System Packet Interface Level 4 (SPI-4) Phase 2 Revision 1: OC-192 System Interface for Physical and Link Layer Devices.

OIF-SPI-4-02.1

October 15, 2003

Implementation Agreement Created and Approved
by the Optical Internetworking Forum
www.oiforum.com

Working Group: **Physical – & Link Layer (PLL)**

TITLE: **System Packet Interface Level 4 (SPI-4) Phase 2 Revision 1: OC-192 System Interface for Physical and Link Layer Devices**
Physical and Link Layer Devices.

SOURCE:

Richard Cam
Technical Editor
PMC-Sierra
105-8555 Baxter Place
Burnaby, B.C.
Canada V5A 4V7
Phone: +1 604 415 6022
Email: richard_cam@pmc-sierra.com

Russ TuckMike Lerer
Working Group Chair
Pluris Terabit Network SystemsXilinx
10455 Bandley DrivePO Box 636
Cupertino, CA 95014Londonderry, NH 03053
USA
Phone: +1 603-548-3704408 861 3360
Email: mlerer@fpga.comtuck@pluris.com

Karl Gass
Working Group Vice Chair
Sandia National Laboratories
PO Box 5800, MS 0874
Albuquerque, NM 87185
USA
Phone: +1 603-548-3704
Email: kgass@sandia.gov

Wayne Nation
Revision Editor
LSI Logic Corporation
3425 40th Avenue NW
Rochester, MN 55901
USA
Phone: +1 507-535-3865
Email: wnation@lsil.com

DATE: **October 15, 2003January 2001**

Document Status: Draft of Revised Implementation Agreement OIF-SPI4-02.00
Project Name: SPI-4 Phase 2
Project Number:

Notice:

For additional information contact:
The Optical Internetworking Forum, 39355 California Street,
Suite 307, Fremont, CA 94538
510-608-5928 phone Φ info@oiforum.com

1.1

1 Table of Contents

2 List of Figures

3 List of Tables

4 Document Revision History

Oif2000.088 23 April 2000
Baseline document produced from adoption of contribution oif2000.026 (“The POS-PHY L4 Proposal: A Packet / Cell System Interface for 10 Gb/s Applications”) at the January/February 2000 OIF meeting.

oif2000.088.1 03 May 2000, 19 June 2000
Revised document reflecting acceptance of changes in contribution oif2000.099 (“Suggested Changes to the SPI-4 Phase 2 Baseline Document”), and recommended changes during the May 2000 OIF meeting. Updated working group chair (formerly Sid Chaudhuri, AT&T / Tellium).

oif2000.088.2 20 June 2000
Removed change markings.

oif2000.088.3 06 October 2000
Revised document reflecting accepted changes at August 2000 meeting.

oif2000.088.4 08 December 2000, 02 January 2001
Revised document reflecting accepted technical changes at November 2000 meeting as well as various editorial clarifications/corrections.

OIF2003.297.00 30 July 2003
Incorporated contents of Implementation Guide (OIF2002.230.07) to form draft of Revision 1 of the SPI4 Phase 2 Implementation Agreement

OIF2003.297.01 15 October 2003
Incorporated editorial changes from Straw Ballot

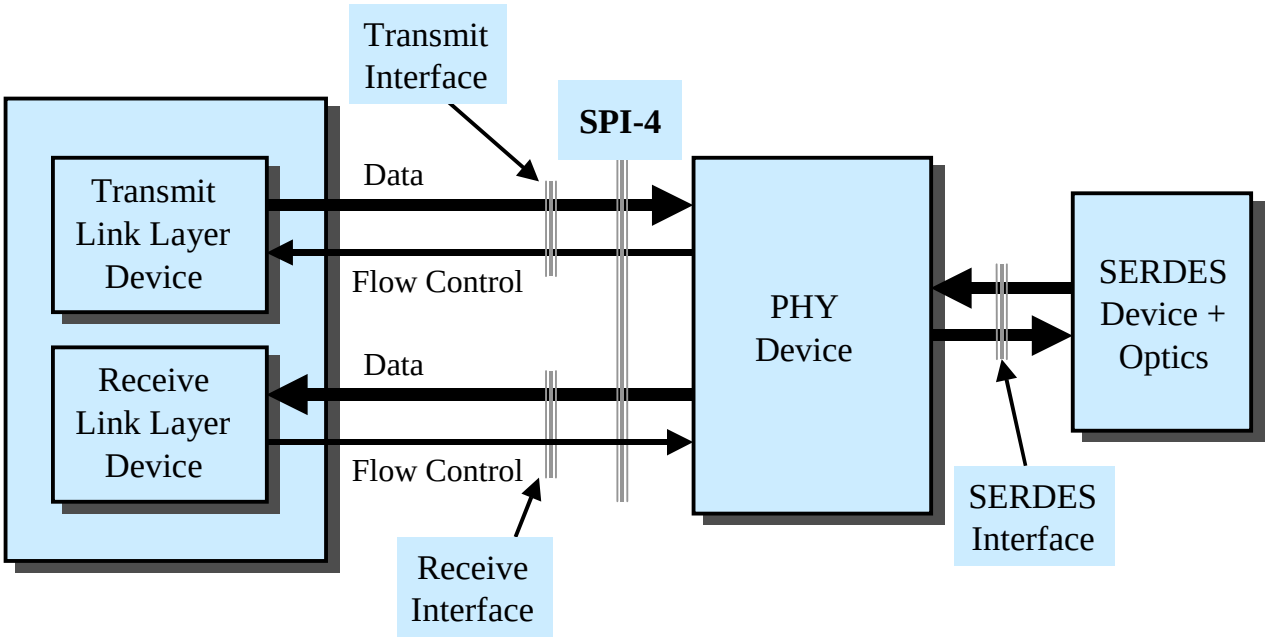
5 Introduction

This document specifies the Optical Internetworking Forum's recommended interface for the interconnection of Physical Layer (PHY) devices to Link Layer devices for 10 Gb/s aggregate bandwidth applications by means of a higher-speed interface than defined in SPI-4 Phase 1. This Phase 2 specification will be referred to hereon for convenience in this document as the SPI-4 interface.

SPI-4 is an interface for packet and cell transfer between a physical layer (PHY) device and a link layer device, for aggregate bandwidths of OC-192 ATM and Packet over SONET/SDH (POS), as well as 10 Gb/s Ethernet applications. This section provides a general overview of the interface. The next section contains more detailed descriptions of the signals and associated operations, data structures, start-up and AC timing parameters. Appendices A through E have been included to provide guidance on implementations; they are for information only and are not normative parts of this specification.

The following is a general synopsis of the SPI-4 interface. For reference, a general block diagram is shown in Fig. 5.1. SPI-4 is the system packet interface for data transfer between the link layer and the PHY device; it is designed to meet requirements of this particular application, although it may be used in other applications as well. “Transmit” and “Receive” refer, respectively, to data flow and associated control/status information for the Link Layer to PHY, and the PHY to Link Layer directions.

Figure 5.1: System Reference Model.



On both the transmit and receive interfaces, FIFO status information is sent separately from the corresponding data path. By taking FIFO status information out-of-band, it is possible to decouple the transmit and receive interfaces so that each operates independently of the other. Such an arrangement makes POS-PHY L4 suitable not only for bidirectional but also for unidirectional link layer devices.

In both the transmit and receive interfaces, the packet's address, delineation information and error control coding is sent in-band with the data.

SPI-4 has the following general characteristics:

- Point-to-point connection (i.e., between single PHY and single Link Layer device).

- Support for 256 ports (suitable for STS-1 granularity in SONET/SDH applications (192 ports) and Fast Ethernet granularity in Ethernet applications (100 ports)).

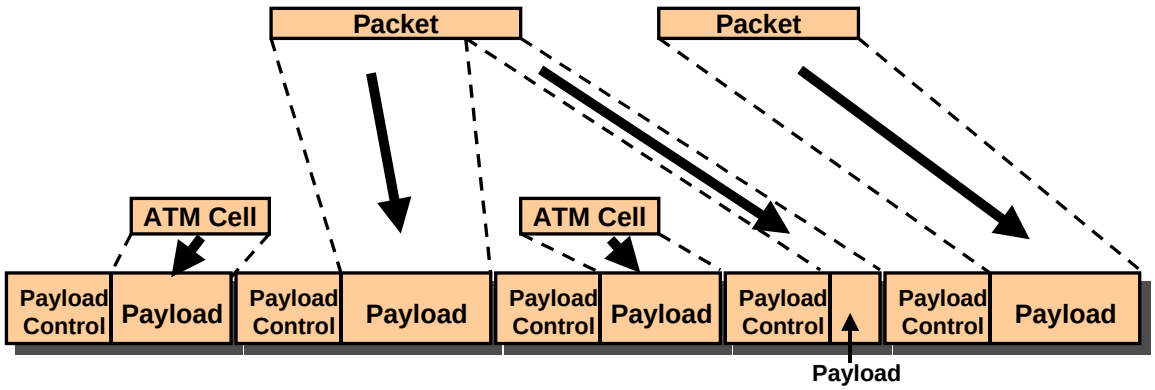
- Transmit / Receive Data Path:
    - 16 bits wide.
    - In-band port address, start/end-of-packet indication, error-control code.
    - LVDS I/O (IEEE 1596.3 – 1996 [1], ANSI/TIA/EIA-644-A-20011995 [12]).

  - 622 Mb/s minimum data rate per line.
  - Source-synchronous double-edge clocking, 311 MHz minimum.

- Transmit / Receive FIFO Status Interface:
  - LVTTL I/O or optional LVDS I/O (IEEE 1596.3 – 1996 [1], ANSI/TIA/EIA-644-A-20011995 [12]).
  - Maximum 1/4 data path clock rate for LVTTL I/O, data path clock rate (double-edge clocking) for LVDS I/O.
  - 2-bit parallel FIFO status indication.
  - In-band Start-of-FIFO Status signal.
  - Source-synchronous clocking.

Data is transferred in bursts that have a provisionable maximum length and either a fixed or a provisionable minimum length. Both the minimum and the maximum burst transfer lengths must be multiples of 16 bytes. The actual burst transfer length must be a multiple of 16 bytes, with the exception of transfers that terminate with an EOPData is transferred in bursts that have a provisionable maximum length, with the exception of transfers that terminate with an EOP. Information associated with each transfer (port address, start/end-of-packet indication and error-control coding) is sent in 16-bit control words described later in this document. Fig. 5.2 shows how ATM cells and variable-length packets map onto the data stream.

Figure 5.2: Mapping of packets and ATM cells onto payload stream.



6 Interface Description

Section 6.1 contains signal definitions for the transmit and receive directions. Section 6.2 describes the signal operation along with the data structures for payload data and in-band control/status information. Section 6.3 describes start-up parameters. Section 6.4 specifies AC timing parameters.

6.1 Signals

A block diagram depicting the interface signals is shown in Fig. 6.1. The transmit and receive data paths include, respectively, (TDCLK, TDAT[15:0], TCTL) and (RDCLK, RDAT[15:0], RCTL). The transmit and receive FIFO status channels include (TSCLK, TSTAT[1:0]) and (RSCLK, RSTAT[1:0]) respectively. Table 6.1 provides a summary of the interface signals.

Fig. 6.1: SPI-4 Interface.

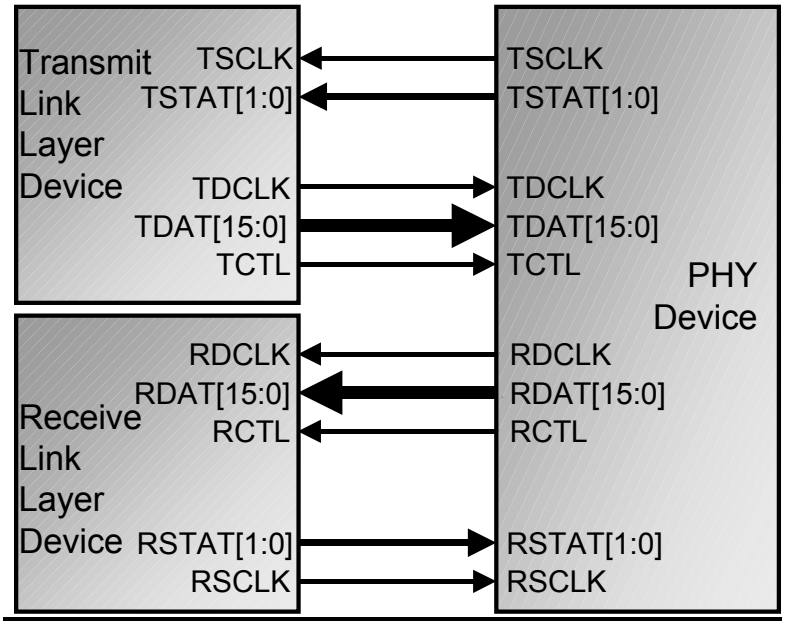


Table 6.1: SPI-4 Interface Signal Summary.

Signal	Direction	Description
TDCLK	Link to PHY	Transmit Data Clock. Clock associated with TDAT and TCTL. Data and control lines are driven off the rising and falling edges of the clock.
TDAT [15:0]	Link to PHY	Transmit Data. Used to carry payload data and in-band control words from the Link Layer to the PHY device. The control word format is described in Section 6.2.
TCTL	Link to PHY	Transmit Control. TCTL is high when a control word is present on TDAT[15:0]. It is low otherwise.
TSCLK	PHY to Link	Transmit Status Clock. Clock associated with TSTAT.
TSTAT [1:0]	PHY to Link	Transmit FIFO Status. Used to carry round-robin FIFO status information, along with associated error detection and framing.
RDCLK	PHY to Link	Receive Data Clock. Clock associated with RDAT and RCTL. Data and control lines are driven off the rising and falling edges of the clock.
RDAT [15:0]	PHY to Link	Receive Data. Carries payload data and in-band control from the PHY to the Link Layer device. The control word format is described in Section 6.2.

Signal	Direction	Description
RCTL	PHY to Link	Receive Control. RCTL is high when a control word is present on RDAT[15:0]. It is low otherwise.
RSCLK	Link to PHY	Receive Status Clock. Clock associated with RSTAT.
RSTAT [1:0]	Link to PHY	Receive FIFO Status. Used to carry round-robin FIFO status information, along with associated error detection and framing.

6.2 Interface Operation and Data Structures

6.2.1 6.2.1 Data Path

Complete packets or shorter bursts may be transferred, as shown in Fig. 5.2. The maximum configured payload data transfer size must be a multiple of 16 bytes. Control words are inserted only between burst transfers; once a transfer has begun, data words are sent uninterrupted until end-of-packet or a multiple of 16 bytes is reached. The interval between the end of a given transfer and the next payload control word (marking the start of another transfer) consists of zero or more idle control words and training patterns (Section 6.2.3).

The minimum and maximum supported packet lengths are determined by the application. For ease of implementation however, successive start-of-packets must occur not less than 8 cycles apart, where a cycle is one control or data word. The gap between shorter packets is filled with idle control words.

Fig. 6.2 shows cycle-by-cycle behavior of the data path for valid state transitions. The states correspond to the type of words transferred on the data path. Transitions from the “Data Burst” state (to “Payload Control” (if data is available for transfer) or “Idle Control” (otherwise)) are possible only on integer multiples of 8 cycles (corresponding to multiple of 16 byte segmentation) or upon end-of-packet. A data burst must follow a payload control word immediately on the next cycle unless control word extensions are supported by the implementation. Only control word extensions are allowed between a payload control word and a data burst. The use of control word extensions is not an integral part of this agreementA data burst must follow a payload control word immediately on the next cycle. Arcs not annotated correspond to single cycles.

Fig. 6.2. Data Path State Diagram.

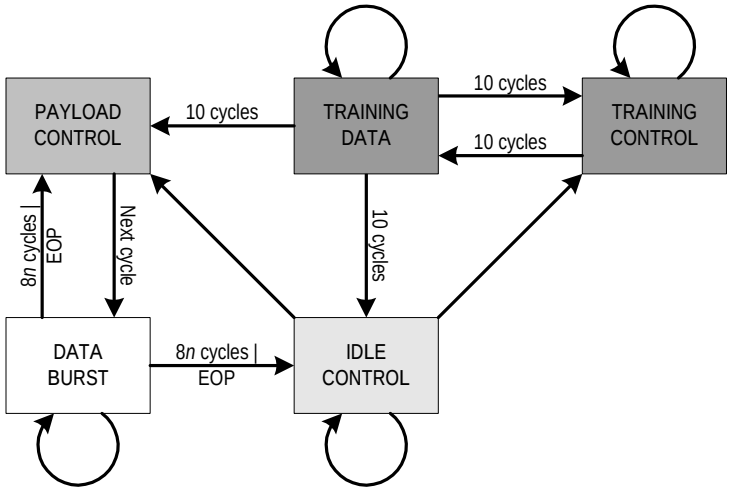
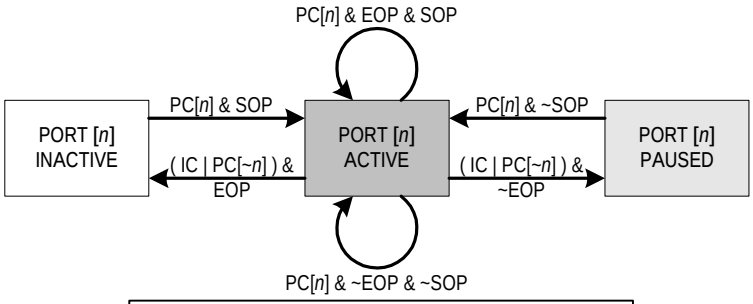


Fig. 6.3 shows per-port state transitions at control word boundaries. At any given time, a given port may be active (sending data), paused (not sending data, but pending the completion of an outstanding packet), or inactive (not sending data, no outstanding packet).

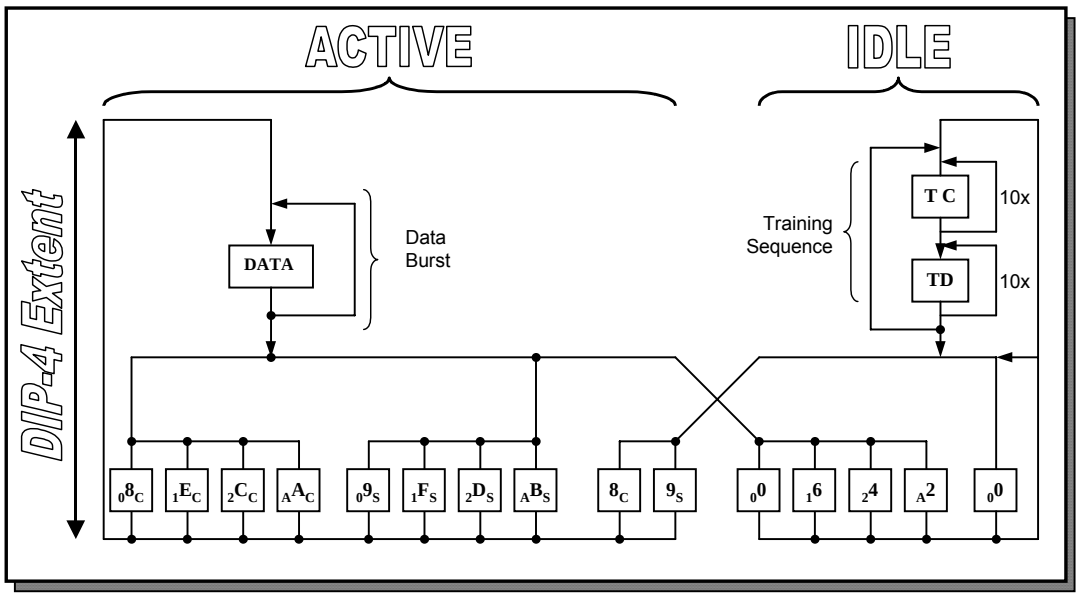
Fig. 6.3. Per-Port State Diagram with Transitions at Control Words.



KEY:	
IC:	Idle Control Word
PC[*n*]:	Payload Control Word for port *n*
PC[~*n*]:	Payload Control Word for a port other than port *n*
SOP:	Start-of-Packet in Payload Control Word
~SOP:	No Start-of-Packet in Payload Control Word
EOP:	End-of-Packet in Control Word
~EOP:	No End-of-Packet in Control Word
&	*AND*
|	*OR*

Fig. 6.4 shows valid sequences of data/control words for both “data burst” transfers and “idle” or “training” intervals. Valid control word encodings on bit positions [15:12] are shown on the bottom row. The left subscript indicates EOPS status (of the preceding transfer), while the right subscript indicates SOP status (of the next transfer).

Fig. 6.4. Sequences of Data / Control Words.



Payload data bytes are transferred over the interface in the same order as they would be transmitted or received on the line side. The order of bytes within a word is shown in Figs. 6.5 and 6.6 (for transfer of arbitrary packets that end on odd and even byte boundaries, respectively). The most significant bits (MSBs) of the constituent bytes correspond to bits 15 and 7. The order in which the bits within each byte are transmitted and received on the line is defined by the Physical Layer and is beyond the scope of this specification. On payload transfers that do not end on an even byte boundary, the unused byte (after the last valid byte) on bit positions 7 through 0 is set to all zeroes, as shown in Fig. 6.5.

Figure 6.5: Example of Payload Transfer Data Structure (43-byte packet, * = set to all zeroes).

	Bit 15 Bit 8	Bit 7 Bit 0
Data Word 1	Byte 1	Byte 2
Data Word 2	Byte 3	Byte 4
Data Word 3	Byte 5	Byte 6
Data Word 4	Byte 7	Byte 8
Data Word 21	Byte 41	Byte 42
Data Word 22	Byte 43	XX*

Figure 6.6: Example of Payload Transfer Data Structure (52-byte packet).

	Bit 15 Bit 8	Bit 7 Bit 0
Data Word 1	Byte1	Byte2
Data Word 2	Byte3	Byte4
Data Word 3	Byte5	Byte6
Data Word 4	Byte7	Byte8
Data Word 25	Byte49	Byte50
Data Word 26	Byte51	Byte52

A common control word format is used in both the transmit and receive interfaces. Table 6.2 describes the fields in the control word. When inserted in the data path, the control word is aligned such that its MSB is sent on the MSB of the transmit or receive data lines. A payload control word that separates two adjacent burst transfers contains status information pertaining to the previous transfer and the following transfer. Table 6.3 shows a list of control word types. Table 6.4 shows some examples of valid control words. Usage of reserved bits is beyond the scope of this specification. The transmitter shall not send Reserved control words. Receivers may ignore Reserved control words and may optionally report an error condition when Reserved control words are observed.

Table 6.2. Description of Fields in the Control Words.

Bit Position	Label	Description
15	Type	Control Word Type. Set to either of the following values: 1: payload control word (payload transfer will immediately follow the control word). 0: idle or training control word (otherwise).
14:13	EOPS	End-of-Packet (EOP) Status. Set to the following values below according to the status of the immediately preceding payload transfer. 0 0: Not an EOP. 0 1: EOP Abort (application-specific error condition). 1 0: EOP Normal termination, 2 bytes valid. 1 1: EOP Normal termination, 1 byte valid. EOPS is valid in the first control word following a burst transfer. It is ignored and set to “0 0” otherwise.
12	SOP	Start-of-Packet. Set to 1 if the payload transfer immediately following the control word corresponds to the start of a packet. Set to 0 otherwise. Set to 0 in all idle and training control words.
11:4	ADR	Port Address. 8-bit port address of the payload data transfer immediately following the control word. None of the addresses are reserved (all are available for payload transfer). Set to all zeroes in all idle control words. Set to all ones in all training control words.
3:0	DIP-4	4-bit Diagonal Interleaved Parity. 4-bit odd parity computed over the current control word and the immediately preceding data words (if any) following the last control word.

Table 6.3 Control Word Type List

	Bit [15:12]	Next Word Status	Prior Word Status	Meaning
0	0000	Idle	Continued	Idle, not EOP, training control word
1	0001	Reserved	Reserved	Reserved*
2	0010	Idle	EOP w/abort	Idle, Abort last packet
3	0011	Reserved	Reserved	Reserved*
4	0100	Idle	EOP w/ 2 bytes	Idle, EOP with 2 bytes valid
5	0101	Reserved	Reserved	Reserved
6	0110	Idle	EOP w/ 1 byte	Idle, EOP with 1 byte valid
7	0111	Reserved	Reserved	Reserved
8	1000	Valid	None	Valid, no SOP, no EOP
9	1001	Valid/SOP	None	Valid, SOP, no EOP
A	1010	Valid	EOP w/abort	Valid, no SOP, abort
B	1011	Valid/SOP	EOP w/abort	Valid, SOP, abort
C	1100	Valid	EOP w/ 2 bytes	Valid, no SOP, EOP with 2 bytes valid
D	1101	Valid	EOP w/ 2 bytes	Valid, SOP, EOP with 2 bytes valid
E	1110	Valid	EOP w/1 byte	Valid, no SOP, EOP with 1 byte valid
F	1111	Valid	EOP w/1 byte	Valid, SOP, EOP with 1 byte valid

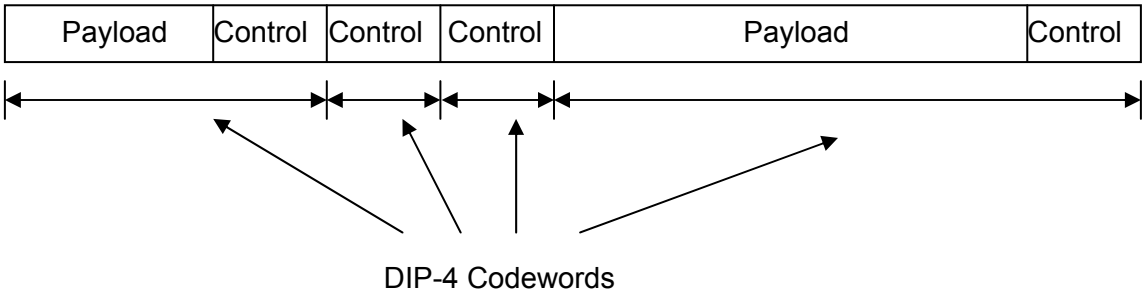
Table 6.4. Some Examples of Control Words.

Control Word	Description
0110 00000000 abcd	Idle control word following end of transfer. End-of-packet, normal termination, 1 byte valid in last data word. (Note: abcd bits depend on contents of this control word and preceding transfer.)
0000 00000000 1111	Idle control word preceded by another (idle) control word.
1101 00000101 abcd	Payload control word following end of transfer. End-of-packet, normal termination, 2 bytes valid in last data word of preceding transfer (abcd bits depend on contents of this control word and preceding transfer). Start-of-packet in next transfer to port 5.

* See Section 7.5 (Appendix E).

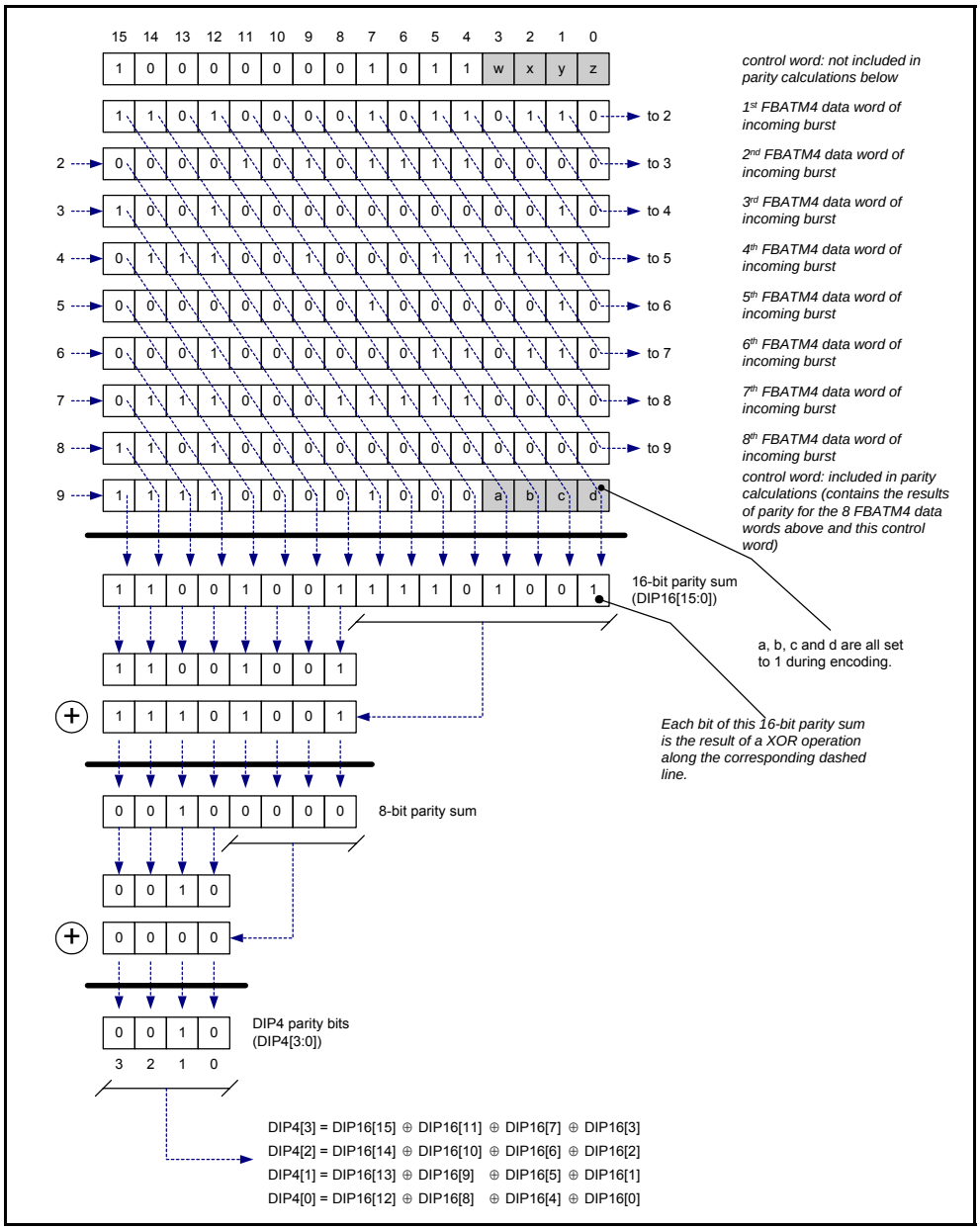
Fig. 6.7 shows the range over which the DIP-4 parity bits are computed. The DIP-4 code is almost as easy to implement as a conventional BIP code. In the presence of random errors, it offers the same error protection capability as a comparable BIP code, but has an additional advantage of spreading single-column errors (as might occur in a single defective line) across the parity bits. Appendix A discusses the error detection performance of this code.

Figure 6.7: Extent of DIP-4 codewords.



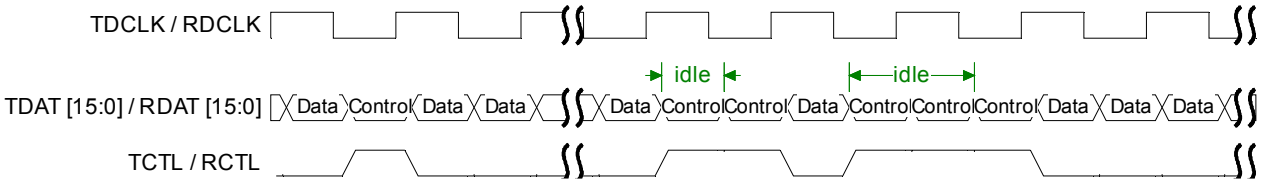
A functional description of calculating the DIP-4 code is given as follows. Assume that the stream of 16-bit data words are arranged as shown in Figure 6.8, MSB at the leftmost column, time moving downward. (The first word received is at the top of the figure; the last word is at the bottom of the figure.) The parity bits are generated by summing diagonally (in the control word, the space occupied by the DIP-4 code (bits a,b,c,d) is set to all 1's during encoding). The first 16-bit checksum is split into two bytes, which are added to each other modulo-2 to produce an 8-bit checksum. The 8-bit checksum is then divided into two 4-bit nibbles, which are added to each other modulo-2 to produce the final DIP-4 code. The procedure described applies to either parity generation on the egress path or to check parity on the ingress path.

Figure 6.8 Example of DIP-4 Encoding (Odd Parity).



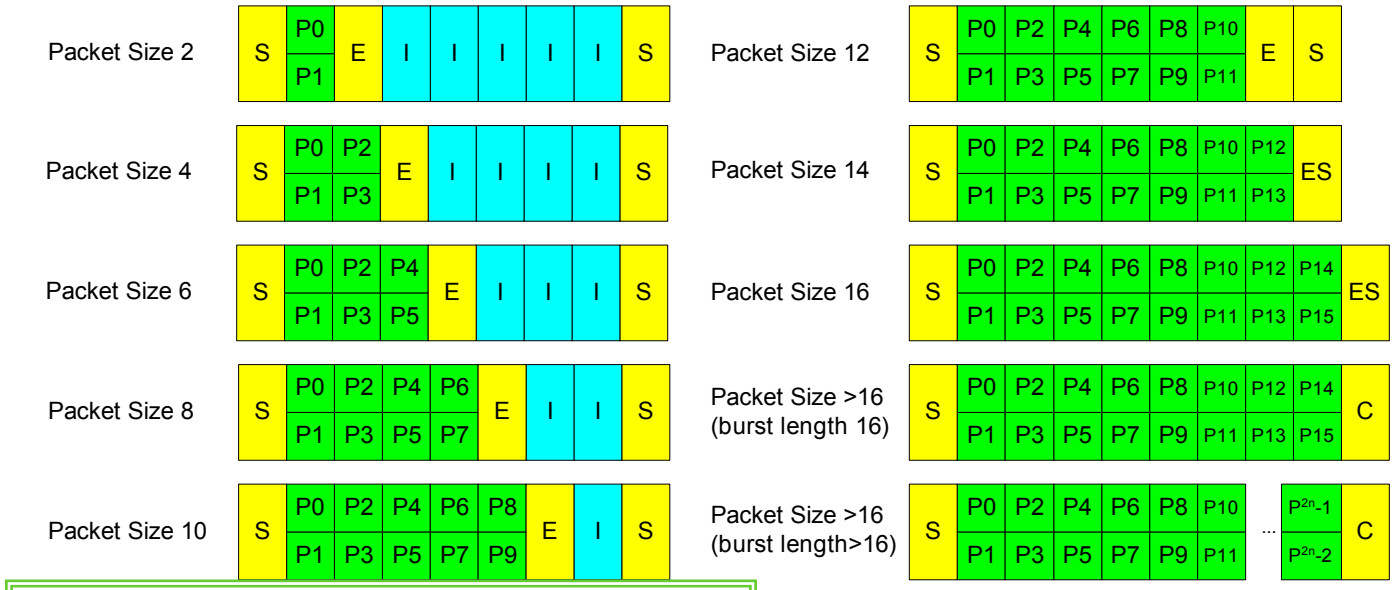
A timing diagram of the data path signals is shown in Fig. 6.9. This diagram is applicable to either the transmit or the receive interface. TCTL/RCTL is high when TDAT/RDAT contain control words. Idle periods correspond to back-to-back control words.

Fig. 6.9. Data Path Timing Diagram.

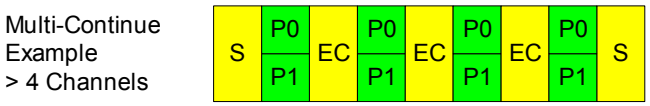


The actual clock rate used in practice is determined by the application at hand. A discussion of the minimum data clock rates required for certain applications is given in Appendix B.

Fig 6.10. SOP Spacing Example



Note: The spacing restriction applies only to SOP words. In multiple channel systems, any of the Idle or EOP characters in the above examples can be turned into continuation characters allowing more payload data to follow immediately.



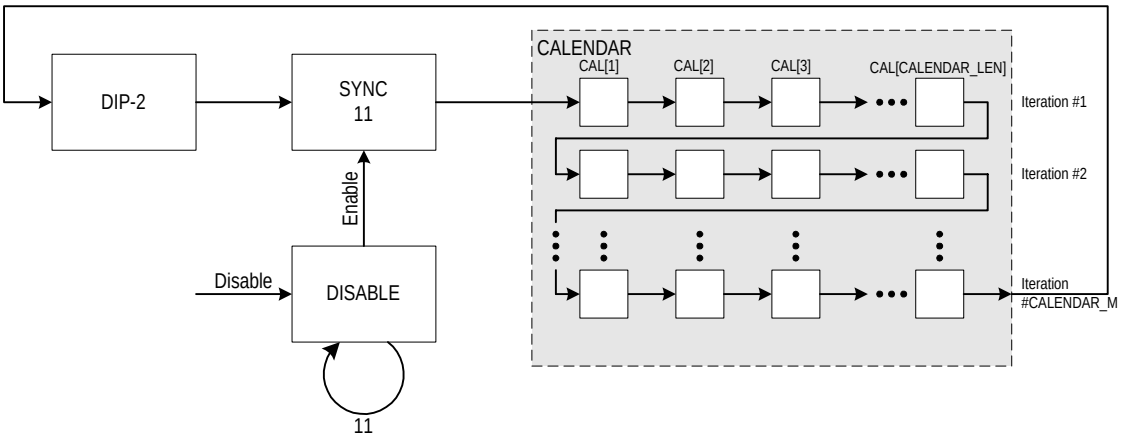
Key:
- S: Control Words with SOP asserted
- E: EOP Control Words with EOP asserted
- ES: Control words with SOP and EOP asserted
- I: Idle control words
- C: Continuation words (bits[15:12]=8)
- EC: Control words with EOP and Contination set
- P0, P1, … : Payload Data bytes

6.2.2.6.2.2 FIFO Status Channel.

FIFO status information is sent periodically over the TSTAT link from the PHY to the Link Layer device, and over the RSTAT link from the Link Layer to the PHY device. Implementation of the FIFO status channel for the transmit interface is mandatory; the corresponding implementation for the receive interface is optional. If both status channels are implemented, they shall operate independently of each other.

Figure 6.106.11 shows the operation of the FIFO status channel. The sending side of the FIFO status channel is initially in the DISABLE state and sends the "1 1" pattern repeatedly. When FIFO status transmission is enabled, there is a transition to the SYNC state and the "1 1" framing pattern is sent. FIFO status words are then sent according to the calendar sequence, repeating the sequence CALENDAR_M times followed by the DIP-2 code. FIFO status reporting can be reset to the "DISABLE" state by an implementation-specific "Disable" command.

Figure 6.106.11 FIFO Status State Diagram (Sending Side)

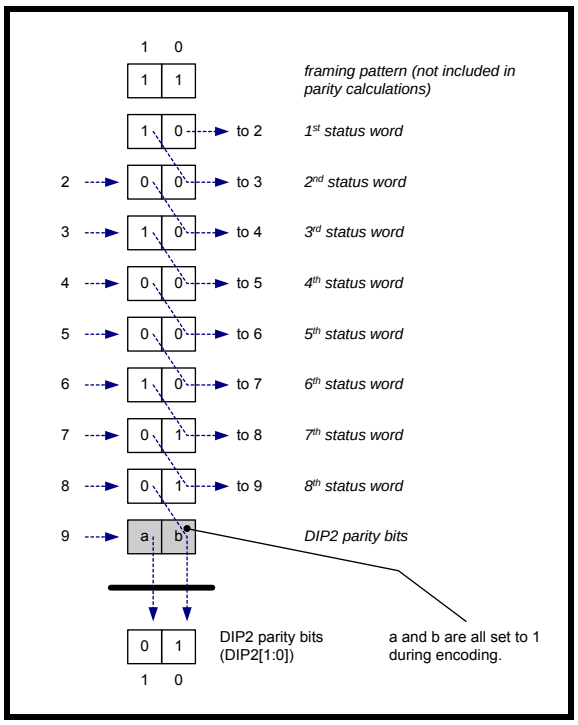


The FIFO status of each port is encoded in a 2-bit data structure, whose format is defined in Table 6.5. The most significant bit of each port status is sent over TSTAT / RSTAT [1], while the least significant bit is sent over TSTAT / RSTAT [0]. Section 6.3 discusses configuration of the FIFO status port sequence. The port sequences on the transmit and receive interfaces may be configured differently from each other. The "1 1" pattern is reserved for in-band framing; it must be sent once prior to the start of the FIFO status sequence.

A DIP-2 odd parity checksum is sent at the end of each complete sequence, immediately before the "1 1" framing pattern. The DIP-2 code is computed

over all preceding FIFO status indications sent after the last “1 1” framing pattern, diagonally over TSTAT / RSTAT [1] and TSTAT / RSTAT [0], as shown in Fig. 6.116.12. The first word is at the top of the figure, while the last word is at the bottom of the figure. The parity bits are computed by summing diagonally. Bits a and b in line 9 correspond to the space occupied by the DIP-2 parity bits; these bits are set to 1 during encoding. Note also that the “1 1” framing pattern is not included in the parity calculation. The procedure described applies to either parity generation on the egress path or to check parity on the ingress path.

Figure 6.116.12 Example of DIP-2 Encoding (Odd Parity).



While the parity bits can mimic the “1 1” pattern, the receiving end can still frame successfully by syncing onto the last cycle in a repeated “1 1” pattern and by making use of the configured length of the sequence (see Section 6.3) when searching for the framing pattern. A timing diagram of the FIFO status channel is shown in Fig. 6.126.13.

To permit more efficient FIFO utilization, the MaxBurst1 and MaxBurst2 credits are granted and consumed in increments of 16-byte blocks. For any given port, these credits correspond to the most recently received FIFO status. They are not cumulative and supersede previously granted credits for the given port, only if they are greater than the amount of credits remaining at the data path source (i.e., if the remaining credits are greater than the credit grant, the grant is ignored). To clarify, a transition to Satisfied is not intended

to cancel any previously granted credits. Likewise, a transition from Starving to Hungry does not reduce the number of credits available at the sending side of the data path. This is made clearer in Table 6.5. This does not prevent the sending side of the data path from treating the receipt of a Satisfied status as an XOFF indicator (and not using available credits); however, the sending side of the FIFO status channel cannot depend upon such They are not cumulative and supersede previously granted credits for the given port. Abehavior. A burst transfer shorter than 16 bytes (e.g., end-of-packet fragment) will consume an entire 16-byte credit.

The sending side of the data path cancels all credits upon detection of continuous “1 1“ in the status channel. The sending side of the data path may also choose to cancel credits as a result of detected errors in the status channel.

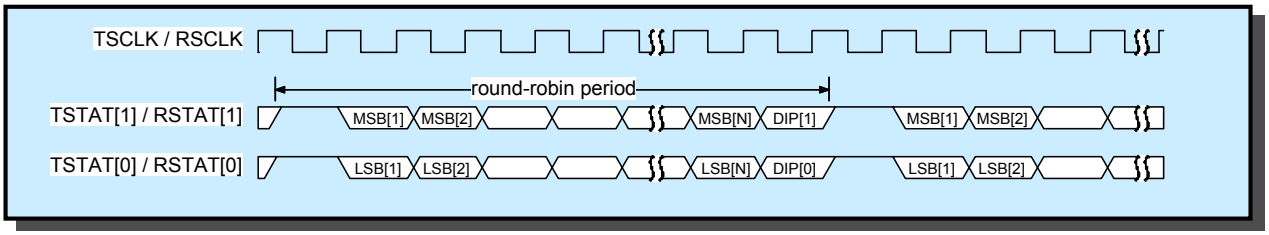
A continuous stream of repeated “1 1” framing patterns indicates a disabled status link. For example, it may be sent to indicate that data path deskew (Section 6.2.3) has not yet been completed or confirmed. When a repeated “1 1” pattern is detected, all outstanding credits are cancelled and set to zero.

Table 6.5. FIFO Status Format.

MSB	LSB	Description
1	1	Reserved for framing or to indicate a disabled status link.
1	0	SATISFIED Indicates that the corresponding port’s FIFO is almost full. When SATISFIED is received, only transfers using the remaining previously granted 16-byte blocks (if any) may be sent to the corresponding port until the next status update. No additional transfers to that port are permitted while SATISFIED is indicated.
0	1	HUNGRY When HUNGRY is received, transfers for up to MaxBurst2 16-byte blocks or the remainder of what was previously granted (whichever is greater) may be sent to the corresponding port until the next status update.
0	0	STARVING Indicates that buffer underflow is imminent in the corresponding PHY port. When STARVING is received, transfers for up to MaxBurst1 16-byte blocks may be sent

MSB	LSB	Description
		to the corresponding port until the next status update.

Fig. 6.126.13. FIFO Status Channel TimingTiming Diagram.
(Note: round-robin period = framing + (CALENDAR_LEN * CALENDAR_M) + DIP2.)



The indicated FIFO status is based on the latest available information. A STARVING indication provides additional feedback information, so that transfers can be scheduled accordingly. Applications which do not need to distinguish between HUNGRY and STARVING may only examine the most significant FIFO status bit. A further discussion on the required clock rate for the FIFO Status Channel is given in Appendix C.

Higher Bandwidth LVDS Status Operation

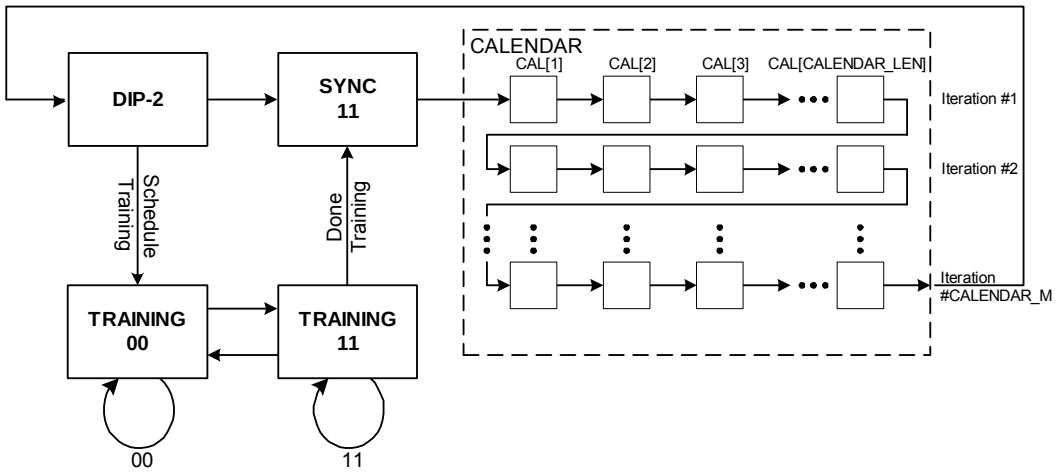
Applications which require higher FIFO Status Channel bandwidths than feasible with LVTTL I/O, may optionally use LVDS I/O instead. If LVDS I/O is used, double-edge clocking is used on TSCLK and RSCLK, running at the same rate as the corresponding data path rate. The framing structure and operation of TSTAT[1:0] and RSTAT[1:0] remain unchanged. AC timing parameters for an LVDS FIFO Status Channel are defined in Section 6.4.1.

A training sequence is scheduled to be sent at least once every preconfigured bounded interval (FIFO_MAX_T) on both the transmit and receive FIFO Status interfaces. These training sequences may be used by the receiving end of each interface for deskewing bit arrival times on the FIFO status and control lines. Training sequences may only be inserted between the DIP-2 code of the preceding cycle and the SYNC (1 1) word of the next calendar sequence. The sequence defined in this section is designed to allow the receiving end to correct for relative skew differences of up to +/- 1 bit time. The training pattern consists of ten words of “0 0” followed by ten words of “1 1”. The length of the training sequence can optionally be extended by repeating the 20 cycle pattern as many times as necessary. When using LVDS I/O and the training sequence on the FIFO Status Channel, the product of CALENDAR_LEN and CALENDAR_M (CALENDAR_LEN * CALENDAR_M) must be greater than or equal to sixteen to be able to distinguish between FIFO status information and the training sequence.

The training sequence is chosen so that it can be distinguished from a valid FIFO Status Channel message. In the absence of bit errors in the training pattern, a receiver should be able to successfully deskew the FIFO Status lines with one training pattern.

Setting FIFO_MAX_T equal to zero will disable the training sequence.

Fig. 6.14. LVDS Status Channel State Diagram.



An error monitor in the FIFO Status Channel receiver continuously verifies the status frames received. After multiple DIP-2 or framing errors, the monitor enters the DIP-alarm state and all previously granted credits are cancelled and will remain cancelled. No alarm is sent to the Status Channel transmitter. Rather, the Status Channel receiver works to reacquire synchronization on the status and regularly scheduled status channel training patterns that it is receiving. After multiple consecutive error-free status frames are received, the monitor returns to the DIP-normal state and credits can be acquired afresh.

Lower Bandwidth LVDS Status Operation

Most systems that need the electrical advantages of LVDS over LVTTL for the status channel do not require the high bandwidth status capability and cannot justify the additional complexity of status training and deskew. For implementations that support the LVDS status channel option, lower bandwidth operation is recommended for compatibility with future devices. Timing and protocol shall be identical to the LVTTL status channel, while the electrical LVDS specification is identical to that for the datapath.

6.2.3 6.2.3. Training Sequence for Data Path Deskew

A training sequence is scheduled to be sent at least once every preconfigured bounded interval (DATA_MAX_T) on both the transmit and receive data paths. These training sequences may be used by the receiving end of each interface for deskewing bit arrival times on the data and control lines. The sequence defined in this section is designed to allow the receiving end to correct for relative skew differences of up to +/- 1 bit time. The training sequence consists of 1 idle control word followed by one or more repetitions of a 20-word training pattern consisting of 10 (repeated) training control words and 10 (repeated) training data words. The initial idle control word removes dependencies of the DIP-4 in the training control words from preceding data words. Assuming a maximum of +/- 1 bit time in bit alignment jitter on each line, and a maximum of +/- 1 bit time relative skew between lines, there will be at least 8 bit times during which a receiver can detect a training control word prior to deskew. The training data word is chosen to be orthogonal to the training control word. In the absence of bit errors in the training pattern, a receiver should be able to successfully deskew the data and control lines with one training pattern.

The sending side of the data path on both the transmit and receive interfaces must schedule the training sequence in Table 6.6 (from cycles 1 through $20\alpha + 1$) at least once every DATA_MAX_T cycles, where DATA_MAX_T and α are configurable on start-up. Training sequences at the transmit and receive interfaces are scheduled independently. They must not be inserted within a payload burst transfer (i.e., not inserted between a payload control word and any of the subsequent data words until the end of transfer). Setting DATA_MAX_T equal to zero will disable the training sequence. Note that the DIP-4 code of the first control word following the training pattern is not affected by the preceding training data words, because an even number of training data words produces no net DIP-4 contribution.

Table 6.6. Training Sequence.

(Note: In cycle 1, XX and abcd depend on the contents of the interval after the last preceding control word.)

Cycle	TCTL / RCTL	TDAT[i] / RDAT[i]															
		15	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
1	1	0	X	X	0	0	0	0	0	0	0	0	0	a	b	c	d
2 to 11	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
12 to 21	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
20α -18 to 20α-9	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
20α-8 to 20α+1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0

6.3 Start-Up Parameters

The sequence of ports at a FIFO status channel is defined in a data structure called CALENDAR, where CALENDAR[i], i = 1,…,CALENDAR_LEN, refers to the ith port in the repeating sequence. CALENDAR_LEN typically corresponds to the number of ports with the lowest data rate that can be accommodated in the total data rate of the given application. CALENDAR_LEN must be at least as large as the number of active ports in the system. The calendar sequence (of length CALENDAR_LEN) is repeated CALENDAR_M times before the DIP-2 parity and "1 1" framing words are inserted. CALENDAR_LEN and CALENDAR_M are both greater than zero.

Examples:

1. Single OC-192 or 10 Gb/s Ethernet port: CALENDAR_LEN = 1, CALENDAR[1] = 1.
2. Four OC-48 ports: CALENDAR_LEN = 4, CALENDAR[i] = 1, 2, 3, 4.
3. Two OC-48 channels (ports 1 and 2), eight OC-12 channels (ports 3 through 10): CALENDAR_LEN = 16, CALENDAR[i] = 1, 2, 3, 4, 1, 2, 5, 6, 1, 2, 7, 8, 1, 2, 9, 10,…. Other combinations are feasible: CALENDAR[i] = 1, 3, 2, 4, 1, 5, 2, 6, 1, 7, 2, 8, 1, 9, 2, 10, ….
4. Ten 1 Gb/s Ethernet ports: CALENDAR_LEN = 10, CALENDAR[i] = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

The values of CALENDAR_LEN, CALENDAR_M and CALENDAR[i], i = 1,…,CALENDAR_LEN, must be identical in both the PHY and Link Layer devices for each interface. They need not be identical on both the transmit and receive FIFO status channels. The maximum supported value of CALENDAR_LEN is contained in the parameter MAX_CALENDAR_LEN, whose upper bound is implementation-specific. MAX_CALENDAR_LEN need not be identical on either side of the transmit or receive FIFO status channels. Users however, must ensure that the value of CALENDAR_LEN on the sending side of a FIFO status channel must not exceed MAX_CALENDAR_LEN on the receiving side.

For the FIFO Status channel(s), MaxBurst1 and MaxBurst2 may be configured to apply globally over all ports, or to apply on a per-port basis. In either case, both parameters must be consistently configured at the PHY and Link Layer devices for each interface, but need not be identical between the transmit and receive interfaces. MaxBurst1 must not be less than the corresponding MaxBurst2 (at the same port and interface).

For the data path deskew procedure, DATA_MAX_T is configured only on the sending side of the data paths on the transmit and receive interfaces. DATA_MAX_T need not be identical over both interfaces.

Start-up parameters are listed below in Table 6.7.

Table 6.7 Summary of Start-up Parameters.

Parameter	Definition	P	CH	Units
CALENDAR[i]	Port address at calendar location i.		I	(N/A)
CALENDAR_LEN	Length of the calendar sequence.		I	(N/A)
CALENDAR_M	Number of times calendar sequence is repeated between insertions of framing pattern.		I	(N/A)
MAX_CALENDAR_LEN	Maximum supported value of CALENDAR_LEN		I	(N/A)
MaxBurst1	Maximum number of 16 byte blocks that the FIFO can accept when FIFO Status channel indicates Starving.		C / I	16 byte blocks
MaxBurst2	Maximum number of 16 byte blocks that the FIFO can accept when FIFO Status channel indicates Hungry. MaxBurst2 <= MaxBurst1		C / I	16 byte blocks
α	Number of repetitions of the data training sequence that must be scheduled every DATA_MAX_T cycles.		I	(N/A)
DATA_MAX_T	Maximum interval between scheduling of training sequences on Data Path interface.		I	Cycles
FIFO_MAX_T	Maximum interval between scheduling of training sequences on FIFO Status Path interface.		I	Cycles

P – Provisionable

CH – Per channel (C) or per interface (I)

DATA_MAX_T − minimum range 0 to 2e16 cycles inclusive, in a power-of-2 granularity not greater than 256.

Alpha − minimum range 0 to 4, inclusive, granularity of 1.

Upon reset, the FIFOs in the datapath receiver are emptied, and any outstanding credits are cleared in the data path transmitter. After reset, but before active traffic is generated, the data transmitter shall send continuous training patterns. Transmission of training patterns shall continue until valid information is received on the FIFO Status Channel. The receiver shall ignore all incoming data until it has observed the training pattern and acquired synchronization with the data. Synchronization may be declared after a provisionable number of consecutive correct DIP-4 codewords are seen. Loss of synchronization may be reported after a provisionable number of consecutive incorrect DIP- 4 codewords is detected.

After reset but before active traffic is generated, the FIFO Status Channel transmitter shall send a continuous “1 1” framing pattern for LVTTL implementations, or continuous training patterns for optional LVDS implementations. Once the corresponding data channel has achieved synchronization, and a calendar has been provisioned, it may begin transmission of FIFO Status information. Once the data transmitter has received valid FIFO Status information (as indicated, for example, by a sufficient number of consecutively correct DIP- 2 codewords), it may begin transmission of data bursts to channels that have been provisioned and have space available.

In the event that the data path receiver is reset but the transmitter is still active, events at the receiver follow the same behavior as above. It shall ignore all incoming data until it has observed the training pattern and acquired synchronization with the data. It shall also send a continuous “1 1” framing pattern for LVTTL implementations (or continuous training patterns for optional LVDS implementations) on its FIFO Status Channel, cancelling previously granted credits and setting them to zero. In this case the transmitter should send continuous training patterns to facilitate reacquistion by the receiver.

In the event that the data path transmitter is reset but the receiver is still active, events at the transmitter follow the same behavior as above. The transmitter shall send continuous training patterns until a calendar is configured and valid status information is received on the FIFO Status Channel. At the same time, the receiver may have lost synchronization with the data, and begun sending continuous framing patterns (or continuous training patterns for optional LVDS implementations) on the FIFO Status Channel. Once the data transmitter has received valid FIFO Status information (as indicated, for example, by a sufficient number of consecutively correct DIP- 2 codewords), it may begin transmission of data bursts to channels that have been provisioned and have space available.

As a means to reacquire synchronization after a loss of data synchronization in dynamic mode, it is required that continuous data training be sent. By setting FIFO status to 11 for LVTTL Status, or continuous training patterns for LVDS Status, continuous training, comprising only of training data and training control words, will be sent out on the data channel until a valid status frame is received.

6.4 AC Timing

System-level reference points for specified parameters in this section are shown in Fig. 6.136.15. Corresponding reference points with respect to the clock edge are shown in Figs. 6.14 and 6.15.

Fig. 6.136.15. System-Level Reference Points.

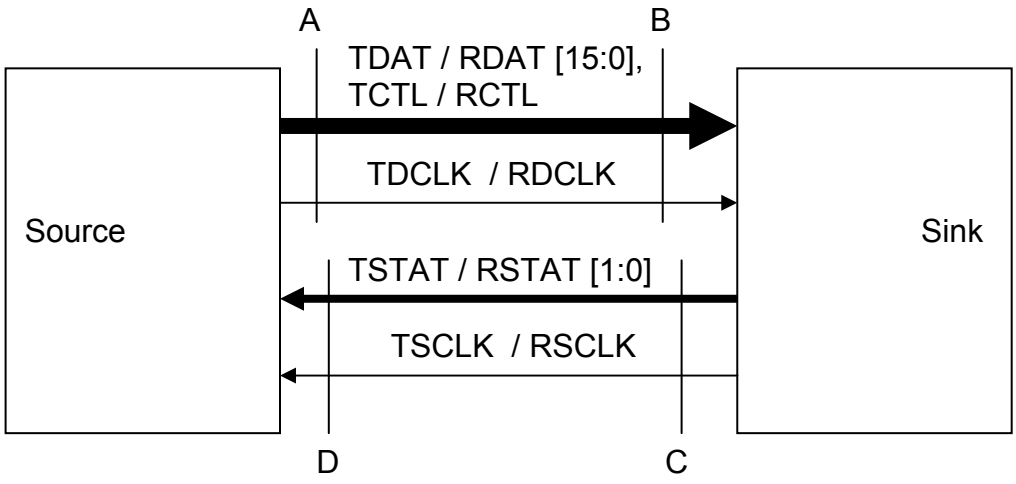


Figure 6.14 Reference Points for Data Path Timing Parameters.

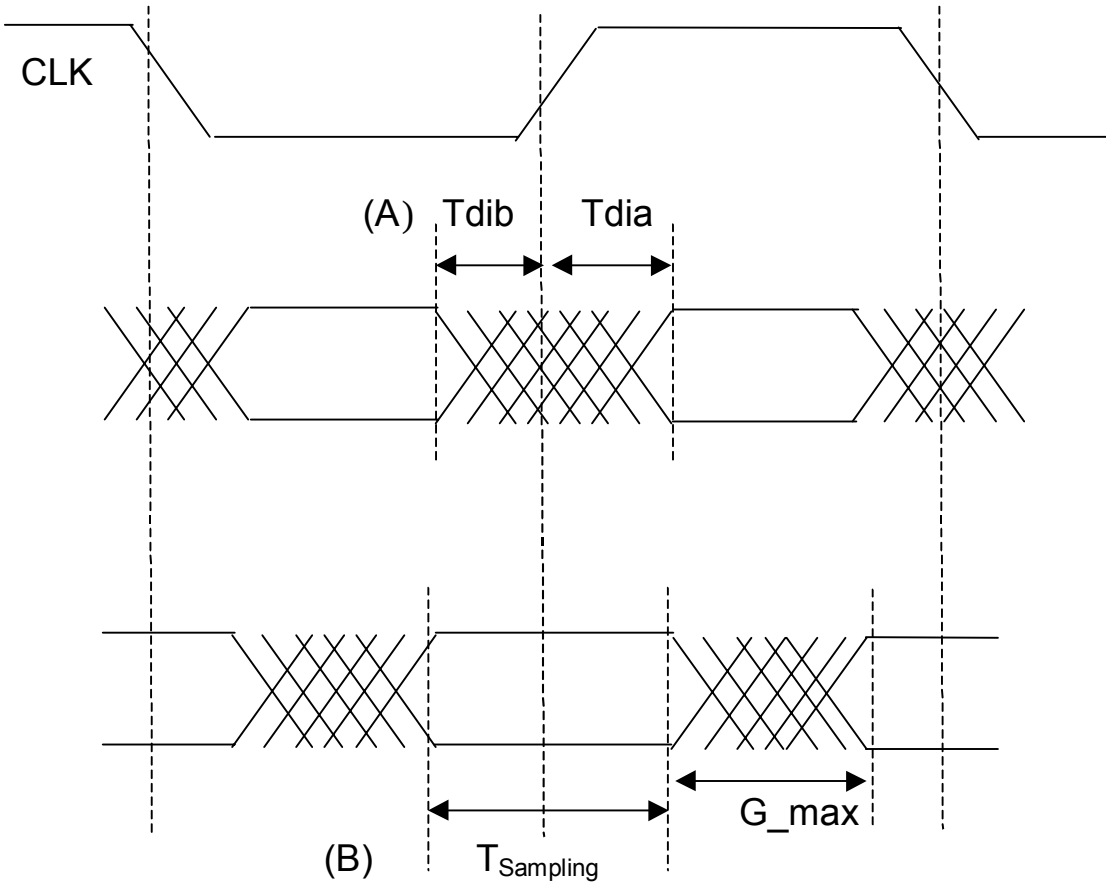


Figure 6.15 Reference Points for FIFO Status Channel Timing Parameters.

6.4.1.6.4.1 Data Path

Two sets of data path timing parameters are specified to support different bit alignment schemes at the receiver. Table 6.8a gives the corresponding parameters for the case of “static alignment”, in which the receiver latches data at a fixed point in time relative to clock (requiring a more precisely specified sampling window). Table 6.8b gives the corresponding parameters for the case of “dynamic alignment”, in which the receiver has the capability of centering the data and control bits relative to clock. From an AC timing perspective, a compliant interface only needs to meet the parameters at the data path for either static or dynamic alignment, but may also comply to both sets of parameters. A compliant driver must meet both timing specifications to be interoperable with both types of receivers. For the case of static alignment, a sample timing budget suitable for up to fD = 350 MHz in Table 6.8a is shown in Appendix D. Also shown in that section is a sample budget for the case of dynamic alignment. Corresponding reference points with respect to the clock edge are shown in Figs. 6.16.

Table 6.8a. Data Path Interface Timing (Static Alignment).

Symbol	Description		Min	Max	Units
fD	TDCLK / RDCLK Frequency		311		MHz
	TDCLK / RDCLK Duty Cycle		45	55	%
T_dia, T_dib	Data invalid window with respect to clock edge. (Reference point A)			280	ps
G_max	Worst-case cumulative skew and jitter contribution. (Reference point B)			790	ps
Tsampling	Data valid window with respect to clock edge. (Reference point B)			$\frac{1}{2fD}$ - G_max	ps
	20% - 80% rise and fall times (UI = 1/2fD)	(Reference point A)	100 ps	0.30 UI	
		(Reference point B)	100 ps	0.36 UI	

Figure 6.16 Reference Points for Data Path Timing Parameters.

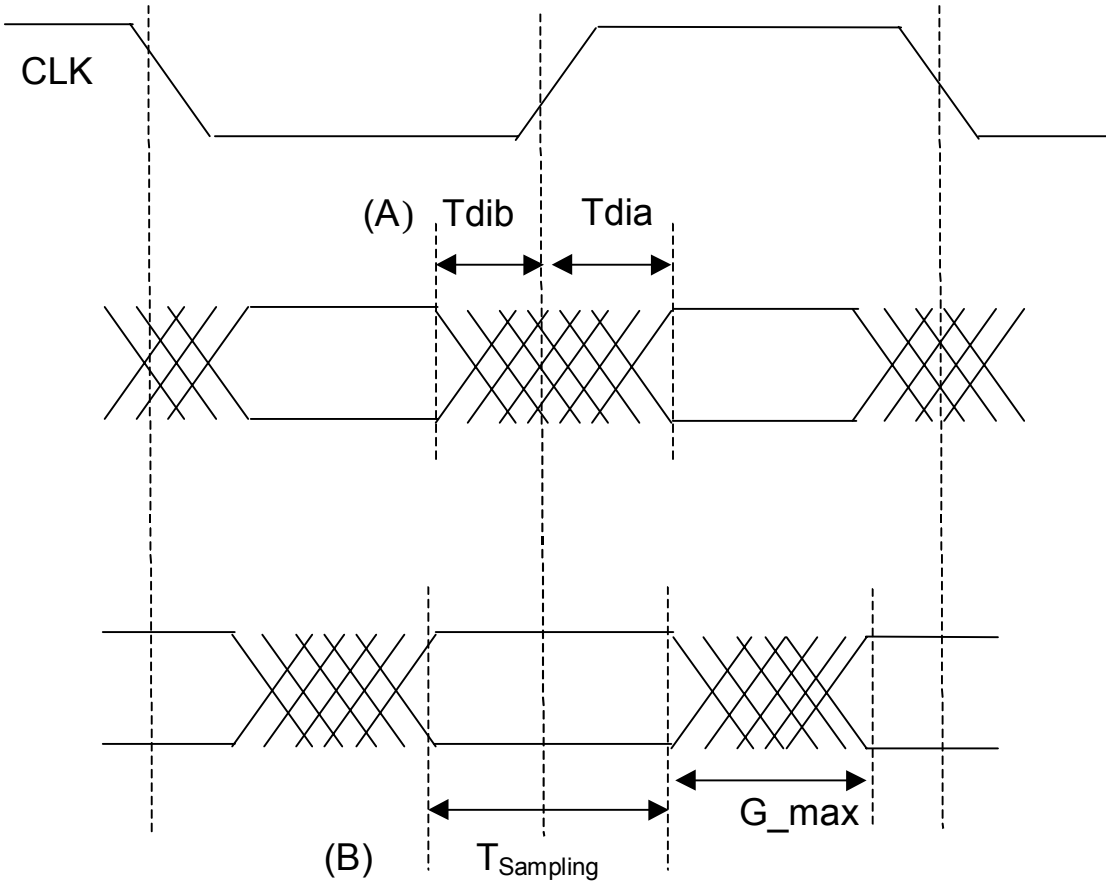


Table 6.8b. Data Path Interface Timing (Dynamic Alignment).

Symbol	Description		Min	Max	Units
fD	TDCLK / RDCLK Frequency		311		MHz
	TDCLK / RDCLK Jitter (at reference point A)			0.10	UI
	TDAT / RDAT / TCTL / RCTL Jitter (at reference point A)			0.24	UI
	20% - 80% rise and fall times (UI = 1/2fD)	(Reference point A)	100 ps	0.30 UI	
		(Reference point B)	100 ps	0.36 UI	

Notes:

1. Rise and fall times assume nominal 100-ohm termination and exclude reflections.
2. All timing parameters are measured relative to the differential crossing point of the corresponding clock signal.
3. Jitter parameters are peak-to-peak, measured above fD / 1000 and below fD.
4. Receiver sensitivity is assumed to be less than or equal to 100 mV.
5. Assumes a 5 pF output load at reference point A, a 10 pF load at reference point B, and a 50-ohm transmission line in between.
6. Assumes up to 20 ps skew between traces of a differential pair.

Table 6.8c gives the normative values for Reference Point A to support both static and dynamic alignment. All notes for Table 6.8b also apply to this table. Compliance is required for all applications.

Table 6.8c Normative values for Reference Point A

Symbol	Description	Min	Max	Units
fD	TDCLK / RDCLK Frequency	311		MHz
	TDCLK / RDCLK duty Cycle	45	55	%
T_dia, T_dib	Data invalid window with respect to clock edge (Reference point A)		280	ps
J_datA	Data Jitter at Point A (Note 2)		0.24	UI
J_clkA	Clock Jitter at Point A (Note 1)		0.1	UI

Notes:
1: Clock jitter applies to TDCLK / RDCLK
2: Data jitter applies to TDAT / TCTL / RDAT / RCTL

Table 6.8d gives the informative values for jitter and skew at Reference Point B to support dynamic alignment. All notes for Table 6.8b also apply to this table. Reduced values are not precluded for particular applications.

Table 6.8d Informative values for Reference Point B

Symbol	Description	Min	Max	Units
J_datB	Data Jitter Tolerance at Point B (Note 2)	0.44		UI
J_clkB	Clock Jitter Tolerance at Point B (Note 1)	0.13		UI
Sk_datB	Data Skew Tolerance at Point B	+/- 1		UI

Notes:
1: Clock jitter applies to TDCLK / RDCLK
2: Data jitter applies to TDAT / TCTL / RDAT / RCTL”

6.4.2.6.4.2 FIFO Status Channel

The following section describes AC timing parameters for a FIFO status channel implemented using LVTTL I/O. For optional LVDS FIFO status channel implementations, the reader is referred to the LVDS data path parameters in Section 6.4.1. As noted in Table 6.9, the maximum clock frequency of the LVTTL FIFO Status Channel shall not exceed one quarter of the selected data path clock rate. Corresponding reference points with respect to the clock edge are shown in Figure 6.17.

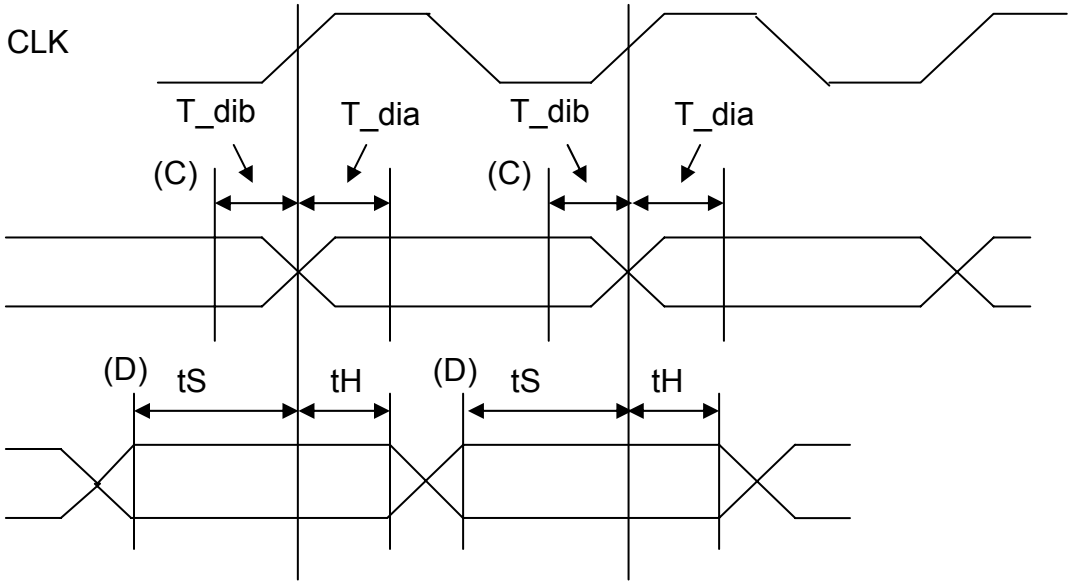
Table 6.9. FIFO Status Channel Interface Timing (LVTTL I/O).

Symbol	Description		Min	Max	Units
fS	TSCLK Frequency			fD / 4	
	TSCLK Duty Cycle		40	60	%
T_dia	Data invalid window with respect to clock edge.	(Reference point C)	2.5		ns
T_dib		(Reference point C)	1		
tS_{SCLK}	TSTAT Setup time to TSCLK , RSTAT Setup time to RSCLK.	(Reference point D)	2		ns
tH_{SCLK}	TSTAT Hold time to TSCLK, RSTAT Hold time to RSCLK.	(Reference point D)	0.5		ns

Notes on LVTTL I/O Timing:

1. When a set-up time is specified between an input and a clock, the set-up time is the time from the 1.4 Volt point of the input to the 1.4 Volt point of the clock.
2. When a hold time is specified between an input and a clock, the hold time is the time from the 1.4 Volt point of the clock to the 1.4 Volt point of the input.
3. Assumes a 25 pF, 500 ohm load.
4. For maximum interoperability and reduced PCB complexity, it is recommended that implementations allow inputs to sample on either clock edge via a configurable setting. The output transitions may be configurable to occur on either edge. The default setting should match Figure 6.17.

Figure 6.17 Reference Points for FIFO Status Channel Timing Parameters.



6.5 DC Parameters

Table 6.10 lists applicable DC thresholds for the LVTTL FIFO Status Channel.

Table 6.10. LVTTL DC Thresholds

Parameter	Symbol	Min (V)	Max (V)
Output High Voltage	V_OH	2.4	3.6
Input High Voltage	V_IH	2.0	3.6
Input Low Voltage	V_IL		0.8
Output Low Voltage	V_OL		0.4

The LVDS data path and FIFO Status channel conform to ANSI/TIA/EIA-644-A-2001 subject to the following recommendations:

1. Vos = 1125 – 1375 mV range may be a typo and should be 1125 – 1275 mV instead.

2. Vgpd_max is not specified, but can be calculated to be +/- 825 mV, if Vos = 1275 mV.

3. Ro @ Tx is not specified (cf. IEEE), but it is recommended that the range specified by IEEE (40 – 140 ohms) be used.

4. Input voltage range = 825 – 1575 mV. Only a subset of the 0 – 2400 mV range specified in ANSI/TIA/EIA-644-A-2001 is permitted, to simplify receiver design. There is no need to support the extended range for a short reach device interface.

5. To conform to SFI-4 Phase 1 electricals, it is recommend that Rin = 80 – 120 ohms.

6. To conform to SFI-4 Phase 1 electricals, it is recommended that Vod = 250 – 600 mV be specified.

Note that recommendations 1 to 4 would also be relevant to the SFI-4 Phase 1 electrical specification.

7 Appendices

7.1 Appendix A. Error Protection Capability of the DIP-4 Code.

7 Appendix A. Error Protection Capability of the DIP-4 Code.

This appendix does not form an integral part of this specification. In this section, the undetected error probability of the DIP-4 code is evaluated for a burst transfer length of 64 bytes, in the presence of random bit errors.

With the DIP-4 code, each constituent parity bit is computed over 66*8/4-1 information bits. These information bits, together with the parity bit, constitute a single-parity codeword. In the assumed case of odd parity, the value of the parity bit is set such that the total number of ones in the codeword is an odd number. It can easily be seen that any odd number of errors can be detected in a single-parity codeword. Hence, an undetected error event for a given codeword occurs whenever an even number of errors falls on that codeword. An undetected error event on a given burst transfer occurs when an undetected error event occurs on at least one of the constituent codewords.

Let P_{UC} and P_{UT} denote, respectively, the undetected error probabilities in a single-parity codeword and in a DIP-4 burst transfer. Let *L* denote the number of bits in a codeword and *p* the probability of a bit error. Enumerating all combinations of even-numbered errors in the codeword, we have,

$$P_{UC} = \sum_{\substack{i=2, \\ i\text{even}}}^{L} \binom{L}{i} p^i (1-p)^{L-i}.$$

Since an undetected error in a burst transfer corresponds to an undetected error in at least one codeword,

$$P_{UT} = 1 - (1 - P_{UC})^4.$$

Table A.1 shows the corresponding probabilities of undetected error in a burst transfer, for the DIP-4 code, over a range of bit error rates. Also shown for comparison are corresponding probabilities without an error-detection code. It can be seen that a DIP-4 code reduces the undetected error probability by several orders of magnitude. While the bit error rates in well-designed implementations may already be very low, the DIP-4 code can reduce the undetected error probability to extremely negligible levels with minimal added complexity to the interface implementation. Note however that longer burst transfers increase the DIP-4 extent, which increases the probability of an undetected error.

Table A.1. Error Detection Capability of DIP-4 Code.

Bit Error Rate (*p*)	Probability of Undetected Error	
	Without Error Detection	With DIP-4 Code
1.0E-6	5.28E-4	3.458E-8
1.0E-7	5.28E-5	3.458E-10
1.0E-8	5.28E-6	3.459E-12
1.0E-9	5.28E-7	3.464E-14
1.0E-10	5.28E-8	3.441E-16
1.0E-11	5.28E-9	(tiny)

7.2.7.2 Appendix B. Minimum Data Path Bandwidth Requirements for Typical Applications.

This appendix does not form an integral part of this specification. This section discusses the minimum bus frequencies required for a number of applications, assuming maximum 64-byte payload data transfers. The numerical results can be used to provide guidance on the minimum operating frequency of the bus. To provide design margin, the actual operating frequency would be higher to account for training and other overhead. Calculations in this section assume that the training bandwidth is insignificant and is not included. For the sake of brevity, the word SONET is used to refer to either SONET or SDH.

Since the PHY device may remove or insert framing overhead to data received from or transmitted to the line side, the actual data rate over the POS-PHY interface may be less than the operating line rate. For a given line rate, the actual packet data rate becomes smaller with the packet length, since the framing overhead becomes a larger proportion of the line bandwidth. The POS-PHY interface, however, inserts a 16-bit control word between payload burst transfers. Hence, the bus tends to operate less efficiently with shorter packet lengths (and therefore tends to require a proportionally higher bus bandwidth for a given data transfer rate). The minimum operating frequency of the bus is the result of an interplay between the control word and the line overhead.

The following cases are considered in this section:

1. ATM cells over SONET.

2. (HDLC-framed) Packet over SONET.

3. 10 Gb/s Ethernet LAN PHY Framing.

In case 1, ATM cells are transported back-to-back in the payload area of a SONET frame. There is an 8-bit HEC field in the ATM cell header, which may or may not be transferred over the POS-PHY interface. In case 2, back-to-back HDLC-framed packets are transported in the SONET payload area. Apart from inter-frame flags, the HDLC framing overhead is assumed to consist of an 8-bit Type field, an 8-bit Address field, and a 32-bit frame check sequence (FCS). Byte-stuffing events will increase the line overhead, but these contributions are ignored since the worst-case in this discussion corresponds to the absence of byte stuffing. For cases 1 and 2, the SONET overhead is assumed to be 3.7% of the line rate. In case 3, the packets are encapsulated in Ethernet frames. In this case, jumbo frames (to the order of 8192 bytes long) are also considered, as these frames may be encountered in some applications even though they are longer than the maximum specified by the IEEE.

In general, the bus frequency of the interface, f_I, can be expressed as,

$f_I = f_S ABC / W,$

Where

f_S = line rate (9.953 Gb/s for SONET, 10 Gb/s for Ethernet 10 Gb/s LAN PHY).

A = line (Layer 1) efficiency,

$$= \begin{cases} 0.963, & \text{SONET/SDH} \\ 1, & \text{10 Gb/s Ethernet LAN PHY} \end{cases}$$

B = Layer 2 framing efficiency

$= L / L'$, where

$$L' = \begin{cases} L, & \text{53 - byte ATM cells} \\ L+1, & \text{52 - byte ATM cells, POS incl HDLC header/FCS} \\ L+7, & \text{POS excl HDLC header/FCS} \\ L+12, & \text{10 Gb/s Ethernet LAN PHY} \end{cases}$$

$$C = \left\{ \underbrace{L}_{PktLen} + \underbrace{\left[\left\lceil \frac{L}{W} \right\rceil - \frac{L}{W}\right] W}_{WidthOverhead} + \underbrace{\left\lceil \frac{L}{M} \right\rceil N}_{BurstOverhead} \right\} / L$$

L = packet length, including packet overhead transferred over interface,

W = interface width (number of parallel data lines),

M = maximum burst transfer size,

N = additional interface overhead per transfer.

Setting W = 16 bits, M = 64 bytes, N = 2 bytes, the minimum bus frequency is obtained by finding the maximum of f_I over the applicable range of L.

For case 1, the minimum bus frequency, $f_{\min}$, for 52-byte cell transfer is 610.37 Mb/s. The corresponding $f_{\min}$ for 53-byte cells is 632.97 Mb/s.

For case 2, considering HDLC header and FCS transfer over the interface, the maximum occurs at L = 65 bytes, where the corresponding $f_{\min}$ is 635.37 Mb/s. In the more common situation where the header and FCS are not transferred over the interface, $f_{\min}$ increases with L (though not monotonically), approaching an asymptotic limit of roughly 618 Mb/s.

Due to the relatively long inter-frame gap (12 bytes) for case 3, the required f_I is fairly small for short *L* but increases with *L* as the gap becomes a smaller proportion of the frame length. Hence, the worst case corresponds to jumbo frame transfers. These frames are much longer than the maximum specified by IEEE, but have been used in some applications. An upper bound to $f_{\min}$ can be obtained by assuming that the overhead due to the preamble and the inter-frame gap is negligible for very long frames. Hence, $L' \approx L$, and $f_{\min} \leq$ 644.53 Mb/s.

Results for the minimum bus frequencies are summarized below in Table B.1. Further information on the relationship between packet length and minimum bus frequency are shown in Figs. B.1 through B.4.

Table B.1 Summary of Minimum Bus Frequencies.

Traffic Type	Min Freq (MHz)
ATM Cells, 52 bytes, no HEC	610.37
ATM Cells, 53 bytes, with HEC	632.97
HDLC, L=65 (with HDR,CRC)	635.37
HDLC, no HDR, CRC	618.00
10G Ethernet LAN, worst case	644.53

Figure B.1 Packet Length vs Frequency Chart

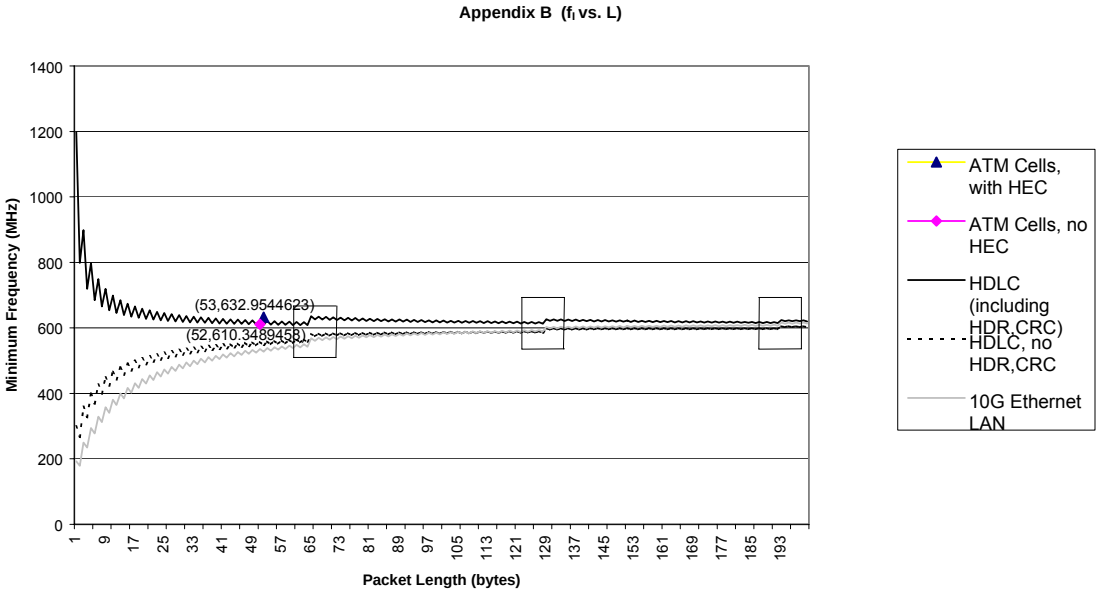


Figure B.2 Packet Length (64 bytes) vs Frequency

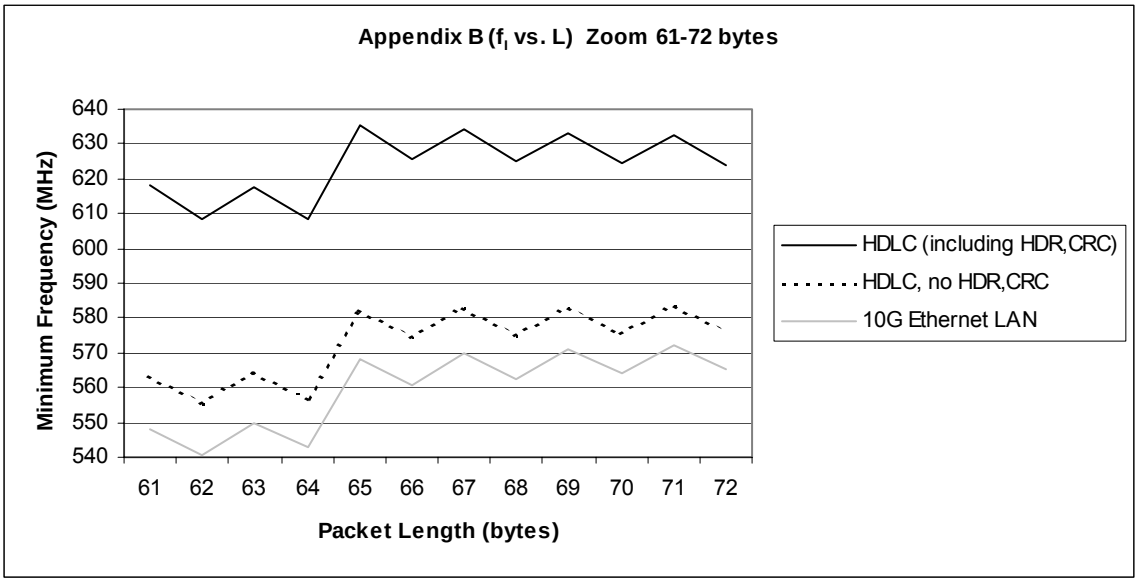


Figure B.3 Packet Length (128 bytes) vs Frequency

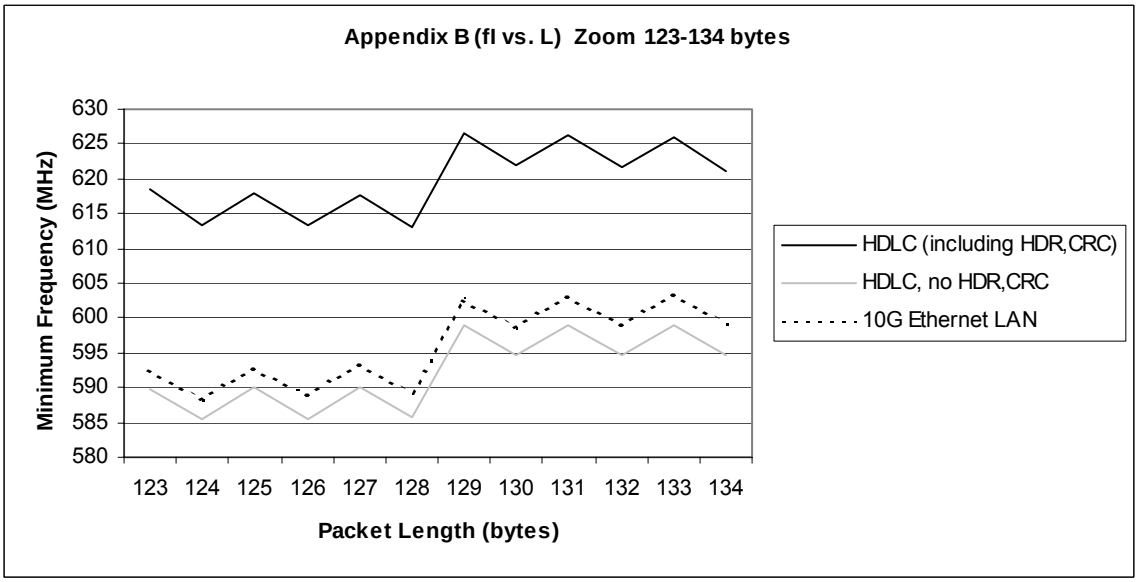
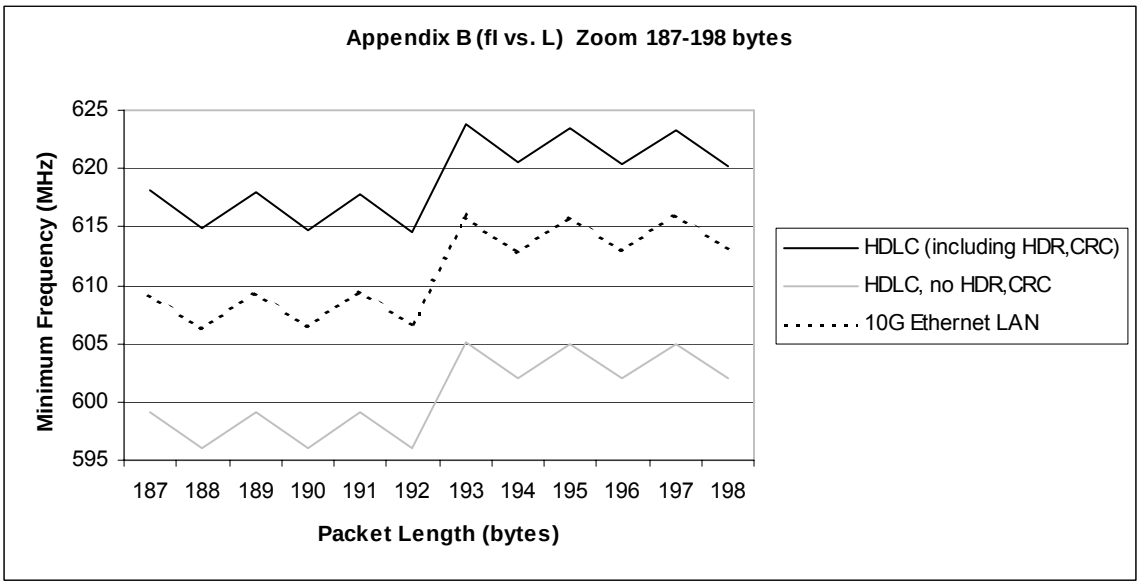


Figure B.4 Packet Length (192 bytes) vs Frequency



7.3.7.3 Appendix C. FIFO Status Bandwidth Requirements.

This appendix does not form an integral part of this specification. It is not a trivial task to estimate a universally applicable required bandwidth on the status channel, as plausible worst-case scenarios are likely to be application-specific, and in any case, dependent on other implementation details such as FIFO depth. However, some general statements can be made without delving too deeply into specifics of particular implementations.

Given the nominal operating frequency of the data path, it is reasonable to expect that the status channel can be run at 1/8th the serial rate of the data path. For the case of very short packets sent in 16-byte bursts, there is roughly one status update opportunity for each minimum transfer period, as eight cycles will have elapsed in the data path for each cycle in the status channel. The overhead in the status channel (from parity and framing) can be made sufficiently small with a suitably long calendar length.

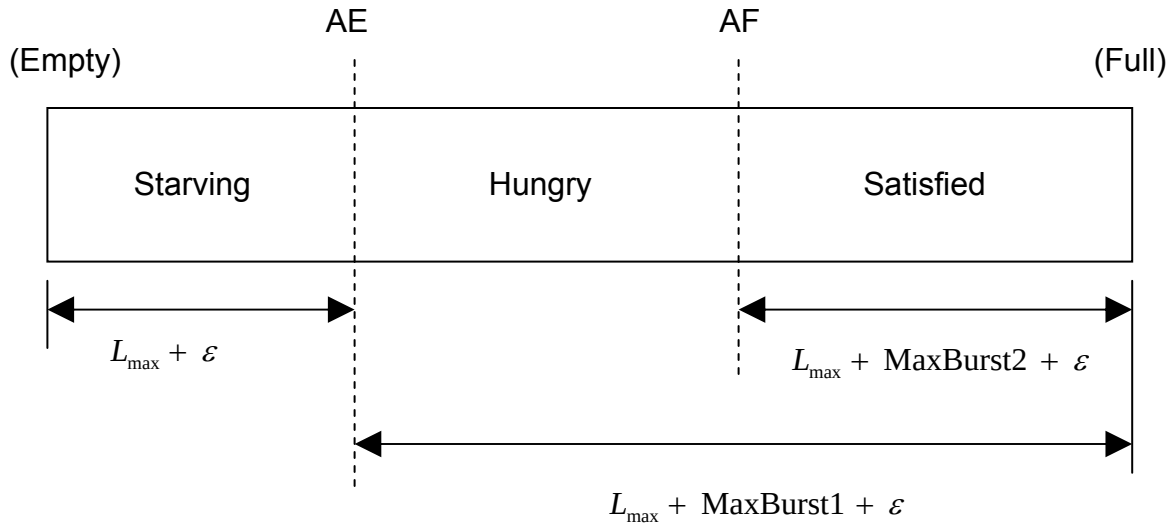
In the more typical cases corresponding to Internet traffic traces, the shortest packets will be roughly 40 bytes long (requiring 21 cycles to transfer, including control overhead). For a 65-byte packet segmented into two transfers, an average of 35 cycles will have elapsed between those transfers (33 cycles for payload and two for control). Hence, such transfers are likely to be in the range of 17.5 ~ 21 cycles, ample time for at least two status update opportunities within the transfer interval. This result does not fundamentally change in multi-port configurations, even with widely different line rates, as each constituent line rate will ultimately bound the amount of data path bandwidth allocated to a given port in the long run. For example, a port running at only a tenth of the total bandwidth supported by the interface can only send and receive data by that same fraction of bandwidth in the long run (excluding transient intervals). By weighting the calendar such that this port has an average of one update out of every ten opportunities, the corresponding status channel bandwidth for this port is scaled accordingly.

The thresholds for STARVING and HUNGRY are set such that the FIFO can accept at least MaxBurst1 and MaxBurst2 (16-byte) blocks respectively, plus an additional amount to account for feedback-response delay. In order to guard against potential buffer underflow, the lowest threshold must be set high enough to allow the other end to respond to transitions to the state of lowest FIFO occupancy in a reasonable length of time (to the order of the status update interval, plus scheduler response time). MaxBurst2 and MaxBurst1 (if applicable) must be provisioned to allow adequate utilization of transfer bandwidth between status updates for the given port.

Fig. C.1 shows one possible way for relating FIFO thresholds to MaxBurst1 and MaxBurst2 as well as the response latency. L_max corresponds to the worst-case response time, from the delay in receiving a status update over the FIFO status channel, until observing the reaction to that update on the corresponding data path. The quantity, ε, corresponds to the difference between the granted credit and the actual data transfer length. This difference arises from various protocol overheads and quantization errors in the packet scheduler and the data path.

Fig. C.1. Sample FIFO Thresholds.

(AE = Almost Empty waterline, AF = Almost Full waterline)



7.4 7.4 Appendix D. Sample LVDS Timing Budgets.

This appendix does not form an integral part of this specification. The calculations in this appendix are conservative; they assume linear addition of timing error sources. Reference points noted are shown in Fig. 6.136.15.

7.4.1 7.4.1 D.1. Static Alignment at the Receiver.

A sample timing budget suitable for operation up to fD=350 MHz is shown in Table D.1.

Table D.1. Timing Budget (Static Alignment).

Description		**Value (peak-to-peak, ps)**
Before Reference Point A	Clock to Data Skew (fixed) [note 1]	200
	Clock Duty Cycle Distortion (random)	140
	Data Duty Cycle Distortion (random)	140
	Data Jitter (random) [note 2]	80
	Subtotal (= Tdia + Tdib)	**560**
Between Reference Points A and B	Clock to Data Skew (fixed) [note 3]	150
	Relative Jitter (random) [note 4]	80
	Subtotal (G_max)	**790**
After Reference Point B	Sampling Error (fixed) [note 5]	550
	Relative Jitter (random)	80
	Total	**1420**

Notes:

1. Includes variations in clock-to-data outputs across LVDS drivers, and across process, voltage and temperature.

2. Includes jitter from all sources, including data-dependent contributions.

3. Includes trim / data-eye centering error (assumes clock offset achieved on board). Also includes 20 ps for sum of package and board length mismatch.

4. Relative jitter between reference points A and B includes contributions from the board as well as connector.

5. Includes setup and hold times of D flip-flop.

7.4.2 7.4.2 D.2. Dynamic Alignment at the Receiver.

A sample timing budget is shown in Table D.2. The margin remaining is 1 – 0.90 = 0.10 UI.

Table D.2. Timing Budget (Dynamic Alignment).

Description		Value (UI, peak-to-peak)
Clock jitter at reference point A		0.10
Data jitter at reference point A		0.24
Data jitter between reference points A and B		0.20
After reference point B	On-chip jitter due to routing.	0.01
	SYNTH Jitter	0.05
	Sampling granularity	0.25
	Variation in sampling position.	0.05
Total		0.90

Notes:

1. Sampling granularity (quantization error) assumes 4 samples per bit period.

Table D.3. Data and Clock Jitter Budget (Dynamic Alignment).

Jitter Source	Jitter (UI)
Clock Jitter at Point A (Notes 4,7)	0.10 (Notes 1,6,8)
Data Jitter at Point A	0.24 (Notes 1,6)
Data Jitter between A and B	0.20 (Notes 2,5,6)
Clock Jitter between A and B	0.03 (Notes 2,6,8)
Data Jitter at Point B	0.44 (Notes 3,6)
Clock Jitter at Point B	0.13 (Notes 3,6)

Notes:

1. Point A numbers are already specified in Table 6.8b.
2. Point A to B jitter taken from Appendix D, table D2
3. Input jitter tolerance calculated by summing jitter sources linearly (pessimistic)
4. No correlation assumed between clock and data jitter (pessimistic)
5. 0.20 UI is pessimistic for most applications (allows for 20” of FR4)
6. Jitter at points A and B is measured to the zero crossing point of each differential pair
7. Point A and Point B are at the pins of the device
8. Clock jitter is measured from the rising to rising edge and from falling to falling edge

7.5 7.5 Appendix E. Control Word Extension.

This appendix does not form an integral part of this specification. It is the

recommended implementation for one or more additional control words following the payload control word that provide control and address extension when more than 256 addresses are required.

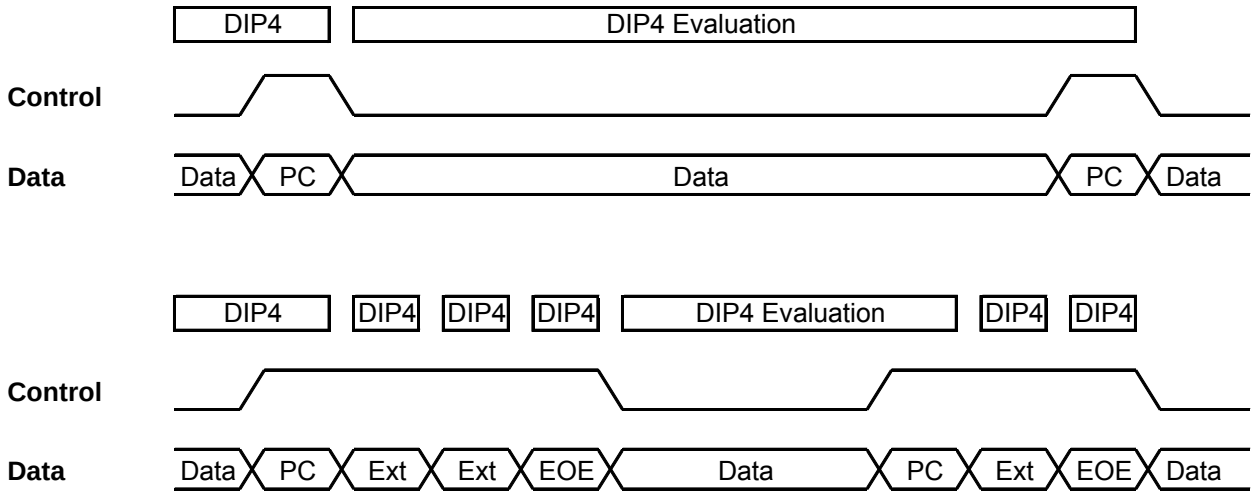
Extended control words are identified by the bit combination of [15:12] = “0111” or [15:12] = “0001”. For a single extension control word, the bit combination of [15:12] = “0111” is used. For more than one extension control word, the bit combination of [15:12] = “0001” is used except for the last extended control word which uses the bit combination of [15:12] = “0111. If used, one or more extension control words must follow the payload control word immediately. Payload data transfer must begin immediately after the extended control word with the bit combination of [15:12] = “0111”. Figure E.1 shows sample timing diagrams for control word extension.

In extension control words, the EOPS and SOP bits are used for type indication only. When using extension control word(s), EOPS and SOP information is contained in the preceding payload control word.

Bits [11:4] of the extension control word immediately following the payload control word indicate the upper eight bits of a 16-bit address. A full address is formed by concatenating the 8-bit extended address with the 8-bit normal address from the payload control word.

If more extension control words are sent, bits [11:4] of the second and following extension control words are user-defined and application-specific. Bits [3:0] of the extension control word is the 4-bit odd parity calculated over this control word only.

The diagram of Figure E.2 illustrates the sequence of possible control words with modifications to support extended addressing.



PC = Payload Control
Ext = Extension Control
EOE = End of Control Word Extension

Figure E.1 Sample Timing Diagrams when using Extended Addressing

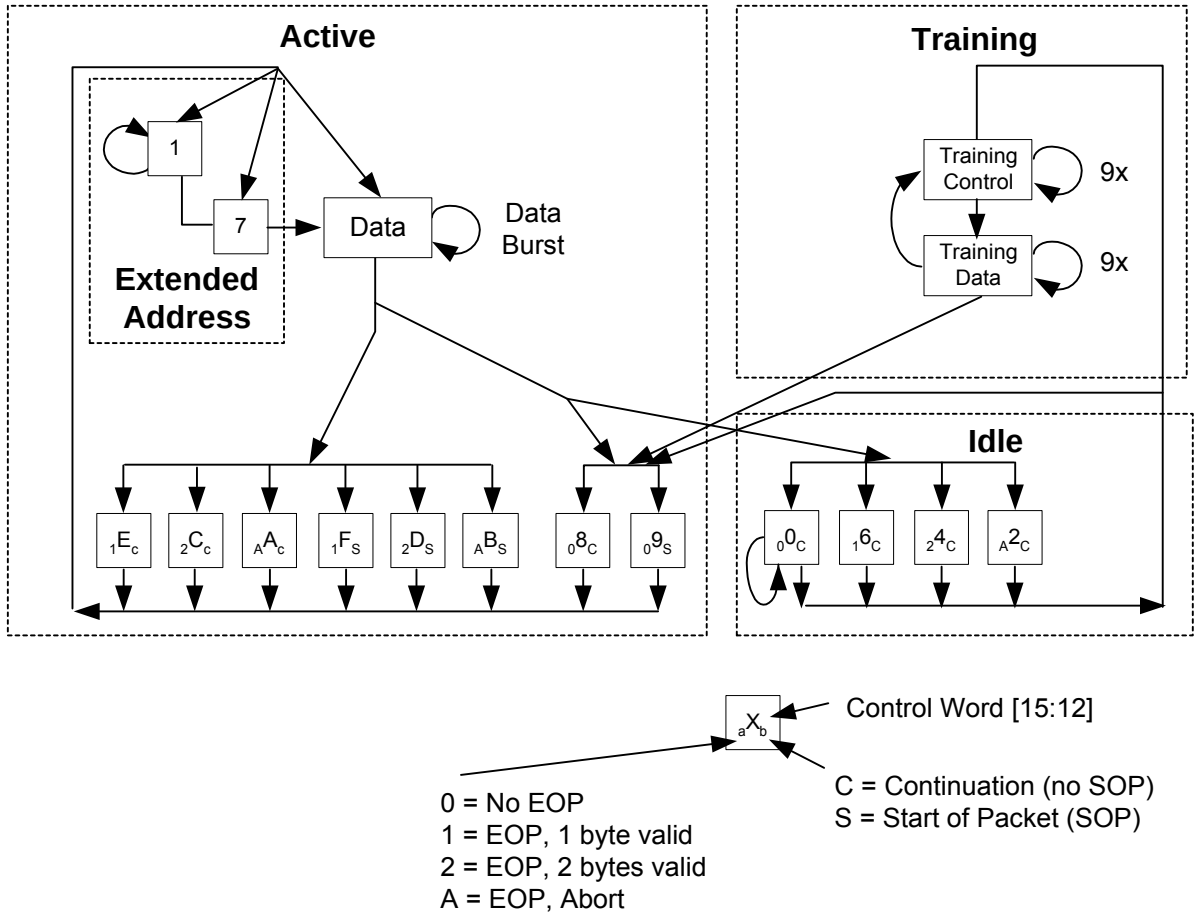


Figure E.2 Control and Data Sequences when using Extended Addressing

7.6 7.6 Appendix F. Narrow Interface Applications.

This appendix does not form an integral part of this specification. It is the recommended implementation for OC-48 applications via a narrow mode of operation. This enables the development of interfaces that are downward compatible with OC-48 data rates and saves pins.

Operation of a narrow-mode interface, if required, may follow the description in the main body of this text. 16-Bit words are divided into four 4-bit nibbles, which are sent across the interface in most-significant to least significant order (i.e. bits[15:12], then bits [11:8], then bits [7:4] then bits [3:0]). The format and sequence of data words and control words is unchanged. Operation of the FIFO Status channel is unchanged. AC timing parameters are also unchanged.

For 16-bit operation, the training sequence consists of ten repetitions of the word 0x0FFF followed by ten repetitions of the word 0xF000. This sequence is easy to detect because the signals on the data bus do not change for 10 cycles. If the 16-bit training sequence is transmitted nibble by nibble across the interface, then it appears at the receiver as the following sequence of nibbles:

F,0,0,0, F,0,0,0, F,0,0,0, F,0,0,0, F,0,0,0,
F,0,0,0, F,0,0,0, F,0,0,0, F,0,0,0, F,0,0,0,
0,F,F,F, 0,F,F,F, 0,F,F,F, 0,F,F,F, 0,F,F,F,
0,F,F,F, 0,F,F,F, 0,F,F,F, 0,F,F,F, 0,F,F,F

The suitability of the above sequence for training a narrow-mode interface has yet to be determined and requires further study. This appendix currently does not make any recommendation regarding the training sequence for narrow mode operation.

7.7 7.7 Appendix G. Hit-Less Bandwidth Reprovisioning.

This appendix does not form an integral part of this specification. It is the recommended implementation for “hit-less” bandwidth reprovisioning on the interface. This appendix describes a mechanism to hitlessly reconfigure an interface in the event that channels need to be added, removed or resized. It is “hit-less” in the sense that active traffic on the interface is unaffected.

The FIFO Status Channel is described in Section 6.2.2. It is a two-bit wide bus that conveys information back to a data source regarding the availability of space within the FIFOs (or other memory devices) that receive data transmitted across the interface. A provisionable *calendar* defines the sequence of these status messages. There are a few other provisionable parameters associated with each channel (for example, MaxBurst1 and MaxBurst2) that must also be provisioned identically on both sides of the interface.

If provisioning for the interface needs to be changed (e.g. to add a channel or service), it may be desirable to avoid any interruption or corruption of service on other channels.

The following modifications can be made to the format of the FIFO Status Channel. Both sides of the interface can have two sets of provisionable parameters: one that is currently in use (active), and one that can be updated and/or changed (shadow). For the purposes of explanation, call them X and Y. The first valid word of the FIFO Status channel following the framing pattern indicates which parameter set to use, and is called the *Calendar Selection Word*. If the first word contains “01,” then the operation of the interface is governed by parameter set X. If the first word contains “10,” then the operation of the interface is governed by parameter set Y. The sender sets the value of this word to correspond to the parameters in use. The receiver looks at this word to determine the meaning of the subsequent bytes. When an update is desired, the shadow parameter sets in both the sender and receiver are updated identically, and then the transmitter starts to use the new set. The receiver automatically switches to the new set when it sees the new *calendar selection word* at the beginning of a FIFO Status Channel message.

7.8 7.8 Appendix H. Error Conditions.

Recommended actions in response to conditions on the interface that differ from that specified in this document are subject to further work (with other standards bodies).

8.1 Summary

This document describes a packet interface for 10 Gb/s aggregate bandwidth applications. It specifies a 16-bit data path, running at a minimum rate of 622 MHz, to support packet and cell transfer over OC-192 SONET/SDH, as well as 10 Gb/s Ethernet applications. The interface supports bidirectional data transfer between two devices, through two different sets of signals that operate independently from the other.

Compatibility Statement

Implementers may choose to support certain optional features beyond those required in this specification; these features are indicated in Appendices E, F and G. Some parts of the specification require support of at least one of two possible alternatives. As noted in Section 6, the FIFO status channel may be implemented in LVTTL or optional LVDS I/O. Support for both modes by a given implementation to improve interoperability is not precluded by this specification. When LVDS I/O is used in either the data path or FIFO status channel, the choice of static or dynamic alignment at the receiver is implementation-specific. The transmitter must meet either the static or dynamic alignment AC timing parameters, but may meet both parameters (Section 6.4) where feasible, for greater interoperability. The supported ranges of various start-up parameters (Section 6.3), where unspecified, are implementation-specific.

8.2 References

8.2.1 Normative references

[1] IEEE Std 1596.3 – 1996, “IEEE Standard for Low-Voltage Differential Signals (LVDS) for Scalable Coherent Interface (SCI)”, approved March 21, 1996.

[12] ANSI/TIA/EIA-644-A-20011995, “Electrical Characteristics of Low Voltage Differential Signaling (LVDS) Interface Circuits”, approved November 15, 1995.

8.2.2 Informative references

8.3 Glossary

ATM	Asynchronous Transfer Mode
FCS	Frame Check Sequence
FIFO	First-In First-Out (queue)
HDLC	High-level Data Link Control
HEC	Header Error Correction
IEEE	Institute of Electrical and Electronics Engineers
I/O	Input / Output
LAN	Local Area Network
LSB	Least Significant Bit
LVDS	Low-Voltage Differential Signalling
LVTTL	Low-Voltage Complementary Metal Oxide Semiconductor
MSB	Most Significant Bit
PHY	Physical Layer Device
POS	Packet-Over-SONET/SDH
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical Network
SYNTH	Clock Synthesis Unit
UDF	User-Defined Field
WAN	Wide Area Network

8.4 Open Issues / current work items

8.5 List of companies belonging to OIF when document was approved

Accelerant Networks	Iolon
Aeluros	Japan Telecom
Agere Systems	JDS Uniphase
Agilent Technologies	Juniper Networks
Agility Communications	KDDI R&D Laboratories
Alcatel	KeyEye Communications
Altera	Kodeos Communications
AMCC	KT Corporation
America Online	Lambda Optical Systems
Analog Devices	Lattice Semiconductor
Analogix Semiconductor	LSI Logic
Anritsu	Lucent
Artisan Components	Lumentis
ASTRI	Marconi Communications
AT&T	MCI
Atrica Inc.	MergeOptics GmbH
Avici Systems	Mindspeed
Azna	Mintera
Big Bear Networks	Mitretek Systems

Bit Blitz Communications	Mitsubishi Electric Corporation
Bookham Technology	Molex
Booz-Allen & Hamilton	Multiplex
Broadcom	Mysticom
Cadence Design Systems	Navtel Communications
Calient Networks	NEC
Caspian Networks	NIST
China Telecom	Nortel Networks
Chunghwa Telecom Labs	NTT Corporation
Ciena Corporation	OpNext
Circadiant Systems	PhotonEx
Cisco Systems	Photuris, Inc.
CIVCOM	Phyworks
CoreOptics	PMC Sierra
Corrigent Systems	Princeton Optronics
Corvis Corporation	Procket Networks
Cypress Semiconductor	Quake Technologies
Data Connection	Quellan
Department of Defense	Qwest Communications
Diablo Technologies	Sandia National Laboratories
ELEMATICS	Santur
Elisa Communications	SBC
Emcore	Scintera Networks
Ericsson	Silicon Access Networks
ETRI	Silicon Laboratories
FCI	Silicon Logic Engineering
Finisar Corporation	StrataLight Communications
Flextronics	Sun Microsystems
Force 10 Networks	Sycamore Networks
Foxconn	Tektronix
France Telecom	Telcordia Technologies
Fujitsu	Telecom Italia Lab
Furukawa America	Tellabs
Galazar Networks	Tellium
Gennum Corporation	Teradyne
Harris Corporation	Texas Instruments
Helix AG	Toshiba Corporation
Hi/fn	TriQuint Semiconductor
Hitachi	T-Systems/ Deutsche Telekom
Ibiden	Turin Networks
IBM Corporation	Tyco Electronics
IDT	US Conec
Industrial Technology Research Institute	Velio Communications
Infineon Technologies	Verizon
Infinera	Vitesse Semiconductor
Innovance Networks	W.L. Gore & Associates
Intel	Winchester Electronics
Intelligent Photonics Control	Xanoptix
Interoute	Xignal Technologies
Intune Technologies, Ltd.	Xilinx