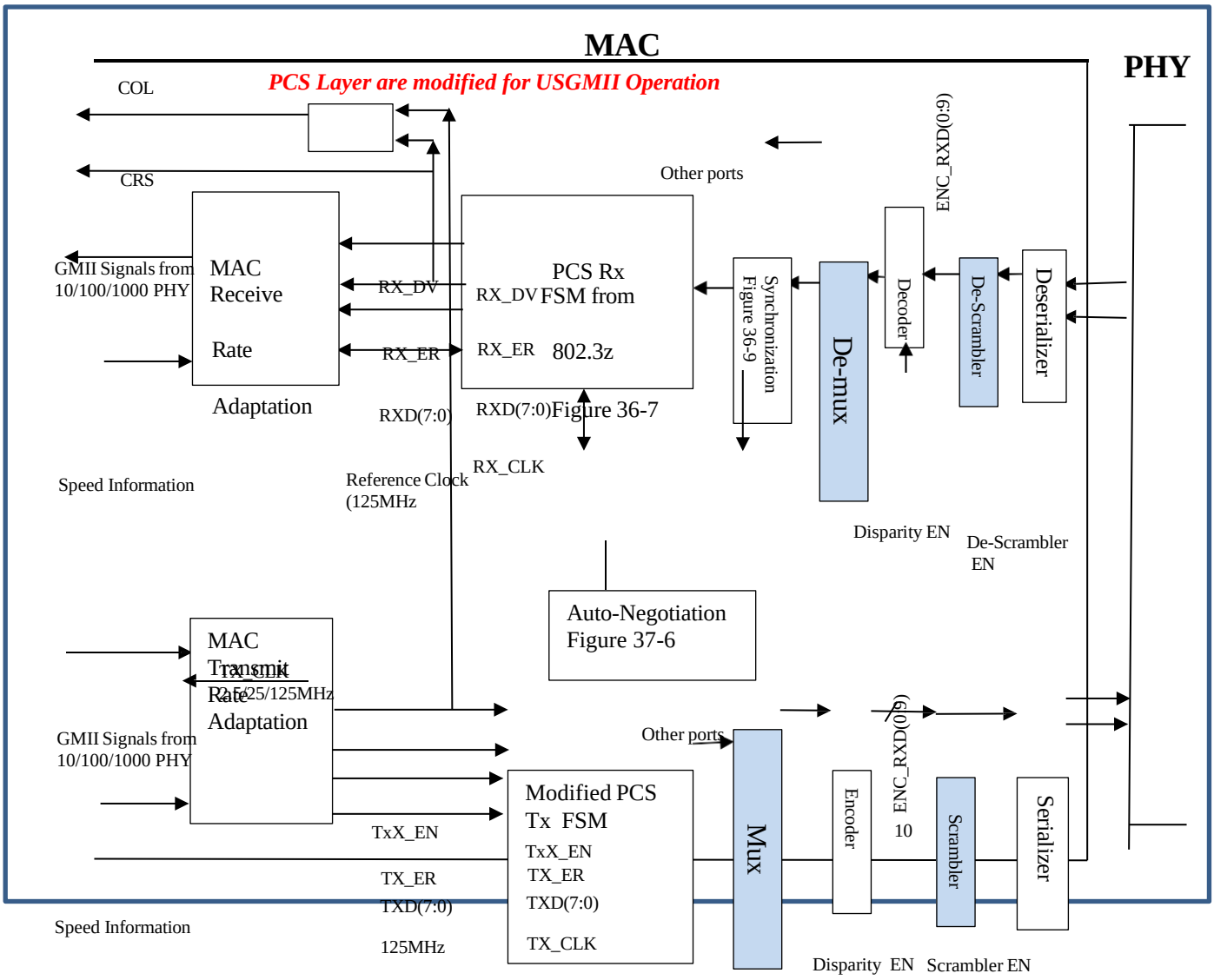


Figure 7: MAC Functional Block (One Port)

At the receive side, the MAC de-serializes and de-scrambles (if enables), RX to recover encoded ENC_RXD [0:9]. The MAC passes ENC_RXD [0:9] through the PCS Receive State Machine to recover the GMII signals. In the meantime, Synchronization block checks ENC_RXD [0:9] to determine the synchronization status between links, and to realign once it detects the loss of

synchronization. The decoded data/control symbols are Demuxed to recover port data/control. GMII signals have to pass the MAC Receive Rate Adaptation block to output data segments according to the PHY port speed, passed from the PHY to MAC via Auto-Negotiation process. The rate adjust bytes are removed/discarded as configured by Rate Adjust Valid register.

At the transmit side, GMII signals come in at 10/100/1000 Mbit/s data clocked at 2.5/25/125 MHz The MAC passes these signals through the MAC Transmit Rate Adaptation to output the 8-bit data TXD[7:0] in 125MHz clock domain. TXD is sent to the PCS Transmit State Machine to generate control symbols.

The rate adjust bytes are added as configured by Rate Adjust Valid register via the Mux. Output of the MUX is passed to Scrambler. If Scrambler is enabled data symbols are scrambled while control symbols are unaltered.

The MAC serializes ENC_TXD [0:9] to create TX and sends it to the PHY N x Maximum port speed.

2.2.1 Auto-neg Control Information Exchanged between Links

As described in Overview, it is necessary for the PHY to pass control information to the MAC to notify the change of the link status. USGMII interface uses Auto-Negotiation block to pass the control information via tx_config_Reg [15:0].

If the PHY detects the link change, it starts its Auto-Negotiation process, switching its Transmit block from “data” to “configuration” state and sending out the updated control information via tx_config_Reg [15:0]. The Receive block in the MAC receives and decodes control information, and starts the MAC’s Auto-Negotiation process. The Transmit block in the MAC acknowledges the update of link status via tx_config_Reg [15:0] with bit 14 asserted, as specified in Table 3. Upon receiving the acknowledgement from the MAC, the PHY completes the auto-negotiation process and returns to the normal data process.

As specified in Overview, inside the USGMII interface the Auto-Negotiation link_timer has been changed from 10 msec to 1.6 msec, ensuring a prompt update of the link status. The expected latency for the update of link is 3.4 msec (two link_timer time + an acknowledgement process).

2.2.2 Data Information Transferred Between Links

Below we briefly describe at receive side how GMII signals get transferred across from the PHY and recovered at the MAC by using the 8B/10B transmission code. The same method applies to the transmit side.

According to the assertion and de-assertion of RX_DV, the PHY encodes the Start_of_Packet delimiter (SPD /S/) and the End_Of_Packet delimiter (EPD) to signal the beginning and end of each packet. The MAC recovers RX_DV signal by detecting these two delimiters.

The PHY encodes the Error_Propagation (/V/) ordered_set to indicate a data transmission error. The MAC asserts RX_ER signal whenever it detects this ordered_set.

CRS is not directly encoded and passed to the MAC. To regenerate CRS, the MAC shall use signal RX_DV before it is being passed to the MAC Receive Rate Adaptation block as shown in figure 6.

The MAC decodes ENC_RXD [0:9] to recover RXD [7:0].

Figure 8, illustrates how the MAC samples data in 100 Mbit/s modes. The GMII data in 100 Mbit/s modes get replicated ten times after passing through the PHY Receive Rate Adaptation to generate RXD [7:0]. The modified PCS Transmit State Machine encodes RXD [7:0] to create ENC_RXD [0:9]. As noted in the Overview, the SPD (/S/) only appears once per frame.

SAMPLE_EN is a MAC *internal* signal to enable the MAC sampling of data starting at the first data segment (/S/) once every ten data segments in 100 Mbit/s modes.

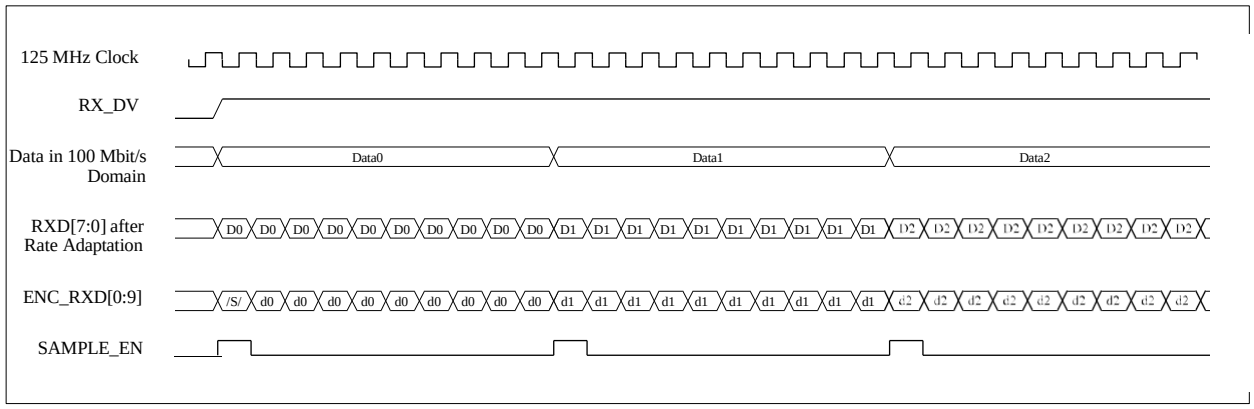


Figure 8: Data sampling in 100 Mbit/s modes

2.2.3 Scrambler/De-scrambler

The purpose of the scrambler is to "randomize" the serial stream that eventually will be transmitted over the serial link. All 8B/10B code-groups are scrambled.

The scrambler used is the same as defined in 1EEE 802.3 Clause 49 for 10GBASE-R.

The payload of the block is scrambled with a self-synchronizing scrambler. The scrambler shall produce the same result as the implementation shown in Figure 8. This implements the scrambler polynomial:

$$G(x) = 1 + x^{39} + x^{58}$$

There is no requirement on the initial value for the scrambler. The scrambler is run continuously on all payload bits.

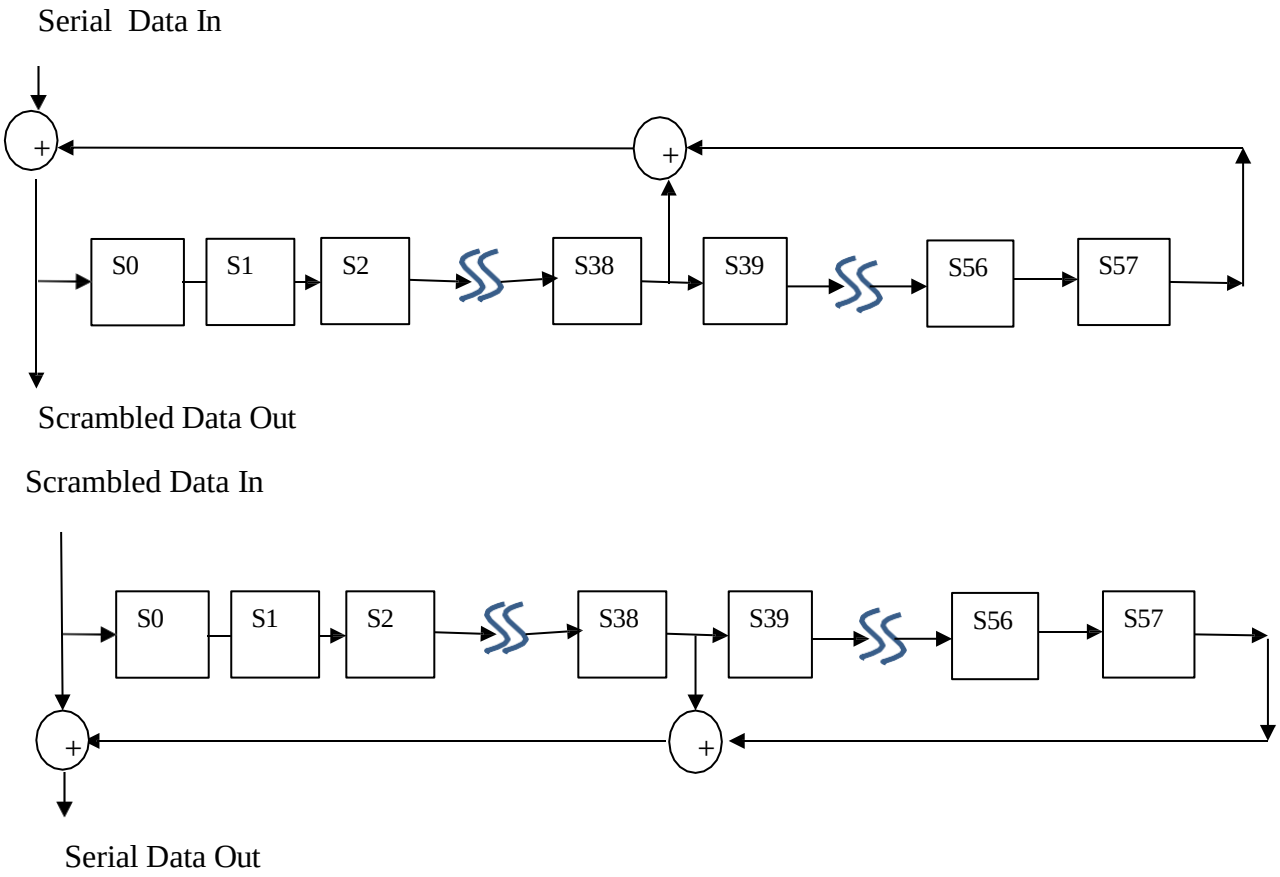


Figure 9: Scrambler/De-scrambler for 10G Mode

2.2.4 Packet Control Header PHY/ASIC

Figure 10, shows the mapping of Packet Control Header (UsgmiiPCH) in the Pre-amble.

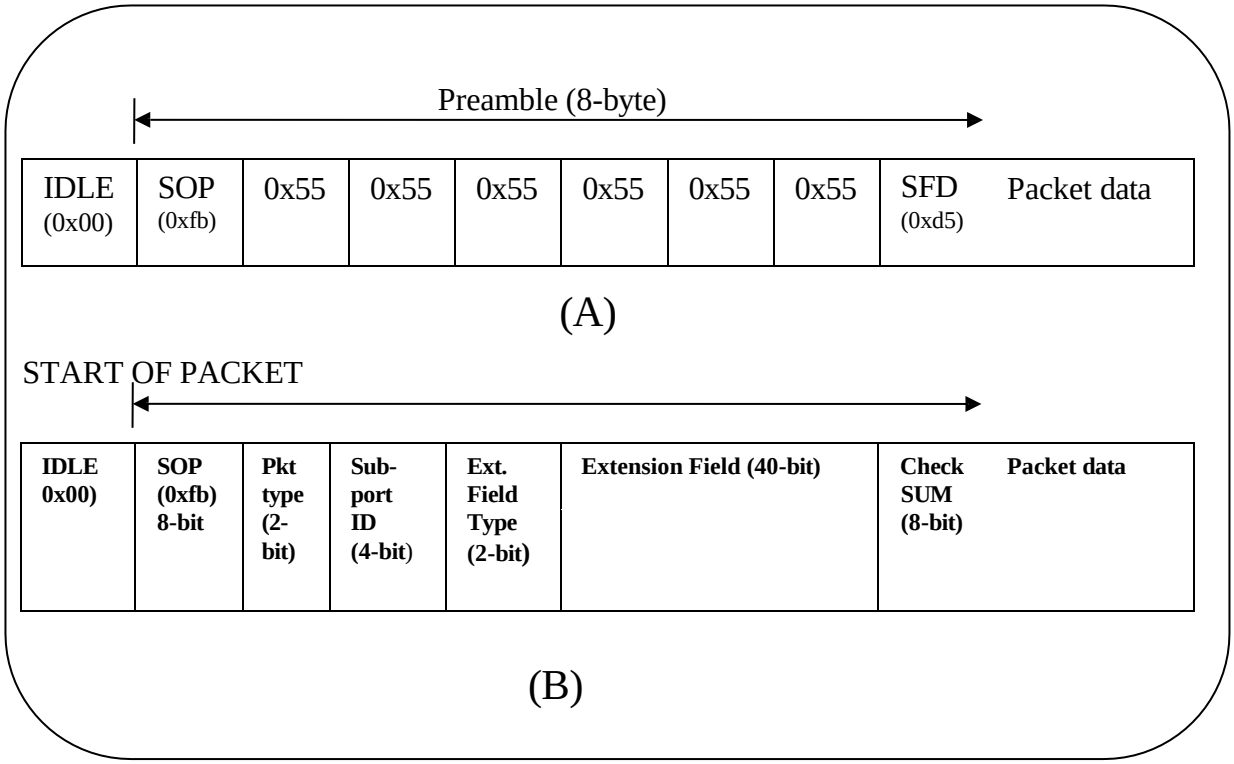


Figure 10: Packet Control Header Mapping to Pre-amble with SOP and Checksum

As specified above the preamble needs to be fixed size of 8 bytes (7bytes + Checksum) independent of even/odd number of bytes in the packet (refer to 35.2.3.2.2 IEEE 802.3). Port ASIC and PHY must provide a configurable option to allow dynamic adjustment of IPG (IDLES) to in this mode.

On the transmit side, when a frame with an odd number of bytes is transmitted, it is followed by an extra /R/ character, for e.g.:

… Payload, 4-byte CRC, T, R, R, Idle-pair, Idle-pair, …

This extra R-character is required to align to an even-boundary since idles are transmitted as pairs and aligned to even boundary. The successive frame has a byte of preamble reduced to offset this extra R-character.

2.2.4.1 Start of Packet (SOP)

The SOP symbol indicates the start of packet. The value of SOP MUST be 0xFB (Hex) as per IEEE 802.3 standard.

2.2.4.2 Packet Type

Packet Type identifies type of packet:

00: Ethernet Packet with status header
01: Ethernet packet, no Status Header (packet information)
10: Idle Packet – Contains status data for a port – no packet data
11: Unused (Reserved)

2.2.4.3 Sub-portID

Sub-portID is provided as additional information to be used at layer protocol layers. For multiple port PHY, Sub-port ID is same as network port number on PHY e.g. 0, 1, 2, 3.

For single port, Sub-port ID is 0.

2.2.4.4 Extension Field Type

This 2-bit field type defines the content of Extension Field. Refer to appropriate PHY for the definition.

2.2.4.5 Extension Field

This 40-bit Extension Filed carries data defined by Extension Field Type, e.g. Time-Stamp.

2.2.4.6 CRC computation

Header CRC is computed using CRC-8 algorithm with polynomial x8+x2+x+1 over the 6-bytes header data (inclusive of Packet Type to Extension Field). The initial value of CRC-8 computation is 0. The CRC remainder is exclusive-or'ed with 0x55 to header.crc value. The result is copied to the header.crc field on header generation and compared to header.crc on header reception. Header CRC generation is based on ITU–T I.432

Readers must pay attention to the differences in bit numbering conventions and bit CISCO Systems Confidential Information [Page 15]. CDL Working Group CDL Specification 17 May 2002 transmission order between ITU-T standard I.432 and this specification.

The I.432 standard numbers the LSB of a multi-bit field as zero and bits within a byte are transmitted MSB first. See the Section "Definitions" for bit numbering conventions and bit transmission order applicable to this specification.

The Header CRC is an 8-bit sequence calculated over 6 bytes PCH but excluding the SOP symbol and the Header CRC. The 48-bits long relevant portion of the CDL header is taken to represent a polynomial of order 47. The coefficients can only have the value 0 or 1. The least significant bit of the first byte of the header represents the coefficient of the highest order (x^47) term. The polynomial operations are performed modulo-2.

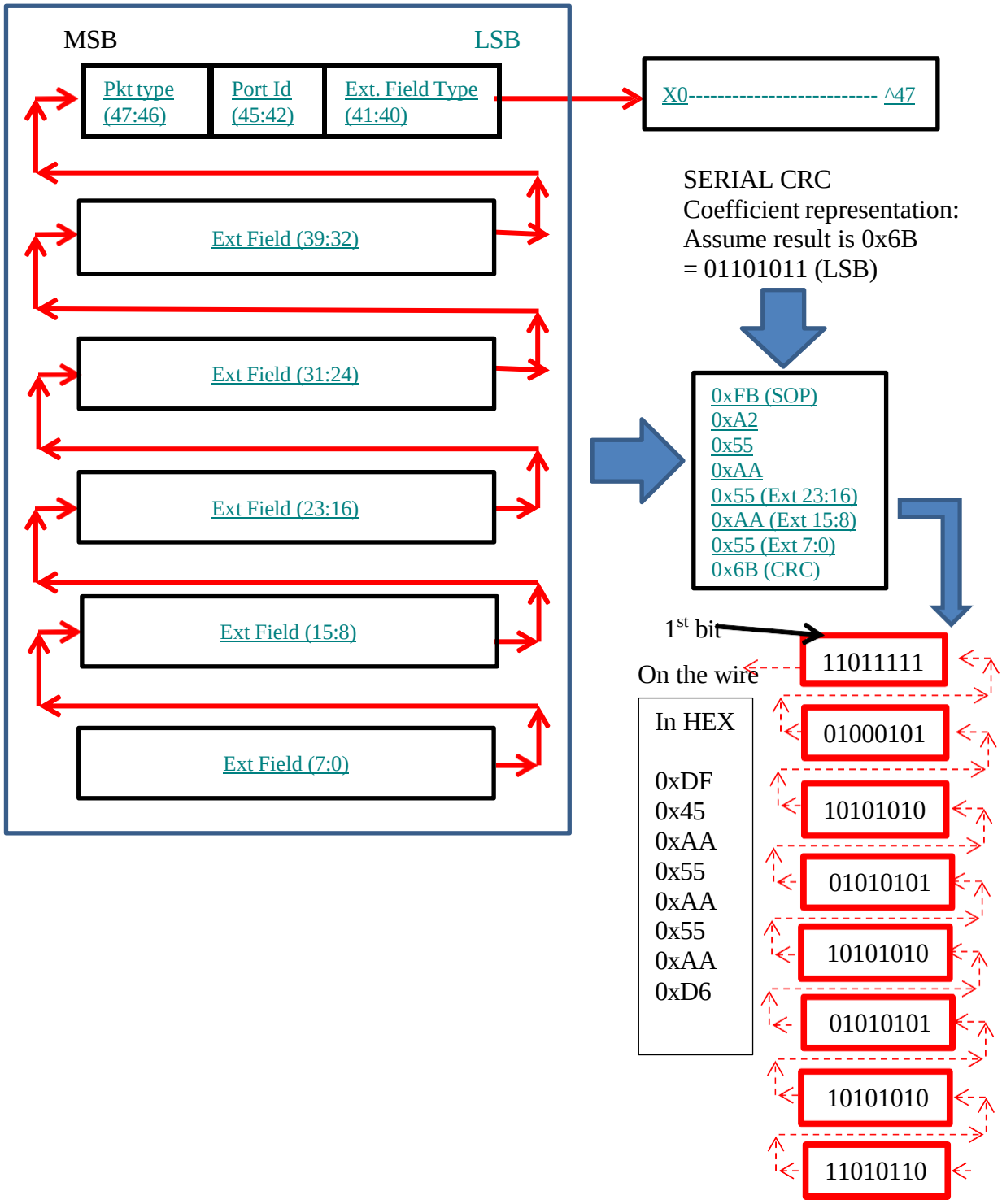


Figure 11: Bit order for Serial CRC Computation

Other examples for CRC:

PCH[47:0] = 0x2910_4602_7710. PCHCRC = 0x0B
PCH[47:0] = 0x2910_4602_7720. PCHCRC = 0x07
PCH[47:0] = 0x2910_4602_7730. PCHCRC = 0x0F
PCH[47:0] = 0x2910_4602_7740. PCHCRC = 0x01

The Header CRC MUST is recomputed whenever any of the fields in the header is changed and MUST be passed transparently whenever the fields of the header do not change.
The receiver in to-CDL-network direction must perform HCRC and maintain count of packets with failed HCRC for performance monitoring purposes. CDL information MUST NOT be extracted from, made use of or inserted into packets that failed HCRC. Such packets MAY also be discarded by the transmitter in from-CDL-network direction.

The linear feedback shift register below can be used to calculate the 8-bit CRC in a bit serial fashion. 8, 16 or 32 bit parallel implementations can be deduced from the bit serial implementation.

NOTE: The feedback on this implementation is different than 32-bit frame CRC and Clause 55

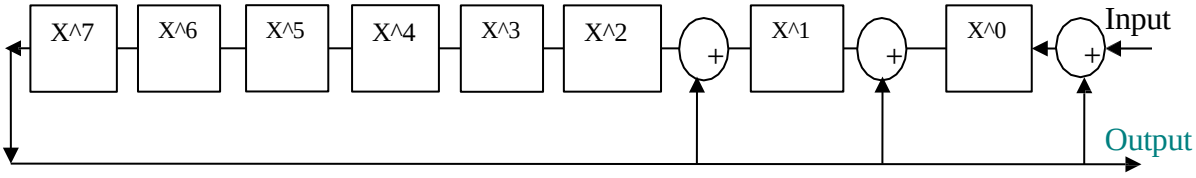


Figure 12: CRC Calculation

2.3 Electrical Specification

CML, (Current Mode Logic) is by far the most common Serdes IO standard in use today. The signal swing provided by the CML output is small, resulting in low power consumption. The driver and receiver are often self-terminated, eliminating external components and minimizing transmission line impedance discontinuity effects on timing and signal integrity.

The following section details the requirements for the high speed electrical interface that will operate at 1.25Gsym/s, 5Gsym/s or 10Gsym/s using NRZ coding (hence 1 bit per symbol at electrical level). Connections are point to point balanced differential pair with 100 Ohm nominal differential impedance and signaling is unidirectional. Clock and data are embedded hence CDR is required in the receiver. The link should operate with a BER of 10^{-15}.

2.3.1 SGMII – 1.25 Gbps Electrical Specification

The USGMII 1.25 Gbps electrical specification will be the same as the 1000BASE-KX electrical characteristics as defined in section 70.7 of the IEEE 802.3-2008 Standard.

It supports both AC and DC coupled operation. However, DC coupling of PHY to MAC is required since it optimizes system cost, complexity, and signal integrity.

2.3.2 Q-USGMII – 5.0 Gbps Electrical Specification

The following section details the requirements for the USGMII - 5 Gbps electrical interface

- 2.5/5Gsym/s using NRZ coding (hence 1 bit per symbol at electrical level)
- Connections are point to point balanced differential pair
- 100 Ohm nominal differential impedance
- Unidirectional signaling on each differential pair
- Clock and data are embedded hence CDR is required in the receiver
- The link should operate with a BER of 10^{-15}
- AC coupling required between driver and receiver

This section is based on the Optical Internetworking Forum’s (OIF) Common Electrical I/O CEI-6G-SR Short Reach Standard IA#OIF-CEI-02.0 with some modifications listed below. 2.5G is based on 5.0Gbps SI requirements.

However, DC coupling of PHY to MAC is required since it optimizes system cost, complexity, and signal integrity.

Parameter	Symbol	Min	Typ	Max	Units	Notes
Baud Rate	T_Baud		5.00		Gsym/s	*1*
Output Differential Voltage (into floating Load Rload=100Ohm)	T_Vdiff	400		900	mVppd	*2*
Differential Resistance	T_Rd	80	100	120	Ohms	
Recommended Output Rise and Fall Times (20% to 80%)	T_tr,T_tf	30			ps	
Differential Output Return Loss (100Mhz to *2.5 GHz*)	T_SDD22			-8	dB	*3*
Differential Output Return Loss (*1.25 GHz* to 5 GHz)	T_SDD22				dB	*4*
Common Mode Return Loss (100Mhz to *2.5 GHz*)	T_SCC22			-6	dB	5
Transmitter Common Mode Noise	T_Ncm			5% of T-Vdiff	mVppd	
Output current into or out of driver pins when either SHORT to GND or each other				100	mA	6
Output Common Mode Voltage See Note [7]	T_Vcm	0.0		1.8	V	

Table 6: Transmitter Output Electrical Specification

[1]CEI-6G-SR is defined to operate between baud rates of 4.976 and 6.375 Gsym/s. Nominal 5 Gsym/s with a tolerance of +/- 100 ppm.

[2]Absolute driver output voltage shall be between -0.1V and 1.9V with respect to local ground. See Figure 9 for details.

[3]See Figure 8

[4]See Figure 8

[5]See Figure 8

[6]+/- 100 mA

[7]R_Rdin=100 Ohms+/- 20 Ohms.For Vcm definition, see Figure 9.

Parameter	Symbol	Min	Typ	Max	Units	Notes
Uncorrelated High Probability Jitter	T_UHPJ			0.15	UIpp	[8]
Duty Cycle Distortion	T_DCD			0.05	UIpp	
Total Jitter	T_TJ			0.30	UIpp	[9]
Eye Mask	T_X1			0.15	UI	[10]
Eye Mask	T_X2			0.40	UI	[11]
Eye Mask	T_Y1	200			mV	[12]
Eye Mask	T_Y2			*450*	mV	[13]

Table 7: Transmitter Output Jitter and Eye Specifications

[8]This parameter is defined as: Jitter distribution where the value of the jitter show no correlation to any signal level being transmitted. Formally defined as deterministic jitter, T_DJ

[9]The link will operate with a BER of 10^{-15}

[10]See Figure 7

[11]See Figure 7

[12]See Figure 7

[13]See Figure 7

Parameter	Symbol	Min	Typ	Max	Units	Notes
RX Baud rate	R_Baud		5.00		GSym/s	[14]
Input Differential Voltage	R_Vdiff	100		900	mVppd	[15]
Differential Resistance	R_Rdin	80	100	120	Ohms	
Differential Input Return Loss (100MHz to *2.5 GHz*)	R_SDD1			-8	dB	[16]
Differential Input Return Loss (*2.5 GHz* to 5 GHz)	R_SDD1				dB	[17]
Common Mode Input Return Loss (100MHz to *2.5 GHz*)	R_SCC1			-6	dB	[18]
Input Common Mode Voltage See Note [19]	R_Vrcm	-0.05		1.85	V	
Wander divider	N			10		[20]

Table 8: Receiver Electrical Input Specifications

[14]CEI-6G-SR is defined to operate between baud rates of 4.976 and 6.375 Gsym/s, Nominal 5 Gsym/s with a tolerance of +/- 100 ppm.

[15]Min Value is changed from the standard and reduced to 100 mV.

[16]See Figure 8

[17]See Figure 8

[18]See Figure 8

[19]For Vcm definition, see Figure 9.

[20]See Figure 2-27 & Figure 2-28 in CEI-6G-SR document for details

Parameter	Symbol	Min	Typ	Max	Units	Notes
Bounded High Probability Jitter	R_BHPJ			0.45	UIpp	[21]
Sinusoidal Jitter,maximum	R_SJ-max			5	UIpp	
Sinusodial Jitter, High Frequency	R_SJ-hf			0.05	UIpp	
Total Jitter(Does not include Sinusodial Jitter)	R_TJ			0.60	UIpp	[22]
Eye Mask	R_X1			0.30	UI	[23]
Eye Mask	R_Y1			*50*	mV	[24]
Eye Mask	R_Y2			*450*	mV	[25]

Table 9: Receiver Input Jitter and Eye Specifications

[21]This is the sum of Uncorrelated Bounded High Probability Jitter (0.15 UI) and Correlated Bounded High Probability Jitter (0.30 UI).

Uncorrelated Bounded High Probability Jitter: Jitter distribution where the value of the jitter show no correlation to any signal level being transmitted. Formally defined as deterministic jitter, T_DJ

Correlated Bounded High Probability Jitter: Jitter distribution where the value of the jitter shows a strong correlation to the signal level being transmitted. This jitter may considered as being equalizable due to its correlation to the signal level

[22]The link will operate with a BER of 10^{-15}

[23]See Figure 7

[24]See Figure 7.

[25]See Figure 7.

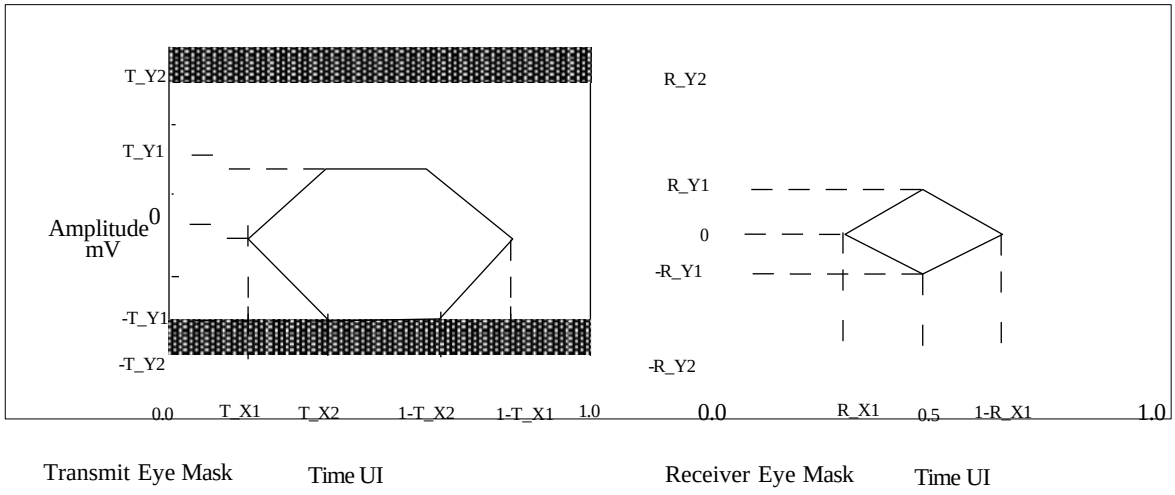
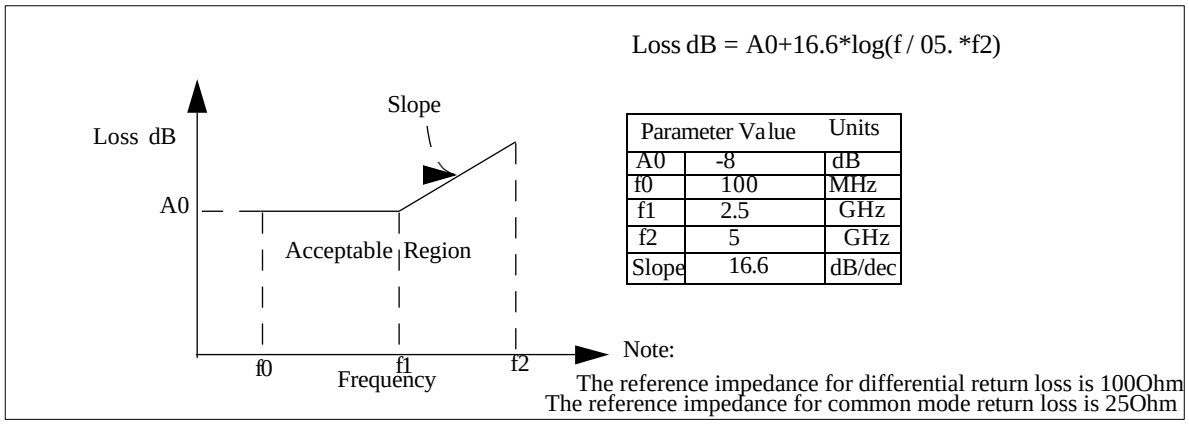


Figure 7 Driver and Receiver Eye Mask

Figure 13: Driver and Receiver Eye Mask



Parameter	Value	Units
A0	-8	dB
f0	100	MHz
f1	2.5	GHz
f2	5	GHz
Slope	16.6	dB/dec

Figure 8 Driver and Receiver Differential Return Loss

Figure 14: Driver and Receiver Differential Return Loss

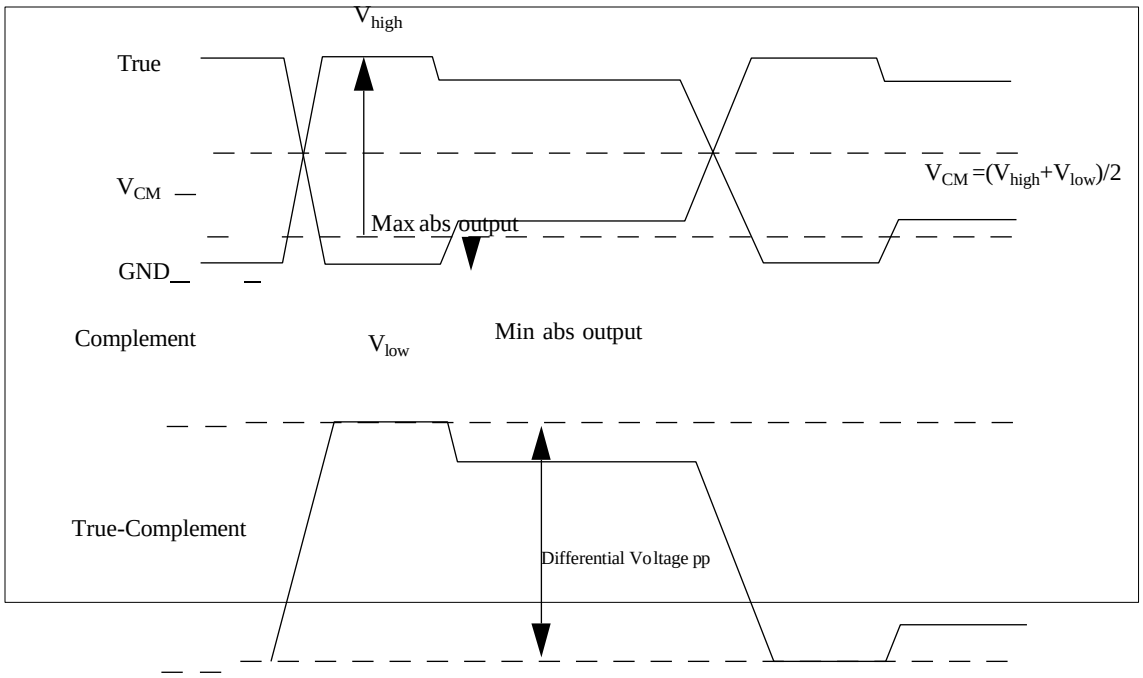


Figure 9 Definition of Driver Amplitude and Swing

Figure 15: Definition of Driver Amplitude and Swing

USGMII – 5 Gbps channel has a total loss budget of 12 dB as shown in **Error! Reference source not found.**

Parameter	**Condition**	**Loss (dB)**
Transmitter package loss margin	at 2.5 Ghz	0.5
Receiver Package loss margin	at 2.5 Ghz	0.5
Channel Interconnect loss margin	at 2.5 Ghz	10.0
Channel Crosstalk, Reflection and other loss margin	at 2.5 Ghz	1.0
Total Channel loss	at 2.5 Ghz	**12.0**

Table 10: Q-USGMII – 5 Gbps Channel Loss Budget

Figure 11, shows the interconnect loss template for the channel. At 2.5 GHz, maximum allowed interconnect loss is -10.0 dB which represents a typical 20 inch trace on FR4 PCB.

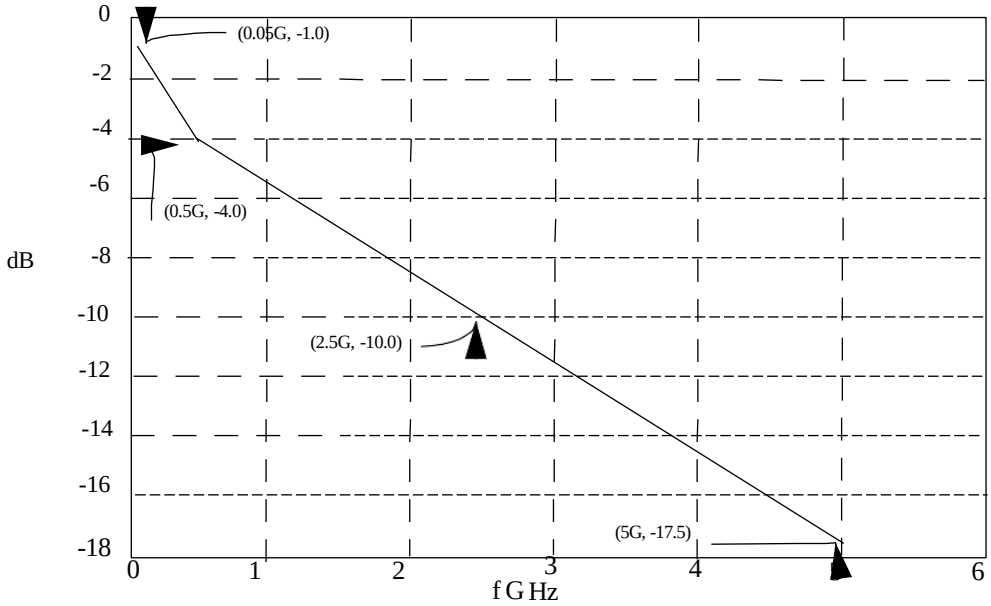


Figure 10 Interconnect Loss Template for QSGMII Channel

Figure 16: Interconnect Loss Template for Q-USGMII Channel

2.3.3 O-USGMII – 10.0 Gbps Electrical Specification

The following section details the requirements for the USGMII – 5/10 Gbps electrical interface

- IP vendor must provide AMI models and support internal eye (ability to measure the eye at the RX sampling latch)
- IP vendor must provide Vertical and Horizontal Eye opening at a BER (used as a PASS/FAIL criteria for AMI simulation and internal eye diagnostic)
- 10Gsym/s using NRZ coding (hence 1 bit per symbol at electrical level)
- Connections are point to point balanced differential pair
- 100 Ohm nominal differential impedance
- Unidirectional signaling on each differential pair
- Clock and data are embedded hence CDR is required in the receiver
- The link should operate with a BER of 10^{-15}
- AC coupling required between driver and receiver
- All new designs should follow section 2.3.3

The USGMII 10 Gbps electrical specification will be the same as the 10GBASE-KR electrical characteristics as defined in section 72.7 and Annex 69B of the IEEE 802.3-2008

Few exceptions will apply: Exception is added to allow vendors implement low cost interface.

1. Maximum insertion loss is 12dB @ 2.5GHz (5.0 Gbps) and 15dB @ 5Ghz (10 Gbps) (vs ~25dB based on the KR spec)

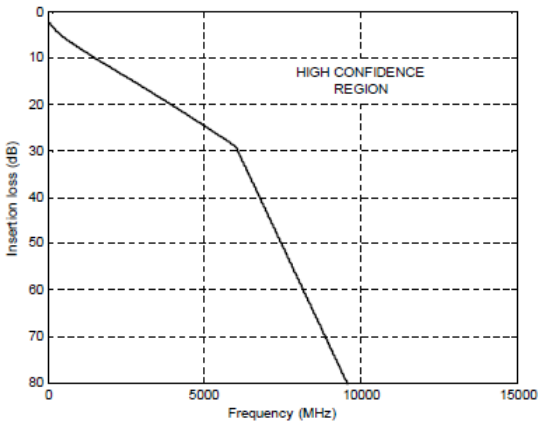


Figure 69B–5—Insertion loss limit for 10GBASE-KR

2. Maximum ICR at 5GHz is 15dB (vs ~25dB)

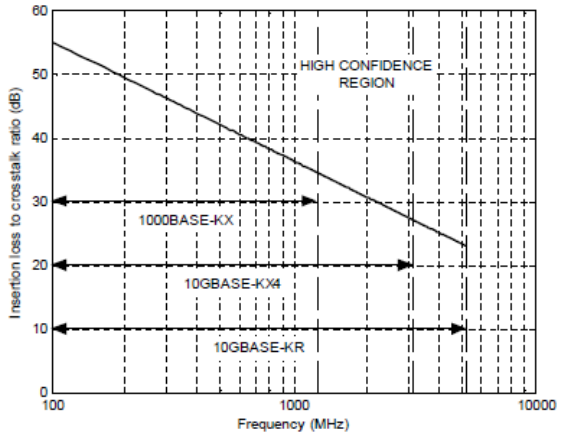


Figure 69B–8—Insertion loss to crosstalk ratio limit

$$ICR(f) = -IL(f) + PSXT(f)$$